aws

用户指南

AWS DataSync



Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS DataSync: 用户指南

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆、贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

Table of Contents

什么是 AWS DataSync？ 1
使用案例 2
优势 2
其他资源 3
工作原理 4
DataSync 传输架构 4
在本地存储和之间传输 AWS 4
在 AWS 存储服务之间传输 5
在 AWS 存储服务与其他云中的存储系统之间传输 6
概念和术语 6
座席 7
位置 7
Task 7
任务执行 7
如何 DataSync 传输文件、对象和目录 7
如何 DataSync 准备数据传输 8
如何 DataSync 传输您的数据 9
如何 DataSync 验证数据的完整性 9
DataSync 如何处理打开和锁定的文件 10
定期传输选项 10
开始使用 11
注册获取 AWS 账户 11
创建具有管理访问权限的用户 11
使用所需的 IAM 权限 DataSync 12
AWS 托管策略 13
客户托管策略 13
我可以在哪里使用 DataSync？ 13
我该怎么用 DataSync？ 13
要 DataSync 花多少钱？ 13
使用的开源组件 DataSync 13
我是否需要代理？ 14
需要 DataSync 代理的情况 14
不需要 DataSync 代理的情况 14
为您的任务模式选择代理 14

使用多个 DataSync 代理 15
后续步骤 16
代理要求 16
虚拟机管理程序要求 16
代理商对 DataSync 转账的要求 17
AWS 区域 分区的代理要求 18
代理管理要求 18
部署代理 18
在上部署代理 VMware 18
在 KVM 上部署代理 19
在 Microsoft Hyper-V 上部署你的基本模式代理 20
部署 Amazon EC2 代理 21
在上部署基本模式代理 AWS Outposts 26
选择代理的服务端点 26
选择一个公共服务端点 27
选择 FIPS 服务端点 27
选择 VPC 服务端点 27
激活 代理 29
先决条件 30
获取激活密钥 30
激活 代理 32
后续步骤 35
验证代理的网络连接 35
访问代理的本地控制台 35
验证代理与存储系统的连接 36
验证您的代理与 DataSync 服务的连接 36
后续步骤 37
连接您的网络 38
1. 存储系统和代理之间的网络连接 38
2. 您的代理和 DataSync 服务之间的网络连接 38
将您的存储网络连接到 AWS 38
选择服务终端节点 39
3. 服务和 AWS 存储 DataSync 服务之间的网络连接 39
不需要 DataSync 代理时联网 39
DataSync 流量通过网络的方式和位置 39
网络安全 DataSync 40

网络要求 40
IPv6 支持 40
本地、自我管理和其他云存储的网络要求 41
AWS 存储服务的网络要求 45
公共或 FIPS 服务端点的网络要求 46
VPC 或 FIPS VPC 服务端点的网络要求 53
用于数据传输的网络接口 55
用于与代理传输的网络接口 56
无需代理即可传输的网络接口 56
查看您的网络接口 57
使用 Direct Connect 的架构和路由示例 57
将 Direct Connect 与 DataSync VPC 服务终端节点结合使用 58
将 Direct Connect 与 DataSync 公共或 FIPS 服务端点配合使用 61
后续步骤 61
为多个代理配置代理 NICs 61
传输数据 63
我可以在哪里传输数据？ 63
支持同一个传输 AWS 账户 64
支持的跨境传输 AWS 账户 65
支持同一个传输 AWS 区域 66
支持之间的传输 AWS 区域 66
确定您的转账是否需要代 DataSync 理 67
传入或传出本地存储 67
配置使用 NFS 文件服务器传输 68
配置使用 SMB 文件服务器传输 71
配置与 HDFS 集群的传输 80
配置与对象存储系统的传输 85
传输到存储器或从 AWS 存储器中传出 89
使用 Amazon S3 配置传输 89
使用 Amazon EFS 配置传输 116
使用为 Window FSx s 文件服务器配置传输 124
为 Lustr FSx e 配置传输 129
为 OpenZF FSx S 配置传输 130
使用 FSx 为 ONTAP 配置传输 133
传入或传出其他云存储 140
计划传入或传出第三方云存储系统 140

使用 Google Cloud Storage 配置传输 142
配置使用 Microsoft Azure Blob Storage 的传输 150
配置使用 Microsoft Azure Files 的传输 166
配置使用其他云对象存储的传输 168
创建一个传输数据的任务 174
创建 任务 174
任务状态 177
使用多个任务对大型数据集进行分区 177
通过多个任务来分割传输的数据 177
为传输选择任务模式 178
选择要传输的数据 182
验证数据完整性 215
设置带宽限值 217
计划任务 218
标记您的任务 223
启动传输数据的任务 225
开始您的任务 226
任务执行状态 227
了解任务何时排队 228
取消执行任务 228
监控数据传输 229
了解数据传输性能计数器 229
使用 CloudWatch 指标监控数据传输 244
DataSync 的 CloudWatch 指标 245
使用任务报告监控数据传输 247
使用案例 248
仅摘要任务报告 248
标准任务报告 248
示例任务报告 251
限制 254
创建任务报告 254
查看任务报告 264
使用 CloudWatch 日志监控数据传输 264
DataSync 允许将日志上传到 CloudWatch 日志组 264
为您的 DataSync 任务配置日志记录 266
查看 DataSync 任务日志 269

使用登录 CloudTrail 270
处理中的 DataSync 信息 CloudTrail 270
了解 DataSync 日志文件条目 271
使用监控 EventBridge 272
DataSync 转移事件 272
使用手动工具监控 274
使用 DataSync 控制台监控您的转账 274
使用以下方法监控您的转账 AWS CLI 275
使用 watch 实用程序监控传输 277
管理资源 278
管理您的 DataSync 代理 278
测试 DataSync 代理的连接和系统资源 278
更换您的 DataSync 代理 278
清理 DataSync 资源 278
重复使用 DataSync 代理的基础架构 278
管理代理 278
代理软件更新 278
代理状态 279
代理故障排除 279
对代理进行维护 280
访问代理的本地控制台 280
配置代理的 DHCP、DNS 和 IP 设置 281
检查代理的系统资源 285
查看和管理代理的系统时间服务器配置 286
运行与代理维护相关的命令 287
更换代理 288
创建新代理 288
通过新代理更新您的位置 289
后续步骤 293
筛选 DataSync 资源 294
用于筛选的参数 294
按位置筛选 295
按任务筛选 296
清理 DataSync 资源 297
删除代 DataSync 理 297
重复使用 DataSync 代理的基础架构 298

删除 DataSync 地点 298
删除 DataSync 任务 299
安全性 301
数据保护 301
传输中加密 302
静态加密 304
互联网络流量隐私 305
Identity and access management 305
访问权限管理 306
AWS 托管策略 311
客户托管策略 317
使用服务关联角色 320
在创建过程中标记资源 323
防止跨服务混淆代理 325
合规性验证 327
恢复能力 327
基础结构安全性 327
保护存储位置凭证 328
使用通过默认密钥加密的服务托管密钥 328
使用使用自定义 AWS KMS 密钥加密的服务管理密钥 328
使用您管理的密钥 330
配额 333
存储系统、文件和对象限制 333
DataSync 配额 333
请求提高限额 336
问题排查 337
排查代理问题 337
如何连接到 Amazon EC2 代理的本地控制台？ 337
无法检索代理激活密钥错误是什么意思？ 338
我仍然无法使用 VPC 服务端点激活代理 338
如果我的代理处于离线状态，我该怎么办？ 338
我不知道我的代理出了什么问题。有人能帮我吗？ 339
排查位置问题 339
我的任务失败，出现 NFS 权限被拒绝错误 340
我的任务失败，出现 NFS 挂载错误 340
我的任务失败，出现 Amazon EFS 挂载错误 341

NFS 传输未保留文件所有权 341
我的任务无法访问使用 Kerberos 的 SMB 位置 341
我的任务因 input/output 错误而失败 342
错误：FsS3UnableToConnectToEndpoint 343
错误：FsS3HeadBucketFailed 344
任务因 Unable to list Azure Blobs on the volume root 错误而失败 344
错误：FsAzureBlobVolRootListBlobsFailed 344
错误：SrcLocHitAccess 344
错误：SyncTaskErrorLocationNotAdded 345
错误：S3 location creation failed with (InvalidRequestException) when calling the CreateLocationS3 operation 345
S3 源位置的任务失败并显示 HeadObject 或 GetObjectTagging 错误 345
解决任务问题 346
错误：SyncOption 值无效。选项： TransferMode，PreserveDeletedFiles，值：全部，移除。 346
任务执行失败并EniNotFound出现错误 346
任务执行失败，出现无法分配内存错误 346
Input/Output error对于 ONTAP 文件 FSx 系统，任务失败 347
任务失Connection Reset by peer败并显示 ONTAP 文件系统的Host is down消息 FSx 348
任务执行处于启动状态，但似乎什么也没发生 349
任务执行似乎卡在准备状态 349
任务执行在传输完成之前停止 349
从 Google 云存储桶传输时，任务执行失败 350
任务执行的时间戳不匹配 350
任务执行失败并出现 NoMem 错误 350
任务执行失败并出现 FsNfsIdMappingEnabled 错误 350
对象无法传输到 Azure Blob Storage，并出现 user metadata key 错误 351
目标位置存在 /.aws-datasync 文件夹 351
无法使用 SMB 在两个位置之间传输符号链接 351
任务报告错误 352
解决数据验证的问题 352
文件内容不匹配 352
文件的 SMB 元数据不匹配 353
要传输的文件已不在源位置 354
DataSync 无法验证目标数据 355

DataSync 无法读取对象元数据 356
对象的系统定义元数据不匹配 356
了解数据验证时间 358
使用 S3 存储成本疑难解答 DataSync 358
教程 360
跨账户从本地传输到 S3 360
概述 360
先决条件：所需的源账户权限 361
先决条件：所需的目标账户权限 364
第 1 步：在您的源账户中，创建 DataSync 代理 364
步骤 2：在您的源账户中，创建用于访问目标存储桶 DataSync 的 IAM 角色 364
第 3 步：在您的目标账户中，更新 S3 存储桶策略 366
步骤 4：在您的目标账户中， ACLs 对您的 S3 存储桶禁用 368
第 5 步：在您的源账户中，为您的本地存储创建 DataSync 源位置 368
步骤 6：在您的源账户中，为您的 S3 存储桶创建 DataSync 目标位置 368
第 7 步：在您的源账户中，创建并启动 DataSync 任务 370
相关资源 370
跨账户在 S3 存储桶之间传输 371
概述 371
先决条件：所需的源账户权限 373
先决条件：所需的目标账户权限 376
步骤 1：在您的源账户中，创建用于访问目标存储桶 DataSync 的 IAM 角色 376
第 2 步：在您的目标账户中，更新 S3 存储桶策略 378
步骤 3：在您的目标账户中， ACLs 对您的 S3 存储桶禁用 379
第 4 步：在您的来源账户中，创建您的 DataSync 营业地点 380
第 5 步：在您的源账户中，创建并启动 DataSync 任务 381
问题排查 382
相关信息：使用服务器端加密的 S3 存储桶进行跨账户传输 381
执行大规模迁移 384
什么是大规模数据迁移？ 384
大规模数据迁移的关键阶段 384
其他资源 385
第 1 阶段：规划迁移 385
收集要求 385
运行概念验证 390
估算迁移时间表 391

第 2 阶段：实施迁移 393
通过分区加速迁移 393
运行你的 DataSync 任务 395
监控传输 396
DataSync API 397
操作 397
CancelTaskExecution 400
CreateAgent 402
CreateLocationAzureBlob 407
CreateLocationEfs 413
CreateLocationFsxLustre 419
CreateLocationFsxOntap 423
CreateLocationFsxOpenZfs 428
CreateLocationFsxWindows 433
CreateLocationHdfs 439
CreateLocationNfs 447
CreateLocationObjectStorage 452
CreateLocationS3 459
CreateLocationSmb 464
CreateTask 473
DeleteAgent 482
DeleteLocation 484
DeleteTask 486
DescribeAgent 488
DescribeLocationAzureBlob 493
DescribeLocationEfs 498
DescribeLocationFsxLustre 503
DescribeLocationFsxOntap 506
DescribeLocationFsxOpenZfs 511
DescribeLocationFsxWindows 515
DescribeLocationHdfs 520
DescribeLocationNfs 526
DescribeLocationObjectStorage 530
DescribeLocationS3 535
DescribeLocationSmb 540
DescribeTask 547

DescribeTaskExecution 556
ListAgents 576
ListLocations 579
ListTagsForResource 582
ListTaskExecutions 585
ListTasks 588
StartTaskExecution 591
TagResource 598
UntagResource 601
UpdateAgent 604
UpdateLocationAzureBlob 606
UpdateLocationEfs 611
UpdateLocationFsxLustre 615
UpdateLocationFsxOntap 618
UpdateLocationFsxOpenZfs 621
UpdateLocationFsxWindows 624
UpdateLocationHdfs 628
UpdateLocationNfs 634
UpdateLocationObjectStorage 637
UpdateLocationS3 643
UpdateLocationSmb 647
UpdateTask 654
UpdateTaskExecution 660
数据类型 662
AgentListEntry 665
AzureBlobSasConfiguration 667
CmkSecretConfig 668
CustomSecretConfig 670
Ec2Config 672
FilterRule 674
FsxProtocol 675
FsxProtocolNfs 676
FsxProtocolSmb 677
FsxUpdateProtocol 680
FsxUpdateProtocolSmb 681
HdfsNameNode 683

LocationFilter 684
LocationListEntry 686
ManagedSecretConfig 688
ManifestConfig 689
NfsMountOptions 691
OnPremConfig 692
Options 693
Platform 701
PrivateLinkConfig 702
QopConfiguration 704
ReportDestination 705
ReportDestinationS3 706
ReportOverride 708
ReportOverrides 709
ReportResult 711
S3Config 712
S3ManifestConfig 713
SmbMountOptions 715
SourceManifestConfig 717
TagListEntry 718
TaskExecutionFilesFailedDetail 719
TaskExecutionFilesListedDetail 721
TaskExecutionFoldersFailedDetail 722
TaskExecutionFoldersListedDetail 724
TaskExecutionListEntry 725
TaskExecutionResultDetail 727
TaskFilter 730
TaskListEntry 732
TaskReportConfig 734
TaskSchedule 736
TaskScheduleDetails 738
常见错误类型 739
常见参数 741
文档历史记录 743
AWS 术语表 754
..... dcclv

什么是 AWS DataSync？

AWS DataSync 是一种安全、可靠、高速的文件传输服务，可帮助您快速、轻松地将文件或对象数据传输到存储服务、从存储服务传输或在 AWS 存储服务之间传输。

本地存储传输

DataSync 适用于以下本地存储系统：

- 网络文件系统（NFS）
- 服务器消息块 (SMB)
- Hadoop Distributed File Systems (HDFS)
- 对象存储

AWS 存储传输

DataSync 可与以下 AWS 存储服务配合使用：

- Amazon S3
- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的 Amazon

其他云存储传输

DataSync 可与其他云中的以下存储服务配合使用：

- Google Cloud Storage
- Microsoft Azure Blob Storage
- Microsoft Azure Files
- Wasabi Cloud Storage
- DigitalOcean Spaces
- Oracle Cloud Infrastructure Object Storage

- Cloudflare R2 Storage
- Backblaze B2 Cloud Storage
- NAVER Cloud Object Storage
- Alibaba Cloud Object Storage Service
- IBM Cloud Object Storage
- Seagate Lyve Cloud

使用案例

以下是以下方面的一些主要用例 DataSync：

- 迁移数据-通过网络将活动数据集快速传输到 AWS 存储服务中。 DataSync 包括自动加密和数据完整性验证，可帮助确保您的数据安全、完好无损且随时可用。
- 存档冷数据 — 将存储在本地存储中的冷数据直接移动到持久且安全的长期存储类别，例如 S3 Glacier Flexible Retrieval 或 S3 Glacier Deep Archive。这样做可以释放本地存储容量并帮助您关闭传统系统。
- 复制数据 — 将数据复制到大多数 Amazon S3 存储类别，根据您的需求选择最具成本效益的存储类别。您也可以向 Amazon EFS 或 Amazon FSx 发送备用文件系统的数据。
- 传输数据以便在云端及时进行处理 — 将数据传入或传出 AWS 以进行处理。这种方法可加快许多行业内的关键混合云工作流。其中包括生命科学行业中的机器学习、媒体和娱乐行业中的视频制作、金融行业中的大数据分析，以及石油和天然气行业中的地震研究。

优势

通过使用 DataSync，您可以获得以下好处：

- 自动移动数据 — DataSync 使通过网络在存储系统和服务之间传输数据变得更加容易。 DataSync 可自动管理数据传输过程以及实现高性能和安全数据传输所需的基础架构。
- 安全地传输数据- DataSync 提供 end-to-end安全性，包括加密和数据完整性验证，以帮助确保您的数据安全、完好无损且随时可用。 DataSync 通过内置 AWS 安全机制（例如 AWS Identity and Access Management (IAM) 角色）访问您的 AWS 存储。它还支持虚拟私有云（VPC）端点，使您可以选择在不通过公共互联网的情况下传输数据，并进一步提高在线复制数据的安全性。
- 更快地移动数据- DataSync 使用专门构建的网络协议和并行多线程架构来加快传输速度。这种方法可加快迁移、用于分析和机器学习的重复数据处理工作流以及数据保护流程。

其他资源

我们建议您阅读以下部分：

- DataSync 资源-包括博客、视频和其他培训材料
- AWS re: pos t — 查看有关的最新讨论 DataSync
- AWS DataSync 定价

如何 AWS DataSync 运作

了解与 AWS DataSync 传输相关的关键概念和术语，包括如何从本地和云端传输数据。

DataSync 传输架构

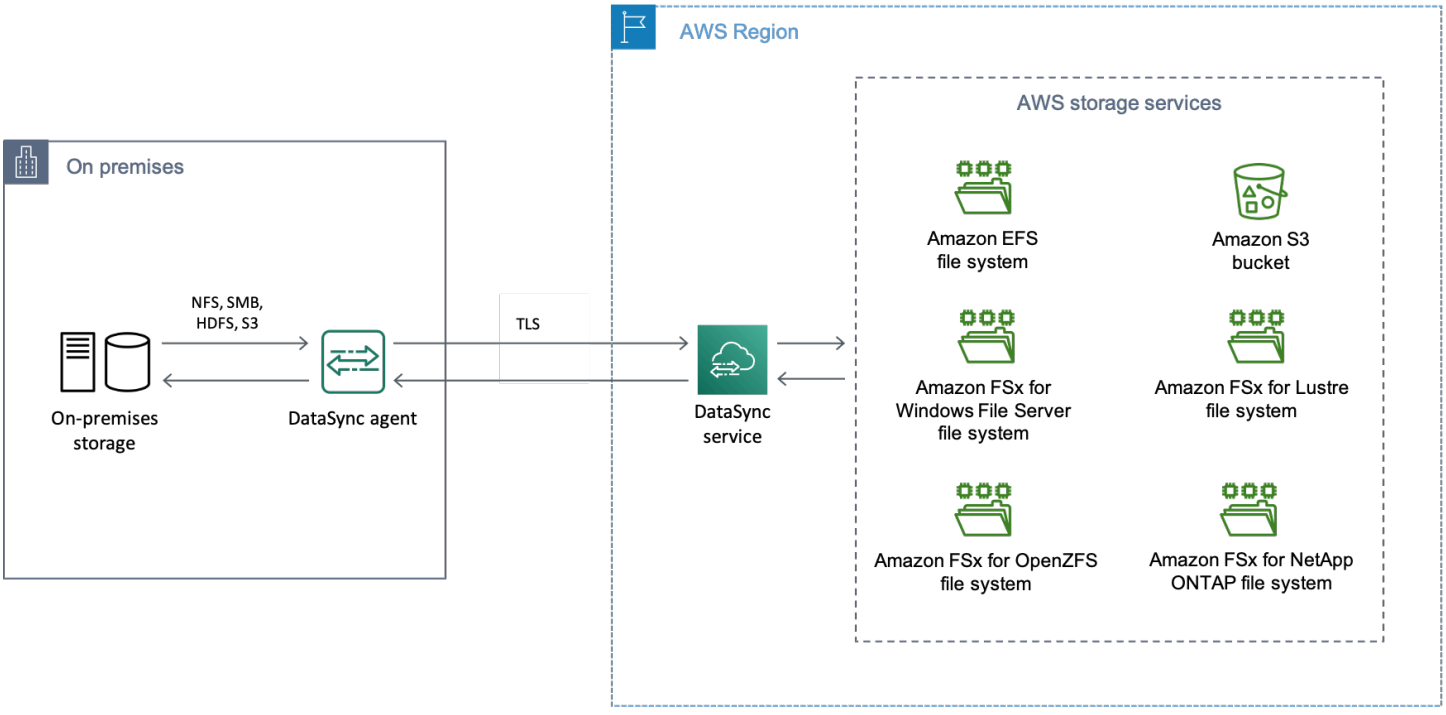
下图显示了 DataSync 通常如何以及在何处传输存储数据。有关 DataSync 支持的存储系统和服务的完整列表，请参见 我可以在哪里传输我的数据 AWS DataSync？

主题

- 在本地存储和之间传输 AWS
- 在 AWS 存储服务之间传输
- 在 AWS 存储服务与其他云中的存储系统之间传输

在本地存储和之间传输 AWS

下图简要概述了在自行管理的本地存储系统和 AWS 服务之间 DataSync 传输文件。



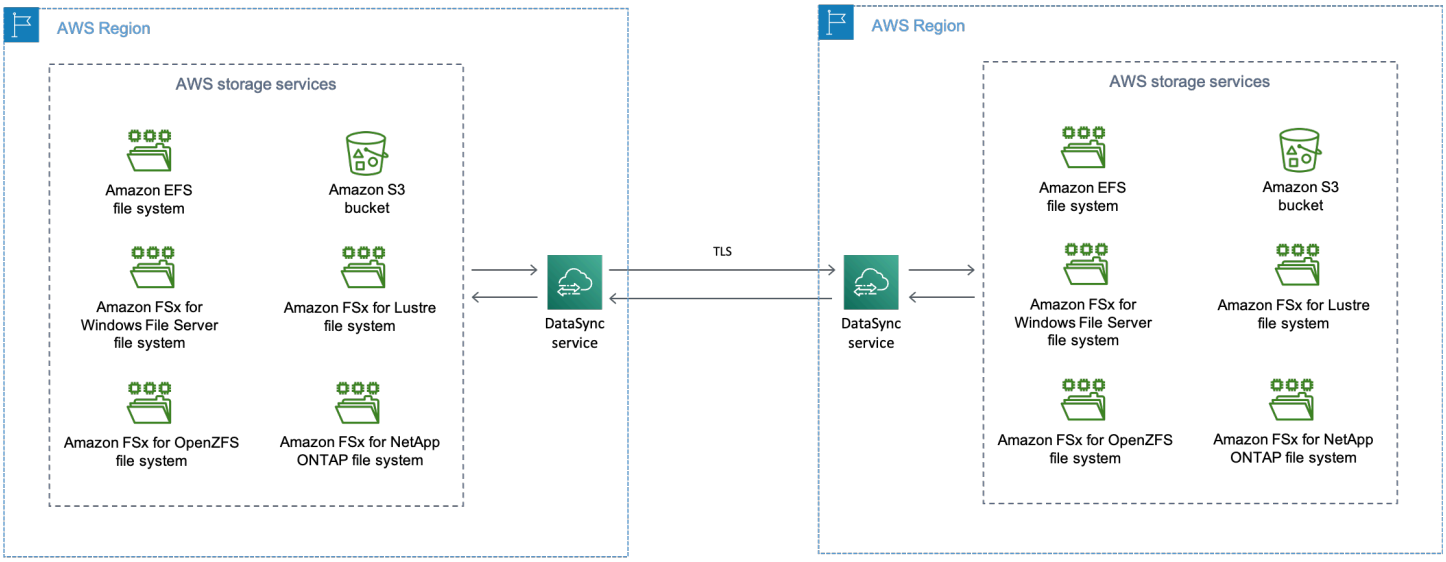
该图说明了一个常见的 DataSync 用例：

- 从本地存储系统复制数据的 DataSync 代理。

- 数据传入 AWS，使用传输层安全 (TLS) 进行加密。
- DataSync 将数据复制到支持的 AWS 存储服务。

在 AWS 存储服务之间传输

下图显示了 AWS 服务 在同一个服务器之间 DataSync 传输文件的高级概述 AWS 账户。



该图说明了一个常见的 DataSync 用例：

- DataSync 从支持的 AWS 存储服务中复制数据。
- 数据传输 AWS 区域，使用 TLS 加密。
- DataSync 将数据复制到支持的 AWS 存储服务。

在同一帐户中的 AWS 存储服务之间传输时（无论是在同 AWS 区域 一个分区 AWS 区域 中还是在同一分区中），不需要代理。您的数据仍保留在 AWS 网络中，不会通过公共互联网。

> **Important**
>
> 您需要为两者之间传输的数据付费 AWS 区域。此次传输按从源区域向外传输到目标区域的数据计费。有关更多信息，请参阅 数据传输定价。

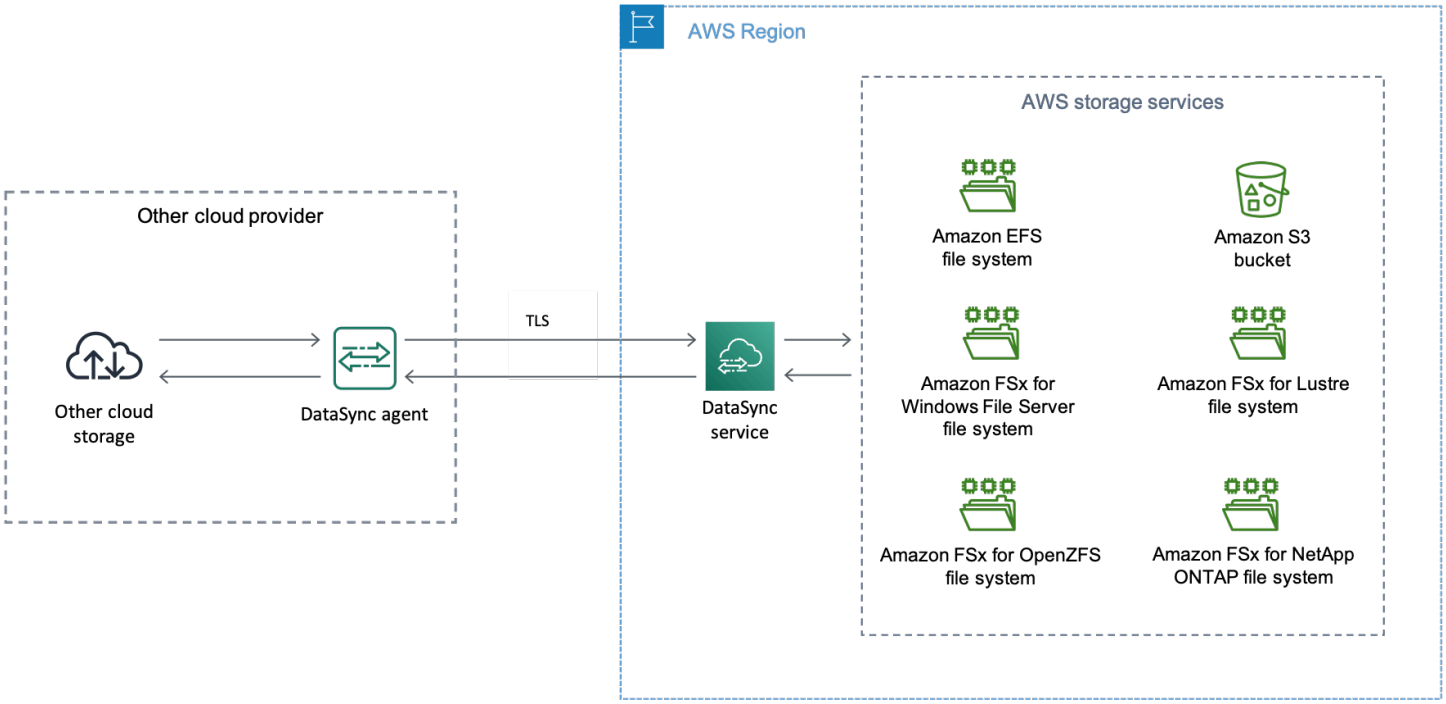
在 AWS 存储服务与其他云中的存储系统之间传输

使用 DataSync，您可以在其他云存储系统和之间传输数据 AWS 服务。在这种情况下，云存储系统可包括：

- 自我管理存储系统，例如 AWS中虚拟私有云（VPC）的 NFS 文件服务器。
- 由其他云提供程序托管的存储系统或服务。有关更多信息，请参阅 使用其他云存储进行传入或传出其他云存储 AWS DataSync。

DataSync 无论是否使用代理，都可以将数据复制到其他云或从其他云中复制数据。有关何时使用代理的更多信息，请参阅我需要 AWS DataSync 代理吗？。

下图显示了在 AWS 存储服务与其他云提供商之间 DataSync传输数据的高级概述。



概念和术语

熟悉 DataSync 传输功能。

主题

- 座席
- 位置
- Task

- 任务执行

座席

代理是一种虚拟机 (VM) 设备， DataSync 用于在传输过程中对存储进行读取和写入。 DataSync 提供两种类型的代理，一种处理基本模式任务，另一种处理增强模式任务。有关为您的用例选择代理的更多信息，请参阅为您的任务模式选择代理。

你可以在基于 Linux 内核的虚拟机 (KVM) VMware ESXi、Nutanix AHV（使用 KVM 代理映像）或 Microsoft Hyper-V 虚拟机管理程序上部署代理。要在中的虚拟私有云 (VPC) 中存储 AWS，您可以将代理部署为 Amazon EC2 实例。

要了解其用法，请参阅 我需要 AWS DataSync 经纪人吗？

位置

位置 描述了您要从何处复制数据或将数据复制到何处。每次 DataSync 传输（也称为任务）都有源位置和目标位置。有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。

Task

任务描述了 DataSync 转移。它标识源位置和目标位置，以及有关如何在这些位置之间复制数据的详细信息。您还可以指定此任务如何处理元数据、已删除的文件和权限。

任务执行

任务执行是 DataSync 传输任务的单独运行。任务执行涉及多阶段。有关更多信息，请参阅 任务执行状态。

如何 DataSync 传输文件、对象和目录

在任务执行期间， DataSync 准备、传输和验证您的数据。这些操作的 DataSync 执行方式取决于您如何配置 DataSync 任务选项，例如任务模式。基本模式任务按顺序准备、传输和验证数据，而增强模式任务则并行执行这些任务。

主题

- 如何 DataSync 准备数据传输
- 如何 DataSync 传输您的数据

- 如何 DataSync 验证数据的完整性
- DataSync 如何处理打开和锁定的文件
- 定期传输选项

如何 DataSync 准备数据传输

DataSync 默认情况下，通过检查来源和目标位置来确定要传输的内容，从而为传输做准备。具体做法是扫描两个位置的内容和元数据，找出两者之间的差异。

> **Note**
>
> 如果将任务配置为传输所有数据，则无需准备。开始任务时，无需比较位置即可 DataSync 立即将所有内容从源头转移到目的地。

如何 DataSync 准备传输还取决于您的任务模式：

增强模式准备	基本模式准备
DataSync 在源位置找到对象时对其进行准备。在整个任务执行过程中，准备过程一直在进行，直至源不再列出任何对象。	准备过程可能只需要几分钟，也可能需要几个小时甚至更长时间，具体取决于两个位置的文件、对象或目录的数量，以及存储的性能。
与基本模式不同，每次执行任务时 DataSync 可以准备几乎无限数量的对象。	在您的来源和目的地中 DataSync 盘点的项目计入您的任务配额。这些配额不是基于每次任务执行期间 DataSync 传输的项目数量。

DataSync 在准备过程中可能会跳过一些文件、对象和目录。其原因可能取决于多个因素，例如您配置任务和存储系统权限的方式。下面是一些示例：

- 一个文件存在于源位置和目标位置。自上次执行任务以来，源位置的文件未被修改。由于您只传输已更改的数据，因此下次运行任务时 DataSync 不会传输该文件。
- 两个位置都存在的某个对象在源位置发生变化。运行任务时，会 DataSync 跳过目标中的此对象，因为您的任务不会覆盖目标中的数据。
- DataSync 跳过源位置中使用存档存储类且未恢复的对象。您必须恢复已存档的对象 DataSync 才能读取它。

- DataSync 跳过源位置中的文件、对象或目录，因为它无法读取它。如果发生这种情况但不是预料之中的，请检查存储空间的访问权限，并确保该权限 DataSync 可以读取跳过的内容。

如何 DataSync 传输您的数据

DataSync 根据您的任务选项将您的数据（包括元数据）从源复制到目标。例如，您可以指定要复制哪些元数据，排除某些文件，限制带宽 DataSync 使用量等。

如何 DataSync 传输数据还取决于您的任务模式：

增强模式传输	基本模式传输
DataSync 准备好每个对象后立即进行传输。	DataSync 准备好所有数据后，传输就开始了。

DataSync在传输过程中可能会跳过一些项目。如果将任务配置为传输所有数据，则源位置中使用存档存储类别且未恢复的对象可能会出现这种情况。

如何 DataSync 验证数据的完整性

DataSync 在传输过程中，始终对您的数据进行完整性检查。在传输结束时， DataSync 还可以仅对两个位置的传输数据或整个数据集执行额外检查。有关更多信息，请参阅 配置如何 AWS DataSync 验证数据完整性。

检查数据完整性时， DataSync 计算和比较您所在位置的文件、对象或目录的校验和和和元数据。如果 DataSync 发现地点之间存在差异，则验证失败并显示错误。例如，您可能会看到 `Checksum failure`、`Metadata failure`、`Files were added` 或 `Files were removed` 错误。

验证的工作方式取决于您的任务模式，以及您是否配置 DataSync为在传输结束时验证数据完整性。

增强模式验证	基本模式验证
DataSync 在将每个对象转移到目的地时对其进行验证。 在增强模式下，仅 DataSync 验证传输的数据。	在传输结束时，请 DataSync 验证数据的完整性。 根据您配置数据验证的方式，处理大型数据集时，这可能会花费大量时间。

DataSync 如何处理打开和锁定的文件

在尝试传输打开（使用中）或锁定的文件时，注意以下几点：

- 通常， DataSync 可以不受限制地传输打开的文件。
- 如果文件在传输过程中处于打开状态并正在写入，则 DataSync 可以在传输任务的验证阶段检测到这种不一致性。要获取该文件的最新版本，您必须再次运行该任务。
- 如果文件被锁定而服务器 DataSync 阻止打开该文件，则会在传输过程中 DataSync 跳过该文件并记录错误。
- DataSync 无法锁定或解锁文件。

定期传输选项

除了一次性传输外， DataSync 还可以定期传输数据。以下是适用于这些情况的一些选项：

- 安排任务的执行时间。
- 只传输自上次任务执行以来发生更改的数据。
- 删除存在于目标位置，但不再存在于源位置的数据。

入门 AWS DataSync

在开始使用之前 AWS DataSync， AWS 账户 如果没有，则需要先注册一个。我们还建议您了解 DataSync 可以在何处使用以及传输数据可能要花多少钱。

注册获取 AWS 账户

如果您没有 AWS 账户，请完成以下步骤来创建一个。

要注册 AWS 账户

1. 打开https://portal.aws.amazon.com/billing/注册。
2. 按照屏幕上的说明操作。

   在注册时，将接到电话或收到短信，要求使用电话键盘输入一个验证码。

   当您注册时 AWS 账户，就会创建AWS 账户根用户一个。根用户有权访问该账户中的所有 AWS 服务 和资源。作为最佳安全实践，请为用户分配管理访问权限，并且只使用根用户来执行需要根用户访问权限的任务。

AWS 注册过程完成后会向您发送一封确认电子邮件。您可以随时前往 https://aws.amazon.com/并选择“我的账户”，查看您当前的账户活动并管理您的账户。

创建具有管理访问权限的用户

注册后，请保护您的安全 AWS 账户 AWS 账户根用户 AWS IAM Identity Center，启用并创建管理用户，这样您就不会使用 root 用户执行日常任务。

保护你的 AWS 账户根用户

1. 选择 Root 用户并输入您的 AWS 账户 电子邮件地址，以账户所有者的身份登录。AWS 管理控制台在下一页上，输入您的密码。

   要获取使用根用户登录方面的帮助，请参阅《AWS 登录 用户指南》中的 Signing in as the root user。
2. 为您的根用户启用多重身份验证（MFA）。

   有关说明，请参阅 I A M 用户指南中的为 AWS 账户 根用户启用虚拟 MFA 设备（控制台）。

创建具有管理访问权限的用户

1. 启用 IAM Identity Center。

   有关说明，请参阅《AWS IAM Identity Center 用户指南》中的启用 AWS IAM Identity Center。
2. 在 IAM Identity Center 中，为用户授予管理访问权限。

   有关使用 IAM Identity Center 目录 作为身份源的教程，请参阅《用户指南》 IAM Identity Center 目录中的使用默认设置配置AWS IAM Identity Center 用户访问权限。

以具有管理访问权限的用户身份登录

- 要使用您的 IAM Identity Center 用户身份登录，请使用您在创建 IAM Identity Center 用户时发送到您的电子邮件地址的登录 URL。

  有关使用 IAM Identity Center 用户登录的帮助，请参阅AWS 登录 用户指南中的登录 AWS 访问门户。

将访问权限分配给其他用户

1. 在 IAM Identity Center 中，创建一个权限集，该权限集遵循应用最低权限的最佳做法。

   有关说明，请参阅《AWS IAM Identity Center 用户指南》中的 Create a permission set。
2. 将用户分配到一个组，然后为该组分配单点登录访问权限。

   有关说明，请参阅《AWS IAM Identity Center 用户指南》中的 Add groups。

使用所需的 IAM 权限 DataSync

DataSync 可以将您的数据传输到或传出亚马逊 S3 存储桶、Amazon EFS 文件系统或亚马逊 FSx 文件系统。要将数据传送到您想要的地方，您需要为您的身份授予正确的 IAM 权限。例如，与之配合使用的 IAM 角色 DataSync 需要获得使用向 S3 存储桶传输数据所需的 Amazon S3 操作的权限。

您可以使用自己提供的 IAM 策略 AWS 或通过创建自己的策略来授予这些权限。

目录

- AWS 托管策略
- 客户托管策略

AWS 托管策略

AWS 为常见 DataSync 用例提供了以下托管策略：

- `AWSDataSyncReadOnlyAccess`— 提供对的只读访问权限 DataSync。
- `AWSDataSyncFullAccess`— 提供对其依赖项的完全访问权限 DataSync和最低限度的访问权限。

有关更多信息，请参阅 AWS 的托管策略 AWS DataSync。

客户托管策略

您可以创建用于的自定义 IAM 策略 DataSync。有关更多信息，请参阅 IAM 客户托管的政策 AWS DataSync。

我可以在哪里使用 DataSync？

有关 DataSync 支持的 AWS 区域 和终端节点的列表，请参阅中的AWS DataSync 终端节点和配额AWS 一般参考。

我该怎么用 DataSync？

有几种使用方法 DataSync：

- DataSync 控制台，它是 AWS 管理控制台.
- DataSync API 或，AWS CLI以编程方式进行配置和管理 DataSync。
- AWS CloudFormation或者 Terraform 来配置你的资源。 DataSync
- AWS SDKs来构建使用 DataSync. 的应用程序。

要 DataSync 花多少钱？

要使用您计划传输的数据量创建自定义估算值，请参阅DataSync 定价。

使用的开源组件 DataSync

要查看使用的开源组件 DataSync，请下载以下链接：

- datasync-open-source-components.zip

我需要 AWS DataSync 经纪人吗？

要使用 AWS DataSync，您可能需要代理。代理是一种虚拟机（VM）设备，您可以将其部署在存储环境中，用于数据传输。

是否需要代理取决于多个因素，包括您要传输的存储类型或从中传输的存储类型、是否要传输以及 AWS 区域 要在 AWS 账户哪些存储之间进行传输。在进一步阅读之前，请检查是否 DataSync 支持您感兴趣的转账。

确定 DataSync 支持您的转移场景后，请查看以下信息，以帮助您了解是否需要代理。

需要 DataSync 代理的情况

大多数需要 DataSync 代理的情况都涉及由您或其他云提供商管理的存储。

- 在 AWS 存储服务和本地存储之间传输
- 在 Amazon EFS 或 Amazon 之间传输 FSx 并存储在其他云中
- 在某些 AWS 存储服务之间传输 AWS 账户（当两个存储服务都不是 Amazon S3 时）
- AWS GovCloud (US) Region 在商业广告与来源 AWS 区域 和目的地均为 Amazon EFS 或 Amazon 的广告之间转移 FSx

不需要 DataSync 代理的情况

无论是在同一 AWS 区域 传输，还是跨区域传输，都存在不需要代理的情况。

- 在同一个 AWS 存储服务之间传输 AWS 账户
- 在 Amazon S3 和不同的 AWS 存储服务之间传输 AWS 账户
- 在 Amazon S3 与其他云的对象存储之间传输
- AWS GovCloud (US) 在商业广告 AWS 区域 和来源或目的地均为 Amazon S3 的广告之间转移

为您的任务模式选择代理

DataSync 任务在 “基本” 模式或 “增强” 模式下运行。基本模式任务需要基本模式代理。增强模式任务需要增强模式代理。

基本模式支持在向以下位置复制或从以下位置复制时使用代理：

- NFS
- SMB
- HDFS
- 对象存储（包括其他云）
- 天蓝色斑点

增强模式支持使用代理向以下位置的 Amazon S3 进行传送或传出：

- NFS
- SMB

有关更多信息，请参阅 为数据传输选择任务模式。

使用多个 DataSync 代理

虽然大多数传输只需要一个代理，但对于包含数百万个文件或对象的大型数据集，使用多个代理可以加快传输速度。在这种情况下，我们建议并行运行传输任务，每个任务使用一个代理。这种方法将传输工作负载分散到多个任务，每个任务都使用自己的代理。它还有助于缩短 DataSync 准备和传输数据所需的时间。有关更多信息，请参阅 使用多个任务对大型数据集进行分区。

另一种方案（尤其是在有数百万个小文件的情况下）是在一个传输位置使用多个代理。例如，您最多可以将四个代理连接到本地网络文件系统 (NFS) 文件服务。尽管准备转移所需的 DataSync 时间不会改变，但此选项可能会加快您的传输速度。

无论采用哪种方法，都要注意，这可能会增加存储 I/O 操作并影响网络带宽。有关使用多个代理进行 DataSync 传输的更多信息，请参阅AWS 存储博客。

如果您正在考虑使用多个代理，请记住以下几点：

- 一个地点最多可以分配四个基本模式代理和最多四个增强模式代理。使用该位置的任务将仅使用与配置的任务模式相对应的代理。
- 在同一个位置使用多个代理并不能提供高可用性。所有与某个位置关联的代理都必须处于联机状态，才能开始传输任务。如果其中一个代理处于脱机状态，您就无法运行任务。
- 如果您使用虚拟私有云 (VPC) 服务端点与 DataSync 服务通信，则所有代理都必须使用相同的终端节点和子网。

后续步骤

- 如果您需要代理，请查看代理要求，以了解对您的存储环境有意义的要素。
- 如果您的传输不需要代理，则可以开始配置传输。

对 AWS DataSync 代理人的要求

在存储环境中部署 AWS DataSync 代理之前，请务必了解代理虚拟机管理程序和资源要求。

虚拟机管理程序要求

DataSync 可以在支持的虚拟机管理程序上部署代理，以便于数据传输。

> **Note**
>
> 仅支持增强模式代理程序 VMware ESXi、KVM、Nutanix AHV 和。EC2

你可以在以下虚拟机管理程序上运行 DataSync 代理：

- VMware ESXi （版本 7.0 或 8.0）： VMware ESXi 可在博通网站上找到。您还需要一个 VMware vSphere 客户端才能连接到主机。
- 基于 Linux 内核的虚拟机 (KVM)：一种免费的开源虚拟化技术。KVM 包含在 Linux 版本 2.6.20 及更高版本中。 DataSync 已在 Centos/RHEL 7 和 8、Ubuntu 16.04 LTS 和 Ubuntu 18.04 LTS 发行版中进行了测试和支持。其他现代 Linux 发行版可能有效，但不能保证功能或性能。您必须在 KVM 主机上启用硬件加速虚拟化才能部署 DataSync 代理。

  如果您已经启动并运行了 KVM 环境并且您已经熟悉 KVM 的工作原理，我们建议使用此选项。

  不支持在 Amazon EC2 上运行 KVM，也不能用于 DataSync代理。
- 微软 Hyper-V（版本 2012 R2、2016 或 2019）：仅限基本模式代理。对于此设置，您需要 Microsoft Windows 客户端计算机上的 Microsoft Hyper-V Manager 才能连接到主机。

  DataSync 代理是第 1 代虚拟机 (VM)。有关第 1 代和第 2 代之间区别的更多信息 VMs，请参阅我应该在 Hyper-V 中创建第 1 代还是第 2 代虚拟机？
- 亚马逊 EC2： DataSync 提供包含该映像的亚马逊系统映 DataSync 像 (AMI)。有关建议的实例类型，请参阅 Amazon EC2 实例要求。

代理商对 DataSync 转账的要求

对于 DataSync 转移，您的代理必须满足以下资源要求。

> **Important**
>
> 请记住，处理多达 2000 万个文件、对象或目录的基本模式代理要求是一般指导原则。由于其他因素，例如您拥有的目录数量和对象元数据大小，您的代理可能需要更多资源。例如，Amazon EC2 代理的 m5.2xlarge 实例可能仍不足以传输少于 2000 万文件。
> 增强模式代理没有文件配额。

目录

- 虚拟机要求
- Amazon EC2 实例要求

虚拟机要求

部署不在 Amazon EC2 实例上的 DataSync 代理时，代理虚拟机需要以下资源，具体取决于您使用的是基本模式代理还是增强模式代理：

资源	基本模式	增强模式
虚拟处理器	分配给 VM 的四个虚拟处理器	为虚拟机分配了八个虚拟处理器
磁盘空间	80 GB 磁盘空间用于安装虚拟机映像和系统数据	80 GB 磁盘空间用于安装虚拟机映像和系统数据
RAM	分配给虚拟机的 32 GB 内存，用于执行任务，处理多达 2,000 万个文件、对象或目录 分配给 VM 的 64 GB 内存，用于执行处理超过 2,000 万个文件、对象或目录的任务	分配给 VM 的 32 GB RAM

Amazon EC2 实例要求

在 Amazon EC2 实例上部署 DataSync 代理时，实例大小必须至少为 2xlarge。我们建议使用以下实例大小之一，具体取决于您使用的是基本模式代理还是增强模式代理：

基本模式代理	增强模式代理
要执行处理多达 2000 万个文件、对象或目录的任务，请使用 m5. 2xlarge。 要执行处理超过 2,000 万个文件、对象或目录的任务，请使用 m5. 4xlarge。	无论数据集中有多少文件、对象或目录，都使用 m6a.2xl arge。

AWS 区域 分区的代理要求

DataSync 代理映像与特定AWS 区域 分区相关联。例如，默认情况下，您无法在广告中下载代理 AWS 区域 然后在广告中将其激活 AWS GovCloud (US) Region。

代理管理要求

激活 DataSync 代理后，即可为您 AWS 管理代理。有关更多信息，请参阅 管理您的 AWS DataSync 代理。

部署您的 AWS DataSync 代理

创建 AWS DataSync 代理时，第一步是在存储环境中部署代理。你可以在 VMware ESXi、基于 Linux 内核的虚拟机 (KVM)、Nutanix AHV（使用 KVM 镜像）和微软 Hyper-V 虚拟机管理程序上部署代理作为虚拟机 (VM)。您还可以在 AWS的虚拟私有云（VPC）中将代理部署为 Amazon EC2 实例。

> **Tip**
>
> 在开始之前，请确认是否需要代 DataSync 理。

在上部署代理 VMware

您可以从 DataSync 控制台下载代理并将其部署到您的 VMware环境中。

开始之前：请确保您的存储环境可以支持 DataSync 代理。有关更多信息，请参阅 虚拟机要求。

要在上部署代理 VMware

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择 代理，然后选择 创建代理。
3. 对于 “虚拟机管理程序”，选择 VMWare ESXi，然后选择 “下载图像”。

   - 增强模式代理下载为`.ova`图像文件。
   - 基本模式代理下载到包含`.ova`图像`.zip`文件的文件中
4. 要最大限度地减少网络延迟，请将代理部署在尽可能靠近 DataSync 需要访问的存储系统附近（如果可能，则部署在同一个本地网络）。有关更多信息，请参阅 本地、自我管理和其他云存储的网络要求。

   如果需要，请参阅虚拟机管理程序的文档，了解如何在主机中部署`.ova`文件。 VMware
5. 打开虚拟机监视器电源，登录代理 VM 并获取代理的 IP 地址。您需要此 IP 地址来激活代理。

   代理 VM 的默认凭证为登录名 **`admin`** 和密码 **`password`**。如果需要，请通过虚拟机的本地控制台更改密码。

下一步为 AWS DataSync 代理选择服务端点：

在 KVM 上部署代理

您可以从 DataSync 控制台下载代理并将其部署到您的 KVM 环境中。

开始之前：请确保您的存储环境可以支持 DataSync 代理。有关更多信息，请参阅 虚拟机要求。

要在 KVM 上部署代理

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择 代理，然后选择 创建代理。
3. 对于虚拟机监控器，请选择基于内核的虚拟机 (KVM)，然后选择下载映像。

   - 增强模式代理下载为`.qcow2`图像文件。
   - 基本模式代理下载到包含`.qcow2`图像`.zip`文件的文件中
4. 要最大限度地减少网络延迟，请将代理部署在尽可能靠近 DataSync 需要访问的存储系统附近（如果可能，则部署在同一个本地网络）。有关更多信息，请参阅 本地、自我管理和其他云存储的网络要求。

5. 运行以下命令安装 .qcow2 映像。

```
virt-install \
    --name "datasync" \
    --description "DataSync agent" \
    --os-type=generic \
    --ram=32768 \
    --vcpus=4 \
    --disk path=datasync-yyyymmdd-x86_64.qcow2,bus=virtio,size=80 \
    --network default,model=virtio \
    --graphics none \
    --virt-type kvm \
    --import
```

   有关如何管理此虚拟机和 KVM 主机的信息，请参阅虚拟机监控器文档。

6. 打开虚拟机监视器电源，登录 VM 并获取代理的 IP 地址。您需要此 IP 地址来激活代理。

   代理 VM 的默认凭证为登录名 **admin** 和密码 **password**。如果需要，请通过虚拟机的本地控制台更改密码。

下一步为 AWS DataSync 代理选择服务端点：

在 Microsoft Hyper-V 上部署你的基本模式代理

你可以从 DataSync 控制台下载基本模式代理并将其部署到你的 Microsoft Hyper-V 环境中。

开始之前：请确保您的存储环境可以支持 DataSync 代理。有关更多信息，请参阅 虚拟机要求。

在 Hyper-V 上部署基本模式代理

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择 代理，然后选择 创建代理。
3. 对于虚拟机监控器，请选择 Microsoft Hyper-V，然后选择下载映像。

   代理下载到包含 .vhdx 图像文件的 .zip 文件中。

4. 要最大限度地减少网络延迟，请将代理部署在尽可能靠近 DataSync 需要访问的存储系统附近（如果可能，则部署在同一个本地网络）。有关更多信息，请参阅 本地、自我管理和其他云存储的网络要求。

   如果需要，请参阅虚拟机监控器文档，了解如何在 Hyper-V 主机部署 .vhdx 文件。

> **Warning**
>
> 如果您在使用 Broadcom 网络适配器的 Hyper-V 主机上启用虚拟机队列 (VMQ)，可能会发现网络性能不佳。有关工作区的更多信息，请参阅 Microsoft 文档。

5. 打开虚拟机监视器电源，登录 VM 并获取代理的 IP 地址。您需要此 IP 地址来激活代理。

   代理 VM 的默认凭证为登录名 **admin** 和密码 **password**。如果需要，请通过虚拟机的本地控制台更改密码。

下一步为 AWS DataSync 代理选择服务端点：

部署 Amazon EC2 代理

在以下两者之间传输数据时，您可以将 DataSync 代理部署为 Amazon EC2 实例：

- 自我管理的云存储系统（例如，中的 NFS 文件服务器 AWS）和 AWS 存储服务。
- 云存储提供商（例如Microsoft Azure Blob Storage或 Google Cloud Storage）和使用 “基本” 模式的 AWS 存储服务。
- 商业版中的 S3 存储桶 AWS 区域 和广告中的 S3 存储桶 AWS GovCloud (US) Region。
- Amazon AWS Outposts S3 已开 AWS 启，存储服务使用基本模式。

> **Warning**
>
> 我们不建议使用带有本地存储的 Amazon EC2 代理，因为网络延迟会增加。相反，应将代理作为 VMware KVM 或 Hyper-V 虚拟机部署在数据中心中，尽可能靠近本地存储。

部署 EC2 代理

为您选择代理 AMI AWS 区域

1. 打开终端并复制以下 AWS CLI 命令以获取要部署 DataSync Amazon EC2 代理的区域的最新亚马逊系统映像 (AMI) ID。

   基本模式代理

```
aws ssm get-parameter --name /aws/service/datasync/ami --region your-region
```

增强模式代理

```
aws ssm get-parameter --name /aws/service/datasync/ami/v3 --region your-region
```

2. 运行 命令。在输出中，记下带有 DataSync AMI ID 的"Value"属性。

Example命令和输出示例

```
aws ssm get-parameter --name /aws/service/datasync/ami --region us-east-1


{
    "Parameter": {
        "Name": "/aws/service/datasync/ami",
        "Type": "String",
        "Value": "ami-1234567890abcdef0",
        "Version": 6,
        "LastModifiedDate": 1569946277.996,
        "ARN": "arn:aws:ssm:us-east-1::parameter/aws/service/datasync/ami"
    }
}
```

部署 Amazon EC2 代理

> **Tip**
>
> 为避免因跨可用区传输而产生费用，部署代理时应确保其不需要在可用区之间传输网络流量。（要详细了解所有人的数据传输价格 AWS 区域，请参阅 Amazon EC2 数据传输定价。）例如，在自我管理的云存储系统所在的可用区部署代理。

1. 复制以下 URL：

```
https://console.aws.amazon.com/ec2/v2/home?region=agent-
region#LaunchInstanceWizard:ami=ami-id
```

- 将 *agent-region* 替换为您要在其中部署代理的区域。
- *ami-id*替换为您获得的 DataSync AMI ID。

2. 将该 URL 粘贴到浏览器中。

   屏幕上 AWS 管理控制台 显示了 Amazon EC2 实例启动页面。

3. 对于实例类型，请选择一个推荐的 Amazon EC2 实例 DataSync。
4. 对于密钥对，选择现有密钥对或新建一个密钥对。
5. 对于 网络设置，选择 编辑，然后执行以下操作：

   a. 对于 VPC，请选择要在其中部署代理的 VPC。

   b. 对于自动分配公有 IP，请选择您是否希望代理能够从公共互联网访问。

      您稍后使用实例的公有或私有 IP 地址来激活代理。

   c. 对于防火墙（安全组），请创建或选择一个执行以下操作的安全组：

      - 如果需要，允许通过端口 80（HTTP）向 Amazon EC2 实例传入的入站流量。获取代理激活密钥的某些选项需要此连接。
      - 在 Amazon EC2 实例与您要向其传入数据或从中传出数据的存储系统之间，允许入站和出站流量。有关更多信息，请参阅 本地、自我管理和其他云存储的网络要求。

      > **Note**
      >
      > 根据代理使用的服务端点类型，还需要配置其他端口。

6. （推荐）要提高从基于云的文件系统传输时的性能，请展开高级详细信息，选择您的存储所在的置放群组值。
7. 选择启动实例来启动您的 Amazon EC2 实例。
8. 一旦实例的状态为正在运行，选择该实例。
9. 如果将实例配置为可从公共互联网访问，请记下该实例的公有 IP 地址。如果没有，请记下私有 IP 地址。

   激活代理时需要此 IP 地址。

示例：将 EC2 代理部署在 AWS 区域

如果您在 AWS 区域中部署 DataSync 代理，以下指南可以帮助解决常见场景问题。

主题

- 部署基本模式代理，以便在云存储和 AWS 存储服务之间进行传输
- 部署您的基本模式代理以在 Amazon S3 和 AWS 文件系统之间进行传输

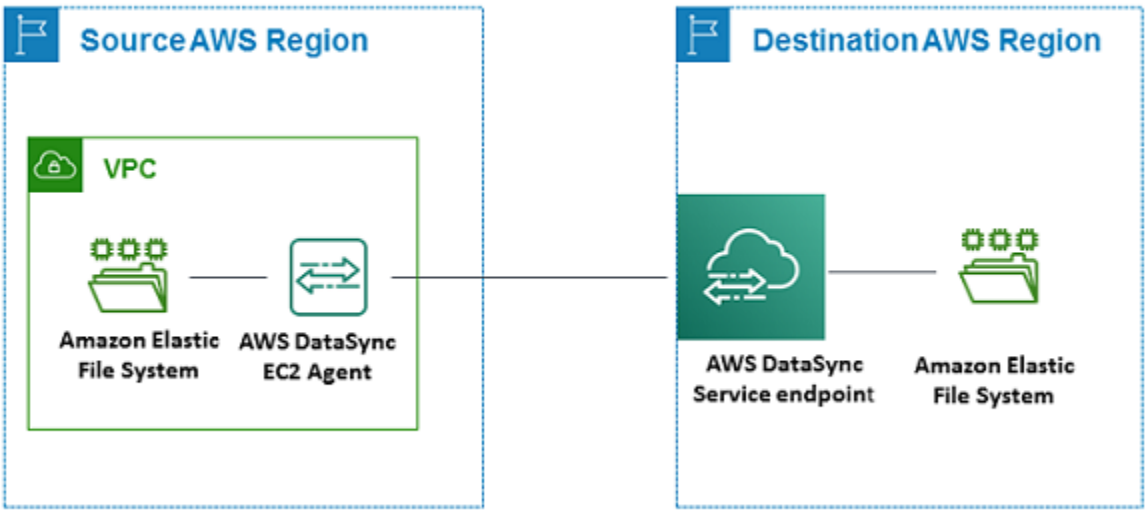
部署基本模式代理，以便在云存储和 AWS 存储服务之间进行传输

要在云存储系统之间或云存储系统之间传输数据， DataSync 代理必须位于源文件系统 AWS 账户 所在的相同 AWS 区域 位置。 AWS 账户这种类型的传输包括以下传输：

- 在 Amazon EFS 或 Amazon 之间传输 FSx 到不同的 AWS 存储空间 AWS 账户。
- 从自行管理的文件系统传输到 AWS 存储服务。

> **Important**
>
> 部署您的代理，使其不需要可用区之间的网络流量（以避免此类流量产生费用）。
>
> - 要访问您的 Amazon EFS 或 FSx Windows 文件服务器文件系统，请在具有文件系统挂载目标的可用区中部署代理。
> - 对于自管理的文件系统，请将代理部署在文件系统所在的可用区中。
>
> 要详细了解所有人的数据传输价格 AWS 区域，请参阅 Amazon EC2 按需定价。

例如，下图显示了将数据从云端网络文件系统 (NFS) 传输到云中 NFS 或 Amazon S3 的 DataSync 架构的高级视图。

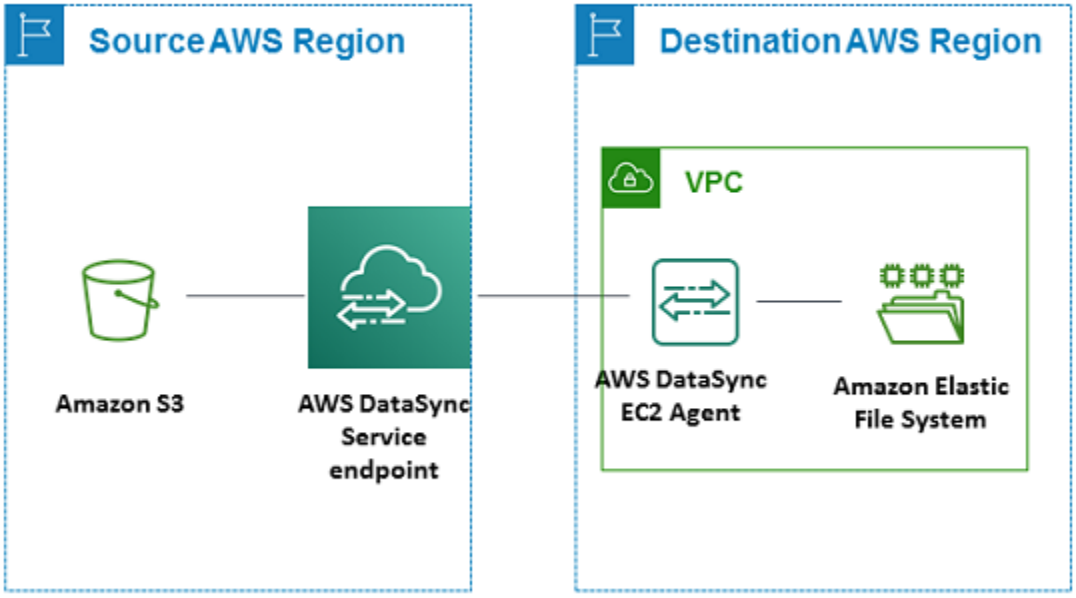


在跨 AWS 存储服务之间传输时，请记住以下几点 AWS 账户：

- 使用 NFS 协议在 Amazon EFS FSx 文件系统或亚马逊文件系统之间传输时，请将源文件系统配置为 NFS 位置。
- 使用 SMB 协议在 Amazon FSx 文件系统之间传输时，请将您的源文件系统配置为 SMB 位置。

部署您的基本模式代理以在 Amazon S3 和 AWS 文件系统之间进行传输

下图提供了将数据从 Amazon S3 传输到 AWS 文件系统（例如 Amazon EFS 或 Amazon）的 DataSync架构的高级视图 FSx。您可以使用此架构将数据从一个架构传输 AWS 账户 到另一个架构，或者将数据从 Amazon S3 传输到自行管理的云端文件系统。



在上部署基本模式代理 AWS Outposts

你可以在你的 Outpost 上启动 DataSync Amazon EC2 实例。要了解有关在上启动 AMI 的更多信息 AWS Outposts，请参阅AWS Outposts 用户指南中的在您的 Outpost 上启动实例。

在 Outposts 上使用 DataSync 访问 Amazon S3 时，您必须使用基本模式代理并在允许访问您的 Amazon S3 接入点的 VPC 中启动该代理，然后在 Outpost 的父区域激活该代理。代理还必须能够路由到存储桶的 Amazon S3 on Outposts 端点 。要了解有关使用 Amazon S3 on Outposts 端点的更多信息，请参阅 Amazon S3 用户指南中的使用 Amazon S3 on Outposts。

为 AWS DataSync 代理选择服务端点

服务端点是您的 AWS DataSync 代理与 DataSync服务通信的方式。 DataSync 支持以下类型的服务端点：

- 公共服务端点：数据通过公共互联网发送。
- 联邦信息处理标准（FIPS）服务端点：使用符合 FIPS 的进程通过公共互联网发送数据。
- 虚拟私有云（VPC）服务端点：数据通过您的 VPC 发送，而不是通过公共互联网发送，从而提高了传输数据的安全性。
- FIPS VPC 服务端点：数据通过符合 FIPS 的流程经由 VPC 发送。

您需要服务端点来激活代理。在选择服务端点时，请记住以下事项：

- 一个代理只能使用一种类型的端点。如果您需要使用不同的端点类型来传输数据，请为每种类型创建一个代理。
- 将存储网络连接至 AWS 的方式决定了您可以使用的服务端点。

选择一个公共服务端点

如果您使用公共服务端点，则您的 DataSync 代理与 DataSync 服务之间的所有通信都通过公共 Internet 进行。

1. 确定要使用的 DataSync 公共服务终端节点。
2. 配置您的网络以允许使用 DataSync 公共服务端点所需的流量。

下一步激活您的 AWS DataSync 代理：

选择 FIPS 服务端点

DataSync 提供了一些符合 FIPS 的服务端点。有关更多信息，请参阅 https://docs.aws.amazon.com/general/latest/gr/rande.html#FIPS-endpoints 中的AWS 一般参考 FIPS 端点。

1. 确定要使用的 DataSync FIPS 服务终端节点。
2. 配置您的网络以允许使用 DataSync FIPS 服务端点所需的流量。

下一步激活您的 AWS DataSync 代理：

选择 VPC 服务端点

如果您使用 VPC 服务终端节点，则您的数据不会通过公共互联网传输。 DataSync 而是通过基于 Amazon VPC 服务的 VPC 传输数据。

目录

- DataSync 代理如何使用 VPC 服务终端节点
- DataSync 局限性 VPCs
- 为创建 VPC 服务终端节点 DataSync

DataSync 代理如何使用 VPC 服务终端节点

VPC 服务终端节点由提供 AWS PrivateLink。这些类型的终端节点允许您以私密方式 AWS 服务连接到支持的 VPC。当您将 VPC 服务终端节点与一起使用时 DataSync，您的 DataSync 代理与 DataSync 服务之间的所有通信都将保留在您的 VPC 中。

VPC 服务终端节点（以及为数据传输流量 DataSync 创建的网络接口）使用只能从您的 VPC 内部访问的私有 IP 地址。有关更多信息，请参阅 连接您的网络进行 AWS DataSync 传输。

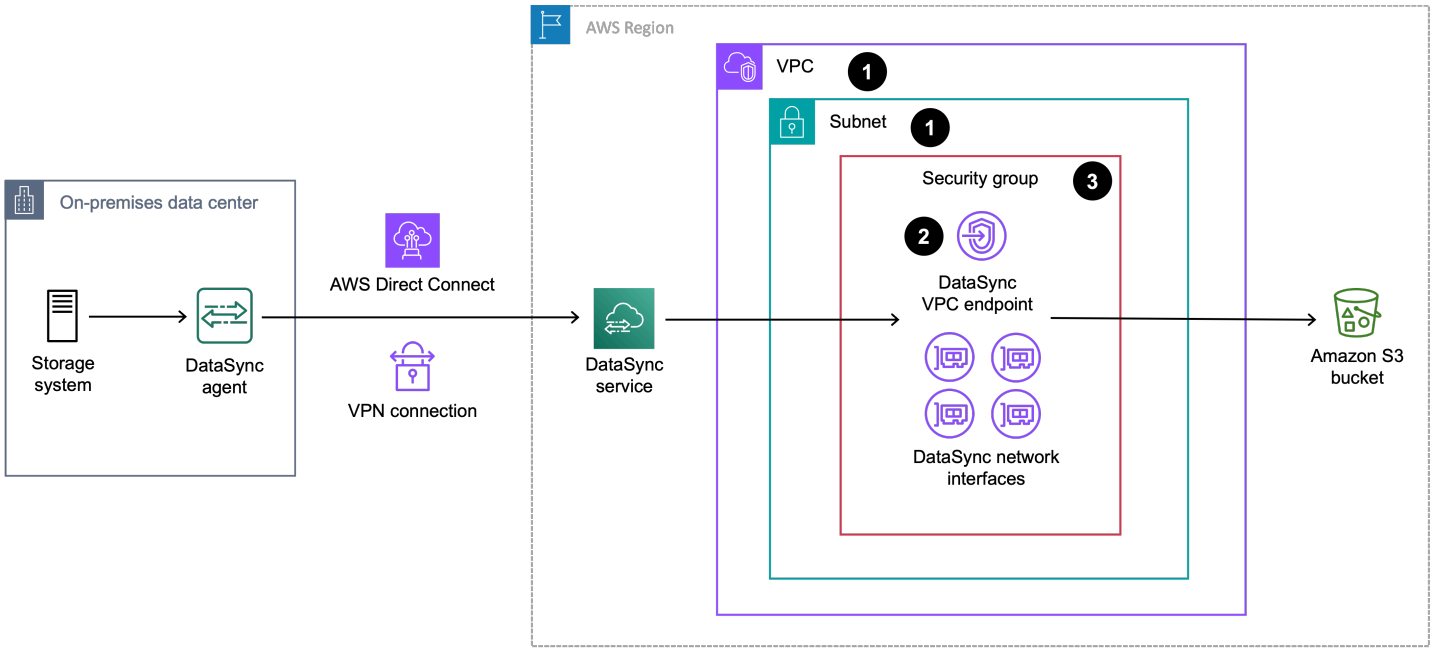
DataSync 局限性 VPCs

- VPCs 与一起使用的 DataSync 必须具有默认租期。 VPCs 不支持专用租赁。
- DataSync 不支持共享 VPCs。

为创建 VPC 服务终端节点 DataSync

您可以在自己管理的 VPC DataSync 中为创建 VPC 服务终端节点。您的服务终端节点、VPC 和 DataSync 代理必须属于同一个终端节点 AWS 账户。

下图显示了 DataSync 使用 VPC 服务终端节点从本地存储系统传输到 Amazon S3 存储桶的示例。编号的标注与 VPC 服务端点的创建步骤对应。



为创建 VPC 服务终端节点 DataSync

1. 创建或确定要在其中创建 VPC 服务端点的 VPC 和子网。

如果您要与外部存储进行传输 AWS，VPC 应扩展到该存储环境（例如，您的存储环境可能是您的本地 NFS 文件服务器所在的数据中心）。您可以通过 Direct Connect 或 VPN 使用路由规则来实现这一点。

2. 通过执行以下操作创建 DataSync VPC 服务终端节点：

   a. 打开位于 https://console.aws.amazon.com/vpc/ 的 Amazon VPC 控制台。

   b. 在左侧导航窗格中，选择端点，然后选择创建端点。

   c. 对于 Service category（服务类别），选择 AWS 服务。

   d. 对于服务，搜索**datasync**并选择您所在 AWS 区域 的终端节点（例如，`com.amazonaws.us-east-1.datasync`或`com.amazonaws.us-east-1.datasync-fips`）。

   e. 对于 VPC，选择您要在其中创建 VPC 服务端点的 VPC。

   f. 展开附加设置并清除启用专用 DNS 名称复选框，以禁用此设置。

      如果同一 VPC 中有代理需要使用公共服务端点，我们建议禁用此设置。启用此设置后，代理无法通过网络到达公共服务端点。

   g. 对于子网，选择您要在其中创建 VPC 服务端点的子网。记下子网 ARN（激活代理时需要）。

   h. 选择创建端点。记下端点 ID（激活代理时需要）。

3. 在您的 VPC 中，配置一个允许使用 DataSync VPC 服务终端节点所需的流量的安全组。记下安全组 ARN（激活代理时需要）。

   安全组必须允许代理连接 VPC 服务端点的私有 IP 地址和您的网络接口（创建任务时会创建）。

下一步激活您的 AWS DataSync 代理：

激活您的 AWS DataSync 代理

要完成 AWS DataSync 代理的创建，必须将其激活。此步骤将代理与您的关联 AWS 账户。

> **Note**
>
> 您不能 AWS 区域 同时激活多个 AWS 账户 代理。

先决条件

要激活您的 DataSync 代理，请确保您有以下信息：

- 您激活代理时使用的DataSync 服务终端节点。

  如果使用的是 VPC 服务端点，则需要这些详细信息：

  - VPC 服务端点 ID。
  - 您的 VPC 服务端点所在子网。
  - 允许使用 DataSync VPC 服务终端节点所需的流量的安全组。
- 您的代理的 IP 地址或域名。

  如何查找此信息取决于您部署的代理类型。例如，如果您的代理是一个 Amazon EC2 实例，可以通过访问 Amazon EC2 控制台的实例页面，找到它的 IP 地址。

> **Note**
>
> 对于 FIPS VPC 终端节点，请使用 AWS CLI 或 DataSync API。

获取激活密钥

您可以通过几种不同的方式获取已部署 DataSync 代理的激活密钥。某些选项需要通过端口 80（HTTP）访问代理。如果您使用这些选项之一，请在激活代理后 DataSync 关闭端口。

> **Note**
>
> 如果未使用，代理激活密钥将在 30 分钟后过期。

DataSync console

在 DataSync 控制台中激活代理时， DataSync 可以使用“自动从代理获取激活密钥”选项来获取激活密钥。

要使用此选项，您的浏览器必须能够通过端口 80 访问您的代理。

Agent local console

与其他获取激活密钥的选项不同，此选项不要求您的代理可通过端口 80 进行访问。

1. 登录到代理虚拟机（VM）或 Amazon EC2 实例的本地控制台。
2. 在 “AWS DataSync 激活-配置” 主菜单上，输入**0**以获取激活密钥。
3. 输入您 AWS 区域 要激活代理的。
4. 输入代理使用的服务端点类型。
5. 复制显示的激活密钥。

   例如：`F0EFT-7FPPR-GG7MC-3I9R3-27DOH`

   您可以在激活代理时指定此密钥。

CLI

使用标准 Unix 工具，您可以向代理的 IP 地址发出 `curl` 请求，以获取其激活密钥。

要使用此选项，您的客户端必须能够通过端口 80 访问您的代理。您可以运行以下命令检查：

```
nc -vz agent-ip-address 80
```

确认可以连接到代理后，根据使用的服务端点类型，运行以下命令之一：

- 公共服务端点：

  ```
  curl "http://agent-ip-address/?gatewayType=SYNC&activationRegion=your-
  region&no_redirect"
  ```

- FIPS 服务端点：

  ```
  curl "http://agent-ip-address/?gatewayType=SYNC&activationRegion=your-
  region&endpointType=FIPS&no_redirect"
  ```

- VPC 服务端点：

  ```
  curl "http://agent-ip-address/?gatewayType=SYNC&activationRegion=your-
  region&privateLinkEndpoint=vpc-endpoint-ip-
  address&endpointType=PRIVATE_LINK&no_redirect"
  ```

- FIPS VPC 服务端点：

```
curl "http://agent-ip-address/?gatewayType=SYNC&activationRegion=your-
region&privateLinkEndpoint=vpc-endpoint-ip-
address&endpointType=FIPS_PRIVATE_LINK&no_redirect"
```

> **Note**
>
> 要查找*vpc-endpoint-ip-address*，请打开 Amazon VPC 控制台，选择终端节点，然后选择您的 DataSync VPC 服务终端节点。在子网选项卡上，找到您的 VPC 服务端点子网的 IP 地址。这是端点的 IP 地址。

此命令将返回激活密钥。例如：

```
F0EFT-7FPPR-GG7MC-3I9R3-27DOH
```

您可以在激活代理时指定此密钥。

激活 代理

您可以通过多种方式激活 DataSync 代理。激活后， AWS 将为您管理代理。

DataSync console

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择 代理，然后选择 创建代理。
3. 在服务端点部分，执行以下操作，指定代理的服务端点：
   - 对于公共服务终端节点，请选择中的公共服务终端节点***your current AWS ##***。
   - 对于 FIPS 服务终端节点，请在中选择 FIPS 服务终端节点。*your current AWS ##*
   - 对于 VPC 服务端点，请执行以下操作：
     - 选择使用 AWS PrivateLink的 VPC 端点。
     - 对于 VPC 端点，选择您希望代理使用的 VPC 服务端点。
     - 对于子网，选择 VPC 服务端点所在的子网。
     - 对于安全组，请选择允许使用 DataSync VPC 服务终端节点所需的流量的安全组。
4. 在激活密钥部分，执行以下任一操作，以指定代理的激活密钥：

- 选择“自动从代理处获取激活密钥” DataSync 以获取密钥。
  - 对于代理地址，输入代理的 IP 地址或域名。
  - 选择 获取密钥。

    如果激活失败，请根据使用的服务端点类型检查网络配置。
- 如果不想在浏览器和代理之间建立连接，请选择手动输入代理激活密钥。
  - 从代理本地控制台或使用 `curl` 命令获取密钥。
  - 回到 DataSync 控制台，在激活密钥字段中输入密钥。

5. （推荐）对于代理名称，请为代理提供一个您能记住的名称。
6. （可选）对于标签，为密钥和值字段输入值以标记您的代理。

   标签可帮助您管理、筛选和搜索 AWS 资源。
7. 选择 Create agent (创建代理)。
8. 在代理页面，验证代理使用的服务端点类型是否正确。

> **Note**
>
> 此时，您可能会注意到代理处于离线状态。激活代理后会短暂出现这种情况。

AWS CLI

1. 获得激活密钥后，根据使用的服务端点类型，复制以下 `create-agent` 命令之一：
   - 公共或 FIPS 服务端点：

```
aws datasync create-agent \
  --activation-key activation-key \
  --agent-name name-for-agent
```

   - VPC 或 FIPS VPC 服务端点：

```
aws datasync create-agent \
  --activation-key activation-key \
  --agent-name name-for-agent \
  --vpc-endpoint-id vpc-endpoint-id \
  --subnet-arns subnet-arn \
```

```
    --security-group-arns security-group-arn
```

2. 对于 `--activation-key`，请指定您的代理激活密钥。
3. （推荐）对于 `--agent-name`，请为代理指定一个您可以记住的名称。
4. 如果使用的是 VPC 服务端点，请指定以下选项：
   - 对于 `--vpc-endpoint-id`，请指定您正在使用的 VPC 服务端点的 ID。
   - 对于 `--subnet-arns`，指定您的 VPC 服务端点所在子网的 ARN。
   - 对于`--security-group-arns`，请指定允许使用 VP DataSync C 服务终端节点所需的流量的安全组的 ARN。
5. 运行 `create-agent` 命令。

   您会收到回复，其中包含您刚刚激活的代理的 ARN。例如：

```
{
    "AgentArn": "arn:aws:datasync:us-east-1:111222333444:agent/
agent-0b0addbeef44baca3"
}
```

6. 通过运行 `list-agents` 命令，验证您的代理是否已激活：

```
aws datasync list-agents
```

> **Note**
>
> 此时，您可能会注意到代理 `Status` 为 `OFFLINE`。激活代理后会短暂出现这种情况。

DataSync API

获得激活密钥后，使用CreateAgent操作激活代理。

> **Note**
>
> 完成后，您可能会注意到代理已离线。激活代理后会短暂出现这种情况。

后续步骤

- 验证代理与存储系统和 DataSync 服务的连接。
- 如果您在尝试激活代理时遇到问题，请通过故障排除页面寻求帮助。
- 创建您要与代理一起使用 DataSync 的地点。这可以是本地位置，或其他云端位置。

验证代理的网络连接

激活 AWS DataSync 代理后，请确保该代理已通过网络连接到您的存储系统和 DataSync 服务。

访问代理的本地控制台

如何访问代理的本地控制台取决于您使用的代理类型。

访问本地控制台（VMware ESXi、Linux KVM 或微软 Hyper-V）

出于安全考虑，您无法远程连接到 DataSync 代理虚拟机 (VM) 的本地控制台。

- 如果这是您首次登录本地控制台，请使用默认凭证登录。默认用户名为 **admin**，密码为 **password**。

  > **Note**
  >
  > 我们建议您更改默认密码。为此，请在控制台主菜单上输入**5**（或 **6** for VMware VMs），然后运行`passwd`命令以更改密码。

访问本地控制台 (Amazon EC2)

要连接到 Amazon EC2 代理的本地控制台，必须使用 SSH。

开始之前：确保您的 EC2 实例的安全组允许通过 SSH 进行访问（TCP 端口 22）。

1. 打开一个终端，并复制以下 `ssh` 命令：

   ```
   ssh -i /path/key-pair-name.pem instance-user-name@instance-public-ip-address
   ```

   - 对于*/path/key-pair-name*，请指定连接到您的实例所需的私有密钥的路径和文件名 (`.pem`)。

- 对于 *instance-user-name*，请指定 admin。
- 对于*instance-public-ip-address*，请指定您的实例的公有 IP 地址。

2. 运行 ssh 命令连接实例。

连接后，将显示代理本地控制台的主菜单。

验证代理与存储系统的连接

测试您的 DataSync 代理是否可以连接到您的存储系统。有关更多信息，请参阅 1. 存储系统和代理之间的网络连接。

1. 访问代理的本地控制台。
2. 在 “AWS DataSync 激活-配置” 主菜单上输入**3**。
3. 输入以下选项之一：

   a. 输入 **1** 以测试 NFS 服务器连接。

   b. 输入 **2** 以测试 SMB 服务器连接。

   c. 输入 **3** 以测试对象存储服务器的连接。

   d. 输入 **4** 以测试 HDFS 连接。

   e. 输入 **5** 以测试 Microsoft Azure Blob Storage 连接。

4. 输入存储服务器的 IP 地址或域名。

   输入 IP 地址或域名时，请记住以下几点：

   - 请不要包含协议。例如，输入 **mystorage.com** 而不是 **https://mystorage.com**。
   - 对于 HDFS，请在 Hadoop 集群 DataNode 中输入 NameNode 或的 IP 地址或域名。

5. 如果需要，请输入用于连接存储服务器的 TCP 端口（例如 **443**）。

查看连接测试是通过还是失败。

验证您的代理与 DataSync 服务的连接

测试您的 DataSync 代理是否可以连接到该 DataSync 服务。有关更多信息，请参阅 2. 您的代理和 DataSync 服务之间的网络连接。

1. 访问代理的本地控制台。

2. 在 “AWS DataSync 激活-配置” 主菜单上**2**，输入开始测试网络连接。

   如果您的代理已激活，无需任何其他用户输入即可启动测试网络连接选项，因为区域和端点类型取自激活的代理信息。

3. 输入您的代理使用的 DataSync 服务端点的类型：

   a. 对于公共服务终端节点**1**，请输入您的代理的激活 AWS 区域 位置。

   b. 对于 FIPS 服务端点，输入 **2** 和在其中激活代理的区域。

   c. 对于 VPC 服务端点，输入 **3**。

   d. 对于 FIPS VPC 服务端点，输入 **4**。

   您会看到指示通过或失败的消息。

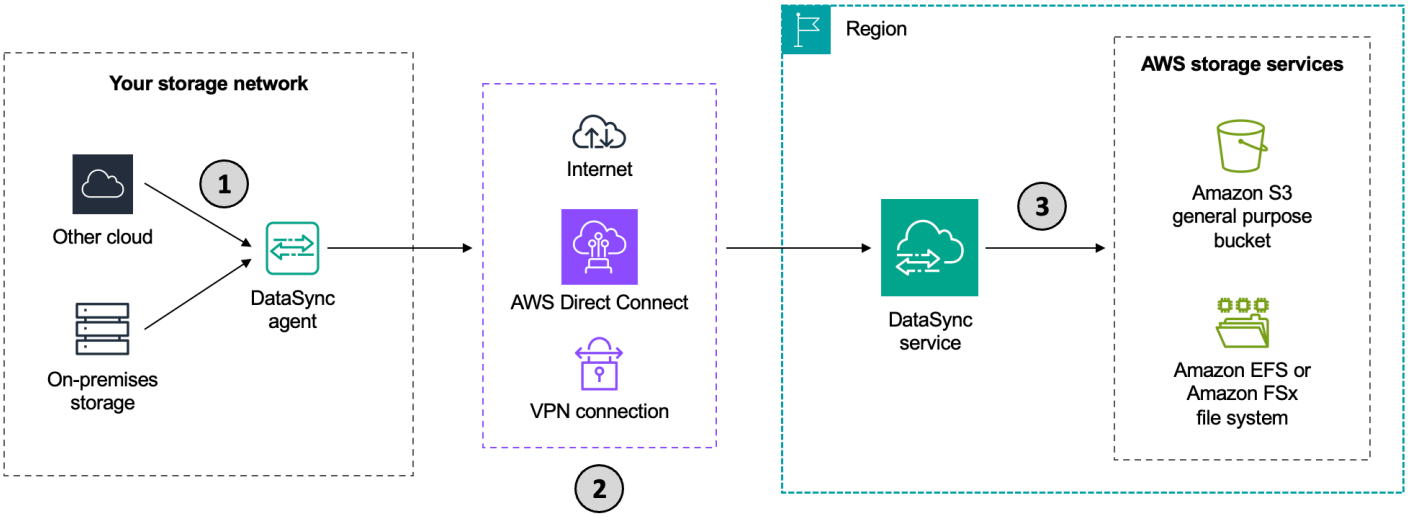
4. 如果您看到失败消息，请检查您的网络配置。有关更多信息，请参阅 AWS DataSync 网络要求。

后续步骤

创建您要与代理一起使用 DataSync 的地点。这可以是本地位置，或其他云端位置。

连接您的网络进行 AWS DataSync 传输

如果需要 AWS DataSync 代理，则必须建立多个网络连接才能进行数据传输。下图显示了从存储系统（可能位于本地、另一云端或边缘）向 AWS 存储服务 DataSync 传输过程中的三个网络连接。



1. 存储系统和代理之间的网络连接

您的 DataSync 代理连接到您的本地存储或其他云中的存储。有关更多信息，请参阅 本地、自我管理和其他云存储的网络要求。

2. 您的代理和 DataSync 服务之间的网络连接

将您的代理连接到 DataSync 服务有几个方面。首先，您必须在存储位置和之间建立网络连接 AWS。其次，您的代理需要一个服务端点来与之通信 DataSync。

目录

- 将您的存储网络连接到 AWS
- 选择服务终端节点

将您的存储网络连接到 AWS

使用时 DataSync，请考虑使用以下选项将存储网络连接到 AWS：

- Direct Connect：使用 Direct Connect，您可以在存储网络和 AWS之间创建专用连接。从 DataSync 角度来看，这可以让你：

- 通过专用路径将数据传输到虚拟私有云（VPC），从而避免通过公共互联网路由。
- 与使用虚拟专用网络 (VPN) 连接存储网络相比，获得更可预测的连接 AWS （尤其是在您的代理是 Amazon EC2 实例的情况下）。
- 使用任何类型的 DataSync 服务终端节点，包括公共、联邦信息处理标准 (FIPS) 或 VPC 终端节点。

有关更多信息，请参阅 DataSync 架构和路由示例 Direct Connect。

- VPN-您可以使用 VPN（例如 AWS Site-to-Site VPN）将存储网络连接到 AWS 。
- 公共互联网-您可以使用公共或 FIPS 服务端点将存储网络直接连接到互联网。 DataSync

选择服务终端节点

您的代理使用服务端点进行通信 DataSync。有关更多信息，请参阅 为 AWS DataSync 代理选择服务端点。

3. 服务和 AWS 存储 DataSync 服务之间的网络连接

要 DataSync 连接到 AWS 存储服务，您只需确保该 DataSync 服务可以访问您的 S3 存储桶或文件系统即可。有关更多信息，请参阅 AWS 存储服务的网络要求。

不需要 DataSync 代理时联网

对于不需要 DataSync代理的传输，您只需要确保该 DataSync 服务可以访问要在它们之间传输的 AWS 存储系统即可。有关更多信息，请参阅 AWS 存储服务的网络要求。

DataSync 流量通过网络的方式和位置

DataSync 有数据平面和控制平面流量。如果您想分离流量，那么了解这些流量是如何流经网络的，这一点非常重要。 DataSync

- 数据面板流量：包括在存储位置之间移动的文件或对象数据。在大多数情况下，数据平面流量会通过网络接口进行路由，这些接口会在您创建任务时 DataSync 自动生成和管理。这些网络接口的创建位置取决于您要传输的 AWS 存储服务类型以及 DataSync 代理使用的服务端点。
- 控制飞机流量-包括 DataSync 资源的管理活动。此种流量通过代理使用的服务端点路由。

网络安全 DataSync

有关在传输过程中如何保护存储数据（包括元数据）的信息，请参阅 AWS DataSync 传输中的加密。

AWS DataSync 网络要求

配置网络是 AWS DataSync设置过程中的一个重要步骤。网络配置取决于多种因素，例如您正在使用的存储系统类型。它还取决于您计划使用的 DataSync 服务端点类型。

IPv6 支持

DataSync 支持双堆栈，可与 IPv6 网络 IPv4 兼容。 IPv6 在提供服务的所有 AWS 区域 地方都提供支持。 DataSync 支持使用具有以下数据源 IPv6 的地址：

- Elastic File System（EFS）
- 网络文件系统 (NFS)
- 服务器消息块 (SMB)
- 对象存储

IPv6-适用于本地存储的兼容代理

要在 IPv6 网络环境 DataSync 中使用，需要使用 IPv6兼容的代理。这些代理同时支持 IPv4 和 IPv6 连接，可适应各种网络环境。

- 对于 IPv6仅限网络的网络，无需更改配置。
- 对于 IPv4仅限网络的网络，无需更改配置。
- 对于双栈（两者 IPv4 兼有 IPv6）网络，让代理根据您的喜好选择协议或手动配置协议。

双堆栈网络的注意事项

可使用以下方式，通过本地控制台自定义代理行为：

- 禁用， IPv6 这样代理就无法使用 IPv6 它来访问本地文件系统或服务。 DataSync
- 设置代理用于数据传输的 IP 版本：
  - 设置为， IPv6 使代理仅 IPv6 用于数据传输。
  - 设置 IPv4 为代理仅 IPv4 用于数据传输。
  - 设置为“自动”（恢复默认值），以便代理自动选择用于数据传输的协议版本（IPv4 或 IPv6）。

有关管理代理的 IP 版本设置的更多信息，请参阅the section called “对代理进行维护”。

> **Important**
>
> 根据 2025 年 7 月 16 日之前下载的图像构建的代理不支持 IPv6。

本地、自我管理和其他云存储的网络要求

以下网络要求可能适用于本地、自我管理和其他云存储系统。这些存储系统通常由您管理，也可能由其他云提供商管理。

> **Note**
>
> 根据您的网络，您可能需要允许此处列出的端口以外的端口上的流量，您的 DataSync 代理才能连接到您的存储。

来源	目标	协议	端口：	它是如何使用的 DataSync
DataSync 代理人	NFS 文件服务器	TCP	2049（适用于 NFS 版本 4.1 和 4.0） 111 和 2049（适用于 NFS 3.x 版本）	装载 NFS 文件服务器。 DataSync 支持 NFS 版本 3.x、4.0 和 4.1。
DataSync 代理人	SMB 文件服务器	TCP	139 或 445	装载 SMB 文件服务器。 DataSync 支持 SMB 版本 1.0 及更高版本。出于安全起见，建议使用 SMB 版本 3.0.2 或更高版本。早期版本（例如 SMB

来源	目标	协议	端口：	它是如何使用的 DataSync
				1.0）包含已知的安全漏洞，攻击者可以借此来窃取您的数据。

来源	目标	协议	端口：	它是如何使用的 DataSync
DataSync 代理人	对象存储	TCP	443 (HTTPS) 或 80 (HTTP) Note 根据对象存储的不同，您可能需要允许非标准 HTTPS 和 HTTP 端口（如 8443	访问与 Amazon S3 兼容的本地对象存储或其他云中的存储。

来源	目标	协议	端口：	它是如何使用的 DataSync
			或 8080）上的流量。	
DataSync 代理人	Hadoop 集群	TCP	NameNode 端口（默认为 8020） 在大多数集群中，可以在`core-site.xml` 文件中的 `fs.default` 或`fs.deft.name` 属性下找到此端口号（取决于 Hadoop 发行版）。	访问您的 Hadoop 集群 NameNodes 中的。指定创建 HDFS 位置的端口。

来源	目标	协议	端口：	它是如何使用的 DataSync
DataSync 代理人	Hadoop 集群	TCP	DataNode 端口（默认为 50010） 在大多数集群中，可以在`dfs.datanode.address` 属性下的 `hdfs-site.xml` 文件中找到此端口号。	访问您的 Hadoop 集群 DataNodes 中的。DataSync代理会自动确定要使用的端口。
DataSync 代理人	Hadoop 密钥管理服务器（KMS）	TCP	KMS 端口（默认值为 9600）	访问您的 Hadoop 集群 KMS。
DataSync 代理人	Kerberos 密钥分配中心（KDC）服务器	TCP	KDC 端口（默认值为 88）	使用 Kerberos 领域执行身份验证。此端口仅适用于使用 Kerberos 身份验证的 HDFS 与 SMB 位置。
DataSync 代理人	存储系统管理界面	TCP	取决于您的网络	连接至您的存储系统。

AWS 存储服务的网络要求

传输期间 DataSync 连接到 AWS 存储服务所需的网络端口各不相同。

来源	目标	协议	端口：
DataSync 服务	Amazon EFS	TCP	2049
DataSync 服务	FSx 适用于 Windows 文件服务器	请参阅 Windows 文件服务器的文件系统访问控制。 FSx	
DataSync 服务	FSx 为了光泽	参见 Lustre 的文件系统访问控制。 FSx	
DataSync 服务	FSx 适用于 OpenZFS	请参阅 OpenZFS 的文件系统访问控制。 FSx	
DataSync 服务	FSx 适用于 ONTAP	TCP	111、635 和 2049 (NFS) 445 (SMB)
DataSync 服务	Amazon S3	不适用（代表您DataSync 连接到 S3 存储桶）	

公共或 FIPS 服务端点的网络要求

使用公共或 FIPS 服务端点时，您的 DataSync 代理需要以下网络访问权限。如果使用防火墙或路由器来筛选或限制网络流量，请配置防火墙和路由器以允许这些端点。

来源	目标	协议	端口：	使用方法	访问的端点
您的 Web 浏览器	DataSync 代理人	TCP	80 (HTTP)	允许您的浏览器获取 DataSync 代理的激活密钥。激活后，DataSync 关闭代理的端口 80。 您的代理不要求端口 80 可供公	不适用

来源	目标	协议	端口：	使用方法	访问的端点
				开访问。端口 80 所需的访问级别取决于网络配置。 Note 您无需在浏览器和代理之间建立连接，即可获取激活密钥。有关更多信息，请参阅 获取激活密钥。	
DataSync 代理人	Amazon CloudFront	TCP	443（HTTPS）	在激活之前帮助引导您的 DataSync 代理。 只有基本模式代理才需要。	AWS 区域: • d3dvvaliwoko8h.cloudfront.net AWS GovCloud (US) 区域： • s3.us-gov-west-1.amazonaws.com/fmrsendpoints-endpointsbucket-go4p5gpna6sk

来源	目标	协议	端口：	使用方法	访问的端点
DataSync 代理人	AWS	TCP	443（HT ）	激活您的 DataSync 代理并将其与您的 AWS 账户代理关联。您可以在激活后阻止公有端点。	*activation-region* 这是您激活 DataSync代理 AWS 区域 的地方。 公有端点激活： • activation.datasync. *activation-region*.amazonaws.com FIPS 端点激活： • activation.datasync-fips. *activation-region*.amazonaws.com

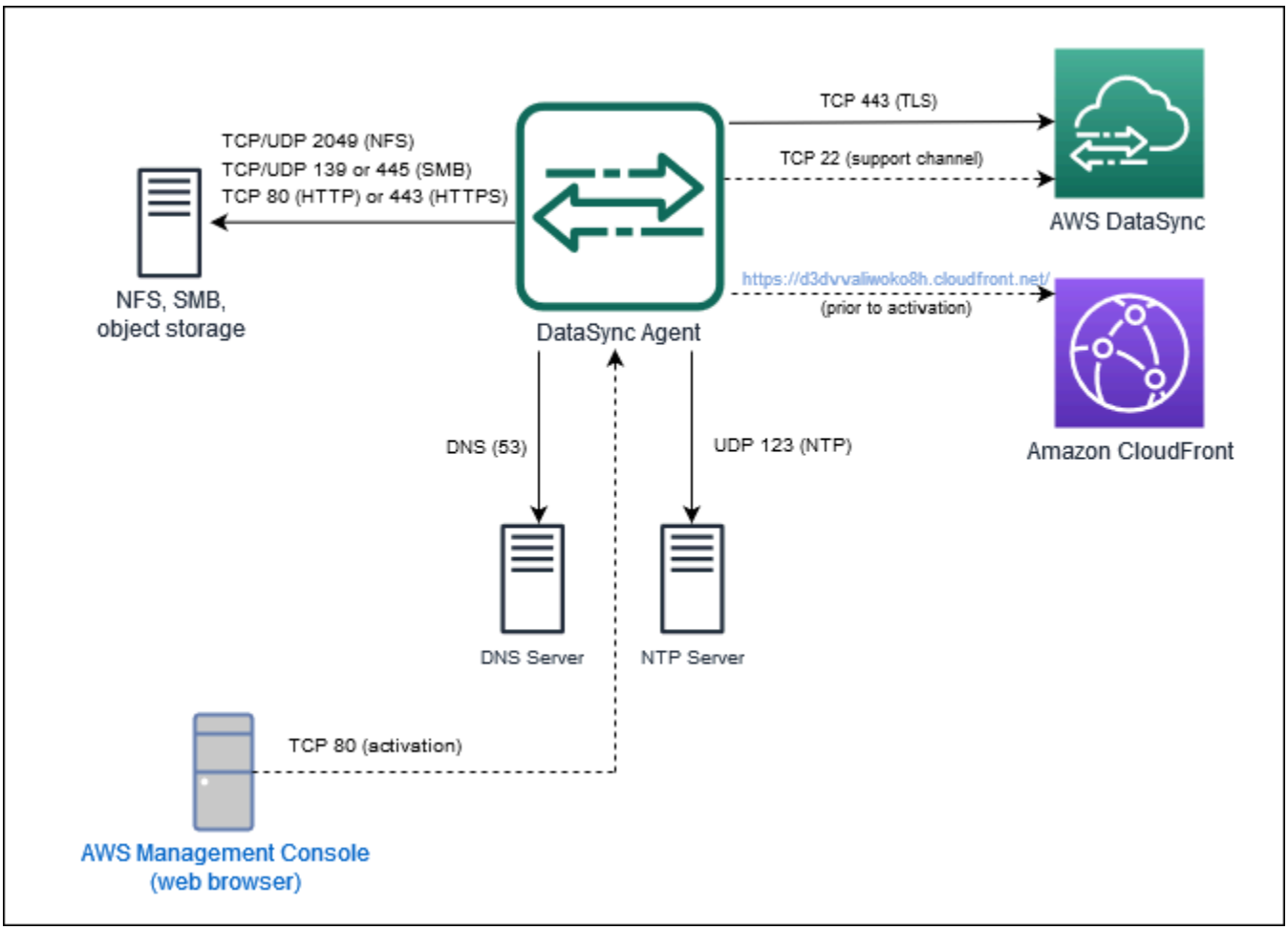
来源	目标	协议	端口：	使用方法	访问的端点
DataSync 代理人	AWS	TCP	443（HT ）	允许 DataSync 代理和 DataSync 服务端点之间的通信。 有关信息，请参阅为 AWS DataSync 代理选择服务端点。	*activation-region* 这是您激活 DataSync 代理 AWS 区域 的地方。根据您的用 DataSync 途，您可能不需要允许访问此处列出的每个端点。 DataSync 控制平面端点： • 公有端点：cp.datasync. *activation-region*.amazonaws.com • FIPS 端点：cp.datasync-fips. *activation-region*.amazonaws.com DataSync 数据平面端点（仅适用于传输任务）： • 基本模式代理：*your-task-id* .datasync-dp. *activation-region*.amazonaws.com 增强模式代理：*.*.*your-task-id.** .datasync-dp. *activation-region*.amazonaws.com

来源	目标	协议	端口：	使用方法	访问的端点
您的客户端	AWS	TCP	443（HT ）	允许您发出 DataSync API 请求。	*activation-region* 这是您激活 DataSync 代理 AWS 区域 的地方。 公有端点： • datasync. *activation-region*.amazonaws.com FIPS 端点： • datasync-fips. *activation-region* .amazonaws.com

来源	目标	协议	端口：	使用方法	访问的端点
DataSync 代理人	AWS	TCP	443（HT）	允许 DataSync 代理从获取更新 AWS。有关更多信息，请参阅 管理您的 AWS DataSync 代理。	*activation-region* 这是您激活 DataSync 代理 AWS 区域 的地方。 基本模式代理： • amazonlinux.default.amazonaws.com • cdn.amazonlinux.com • amazonlinux-2-repos- *activation-region* .s3.dualstack. *activation-region*.amazonaws.com • amazonlinux-2-repos- *activation-region*.s3.*activation-region*.amazonaws.com • *.s3.*activation-region*.amazonaws.com • *.s3.dualstack. *activation-region*.amazonaws.com 增强模式代理： • *.s3.dualstack. *activation-region*.amazonaws.com
DataSync 代理人	域名服务 (DNS) 服务器	TCP/UDP	53 (DNS)	允许 DataSync 代理与 DNS 服务器之间的通信。	不适用

来源	目标	协议	端口：	使用方法	访问的端点
DataSync 代理人	AWS	TCP	22 (支持渠道)	AWS 支持 允许访问您的 DataSync 代理以帮助您解决问题。您不需要打开此端口即可正常操作。	基本模式代理： • 商业区域：54.201.223.107 • GovCloud 区域：52.222.99.142 增强模式代理： • 商业区域：support.datasync.us-east-1.amazonaws.com • GovCloud 区域：support.datasync.us-gov-west-1.amazonaws.com
DataSync 代理人	网络时间协议 (NTP) 服务器	UDP	123 (NTP)	允许本地系统使用以将虚拟机时间同步到主机时间。	NTP： • 0.amazon.pool.ntp.org • 1.amazon.pool.ntp.org • 2.amazon.pool.ntp.org • 3.amazon.pool.ntp.org **Note** 要使用本地控制台更改 VM 代理的默认 NTP 配置以使用其他 NTP 服务器，请参阅查看和管理代理的系统时间服务器配置。

下图显示了使用公共或 FIPS 服务端点 DataSync 时所需的端口。



VPC 或 FIPS VPC 服务端点的网络要求

虚拟私有云 (VPC) 端点在您的代理之间提供私有连接 AWS ，该连接不通过互联网或使用公有 IP 地址。这还有助于防止数据包进入或者离开网络。有关更多信息，请参阅 选择 VPC 服务端点。

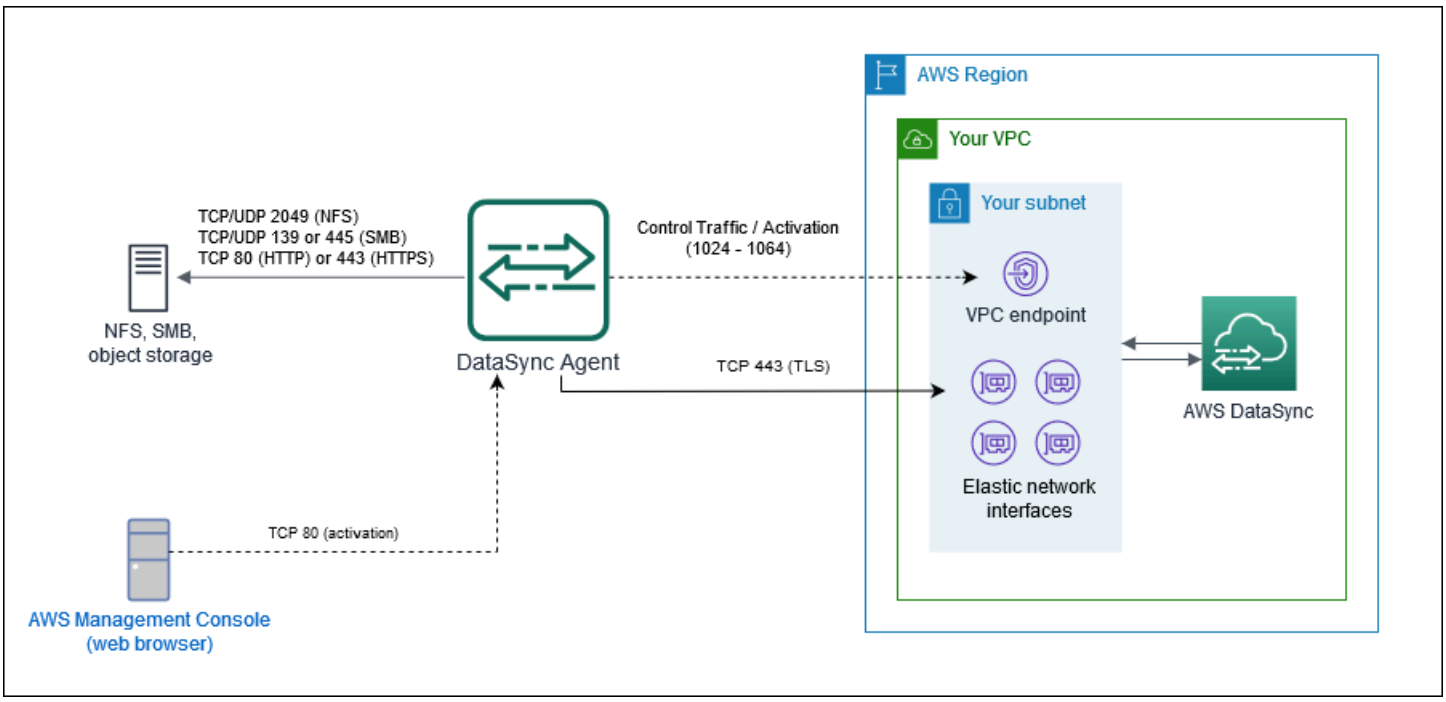
DataSync 您的代理需要以下端口才能使用 VPC 服务终端节点。

来源	目标	协议	端口：	使用方法
您的 Web 浏览器	你的 DataSync 经纪人	TCP	80 (HTTP)	允许浏览器获取代理的激活密钥。激活后，DataSync 关闭代理的端口 80。 您的代理不要求端口 80 可供公开访问。端口 80

来源	目标	协议	端口：	使用方法
				所需的访问级别取决于网络配置。 **Note** 您无需在浏览器和代理之间建立连接，即可获取激活密钥。有关更多信息，请参阅 获取激活密钥。
DataSync 代理人	您的 DataSync VPC 服务终端节点 要查找终端节点的 IP 地址，请打开 Amazon VPC 控制台，选择终端节点，然后选择您的 DataSync VPC 服务终端节点。在子网选项卡上，找到您的 VPC 服务端点子网的 IP 地址。这是端点的 IP 地址。	TCP	1024-1064	用于控制面板流量。
DataSync 代理人	您的 DataSync 任务的网络接口 要查找这些接口的 IP 地址，请参阅 查看您的网络接口。	TCP	443（HTTPS）	用于数据面板流量。

来源	目标	协议	端口：	使用方法
DataSync 代理人	您的 DataSync VPC 服务终端节点	TCP	22 (支持渠道)	允许 AWS 支持 访问您的 DataSync 代理进行故障排除。 只有基本模式代理才需要。 您不需要打开此端口即可正常操作。

下图显示了使用 VPC 服务终端节点 DataSync 时所需的端口。



用于 AWS DataSync 传输的网络接口

对于您创建的每项任务，都会 AWS DataSync 自动生成和管理用于数据传输流量的网络接口。创建多少网络接口及其 DataSync创建位置取决于您的传输任务的以下详细信息：

- 您的任务是否需要代 DataSync理。
- 您的源位置和目标位置（您将数据往来复制的位置）。
- 您的代理使用的服务端点类型。

每个网络接口在您的子网中使用一个 IP 地址（网络接口越多，您需要的 IP 地址就越多）。使用下表，确保您的子网有足够的 IP 地址来完成您的任务。

用于与代理传输的网络接口

通常，在 AWS 存储服务和存储系统之间复制数据时需要 DataSync 代理 AWS。

位置	默认创建网络接口	使用公共或 FIPS 端点时创建网络接口的位置	使用私有（VPC）端点时创建网络接口的位置
Amazon S3	4	不适用 [1]	您在激活 DataSync代理时指定的子网。
Amazon EFS	4	您在创建 Amazon EFS 位置时的指定子网。	
FSx 适用于 Windows 文件服务器的亚马逊	4	与文件系统首选文件服务器相同的子网。	
亚马逊 f FSx or Lustre	4	与文件系统相同的子网。	
FSx 适用于 OpenZFS 的亚马逊	4	与文件系统相同的子网。	
FSx 适用于 NetApp ONTAP 的亚马逊	4	与文件系统相同的子网。	

[1] 不需要网络接口，因为该 DataSync服务直接与 S3 存储桶通信。

无需代理即可传输的网络接口

在两者之间复制数据时不需要 DataSync 代理 AWS 服务。

网络接口的总数取决于传输中的 DataSync 位置。例如，在 Amazon EFS 和 FSx Lustre 文件系统之间传输需要四个网络接口。同时，在 Window FSx s 文件服务器和 S3 存储桶之间进行传输需要两个网络接口。

位置	默认创建网络接口	网络接口的创建位置
Amazon S3	不适用 [1]	不适用 [1]
Amazon EFS	2	您在创建 Amazon EFS 位置时的指定子网。
FSx 适用于 Windows 文件服务器	2	与文件系统首选文件服务器相同的子网。
FSx 为了光泽	2	与文件系统相同的子网。
FSx 适用于 OpenZFS	2	与文件系统相同的子网。
FSx 适用于 ONTAP	2	与文件系统相同的子网。

[1] 不需要网络接口，因为该 DataSync服务直接与 S3 存储桶通信。

查看您的网络接口

要查看分配给您的 DataSync 传输任务的网络接口，请执行以下任一操作：

- 使用 DescribeTask 操作。操作将返回SourceNetworkInterfaceArns 和 DestinationNetworkInterfaceArns，其响应如下所示：

  ```
  arn:aws:ec2:your-region:your-account-id:network-interface/eni-f012345678abcdef0
  ```

  在此示例中，网络接口 ID 为 eni-f012345678abcdef0。
- 在 Amazon EC2 控制台中，搜索您的任务 ID（例如 task-f012345678abcdef0）以查找其网络接口。

DataSync 架构和路由示例 Direct Connect

在使用 Direct Connect 进行 AWS DataSync 传输时，请考虑以下网络架构。

> **Tip**
>
> 如果您的网络使用传输网关，我们建议将 DataSync 传输的逻辑路径分开，以优化成本（尤其是在迁移大量数据的情况下）。
> 例如，如果您使用AWS Transit Gateway本地网络和虚拟私有云之间的普通流量（VPCs），则可以配置网络，使 DataSync 流量绕过传输网关及其数据处理费用。

将 Direct Connect 与 DataSync VPC 服务终端节点结合使用

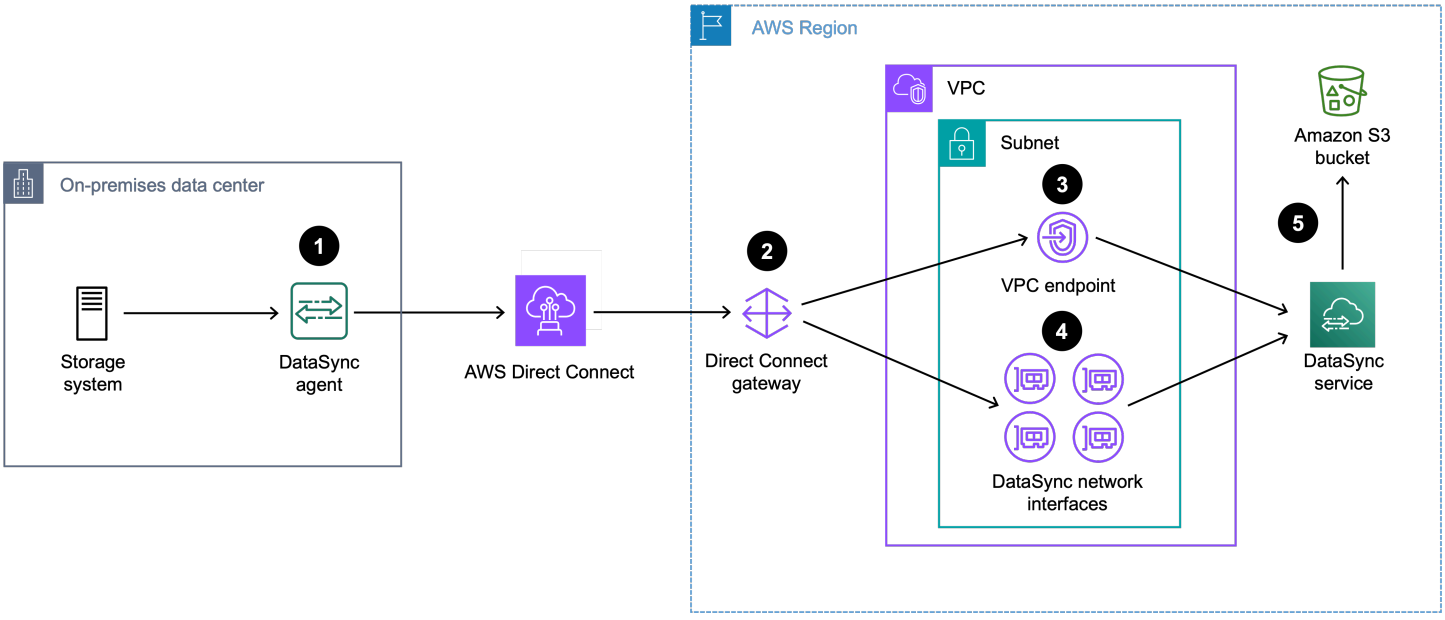
如果您的 DataSync 代理使用 VPC 服务终端节点，则需要一个 Direct Connect 网关才能连接到您的 VPC。

目录

- 带有 VPC 端点和 S3 目标的 Direct Connect 架构
- Direct Connect 架构，VPC 端点和文件系统目标位于同一子网中
- Direct Connect 架构，VPC 端点和文件系统目标位于不同子网

带有 VPC 端点和 S3 目标的 Direct Connect 架构

以下 Direct Connect 架构显示了从本地存储系统到 S3 存储桶的 DataSync 传输。



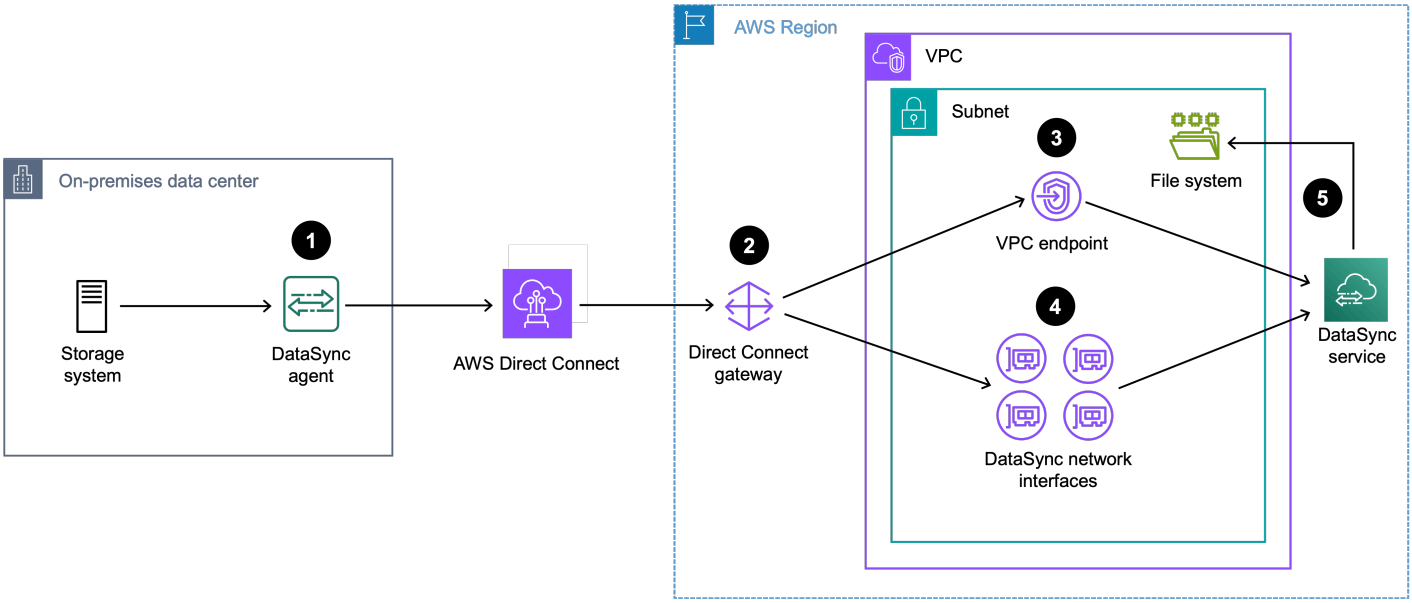
1. DataSync 代理将 DataSync 流量从本地存储系统（源位置）路由到 Direct Connect 连接。

2. DataSync 流量会路由到用于传输的 Direct Connect 网关。要对此进行设置，您必须：
   a. 将 Direct Connect 网关与 VPC 的虚拟专用网关关联。这是 VP DataSync C 终端节点所在的 VPC，也是 DataSync 任务创建网络接口的地方。
   b. 创建将此 VPC 连接到 Direct Connect 网关的专用虚拟接口。
3. DataSync 流量（控制平面）通过 DataSync VPC 终端节点路由。
4. DataSync 流量（数据平面）通过您在创建 DataSync 代理时指定的子网中的 DataSync 网络接口进行路由。
5. DataSync 流量通过 DataSync 服务路由到 S3 存储桶（目标位置）。

Direct Connect 架构，VPC 端点和文件系统目标位于同一子网中

在与 Amazon EFS 或 Amazon FSx 文件系统进行传输或传出时，您的文件系统和 DataSync VPC 终端节点可以位于同一个子网中。

以下 Direct Connect 架构显示了从本地存储系统到 Amazon EFS 或 Amazon FSx 文件系统的 DataSync 传输。



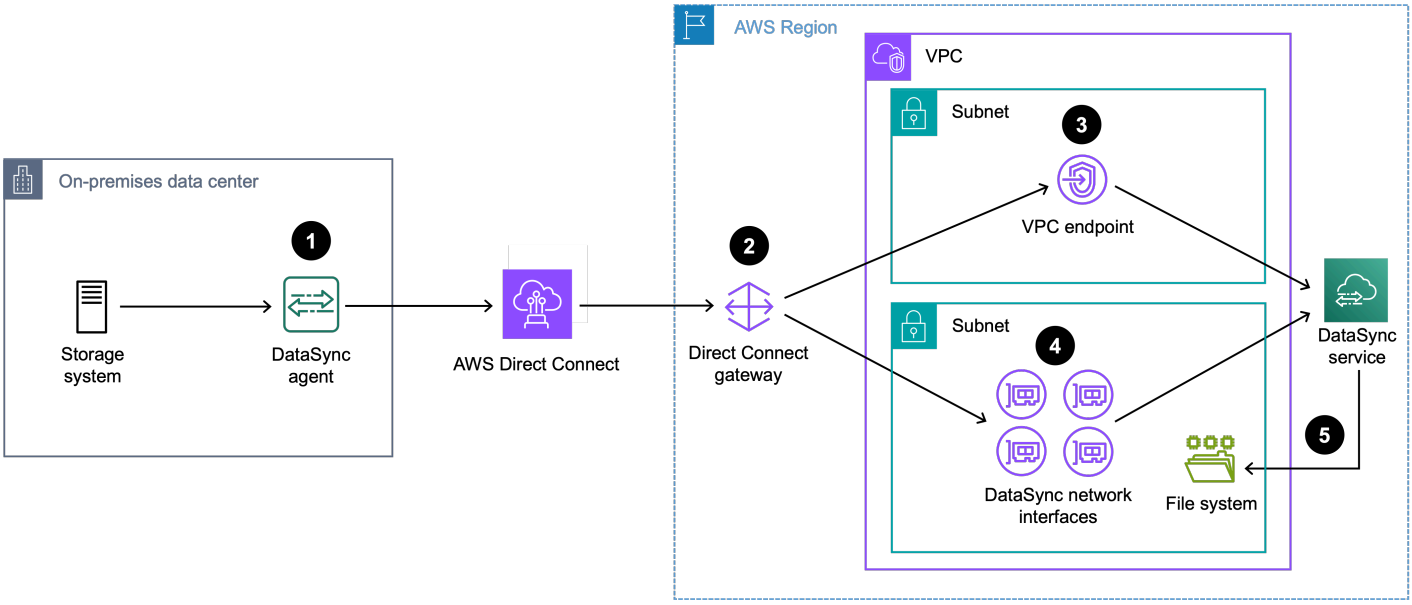
1. DataSync 代理将 DataSync 流量从本地存储系统（源位置）路由到 Direct Connect 连接。
2. DataSync 流量会路由到用于传输的 Direct Connect 网关。要对此进行设置，您必须：
   a. 将 Direct Connect 网关与 VPC 的虚拟专用网关关联。这是 VP DataSync C 终端节点所在的 VPC，也是 DataSync 任务为文件系统创建网络接口的地方（目标位置）。
   b. 创建将此 VPC 连接到 Direct Connect 网关的专用虚拟接口。
3. DataSync 流量（控制平面）通过 DataSync VPC 终端节点路由。

4. DataSync 流量（数据平面）通过文件系统子网中的 DataSync 网络接口进行路由。这与 DataSync VPC 终端节点所在的子网相同。
5. DataSync 流量通过 DataSync 服务路由到文件系统（目标位置）。

Direct Connect 架构，VPC 端点和文件系统目标位于不同子网

在与 Amazon EFS 或 Amazon FSx 文件系统进行传输或传出时，您的文件系统和 DataSync VPC 终端节点可以位于不同的子网中。

以下 Direct Connect 架构显示了从本地存储系统到 Amazon EFS 或 Amazon FSx 文件系统的 DataSync 传输。



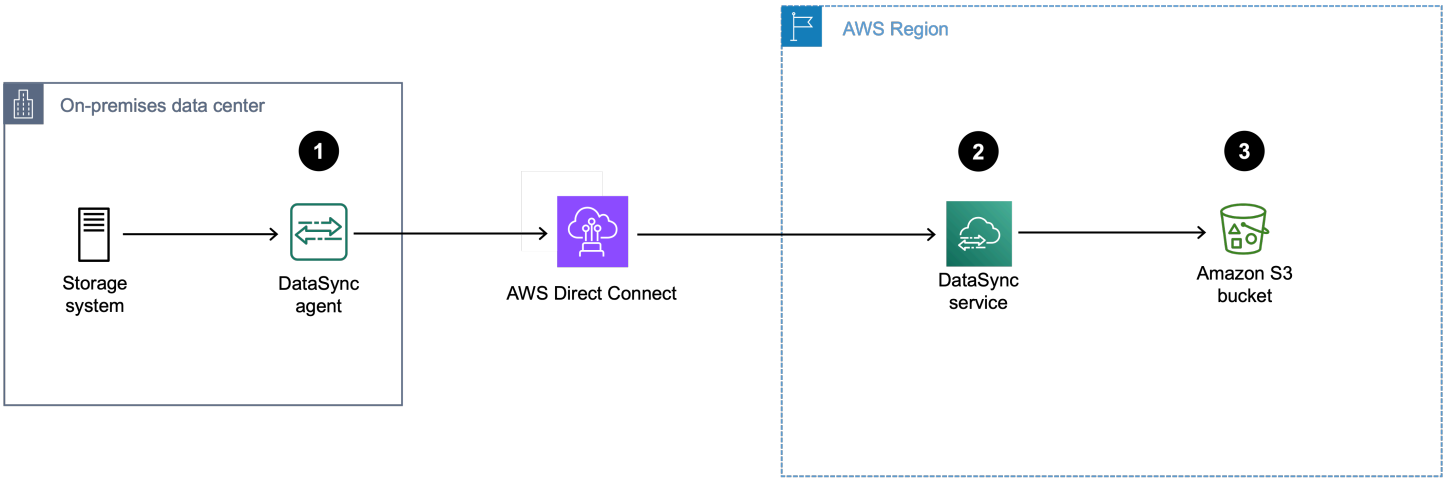
1. DataSync 代理将 DataSync 流量从本地存储系统（源位置）路由到 Direct Connect 连接。
2. DataSync 流量会路由到用于传输的 Direct Connect 网关。要对此进行设置，您必须：
   a. 将 Direct Connect 网关与 VPC 的虚拟专用网关关联。这是 VP DataSync C 终端节点所在的 VPC，也是 DataSync 任务为文件系统创建网络接口的地方（目标位置）。
   b. 创建一个私有虚拟接口， VPCs 将其连接到 Direct Connect 网关。
3. DataSync 流量（控制平面）通过 DataSync VPC 终端节点路由。
4. DataSync 流量（数据平面）通过文件系统子网中的 DataSync 网络接口进行路由。这与 DataSync VPC 终端节点所在的子网不同。
5. DataSync 流量通过 DataSync 服务路由到文件系统（目标位置）。

将 Direct Connect 与 DataSync 公共或 FIPS 服务端点配合使用

如果您的 DataSync 代理使用公共或联邦信息处理标准 (FIPS) 服务端点，则可以使用公共虚拟接口通过 Direct Connect 连接路由数据传输流量。

虽然 Direct Connect 默认会通告所有本地和远程 AWS 区域 前缀，但您可以使用 BGP 社区标签来控制公共虚拟接口上流量的范围（区域或全局）和路由首选项。您必须宣传至少一个公共前缀才能创建您的 DataSync 代理。

以下 Direct Connect 架构显示了从本地存储系统通过公共或 FIPS 端点到 S3 存储桶的 DataSync 传输。



1. DataSync 代理将 DataSync 流量从本地存储系统（源位置）路由到 Direct Connect 连接。
2. DataSync 流量通过公共虚拟接口路由到 DataSync 服务。
3. DataSync 到 S3 存储桶（目标位置）的流量。

后续步骤

如果您需要 DataSync 代理但尚未创建代理，请部署代理，为代理选择服务端点，然后激活该代理。

创建代理后，您可以为配置网络 DataSync。

为多个 AWS DataSync 代理配置代理 NICs

如果将 AWS DataSync 代理配置为使用多个网络适配器 (NICs)，则可以通过多个 IP 地址访问该代理。您可能希望在以下情况下执行此操作：

- 最大程度地增加吞吐量 - 当网络适配器成为瓶颈时，您可能希望最大程度地增加代理的吞吐量。

- 网络隔离 — 网络文件系统 (NFS)、服务器消息块 (SMB)、Hadoop Distributed File System (HDFS) 或对象存储服务器，可能位于出于安全原因而缺少互联网连接的虚拟局域网 (VLAN) 上。

在典型的多适配器用例中，将一个适配器配置为代理与之通信的路由 AWS （作为默认代理）。除了这个适配器之外，NFS、SMB、HDFS 或自行管理对象存储位置必须与连接到它们的适配器位于同一子网中。

否则，可能无法与预期 NFS、SMB、HDFS 或对象存储位置通信。在某些情况下，您可以在用于与通信的同一适配器上配置 NFS、SMB、HDFS 或对象存储位置。 AWS在这些情况下，该服务器的 NFS、SMB、HDFS 或对象存储流量以及 AWS 流量会流经同一个适配器。

在某些情况下，您可以将一个适配器配置为连接到 AWS DataSync 控制台，然后再添加第二个适配器。在这种情况下， DataSync 自动将路由表配置为使用第二个适配器作为首选路由。

使用传输您的数据 AWS DataSync

借 AWS DataSync助，您可以将数据传输到本地、内部或其他云中的存储 AWS，或从中传输数据。

设置 DataSync 转账通常涉及以下步骤：

1. 确定是否 DataSync 支持您的转移。
2. 如果需要 DataSync 代理来进行传输，请在尽可能靠近其中一个存储系统的地方部署和激活代理。

   例如，如果要从本地网络文件系统（NFS）文件服务器传输数据，应尽可能将代理部署在靠近该文件服务器的位置。
3. 提供对存储系统的 DataSync 访问权限。

   DataSync 需要权限才能读取或写入您的存储空间（取决于您的存储是源位置还是目标位置）。例如，学习如何提供 DataSync 对 NFS 文件服务器的访问权限。
4. 连接您的网络以获取存储系统和之间的流量 DataSync。
5. 使用 DataSync 控制台、 AWS CLI或 DataSync API 为您的源存储系统创建位置。

   例如，了解如何创建 NFS 位置或 Amazon S3 位置。
6. 重复步骤 3-5，创建转账的目的地地点。
7. 创建并启动包括您的来源和目标位置的 DataSync 转移任务。

主题

- 我可以在哪里传输我的数据 AWS DataSync？
- 使用本地存储与本地存储进行传入或传出本地存储 AWS DataSync
- 使用传入或传出 AWS 存储 AWS DataSync
- 使用其他云存储进行传入或传出其他云存储 AWS DataSync
- 创建一个传输数据的任务
- 启动传输数据的任务

我可以在哪里传输我的数据 AWS DataSync？

您可以将数据传输到何处 AWS DataSync 取决于以下因素：

- 传输的源位置和目标位置

- 如果你的位置不同 AWS 账户
- 如果你的位置不同 AWS 区域
- 如果您使用的是基本模式或增强模式

支持同一个传输 AWS 账户

DataSync 支持在以下与之关联的存储资源之间进行传输 AWS 账户。

来源	目标位置	需要代理吗？	支持的任务模式
• NFS • SMB	• Amazon S3	是	基本、增强
• NFS • SMB	• Amazon EFS • Amazon FSx	是	仅限基础版
• HDFS • 对象存储	• Amazon S3 • Amazon EFS • Amazon FSx	是	仅限基础版
• 其他云存储	• Amazon S3	仅适用于基本模式	基本、增强
• 其他云存储	• Amazon EFS • Amazon FSx	是	仅限基础版
• Amazon S3	• Amazon S3	否	基本、增强
• Amazon S3	• Amazon EFS • Amazon FSx	否	仅限基础版
• Amazon S3	• NFS • SMB	是	基本、增强
• Amazon S3	• HDFS • 对象存储	是	仅限基础版
• Amazon S3	• 其他云存储	仅适用于基本模式	基本、增强

来源	目标位置	需要代理吗？	支持的任务模式
• Amazon EFS • Amazon FSx	• NFS • SMB	是	仅限基础版
• Amazon EFS • Amazon FSx	• HDFS • 对象存储	是	仅限基础版
• Amazon EFS • Amazon FSx	• 其他云存储	是	仅限基础版
• Amazon EFS • Amazon FSx	• Amazon S3	否	仅限基础版
• Amazon EFS • Amazon FSx	• Amazon EFS • Amazon FSx	否	仅限基础版
• S3 on Outposts	• S3（英寸 AWS 区域）	是	仅限基础版
• Amazon S3（在 AWS 区域中）	• S3 on Outposts	是	仅限基础版

支持的跨境传输 AWS 账户

DataSync 支持在与不同存储资源关联的存储资源之间进行某些传输 AWS 账户。

来源	目标位置	需要代理吗？	支持的任务模式
• NFS • SMB	• Amazon S3	是	基本、增强
• HDFS • 对象存储	• Amazon S3	是	仅限基础版
• Amazon S3	• Amazon S3	否	基本、增强

来源	目标位置	需要代理吗？	支持的任务模式
• Amazon S3	• Amazon EFS • Amazon FSx	否	仅限基础版
• Amazon S3	• NFS • SMB	是	基本、增强
• Amazon S3	• HDFS • 对象存储	是	仅限基础版
• Amazon EFS • Amazon FSx	• Amazon S3	否	仅限基础版
• Amazon EFS[1] • FSx 适用于 OpenZFS 的亚马逊 1 • FSx 适用于 Windows 的亚马逊 File Server [2] • FSx 适用于 NetApp ONTAP 的 Amazon 3	• Amazon EFS • Amazon FSx	是（用作 NFS/SMB 位置时）	仅限基础版

[1] 配置为 NFS 位置。

[2] 配置为 SMB 位置。

[3] 配置为 NFS 或 SMB 位置。

支持同一个传输 AWS 区域

在同一 AWS 区域 中传输数据时没有任何限制（包括选择加入的区域）。有关更多信息，请参阅AWS 区域 支持 DataSync。

支持之间的传输 AWS 区域

在AWS 区域 支持的之间传输数据时，请注意以下几点 DataSync：

- 在不同 AWS 存储服务之间传输时 AWS 区域，两个位置中的一个必须位于您使用的区域 DataSync。
- 您无法使用 NFS、SMB、HDFS 或对象存储位置进行跨区域传输。在这种情况下，您的两个转账地点必须位于您激活 DataSync代理的同一区域。
- 使用 AWS GovCloud (US) 区域，您可以：
  - 在 AWS GovCloud （美国东部）和 AWS GovCloud （美国西部）地区之间转移。
  - 在某个 AWS GovCloud (US) 地区和商业 AWS 区域区域（例如美国东部（弗吉尼亚北部）之间转移。这种类型的传输在 Amazon EFS 或 Amazon FSx 文件系统之间传输时需要代理。

> **Important**
>
> 您需要为两者之间传输的数据付费 AWS 区域。这种传输按从源区域到目标位置区域的数据传出进行计费。有关更多信息，请参阅AWS DataSync 定价。

确定您的转账是否需要代 DataSync 理

根据您的转账情况，您可能需要代 DataSync 理。有关更多信息，请参阅 我需要 AWS DataSync 经纪人吗？。

使用本地存储与本地存储进行传入或传出本地存储 AWS DataSync

使用 AWS DataSync，您可以在多个本地或自行管理的存储系统与以下 AWS 存储服务之间传输文件和对象：

- Amazon S3
- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的亚马逊

主题

- 使用 NFS 文件服务器配置 AWS DataSync 传输

- 使用 SMB 文件服务器配置 AWS DataSync 传输
- 使用 HDFS 集群配置 AWS DataSync 传输
- 使用对象存储系统配置 DataSync 传输

使用 NFS 文件服务器配置 AWS DataSync 传输

使用 AWS DataSync，您可以在网络文件系统 (NFS) 文件服务器和以下 AWS 存储服务之间传输数据。支持的存储服务取决于您的任务模式，如下所示：

基本模式	增强模式
• Amazon S3 • Amazon EFS • FSx 适用于 Windows 文件服务器的亚马逊 • 亚马逊 f FSx or Lustre • FSx 适用于 OpenZFS 的亚马逊 • FSx 适用于 NetApp ONTAP 的亚马逊	• Amazon S3

要设置此类传输，您需要为 NFS 文件服务器创建一个位置。您可以将此位置用作传输源或传输目标位置。

提供 DataSync 对 NFS 文件服务器的访问权限

DataSync 要访问您的 NFS 文件服务器，您需要一个 DataSync 代理。代理使用 NFS 协议在您的文件服务器上装载导出。请务必使用与所需任务模式相对应的代理。

主题

- 配置 NFS 导出
- 支持的 NFS 版本

配置 NFS 导出

传输 DataSync 所需的导出取决于您的 NFS 文件服务器是源位置还是目标位置，以及文件服务器权限的配置方式。

如果您的文件服务器是源位置，则 DataSync 只需读取和遍历您的文件和文件夹即可。如果是目标位置，则 DataSync 需要 root 访问权限才能写入该位置，并对要复制的文件和文件夹设置所有权、权限和其他元数据。您可以使用 `no_root_squash` 选项为导出提供根访问权限。

以下示例介绍如何配置提供访问权限的 NFS 导出。 DataSync

当您的 NFS 文件服务器是源位置时（根访问权限）

使用以下命令配置导出，该命令提供 DataSync只读权限 (`ro`) 和 root 访问权限 (`no_root_squash`)：

```
export-path datasync-agent-ip-address(ro,no_root_squash)
```

当您的 NFS 文件服务器是目标位置时

使用以下命令配置导出，该命令提供 DataSync写入权限 (`rw`) 和 root 访问权限 (`no_root_squash`)：

```
export-path datasync-agent-ip-address(rw,no_root_squash)
```

当您的 NFS 文件服务器是源位置时（没有根访问权限）

使用以下命令配置您的导出，该命令指定您知道将为导出提供 DataSync只读权限的 POSIX 用户 ID (UID) 和组 ID (GID)：

```
export-path datasync-agent-ip-address(ro,all_squash,anonuid=uid,anongid=gid)
```

支持的 NFS 版本

默认情况下， DataSync 使用 NFS 版本 4.1。 DataSync 还支持 NFS 4.0 和 3.x。

为 NFS 传输配置网络

要进行 DataSync 传输，必须为几个网络连接配置流量：

1. 允许通过以下端口从 DataSync 代理到 NFS 文件服务器的流量：

   - 对于 NFS 版本 4.1 和 4.0：TCP 端口 2049
   - 对于 NFS 版本 3.x：TCP 端口 111 和 2049

   您网络中的其他 NFS 客户端应该能够装载您用来传输数据的 NFS 导出。无需 Kerberos 身份验证即可访问导出。

2. 为服务端点连接（如 VPC、公共端点或 FIPS 端点）配置流量。
3. 允许流量从 DataSync 服务传输到您要传输或传输的AWS 存储服务。

创建您的 NFS 传输位置

在开始之前，请注意以下事项：

- 您需要一个 NFS 文件服务器来传输数据。
- 您需要一个可以访问您的文件服务器的 DataSync 代理。
- DataSync 不支持复制 NFS 版本 4 访问控制列表 (ACLs)。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，请选择 网络文件系统 (NFS)。
4. 对于代理，请选择可以连接到 NFS 文件服务器的 DataSync 代理。

   可以选择多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。
5. 对于 NFS 服务器，请输入您的 DataSync代理所连接的 NFS 文件服务器的域名系统 (DNS) 名称或 IP 地址。
6. 在装载路径中，输入要 DataSync 装载的 NFS 导出路径。

   此路径（或路径的子目录）是传入或 DataSync 传出数据的地方。有关更多信息，请参阅 配置 NFS 导出。
7. （可选）展开 “其他设置”，然后选择 DataSync 要在访问文件服务器时使用的特定 NFS 版本。

   有关更多信息，请参阅 支持的 NFS 版本。
8. 或者，选择添加标签以标记您的 NFS 位置。

   标签 是帮助您管理、筛选和搜索位置的键值对。我们建议至少为您的位置创建一个名称标签。
9. 选择创建位置。

使用 AWS CLI

- 使用以下命令创建 NFS 位置。

```
aws datasync create-location-nfs \
    --server-hostname nfs-server-address \
    --on-prem-config AgentArns=datasync-agent-arns \
    --subdirectory nfs-export-path
```

有关创建位置的更多信息，请参阅 提供 DataSync 对 NFS 文件服务器的访问权限。

DataSync 自动选择用于从 NFS 位置读取的 NFS 版本。要指定 NFS 版本，请在 NfsMountOptions API 操作中使用可选 `Version` 参数。

此命令返回 NFS 位置的 Amazon 资源名称 (ARN)，与以下所示的 ARN 类似。

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:111222333444:location/
loc-0f01451b140b2af49"
}
```

要确保可以挂载目录，您可以连接到与代理具有相同网络配置的任何计算机，并运行以下命令。

```
mount -t nfs -o nfsvers=<nfs-server-version <nfs-server-address:<nfs-export-path <test-
folder
```

以下是 命令的示例。

```
mount -t nfs -o nfsvers=3 198.51.100.123:/path_for_sync_to_read_from /
temp_folder_to_test_mount_on_local_machine
```

使用 SMB 文件服务器配置 AWS DataSync 传输

使用 AWS DataSync，您可以在服务器消息块 (SMB) 文件服务器和以下 AWS 存储服务之间传输数据。支持的存储服务取决于您的任务模式，如下所示：

基本模式	增强模式
• Amazon S3 • Amazon EFS • FSx 适用于 Windows 文件服务器的亚马逊	• Amazon S3

基本模式	增强模式
• 亚马逊 f FSx or Lustre • FSx 适用于 OpenZFS 的亚马逊 • FSx 适用于 NetApp ONTAP 的亚马逊	

要设置此类传输，您需要为 SMB 文件服务器创建一个位置。您可以将此位置用作传输源或传输目标位置。请务必使用与所需任务模式相对应的代理。

提供 DataSync 对 SMB 文件服务器的访问权限

DataSync 使用 SMB 协议连接到您的文件服务器，并且可以使用 NTLM 或 Kerberos 进行身份验证。

主题

- 支持的 SMB 版本
- 使用 NTLM 身份验证
- 使用 Kerberos 身份验证
- 所需的权限
- DFS 命名空间

支持的 SMB 版本

默认情况下，根据与 SMB 文件服务器的协商 DataSync 自动选择 SMB 协议的版本。

您也可以配置 DataSync 为使用特定的 SMB 版本，但我们建议只有 DataSync 在无法自动与 SMB 文件服务器协商时才这样做。 DataSync 支持 SMB 版本 1.0 及更高版本。出于安全起见，建议使用 SMB 版本 3.0.2 或更高版本。早期版本（例如 SMB 1.0）包含已知的安全漏洞，攻击者可以借此来窃取您的数据。

有关 DataSync控制台和 API 中的选项列表，请参阅下表：

控制台选项	API 选项	说明
自动	`AUTOMATIC`	DataSync 和 SMB 文件服务器协商它们在 2.1 和 3.1.1 之间相互支持的最高版本的 SMB。

控制台选项	API 选项	说明
		这是默认选项，也是推荐的选项。如果您改为选择文件服务器不支持的特定版本，则可能会出现 `Operation Not Supported` 错误。
SMB 3.0.2	`SMB3`	将协议协商限制为仅 SMB 版本 3.0.2。
SMB 2.1	`SMB2`	将协议协商限制为仅 SMB 版本 2.1。
SMB 2.0	`SMB2_0`	将协议协商限制为仅 SMB 版本 2.0。
SMB 1.0	`SMB1`	将协议协商限制为仅 SMB 版本 1.0。

使用 NTLM 身份验证

要使用 NTLM 身份验证，您需要提供一个用户名和密码，允许您访问 DataSync 要传输到或从中传输的 SMB 文件服务器。该用户可以是文件服务器上的本地用户，也可以是 Microsoft Active Directory 中的域用户。

使用 Kerberos 身份验证

要使用 Kerberos 身份验证，您需要提供 Kerberos 主体、Kerberos 密钥表 (keytab) 文件和 Kerberos 配置文件，该文件允许访问 DataSync 要传输到或从中传输的 SMB 文件服务器。

主题

- 先决条件
- DataSync Kerberos 的配置选项

先决条件

您需要创建几个 Kerberos 工件并配置您的网络，以便 DataSync 可以访问您的 SMB 文件服务器。

- 使用 ktpass 或 kutil 实用程序创建 Kerberos 的 keytab 文件。

  以下示例使用 `ktpass` 创建 keytab 文件。您指定的 Kerberos 领域 (`MYDOMAIN.ORG`) 必须采用大写字母。

```
ktpass /out C:\YOUR_KEYTAB.keytab /princ HOST/kerberosuser@MYDOMAIN.ORG /mapuser
 kerberosuser /pass * /crypto AES256-SHA1 /ptype KRB5_NT_PRINCIPAL
```

- 准备 Kerberos 配置文件 (`krb5.conf`) 的简化版本。包括有关领域、域管理服务器位置以及主机名映射到 Kerberos 领域的信息。

  验证 `krb5.conf` 内容中领域和域名的大小写格式是否正确。例如：

```
[libdefaults]
  dns_lookup_realm = true
  dns_lookup_kdc = true
  forwardable = true
  default_realm = MYDOMAIN.ORG

[realms]
  MYDOMAIN.ORG = {
    kdc = mydomain.org
    admin_server = mydomain.org
  }

[domain_realm]
  .mydomain.org = MYDOMAIN.ORG
  mydomain.org = MYDOMAIN.ORG
```

- 在网络配置中，确保 Kerberos 密钥分配中心（KDC）服务器端口已打开。KDC 端口通常为 TCP 端口 88。

DataSync Kerberos 的配置选项

在创建使用 Kerberos 的 SMB 位置时，需配置以下选项。

控制台选项	API 选项	说明
中小型企业服务器	`ServerHostName`	您的 DataSync代理将挂载的 SMB 文件服务器的域名。对于 Kerberos，无法指定文件服务器的 IP 地址。

控制台选项	API 选项	说明
Kerberos 主体	`KerberosPrincipal`	Kerberos 领域中有权访问 SMB 文件服务器中文件、文件夹和文件元数据的身份。 Kerberos 主体可能看起来像 `HOST/kerberosuser@MYDOMAIN.ORG` 。 主体名称区分大小写。
Keytab 文件	`KerberosKeytab`	Kerberos 密钥表（keytab）文件，其中包括 Kerberos 主体和加密密钥之间的映射。
Kerberos 配置文件	`KerberosKrbConf`	定义 Kerberos 领域配置的 `krb5.conf` 文件。
DNS IP 地址（可选）	`DnsIpAddresses`	您的 SMB 文件服务器所属的 DNS 服务器 IPv4 的地址。 如果您的环境中有多个域，则配置此域可确保 DataSync 连接到正确的 SMB 文件服务器。

所需的权限

您提供的身份 DataSync 必须具有装载和访问您的 SMB 文件服务器的文件、文件夹和文件元数据的权限。

如果您在 Active Directory 中提供身份，则该身份必须是具有以下一种或两种用户权限的 Active Directory 组的成员（取决于[您 DataSync 要复制的元数据]）：

用户权限	说明
恢复文件和目录（`SE_RESTORE_NAME` ）	DataSync 允许复制对象所有权、权限、文件元数据和 NTFS 自由访问列表 () DACLs。

用户权限	说明
	此用户权限通常授予域管理员组和备份操作员组（均为默认 Active Directory 组）的成员。
管理审计和安全日志（SE_SECURITY_NAME）	DataSync 允许复制 NTFS 系统访问控制列表 (SACLs)。 此用户权限通常授予域管理员组的成员。

如果您要复制 Windows， ACLs 并且要在 SMB 文件服务器和另一个使用 SMB 的存储系统（例如适用于 Windows 文件服务器的 Amazon 或 FSx ONTAP）之间 FSx 进行传输，则您提供的身份 DataSync 必须属于同一 Active Directory 域或者它们的域之间有 Active Directory 信任关系。

DFS 命名空间

DataSync 不支持 Microsoft 分布式文件系统 (DFS) 命名空间。我们建议您在创建 DataSync 位置时改为指定底层文件服务器或共享。

创建您的 SMB 传输位置

在开始之前，您需要一个要从中传输数据的 SMB 文件服务器。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，选择 Server Message Block (SMB) (服务器消息块 (SMB))。

   您可在稍后将此位置配置为源或目标。
4. 对于代理，请选择可以连接到 SMB 文件服务器的 DataSync 代理。

   可以选择多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。
5. 对于 SMB 服务器，请输入您的 DataSync 代理将挂载的 SMB 文件服务器的域名或 IP 地址。

   对于此设置，请记住以下几点：

   - 您无法指定 IP 版本 6 (IPv6) 地址。
   - 如果使用的是 Kerberos 身份验证，则必须指定域名。

6. 在“共享名称”中，输入 SMB 文件服务器导出的共享名称，用于读 DataSync 取或写入数据。

   您可以在共享路径中包含子目录（例如，`/path/to/subdirectory`）。确保网络中的其他 SMB 客户端也可以挂载此路径。

   要复制子目录中的所有数据， DataSync 必须能够装载 SMB 共享并访问其所有数据。有关更多信息，请参阅 所需的权限。

7. （可选）展开“其他设置”，然后选择 DataSync 要在访问文件服务器时使用的 SMB 版本。

   默认情况下，根据与 SMB 文件服务器的协商 DataSync 自动选择版本。有关信息，请参阅支持的 SMB 版本。

8. 对于身份验证类型，选择 NTLM 或 Kerberos。

9. 根据您的身份验证类型，执行以下操作之一：

   NTLM

   - 对于 用户，输入可装载 SMB 文件服务器并有权访问传输中涉及文件和文件夹的用户名。

     有关更多信息，请参阅 所需的权限。

   - 对于密码，输入可以挂载 SMB 文件服务器并有权访问传输过程中涉及的文件和文件夹的用户的密码。

   - 或者，对于 域，输入 SMB 文件服务器所属的 Windows 域名。

     如果您的环境中有多个域，则配置此设置 DataSync 可确保连接到正确的 SMB 文件服务器。

   Kerberos

   - 对于 Kerberos 主体，请指定 Kerberos 领域中有权访问 SMB 文件服务器中文件、文件夹和文件元数据的主体。

     Kerberos 主体可能看起来像 `HOST/kerberosuser@MYDOMAIN.ORG`。

     主体名称区分大小写。如果您为此设置指定的主体与您用于创建 keytab 文件的主体不完全匹配，则 DataSync任务执行将失败。

   - 对于 Keytab 文件，请上传一个包含 Kerberos 主体和加密密钥之间映射关系的 keytab 文件。

   - 对于 Kerberos 配置文件，请上传一个定义 Kerberos 领域配置的 `krb5.conf` 文件。

- （可选）对于 DNS IP 地址，请为 SMB 文件服务器所属的 DNS 服务器最多指定两个 IPv4 地址。

  如果您的环境中有多个域，则配置此参数 DataSync 可确保连接到正确的 SMB 文件服务器。

10. 或者，选择 添加标签以标记您的 SMB 位置。

    标签 是帮助您管理、筛选和搜索位置的键值对。我们建议至少为您的位置创建一个名称标签。

11. 选择创建位置。

使用 AWS CLI

以下说明介绍如何使用 NTLM 或 Kerberos 身份验证来创建 SMB 位置。

NTLM

1. 复制以下 `create-location-smb` 命令。

```
aws datasync create-location-smb \
    --agent-arns datasync-agent-arns \
    --server-hostname smb-server-address \
    --subdirectory smb-export-path \
    --authentication-type "NTLM" \
    --user user-who-can-mount-share \
    --password user-password \
    --domain windows-domain-of-smb-server
```

2. 对于`--agent-arns`，请指定可以连接到 SMB 文件服务器的 DataSync 代理。

   可以选择多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。

3. 对于`--server-hostname`，请指定要装载 DataSync 代理的 SMB 文件服务器的域名或 IPv4 地址。

4. 对于`--subdirectory`，请指定 SMB 文件服务器导出的共享的名称，该共享 DataSync将在其中读取或写入数据。

   您可以在共享路径中包含子目录（例如，`/path/to/subdirectory`）。确保网络中的其他 SMB 客户端也可以挂载此路径。

   要复制子目录中的所有数据， DataSync 必须能够装载 SMB 共享并访问其所有数据。有关更多信息，请参阅 所需的权限。

5. 对于 `--user`，指定可挂载 SMB 文件服务器并有权访问传输中涉及文件和文件夹的用户名。

   有关更多信息，请参阅 所需的权限。

6. 对于 `--password`，指定可以挂载 SMB 文件服务器并有权限访问传输中涉及的文件和文件夹的用户的密码。

7. （可选）对于 `--domain`，指定 SMB 文件服务器所属的 Windows 域名。

   如果您的环境中有多个域，则配置此设置 DataSync 可确保连接到正确的 SMB 文件服务器。

8. （可选）如果 DataSync 要使用特定的 SMB 版本，请添加该`--version`选项。有关更多信息，请参阅 支持的 SMB 版本。

9. 运行 `create-location-smb` 命令。

   如果命令成功，您将收到一条响应，显示您创建位置的 ARN。例如：

```
{
    "arn:aws:datasync:us-east-1:123456789012:location/loc-01234567890example"
}
```

Kerberos

1. 复制以下 `create-location-smb` 命令。

```
aws datasync create-location-smb \
    --agent-arns datasync-agent-arns \
    --server-hostname smb-server-address \
    --subdirectory smb-export-path \
    --authentication-type "KERBEROS" \
    --kerberos-principal "HOST/kerberosuser@EXAMPLE.COM" \
    --kerberos-keytab "fileb://path/to/file.keytab" \
    --kerberos-krb5-conf "file://path/to/krb5.conf" \
    --dns-ip-addresses array-of-ipv4-addresses
```

2. 对于`--agent-arns`，请指定可以连接到 SMB 文件服务器的 DataSync 代理。

   可以选择多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。

3. 对于`--server-hostname`，请指定您的 DataSync 代理将装载的 SMB 文件服务器的域名。

4. 对于`--subdirectory`，请指定 SMB 文件服务器导出的共享的名称，该共享 DataSync将在其中读取或写入数据。

您可以在共享路径中包含子目录（例如，`/path/to/subdirectory`）。确保网络中的其他 SMB 客户端也可以挂载此路径。

要复制子目录中的所有数据，DataSync 必须能够装载 SMB 共享并访问其所有数据。有关更多信息，请参阅 所需的权限。

5. 对于 Kerberos 选项，执行下列操作：

   - `--kerberos-principal`：指定 Kerberos 领域中有权访问 SMB 文件服务器中文件、文件夹和文件元数据的主体。

     Kerberos 主体可能看起来像 `HOST/kerberosuser@MYDOMAIN.ORG`。

     主体名称区分大小写。如果您为此选项指定的主体与您用于创建密钥表文件的主体不完全匹配，则 DataSync任务执行将失败。

   - `--kerberos-keytab`：指定一个包含 Kerberos 主体和加密密钥之间映射关系的 keytab 文件。

   - `--kerberos-krb5-conf`：指定一个定义 Kerberos 领域配置的 `krb5.conf` 文件。

   - （可选）`--dns-ip-addresses`：为您的 SMB 文件服务器所属的 DNS 服务器最多指定两个 IPv4 地址。

     如果您的环境中有多个域，则配置此参数 DataSync 可确保连接到正确的 SMB 文件服务器。

6. （可选）如果 DataSync 要使用特定的 SMB 版本，请添加该`--version`选项。有关更多信息，请参阅 支持的 SMB 版本。

7. 运行 `create-location-smb` 命令。

   如果命令成功，您将收到一条响应，显示您创建位置的 ARN。例如：

```
{
    "arn:aws:datasync:us-east-1:123456789012:location/loc-01234567890example"
}
```

使用 HDFS 集群配置 AWS DataSync 传输

使用 AWS DataSync，您可以使用基本模式任务在 Hadoop 分布式文件系统 (HDFS) 集群和以下 AWS 存储服务之一之间传输数据：

- Amazon S3
- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的亚马逊

要设置此类传输，您需要为 HDFS 集群创建一个位置。您可以将此位置用作传输源或传输目标位置。

提供 DataSync 对 HDFS 集群的访问权限

要连接到 HDFS 集群，请 DataSync 使用尽可能靠近 HDFS 集群部署的基本模式代理代理。该 DataSync代理充当 HDFS 客户端，与集群 DataNodes 中的 NameNodes 和进行通信。

启动传输任务时，会 DataSync NameNode 查询群集上文件和文件夹的位置。如果您将 HDFS 位置配置为源位置，则会从集群 DataNodes 中 DataSync 读取文件和文件夹数据，并将这些数据复制到目标。如果您将 HDFS 位置配置为目标位置，则 DataSync 会将文件和文件夹从源位置写入集群 DataNodes 中。

身份验证

连接到 HDFS 集群时， DataSync 支持简单身份验证或 Kerberos 身份验证。要使用简单身份验证，请提供对 HDFS 集群具有读写权限的用户的用户名。要使用 Kerberos 身份验证，请提供 Kerberos 配置文件、Kerberos 密钥表（密钥表）文件和 Kerberos 主体名称。Kerberos 主体的凭证必须位于提供的密钥表文件中。

加密

使用 Kerberos 身份验证时， DataSync 支持在 DataSync 代理和 HDFS 集群之间传输的数据时对其进行加密。使用 HDFS 集群上的保护质量 (QOP) 配置设置，并在创建 HDFS 位置时指定 QOP 设置，从而加密数据。QOP 配置包括数据传输保护和远程过程调用 (RPC) 保护设置。

DataSync 支持以下 Kerberos 加密类型：

- `des-cbc-crc`
- `des-cbc-md4`
- `des-cbc-md5`

- des3-cbc-sha1
- arcfour-hmac
- arcfour-hmac-exp
- aes128-cts-hmac-sha1-96
- aes256-cts-hmac-sha1-96
- aes128-cts-hmac-sha256-128
- aes256-cts-hmac-sha384-192
- camellia128-cts-cmac
- camellia256-cts-cmac

您还可以使用透明数据加密 (TDE) 将 HDFS 集群配置为静态加密。使用简单身份验证时，对启用 TDE 的集群进行 DataSync 读取和写入。如果您使用将数据复制 DataSync 到启用 TDE 的集群，请先在 HDFS 集群上配置加密区域。 DataSync 不创建加密区域。

不支持的 HDFS 功能

目前不支持以下 HDFS 功能： DataSync

- 使用 Kerberos 身份验证时的透明数据加密 (TDE)
- 配置多个 NameNodes
- 通过 HTTPS 的 Hadoop HDFS (httpFS)
- POSIX 访问控制列表 () ACLs
- HDFS 扩展属性 (xattrs)
- 使用 Apache 的 HDFS 集群 HBase

创建您的 HDFS 传输位置

您可以使用您的位置作为 DataSync转账的来源或目的地。

开始之前：通过执行以下操作验证代理与 Hadoop 集群之间的网络连接：

- 测试对 本地、自我管理和其他云存储的网络要求 中列出的 TCP 端口的访问权限。
- 测试本地代理与 Hadoop 集群之间的访问权限。有关说明，请参阅验证代理与存储系统的连接。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，选择 Hadoop Distributed File System (HDFS)。

   您可在稍后将此位置配置为源或目标。
4. 对于代理，请选择可以连接到 HDFS 集群的代理。

   可以选择多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。
5. 对于 NameNode，请提供 HDFS 群集的主 NameNode群集的域名或 IP 地址。
6. 在 “文件夹” 中，输入 HDFS 集群上 DataSync 要用于数据传输的文件夹。

   如果您的 HDFS 位置是源，请将此文件夹中的文件 DataSync 复制到目标。如果您的位置是目的地，则 DataSync会将文件写入此文件夹。
7. 要设置 区块大小或 复制因子，请选择 其他设置。

   默认块大小为 128MiB。您提供的块大小必须是 512 字节的倍数。

   传输到 HDFS 集群 DataNodes 时，默认重复因子为三。
8. 在 安全部分中，选择 HDFS 群集上使用的 身份验证类型。

   - 简单 - 对于 用户，在 HDFS 集群上指定具有以下权限的用户名（取决于您的用例）：
     - 如果您计划将此位置用作源位置，请指定仅具有读取权限的用户。
     - 如果您计划将此位置用作目标位置，请指定具有读写权限的用户。

     您可以选择为 HDFS 集群的密钥管理服务器（KMS）指定 URI。
   - Kerberos – 指定有权访问您的 HDFS 集群的 Kerberos 主体。接下来，提供包含所提供的 Kerberos 主体的KeyTab文件。然后，提供 Kerberos 配置文件。最后，在 RPC 保护和 数据传输保护下拉列表中指定传输中加密保护的类型。
9. 或者，选择 添加标签来标记您的 HDFS 位置。

   标签 是帮助您管理、筛选和搜索位置的键值对。我们建议至少为您的位置创建一个名称标签。
10. 选择创建位置。

使用 AWS CLI

1. 复制以下 `create-location-hdfs` 命令。

```
aws datasync create-location-hdfs --name-nodes [{"Hostname":"host1", "Port": 8020}]
 \
    --authentication-type "SIMPLE|KERBEROS" \
    --agent-arns [arn:aws:datasync:us-east-1:123456789012:agent/
agent-01234567890example] \
    --subdirectory "/path/to/my/data"
```

2. 在`--name-nodes`参数中，指定 HDFS 集群的主集群的主机名或 IP 地址 NameNode 以及 NameNode 正在监听的 TCP 端口。

3. 对于 `--authentication-type` 参数，指定连接 Hadoop 集群时使用的身份验证类型。您可以指定 `SIMPLE` 或 `KERBEROS`。

   如果您采用 `SIMPLE` 身份验证，请使用 `--simple-user` 参数指定用户的用户名。如果您采用 `KERBEROS` 身份验证，请使用 `--kerberos-principal`、`--kerberos-keytab` 和`--kerberos-krb5-conf` 参数。有关更多信息，请参阅 create-location-hdfs。

4. 对于`--agent-arns`参数，请指定可以连接到您的 HDFS DataSync 集群的代理的 ARN。

   可以选择多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。

5. （可选）在`--subdirectory`参数中，在 HDFS 集群上指定 DataSync 要用于数据传输的文件夹。

   如果您的 HDFS 位置是源，请将此文件夹中的文件 DataSync 复制到目标。如果您的位置是目的地，则 DataSync会将文件写入此文件夹。

6. 运行 `create-location-hdfs` 命令。

   如果命令成功，您将收到一条响应，显示您创建位置的 ARN。例如：

```
{
    "arn:aws:datasync:us-east-1:123456789012:location/loc-01234567890example"
}
```

使用对象存储系统配置 DataSync 传输

使用 AWS DataSync，您可以使用基本模式任务在对象存储系统和以下 AWS 存储服务之一之间传输数据：

- Amazon S3
- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的亚马逊

要设置此类传输，您需要为对象存储系统创建一个位置。您可以将此位置用作传输源或传输目标位置。在本地对象存储之间传输数据需要使用基本模式 DataSync 代理。

先决条件

您的对象存储系统必须与以下 Amazon S3 API 操作兼容 DataSync 才能连接到该系统：

- `AbortMultipartUpload`
- `CompleteMultipartUpload`
- `CopyObject`
- `CreateMultipartUpload`
- `DeleteObject`
- `DeleteObjects`
- `DeleteObjectTagging`
- `GetBucketLocation`
- `GetObject`
- `GetObjectTagging`
- `HeadBucket`
- `HeadObject`
- `ListObjectsV2`

- `PutObject`
- `PutObjectTagging`
- `UploadPart`

创建您的对象存储传输位置

在开始之前，您需要拥有计划向其传入数据或从中传出数据的对象存储系统。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，选择 对象存储。

   您可在稍后将此位置配置为源或目标。
4. 对于 服务器，提供对象存储服务器的域名或 IP 地址。
5. 对于 存储桶名称，输入传输中涉及的对象存储桶的名称。
6. 对于 文件夹，请输入对象前缀。

   DataSync 仅复制带有此前缀的对象。
7. 如果传输需要代理，请选择使用代理，然后选择连接到您的对象存储系统的 DataSync 代理。

   有些传输不需要代理。在其他情况中，可能需要使用多个代理。有关更多信息，请参阅不需要 DataSync 代理的情况和使用多个 DataSync 代理。
8. 要配置与对象存储服务器的连接，请展开 其他设置并执行以下操作：

   a. 对于 服务器协议，选择 HTTP 或 HTTPS。

   b. 对于 服务器端口，请使用默认端口（HTTP 为 80，HTTPS 为 443），或者根据需要指定自定义端口。

   c. 对于证书，如果对象存储系统使用私有或自签名证书颁发机构（CA），请选择选择文件，并指定带有完整证书链的单个 `.pem` 文件。

      证书链可能包括：

      - 对象存储系统的证书
      - 所有中间证书（如果有）

- 签名 CA 的根证书

  您可以将证书串联到一个 .pem 文件中（在进行 base64 编码前，文件最大长度可以为 32768 字节）。以下示例 cat 命令创建了一个包含三个证书的 *object_storage_certificates*.pem 文件：

  ```
  cat object_server_certificate.pem intermediate_certificate.pem ca_root_certificate.pem
   > object_storage_certificates.pem
  ```

9. 如果需要凭证才能访问对象存储服务器，请选择需要凭证，然后输入用于访问存储桶的访问密钥。然后，要么直接输入密钥，要么指定包含该 AWS Secrets Manager 密钥的密钥。有关更多信息，请参阅为存储位置提供凭证。

   访问密钥和密钥可以分别是用户名和密码。

10. 或者，选择 添加标签以标记您的对象存储位置。

    标签 是帮助您管理、筛选和搜索位置的键值对。我们建议至少为您的位置创建一个名称标签。

11. 选择创建位置。

使用 AWS CLI

1. 复制以下 create-location-object-storage 命令：

   ```
   aws datasync create-location-object-storage \
       --server-hostname object-storage-server.example.com \
       --bucket-name your-bucket \
       --agent-arns arn:aws:datasync:us-east-1:123456789012:agent/
   agent-01234567890deadfb
   ```

2. 在命令中指定以下必需的参数：

   - --server-hostname – 指定您的对象存储服务器的域名或 IP 地址。
   - --bucket-name — 指定您要向其传输或从其传输的对象存储服务器上的存储桶的名称。

3. 或者，在命令中添加以下任意参数：

   - --agent-arns— 指定要连接到对象存储服务器的 DataSync代理。
   - --server-port – 指定对象存储服务器接受入站网络流量的端口（例如，端口 443）。
   - --server-protocol – 指定对象存储服务器通信时使用的协议（HTTP 或 HTTPS）。

- `--access-key` – 如果需要凭证来向对象存储服务器进行身份验证，则指定访问密钥（例如，用户名）。
- `--secret-key` – 如果需要凭证来向对象存储服务器进行身份验证，则指定私有密钥（例如，密码）。

  您还可以使用 AWS Secrets Manager提供其他参数来保护密钥。有关更多信息，请参阅为存储位置提供凭证。
- `--server-certificate`— 如果您的对象存储系统使用私有或自签名证书颁发机构 (CA)，则指定一个证书链， DataSync 以便在对象存储系统中进行身份验证。您必须指定具有完整证书链的单个 `.pem` 文件（例如，`file:///home/user/.ssh/object_storage_certificates.pem`）。

  证书链可能包括：

  - 对象存储系统的证书
  - 所有中间证书（如果有）
  - 签名 CA 的根证书

  您可以将证书串联到一个 `.pem` 文件中（在进行 base64 编码前，文件最大长度可以为 32768 字节）。以下示例 `cat` 命令创建了一个包含三个证书的 *`object_storage_certificates`*`.pem` 文件：

  ```
  cat object_server_certificate.pem intermediate_certificate.pem ca_root_certificate.pem
   > object_storage_certificates.pem
  ```
- `--subdirectory` – 指定对象存储服务器的对象前缀。

  DataSync 仅复制带有此前缀的对象。
- `--tags` – 指定表示要添加到位置资源的标签的键值对。

  标签可帮助您管理、筛选和搜索资源。我们建议为您的位置创建一个名称标签。

4. 运行 `create-location-object-storage` 命令。

   您会收到一条回复，表明您刚刚创建的位置 ARN。

   ```
   {
       "LocationArn": "arn:aws:datasync:us-east-1:123456789012:location/
   loc-01234567890abcdef"
   }
   ```

使用传入或传出 AWS 存储 AWS DataSync

使用 AWS DataSync，您可以将数据传输到多个存储服务或从多个 AWS 存储服务中传输数据。有关更多信息，请参阅我可以在哪里传输数据 DataSync？

主题

- 使用 Amazon S3 配置 AWS DataSync 传输
- 使用 Amazon EFS 配置 AWS DataSync 传输
- 使用为 Window FSx s 文件服务器配置传输
- 为 Lustr FSx e 配置 DataSync 传输
- 为 OpenZFS 配置与亚马逊 FSx 的 DataSync 转账
- 在 AMAZON 上 FSx 为 NetApp ONTAP 配置传输

使用 Amazon S3 配置 AWS DataSync 传输

要将数据传输到您的 Amazon S3 存储桶或从中传输数据，您需要创建一个 AWS DataSync 传输位置。 DataSync 可以将此位置用作传输数据的源或目的地。

提供 DataSync 对 S3 存储桶的访问权限

DataSync 需要访问您要传入或传出的 S3 存储桶。为此，您必须创建一个具有访问存储桶所需权限的 AWS Identity and Access Management (IAM) 角色。 DataSync 然后，您可以在为创建 Amazon S3 营业地点时指定此角色 DataSync。

目录

- 所需的权限
- 为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置
- 使用服务器端加密访问 S3 存储桶
- 访问受限的 S3 存储桶
- 通过受限 VPC 的访问权限访问 S3 存储库

所需的权限

您的 IAM 角色所需的权限可能取决于存储桶是 DataSync 源位置还是目标位置。Outposts 上的 Amazon S3 需要一组不同的权限。

Amazon S3 (source location)

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "s3:GetBucketLocation",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:s3:::amzn-s3-demo-bucket"
        },
        {
            "Action": [
                "s3:GetObject",
                "s3:GetObjectTagging",
                "s3:GetObjectVersion",
                "s3:GetObjectVersionTagging",
                "s3:ListMultipartUploadParts"
              ],
            "Effect": "Allow",
            "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/*"
        }
    ]
}
```

Amazon S3 (destination location)

```
{
 "Version": "2012-10-17",
 "Statement": [
     {
         "Action": [
             "s3:GetBucketLocation",
             "s3:ListBucket",
             "s3:ListBucketMultipartUploads"
         ],
         "Effect": "Allow",
         "Resource": "arn:aws:s3:::amzn-s3-demo-bucket",
         "Condition": {
             "StringEquals": {
```

```
                "aws:ResourceAccount": "123456789012"
            }
        }
    },
    {
        "Action": [
            "s3:AbortMultipartUpload",
            "s3:DeleteObject",
            "s3:GetObject",
            "s3:GetObjectTagging",
            "s3:GetObjectVersion",
            "s3:GetObjectVersionTagging",
            "s3:ListMultipartUploadParts",
            "s3:PutObject",
            "s3:PutObjectTagging"
          ],
        "Effect": "Allow",
        "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/*",
        "Condition": {
            "StringEquals": {
                "aws:ResourceAccount": "123456789012"
            }
        }
    }
 ]
}
```

Amazon S3 on Outposts

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Action": [
                "s3-outposts:ListBucket",
                "s3-outposts:ListBucketMultipartUploads"
            ],
            "Effect": "Allow",
            "Resource": [
            "arn:aws:s3-outposts:us-east-1:123456789012:outpost/outpost-id/
bucket/amzn-s3-demo-bucket",
    "arn:aws:s3-outposts:us-east-1:123456789012:outpost/outpost-id/
accesspoint/bucket-access-point-name"
```

```
            ]
        },
        {
            "Action": [
                "s3-outposts:AbortMultipartUpload",
                "s3-outposts:DeleteObject",
                "s3-outposts:GetObject",
                "s3-outposts:GetObjectTagging",
                "s3-outposts:GetObjectVersion",
                "s3-outposts:GetObjectVersionTagging",
                "s3-outposts:ListMultipartUploadParts",
                "s3-outposts:PutObject",
                "s3-outposts:PutObjectTagging"
            ],
            "Effect": "Allow",
            "Resource": [
            "arn:aws:s3-outposts:us-east-1:123456789012:outpost/outpost-id/
bucket/amzn-s3-demo-bucket/*",
    "arn:aws:s3-outposts:us-east-1:123456789012:outpost/outpost-id/
accesspoint/bucket-access-point-name/*"
            ]
        },
        {
            "Action": "s3-outposts:GetAccessPoint",
            "Effect": "Allow",
            "Resource": "arn:aws:s3-outposts:us-
east-1:123456789012:outpost/outpost-id/accesspoint/bucket-access-point-name"
        }
    ]
}
```

为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置

在控制台中创建您的 Amazon S3 位置时， DataSync 可以自动创建并代入一个 IAM 角色，该角色通常具有访问您的 S3 存储桶的正确权限。

在某些情况下，您可能需要手动创建此角色（例如，访问具有额外安全层的存储分区，或者在其他存储桶中向存储桶传送或从存储桶传出 AWS 账户）。

为手动创建 IAM 角色 DataSync

1. 使用 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。

2. 在左侧导航窗格的访问管理下，选择角色，然后选择创建角色。
3. 在选择可信实体页面中，为可信实体类型选择 AWS 服务。
4. 对于“用例”，DataSync在下拉列表中进行选择，然后选择DataSync。选择下一步。
5. 在添加权限页面上，选择下一步。输入角色名称，然后选择创建角色。
6. 在角色页面上，搜索您刚刚创建的角色并选择其名称。
7. 在角色的详情页面上，选择权限选项卡。选择添加权限，然后选择创建内联策略。
8. 选择 JSON 选项卡，然后在策略编辑器中添加访问存储桶所需的权限。
9. 选择下一步。为您的策略输入名称，然后选择创建策略。
10. （推荐）为防止出现跨服务混淆代理问题，请执行以下操作：

    a. 在角色的详情页面上，选择信任关系选项卡。选择编辑信任策略。

    b. 使用以下示例更新信任策略，其中包括 `aws:SourceArn` 和 `aws:SourceAccount` 全局条件上下文键：

    JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Principal": {
            "Service": "datasync.amazonaws.com"
        },
        "Action": "sts:AssumeRole",
        "Condition": {
            "StringEquals": {
            "aws:SourceAccount": "444455556666"
            },
            "ArnLike": {
            "aws:SourceArn": "arn:aws:datasync:us-east-1:444455556666:*"
            }
        }
      }
  ]
}
```

    c. 选择更新策略。

您可以在创建您的 Amazon S3 位置时指定此角色。

使用服务器端加密访问 S3 存储桶

DataSync 可以将数据传输到使用服务器端加密的 S3 存储桶或从中传输数据。存储桶使用的加密密钥类型可以决定您是否需要 DataSync 允许访问存储桶的自定义策略。

DataSync 与使用服务器端加密的 S3 存储桶一起使用时，请记住以下几点：

- 如果您的 S3 存储桶使用 AWS 托管密钥加密，则默认情况下，如果您的所有资源都在同一个存储桶中，则 DataSync 可以访问该存储桶的对象 AWS 账户。
- 如果您的 S3 存储桶使用客户托管 AWS Key Management Service (AWS KMS) 密钥 (SSE-KMS) 加密，则密钥的策略必须包含 DataSync 用于访问存储桶的 IAM 角色。
- 如果您的 S3 存储桶使用客户托管的 SSE-KMS 密钥和其他密钥进行加密，则 DataSync 需要获得访问另一个 AWS 账户存储桶的权限。 AWS 账户您可以通过以下步骤完成上述设置：
  - 在使用的 IAM 角色中，您必须 DataSync 使用该密钥的完全限定的 Amazon 资源名称 (ARN) 来指定跨账户存储桶的 SSE-KMS 密钥。该密钥 ARN 与您用来配置存储桶默认加密的密钥 ARN 相同。在这种情况下，您不能指定密钥 ID、别名名称或别名 ARN。

    下面是一个密钥 ARN 示例：

    ```
    arn:aws:kms:us-
    west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab
    ```

    有关在 IAM 策略声明中指定 KMS 密钥的更多信息，请参阅 AWS Key Management Service 开发人员指南。
  - 在 SSE-KMS 密钥策略中，指定使用的 IAM 角色。 DataSync
- 如果您的 S3 存储桶使用客户托管 AWS KMS 密钥 (DSSE-KMS) 进行加密以进行双层服务器端加密，则密钥的策略必须包括 DataSync 用于访问存储桶的 IAM 角色。（请记住，DSSE-KMS 不支持 S3 存储桶密钥，这可以降低 AWS KMS 请求成本。）
- 如果您的 S3 存储桶使用客户提供的加密密钥 (SSE-C) 进行加密，则 DataSync无法访问此存储桶。

示例：适用于 SSE-KMS 的密钥策略 DataSync

以下示例是客户托管的 SSE-KMS 密钥的密钥政策。此政策与使用服务器端加密的 S3 存储桶关联。

如果要使用此示例，请将以下值替换为您自己的值：

- *account-id*— 你的 AWS 账户。

- *admin-role-name*— 可以管理密钥的 IAM 角色的名称。
- *datasync-role-name*— 允许 DataSync 在访问存储桶时使用密钥的 IAM 角色的名称。

JSON

```
{
    "Id": "key-consolepolicy-3",
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Enable IAM Permissions",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::111122223333:root"
            },
            "Action": "kms:*",
            "Resource": "*"
        },
        {
            "Sid": "Allow access for Key Administrators",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::111122223333:role/admin-role-name"
            },
            "Action": [
                "kms:Create*",
                "kms:Describe*",
                "kms:Enable*",
                "kms:List*",
                "kms:Put*",
                "kms:Update*",
                "kms:Revoke*",
                "kms:Disable*",
                "kms:Get*",
                "kms:Delete*",
                "kms:TagResource",
                "kms:UntagResource",
                "kms:ScheduleKeyDeletion",
                "kms:CancelKeyDeletion"
            ],
            "Resource": "*"
        },
```

```
        {
            "Sid": "Allow use of the key",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::111122223333:role/datasync-role-name"
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*"
            ],
            "Resource": "*"
        }
    ]
}
```

访问受限的 S3 存储桶

如果您需要向通常拒绝所有访问权限的 S3 存储桶进行传输，则可以编辑存储桶策略， DataSync 使其只能在传输时访问该存储桶。

示例：允许基于 IAM 角色的访问

1. 复制以下 S3 存储桶策略。

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [{
        "Sid": "Deny-access-to-bucket",
        "Effect": "Deny",
        "Principal": "*",
        "Action": "s3:*",
        "Resource": [
            "arn:aws:s3:::amzn-s3-demo-bucket",
            "arn:aws:s3:::amzn-s3-demo-bucket/*"
        ],
        "Condition": {
            "StringNotLike": {
                "aws:userid": [
```

```
                    "datasync-iam-role-id:*",
                    "your-iam-role-id"
                ]
            }
        }
    }]
}
```

2. 在策略中，替换以下值：

   - *amzn-s3-demo-bucket*：指定受限 S3 存储桶的名称。
   - *datasync-iam-role-id*— 指定用于访问存储桶的 IAM 角色的 ID。 DataSync

     运行以下 AWS CLI 命令获取 IAM 角色 ID：

     ```
     aws iam get-role --role-name datasync-iam-role-name
     ```

     在输出中，查找 RoleId 值：

     ```
     "RoleId": "ANPAJ2UCCR6DPCEXAMPLE"
     ```
   - *your-iam-role-id*— 指定用于为存储桶创建 DataSync 位置的 IAM 角色的 ID。

     运行以下命令以创建 IAM 角色 ID：

     ```
     aws iam get-role --role-name your-iam-role-name
     ```

     在输出中，查找 RoleId 值：

     ```
     "RoleId": "AIDACKCEVSQ6C2EXAMPLE"
     ```
3. 将此策略添加到 S3 存储桶策略中。
4. 使用 DataSync 完受限存储桶后，请从存储桶策略中移除两个 IAM 角色的条件。

通过受限 VPC 的访问权限访问 S3 存储库

一个 Amazon S3 存储桶，用于限制对特定虚拟私有云 (VPC) 终端节点的访问或 VPCs拒绝 DataSync 向该存储桶传输或从该存储桶传输数据。要在这些情况下启用传输，您可以更新存储桶的策略，将您在 DataSync 位置中指定的 IAM 角色包括在内。

Option 1: Allowing access based on DataSync location role ARN

在 S3 存储桶策略中，您可以指定您的 DataSync 位置 IAM 角色的亚马逊资源名称 (ARN)。

以下示例是一个 S3 存储桶策略，该策略拒绝除两个 VPCs（vpc-1234567890abcdef0和vpc-abcdef01234567890）之外的所有存储桶的访问。但是，该策略还包括ArnNotLikeIfExists条件和 a ws: PrincipalArn 条件键，它们允许 DataSync 位置角色的 ARN 访问存储桶。

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Access-to-specific-VPCs-only",
            "Effect": "Deny",
            "Principal": "*",
            "Action": "s3:*",
            "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/*",
            "Condition": {
                "StringNotEqualsIfExists": {
                    "aws:SourceVpc": [
                        "vpc-1234567890abcdef0",
                        "vpc-abcdef01234567890"
                    ]
                },
                "ArnNotLikeIfExists": {
                    "aws:PrincipalArn": [
                        "arn:aws:iam::111122223333:role/datasync-location-role-name"
                    ]
                }
            }
        }
    ]
}
```

Option 2: Allowing access based on DataSync location role tag

在 S3 存储桶策略中，您可以指定附加到您的 DataSync 位置 IAM 角色的标签。

以下示例是一个 S3 存储桶策略，该策略拒绝除两个 VPCs（vpc-1234567890abcdef0和vpc-abcdef01234567890）之外的所有存储桶的访问。但是，该策略还包括StringNotEqualsIfExists条件和 a ws: PrincipalTag 条件键，它们允许委托人使用标签键exclude-from-vpc-restriction和值true。您可以通过指定附加到您的 DataSync 位置角色的标签来尝试在存储桶策略中使用类似的方法。

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Access-to-specific-VPCs-only",
            "Effect": "Deny",
            "Principal": "*",
            "Action": "s3:*",
            "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/*",
            "Condition": {
                "StringNotEqualsIfExists": {
                    "aws:SourceVpc": [
                        "vpc-1234567890abcdef0",
                        "vpc-abcdef01234567890"
                    ],
                    "aws:PrincipalTag/exclude-from-vpc-restriction": "true"
                }
            }
        }
    ]
}
```

Amazon S3 传输的存储类别注意事项

当 Amazon S3 是您的目标位置时， DataSync 可以将您的数据直接传输到特定的 Amazon S3 存储类别。

有些存储类别的行为会影响您的 Amazon S3 存储成本。使用可能会因覆盖、删除或检索对象而产生额外费用的存储类别时，更改对象数据或元数据会产生此类费用。有关更多信息，请参阅 Amazon S3 定价。

> **Important**
>
> 传输到 Amazon S3 目标位置的新对象使用您在创建位置时指定的存储类别进行存储。默认情况下，除非您将任务配置为传输所有数据，否则会 DataSync 保留目标位置中现有对象的存储类别。在这种情况下，创建位置时指定的存储类别适用于所有对象。

Amazon S3 存储类	注意事项
S3 标准	选择“S3 标准”将您经常访问的文件冗余存储在地理上分开的多个可用区中。如果您没有指定存储类别，则这是默认设置。
S3 智能分层	选择“S3 智能分层”，可通过自动将数据移动到最具成本效益的存储访问层来优化存储成本。 您需要为存储在 S3 智能分层存储类别中的每个对象按月支付相关费用。这项 Amazon S3 费用包括监控数据访问模式和在层之间移动对象。
S3 标准 - IA	选择“S3 标准 - IA”将您不常访问的对象冗余存储在地理上分开的多个可用区中。 存储在“S3 标准 - IA”存储类别中的对象可能会因覆盖、删除或检索而产生额外费用。考虑这些对象的更改频率，计划保留这些对象的时间以及需要访问的频率。更改对象数据或元数据等同于删除一个对象后再创建一个新对象来替换它。这将导致存储在“S3 标准 - IA”存储类别中的对象产生额外费用。 小于 128 KB 的对象小于“S3 标准 - IA”存储类别中每个对象的最小容量费用。这些对象存储在 S3 标准存储类别中。
S3 单区 - IA	选择“S3 单区 - IA”：将您不常访问的对象存储在单个可用区中。 存储在“S3 单区 - IA”存储类别中的对象可能会因覆盖、删除或检索而产生额外费用。考虑这些对象的更改频率，计划保留这些对象的时间以及需要访问的频率。更改对象数据或元数据等同于删除一个对象后再创建一个新对象来替换它。这会导致存储在“S3 单区 - IA”存储类别中的对象产生额外费用。 小于 128 KB 的对象小于“S3 单区 - IA”存储类别中每个对象的最小容量费用。这些对象存储在 S3 标准存储类别中。
S3 Glacier Instant Retrieval	选择“S3 Glacier 即时检索”可存档很少访问但需要在毫秒内检索的对象。 与“S3 标准-IA”存储类别相比，存储在“S3 Glacier 即时检索”存储类中的数据可以节省成本，同时享有与“S3 标准-IA”存储类别相同的延迟和吞吐量性能。“S3 Glacier 即时检索”的数据访问成本高于“S3 标准-IA”。

Amazon S3 存储类	注意事项
	存储在“S3 Glacier 即时检索”存储类中的对象可能会因覆盖、删除或检索而产生额外费用。考虑这些对象的更改频率，计划保留这些对象的时间以及需要访问的频率。更改对象数据或元数据等同于删除一个对象后再创建一个新对象来替换它。这会导致存储在“S3 Glacier 即时检索”存储类别中的对象产生额外费用。 小于 128 KB 的对象小于“S3 Glacier 即时检索”存储类别中每个对象的最小容量费用。这些对象存储在 S3 标准存储类别中。
S3 Glacier 灵活检索	选择“S3 Glacier 灵活检索”，获取更多活跃档案。 存储在“S3 Glacier 灵活检索”存储类中的对象可能会因覆盖、删除或检索而产生额外费用。考虑这些对象的更改频率，计划保留这些对象的时间以及需要访问的频率。更改对象数据或元数据等同于删除一个对象后再创建一个新对象来替换它。这会导致存储在“S3 Glacier 灵活检索”存储类别中的对象产生额外费用。 S3 Glacier 灵活检索存储类需要为每个存档对象提供 40 KB 的额外元数据。 DataSync 将小于 40 KB 的对象归入 S3 标准存储类别。 必须先恢复以该存储类别存档的对象，然后 DataSync 才能读取它们。有关信息，请参阅《Amazon S3 用户指南》中的使用已归档的对象。 使用“S3 Glacier 灵活检索”时，请选择仅验证传输的数据任务选项，以便在传输结束时比较数据和元数据校验和。不能对此存储类别使用验证目标中的所有数据选项，因为它需要检索目标中的所有现有对象。

Amazon S3 存储类	注意事项
S3 Glacier 深度存档	选择“S3 Glacier 深度存档”来存档文件，以提供长期数据留存和数字保留，每年可以在其中访问一次或两次数据。 存储在“S3 Glacier 深度存档”存储类中的对象可能会因覆盖、删除或检索而产生额外费用。考虑这些对象的更改频率，计划保留这些对象的时间以及需要访问的频率。更改对象数据或元数据等同于删除一个对象后再创建一个新对象来替换它。这会导致存储在“S3 Glacier 深度存档”存储类别中的对象产生额外费用。 S3 Glacier Deep Archive Deep Archive 存储类需要为每个存档对象提供 40 KB 的额外元数据。 DataSync 将小于 40 KB 的对象归入 S3 标准存储类别。 必须先恢复以该存储类别存档的对象，然后 DataSync 才能读取它们。有关信息，请参阅《Amazon S3 用户指南》中的使用已归档的对象。 使用“S3 Glacier 深度存档”时，请选择仅验证传输的数据任务选项，以便在传输结束时比较数据和元数据校验和。不能对此存储类别使用验证目标中的所有数据选项，因为它需要检索目标中的所有现有对象。
S3 Outposts	Amazon S3 on Outposts 的存储类别。

使用时评估 S3 请求成本 DataSync

在 Amazon S3 分支机构 DataSync中，您会产生与发出 S3 API 请求相关的费用。本节可帮助您了解这些请求的 DataSync 使用方式以及它们可能如何影响您的 Amazon S3 成本。

主题

- 发出的 S3 请求由 DataSync
- 成本注意事项

发出的 S3 请求由 DataSync

下表描述了您在向 Amazon S3 位置复制数据或从中复制数据时 DataSync 可能发出的 S3 请求。

S3 请求	怎么 DataSync 用
ListObjectV2	DataSync 对每个以正斜杠 (/) 结尾的对象发出至少一个LIST请求，以列出以该前缀开头的对象。此请求是在任务的准备阶段调用的。
HeadObject	DataSync 在任务的准备和验证阶段发出检索对象元数据的HEAD请求。每个对象可能有多个HEAD请求，具体取决于您希望 DataSync 如何验证其传输的数据的完整性。
GetObject	DataSync 在任务的传输阶段发出从对象读取数据的GET请求。对于大型对象可能会发出多个 GET 请求。
GetObjectTagging	如果您将任务配置为复制对象标签，则会在任务的准备和传输阶段 DataSync 发出这些GET请求以检查对象标签。
PutObject	DataSync 在任务的传输阶段，PUT请求在目标 S3 存储桶中创建对象和前缀。由于 DataSync 使用 Amazon S3 分段上传功能，因此可能会有多个针对大型对象的PUT请求。为了最大限度地降低您的存储成本，我们建议您使用生命周期配置来停止未完成的分段上传。
PutObjectTagging	如果您的源对象有标签，并且您将任务配置为复制对象标签， DataSync 则在传输这些标签时会PUT发出这些请求。
CopyObject	DataSync 仅当对象的元数据发生变化时，才会COPY请求创建该对象的副本。如果您最初使用未保留其元数据的其他服务或工具将数据复制到 S3 存储桶，则可能会发生这种情况。

成本注意事项

DataSync 每次运行任务时，都会在 S3 存储桶上发出 S3 请求。在某些情况下，这可能会导致费用增加。例如：

- 您经常将对象传入或传出 S3 存储桶。
- 您可能没有传输太多数据，但是您的 S3 存储桶里有很多对象。在这种情况下，您仍然可以看到高额费用，因为 DataSync 对存储桶的每个对象发出 S3 请求。
- 您在 S3 存储桶之间进行传输，在源存储桶和目标上发出 S3 请求 DataSync 也是如此。

为了帮助最大限度地降低与之相关的 S3 请求成本 DataSync，请考虑以下几点：

主题

- 我使用的是哪些 S3 存储类别？
- 我需要多久传输一次数据？

我使用的是哪些 S3 存储类别？

S3 请求费用可能因对象使用的 Amazon S3 存储类别而有所不同，特别是对于存储对象的存储类别（例如 S3 Glacier 即时检索、S3 Glacier 灵活检索和 S3 Glacier 深度存档）。

以下是一些使用存储类会影响您的 S3 请求费用的场景 DataSync：

- 每次运行任务时，都会 DataSync 发出检索对象元数据的HEAD请求。即使您没有移动任何对象，这些请求也会产生费用。这些请求对账单的影响程度取决于您的对象使用的存储类别以及 DataSync 扫描的对象数量。
- 如果您将对象移至 S3 Glacier 即时检索存储类别（直接或通过存储桶生命周期配置），则对该类中对象的请求比其他存储类别中的对象更昂贵。
- 如果您将 DataSync 任务配置为验证源位置和目标位置是否完全同步，则所有存储类别（S3 Glacier 灵活检索和 S3 Glacier Deep Archive Deep Archive 除外）中的每个对象都将GET收到请求。
- 除了 GET 请求之外，您还会产生 S3 标准-IA、S3 单区-IA 或 S3 Glacier 即时检索存储类别中的对象的数据检索成本。

有关更多信息，请参阅 Amazon S3 定价。

我需要多久传输一次数据？

如果您需要定期移动数据，请考虑一个不会超出所需任务的时间表。

您也可以考虑限制传输范围。例如，您可以配置为聚焦 DataSync 于某些前缀中的对象，或者筛选要传输的数据。这些选项可以帮助减少每次运行 DataSync 任务时发出的 S3 请求数量。

Amazon S3 传输的对象注意事项

- 如果您要从 S3 存储桶中传输，请使用 S3 Storage Lens 存储分析功能确定要移动的对象数量。
- 在 S3 存储桶之间传输时，我们建议使用增强型任务模式，因为您不受 DataSync 任务配额的限制。
- DataSync 可能无法传输名称中包含非标准字符的对象。有关更多信息，请参阅《Amazon S3 用户指南》中的对象键命名指南。
- DataSync 与使用版本控制的 S3 存储桶一起使用时，请记住以下几点：
  - 传输到 S3 存储桶时，如果在源位置修改了对象，则 DataSync 会创建该对象的新版本。这会导致额外收费。
  - 对象在源存储桶和目标存储桶 IDs 中的版本不同。
  - 只有每个对象的最新版本才会从源存储桶传输。早期版本不会复制到目标。
- 最初将数据从 S3 存储桶传输到文件系统（例如 NFS 或 Amazon FSx）后，后续运行的同一 DataSync 任务将不包括已修改但大小与第一次传输时大小相同的对象。

为 Amazon S3 通用存储桶创建传输位置

要创建用于传输的位置，您需要一个现有的 S3 通用存储桶。如果还没有存储桶，请参阅《Amazon S3 用户指南》。

> **Important**
>
> 在创建位置之前，请确保阅读以下部分：
>
> - Amazon S3 传输的存储类别注意事项
> - 使用时评估 S3 请求成本 DataSync

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。

2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于位置类型，选择 Amazon S3，然后选择通用存储桶。
4. 对于 S3 URI，输入或选择要用于相应位置的存储桶和前缀。

> **Warning**
>
> DataSync 无法传输前缀以斜杠 (/) 开头或包含`///./`、或`/../`模式的对象。例如：
>
> - `/photos`
> - `photos//2006/January`
> - `photos/./2006/February`
> - `photos/../2006/March`

5. 对于 用作目标时的 S3 存储类别，请选择您希望对象在以 Amazon S3 为传输目标时使用的存储类别。

   有关更多信息，请参阅 Amazon S3 传输的存储类别注意事项。
6. 对于 IAM 角色，执行以下操作之一：

   - 选择 “自动生成” DataSync 以自动创建具有访问 S3 存储桶所需权限的 IAM 角色。

     如果 DataSync 之前为此 S3 存储桶创建了 IAM 角色，则默认情况下会选择该角色。
   - 选择您创建的自定义 IAM 角色。有关更多信息，请参阅 为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置。
7. （可选）选择添加新标签以标记您的 Amazon S3 位置。

   标签可帮助您管理、筛选和搜索资源。我们建议为您的位置创建一个名称标签。
8. 选择创建位置。

使用 AWS CLI

1. 复制以下 `create-location-s3` 命令：

```
aws datasync create-location-s3 \
    --s3-bucket-arn 'arn:aws:s3:::amzn-s3-demo-bucket' \
    --s3-storage-class 'your-S3-storage-class' \
    --s3-config 'BucketAccessRoleArn=arn:aws:iam::account-id:role/role-allowing-
datasync-operations' \
```

```
    --subdirectory /your-prefix-name
```

2. 对于 `--s3-bucket-arn`，指定要用作位置的 S3 存储桶的 ARN。
3. 对于 `--s3-storage-class`，指定您希望对象在以 Amazon S3 为传输目标时使用的存储类别。
4. 对于`--s3-config`，请指定 DataSync 需要访问您的存储桶的 IAM 角色的 ARN。

   有关更多信息，请参阅 为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置。
5. 对于`--subdirectory`，在 S3 存储桶中指定 DataSync 读取或写入的前缀（取决于存储桶是源位置还是目标位置）。

   > **Warning**
   >
   > DataSync 无法传输前缀以斜杠 (/) 开头或包含`///./`、或`/../`模式的对象。例如：
   >
   > - `/photos`
   > - `photos//2006/January`
   > - `photos/./2006/February`
   > - `photos/../2006/March`
6. 运行 `create-location-s3` 命令。

   如果命令成功，您将收到一条响应，显示您创建位置的 ARN。例如：

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:111222333444:location/
loc-0b3017fc4ba4a2d8d"
}
```

您可以将此位置用作 DataSync 任务的来源或目的地。

为 S3 on Outposts 存储桶创建传输位置

要创建用于传输的位置，您需要一个现有的 Amazon S3 on Outposts 存储桶。如果还没有存储桶，请参阅《Amazon S3 on Outposts 用户指南》。

你还需要一个 DataSync 代理。有关更多信息，请参阅 在上部署基本模式代理 AWS Outposts。

从包含大型数据集（例如数十万或数百万个对象）的 S3 on Outposts 存储桶前缀传输时，您的 DataSync 任务可能会超时。为避免这种情况，可以考虑使用DataSync 清单，它允许您指定需要传输的确切对象。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于位置类型，选择 Amazon S3，然后选择 Outposts 存储桶。
4. 对于 S3 存储桶，请选择一个可以访问 S3 on Outposts 存储桶的 Amazon S3 接入点。

   有关更多信息，请参阅 Amazon S3 用户指南。
5. 对于 用作目标时的 S3 存储类别，请选择您希望对象在以 Amazon S3 为传输目标时使用的存储类别。

   有关更多信息，请参阅Amazon S3 传输的存储类别注意事项。 DataSync 默认情况下，在 Outposts 上使用亚马逊 S3 的 S3 Outposts 存储类。
6. 对于代理，请在 Outpos DataSync t 上指定代理的亚马逊资源名称 (ARN)。
7. 对于文件夹，在 S3 存储桶中输入 DataSync 读取或写入的前缀（取决于存储桶是源位置还是目标位置）。

   > **Warning**
   >
   > DataSync 无法传输前缀以斜杠 (/) 开头或包含`///./`、或`/../`模式的对象。例如：
   >
   > - `/photos`
   > - `photos//2006/January`
   > - `photos/./2006/February`
   > - `photos/../2006/March`
8. 对于 IAM 角色，执行以下操作之一：

   - 选择 “自动生成” DataSync 以自动创建具有访问 S3 存储桶所需权限的 IAM 角色。

     如果 DataSync 之前为此 S3 存储桶创建了 IAM 角色，则默认情况下会选择该角色。
   - 选择您创建的自定义 IAM 角色。有关更多信息，请参阅 为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置。

9. （可选）选择添加新标签以标记您的 Amazon S3 位置。

   标签可帮助您管理、筛选和搜索资源。我们建议为您的位置创建一个名称标签。

10. 选择创建位置。

使用 AWS CLI

1. 复制以下 `create-location-s3` 命令：

```
aws datasync create-location-s3 \
    --s3-bucket-arn 'bucket-access-point' \
    --s3-storage-class 'your-S3-storage-class' \
    --s3-config 'BucketAccessRoleArn=arn:aws:iam::account-id:role/role-allowing-
datasync-operations' \
    --subdirectory /your-folder \
    --agent-arns 'arn:aws:datasync:your-region:account-id:agent/agent-agent-id'
```

2. 对于 `--s3-bucket-arn`，指定一个可访问 Amazon S3 on Outposts 存储桶的 Amazon S3 接入点。

   有关更多信息，请参阅 Amazon S3 用户指南。

3. 对于 `--s3-storage-class`，指定您希望对象在以 Amazon S3 为传输目标时使用的存储类别。

   有关更多信息，请参阅Amazon S3 传输的存储类别注意事项。 DataSync 默认情况下，在 Outposts 上使用 S3 Outposts 存储类作为 S3。

4. 对于`--s3-config`，请指定 DataSync 需要访问您的存储桶的 IAM 角色的 ARN。

   有关更多信息，请参阅 为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置。

5. 对于`--subdirectory`，在 S3 存储桶中指定 DataSync 读取或写入的前缀（取决于存储桶是源位置还是目标位置）。

   > **Warning**
   >
   > DataSync 无法传输前缀以斜杠 (/) 开头或包含`///./`、或`/../`模式的对象。例如：
   >
   > - `/photos`
   > - `photos//2006/January`
   > - `photos/./2006/February`

- photos/../2006/March

6. 对于--agent-arns，请指定前哨基地上 DataSync 特工的 ARN。
7. 运行 create-location-s3 命令。

   如果命令成功，您将收到一条响应，显示您创建位置的 ARN。例如：

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:111222333444:location/
loc-0b3017fc4ba4a2d8d"
}
```

您可以将此位置用作 DataSync 任务的来源或目的地。

亚马逊 S3 跨境传输 AWS 账户

使用 DataSync，您可以将数据移入或移出不同 AWS 账户的 S3 存储桶。有关更多信息，请参阅以下教程：

- 将数据从本地存储传输到 Amazon S3 AWS 账户
- 将数据从亚马逊 S3 传输到亚马逊 S3 AWS 账户

Amazon S3 在商业版和商业版之间转移 AWS GovCloud (US) Regions

默认情况下， DataSync 不会在商业版和 S3 存储桶之间进行传输。 AWS GovCloud (US) Regions但您仍然可以在传输中为一个 S3 存储桶创建对象存储位置，从而设置这种传输。无论使用不使用代理，都可以执行这种类型的传输。如果使用代理，必须将任务配置为基本模式。要在没有代理的情况下进行传输，就必须使用增强模式。

开始之前：请确保了解在区域间进行传输的成本影响。有关更多信息，请参阅AWS DataSync 定价。

目录

- 提供对您的对象存储位置的存储桶的 DataSync 访问权限
- 创建您的 DataSync 代理（可选）
- 为 S3 存储桶创建对象存储位置

提供对您的对象存储位置的存储桶的 DataSync 访问权限

在为此转移创建对象存储位置时，您必须提供有权访问 DataSync 该位置的 S3 存储桶的 IAM 用户的证书。有关更多信息，请参阅 所需的权限。

> **Warning**
>
> IAM 用户具有长期凭证，这会带来安全风险。为帮助减轻这种风险，我们建议仅向这些用户提供执行任务所需的权限，并在不再需要这些用户时将其移除。

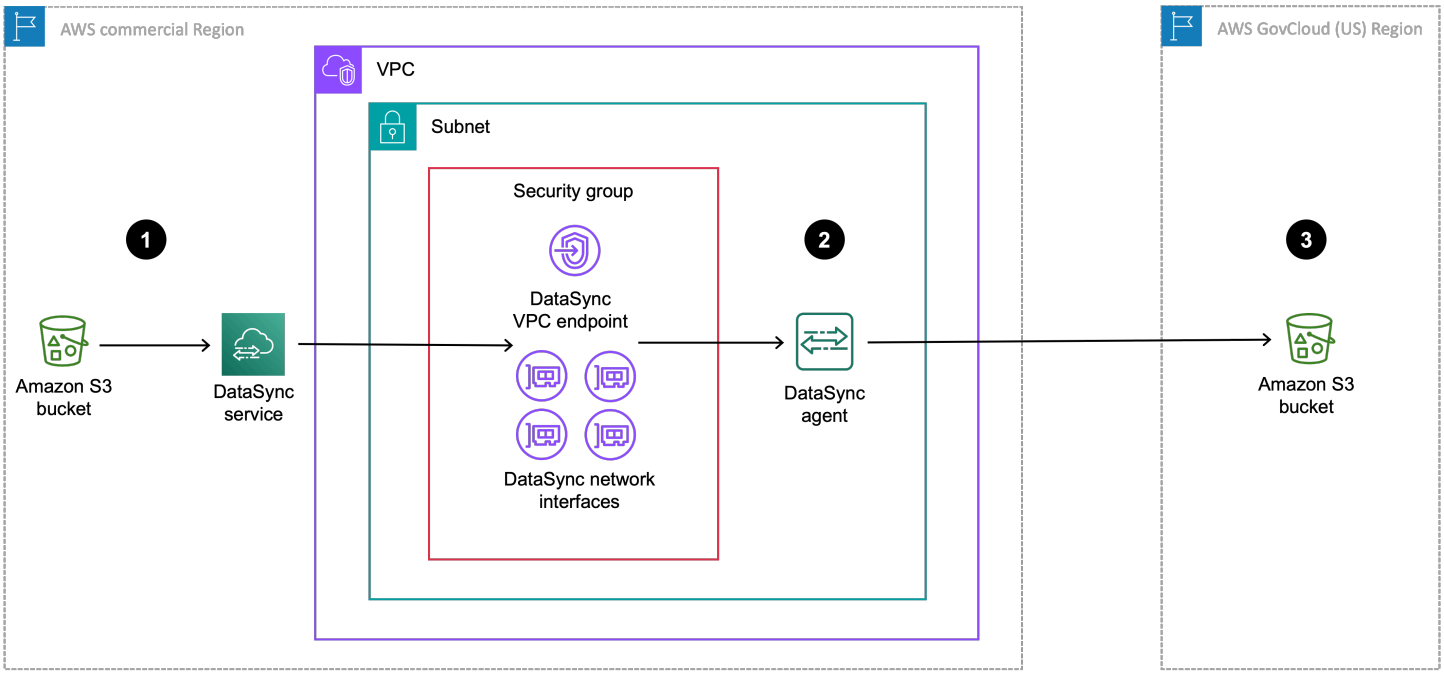
创建您的 DataSync 代理（可选）

如果要使用基本模式运行传输，就需要使用代理。由于您要在商业版和之间进行转移 AWS GovCloud (US) Region，因此您可以将 DataSync 代理作为 Amazon EC2 实例部署到其中一个区域。我们建议您的代理使用 VPC 服务端点，以避免将数据传输至公共互联网时产生的费用。有关更多信息，请参阅 Amazon EC2 数据传输定价。

选择以下情景之一，描述如何根据您计划运行 DataSync 任务的区域创建代理。

在商业区域运行 DataSync 任务时

下图显示了您的 DataSync 任务和代理位于商业区域的转移。



参考	说明
1	在您运行 DataSync任务的商业区域中，数据从源 S3 存储桶传输。源存储桶配置为商业区域中的 Amazon S3 位置。
2	通过代理传输数据， DataSync 代理位于 VPC 服务终端节点和网络接口所在的 VPC 和子网中。
3	数据传输至 AWS GovCloud (US) Region中的目标 S3 存储桶。目标存储桶配置为商业区域中的对象存储位置。

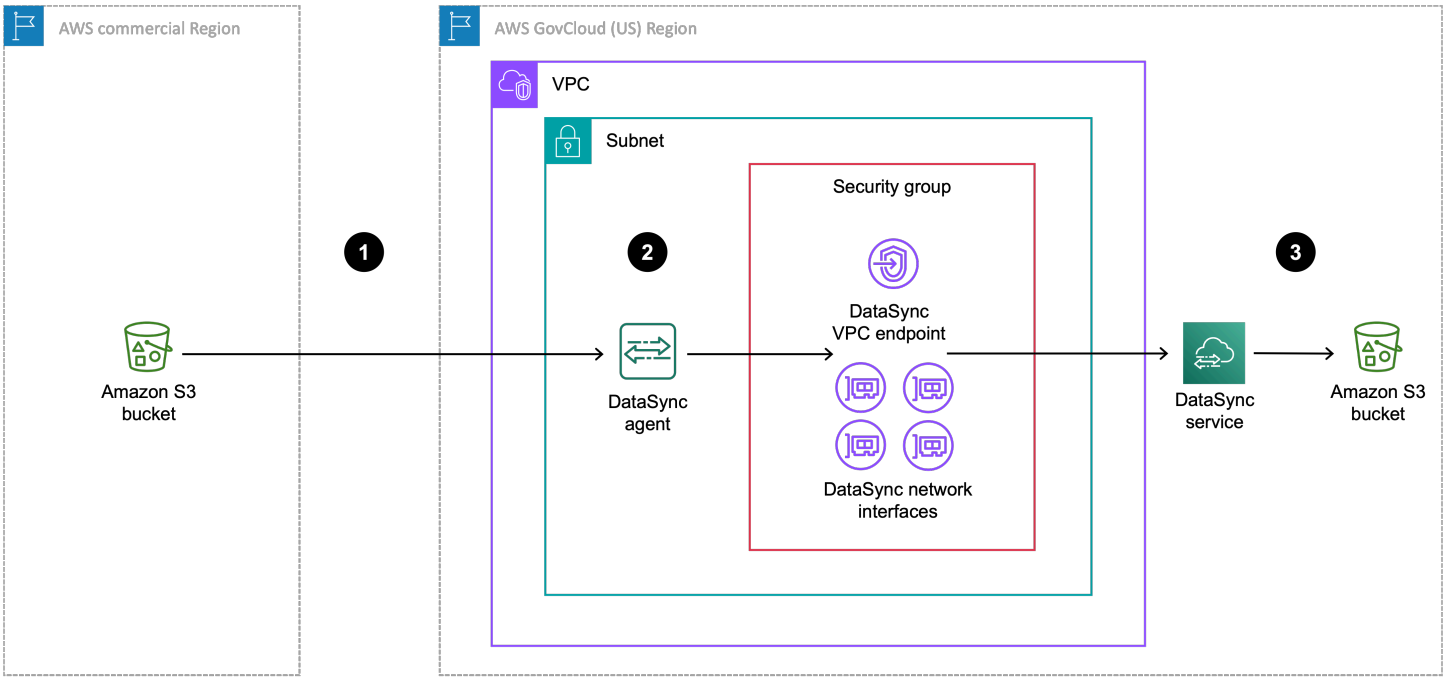
您也可以使用相同的设置将相反的方向从转移 AWS GovCloud (US) Region 到商业区域。

创建您的 DataSync 代理

1. 在您的商业区域部署 Amazon EC2 代理。
2. 将您的代理配置为使用 VPC 服务端点。
3. 激活您的代理。

在 GovCloud （美国）地区运行 DataSync 任务时

下图显示了您的 DataSync 任务和代理所在的转移 AWS GovCloud (US) Region。



参考	说明
1	数据从商业区域的源 S3 存储桶传输到您正在运行 DataSync 任务 AWS GovCloud (US) Region 的地方。源存储桶配置为 AWS GovCloud (US) Region中的对象存储位置。
2	在中 AWS GovCloud (US) Region，通过 VPC 服务终端节点和网络接口所在的同一 VPC 和子网中的 DataSync 代理进行数据传输。
3	数据传输至 AWS GovCloud (US) Region中的目标 S3 存储桶。目标存储桶被配置为 AWS GovCloud (US) Region中的 Amazon S3 位置。

您也可以使用相同的设置将相反的方向从转移 AWS GovCloud (US) Region 到商业区域。

创建您的 DataSync 代理

1. 在您的@@ 中部署 Amazon EC2 代理 AWS GovCloud (US) Region。
2. 将您的代理配置为使用 VPC 服务端点。
3. 激活您的代理。

如果您的数据集具有高度可压缩性，您可以在商业区域创建代理，同时在 AWS GovCloud (US) Region运行任务，这样可能会降低成本。与通常情况相比，创建此代理需要进行更多的设置，包括为在商业区域使用代理做好准备。有关为此设置创建代理的信息，请参阅AWS GovCloud (US) 使用迁入和移出数据 AWS DataSync博客。

为 S3 存储桶创建对象存储位置

您需要为未运行 DataSync 任务的区域中的 S3 存储桶设置一个对象存储位置。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 确保您处于计划运行任务的同一区域。
3. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
4. 对于 位置类型，选择 对象存储。

5. 对于代理，请选择您为此转移创建的 DataSync 代理。
6. 对于服务器，使用以下格式之一为存储桶输入 Amazon S3 端点：

   - 商业区域存储桶：`s3.your-region.amazonaws.com`
   - AWS GovCloud (US) Region 存储桶：`s3.your-gov-region.amazonaws.com`

   有关 Amazon S3 端点的列表，请参阅 AWS 一般参考。
7. 对于存储桶名称，请输入 S3 存储桶的名称。
8. 对于文件夹，在 S3 存储桶中输入 DataSync 读取或写入的前缀（取决于存储桶是源位置还是目标位置）。

   > **Warning**
   >
   > DataSync 无法传输前缀以斜杠 (/) 开头或包含`///./`、或`/../`模式的对象。例如：
   >
   > - `/photos`
   > - `photos//2006/January`
   > - `photos/./2006/February`
   > - `photos/../2006/March`

9. 选择需要凭证，然后执行以下操作：

   - 对于访问密钥，输入可以访问存储桶的 IAM 用户的访问密钥。
   - 对于私有密钥，输入同一 IAM 用户的私有密钥。
10. （可选）选择 添加标签来标记您的位置。

    标签可帮助您管理、筛选和搜索资源。我们建议为您的位置创建一个名称标签。
11. 选择创建位置。

使用 AWS CLI

1. 复制以下 `create-location-object-storage` 命令：

```
aws datasync create-location-object-storage \
    --server-hostname s3-endpoint \
    --bucket-name amzn-s3-demo-bucket \
```

```
    --agent-arns arn:aws:datasync:your-region:123456789012:agent/
agent-01234567890deadfb
```

2. 对于 `--server-hostname` 参数，使用以下格式之一，为存储桶指定 Amazon S3 端点：

   - 商业区域存储桶：`s3.your-region.amazonaws.com`
   - AWS GovCloud (US) Region 存储桶：`s3.your-gov-region.amazonaws.com`

   对于端点中的区域，确保指定的区域与计划运行任务的区域相同。

   有关 Amazon S3 端点的列表，请参阅 AWS 一般参考。
3. 对于 `--bucket-name` 参数，指定 S3 存储桶的名称。
4. 在`--agent-arns`参数中，指定您为此传输创建的 DataSync代理。
5. 对于 `--access-key` 参数，指定可以访问存储桶的 IAM 用户的访问密钥。
6. 对于 `--secret-key` 参数，输入同一 IAM 用户的私有密钥。
7. （可选）在`--subdirectory`参数中，在 S3 存储桶中指定 DataSync 读取或写入的前缀（取决于存储桶是源位置还是目标位置）。

   > **Warning**
   >
   > DataSync 无法传输前缀以斜杠 (/) 开头或包含`///./`、或`/../`模式的对象。例如：
   >
   > - `/photos`
   > - `photos//2006/January`
   > - `photos/./2006/February`
   > - `photos/../2006/March`

8. （可选）对于 `--tags` 参数，指定代表位置资源标签的键值对。

   标签可帮助您管理、筛选和搜索资源。我们建议为您的位置创建一个名称标签。
9. 运行 `create-location-object-storage` 命令。

   您会收到一条回复，表明您刚刚创建的位置 ARN。

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:123456789012:location/
loc-01234567890abcdef"
```

```
}
```

您可以将此位置用作 DataSync 任务的来源或目的地。对于此传输中的其他 S3 存储桶，创建一个 Amazon S3 位置。

后续步骤

后续可能采取的一些步骤包括：

1. 如果需要，创建其他位置。有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
2. 配置 DataSync 任务设置，例如要传输哪些文件、如何处理元数据以及其他选项。
3. 为您的 DataSync任务@@ 设定时间表。
4. 为您的 DataSync 任务@@ 配置监控。
5. 开始任务。

使用 Amazon EFS 配置 AWS DataSync 传输

要将数据传输到您的 Amazon EFS 文件系统或从中传输数据，您必须创建一个 AWS DataSync 传输位置。 DataSync 可以将此位置用作传输数据的源或目的地。

提供 DataSync 对 Amazon EFS 文件系统的访问权限

创建位置需要了解 DataSync 如何访问您的存储空间。对于 Amazon EFS，使用网络接口以根用户身份从虚拟私有云 (VPC) DataSync 挂载您的文件系统。

目录

- 确定挂载目标的子网和安全组
- 访问受限文件系统
  - 为文件系统访问创建 DataSync IAM 角色
  - 允许 DataSync 访问的文件系统策略示例

确定挂载目标的子网和安全组

创建位置时，您可以指定允许连接 DataSync 到 Amazon EFS 文件系统的挂载目标之一的子网和安全组。

指定的子网必须位于：

- 文件系统所在的相同 VPC 中。
- 文件系统至少一个挂载目标所在的相同可用区中。

> **Note**
>
> 您无需指定包含文件系统挂载目标的子网。

您指定的安全组必须允许网络文件系统（NFS）端口 2049 上的入站流量。有关为挂载目标创建和更新安全组的信息，请参阅 Amazon EFS 用户指南。

指定与挂载目标相关联的安全组

您可以指定与文件系统挂载目标之一相关联的安全组。从网络管理的角度来看，我们推荐这种方法。

指定与挂载目标无关联的安全组

您还可以指定与文件系统的挂载目标无关联的安全组。但是，此安全组必须能够与挂载目标的安全组进行通信。

例如，您可以通过以下方式在安全组 D（用于 DataSync）和安全组 M（用于挂载目标）之间创建关系：

- 您在创建位置时指定的安全组 D 必须具有一条规则，允许 NFS 端口 2049 上与安全组 M 的出站连接。
- 与挂载目标相关联的安全组 M 必须允许安全组 D 通过 NFS 端口 2049 进行的入站访问。

查找挂载目标的安全组

以下说明可以帮助您识别要 DataSync 用于传输的 Amazon EFS 文件系统挂载目标的安全组。

1. 在中 AWS CLI，运行以下`describe-mount-targets`命令。

```
aws efs describe-mount-targets \
    --region file-system-region  \
    --file-system-id file-system-id
```

此命令将返回有关文件系统挂载目标的信息（与以下示例输出类似）。

```
{
    "MountTargets": [
        {
            "OwnerId": "111222333444",
            "MountTargetId": "fsmt-22334a10",
            "FileSystemId": "fs-123456ab",
            "SubnetId": "subnet-f12a0e34",
            "LifeCycleState": "available",
            "IpAddress": "11.222.0.123",
            "NetworkInterfaceId": "eni-1234a044"
        }
    ]
}
```

2. 记下要使用的 MountTargetId 值。
3. 使用 MountTargetId 运行以下 describe-mount-target-security-groups 命令，查看挂载目标的安全组。

```
aws efs describe-mount-target-security-groups \
    --region file-system-region \
    --mount-target-id mount-target-id
```

在创建位置时，指定此安全组。

访问受限文件系统

DataSync 可以与限制通过接入点和 IAM 策略进行访问的 Amazon EFS 文件系统进行传输或传出。

> **Note**
>
> 如果通过强制 DataSync 使用用户身份的接入点访问目标文件系统，则如果将 DataSync 任务配置 IDs 为复制所有权，则不会保留源数据的 POSIX 用户和组。而是将传输的文件和文件夹设置为接入点的用户和群组 IDs。发生这种情况时，任务验证会失败，因为 DataSync 检测到源位置和目标位置的元数据不匹配。

目录

- 为文件系统访问创建 DataSync IAM 角色
- 允许 DataSync 访问的文件系统策略示例

为文件系统访问创建 DataSync IAM 角色

如果您的 Amazon EFS 文件系统通过 IAM 策略限制访问，则可以创建一个 IAM 角色来提供从文件系统读取或写入数据的 DataSync 权限。然后，您可能需要在文件系统策略中指定该角色。

创建 DataSync IAM 角色

1. 使用 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。
2. 在左侧导航窗格的访问管理下，选择角色，然后选择创建角色。
3. 在选择可信实体中的可信实体类型下，选择自定义信任策略。
4. 将以下 JSON 粘贴到策略编辑器中：

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Principal": {
            "Service": "datasync.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

5. 选择下一步。在添加权限页面上，选择下一步。
6. 输入角色名称，然后选择创建角色。

请在创建位置时指定此角色。

允许 DataSync 访问的文件系统策略示例

以下示例文件系统策略显示了如何限制对 Amazon EFS 文件系统的访问（在策略中标识为`fs-`*`1234567890abcdef0`*），但仍允许 DataSync 通过名为的 IAM 角色进行访问*`MyDataSyncRole`*：

JSON

```
{
    "Version":"2012-10-17",
    "Id": "ExampleEFSFileSystemPolicy",
    "Statement": [{
        "Sid": "AccessEFSFileSystem",
        "Effect": "Allow",
        "Principal": {
            "AWS": "arn:aws:iam::111122223333:role/MyDataSyncRole"
        },
        "Action": [
            "elasticfilesystem:ClientMount",
            "elasticfilesystem:ClientWrite",
            "elasticfilesystem:ClientRootAccess"
        ],
        "Resource": "arn:aws:elasticfilesystem:us-east-1:111122223333:file-
system/fs-1234567890abcdef0",
        "Condition": {
            "Bool": {
                "aws:SecureTransport": "true"
            },
            "StringEquals": {
                "elasticfilesystem:AccessPointArn":
 "arn:aws:elasticfilesystem:us-east-1:111122223333:access-point/
fsap-abcdef01234567890"
            }
        }
    }]
}
```

- `Principal`— 指定一个授予文件系统访问 DataSync 权限的 IAM 角色。
- `Action`— 授予 DataSync root 用户访问权限并允许其读取和写入文件系统。
- `aws:SecureTransport`：要求 NFS 客户端在连接到文件系统时使用 TLS。
- `elasticfilesystem:AccessPointArn`：仅允许通过特定的接入点访问文件系统。

Amazon EFS 传输的网络注意事项

VPCs 与一起使用的 DataSync 必须具有默认租期。 VPCs 不支持专用租赁。

Amazon EFS 传输的性能注意事项

Amazon EFS 文件系统的吞吐量模式可能会在传输期间影响传输持续时长和文件系统性能。请考虑以下事项：

- 为了获得最佳效果，建议使用弹性吞吐量模式。如果不使用弹性吞吐量模式，传输可能需要更长的时间。
- 如果您使用突增吞吐量模式，则文件系统应用程序的性能可能会受到影响，因为会 DataSync 消耗文件系统的突发积分。
- 如何配置 DataSync 以验证传输的数据可能会影响文件系统的性能和数据访问成本。

有关更多信息，请参阅《Amazon Elastic File System 用户指南》中的 Amazon EFS 性能，以及 Amazon EFS 定价页面。

创建您的 Amazon EFS 传输位置

要创建传输位置，您需要一个现有的 Amazon EFS 文件系统。如果您还没有 Amazon EFS 文件系统，请参阅《Amazon Elastic File System 用户指南》中的 Amazon EFS 入门。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于位置类型，请选择 Amazon EFS 文件系统。

   您可在稍后将此位置配置为源或目标。
4. 对于文件系统，选择要用作端点的 Amazon EFS 文件系统。
5. 对于挂载路径，输入 Amazon EFS 文件系统的挂载路径。

   这指定了文件系统上 DataSync 读取或写入数据的位置（取决于这是源位置还是目标位置）。

   默认情况下， DataSync 使用根目录（如果您为 EFS 接入点设置提供了一个接入点，则使用接入点）。也可以使用正斜杠指定子目录（例如，`/path/to/directory`）。
6. 对于子网，请选择 DataSync 要在其中创建用于管理数据传输流量的网络接口的子网。

   子网必须位于：

   - 文件系统所在的相同 VPC 中。

- 与至少一个文件系统挂载目标位于同一可用区中。

> **Note**
>
> 您无需指定包含文件系统挂载目标的子网。

7. 对于安全组，选择与您的 Amazon EFS 文件系统的挂载目标相关联的安全组。您可以选择多个安全组。

> **Note**
>
> 您指定的安全组必须允许 NFS 端口 2049 上的入站流量。有关更多信息，请参阅 确定挂载目标的子网和安全组。

8. 对于传输中加密，请选择在 DataSync 向文件系统传输数据或从文件系统传输数据时是否要使用传输层安全 (TLS) 加密。

> **Note**
>
> 您必须启用此设置，才能在您的 Amazon EFS 位置配置接入点、IAM 角色或两者。

9. （可选）对于 EFS 接入点，请选择 DataSync 可用于装载文件系统的接入点。

   有关更多信息，请参阅 访问受限文件系统。

10. （可选）对于 IAM 角色，请指定 DataSync 允许访问您的文件系统的角色。

    有关创建该角色的信息，请参阅为文件系统访问创建 DataSync IAM 角色。

11. （可选）选择添加标签来标记您的文件系统。

    标签是帮助您管理、筛选和搜索位置的键值对。

12. 选择创建位置。

使用 AWS CLI

1. 复制以下 `create-location-efs` 命令：

```
aws datasync create-location-efs \
```

```
    --efs-filesystem-arn 'arn:aws:elasticfilesystem:region:account-id:file-
system/file-system-id' \
    --subdirectory /path/to/your/subdirectory \
    --ec2-config SecurityGroupArns='arn:aws:ec2:region:account-id:security-
group/security-group-id',SubnetArn='arn:aws:ec2:region:account-id:subnet/subnet-id'
 \
    --in-transit-encryption TLS1_2 \
    --access-point-arn 'arn:aws:elasticfilesystem:region:account-id:access-
point/access-point-id' \
    --file-system-access-role-arn 'arn:aws:iam::account-id:role/datasync-efs-
access-role
```

2. 对于 `--efs-filesystem-arn`，指定要向其传输数据或从中传出数据的 Amazon EFS 文件系统的 Amazon 资源名称（ARN）。
3. 对于 `--subdirectory`，指定文件系统的挂载路径。

   这是在文件系统上 DataSync 读取或写入数据的地方（取决于这是源位置还是目标位置）。

   默认情况下， DataSync 使用根目录（或接入点，如果您提供了接入点`--access-point-arn`）。也可以使用正斜杠指定子目录（例如，`/path/to/directory`）。
4. 对于 `--ec2-config`，执行以下操作：

   - 对于 `SecurityGroupArns`，指定与文件系统挂载目标相关联的安全组 ARN。您可以指定多个安全组。

     > **Note**
     >
     > 您指定的安全组必须允许 NFS 端口 2049 上的入站流量。有关更多信息，请参阅 确定挂载目标的子网和安全组。

   - 对于`SubnetArn`，请指定 DataSync 要在其中创建用于管理数据传输流量的网络接口的子网的 ARN。

     子网必须位于：

     - 文件系统所在的相同 VPC 中。
     - 与至少一个文件系统挂载目标位于同一可用区中。

> **Note**
>
> 您无需指定包含文件系统挂载目标的子网。

5. 对于`--in-transit-encryption`，请指定在传输层安全 (TLS) DataSync 向文件系统传输数据或从文件系统传输数据时是否要使用传输层安全 (TLS) 加密。

> **Note**
>
> 您必须将其设置为 `TLS1_2`，才能在您的 Amazon EFS 位置配置接入点、IAM 角色或两者。

6. （可选）对于`--access-point-arn`，请指定 DataSync 可用于挂载文件系统的接入点的 ARN。

   有关更多信息，请参阅 访问受限文件系统。

7. （可选）对于`--file-system-access-role-arn`，请指定允许 DataSync 访问您的文件系统的 IAM 角色的 ARN。

   有关创建该角色的信息，请参阅为文件系统访问创建 DataSync IAM 角色。

8. 运行 `create-location-efs` 命令。

   如果命令成功，您将收到一条响应，显示您创建位置的 ARN。例如：

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:111222333444:location/
loc-0b3017fc4ba4a2d8d"
}
```

使用为 Window FSx s 文件服务器配置传输

要将数据传输到您的 Amazon f FSx or Windows 文件服务器文件系统或从中传输数据，您必须创建一个 AWS DataSync 传输位置。 DataSync 可以将此位置用作传输数据的源或目的地。

FSx 为 Windows 文件服务器文件系统提供 DataSync 访问权限

DataSync 使用服务器消息块 (SMB) 协议连接到 FSx 适用于 Windows 的文件服务器文件系统，并使用网络接口将其从您的虚拟私有云 (VPC) 挂载。

> **Note**
>
> VPCs 与一起使用的 DataSync 必须具有默认租期。 VPCs 不支持专用租赁。

主题

- 所需的权限
- 所需的身份验证协议
- DFS 命名空间

所需的权限

您必须 DataSync 为用户提供必要的权限才能装载和访问您 FSx 的 Windows 文件服务器文件、文件夹和文件元数据。

建议使用属于 Microsoft Active Directory 组的用户，以便管理文件系统。该组的具体情况取决于您的 Active Directory 设置：

- 如果你在 Windows 文件服务器中 AWS Directory Service for Microsoft Active Directory 使用 FSx 和，则该用户必须是AWS 委派 FSx 管理员组的成员。
- 如果你使用的是 Windows 文件服务器的自我管理的 Active Directory，则该用户必须是以下两个群组之一的成员： FSx
  - 域管理员组，这是默认的委派管理员组。
  - 一个自定义的委派管理员组，其用户权限 DataSync 允许复制对象所有权权限和 Windows 访问控制列表 (ACLs)。

> **Important**
>
> 部署文件系统后，无法更改委派管理员组。您必须重新部署文件系统或从备份中恢复文件系统，才能使用具有以下用户权限的自定义委派管理员组来复制元数据。 DataSync

用户权限	说明
恢复文件和目录（`SE_RESTORE_NAME` ）	DataSync 允许复制对象所有权、权限、文件元数据和 NTFS 自由访问列表 () DACLs。

用户权限	说明
	此用户权限通常授予域管理员组和备份操作员组（均为默认 Active Directory 组）的成员。
管理审计和安全日志（SE_SECURITY_NAME ）	DataSync 允许复制 NTFS 系统访问控制列表 (SACLs)。 此用户权限通常授予域管理员组的成员。

- 如果要复制 Windows， ACLs 并且要在 SMB 文件服务器和 Windows 文件服务器文件系统之间 FSx 进行传输，或者在 Windows 文件服务器文件系统之间 FSx 进行传输，则您提供的用户 DataSync 必须属于同一 Active Directory 域或者其域之间具有 Active Directory 信任关系。

> **Warning**
>
> 您 FSx 的 Windows 文件服务器文件系统的系统用户必须对文件系统中的所有文件夹具有完全控制权限。请勿更改此用户在您的文件夹上的 NTFS ACL 权限。如果这样做，则 DataSync 可以更改文件系统的权限，从而使您的文件共享无法访问并阻止文件系统备份可用。有关文件和文件夹级访问的更多信息，请参阅 A mazon for Window FSx s 文件服务器用户指南。

所需的身份验证协议

你 FSx 的 Windows 文件服务器必须使用 NTLM 身份验证 DataSync 才能访问它。 DataSync无法访问使用 Kerberos 身份验证的文件服务器。

DFS 命名空间

DataSync 不支持 Microsoft 分布式文件系统 (DFS) 命名空间。我们建议您在创建 DataSync 位置时改为指定底层文件服务器或共享。

有关更多信息，请参阅《A mazon for Windows 文件服务器用户指南》中的使用 DFS 命名空间对多个文件系统 FSx 进行分组。

创建你 FSx 的 Windows 文件服务器传输位置

在开始之前，请确保您的文件服务器中已有 FSx 适用于 Windows 的文件服务器 AWS 区域。有关更多信息，请参阅亚马逊 Windows 文件服务器用户指南 FSx 中的亚马逊 FSx 入门。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于位置类型，请选择 Amazon FSx。
4. 对于FSx 文件系统， FSx 请选择要用作位置的 Windows 文件服务器文件系统。
5. 在 “共享名称” 中，使用正斜杠输入适用 FSx 于 Windows 文件服务器的挂载路径。

   这指定了 DataSync 读取或写入数据的路径（取决于这是源位置还是目标位置）。

   您也可以包括子目录（例如，`/path/to/directory`）。
6. 对于安全组，最多选择五个 Amazon EC2 安全组，这些安全组提供对文件系统首选子网的访问权限。

   您选择的安全组必须能够与文件系统的安全组通信。有关为文件系统访问配置安全组的信息，请参阅 Amazon f FSx or Windows 文件服务器用户指南。

   > **Note**
   >
   > 如果您选择的安全组不允许从其自身进行连接，请执行以下操作之一：
   >
   > - 将安全组配置为允许其在自身内部进行通信。
   > - 选择可与挂载目标的安全组进行通信的其他安全组。
7. 在 “用户” 中，输入可以访问您的 Windows 文件服务器 FSx 的用户名。

   有关更多信息，请参阅 所需的权限。
8. 对于密码，输入上述用户名所对应的密码。
9. （可选）在 “域” 中，输入您 FSx 的 Windows 文件服务器文件系统所属的 Windows 域的名称。

   如果您的环境中有多个 Active Directory 域，则配置此设置可确保这些域 DataSync 连接到正确的文件系统。
10. （可选）在 “密钥” 和 “值” 字段中输入值 FSx 以标记 Windows 文件服务器。

    标签可帮助您管理、筛选和搜索 AWS 资源。我们建议至少为您的位置创建一个名称标签。
11. 选择创建位置。

使用 AWS CLI

要创建 FSx 适用于 Windows 文件服务器的位置，请使用 AWS CLI

- 使用以下命令创建 Amazon FSx 营业地点。

```
aws datasync create-location-fsx-windows \
    --fsx-filesystem-arn arn:aws:fsx:region:account-id:file-system/filesystem-id \
    --security-group-arns arn:aws:ec2:region:account-id:security-group/group-id \
    --user smb-user --password password
```

在 `create-location-fsx-windows` 命令中，执行以下操作：

- `fsx-filesystem-arn`：指定要向其传输或从中传出数据的文件系统的 Amazon 资源名称（ARN）。
- `security-group-arns`：指定最多五个 Amazon EC2 安全组的 ARN，这些安全组提供对文件系统首选子网的访问权限。

  您指定的安全组必须能够与文件系统的安全组通信。有关为文件系统访问配置安全组的信息，请参阅 Amazon f FSx or Windows 文件服务器用户指南。

  > **Note**
  >
  > 如果您选择的安全组不允许从其自身进行连接，请执行以下操作之一：
  >
  > - 将安全组配置为允许其在自身内部进行通信。
  > - 选择可与挂载目标的安全组进行通信的其他安全组。

- AWS 区域 — 您指定的区域就是您的目标 Amazon FSx 文件系统所在的区域。

上一命令返回一个如下所示的位置 ARN。

```
{
    "LocationArn": "arn:aws:datasync:us-west-2:111222333444:location/
loc-07db7abfc326c50fb"
}
```

为 Lustr FSx e 配置 DataSync 传输

要将数据传输到或传出您的 Amazon FSx for Lustre 文件系统，您必须创建一个 AWS DataSync 传输位置。 DataSync 可以将此位置用作传输数据的源或目的地。

FSx 为 Lustre 文件系统提供 DataSync 访问权限

DataSync 使用 Lustre 客户端访问你 FSx 的 for Lustre 文件系统。 DataSync需要访问您 FSx 的 for Lustre 文件系统上的所有数据。要获得此级别的访问权限，请使用用户 ID (UID) 和组 ID (GID) 以 root 用户身份 DataSync 装载文件系统。0

DataSync 使用网络接口从您的虚拟私有云 (VPC) 挂载您的文件系统。 DataSync 代表您全面管理这些网络接口的创建、使用和删除。

> **Note**
>
> VPCs 与一起使用的 DataSync 必须具有默认租期。 VPCs 不支持专用租赁。

创建您的 f FSx or Lustre 接送地点

要创建传输位置，您需要一个现有 FSx 的 For Lustre 文件系统。有关更多信息，请参阅《亚马逊 for Lustre 用户指南》中的 Ama FSx z FSx on for Lustre 入门。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于位置类型，请选择 Amazon FSx。

   您可在稍后将此位置配置为源或目标。
4. 对于FSx 文件系统，请选择要用作位置的 Lustre 文件系统。 FSx
5. 在装载路径中，输入 for Lustre 文件系统的装载路径。 FSx

   该路径可包含子目录。当该位置用作源时，会从装载路径 DataSync 读取数据。将该位置用作目标时， DataSync 会将所有数据写入装载路径。如果未提供子目录，则 DataSync 使用根目录 (/)。
6. 对于安全组，最多可选择五个安全组来访问您 FSx 的 for Lustre 文件系统。

安全组必须能够访问文件系统的端口。文件系统还必须允许来自安全组的访问。

有关安全组的更多信息，请参阅 Amazon for Lustre 用户指南中的使用 Amazon VPC FSx 进行文件系统访问控制。

7. （可选）在“密钥”和“值”字段中输入值以标记 Lustre 文件系统。 FSx

标签可帮助您管理、筛选和搜索 AWS 资源。我们建议至少为您的位置创建一个名称标签。

8. 选择创建位置。

使用 AWS CLI

要创建 fo FSx r Lustre 位置，请使用 AWS CLI

- 使用以下命令创建 for Lustre 位置。 FSx

```
aws datasync create-location-fsx-lustre \
    --fsx-filesystem-arn arn:aws:fsx:region:account-id:file-system:filesystem-id \
    --security-group-arns arn:aws:ec2:region:account-id:security-group/group-id
```

`create-location-fsx-lustre` 命令需要使用以下参数。

- `fsx-filesystem-arn` – 要在其中读取或写入的文件系统的完全限定的 Amazon 资源名称 (ARN)。
- `security-group-arns` – 将应用于文件系统首选子网的网络接口的 Amazon EC2 安全组的 ARN。

上一命令返回位置 ARN，如下所示。

```
{
    "LocationArn": "arn:aws:datasync:us-west-2:111222333444:location/
loc-07sb7abfc326c50fb"
}
```

为 OpenZFS 配置与亚马逊 FSx 的 DataSync 转账

要将数据传输到或传出您的亚马逊版 FSx OpenZFS 文件系统，您必须创建一个 AWS DataSync 传输位置。 DataSync 可以将此位置用作传输数据的源或目的地。

FSx 为 OpenZFS 文件系统提供 DataSync 访问权限

DataSync 使用网络接口从您的 FSx 虚拟私有云 (VPC) 挂载您的 for OpenZFS 文件系统。 DataSync 代表您全面管理这些网络接口的创建、使用和删除。

> **Note**
>
> VPCs 与一起使用的 DataSync 必须具有默认租期。 VPCs 不支持专用租赁。

配置 O FSx penZFS 文件系统授权

DataSync 以 NFS 客户端 FSx 的身份访问你的 for OpenZFS 文件系统，以 root 用户身份挂载文件系统，用户 ID (UID) 和组 ID (GID) 为。`0`

DataSync 要复制所有文件元数据，必须使用`no_root_squash`在文件系统卷上配置 NFS 导出设置。但是，您可以将此级别的访问权限限制为仅限于特定的 DataSync 任务。

有关更多信息，请参阅 Amazon FSx for OpenZFS 用户指南中的卷属性。

配置特定于 DataSync （推荐）的 NFS 导出

您可以为每个只能由您的 DataSync 任务访问的卷配置 NFS 导出。对您在 FSx 创建 for OpenZFS 位置时指定的装载路径的最新祖先卷执行此操作。

要配置特定于 NFS 导出 DataSync

1. 创建您的DataSync任务。

   这将创建您将在 NFS 导出设置中指定的任务的网络接口。
2. 使用 Amazon EC2 控制台或者，找到任务网络接口的私有 IP 地址 AWS CLI。
3. 对 FSx 于 OpenZFS 文件系统卷，请为任务的每个网络接口配置以下 NFS 导出设置：

   - 客户端地址：输入网络接口的私有 IP 地址（例如 *`10.24.34.0`*）。
   - NFS 选项：输入 `rw,no_root_squash`。

为所有客户端配置 NFS 导出

您可以指定允许所有客户端进行根访问的 NFS 导出。

为所有客户端配置 NFS 导出

- 对于适用于 O FSx penZFS 的文件系统卷，请配置以下 NFS 导出设置：
  - 客户端地址：输入 *。
  - NFS 选项：输入 `rw,no_root_squash`。

FSx 为你的 OpenZFS 创建转账地点

要创建该位置，您需要一个现有 FSx 的 OpenZFS 文件系统。如果你还没有，请参阅《亚马逊版 OpenZFS 用户指南》中的 Ama FSx zon for O penZ FS 入门。 FSx

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择位置，然后选择创建位置。
3. 对于位置类型，请选择 Amazon FSx。

   您可在稍后将此位置配置为源或目标。
4. 对于FSx 文件系统，请选择要用作位置的 OpenZFS 文件系统。 FSx
5. 在装载路径中，输入您 FSx 的 for OpenZFS 文件系统的装载路径。

   路径必须以 `/fsx` 开头，并且可以是文件系统中任何现有的目录路径。当该位置用作源时，会从装载路径 DataSync 读取数据。将该位置用作目标时， DataSync 会将所有数据写入装载路径。如果未提供子目录，则 DataSync 使用根卷目录（例如`/fsx`）。
6. 对于安全组，最多可选择五个安全组，为你 FSx 的 openZFS 文件系统提供网络访问权限。

   安全组必须提供对用于 OpenZFS 文件系统的网络端口 FSx 的访问权限。文件系统必须允许来自安全组的网络访问。

   有关安全组的更多信息，请参阅亚马逊版 OpenZFS 用户指南中的使用 Amazon VPC FSx 进行文件系统访问控制。
7. （可选）展开其他设置，对于 NFS 版本，选择用于访问文件系统的 DataSync NFS 版本。

   默认情况下， DataSync 使用 NFS 版本 4.1。
8. （可选）在 “密钥” 和 “值” 字段中输入值以标记 OpenZFS 文件系统。 FSx

   标签可帮助您管理、筛选和搜索位置。我们建议至少为您的位置创建一个名称标签。

9. 选择创建位置。

使用 AWS CLI

要创建 FSx 适用于 OpenZFS 的位置，请使用 AWS CLI

1. 复制以下 `create-location-fsx-open-zfs` 命令：

```
aws datasync create-location-fsx-open-zfs \
   --fsx-filesystem-arn arn:aws:fsx:region:account-id:file-system/filesystem-id \
   --security-group-arns arn:aws:ec2:region:account-id:security-group/group-id \
   --protocol NFS={}
```

2. 在命令中指定以下必要的选项：

   - 对于 `fsx-filesystem-arn`，请指定位置文件系统的完全限定的 Amazon 资源名称（ARN）。这包括您的文件系统所在 AWS 区域 的位置 AWS 账户、您的和文件系统 ID。
   - 对于`security-group-arns`，请指定 Amazon EC2 安全组的 ARN，该安全组提供对 OpenZFS 文件系统首选 FSx 子网的网络接口的访问权限。这包括您的 Amazon EC2 实例所在的位置 AWS 账户、您的和安全组 ID。 AWS 区域

     有关安全组的更多信息，请参阅《亚马逊版 OpenZFS 用户指南》中的 Amazon FSx VPC 文件系统访问控制。
   - 对于`protocol`，请指定 DataSync用于访问您的文件系统的协议。（DataSync 目前仅支持 NFS。）

3. 运行 命令。您会收到一条回复，表明您刚刚创建的位置。

```
{
    "LocationArn": "arn:aws:datasync:us-west-2:123456789012:location/loc-
abcdef01234567890"
}
```

在 AMAZON 上 FSx 为 NetApp ONTAP 配置传输

要将数据传输到或传出您的 Amazon FSx for NetApp ONTAP 文件系统，您必须创建一个 AWS DataSync 传输位置。 DataSync 可以将此位置用作传输数据的源或目的地。

FSx 为 ONTAP 文件系统提供 DataSync 访问权限

要访问 FSx 适用于 ONTAP 的文件系统，请使用虚拟私有云 (VPC) 中的网络接口在文件系统上 DataSync 安装存储虚拟机 (SVM)。 DataSync 只有在创建包含您的 for ONTAP 位置的任务时，才会在文件系统的首选子网中创建这些网络接口。 FSx

> **Note**
>
> VPCs 与一起使用的 DataSync 必须具有默认租期。 VPCs 不支持专用租赁。

DataSync 可以使用网络文件系统 (NFS) 或服务器消息块 (SMB) 协议连接到 for ONTAP 文件系统的 SVM 并复制数据。 FSx

主题

- 使用 NFS 协议
- 使用 SMB 协议
- 协议不受支持
- 选择正确的协议
- 访问 SnapLock 卷

使用 NFS 协议

在 NFS 协议中， DataSync 使用用户 ID (UID) 和组 ID (GID) 为AUTH_SYS的安全机制向您的 SVM 0 进行身份验证。

> **Note**
>
> DataSync 目前仅支持 NFS 版本 3 和 ONTAP 位置。 FSx

使用 SMB 协议

在 SMB 协议中， DataSync 使用您提供的凭据对 SVM 进行身份验证。

支持的 SMB 版本

默认情况下，根据与 SMB 文件服务器的协商 DataSync 自动选择 SMB 协议的版本。您也可以配置 DataSync 为使用特定版本，但我们建议只有 DataSync 在无法自动与 SMB 文件服务器协商时才这样做。出于安全起见，建议使用 SMB 版本 3.0.2 或更高版本。

有关 DataSync 控制台和 API 中使用您的 for ONTAP 位置配置 SMB 版本的选项列表， FSx 请参阅下表：

控制台选项	API 选项	说明
自动	`AUTOMATIC`	DataSync 和 SMB 文件服务器协商它们在 2.1 和 3.1.1 之间相互支持的最高版本的 SMB。 这是默认选项，也是推荐的选项。如果您改为选择文件服务器不支持的特定版本，则可能会出现 `Operation Not Supported` 错误。
SMB 3.0.2	`SMB3`	将协议协商限制为仅 SMB 版本 3.0.2。
SMB 2.1	`SMB2`	将协议协商限制为仅 SMB 版本 2.1。
SMB 2.0	`SMB2_0`	将协议协商限制为仅 SMB 版本 2.0。

所需的权限

您必须为 SVM 中的本地用户或 Microsoft Active Directory 中的域用户提供 DataSync 挂载和访问您的文件、文件夹和文件元数据的必要权限。

如果您在 Active Directory 中提供用户，请注意以下事项：

- 如果您正在使用 AWS Directory Service for Microsoft Active Directory，则该用户必须是AWS 委派 FSx管理员组的成员。
- 如果您使用的是自我管理的 Active Directory，则该用户必须属于以下两个组中的一个：
  - 域管理员组，这是默认的委派管理员组。
  - 一个自定义的委派管理员组，其用户权限 DataSync 允许复制对象所有权权限和 Windows 访问控制列表 (ACLs)。

> **Important**
>
> 部署文件系统后，无法更改委派管理员组。您必须重新部署文件系统或从备份中恢复文件系统，才能使用具有以下用户权限的自定义委派管理员组来复制元数据。 DataSync

用户权限	说明
作为操作系统的一部分（SE_TCB_NAME ）	DataSync 允许复制对象所有权、权限、文件元数据和 NTFS 自由访问列表 () DACLs。 此用户权限通常授予域管理员组和备份操作员组（均为默认 Active Directory 组）的成员。
管理审计和安全日志（SE_SECURITY_NAME ）	DataSync 允许复制 NTFS 系统访问控制列表 (SACLs)。 此用户权限通常授予域管理员组的成员。

- 如果要复制 Windows ACLs 并在使用 SMB 的 ONTAP 文件系统（或其他使用 SMB 的文件系统）之间 FSx 进行传输，则您提供的用户 DataSync必须属于同一 Active Directory 域或其域之间具有 Active Directory 信任关系。

所需的身份验证协议

DataSync 要访问您的 SMB 共享，您的 FSx for ONTAP 文件系统必须使用 NTLM 身份验证。DataSync 无法访问 FSx 使用 Kerberos 身份验证的 ONTAP 文件系统。

DFS 命名空间

DataSync 不支持 Microsoft 分布式文件系统 (DFS) 命名空间。我们建议您在创建 DataSync 位置时改为指定底层文件服务器或共享。

协议不受支持

DataSync 无法使用 iSCSI（互联网小型计算机系统接口）协议访问 FSx ONTAP 文件系统。

选择正确的协议

要在 ONTAP 迁移中 FSx 保留文件元数据，请将 DataSync源位置和目标位置配置为使用相同的协议。在支持的协议之间，SMB 保留保真度最高的元数据（有关详细信息，请参阅 了解如何 DataSync 处理文件和对象元数据）。

从通过 NFS 为用户提供服务的 Unix (Linux) 服务器或网络附属存储（NAS）共享迁移时，请执行以下操作：

1. 为 Unix (Linux) 服务器或 NAS 共享创建一个 NFS 位置。（这将是您的源位置。）
2. 使用 Un FSx i x 安全模式配置要将数据传输到的 ONTAP 卷。
3. FSx 为您的 for ONTAP 文件系统创建一个为 NFS 配置的位置。（这将是您的目标位置。）

从通过 SMB 为用户提供服务的 Windows 服务器或 NAS 共享迁移时，请执行以下操作：

1. 为 Windows 服务器或 NAS 共享创建一个 SMB 位置。（这将是您的源位置。）
2. 使用 NTFS 安全模式配置要将数据传输到的 ONTAP 卷。 FSx
3. FSx 为您的 onTAP 文件系统创建一个为 SMB 配置的位置。（这将是您的目标位置。）

如果您 FSx 的 onTAP 环境使用多个协议，我们建议您与 AWS 存储专家合作。要了解多协议访问的最佳实践，请参阅使用 Amazon for ONTAP 启用多协议工作负载 FSx 。 NetApp

访问 SnapLock 卷

如果您要将数据传输到 for ONTAP 文件系统上的SnapLock 卷，请确保在传输期间该卷上禁用了自动提交和卷追加模式的 SnapLock 设置。 FSx 完成数据传输后，可以重新启用这些设置。

FSx 为 ONTAP 创建您的传输位置

要创建该位置，您需要一个现有的 F FSx or ONTAP 文件系统。如果您没有，请参阅《亚马逊版 ONTAP 用户指南》中的亚马 FSx 逊 NetApp ONTA NetApp P FSx 版入门。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于位置类型，请选择 Amazon FSx。

   您可在稍后将此位置配置为源或目标。

4. 对于FSx 文件系统，请 FSx 选择要用作位置的 ONTAP 文件系统。
5. 对于存储虚拟机，选择要将数据复制到或从中复制数据的文件系统中的存储虚拟机（SVM）。
6. 对于挂载路径，指定您将在其中复制数据的指向该 SVM 中的文件共享路径。

   您可以指定连接路径（也称为挂载点）、qtree 路径（用于 NFS 文件共享）或共享名称（用于 SMB 文件共享）。例如，您的挂载路径可能是 `/vol1`、`/vol1/tree1` 或 `/share1`。

   > **Tip**
   >
   > 不要在 SVM 的根卷中指定路径。有关更多信息，请参阅《Amazon FSx FSx for ONTAP 用户指南》中的管理 NetApp ONTAP 存储虚拟机。

7. 对于安全组，最多选择五个 Amazon EC2 安全组，这些安全组提供对文件系统首选子网的访问权限。

   安全组必须允许以下端口的出站流量（取决于您使用的协议）：

   - NFS — TCP 端口 111、635 和 2049
   - SMB — TCP 端口 445

   文件系统的安全组还必须允许同一端口上的入站流量。
8. 对于协议，选择 DataSync 用于访问文件系统 SVM 的数据传输协议。

   有关更多信息，请参阅 选择正确的协议。

   NFS

   DataSync 使用 NFS 版本 3。

   SMB

   配置 SMB 版本、用户、密码和 Active Directory 域名（如需要），以访问 SVM。

   - （可选）展开其他设置并选择 DataSync 要在访问 SVM 时使用的 SMB 版本。

     默认情况下，根据与 SMB 文件服务器的协商 DataSync 自动选择版本。有关更多信息，请参阅 使用 SMB 协议。
   - 对于用户，输入一个用户名称，该用户可以挂载和访问 SVM 中要传输的文件、文件夹和元数据。

有关更多信息，请参阅 使用 SMB 协议。

- 对于密码，输入您指定可访问 SVM 的用户的密码。
- （可选）对于 Active Directory 域名，请输入 SVM 所属的 Active Directory 的完全限定域名 (FQDN)。

  如果您的环境中有多个域，则配置此设置 DataSync 可确保连接到正确的 SVM。

9. （可选）在 “密钥” 和 “值” 字段中输入值 FSx 以标记 ONTAP 文件系统。

   标签可帮助您管理、筛选和搜索 AWS 资源。我们建议至少为您的位置创建一个名称标签。

10. 选择创建位置。

使用 AWS CLI

要创建 FSx 适用于 ONTAP 的位置，请使用 AWS CLI

1. 复制以下 `create-location-fsx-ontap` 命令：

```
aws datasync create-location-fsx-ontap \
   --storage-virtual-machine-arn arn:aws:fsx:region:account-id:storage-virtual-
machine/fs-file-system-id \
   --security-group-arns arn:aws:ec2:region:account-id:security-group/group-id \
   --protocol data-transfer-protocol={}
```

2. 在命令中指定以下必要的选项：

   - 对于 `storage-virtual-machine-arn`，请指定您要向其或从其数据复制的文件系统中的存储虚拟机（SVM）的完全限定的 Amazon 资源名称（ARN）。

     此 ARN 包括您的文件系统所在 AWS 区域 的位置、您的 AWS 账户、文件系统和 SVM。 IDs

   - 对于 `security-group-arns`，请指定 Amazon EC2 安全组的 ARN，这些安全组提供对文件系统首选子网的 网络接口 的访问权限。

     这包括您的 Amazon EC2 实例所在的位置 AWS 账户、您的和您的安全组 IDs。 AWS 区域 您最多可以指定五个安全组 ARNs。

     有关安全组的更多信息，请参阅《Amazon for NetApp ONTAP 用户指南》中的使用 Amazon VPC FSx 进行文件系统访问控制。

   - 对于`protocol`，请配置 DataSync 用于访问文件系统 SVM 的协议。

- 对于 NFS，您可以使用以下默认配置：

  ```
  --protocol NFS={}
  ```

- 对于 SMB，您必须指定可以访问 SVM 的用户名和密码：

  ```
  --protocol SMB={User=smb-user,Password=smb-password}
  ```

3. 运行命令。

   您会收到一条回复，表明您刚刚创建的位置。

   ```
   {
       "LocationArn": "arn:aws:datasync:us-west-2:123456789012:location/loc-
   abcdef01234567890"
   }
   ```

使用其他云存储进行传入或传出其他云存储 AWS DataSync

使用 AWS DataSync，您可以在其他一些云提供商和 AWS 存储服务之间传输数据。有关更多信息，请参阅我可以在哪里传输数据 DataSync？

主题

- 计划传入或传出第三方云存储系统
- 使用谷歌云存储配置 AWS DataSync 传输
- 配置使用 Microsoft Azure Blob Storage 的传输
- 使用 Microsoft Azure Files SMB 共享配置 AWS DataSync 传输
- 配置使用其他云对象存储的 传输

计划传入或传出第三方云存储系统

在计划跨云数据传输时，请考虑以下几点：

- 使用代理：只有在使用基本模式任务时，才需要代理来访问其他云中的存储。增强模式任务无需代理。如果决定使用代理，在从云提供商的 S3 兼容对象存储中进行传输时可以将代理部署为 Amazon EC2 实例，或分别作为 Google 计算引擎或 Azure 虚拟机，用于从这些特定的存储服务进行传输。从 Google 和 Azure 中的文件系统进行传输时，建议将代理部署为 Google 或 Azure 虚拟机，以使代

理尽可能靠近文件系统。此外，还可以将代理中的数据 DataSync 压缩到 AWS，这有助于降低出口成本。 DataSync 提供了经过验证的云位置列表，这些位置提供了所需的 Amazon S3 API 兼容性。

- 其他云的对象存储端点：第三方云提供商的存储端点通常是特定区域或账户专属的。区域终端节点与指定的存储桶名称一起用作 DataSync 对象存储位置的服务器。
- 源对象的存储类别：与 Amazon S3 一样，一些云提供商支持使用存档层，要求先进行还原操作才能访问存档对象。例如，必须检索 Azure Blob 存档层中的对象以进行标准访问，然后才能进行数据传输。Google Cloud Storage 存档层中的对象可立即访问且无需恢复，但是直接访问存档层会产生检索费用。在开始数据传输之前，请查阅跨云存储类别文档，确定访问要求和检索费用。有关在 Amazon S3 中还原已归档对象的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的恢复已归档的对象。
- 对象存储访问权限：在第三方云提供商之间传输数据需要通过身份验证密钥的形式，访问其他云的对象存储。例如，要提供对 Google Cloud Storage 的访问权限，您需要配置一个 DataSync 对象存储位置，该位置连接到 Google Cloud Storage XML API，并使用基于哈希的消息身份验证码 (HMAC) 密钥为您的服务帐号进行身份验证。对于 Azure Blob 存储，你可以配置一个专用 Azure Blob DataSync 位置，该位置使用 SAS 令牌进行身份验证。 DataSync AWS Secrets Manager 用于安全存储您的对象存储凭证。有关更多信息，请参阅保护存储位置凭证。
- 对象标签支持：
  - 与 Amazon S3 不同，并非所有云提供商都支持对象标签。 DataSync 如果云提供商不支持通过 Amazon S3 API 进行对象标签，或者您提供的凭证不足以检索标签，则在尝试从源位置读取标签时，任务可能会失败。 DataSync提供了一个任务选项，用于在不支持对象标签或您不想保留对象标签时关闭对对象标签的读取和复制。请查阅云提供商文档确定是否支持对象标签，并在验证传输任务的对象标签设置后再启动传输。
  - 可使用 Amazon S3 API 来检查云提供商是否会返回 `get-object-tagging` 请求。有关更多信息，请参阅《AWS CLI 命令参考》中的 get-object-tagging。

    支持对象标签的云提供商将返回类似以下示例的响应：

    ```
    aws s3api get-object-tagging --bucket BUCKET_NAME --endpoint- url=https://
    BUCKET_ENDPOINT --key prefix/file1

    {

        "TagSet": []

    }
    ```

    不支持 `get-object-tagging` 的云提供商将返回以下消息：

```
aws s3api get-object-tagging --bucket BUCKET_NAME --endpoint- url=https://
BUCKET_ENDPOINT --key prefix/file1

An error occurred (OperationNotSupported) when calling the GetObjectTagging
 operation: The operation is not supported for this resource
```

- 请求和数据传出的相关费用：从云对象存储传输数据会产生与数据读取和传出相关的请求和传出费用。不同云提供商以及不同存储类别（如果适用）的请求费用会有所不同。请查阅云提供商文档，了解与计划读取的存储类别相关的请求的具体费用。有关数据传输的请求费用的 DataSync 概述，请参阅使用时评估 S3 请求费用 DataSync和AWS DataSync 定价。从特定的云提供商传出数据会产生传出费用。数据传输成本因云提供商而异，此外也取决于存储数据的区域。
- 对象存储请求速率：云提供商对象存储平台的性能和请求速率特征各不相同。请查看其他云提供商的请求速率，确定请求限制的适用范围。对于由多个代理组成的高度并行化传输，需要提前规划，以满足特定的分区需求或性能提升要求。

  Amazon S3 记录了请求速率，您可以据此构建解决方案。Amazon S3 的请求速率按分区前缀计算，并且可以跨多个前缀扩展。有关更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的最佳实践设计模式：优化 Amazon S3 性能。

使用谷歌云存储配置 AWS DataSync 传输

借 AWS DataSync助，您可以在 Google 云存储空间和以下 AWS 存储服务之间传输数据：

- Amazon S3
- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的亚马逊

要开始传输设置，请为 Google Cloud Storage 创建一个位置。此位置可以用作传输源或传输目标位置。仅当您在谷歌云存储与 Amazon EFS 或 Amazon 之间传输数据时 FSx，或者使用基本模式任务时，才需要 DataSync 代理。Google Cloud Storage 和 Amazon S3 之间的增强模式数据传输不需要代理。

> **Note**
>
> 要在 Google Cloud Storage 与之间建立私有云连接 AWS，请使用带代理的基本模式。

概述

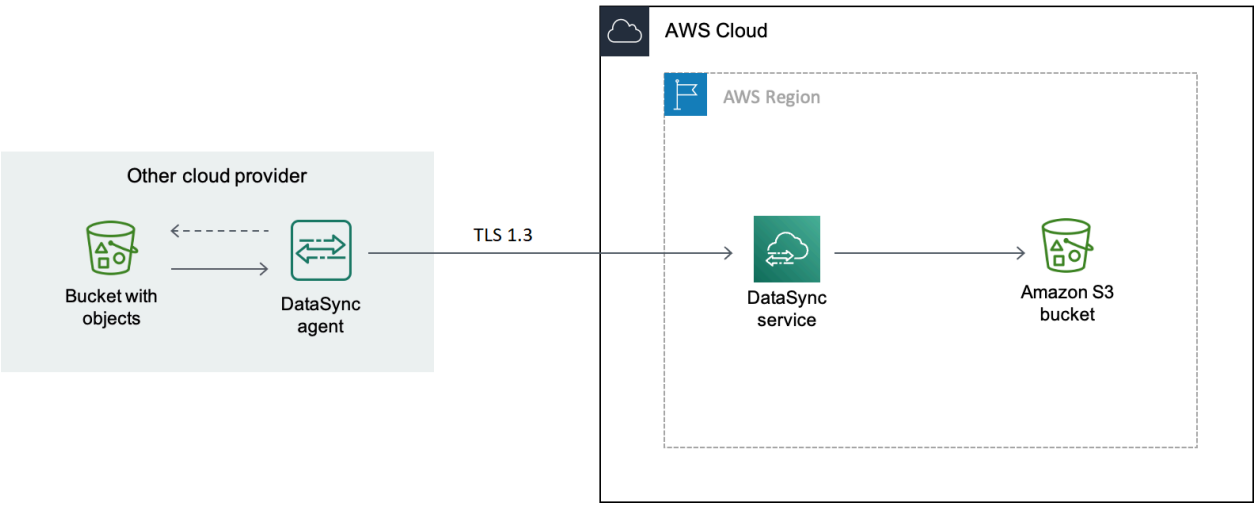
DataSync 使用谷歌云存储 XML API 进行数据传输。此 API 提供了兼容 Amazon S3 的接口，用于通过 Google Cloud Storage 存储桶读取和写入数据。

使用基本模式进行传输时，可以在 Google Cloud Storage 或 Amazon VPC 中部署代理。

Agent in Google Cloud

1. 您在自己的 Google 云环境中部署 DataSync 代理。
2. 代理使用 HMAC 散列消息认证码（HMAC）密钥读取 Google Cloud Storage 存储桶。
3. Google Cloud Storage 存储分区中的对象使用公共终端节点通过 TLS 1.3 安全地传输到中。AWS Cloud
4. 该 DataSync 服务将数据写入您的 S3 存储桶。

下图说明了该传输。

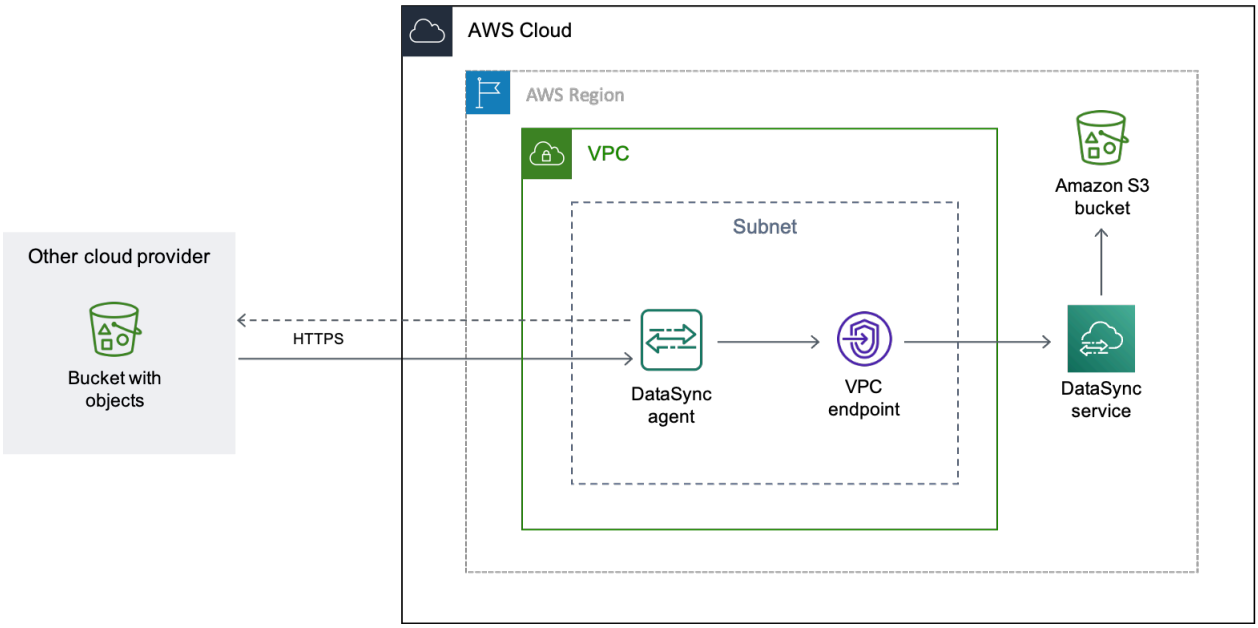


Agent in your VPC

1. 您在 AWS 环境中的虚拟私有云 (VPC) 中部署 DataSync 代理。
2. 代理使用 HMAC 散列消息认证码（HMAC）密钥读取 Google Cloud Storage 存储桶。

3. 您的 Google 云存储分区中的对象使用私有 VPC 终端节点通过 TLS 1.3 安全地传输到中。 AWS Cloud
4. 该 DataSync 服务将数据写入您的 S3 存储桶。

下图说明了该传输。



成本

与此迁移相关的可能费用包括：

- 运行 Google Compute En gine 虚拟机 (VM) 实例（如果您在 Google Cloud 中部署 DataSync 代理）
- 运行 A mazon EC2 实例（如果您在其中的 VPC 中部署 DataSync 代理 AWS）
- 使用传输数据 DataSync，包括与 Google 云存储和 Amazon S3 相关的请求费用（如果 S3 是您的传输地点之一）
- 将数据从 Google Cloud Storage 中传出
- 在 Amazon S3 中存储数据

先决条件

开始之前，请完成以下先决条件（如果尚未完成）：

- 使用要传入 AWS的对象，创建 Google Cloud Storage 存储桶。
- 注册 AWS 账户。
- 创建 Amazon S3 存储桶，用于在对象存入 AWS后存储这些对象。

为 Google Cloud Storage 存储桶创建 HMAC 密钥

DataSync 使用与您的 Google 服务帐号关联的 HMAC 密钥对您要从中传输数据的存储桶进行身份验证和读取。（有关如何创建 HMAC 密钥的详细说明，请参阅 Google Cloud Storage 文档。）

创建 HMAC 密钥

1. 为您的 Google 服务帐号创建 HMAC 密钥。
2. 请确保您的 Google 服务帐号至少具有 `Storage Object Viewer` 权限。
3. 将 HMAC 密钥访问 ID 和私有密钥存储在安全位置。

   稍后您将需要这些物品来配置您的 DataSync 来源位置。

第 2 步：配置网络

仅当使用 DataSync 代理进行传输时，才需要进行网络配置。此迁移的网络要求取决于您选择部署代理的位置。

对于谷歌云端的 DataSync 代理

如果您想在 Google Cloud 中托管 DataSync 代理，请将您的网络配置为允许通过公共端点进行 DataSync 传输。

对于您的 VPC 中的 DataSync 代理

如果要在中托管代理 AWS，则需要一个带有接口终端节点的 VPC。 DataSync 使用 VPC 终端节点来促进传输。

为 VPC 端点配置您的网络

1. 如果您没有 VPC，请在与您的 S3 存储桶 AWS 区域 相同的 VPC 中创建。
2. 为 VPC 创建私有子网。
3. 为创建 VPC 服务终端节点 DataSync。
4. 将您的网络配置为允许通过 VPC 服务终端节点进行 DataSync 传输。

为此，请修改与 VPC 服务端点关联的安全组。

步骤 3：创建 DataSync 代理（可选）

只有在使用基本模式任务时才需要 DataSync 代理。如果使用增强模式在 Google Cloud Storage（GCS）和 Amazon S3 之间进行传输，则无需代理。如果您想使用基本模式，则需要一个可以访问您的 GCS 存储桶的 DataSync 代理。

对于 Google Cloud

在这种情况下， DataSync 代理将在您的 Google Cloud 环境中运行。

开始之前：安装 Google Cloud CLI。

为 Google Cloud 创建代理

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择 代理，然后选择 创建代理。
3. 对于 Hypervisor VMware ESXi，选择，然后选择 “下载映像” 以下载包含代理的`.zip`文件。
4. 打开终端。运行以下命令来对映像解压缩：

```
unzip AWS-DataSync-Agent-VMWare.zip
```

5. 运行以下命令提取以 `aws-datasync` 开头的代理 `.ova` 文件中的内容：

```
tar -xvf aws-datasync-2.0.1655755445.1-x86_64.xfs.gpt.ova
```

6. 运行以下 Google Cloud CLI 命令，将代理的 `.vmdk` 文件导入到 Google Cloud 中：

```
gcloud compute images import aws-datasync-2-test \
   --source-file INCOMPLETE-aws-datasync-2.0.1655755445.1-x86_64.xfs.gpt-disk1.vmdk
 \
   --os centos-7
```

> **Note**
>
> 导入 `.vmdk` 文件最多可能需要两个小时。

7. 为您刚刚导入的代理映像创建并启动 VM 实例。

   该实例需要为代理进行以下配置。（有关如何创建实例的详细说明，请参阅 Google 云计算引擎文档。）

   - 对于计算机类型，请选择下列选项之一：
     - e2-standard-8 — 适用于处理多达 2000 万个对象的 DataSync任务执行。
     - e2-standard-16 — 适用于处理超过 2000 万个对象的 DataSync任务执行。
   - 对于启动磁盘设置，请转到自定义映像部分。然后选择您刚刚导入的 DataSync 代理映像。
   - 对于服务帐号设置，请选择 Google 服务帐号（与您在第 1 步 中使用的帐号相同）。
   - 对于防火墙设置，请选择允许 HTTP（端口 80）通信的选项。

     要激活您的 DataSync 代理，必须在代理上打开端口 80。不要求该端口可供公开访问。激活后， DataSync 关闭端口。

8. VM 实例运行后，记录其公有 IP 地址。

   您需要此 IP 地址来激活代理。

9. 返回 DataSync 控制台。在下载代理映像的 创建代理屏幕上，执行以下操作以激活代理：

   - 对于端点类型，选择公共服务端点选项（例如，美国东部俄亥俄州的公共服务端点）。
   - 对于激活密钥，选择自动从代理处获取激活密钥。
   - 对于代理地址，输入您刚创建的代理 VM 实例的公有 IP 地址。
   - 选择 获取密钥。

10. 为您的代理指定名称，然后选择 创建代理。

代理处于在线状态，可以随时传输数据。

对于 VPC

在这种情况下，代理在与您的关联的 VPC 中作为 Amazon EC2 实例运行 AWS 账户。

开始之前：设置 AWS Command Line Interface（AWS CLI）。

为 VPC 创建代理

1. 打开终端。请务必将您的 AWS CLI 个人资料配置为使用与您的 S3 存储桶关联的账户。
2. 复制以下 命令。将 *vpc-region* 替换为您的 VPC 所在的 AWS 区域（例如，`us-east-1`）。

```
aws ssm get-parameter --name /aws/service/datasync/ami --region vpc-region
```

3. 运行命令。在输出中，记录 "Value" 属性。

   此值是您指定区域的 DataSync 亚马逊系统映像 (AMI) ID。例如，AMI ID 可能类似于 ami-1234567890abcdef0。

4. 复制以下 URL。再次将 *vpc-region* 替换为您的 VPC 所在的 AWS 区域 。接着，将 *ami-id* 替换为您在上一步中记录的 AMI ID。

```
https://console.aws.amazon.com/ec2/v2/home?region=vpc-
region#LaunchInstanceWizard:ami=ami-id
```

5. 将该 URL 粘贴到浏览器中。

   屏幕上 AWS 管理控制台 显示了 Amazon EC2 实例启动页面。

6. 对于实例类型，请为 DataSync代理选择一个推荐的 Amazon EC2 实例。
7. 对于密钥对，选择现有密钥对或新建一个密钥对。
8. 对于网络设置，选择要部署代理的 VPC 和子网。
9. 选择 启动实例。
10. 在 Amazon EC2 实例运行后，选择您的 VPC 端点。
11. 激活您的代理。

第 4 步：为您的 Google 云存储分区创建 DataSync 源位置

要为您的 Google Cloud Storage 存储分区设置 DataSync 位置，您需要在步骤 1 中创建的 HMAC 密钥的访问权限 ID 和密钥。

创建 DataSync 源位置

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，选择 对象存储。
4. 对于服务器，输入 **storage.googleapis.com**。
5. 对于存储桶名称，输入您的 Google Cloud Storage 存储桶的名称。
6. 对于 文件夹，请输入对象前缀。

DataSync 仅复制带有此前缀的对象。

7. 如果传输需要代理，请选择使用代理，然后选择在步骤 3 中创建的代理。
8. 展开其他设置。对于服务器协议，请选择 HTTPS。对于服务器端口，请选择 443。
9. 向下滚动到身份验证部分。确保选中 需要凭据复选框，然后执行以下操作：
   - 对于访问密钥，输入您的 HMAC 密钥访问 ID。
   - 对于密钥，可以直接输入 HMAC 密钥的密钥，也可以指定包含 AWS Secrets Manager 密钥的密钥。有关更多信息，请参阅为存储位置提供凭证。
10. 选择创建位置。

步骤 5：为您的 S3 存储桶创建 DataSync 目标位置

你需要一个 DataSync 位置来存放你想要的数据存放位置。

创建 DataSync 目标位置

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 为 S3 存储桶创建 DataSync 位置。

   如果您在您的 VPC 中部署了 DataSync 代理，则本教程假定 S3 存储桶与您的 VPC 和 DataSync 代理位于同一 AWS 区域 位置。

步骤 6：创建并启动 DataSync 任务

配置好源位置和目标位置后，就可以开始将数据移入 AWS了。

创建并启动 DataSync 任务

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 在配置源位置页面上，执行以下操作：
   a. 选择 选择现有位置。
   b. 选择您在第 4 步 中创建的源位置，然后选择 下一步。

4. 在配置目标位置页面上，执行以下操作：

   a. 选择 选择现有位置。

   b. 选择您在第 5 步 中创建的目标位置，然后选择 下一步。

5. 在 配置设置 页面上，执行以下操作：

   a. 在 数据传输配置下，展开 其他设置，然后清除 复制对象标签复选框。

   > **Important**
   >
   > 由于 Google Cloud Storage XML API 不支持读取或写入对象标签，因此如果您尝试复制对象标签，您的 DataSync 任务可能会失败。

   b. 配置所需的任何其他任务设置，然后选择 下一步。

6. 在 审核页面上，审核您的设置，然后选择 创建任务。

7. 在任务的详细信息页面上，选择开始，然后选择以下选项之一：

   - 如需运行任务而不进行修改，请选择从默认值开始。
   - 如需修改任务再运行，请选择使用覆盖选项开始。

任务完成后，您会在 S3 存储桶中看到 Google Cloud Storage 存储桶中的对象。

配置使用 Microsoft Azure Blob Storage 的传输

使用 AWS DataSync，您可以在Microsoft Azure Blob Storage（包括 Azure Data Lake Storage Gen2 blob 存储）和以下 AWS 存储服务之间传输数据：

- Amazon S3
- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的亚马逊

要设置此类传输，您需要为您的 Azure Blob Storage 创建一个位置。您可以将此位置用作传输源或传输目标位置。只有在 Azure Blob 和 Amazon EFS 或 Amazon 之间传输数据时 FSx，或者使用基本模

式任务时，才需要 DataSync 代理。使用增强模式在 Azure Blob 和 Amazon S3 之间传输数据时不需要代理。

提供对您的 DataSync 访问权限 Azure Blob Storage

您的 DataSync 访问方式Azure Blob Storage取决于多个因素，包括您是向 blob 存储传输还是从 blob 存储进行传输，以及您使用的共享访问签名 (SAS) 令牌的类型。您的对象还必须位于 DataSync 可以使用的访问层中。

主题

- SAS 令牌
- 访问层

SAS 令牌

SAS 令牌指定您的 blob 存储的访问权限。（有关 SAS 的详细信息，请参阅 Azure Blob Storage文档。）

您可以生成 SAS 令牌以提供不同级别的访问权限。 DataSync 支持具有以下访问权限级别的令牌：

- Account
- Container

DataSync 所需的访问权限取决于您的令牌的范围。没有正确的权限可能会导致您的传输失败。例如，如果您要将带有标签的对象移至 Azure Blob Storage，但您的 SAS 令牌没有标签权限，则传输将不会成功。

主题

- 用于账户级访问的 SAS 令牌权限
- 容器级访问的 SAS 令牌权限
- SAS 过期策略

用于账户级访问的 SAS 令牌权限

DataSync 需要具有以下权限的账户级访问令牌（取决于您是向还是从中Azure Blob Storage转移）。

Transfers from blob storage

- 允许的服务 – Blob
- 允许的资源类型 - 容器、对象

  如果您不包括这些权限，则 DataSync 无法传输您的对象元数据，包括对象标签。
- 允许的权限 - 读取、列出
- 允许的 blob 索引权限 — Read/Write （如果 DataSync 要复制对象标签）

Transfers to blob storage

- 允许的服务 – Blob
- 允许的资源类型 - 容器、对象

  如果您不包括这些权限，则 DataSync 无法传输您的对象元数据，包括对象标签。
- 允许的权限-读取、写入、列出、删除（如果您 DataSync 想移除传输源中不存在的文件）
- 允许的 blob 索引权限 — Read/Write （如果 DataSync 要复制对象标签）

容器级访问的 SAS 令牌权限

DataSync 需要具有以下权限的容器级访问令牌（取决于您是向还是从中Azure Blob Storage转移）。

Transfers from blob storage

- 读取
- 列表
- 标签（如果 DataSync 要复制对象标签）

> **Note**
>
> 在 Azure 门户中生成 SAS 令牌时，您无法添加标签权限。要添加标签权限，请改为使用 Azure Storage Explorer 应用程序生成令牌或生成提供账户级访问权限的 SAS 令牌。

Transfers to blob storage

- 读取

- 写入
- 列表
- 删除（如果您 DataSync 要删除不在传输源中的文件）
- 标签（如果 DataSync 要复制对象标签）

> **Note**
>
> 在 Azure 门户中生成 SAS 令牌时，您无法添加标签权限。要添加标签权限，请改为使用 Azure Storage Explorer 应用程序生成令牌或生成提供账户级访问权限的 SAS 令牌。

SAS 过期策略

在您预计完成传输之前，请确保您的 SAS 不会过期。有关配置 SAS 过期策略的信息，请参阅 Azure Blob Storage 文档。

如果 SAS 在转移期间过期，则 DataSync 无法再访问您的Azure Blob Storage位置。（您可能会看到无法打开目录错误。）如果发生这种情况，请使用新的 SAS 令牌更新您的位置并重新启动 DataSync 任务。

访问层

从中传输时Azure Blob Storage， DataSync 可以复制热层和冷层中的对象。对于存档访问层中的对象，必须先将这些对象重新水合到热层或冷层，然后才能复制它们。

传输到时Azure Blob Storage， DataSync 可以将对象复制到热、冷和存档访问层。如果您要将对象复制到存档访问层中，如果您正在尝试验证目标中的所有数据，则 DataSync 无法验证传输。

DataSync 不支持冷访问层。有关访问层的更多信息，请参阅 Azure Blob Storage文档。

使用 Azure Blob Storage 传输的注意事项

在计划使用或从中Azure Blob Storage传输数据时 DataSync，需要记住一些注意事项。

主题

- 成本
- Blob 类型
- AWS 区域 可用性
- 复制对象标签

- 传输到 Amazon S3
- 删除传输目标中的目录
- 限制

成本

将数据移入或移出 Azure Blob Storage 的相关费用可能包括：

- 运行Azure虚拟机 (VM)（如果您在中部署 DataSync 代理Azure）
- 运行 A mazon EC2 实例（如果您在其中的 VPC 中部署 DataSync 代理 AWS）
- 使用传输数据 DataSync，包括与Azure Blob Storage和 Amazon S3 相关的请求费用（如果 S3 是您的传输地点之一）
- 将数据传入或传出 Azure Blob Storage
- 将数据存储在支持的AWS 存储服务中 DataSync

Blob 类型

DataSync 如何处理 blob 类型取决于你是转移到还是从中Azure Blob Storage传输。将数据移入 blob 存储时， DataSync 传输的对象或文件只能是块 blob。将数据移出 Blob 存储空间时， DataSync 可以传输块、页面和附加 blob。

有关 Blob 类型的更多信息，请参阅 Azure Blob Storage文档。

AWS 区域 可用性

您可以在支持的任意AWS 区域 位置创建Azure Blob Storage转账地点 DataSync。

复制对象标签

在传 DataSync 入或传出时保留对象标签的能力Azure Blob Storage取决于以下因素：

- 对象标签的大小 — DataSync 无法传输标签超过 2 KB 的对象。
- DataSync 是否配置为复制对象标签 — 默认情况下 DataSync 复制对象标签。
- 您的Azure存储账户使用的命名空间 — 如果您的Azure存储账户使用平面命名空间，则 DataSync 可以复制对象标签，但如果您的账户使用分层命名空间（的一项功能Azure Data Lake Storage Gen2），则无法复制对象标签。如果您尝试复制对象标签并且您的存储帐户使用分层命名空间，则 DataSync 任务将失败。

- 您的 SAS 令牌是否授权标记 – 复制对象标签所需的权限因令牌提供的访问级别而异。如果您尝试复制对象标签，但您的令牌没有正确的标记权限，则任务将失败。有关更多信息，请查看账户级访问令牌或容器级访问令牌的权限要求。

传输到 Amazon S3

传输到 Amazon S3 时， DataSync 不会传输Azure Blob Storage大于 5 TB 的对象或元数据大于 2 KB 的对象。

删除传输目标中的目录

传输到时Azure Blob Storage， DataSync 可以移除 blob 存储中传输源中不存在的对象。（您可以通过在 DataSync 控制台中清除 “保留已删除的文件” 设置来配置此选项。 您的 SAS 令牌还必须具有删除权限。 ）

以这种方式配置传输时，如果您的存储帐户使用的是分层命名空间，则 DataSync 不会删除 blob Azure 存储中的目录。在这种情况下，必须手动删除目录（例如使用 Azure Storage Explorer ）。

限制

在与 Azure Blob Storage 互相传输数据时，请记住以下限制：

- DataSync 在某个位置创建一些目录以帮助您进行传输。如果Azure Blob Storage是目标位置，并且您的存储帐户使用分层命名空间，则您可能会注意到该文件夹中存在特定于任务的子目录（例如`task-000011112222abcde`）。`/.aws-datasync` DataSync 通常会在传输后删除这些子目录。如果不这样做，只要任务未运行，您就可以自己删除这些特定任务的目录。
- DataSync 不支持使用 SAS 令牌仅访问Azure Blob Storage容器中的特定文件夹。
- 您不能提供 DataSync 用于访问您的 blob 存储空间的用户委托 SAS 令牌。

创建您的 DataSync 代理（可选）

只有在 Azure Blob 和 Amazon EFS 或 Amazon 之间传输数据时 FSx，或者使用基本模式任务时，才需要 DataSync 代理。使用增强模式在 Azure Blob 和 Amazon S3 之间传输数据时不需要代理。本部分介绍如何部署和激活代理。

> **Tip**
>
> 尽管您可以在 Amazon EC2 实例上部署代理，但使用 Microsoft Hyper-V 代理可能会减少网络延迟并提高数据压缩率。

Microsoft Hyper-V 代理

您可以使用Microsoft Hyper-V映像直接在中Azure部署 DataSync 代理。

> **Tip**
>
> 在继续之前，请考虑使用可以帮助您更快地在 Azure 中部署 Hyper-V 代理的 Shell 脚本。您可以获取更多信息并下载代码GitHub。
> 如果您使用脚本，则可以向前跳至有关 获取代理的激活密钥 的部分。

主题

- 先决条件
- 正在下载并准备您的代理
- 在 Azure 中部署您的代理
- 获取代理的激活密钥
- 激活您的代理

先决条件

要准备 DataSync 代理并将其部署到中Azure，必须执行以下操作：

- 在本地机器上启用 Hyper-V。
- 安装 PowerShell（包括 Hyper-V Module）。
- 安装 Azure CLI。
- 安装 AzCopy。

正在下载并准备您的代理

从 DataSync 控制台下载代理。在 Azure 中部署代理之前，必须将其转换为固定大小的虚拟硬盘 (VHD)。有关详情，请参阅 Azure 文档。

下载并准备您的代理

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择 代理，然后选择 创建代理。
3. 对于 虚拟机监控器，选择 Microsoft Hyper-V，然后选择 下载映像。

代理下载到包含 .vhdx 文件的 .zip 文件中。

4. 在本地机器上解压缩 .vhdx 文件。
5. 打开 PowerShell 并执行以下操作：

   a. 复制以下 Convert-VHD cmdlet：

   ```
   Convert-VHD -Path .\local-path-to-vhdx-file\aws-datasync-2.0.1686143940.1-
   x86_64.xfs.gpt.vhdx `
   -DestinationPath .\local-path-to-vhdx-file\aws-datasync-2016861439401-
   x86_64.vhd -VHDType Fixed
   ```

   b. 将 *local-path-to-vhdx-file* 的每个实例替换为本地计算机上的 .vhdx 文件位置。

   c. 运行 命令。

   您的代理现在是固定大小的 VHD（采用 .vhd 文件格式），可以部署到 Azure 中。

在 Azure 中部署您的代理

在中部署 DataSync 代理涉Azure及：

- 在 Azure 中创建托管磁盘
- 将您的代理上传到该托管磁盘
- 将托管磁盘连接到 Linux 虚拟机

要在 Azure 中部署您的代理

1. 在 PowerShell 中，转到包含您的代理 .vhd 文件的目录。
2. 运行 ls 命令并保存 Length 值（例如 85899346432）。

   这是代理映像的大小（以字节为单位），在创建可以容纳映像的托管磁盘时，您需要使用该大小。

3. 执行以下操作来创建托管磁盘：

   a. 复制以下 Azure CLI 命令：

   ```
   az disk create -n your-managed-disk `
   -g your-resource-group `
   -l your-azure-region `
   ```

```
--upload-type Upload `
--upload-size-bytes agent-size-bytes `
--sku standard_lrs
```

b. 将 *your-managed-disk* 替换为您的托管磁盘的名称。

c. 将 *your-resource-group* 替换为您的存储账户所属 Azure 资源组的名称。

d. 将 *your-azure-region* 替换为资源组所在的 Azure 区域。

e. 将 *agent-size-bytes* 替换为代理映像的大小。

f. 运行 命令。

此命令会创建一个带有标准 SKU 的空托管磁盘，您可以在其中上传 DataSync代理。

4. 要生成允许对托管磁盘进行写入访问的共享访问签名 (SAS)，请执行以下操作：

a. 复制以下 Azure CLI 命令：

```
az disk grant-access -n your-managed-disk `
-g your-resource-group `
--access-level Write `
--duration-in-seconds 86400
```

b. 将 *your-managed-disk* 替换为您创建的托管磁盘的名称。

c. 将 *your-resource-group* 替换为您的存储账户所属 Azure 资源组的名称。

d. 运行命令。

在输出中，记录 SAS URI。将代理上传到 Azure 时需要此 URI。

SAS 允许您向磁盘写入长达一个小时。这意味着您有一小时的时间将代理上传到托管磁盘。

5. 要在 Azure 中将代理上传到托管磁盘，请执行以下操作：

a. 复制以下 AzCopy 命令：

```
.\azcopy copy local-path-to-vhd-file sas-uri --blob-type PageBlob
```

b. 将 *local-path-to-vhd-file* 替换为代理 .vhd 文件在本地计算机上的位置。

c. 将 *sas-uri* 替换为运行 `az disk grant-access` 命令时获得的 SAS URI。

d. 运行命令。

6. 代理上传完成后，撤消对托管磁盘的访问权限。要执行这项操作，请复制以下 Azure CLI 命令：

```
az disk revoke-access -n your-managed-disk -g your-resource-group
```

a. 将 *your-resource-group* 替换为您的存储账户所属 Azure 资源组的名称。

b. 将 *your-managed-disk* 替换为您创建的托管磁盘的名称。

c. 运行 命令。

7. 执行以下操作将托管磁盘连接到新的 Linux 虚拟机：

a. 复制以下 Azure CLI 命令：

```
az vm create --resource-group your-resource-group `
--location eastus `
--name your-agent-vm `
--size Standard_E4as_v4 `
--os-type linux `
--attach-os-disk your-managed-disk
```

b. 将 *your-resource-group* 替换为您的存储账户所属 Azure 资源组的名称。

c. 将 *your-agent-vm* 替换为您可以记住的虚拟机名称。

d. 将 *your-managed-disk* 替换为您要装载到虚拟机的托管磁盘的名称。

e. 运行 命令。

您已经部署了代理。必须先激活代理，然后才能开始配置数据传输。

获取代理的激活密钥

要手动获取 DataSync 代理的激活密钥，请按照以下步骤操作。

或者，DataSync 可以自动为您获取激活密钥，但是这种方法需要一些网络配置。

获取代理的激活密钥

1. 在 Azure 门户中，通过选择 启用自定义存储账户设置并指定您的 Azure 存储账户，为代理启用虚拟机的启动诊断。

   为代理启用虚拟机的启动诊断后，您可以访问代理的本地控制台以获取激活密钥。

2. 仍在 Azure 门户网站时，前往您的虚拟机并选择 串行控制台。

3. 在代理的本地控制台中，使用以下默认凭证登录：

- 用户名 – **admin**
- 密码 – **password**

我们建议在某个时候至少更改代理的密码。在代理的本地控制台中，在主菜单上输入 **5**，然后使用 passwd 命令更改密码。

4. 输入 **0** 以获取代理的激活密钥。
5. 输入你正在使用 AWS 区域 的地方 DataSync （例如，**us-east-1**）。
6. 选择代理将用于连接的服务端点 AWS。
7. 保存 Activation key 输出的值。

激活您的代理

获得激活密钥后，即可完成 DataSync代理的创建。

激活您的代理

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧的导航窗格中，选择 代理，然后选择 创建代理。
3. 对于 虚拟机监控器，选择 Microsoft Hyper-V。
4. 对于 Endp oint 类型，请选择您在获取代理激活密钥时指定的相同类型的服务端点（例如，在中选择公共服务终端节点 ***Region name***）。
5. 将您的网络配置为与您的代理正在使用的服务端点类型配合使用。有关服务端点网络要求，请参阅以下主题：
   - VPC 端点
   - 公有端点
   - 美国联邦信息处理标准 (FIPS) 端点
6. 对于 激活密钥，执行以下操作：
   a. 选择 手动输入代理的激活密钥。
   b. 输入您从代理的本地控制台获得的激活密钥。
7. 选择 Create agent (创建代理)。

您的代理已准备好与您的 Azure Blob Storage 连接。有关更多信息，请参阅 创建您的 Azure Blob Storage 传输位置。

Amazon EC2 代理

您可以在 Amazon EC2 实例上部署您的 DataSync 代理。

创建 Amazon EC2 代理

1. 部署 Amazon EC2 代理。
2. 选择代理用来与 AWS通信的服务端点。

   在这种情况下，我们建议使用虚拟私有云（VPC）服务端点。
3. 将您的网络配置为使用 VPC 服务端点。
4. 激活代理。

创建您的 Azure Blob Storage 传输位置

您可以配置 DataSync 为使用您的Azure Blob Storage传输源或目的地。

开始前的准备工作

确保您知道 DataSync 访问层Azure Blob Storage和 blob 类型的访问方式和使用方式。您还需要一个可以连接到您的Azure Blob Storage容器的DataSync 代理。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，请选择 Microsoft Azure Blob Storage
4. 对于 容器 URL，输入涉及传输的容器的 URL。
5. 或者，对于 用作目标的访问层级，请选择要将对象或文件传输到的访问层级。
6. 对于 文件夹，如果要将传输限制到容器中的虚拟目录（例如 `/my/images`），请输入路径段。
7. 如果您的转移需要代理，请选择使用代理，然后选择可以与您的Azure Blob Storage容器连接的DataSync 代理。
8. 对于 SAS 令牌，请提供访问您的 DataSync blob 存储所需的凭据。Azure Blob 存储上的某些公共数据集不需要凭证。您可以直接输入 SAS 令牌，也可以指定包含该令牌的 AWS Secrets Manager 密钥。有关更多信息，请参阅为存储位置提供凭证。

SAS 令牌是 SAS URI 字符串的一部分，位于存储资源 URI 和问号（？）之后。令牌如下所示：

```
sp=r&st=2023-12-20T14:54:52Z&se=2023-12-20T22:54:52Z&spr=https&sv=2021-06-08&sr=c&sig=aBBKD
%2FXTI9E%2F%2Fmq171%2BZU178wcwqU%3D
```

9. （可选）在键和值字段中输入值以标记位置。

   标签可帮助您管理、筛选和搜索 AWS 资源。我们建议至少为您的位置创建一个名称标签。

10. 选择创建位置。

使用 AWS CLI

1. 复制以下 `create-location-azure-blob` 命令：

```
aws datasync create-location-azure-blob \
  --container-url "https://path/to/container" \
  --authentication-type "SAS" \
  --sas-configuration '{
      "Token": "your-sas-token"
    }' \
  --agent-arns my-datasync-agent-arn \
  --subdirectory "/path/to/my/data" \
  --access-tier "access-tier-for-destination" \
  --tags [{"Key": "key1","Value": "value1"}]
```

2. 对于 `--container-url` 参数，请指定传输中涉及的 Azure Blob Storage 容器的 URL。
3. 对于 `--authentication-type` 参数，请指定 SAS。如果访问不需要身份验证的公共数据集，请指定 NONE。
4. 对于`--sas-configuration`参数的Token选项，请指定允许 DataSync访问您的 blob 存储的 SAS 令牌。

   您还可以使用 AWS Secrets Manager提供其他参数来保护密钥。有关更多信息，请参阅为存储位置提供凭证。

   SAS 令牌是 SAS URI 字符串的一部分，位于存储资源 URI 和问号（？）之后。令牌如下所示：

```
sp=r&st=2023-12-20T14:54:52Z&se=2023-12-20T22:54:52Z&spr=https&sv=2021-06-08&sr=c&sig=aBBKD
%2FXTI9E%2F%2Fmq171%2BZU178wcwqU%3D
```

5. （可选）在--agent-arns参数中，指定可以连接到您的容器的 DataSync 代理的 Amazon 资源名称 (ARN)。

   下面是一个示例代理 ARN：arn:aws:datasync:*us-east-1*:*123456789012*:agent/agent-*01234567890aaabfb*

   您可以指定多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。

6. 对于 --subdirectory 参数，如果要将传输限制到容器中的虚拟目录（例如 /my/images），请指定路径段。

7. 或者，对于 --access-tier 参数，指定要将对象或文件传输到的访问层级（HOT、COOL 或 ARCHIVE）。

   此参数仅适用于使用该位置作为传输目标的情况。

8. 或者，对于 --tags 参数，指定可帮助您管理、筛选和搜索位置的键值对。

   我们建议为您的位置创建一个名称标签。

9. 运行 create-location-azure-blob 命令。

   如果命令成功，您将收到一条响应，显示您创建位置的 ARN。例如：

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:123456789012:location/
loc-12345678abcdefgh"
}
```

查看您的 Azure Blob Storage 传输位置

您可以获取有关您的现有 DataSync 转账地点的详细信息Azure Blob Storage。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择位置。
3. 选择您的 Azure Blob Storage 位置。

   您可以查看有关您所在位置的详细信息，包括使用该位置的所有 DataSync 转移任务。

使用 AWS CLI

1. 复制以下 describe-location-azure-blob 命令：

```
aws datasync describe-location-azure-blob \
  --location-arn "your-azure-blob-location-arn"
```

2. 对于 --location-arn 参数，指定您创建 Azure Blob Storage 位置的 ARN（例如 arn:aws:datasync:*us-east-1*:*123456789012*:location/loc-*12345678abcdefgh*）。
3. 运行 describe-location-azure-blob 命令。

   您会收到一条回复，显示有关您所在位置的详细信息。例如：

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:123456789012:location/
loc-12345678abcdefgh",
    "LocationUri": "azure-blob://my-user.blob.core.windows.net/container-1",
    "AuthenticationType": "SAS",
    "Subdirectory": "/my/images",
    "AgentArns": ["arn:aws:datasync:us-east-1:123456789012:agent/
agent-01234567890deadfb"],
}
```

更新您的 Azure Blob Storage 传输位置

如有必要，您可以在控制台或使用 AWS CLI修改位置的配置。

使用 AWS CLI

1. 复制以下 update-location-azure-blob 命令。

```
aws datasync update-location-azure-blob \
  --location-arn "your-azure-blob-location-arn" \
  --authentication-type "SAS" \
  --sas-configuration '{
      "Token": "your-sas-token"
    }' \
  --agent-arns my-datasync-agent-arn \
  --subdirectory "/path/to/my/data" \
```

```
  --access-tier "access-tier-for-destination"
```

2. 对于 `--location-arn` 参数，请指定要更新的 Azure Blob Storage 位置的 ARN（例如 `arn:aws:datasync:us-east-1:123456789012:location/loc-12345678abcdefgh`）。
3. 对于 `--authentication-type` 参数，请指定 SAS。
4. 对于`--sas-configuration`参数的Token选项，请指定允许 DataSync访问您的 blob 存储的 SAS 令牌。

   该令牌是 SAS URI 字符串的一部分，位于存储资源 URI 和问号 (?) 之后。令牌如下所示：

   ```
   sp=r&st=2022-12-20T14:54:52Z&se=2022-12-20T22:54:52Z&spr=https&sv=2021-06-08&sr=c&sig=qCBKD
   %2FXTI9E%2F%2Fmq171%2BZU178wcwqU%3D
   ```

5. 对于`--agent-arns`参数，请指定要连接到容器的 DataSync 代理的 Amazon 资源名称 (ARN)。

   下面是一个示例代理 ARN：`arn:aws:datasync:us-east-1:123456789012:agent/agent-01234567890aaabfb`

   您可以指定多个代理。有关更多信息，请参阅 使用多个 DataSync 代理。
6. 对于 `--subdirectory` 参数，如果要将传输限制到容器中的虚拟目录（例如 `/my/images`），请指定路径段。
7. 或者，对于 `--access-tier` 参数，指定要将对象传输到的访问层级（`HOT`、`COOL` 或 `ARCHIVE`）。

   此参数仅适用于使用该位置作为传输目标的情况。

后续步骤

为你的 DataSync 位置创建完毕后Azure Blob Storage，你可以继续设置接送服务。下面是要考虑的一些后续步骤：

1. 如果您还没有准备好，请创建另一个您计划将数据传入或传出 Azure Blob Storage 的位置。
2. 了解如何 DataSync 处理元数据和特殊文件，尤其是在您的传输地点没有类似的元数据结构的情况下。
3. 配置您数据的传输方式。例如，您可以只传输数据子集或删除 blob 存储中不在源位置的文件（只要您的 SAS 令牌具有删除权限）。
4. 开始传输。

使用 Microsoft Azure Files SMB 共享配置 AWS DataSync 传输

您可以配置 AWS DataSync 为向服务器消息块 (SMB) 共享传输数据或从Microsoft Azure Files服务器消息块 (SMB) 共享中传输数据。

> **Tip**
>
> 有关将数据从 Azure 文件 SMB 共享移动到的完整演练 AWS，请参阅AWS 存储博客。

提供对中小型企业股票的 DataSync 访问权限

DataSync 使用 SMB 协议连接到您的 SMB 共享，并使用您提供的凭据进行身份验证。

主题

- 支持的 SMB 协议版本
- 所需的权限

支持的 SMB 协议版本

默认情况下，根据与 SMB 文件服务器的协商 DataSync 自动选择 SMB 协议的版本。

您也可以配置 DataSync 为使用特定的 SMB 版本，但我们建议只有 DataSync 在无法自动与 SMB 文件服务器协商时才这样做。 DataSync 支持 SMB 版本 1.0 及更高版本。出于安全起见，建议使用 SMB 版本 3.0.2 或更高版本。早期版本（例如 SMB 1.0）包含已知的安全漏洞，攻击者可以借此来窃取您的数据。

有关 DataSync控制台和 API 中的选项列表，请参阅下表：

控制台选项	API 选项	说明
自动	`AUTOMATIC`	DataSync 和 SMB 文件服务器协商它们在 2.1 和 3.1.1 之间相互支持的最高版本的 SMB。 这是默认选项，也是推荐的选项。如果您改为选择文件服务器不支持的特定版本，则可能会出现 `Operation Not Supported` 错误。

控制台选项	API 选项	说明
SMB 3.0.2	SMB3	将协议协商限制为仅 SMB 版本 3.0.2。
SMB 2.1	SMB2	将协议协商限制为仅 SMB 版本 2.1。
SMB 2.0	SMB2_0	将协议协商限制为仅 SMB 版本 2.0。
SMB 1.0	SMB1	将协议协商限制为仅 SMB 版本 1.0。

所需的权限

DataSync 需要有权装载和访问您的 SMB 位置的用户。此用户可以是 Windows File Server 上的本地用户或 Microsoft Active Directory 中定义的域用户。

要设置对象所有权， DataSync 需要权限，该SE_RESTORE_NAME权限通常授予内置 Active Directory 组的成员 Backup Operators 和 Domain Admins。为用户提供此 DataSync 权限还有助于确保对文件、文件夹和文件元数据的足够权限，NTFS 系统访问控制列表除外 (SACLs)。

需要其他权限才能复制 SACLs。此操作特别需要 Windows SE_SECURITY_NAME权限，该权限授予域管理员组的成员。如果将任务配置为复制 SACLs，请确保用户具有所需的权限。要了解有关配置要复制的任务的更多信息 SACLs，请参阅配置文件、对象和元数据的处理方式。

在 SMB 文件服务器和 Amazon FSx for Windows 文件服务器文件系统之间复制数据时，源位置和目标位置必须属于相同的 Microsoft Active Directory 域，或者它们的域之间必须有 Active Directory 信任关系。

使用控制台创建 Azure 文件传输位置

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，选择 Server Message Block (SMB) (服务器消息块 (SMB))。

   您可在稍后将此位置配置为源或目标。
4. 对于代理，请选择一个或多个要连接到 SMB 共享的 DataSync 代理。

   如果您选择多个代理，请确保您了解如何对位置使用多个代理。
5. 对于 SMB 服务器，输入您的 DataSync 代理将挂载的 SMB 共享的域名系统 (DNS) 名称或 IP 地址。

> **Note**
>
> 您无法指定 IP 版本 6 (IPv6) 地址。

6. 在共享名称中，输入您的 SMB 共享导出的共享名称，该共享 DataSync 将在其中读取或写入数据。

   您可以在共享路径中包含子目录（例如，`/path/to/subdirectory`）。确保网络中的其他 SMB 客户端也可以挂载此路径。

   要复制子目录中的所有数据， DataSync 必须能够装载 SMB 共享并访问其所有数据。有关更多信息，请参阅 所需的权限。
7. （可选）展开 “其他设置”，然后选择 DataSync 要在访问您的 SMB 共享时使用的 SMB 版本。

   默认情况下，根据与 SMB 共享的协商 DataSync 自动选择版本。有关信息，请参阅支持的 SMB 版本。
8. 对于用户，输入可挂载 SMB 共享并有权访问传输中涉及文件和文件夹的用户名。

   有关更多信息，请参阅 所需的权限。
9. 对于密码，输入可挂载 SMB 共享并有权访问传输中涉及文件和文件夹的用户的密码。
10. （可选）对于域，输入 SMB 共享所属的 Windows 域名。

    如果您的环境中有多个域，则配置此设置 DataSync 可确保这些域连接到正确的共享。
11. （可选）选择 添加标签来标记您的位置。

    标签 是帮助您管理、筛选和搜索位置的键值对。我们建议至少为您的位置创建一个名称标签。
12. 选择创建位置。

配置使用其他云对象存储的 传输

借 AWS DataSync助，您可以在AWS 存储服务与以下云对象存储提供商之间传输数据：

- Wasabi Cloud Storage
- DigitalOcean Spaces
- Oracle Cloud Infrastructure Object Storage
- Cloudflare R2 Storage

- Backblaze B2 Cloud Storage
- NAVER Cloud Object Storage
- Alibaba Cloud Object Storage Service
- IBM Cloud Object Storage
- Seagate Lyve Cloud

只有在其他云中的存储系统与 Amazon EFS 或 Amazon 之间传输数据时 FSx，或者使用基本模式任务时，才需要 DataSync 代理。使用增强模式在其他云中的存储系统和 Amazon S3 之间传输数据时不需要代理。

无论您是否使用代理，都必须为云对象存储创建一个传输位置（特别是对象存储位置）。 DataSync 可以将此位置用作转账的来源或目的地。

提供对其他云对象存储的 DataSync 访问权限

如何 DataSync 访问您的云对象存储取决于多个因素，包括您的存储是否与 Amazon S3 API 兼容，以及访问存储 DataSync 所需的权限和证书。

主题

- Amazon S3 API 兼容性
- 存储权限和端点
- 存储凭证

Amazon S3 API 兼容性

您的云对象存储必须与以下 Amazon S3 API 操作兼容 DataSync 才能连接到它：

- `AbortMultipartUpload`
- `CompleteMultipartUpload`
- `CopyObject`
- `CreateMultipartUpload`
- `DeleteObject`
- `DeleteObjects`
- `DeleteObjectTagging`

- `GetBucketLocation`
- `GetObject`
- `GetObjectTagging`
- `HeadBucket`
- `HeadObject`
- `ListObjectsV2`
- `PutObject`
- `PutObjectTagging`
- `UploadPart`

存储权限和端点

您必须配置允许 DataSync 访问您的云对象存储的权限。如果您的对象存储是源位置，则 DataSync 需要对从中传输数据的存储桶具有读取和列出权限。如果您的对象存储是目标位置，则 DataSync 需要存储桶的读取、列出、写入和删除权限。

DataSync 还需要一个端点（或服务器）才能连接到您的存储。下表描述了 DataSync 可用于访问其他云对象存储的终端节点：

其他云提供商	端点
Wasabi Cloud Storage	`S3.region.wasabisys.com`
DigitalOcean Spaces	`region.digitaloceanspaces.com`
Oracle Cloud Infrastructure Object Storage	`namespace .compat.objectstorage. region.oraclecloud.com`
Cloudflare R2 Storage	`account-id .r2.cloudflarestorage.com`
Backblaze B2 Cloud Storage	`S3.region.backblazeb2.com`
NAVER Cloud Object Storage	`region.object.ncloudstorage.com` （大多数地区）
Alibaba Cloud Object Storage Service	`region.aliyuncs.com`

其他云提供商	端点
IBM Cloud Object Storage	`s3.region.cloud-object-storage.appdomain.cloud`
Seagate Lyve Cloud	`s3.region.lyvecloud.seagate.com`

> **Important**
>
> 有关如何配置存储桶权限的详细信息以及存储端点的更新信息，请参阅您的云提供商的文档。

存储凭证

DataSync 还需要凭证才能访问传输中涉及的对象存储桶。凭证可能是访问密钥和私有密钥或类似的东西，具体取决于您的云存储提供商如何引用这些凭证。

有关更多信息，请参阅您的云提供商的文档。

从其他云对象存储传输时的注意事项

在计划通过使用向其他云存储提供商传输对象或从其他云存储提供商传输对象时 DataSync，需要记住一些注意事项。

主题

- 成本
- 存储类
- 对象标签
- 传输到 Amazon S3

成本

与将数据移动入和移出其他云存储提供商相关的费用可能包括：

- 为您的 DataSync 代理运行 A mazon EC2 实例
- 使用传输数据 DataSync，包括与您的云对象存储和 Amazon S3（如果 S3 是您的传输目的地）相关的请求费用

- 将数据传入或传出您的云存储（请查看您的云提供商的定价）
- 将数据存储在支持的AWS 存储服务中 DataSync
- 将数据存储在其他云提供商中（请查看您的云提供商的定价）

存储类

一些云存储提供商的存储类别（类似于 Amazon S3），如果不先恢复，则 DataSync 无法读取。例如，Oracle Cloud Infrastructure Object Storage 具有档案存储类别。您需要先恢复该存储类别中的对象，然后 DataSync 才能传输它们。有关更多信息，请参阅您的云提供商的文档。

对象标签

并非所有云提供商都支持对象标签。即使支持对象标签，部分云提供商可能也不允许通过 Amazon S3 API 查询标签。无论哪种情况，如果您尝试复制对象标签，您的 DataSync 传输任务都可能失败。

您可以通过在创建、启动或更新任务时清除 DataSync 控制台中的 “复制对象标签” 复选框来避免这种情况。

传输到 Amazon S3

传输到 Amazon S3 时， DataSync 无法传输大于 5 TB 的对象。 DataSync也只能复制最多 2 KB 的对象元数据。

创建您的 DataSync 代理

只有在其他云中的存储系统与 Amazon EFS 或 Amazon 之间传输数据时 FSx，或者使用基本模式任务时，才需要 DataSync 代理。使用增强模式在其他云中的存储系统和 Amazon S3 之间传输数据时不需要代理。本部分介绍了如何在 AWS中的虚拟私有云（VPC）中的 Amazon EC2 实例上部署和激活代理。

创建 Amazon EC2 代理

1. 部署 Amazon EC2 代理。
2. 选择代理用来与之通信的服务端点 AWS。

   在这种情况下，我们建议使用 VPC 服务端点。
3. 将您的网络配置为使用 VPC 服务端点。
4. 激活代理。

为您的其他云对象存储创建传输位置

您可以配置 DataSync 为使用云对象存储作为源位置或目标位置。

开始前的准备工作

确保您知道如何 DataSync 访问您的云对象存储。您还需要一个可以连接到您的云对象存储的DataSync 代理。

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择 位置和 创建位置。
3. 对于 位置类型，选择 对象存储。
4. 对于 Server，输入 DataSync 可用于访问您的云对象存储的终端节点：
   - Wasabi Cloud Storage – `S3.region.wasabisys.com`
   - DigitalOcean Spaces – `region.digitaloceanspaces.com`
   - Oracle Cloud Infrastructure Object Storage – `namespace.compat.objectstorage.region.oraclecloud.com`
   - Cloudflare R2 Storage – `account-id.r2.cloudflarestorage.com`
   - Backblaze B2 Cloud Storage – `S3.region.backblazeb2.com`
   - NAVER Cloud Object Storage — `region.object.ncloudstorage.com`（大多数地区）
   - Alibaba Cloud Object Storage Service – `region.aliyuncs.com`
   - IBM Cloud Object Storage – `s3.region.cloud-object-storage.appdomain.cloud`
   - Seagate Lyve Cloud – `s3.region.lyvecloud.seagate.com`
5. 在存储桶名称中，输入您要向其传输数据或从中传输数据的对象存储桶的名称。
6. 对于“文件夹”，输入对象前缀。 DataSync 仅传输带有此前缀的对象。
7. 如果您的转移需要代理，请选择使用代理，然后选择可以与您的云对象存储连接的 DataSync 代理。
8. 展开其他设置。对于服务器协议，请选择 HTTPS。对于服务器端口，请选择 443。
9. 向下滚动到身份验证部分。确保选中“需要凭据”复选框，然后提供 DataSync 您的存储凭据。
   - 在访问密钥中，输入用于访问您的云对象存储的 ID。
   - 在私有密钥中，提供用于访问您的云对象存储的私有密钥。您可以直接输入密钥，也可以指定包含 AWS Secrets Manager 密钥的密钥。有关更多信息，请参阅为存储位置提供凭证。

10. （可选）在键和值字段中输入值以标记位置。

    标签可帮助您管理、筛选和搜索 AWS 资源。我们建议至少为您的位置创建一个名称标签。

11. 选择创建位置。

后续步骤

创建完云对象存储 DataSync 位置后，您可以继续设置传输。下面是要考虑的一些后续步骤：

1. 如果您还没有准备好，请创建另一个您计划将数据传入或传出 AWS的位置。
2. 了解如何 DataSync 处理对象存储位置的元数据和特殊文件。
3. 配置您数据的传输方式。例如，也许您只想传输数据的子集。

> **Important**
>
> 确保正确配置 DataSync 复制对象标签的方式。有关更多信息，请参阅有关对象标签的注意事项。

4. 开始传输。

创建一个传输数据的任务

任务描述了数据 AWS DataSync 传输的地点和方式。任务包括以下部分：

- 源位置-从中 DataSync 传输数据的存储系统或服务。
- 目标位置-将数据 DataSync 传输到的存储系统或服务。
- 任务选项 — 诸如要传输哪些文件、如何验证数据、任务何时运行等设置。
- 任务执行 — 当您运行任务时，它被称为任务执行。

创建 任务

创建 DataSync 任务时，需要指定源位置和目标位置。您还可以通过选择要传输的文件、处理元数据的方式，并设置时间表等，对任务进行自定义。

在创建任务之前，请务必了解 DataSync 转移的工作原理并查看任务配额。

> **Important**
>
> 如果您计划在 Amazon S3 地点之间传输数据，请在开始之前查看 DataSync 如何影响您的 S3 请求费用和DataSync 定价页面。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 确保您位于计划传输数据 AWS 区域 的地点之一。
3. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
4. 在配置源位置页面，创建或选择一个源位置，然后选择下一步。
5. 在配置目标位置页面，创建或选择一个目标位置，然后选择下一步。
6. （推荐）在配置设置页面，为任务指定一个您可以记住的名称。
7. 同样在配置设置页面，选择任务选项，或使用默认设置。

   您可能对以下一些选项感兴趣：

   - 指定要使用的任务模式。
   - 使用清单或筛选条件指定要传输的数据。
   - 配置如何处理文件元数据和验证数据完整性。
   - 通过任务报告或 Amazon 监控您的转账 CloudWatch。我们建议为任务设置某种监控。

   完成此操作后，选择下一步。
8. 检查任务配置，然后选择创建任务。

您已经准备好，可以开始任务。

使用 AWS CLI

创建 DataSync 源位置和目标位置后，即可创建任务。

1. 在你的 AWS CLI 设置中，确保你使用的是你计划传输数据的 AWS 区域 其中一个位置。
2. 复制以下 `create-task` 命令：

```
aws datasync create-task \
```

```
  --source-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-
id" \
  --destination-location-arn "arn:aws:datasync:us-east-1:account-
id:location/location-id" \
  --name "task-name"
```

3. 对于 `--source-location-arn`，指定源位置的 Amazon 资源名称（ARN）。
4. 对于 `--destination-location-arn`，指定目标位置的 ARN。

   如果您要跨 AWS 区域 账户转账，请确保 ARN 包含其他地区或账户 ID。
5. （推荐）对于 `--name`，请为任务指定一个您可以记住的名称。
6. 根据需要指定其他任务选项。您可能对以下一些选项感兴趣：

   - 使用清单或筛选条件指定要传输的数据。
   - 配置如何处理文件元数据和验证数据完整性。
   - 通过任务报告或 Amazon 监控您的转账 CloudWatch。我们建议为任务设置某种监控。

   有关更多选项，请参阅 create-task。以下是指定了多个选项的 `create-task` 命令示例：

```
aws datasync create-task \
  --source-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-
id" \
  --destination-location-arn "arn:aws:datasync:us-east-1:account-
id:location/location-id" \
  --cloud-watch-log-group-arn "arn:aws:logs:region:account-id" \
  --name "task-name" \
  --options
 VerifyMode=NONE,OverwriteMode=NEVER,Atime=BEST_EFFORT,Mtime=PRESERVE,Uid=INT_VALUE,Gid=INT
```

7. 运行 `create-task` 命令。

   如果命令成功，您将收到一条响应，显示所创建任务的 ARN。例如：

```
{
    "TaskArn": "arn:aws:datasync:us-east-1:111222333444:task/
task-08de6e6697796f026"
}
```

您已经准备好，可以开始任务。

任务状态

创建 DataSync 任务时，您可以检查其状态以查看其是否已准备好运行。

控制台状态	API 状态	说明
Available	AVAILABLE	任务准备就绪，可以开始传输数据。
运行	RUNNING	任务正在执行中。有关更多信息，请参阅 任务执行状态。
不可用	UNAVAILABLE	任务使用的 DataSync 代理处于脱机状态。有关更多信息，请参阅 如果我的代理处于离线状态，我该怎么办？。
已排队	QUEUED	另一个使用相同 DataSync 代理的任务正在执行中。有关更多信息，请参阅 了解任务何时排队。

使用多个任务对大型数据集进行分区

如果您要传输大型数据集，例如迁移数百万个文件或对象，我们建议您使用 DataSync 增强模式进行传输，该模式可以传输文件数量几乎不受限制的数据集。对于包含数十亿个文件的超大型数据集，您应该考虑使用多个 DataSync 任务对数据集进行分区。将数据分成多个任务（可能还有代理，具体取决于您所在的位置）有助于缩短 DataSync 准备和传输数据所需的时间。

考虑一下将大型数据集分成多个 DataSync任务的一些方法：

- 创建传输单独文件夹的任务。例如，可以在源存储中分别创建两个以 /FolderA 和 /FolderB 为目标的任务。
- 使用清单或筛选条件创建任务，传输文件、对象和文件夹的子集。

请注意，这种方法可能会增加存储 I/O 操作并影响网络带宽。有关更多信息，请参阅关于如何使用横向 DataSync 扩展架构加速数据传输的博客。

通过多个任务来分割传输的数据

如果要将不同的数据集传输到同一目标，您可以创建多个任务来帮助分割要传输的数据。

例如，如果要传输到名为 MyBucket 的同一个 S3 存储桶，您可以在存储桶中创建与每个任务相对应的不同前缀。这种方法可以防止数据集的文件名冲突，并允许您为每个前缀设置不同的权限。下面是设置方法：

1. 在目标 MyBucket 中创建三个前缀，分别命名为 task1、task2 和 task3：
   - s3://MyBucket/task1
   - s3://MyBucket/task2
   - s3://MyBucket/task3
2. 创建三个名为task1task2、的 DataSync 任务task3，并将它们转移到中相应的前缀MyBucket。

为数据传输选择任务模式

您的 AWS DataSync 任务可以在以下模式之一下运行：

- 增强模式-传输几乎无限数量的文件或对象，其性能高于基本模式。增强模式任务通过并行列示、准备、传输和验证数据来优化数据传输过程。增强模式目前可用于在 Amazon S3 位置之间进行传输，在没有代理的情况下在 Amazon S3 之间Azure Blob进行传输，在没有代理的情况下在其他云和 Amazon S3 之间传输，以及使用增强模式代理在 NFS 或 SMB 文件服务器与 Amazon S3 之间传输。
- 基本模式-在 AWS 存储和所有其他支持的 DataSync 位置之间传输文件或对象。基本模式任务在数据集中文件、对象和目录数量方面存在配额限制。基本模式按顺序准备、传输和验证数据，因此对于大多数工作负载来说，基本模式的处理速度慢于增强模式。

了解任务模式的差异

以下信息有助于您确定使用哪种任务模式。

能力	增强模式行为	基本模式行为
性能	DataSync 并行列出、准备、传输和验证您的数据。对于大多数工作负载（例如传输大型对象），可提供比基本模式更高的性能	DataSync 按顺序准备、传输和验证您的数据。对于大多数工作负载，性能比增强模式更慢

能力	增强模式行为	基本模式行为
数据集中每次执行任务时 DataSync 可以处理的项目数	对象数量几乎不受限制	适用配额
数据传输计数器和指标	与“基本”模式相比，计数器和指标更多，例如在源位置 DataSync找到的对象数量、每次任务执行期间准备的对象数量，以及类似于文件和对象计数器的文件夹计数器	计数器和指标低于增强模式
日志记录	结构化日志（JSON 格式）	非结构化日志
支持的位置	目前用于在 Amazon S3 位置之间进行传输，在没有代理的情况下在 Amazon S3 Azure Blob 和 Amazon S3 之间传输，在没有代理的情况下在其他云和 Amazon S3 之间传输，以及使用增强模式代理在 NFS 或 SMB 文件服务器与 Amazon S3 之间传输。	适用于所有 DataSync 支持的地点之间的传输
数据验证选项	DataSync 仅验证传输的数据	DataSync 默认情况下会验证所有数据
成本	如需了解更多信息，请参阅定DataSync 价页面	如需了解更多信息，请参阅定DataSync 价页面
处理不支持的对象标签时失败	云存储传入或传出不支持对象标记的位置时，如果 `ObjectTags` 选项未指定或设置为 `PRESERVE`，任务执行会立即失败。	云存储传入或传出不支持对象标记的位置时，任务执行会正常运行，但如果 `ObjectTags` 选项未指定或设置为 `PRESERVE`，则会逐对象报告已标记对象的故障。

选择任务模式

您只能为 Amazon S3 位置之间的传输、在没有代理的情况下在 Amazon S3 之间Azure Blob进行传输、在没有代理的情况下在其他云和 Amazon S3 之间传输以及使用增强模式代理在 NFS 或 SMB 文件服务器与 Amazon S3 之间的传输选择增强模式。否则，必须使用基本模式。例如，从本地 HDFS 位置传输到 S3 位置需要基本模式。

根据所选择的任务模式，任务选项和性能可能有所不同。一旦创建任务后，就无法更改任务模式。

所需的权限

要创建增强模式任务，与之 DataSync 配合使用的 IAM 角色必须具有`iam:CreateServiceLinkedRole`权限。

要获得 DataSync 用户权限，请考虑使用AWSDataSyncFullAccess。这是一项 AWS 托管策略，可为用户提供对其依赖项的完全访问权限 DataSync 和最少访问权限。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 对于任务模式，请选择下列选项之一：

   - 增强型
   - 基本

   有关更多信息，请参阅 了解任务模式的差异。
5. 同样在配置设置页面，选择其他任务选项，或使用默认设置。

   您可能对以下一些选项感兴趣：

   - 使用清单或筛选条件指定要传输的数据。
   - 配置如何处理文件元数据和验证数据完整性。
   - 使用任务报告或 Amazon CloudWatch 日志监控您的转账。

完成此操作后，选择下一步。

6. 检查任务配置，然后选择创建任务。

使用 AWS CLI

1. 在你的 AWS CLI 设置中，确保你使用的是你计划传输数据的 AWS 区域 位置之一。
2. 复制以下 `create-task` 命令：

```
aws datasync create-task \
  --source-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-id" \
  --destination-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-id" \
  --task-mode "ENHANCED-or-BASIC"
```

3. 对于 `--source-location-arn`，指定源位置的 Amazon 资源名称（ARN）。
4. 对于 `--destination-location-arn`，指定目标位置的 ARN。

   如果您要跨 AWS 区域 账户转账，请确保 ARN 包含其他地区或账户 ID。
5. 对于 `--task-mode`，指定 `ENHANCED` 或 `BASIC`。

   有关更多信息，请参阅 了解任务模式的差异。
6. 根据需要指定其他任务选项。您可能对以下一些选项感兴趣：

   - 使用清单或筛选条件指定要传输的数据。
   - 配置如何处理文件元数据和验证数据完整性。
   - 使用任务报告或 Amazon CloudWatch 日志监控您的转账。

   有关更多选项，请参阅 create-task。以下是指定了增强模式和其他多个选项的 `create-task` 命令示例：

```
aws datasync create-task \
  --source-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-id" \
  --destination-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-id" \
  --name "task-name" \
```

```
  --task-mode "ENHANCED" \
  --options
 TransferMode=CHANGED,VerifyMode=ONLY_FILES_TRANSFERRED,ObjectTags=PRESERVE,LogLevel=TRANSF
```

7. 运行 `create-task` 命令。

   如果命令成功，您将收到一条响应，显示所创建任务的 ARN。例如：

```
{
    "TaskArn": "arn:aws:datasync:us-east-1:111222333444:task/
task-08de6e6697796f026"
}
```

使用 DataSync API

您可以通过在CreateTask操作中配置`TaskMode`参数来指定 DataSync 任务模式。

选择什么 AWS DataSync 转账

AWS DataSync 允许您选择要传输的内容以及如何处理数据。一些选项包括：

- 使用清单传输确切的文件或对象列表。
- 使用筛选条件在传输中包含或排除某些类型的数据。
- 对于周期性传输，仅移动自上次传输以来发生变化的数据
- 覆盖目标位置的数据，使其与源位置中的数据匹配。
- 选择要在存储位置之间保留的文件或对象元数据。

主题

- 使用清单传输特定的文件或对象
- 使用筛选条件传输特定文件、对象和文件夹
- 了解如何 DataSync 处理文件和对象元数据
- 由复制的链接和目录 AWS DataSync
- 配置文件、对象和元数据的处理方式

使用清单传输特定的文件或对象

清单是您要传输的文件或对象 AWS DataSync 的列表。例如，不必传输可能包含数百万个对象的 S3 存储桶中的所有内容，而是只 DataSync 传输您在清单中列出的对象。

清单与筛选条件类似，但清单可让您准确确定要传输的文件或对象，而不是符合筛选模式的数据。

> **Note**
>
> 增强模式任务清单文件允许的最大大小为 20 GB。

创建清单

清单是逗号分隔值 (CSV) 格式的文件，它列出了源位置中要 DataSync 传输的文件或对象。如果源位置是 S3 存储桶，您还可以包含要传输的对象版本。

主题

- 指南要求
- 示例清单

指南要求

使用这些指南来帮助您创建适用于的清单 DataSync。

Do

- 指定要传输的每个文件或对象的完整路径。

  您不能只指定一个目录或文件夹，然后希望传输其中的所有内容。对于这种情况，请考虑使用包含筛选条件，而不是清单。
- 确保每个文件或对象路径都与您在配置 DataSync 源位置时指定的挂载路径、文件夹、目录或前缀相关。

  例如，假设您配置了一个 S3 位置，其前缀名为 `photos`。该前缀包括一个要传输的对象 `my-picture.png`。然后，在清单中，您只需要指定对象（`my-picture.png`），而不是前缀和对象（`photos/my-picture.png`）。
- 要指定 Amazon S3 对象版本 IDs，请使用逗号分隔对象的路径和版本 ID。

下面的示例展示了包含两个字段的清单条目。第一个字段包含一个名为 picture1.png 的对象。第二个字段用逗号分隔，包含版本 ID 111111：

```
picture1.png,111111
```

- 在以下情况下使用引号：
  - 当路径包含特殊字符（逗号、引号和行尾）时：

    ```
    "filename,with,commas.txt"
    ```

  - 当路径跨越多行时：

    ```
    "this
    is
    a
    filename.txt"
    ```

  - 当路径包含引号时：

    ```
    filename""with""quotes.txt
    ```

    这表示名为 filename"with"quotes.txt 的路径。

  这些引号规则也适用于版本 ID 字段。通常，如果清单字段有引号，必须使用另一个引号将其转义。
- 用新行分隔每个文件或对象条目。

  您可以使用 Linux 风格的换行（换行符或回车符）或 Windows 风格的换行（回车符后跟换行符）来分隔行。
- 保存您的清单（例如，my-manifest.csv 或 my-manifest.txt）。
- 将清单上传到DataSync 可以访问的 S3 存储桶。

  此存储桶不必位于您正在使用的同一个 AWS 区域 或账户中 DataSync。

Don't

- 只指定一个目录或文件夹，然后希望传输其中的所有内容。

  清单只能包含要传输的文件或对象的完整路径。如果您将源位置配置为使用特定的挂载路径、文件夹、目录或前缀，您不必在清单中包含这些内容。

- 指定超过 4096 个字符的文件或对象路径。
- 指定超过 1024 字节的文件路径、对象路径或 Amazon S3 对象版本 ID。
- 指定重复的文件或对象路径。
- 包含对象版本 ID（如果您的源位置不是 S3 存储桶）。
- 在一个清单条目中包含两个以上的字段。

  一个条目只能包含一个文件或对象路径，以及一个 Amazon S3 对象版本 ID（如适用）。
- 包含不符合 UTF-8 编码的字符。
- 在条目字段中，在引号外包含非预期的空格。

示例清单

使用这些示例来帮助您创建适用于的清单 DataSync。

包含完整文件或对象路径的清单

下面的示例展示了带有完整文件或对象传输路径的清单。

```
photos/picture1.png
photos/picture2.png
photos/picture3.png
```

仅包含对象键的清单

下面的示例展示了一份清单，其中列出了要从 Amazon S3 源位置传输的对象。由于位置已配置了前缀 `photos`，因此仅指定了对象键。

```
picture1.png
picture2.png
picture3.png
```

包含对象路径和版本的清单 IDs

以下清单示例中的前两个条目包括要传输的特定 Amazon S3 对象版本。

```
photos/picture1.png,111111
photos/picture2.png,121212
photos/picture3.png
```

包含 UTF-8 字符的清单

以下示例展示了包含 UTF-8 字符的文件清单。

```
documents/résumé1.pdf
documents/résumé2.pdf
documents/résumé3.pdf
```

提供对清单的 DataSync 访问权限

您需要一个 AWS Identity and Access Management (IAM) 角色来授予对其 S3 存储桶中清单的 DataSync 访问权限。此角色必须包括以下权限：

- `s3:GetObject`
- `s3:GetObjectVersion`

您可以在 DataSync 控制台中自动生成此角色，也可以自己创建该角色。

> **Note**
>
> 如果您的清单位于不同的清单中 AWS 账户，则必须手动创建此角色。

自动创建 IAM 角色

在控制台中创建或启动转移任务时， DataSync 可以为您创建一个 IAM 角色，该角色具有访问清单所需的`s3:GetObject`和`s3:GetObjectVersion`权限。

自动创建角色所需的权限

要自动创建角色，请确保您用于访问 DataSync 控制台的角色具有以下权限：

- `iam:CreateRole`
- `iam:CreatePolicy`
- `iam:AttachRolePolicy`

创建 IAM 角色（同一账户）

您可以手动创建访问清单 DataSync 所需的 IAM 角色。以下说明假设您 AWS 账户 所在的使用位置相同， DataSync 并且清单的 S3 存储桶位于同一位置。

1. 使用 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。
2. 在左侧导航窗格的访问管理下，选择角色，然后选择创建角色。
3. 在选择可信实体页面中，为可信实体类型选择 AWS 服务。
4. 对于 “用例”，DataSync在下拉列表中进行选择，然后选择DataSync。选择下一步。
5. 在添加权限页面上，选择下一步。输入角色名称，然后选择创建角色。
6. 在角色页面上，搜索您刚刚创建的角色并选择其名称。
7. 在角色的详情页面上，选择权限选项卡。选择添加权限，然后选择创建内联策略。
8. 选择 JSON 选项卡，然后将以下示例策略粘贴到策略编辑器中：

   JSON

   ```
   {
       "Version":"2012-10-17",
       "Statement": [{
           "Sid": "DataSyncAccessManifest",
           "Effect": "Allow",
           "Action": [
               "s3:GetObject",
               "s3:GetObjectVersion"
           ],
           "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/my-manifest.csv"
       }]
   }
   ```

9. 在刚才粘贴的示例策略中，将以下值替换为您自己的值：

   a. 将 *amzn-s3-demo-bucket* 替换为托管清单的 S3 存储桶名称。

   b. 将 *my-manifest.csv* 替换为清单的文件名称。

10. 选择下一步。为您的策略输入名称，然后选择创建策略。
11. （推荐）为防止出现跨服务混淆代理问题，请执行以下操作：

    a. 在角色的详情页面上，选择信任关系选项卡。选择编辑信任策略。

b. 使用以下示例更新信任策略，其中包括 `aws:SourceArn` 和 `aws:SourceAccount` 全局条件上下文键：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Principal": {
            "Service": "datasync.amazonaws.com"
        },
        "Action": "sts:AssumeRole",
        "Condition": {
            "StringEquals": {
            "aws:SourceAccount": "555555555555"
            },
            "ArnLike": {
            "aws:SourceArn": "arn:aws:datasync:us-east-1:555555555555:*"
            }
        }
      }
  ]
}
```

- 将每个实例*account-id*替换为您正在使用的 AWS 账户 ID DataSync。
- *region*替换为你正在使用 AWS 区域 的地方 DataSync。

c. 选择更新策略。

您已经创建了一个允许 DataSync 访问清单的 IAM 角色。在创建或启动任务时指定此角色。

创建 IAM 角色（不同账户）

如果您的清单位于属于不同的 S3 存储桶中 AWS 账户，则必须手动创建 DataSync 用于访问清单的 IAM 角色。然后，在清单 AWS 账户 所在的位置中，您需要在 S3 存储桶策略中包含该角色。

创建角色

1. 使用 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。
2. 在左侧导航窗格的访问管理下，选择角色，然后选择创建角色。
3. 在选择可信实体页面中，为可信实体类型选择 AWS 服务。
4. 对于“用例”，DataSync在下拉列表中进行选择，然后选择DataSync。选择下一步。
5. 在添加权限页面上，选择下一步。输入角色名称，然后选择创建角色。
6. 在角色页面上，搜索您刚刚创建的角色并选择其名称。
7. 在角色的详情页面上，选择权限选项卡。选择添加权限，然后选择创建内联策略。
8. 选择 JSON 选项卡，然后将以下示例策略粘贴到策略编辑器中：

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [{
        "Sid": "DataSyncAccessManifest",
        "Effect": "Allow",
        "Action": [
            "s3:GetObject",
            "s3:GetObjectVersion"
        ],
        "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/my-manifest.csv"
    }]
}
```

9. 在刚才粘贴的示例策略中，将以下值替换为您自己的值：
   a. 将 *amzn-s3-demo-bucket* 替换为托管清单的 S3 存储桶名称。
   b. 将 *my-manifest.csv* 替换为清单的文件名称。
10. 选择下一步。为您的策略输入名称，然后选择创建策略。
11. （推荐）为防止出现跨服务混淆代理问题，请执行以下操作：
    a. 在角色的详情页面上，选择信任关系选项卡。选择编辑信任策略。
    b. 使用以下示例更新信任策略，其中包括 `aws:SourceArn` 和 `aws:SourceAccount` 全局条件上下文键：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Principal": {
            "Service": "datasync.amazonaws.com"
        },
        "Action": "sts:AssumeRole",
        "Condition": {
            "StringEquals": {
            "aws:SourceAccount": "000000000000"
            },
            "ArnLike": {
            "aws:SourceArn": "arn:aws:datasync:us-east-1:000000000000:*"
            }
        }
     }
  ]
}
```

- 将的*account-id*每个实例替换为您正在使用的 AWS 账户 ID DataSync。
- *region*替换为你正在使用 AWS 区域 的地方 DataSync。

c. 选择更新策略。

您创建了可以包含在 S3 存储桶策略中的 IAM 角色。

更新 S3 存储桶策略以包含角色

创建 IAM 角色后，必须将其添加到清单 AWS 账户 所在的另一个存储桶策略中的 S3 存储桶策略中。

1. 在中 AWS 管理控制台，切换到包含你的 manfiest 的 S3 存储桶的账户。
2. 打开 Amazon S3 控制台，网址为 https://console.aws.amazon.com/s3/。
3. 在存储桶的详情页面上，选择权限选项卡。
4. 在 存储桶策略下，选择 编辑并执行以下操作来修改您的 S3 存储桶策略：

a. 更新编辑器中的内容，以包含以下策略声明：

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "DataSyncAccessManifestBucket",
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:GetObjectVersion"
      ],
      "Resource": "arn:aws:s3:::amzn-s3-demo-bucket"
    }
  ]
}
```

b. *account-id*替换为您正在使用的账户 DataSync 的 AWS 账户 ID。

c. *datasync-role*替换为您刚刚创建的 DataSync 允许访问清单的 IAM 角色。

d. 将 *amzn-s3-demo-bucket* 替换为在其他 AWS 账户中托管清单的 S3 存储桶名称。

5. 选择保存更改。

您已经创建了一个 IAM 角色， DataSync 允许您在其他账户中访问您的清单。在创建或启动任务时指定此角色。

在创建任务时指定清单

您可以指定 DataSync 要在创建任务时使用的清单。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 对于要扫描的内容，请选择特定文件、对象和文件夹，然后选择使用清单。

5. 对于 S3 URI，请选择托管在 S3 存储桶上的清单。

   或者，您也可以输入 URI（例如，`s3://bucket/prefix/my-manifest.csv`）。
6. 对于对象版本，请选择 DataSync 要使用的清单版本。

   默认情况下， DataSync 使用对象的最新版本。
7. 对于清单访问角色，执行下列操作之一：

   - 选择 “自动生成” DataSync 以自动创建一个 IAM 角色，该角色具有访问其 S3 存储桶中的清单所需的权限。
   - 选择可以访问清单的现有 IAM 角色。

   有关更多信息，请参阅 提供对清单的 DataSync 访问权限。
8. 配置所需的任何其他任务设置，然后选择下一步。
9. 选择创建任务。

使用 AWS CLI

1. 复制以下 `create-task` 命令：

```
aws datasync create-task \
  --source-location-arn arn:aws:datasync:us-east-1:123456789012:location/
loc-12345678abcdefgh \
  --destination-location-arn arn:aws:datasync:us-east-1:123456789012:location/loc-
abcdefgh12345678 \
  --manifest-config {
      "Source": {
        "S3": {
            "ManifestObjectPath": "s3-object-key-of-manifest",
            "BucketAccessRoleArn": "bucket-iam-role",
            "S3BucketArn": "amzn-s3-demo-bucket-arn",
            "ManifestObjectVersionId": "manifest-version-to-use"
        }
      }
  }
```

2. 对于 `--source-location-arn` 参数，指定数据传输源位置的 Amazon 资源名称（ARN）。
3. 对于 `--destination-location-arn` 参数，指定数据传输目标位置的 ARN。
4. 对于 `--manifest-config` 参数，请执行以下操作：

- `ManifestObjectPath`：指定清单的 S3 对象键。
- `BucketAccessRoleArn`— 指定允许 DataSync 访问其 S3 存储桶中的清单的 IAM 角色。

  有关更多信息，请参阅 提供对清单的 DataSync 访问权限。

- `S3BucketArn`：指定托管清单的 S3 存储桶的 ARN。
- `ManifestObjectVersionId`— 指定 DataSync 要使用的清单版本。

  默认情况下， DataSync 使用对象的最新版本。

5. 运行 `create-task` 命令来创建您的任务。

准备就绪后，您可以启动传输任务。

在启动任务时指定清单

您可以指定 DataSync 要在执行任务时使用的清单。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，选择任务，然后选择要启动的任务。
3. 在任务概览页面，选择开始，然后选择使用覆盖选项开始。
4. 对于要扫描的内容，请选择特定文件、对象和文件夹，然后选择使用清单。
5. 对于 S3 URI，请选择托管在 S3 存储桶上的清单。

   或者，您也可以输入 URI（例如，`s3://bucket/prefix/my-manifest.csv`）。
6. 对于对象版本，请选择 DataSync 要使用的清单版本。

   默认情况下， DataSync 使用对象的最新版本。
7. 对于清单访问角色，执行下列操作之一：

   - 选择 “自动生成” DataSync 以自动创建 IAM 角色来访问其 S3 存储桶中的清单。
   - 选择可以访问清单的现有 IAM 角色。

   有关更多信息，请参阅 提供对清单的 DataSync 访问权限。
8. 选择开始，以开始传输。

使用 AWS CLI

1. 复制以下 start-task-execution 命令：

```
aws datasync start-task-execution \
  --task-arn arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh \
  --manifest-config {
      "Source": {
        "S3": {
            "ManifestObjectPath": "s3-object-key-of-manifest",
            "BucketAccessRoleArn": "bucket-iam-role",
            "S3BucketArn": "amzn-s3-demo-bucket-arn",
            "ManifestObjectVersionId": "manifest-version-to-use"
        }
      }
  }
```

2. 对于 --task-arn 参数，请指定要启动的任务的 Amazon 资源名称（ARN）。
3. 对于 --manifest-config 参数，请执行以下操作：

   - ManifestObjectPath：指定清单的 S3 对象键。
   - BucketAccessRoleArn— 指定允许 DataSync 访问其 S3 存储桶中的清单的 IAM 角色。

     有关更多信息，请参阅 提供对清单的 DataSync 访问权限。

   - S3BucketArn：指定托管清单的 S3 存储桶的 ARN。
   - ManifestObjectVersionId— 指定 DataSync 要使用的清单版本。

     默认情况下， DataSync 使用对象的最新版本。
4. 运行 start-task-execution 命令开始传输。

限制

- 您不能将清单与筛选条件一起使用。
- 您不能只指定一个目录或文件夹，然后希望传输其中的所有内容。对于这种情况，请考虑使用包含筛选条件，而不是清单。
- 您不能使用 “保留已删除文件” 任务选项（PreserveDeletedFiles在 API 中）来维护目标中不在源中的文件或对象。 DataSync 仅传输清单中列出的内容，不会删除目的地中的任何内容。

问题排查

与 **HeadObject** 或 **GetObjectTagging** 相关的错误

如果您要 IDs 从 S3 存储桶传输具有特定版本的对象，则可能会看到与HeadObject或相关的错误GetObjectTagging。例如，以下是与 GetObjectTagging 相关的错误：

```
[WARN] Failed to read metadata for file /picture1.png (versionId: 111111): S3 Get
 Object Tagging Failed
[ERROR] S3 Exception: op=GetObjectTagging photos/picture1.png, code=403, type=15,
 exception=AccessDenied,
msg=Access Denied req-hdrs: content-type=application/xml, x-amz-api-version=2006-03-01
 rsp-hdrs: content-type=application/xml,
date=Wed, 07 Feb 2024 20:16:14 GMT, server=AmazonS3, transfer-encoding=chunked,
x-amz-id-2=IOWQ4fDEXAMPLEQM+ey7N9WgVhSnQ6JEXAMPLEZb7hSQDASK+Jd1vEXAMPLEa3Km, x-amz-
request-id=79104EXAMPLEB723
```

如果您看到其中任何一个错误，请验证用于访问您的 DataSync S3 源位置的 IAM 角色是否具有以下权限：

- s3:GetObjectVersion
- s3:GetObjectVersionTagging

如果您需要更新角色，使其具有这些权限，请参阅 为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置。

错误：**ManifestFileDoesNotExist**

此错误表示在源中找不到清单中的文件。查看创建清单的指南。

后续步骤

如果尚未开始，请启动任务。否则，请监视任务的活动。

使用筛选条件传输特定文件、对象和文件夹

AWS DataSync 允许您应用筛选器在传输中包含或排除来自源位置的数据。例如，如果您不想传输以 .tmp 结尾的临时文件，可以创建一个排除筛选条件，这些文件就不会传输到目标位置。

您可以在同一个传输任务中组合使用排除和包含筛选条件。如果您修改了任务的筛选条件，这些更改将在下次运行该任务时生效。

筛选术语、定义和语法

熟悉与筛选相关的概念： DataSync

筛选条件

构成特定筛选条件的整个字符串（例如，`*.tmp|*.temp` 或 `/folderA|/folderB`）。

筛选条件由使用管道符（|）分隔的模式组成。在 DataSync 控制台中添加模式时不需要分隔符，因为每个模式都是单独添加的。

> **Note**
>
> 筛选区分大小写。例如，筛选条件 `/folderA` 与 `/FolderA` 不匹配。

模式

筛选条件中的模式。例如，`*.tmp` 是一个模式，它是 `*.tmp|*.temp` 筛选条件的组成部分。如果您的筛选条件有多个模式，则使用管道符（|）划分各个模式。

文件夹

- 所有筛选条件都相对于源位置路径。例如，假设您在创建源位置和任务时指定 `/my_source/` 作为源路径，并要求包含筛选条件 `/transfer_this/`。在这种情况下，仅 DataSync 传输目录 `/my_source/transfer_this/`及其内容。
- 要指定直接位于源位置下的文件夹，请在文件夹名称前面加入一个正斜杠 (/)。在上述示例中，模式使用 `/transfer_this`，而不是 `transfer_this`。
- DataSync 以相同的方式解释以下模式，并匹配文件夹及其内容。

  ```
  /dir
  ```

  ```
  /dir/
  ```

- 当您从 Amazon S3 存储桶传输数据或将数据传输到 Amazon S3 存储桶时，会将对象密钥中的/字符 DataSync视为文件系统上的文件夹。

特殊字符

以下是用于筛选的特殊字符。

特殊字符	说明
*（通配符）	一个用于匹配零个或多个字符的字符。例如，/movies_folder* 匹配 /movies_folder 和 /movies_folder1 这两者。
|（竖线分隔符）	用作模式之间分隔符的字符。它支持指定多个模式，其中的任何模式都可以与筛选条件匹配。例如，*.tmp|*.temp 匹配以 tmp 或 temp 结尾的文件。 **Note** 在控制台上添加模式时不需要此分隔符，因为每个模式都在单独的行上添加。
\（反斜杠）	用于转义文件或对象名称中的特殊字符（*、|、\）的字符。 当反斜杠是文件名的组成部分时，需要使用双反斜杠 (\\)。同样，\\\\ 代表文件名中两个连续的反斜杠。 当竖线是文件名的组成部分时，需要使用反斜杠后跟竖线 (|) 字符。 后跟任何其他字符或位于模式末尾的反斜杠 (\) 将被忽略。

示例筛选条件：

以下示例显示了您可以使用的常用过滤器 DataSync。

Note

筛选条件的字符数是有限制的。有关更多信息，请参阅 DataSync 配额。

排除源位置中的某些文件夹

在某些情况下，您可能需要排除源位置中的文件夹，以防止将它们复制到目标位置。例如，如果您有临时 work-in-progress文件夹，则可以使用类似以下筛选器的内容：

*/.temp

要排除内容相似的文件夹（例如 /reports2021 和 /reports2022），您可以使用与下述类似的排除筛选条件：

/reports*

要排除文件层次结构中任何级别的文件夹，您可以使用排除筛选条件，如下所示。

/folder-to-exclude-1|/folder-to-exclude-2

要排除源位置的顶级文件夹，您可以使用排除筛选条件，如下所示。

/top-level-folder-to-exclude-1|/top-level-folder-to-exclude-2

包含源位置中的文件夹子集

在某些情况下，您的源位置可能是大型共享的，并且您需要传输根文件夹下的文件夹子集。要包含特定文件夹，使用如下所示的包含筛选条件开始任务执行。

/folder-to-transfer/*

排除特定文件类型

要从传输中排除某些文件类型，您可以使用诸如 *.temp 的排除筛选条件创建一个任务执行。

传输您指定的个别文件

要传输个别文件的列表，使用如下所示的包含筛选条件开始任务执行：“/folder/subfolder/file1.txt|/folder/subfolder/file2.txt|/folder/subfolder/file2.txt”

创建包含筛选条件

包含筛选器定义要 DataSync 传输的文件、对象和文件夹。您可以在创建、编辑或启动任务时配置以包括这些筛选条件。

DataSync 仅扫描和传输与包含过滤器匹配的文件和文件夹。例如，要包括源文件夹的子集，您可以指定 /important_folder_1|/important_folder_2。

> **Note**
>
> 包括筛选条件仅支持将通配符 (*) 字符用作模式中最右侧的字符。例如，支持 /documents*|/code*，但不支持 *.txt。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 对于要扫描的内容，请选择特定文件、对象和文件夹，然后选择使用筛选条件。
5. 对于包含，输入您的筛选条件（例如，输入 `/important_folders` 以包含重要目录），然后选择添加模式。
6. 根据需要，添加其他包含筛选条件。

使用 AWS CLI

使用时 AWS CLI，必须在筛选器周围使用单引号 (')，如果有多个筛选器，则必须使用|（竖线）作为分隔符。

以下示例指定了运行 `create-task` 命令时使用的两个包含筛选条件 `/important_folder1` 和 `/important_folder2`。

```
aws datasync create-task
   --source-location-arn 'arn:aws:datasync:region:account-id:location/location-id' \
   --destination-location-arn 'arn:aws:datasync:region:account-id:location/location-id'
 \
   --includes FilterType=SIMPLE_PATTERN,Value='/important_folder1|/important_folder2'
```

创建排除筛选条件

排除过滤器定义源位置中您不想 DataSync 传输的文件、对象和文件夹。您可以在创建、编辑或启动任务时配置这些筛选条件。

主题

- 默认排除的数据

默认排除的数据

DataSync 自动将某些数据排除在传输范围之外：

- `.snapshot`— DataSync 忽略任何以结尾的路径`.snapshot`，该路径通常用于存储系统的文件或目录的 point-in-time快照。
- `/.aws-datasync`和 `/.awssync` — 在您所在的位置 DataSync创建这些文件夹，以方便您进行传输。
- `/.zfs`— 你可能会在 Amazon FSx 上看到这个文件夹，里面有 OpenZFS 位置。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 对于排除，输入您的筛选条件（例如，输入 `*/temp` 以排除临时文件夹），然后选择添加模式。
5. 根据需要，添加其他排除筛选条件。
6. 如果需要，添加包含筛选条件。

使用 AWS CLI

使用时 AWS CLI，必须在筛选器周围使用单引号 (')，如果有多个筛选器，则必须使用| （竖线）作为分隔符。

以下示例指定了运行 `create-task` 命令时使用的两个排除筛选条件 `*/temp` 和 `*/tmp`。

```
aws datasync create-task \
   --source-location-arn 'arn:aws:datasync:region:account-id:location/location-id' \
   --destination-location-arn 'arn:aws:datasync:region:account-id:location/location-id'
 \
   --excludes FilterType=SIMPLE_PATTERN,Value='*/temp|*/tmp'
```

了解如何 DataSync 处理文件和对象元数据

AWS DataSync 可以在数据传输期间保留您的文件或对象元数据。如何复制元数据取决于您的传输位置以及这些位置是否使用类似类型的元数据。

系统级元数据

通常，DataSync 不会复制系统级元数据。例如，从 SMB 文件服务器传输时，您在文件系统级别配置的权限不会复制到目标存储系统。

有一些例外。在 Amazon S3 和其他对象存储之间传输时，DataSync 会复制一些系统定义的对象元数据。

在 Amazon S3 传输中复制的元数据

下表描述了当传输涉及 Amazon S3 位置时 DataSync 可以复制哪些元数据。

主题

- 到 Amazon S3
- 在 Amazon S3 和其他对象存储之间
- 在 Amazon S3 和 HDFS 之间

到 Amazon S3

从其中一个位置复制时	到此位置	DataSync 可以复制
• NFS • Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS）	• Amazon S3	以下是 Amazon S3 用户元数据： • 文件和文件夹修改时间戳 • 文件和文件夹访问时间戳（DataSync 只能在尽力而为的基础上执行此操作） • 用户 ID 和组 ID • POSIX 权限 存储在 Amazon S3 用户元数据中的文件元数据可使用 AWS Storage Gateway与文件网关上 NFS 共享互操作。文件网关支持从本地网络对复制到 Amazon S3 的数据进行低延迟

从其中一个位置复制时	到此位置	DataSync 可以复制
		访问。 DataSync此元数据也可以与 Lustre 互操作 FSx 。 将包含此元数据的对象 DataSync 复制回 NFS 服务器时，文件元数据将被恢复。还原元数据要求向 NFS 服务器授予提升的权限。有关更多信息，请参阅 使用 NFS 文件服务器配置 AWS DataSync 传输。

在 Amazon S3 和其他对象存储之间

在这些位置之间复制时	DataSync 可以复制

• 对象存储
• Amazon S3
	• 用户定义的对象元数据
	• 对象标签
	• 以下系统定义的对象元数据：
	  • Content-Disposition
	  • Content-Encoding
	  • Content-Language
	  • Content-Type
	
	注意：在初始传输期间 DataSync复制所有对象的系统级元数据。如果您将任务配置为仅传输已更改的数据，则除非还修改了对象的内容或用户元数据，否则 DataSync 不会在后续传输中复制系统元数据。
	
	DataSync 不会复制其他对象元数据，例如对象访问控制列表 (ACLs)、之前的对象版本或 Last-Modified 密钥。

• Microsoft Azure Blob Storage • Amazon S3	

在 Amazon S3 和 HDFS 之间

在这些位置之间复制时	DataSync 可以复制
• Hadoop Distributed File System（HDFS） • Amazon S3	以下是 Amazon S3 用户元数据： • 文件和文件夹修改时间戳 • 文件和文件夹访问时间戳（DataSync 只能在尽力而为的基础上执行此操作） • 用户 ID 和组 ID • POSIX 权限 HDFS 使用字符串来存储文件和文件夹的用户和群组所有权，而不是数字标识符，例如 UIDs 和 GIDs。

在 NFS 传输中复制的元数据

下表描述了 DataSync 可以在使用网络文件系统 (NFS) 的位置之间复制哪些元数据。

在这些位置之间复制时	DataSync 可以复制
• NFS • Amazon EFS • 亚马逊 f FSx or Lustre • FSx 适用于 OpenZFS 的亚马逊 • FSx 适用于 NetApp ONTAP 的亚马逊（使用 NFS）	• 文件和文件夹修改时间戳 • 文件和文件夹访问时间戳（DataSync 只能在尽力而为的基础上执行此操作） • 用户 ID (UID) 和组 ID (GID) • POSIX 权限

在 SMB 传输中复制的元数据

下表描述了 DataSync 可以在使用服务器消息块 (SMB) 的位置之间复制哪些元数据。

在这些位置之间复制时	DataSync 可以复制
• SMB • FSx 适用于 Windows 文件服务器的亚马逊 • FSx 适用于 ONTAP（使用 SMB）	• 文件时间戳：访问时间、修改时间和创建时间 • 文件所有者安全标识符 (SID) • 标准文件属性：只读 (R)、存档 (A)、系统 (S)、隐藏 (H)、压缩 (C)、未编制内容索引 (I)、加密 (E)、临时 (T)、脱机 (O) 和稀疏文件 (P) DataSync 尝试尽力复制存档 (A)、压缩 (C)、非上下文索引 (I)、稀疏属性 (P) 和临时 (T) 属性。如果这些属性未应用于目标位置，则在任务验证期间会忽略这些属性。 • NTFS 自由访问列表 (DACLs)，用于决定是否授予对对象的访问权限。 • NTFS 系统访问控制列表 (SACLs)，管理员使用这些列表记录访问安全对象的尝试。 注意： SACLs 如果您使用 SMB 版本 1.0，则不会被复制。 正在复制， DACLs 并且 SACLs 需要向 DataSync 使用 SMB 访问您的位置的 Windows 用户授予特定权限。有关更多信息，请参阅为 SMB、Window FSx s 文件服务器或 FSx ONTAP 创建位置（取决于传输中的位置类型）。

在其他传输场景中复制的元数据

DataSync 在这些存储系统（其中大多数具有不同的元数据结构）之间进行复制时，使用以下方式处理元数据。

从其中一个位置复制时	到其中一个位置	DataSync 可以复制
• SMB • FSx 适用于 Windows 文件服务器 • FSx 适用于 ONTAP（使用 SMB）	• Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS） • Amazon S3 • 对象存储 • Azure Blob Storage • NFS	目标文件系统上所有文件和文件夹或目标 S3 存储桶中对象的 默认 POSIX 元数据。该方法包括使用默认 POSIX 用户 ID 和组 ID 值。 基于 Windows 的元数据（例如 ACLs）不会被保留。
• 对象存储 • Amazon S3 • Azure Blob Storage	• Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS）	目标文件和文件夹上的默认 POSIX 元数据。该方法包括使用默认 POSIX 用户 ID 和组 ID 值。
• Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS）	• Azure Blob Storage	以下是用户定义元数据： • 文件和文件夹修改时间戳 • 文件和文件夹访问时间戳（DataSync 只能在尽力而为的基础上执行此操作） • 用户 ID 和组 ID • POSIX 权限
• HDFS	• Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS）	• 文件和文件夹修改时间戳 • 文件和文件夹访问时间戳（DataSync 只能在尽力而为的基础上执行此操作） • POSIX 权限

从其中一个位置复制时	到其中一个位置	DataSync 可以复制
		HDFS 将文件和文件夹用户和组的所有权存储为字符串，而不是数字标识符（例如 UIDs 和 GIDs）。 UIDs 和的默认 GIDs 值应用于目标文件系统。有关更多信息，请参阅 了解何时以及如何 DataSync 应用默认 POSIX 元数据。
• Amazon S3 • Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 Windows 文件服务器 • FSx 适用于 ONTAP	• HDFS	源位置的文件和文件夹时间戳。文件或文件夹所有者根据您在创建 HDFS 传输位置时指定的 HDFS 用户或 Kerberos 主体来设置。Hadoop 集群上的组映射配置决定了该组。
• Amazon S3 • Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS） • 对象存储 • NFS • HDFS	• SMB • FSx 适用于 Windows 文件服务器 • FSx 适用于 ONTAP（使用 SMB）	源位置的文件和文件夹时间戳。所有权是根据中指定的访问亚马逊 FSx 或 SMB 共享的 Windows 用户 DataSync 来设置的。权限继承自父目录。
• Azure Blob Storage	• FSx 适用于 Windows 文件服务器 • FSx 适用于 ONTAP（使用 SMB）	

了解何时以及如何 DataSync 应用默认 POSIX 元数据

DataSync 在以下情况下应用默认 POSIX 元数据：

- 当传输的源位置和目标位置不具有类似的元数据结构时
- 当源位置缺少元数据时

下表描述了在这些类型的传输期间如何 DataSync 应用默认 POSIX 元数据：

来源	目标位置	文件权限	文件夹权限	UID	GID
• Amazon S3[1] • 对象存储 [1] • Microsoft Azure Blob Storage [1]	• Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS） • NFS	0755	0755	65534	65534
• SMB	• Amazon S3 • 对象存储 • Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS） • NFS	0644	0755	65534	65534

来源	目标位置	文件权限	文件夹权限	UID	GID
• HDFS	• Amazon EFS • FSx 为了光泽 • FSx 适用于 OpenZFS • FSx 适用于 ONTAP（使用 NFS） • NFS	0644	0755	65534	65534

[1] 如果对象没有先前应用的元数据 DataSync。

由复制的链接和目录 AWS DataSync

AWS DataSync 根据传输中涉及的存储位置，硬链接、符号链接和目录的处理方式会有所不同。

硬链接

以下是在一些常见的传输场景中 DataSync 处理硬链接的方法：

- 在 NFS 文件服务器、Lustre、 FSx OpenZFS、 FSx FSx ONTAP（使用 NFS）和 Amazon EFS 之间传输时，会保留硬链接。
- 传输到 Amazon S3 时，硬链接引用的每个底层文件仅传输一次。在增量传输期间，将在您的 S3 存储桶中创建单独的对象。如果 Amazon S3 中的硬链接未更改，则在传输到 NFS 文件服务器、Lustre、OpenZFS、 FSx ONTAP（使用 NFS）或 Amazon EFS 文件系统时，硬链接会正确恢复。 FSx FSx
- 传输到 Microsoft Azure Blob Storage 时，硬链接引用的每个底层文件仅传输一次。在增量传输期间，如果源中有新的引用，则会在您的 blob 存储中创建单独的对象。从中传输时Azure Blob Storage，将硬链接当作单个文件进行 DataSync 传输。
- 在 SMB 文件服务器、 FSx Windows 文件服务器和 FSx ONTAP（使用 SMB）之间传输时，不支持硬链接。如果在这些情况下 DataSync 遇到硬链接，则传输任务完成时会出现错误。要了解更多信息，请查看您的 CloudWatch 日志。
- 传输到 HDFS 时，不支持硬链接。 CloudWatch 日志显示这些链接已跳过。

符号链接

以下是在一些常见的传输场景中如何 DataSync 处理符号链接：

- 在 NFS 文件服务器、Lustre、 FSx OpenZFS、 FSx ONTAP（使用 NFS）和 Amazon EFS 之间传输时，符号链接会保留。 FSx
- 在传输到 Amazon S3 时，链接目标路径将存储在 Amazon S3 对象中。在传输到 NFS 文件服务器、Lustre、OpenZFS、 FSx ONTAP 或 Amazon EFS 文件系统时， FSx 链接可以正确恢复。 FSx
- 传输到时 Azure Blob Storage，不支持符号链接。 CloudWatch 日志显示这些链接已跳过。
- 在 SMB 文件服务器、 FSx Windows 文件服务器和 FSx ONTAP（使用 SMB）之间传输时，不支持符号链接。 DataSync 不会传输符号链接本身，而是传输符号链接引用的文件。要识别重复文件并使用符号链接去重，您必须在目标文件系统上配置去重。
- 传输到 HDFS 时，不支持符号链接。 CloudWatch 日志显示这些链接已跳过。

目录

通常，在存储系统之间传输时会 DataSync 保留目录。以下情况并非如此：

- 传输到 Amazon S3 时，目录表示为带有前缀的空对象，并以正斜杠（/）结尾。
- 在传输到没有分层名称空间的 Azure Blob Storage 时，目录不存在。看起来像目录的内容只是对象名称的一部分。

配置文件、对象和元数据的处理方式

您可以配置在不同位置之间传输时如何 AWS DataSync 处理文件、对象及其关联的元数据。

例如，对于定期传输，您可能希望使用源中的更改覆盖目标中的文件，以保持位置同步。您可以复制诸如文件和文件夹的 POSIX 权限、与对象关联的标签以及访问控制列表 (ACLs) 之类的属性。

传输模式选项

您可以配置是仅 DataSync 传输初始副本后更改的数据（包括元数据），还是每次运行任务时传输所有数据。如果计划进行定期传输，则可能仅需要传输自上次执行任务以来发生更改的内容。

控制台中的选项	API 中的选项	说明
仅传输更改的数据	TransferMode 设置为 CHANGED	初次完全传输后，仅 DataSync 复制源位置和目标位置不同的数据和元数据。
传输所有数据	TransferMode 设置为 ALL	DataSync 将源中的所有内容复制到目的地，而不比较位置之间的差异。

文件和对象处理选项

您可以控制如何 DataSync 处理目标位置中的文件或对象的某些方面。例如， DataSync 可以删除目标中不在源中的文件。

控制台中的选项	API 中的选项	说明
保留已删除数据	PreserveDeletedFiles	指定是否在目标位置 DataSync 维护源中不存在的文件或对象。 如果将任务配置为从 Amazon S3 存储桶中删除对象，则可能会为某些存储类产生最低存储持续时间费用。有关详细信息，请参阅 Amazon S3 传输的存储类别注意事项。 **Warning** 无法将任务配置为删除目标中的数据，同时又传输所有数据。当您传输所有数据时，DataSync不会扫描您

控制台中的选项	API 中的选项	说明
		的目标位置，也不知道要删除什么。
覆盖文件	OverwriteMode	指定在源数据或元数据发生更改时是否 DataSync 修改目标位置的数据。如果未将任务配置为覆盖数据，即使源数据不同，目标数据也不会覆盖。 如果您的任务覆盖对象，则您可能需要为某些存储类别（例如检索或提前删除）支付额外费用。有关详细信息，请参阅 Amazon S3 传输的存储类别注意事项。

元数据处理选项

DataSync 可以在传输过程中保留文件和对象的元数据。 DataSync 可以保留的元数据取决于所涉及的存储系统以及这些系统是否使用类似的元数据结构。

在配置任务之前，请务必了解在源位置和目标位置之间传输时如何 DataSync 处理元数据和特殊文件。

> **Important**
>
> DataSync 支持与某些第三方云存储系统（例如 Google Cloud Storage 和 IBM Cloud Object Storage）之间的传输，这些系统以不完全兼容 S3 的方式处理系统元数据。对于这些传 DataSync 输，会尽力复制元数据属性 `ContentType` `ContentEncoding`、`ContentLanguage`、例如、和`CacheControl`。如果目标存储系统未应用这些属性，则在任务验证过程中会将其忽略。

控制台中的选项	API 中的选项	说明
副本所有权	Gid 和 Uid	指定是否 DataSync 复制 POSIX 文件和文件夹所有权，例如文件所有者的组 ID 和文件所有者的用户 ID。
复制权限	PosixPermissions	指定是否 DataSync 将文件和文件夹的 POSIX 权限从源复制到目标。
复制时间戳	Atime 和 Mtime	指定是否 DataSync 将时间戳元数据从源复制到目标。当您需要多次运行任务时，必须使用这些选项。
复制对象标签	ObjectTags	指定在对象存储系统之间传输时是否 DataSync 保留与您的对象关联的标签。
副本所有权 DACLs、和 SACLs	SecurityDescriptorCopyFlags 设置为 `OWNER_DACL_SACL`	DataSync 复制以下内容： • 对象所有者。 • NTFS 自由访问列表 (DACLs)，用于决定是否授予对对象的访问权限。 • NTFS 系统访问控制列表 (SACLs)，管理员使用这些列表记录访问安全对象的尝试。 注意： SACLs 如果您使用 SMB 版本 1.0，则不会被复制。 正在复制， DACLs 并且 SACLs 需要向 DataSync 使用 SMB 访问您的位置的

控制台中的选项	API 中的选项	说明
		Windows 用户授予特定权限。有关更多信息，请参阅为 SMB、Window FSx s 文件服务器或 FSx ONTAP 创建位置（取决于传输中的位置类型）。
副本所有权和 DACLs	SecurityDescriptorCopyFlags 设置为 OWNER_DACL	DataSync 复制以下内容： • 对象所有者。 • DACLs，它决定是否授予对对象的访问权限。 DataSync 选择此选项 SACLs 时不会复制。
不要复制所有权或 ACLs	SecurityDescriptorCopyFlags 设置为 NONE	DataSync 不复制任何所有权或权限数据。 DataSync 写入目标位置的对象归提供访问目标的凭证 DataSync的用户所有。目标对象权限是根据目标服务器上配置的权限确定的。

配置文件、对象和元数据处理选项

您可以配置在创建、编辑或启动传输任务时如何 DataSync 处理文件、对象和元数据。

使用控制 DataSync 台

以下说明介绍了如何在创建任务时配置文件、对象和元数据处理选项。

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。

4. 对于传输方式，请选择下列选项之一：

   - 仅传输更改的数据
   - 传输所有数据

   有关这些选项的详细信息，请参阅传输模式选项。

5. 如果 DataSync 要在目标位置维护源中不存在的文件或对象，请选择 “保留已删除的文件”。

   如果您不选择此选项，并且您的任务从 Amazon S3 存储桶中删除对象，则可能会为某些存储类产生最低存储持续时间费用。有关详细信息，请参阅 Amazon S3 传输的存储类别注意事项。

   > **Warning**
   >
   > 您不能取消选择此选项并启用传输所有数据。当您传输所有数据时， DataSync 不会扫描您的目标位置，也不知道要删除什么。

6. 如果 DataSync要在源数据或元数据发生更改时修改目标位置的数据，请选择 “覆盖文件”。

   如果您的任务覆盖对象，则您可能需要为某些存储类别（例如检索或提前删除）支付额外费用。有关详细信息，请参阅 Amazon S3 传输的存储类别注意事项。

   如果不选择此选项，即使源数据不同，目标数据也不会被覆盖。

7. 在 “传输选项” 下，选择 DataSync 要如何处理元数据。有关选项的更多信息，请参阅元数据处理选项。

   > **Important**
   >
   > 控制台中的选项取决于任务的源位置和目标位置。您可能需要展开其他设置，才能看到部分选项。

   - 副本所有权
   - 复制权限
   - 复制时间戳
   - 复制对象标签
   - 副本所有权 DACLs、和 SACLs
   - 副本所有权和 DACLs

- 不要复制所有权或 ACLs

使用 DataSync API

可通过在以下任何操作中使用 `Options` 参数来配置文件、对象和元数据处理选项：

- CreateTask
- StartTaskExecution
- UpdateTask

配置如何 AWS DataSync 验证数据完整性

在传输过程中， AWS DataSync 使用校验和验证来验证您在不同位置之间复制的数据的完整性。您还可以配置 DataSync 为在传输结束时执行其他验证。

数据验证选项

使用以下信息来帮助您决定是否以及 DataSync 如何执行这些额外检查。

控制台选项	API 选项	说明
仅验证传输数据（推荐）	VerifyMode 设置为 `ONLY_FILES_TRANSFERRED`	DataSync 计算源位置传输的数据（包括元数据）的校验和。在传输结束时， DataSync 将此校验和与在目的地根据相同数据计算的校验和进行比较。 在传输到 S3 Glacier Flexible Retrieval 或 S3 Glacier Deep Archive 存储类时，推荐使用此选项。有关更多信息，请参阅 Amazon S3 传输的存储类别注意事项。
验证所有数据	VerifyMode 设置为 `POINT_IN_TIME_CONSISTENT`	传输结束时， DataSync 检查整个来源和目的地，以验证两个位置是否完全同步。

控制台选项	API 选项	说明
		Note 当任务使用增强模式时不支持。 如果您使用清单，则 DataSync 仅扫描和验证清单中列出的内容。 在传输到 S3 Glacier Flexible Retrieval 或 S3 Glacier Deep Archive 存储类时，不支持此选项。有关更多信息，请参阅 Amazon S3 传输的存储类别注意事项。
传输后不要验证数据	VerifyMode 设置为 NONE	DataSync 仅在传输期间执行数据完整性检查。与其他选项不同，在传输结束时没有额外验证。

配置数据验证

可在创建任务、更新任务或开始执行任务时配置数据验证选项。

使用控制 DataSync 台

以下说明介绍了如何在创建任务时配置数据验证选项。

使用控制台配置数据验证

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。

4. 对于验证，请选择下列选项之一：

   - 仅验证传输数据（推荐）
   - 验证所有数据
   - 传输后不要验证数据

使用 DataSync API

您可以通过将`VerifyMode`参数与以下任一操作一起使用来配置 DataSync 验证数据的方式：

- CreateTask
- UpdateTask
- StartTaskExecution

为 AWS DataSync 任务设置带宽限制

您可以为 AWS DataSync 任务及其每次执行配置网络带宽限制。

限制任务带宽

创建、编辑或启动任务时设置带宽限制。

使用控制 DataSync 台

以下说明介绍了如何在创建任务时为其配置带宽限制。

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 在带宽限制中，选择以下选项之一：

   - 选择使用可用可将所有可用网络带宽用于每个任务执行。
   - 选择 “设置带宽限制 (MiB/s)”，然后输入 DataSync 要用于每次任务执行的最大带宽。

使用 DataSync API

您可以通过在以下任何操作中使用 `BytesPerSecond` 参数配置带宽：

- CreateTask
- UpdateTask
- StartTaskExecution

限制任务执行带宽

您可以修改正在运行或排队的任务执行的带宽限制。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务。
3. 选择任务，然后选择历史记录以查看任务的执行情况。
4. 选择要修改的任务执行，然后选择编辑。
5. 在对话框中，执行下列操作之一：
   - 选择使用可用可将所有可用网络带宽用于任务执行。
   - 选择 “设置带宽限制 (MiB/s)”，然后输入 DataSync 要用于执行任务的最大带宽。
6. 选择保存更改。

   新的带宽限值将在 60 秒内生效。

使用 DataSync API

您可以通过在UpdateTaskExecution操作中使用`BytesPerSecond`参数来修改正在运行或排队的任务执行的带宽限制。

计划 AWS DataSync 任务的运行时间

您可以设置 AWS DataSync 任务计划，定期在存储位置之间传输数据。

DataSync 任务计划的工作方式

计划的 DataSync 任务按您指定的频率运行，最小间隔为 1 小时。您可以使用 cron 或 rate 表达式创建任务计划。

> **Important**
>
> 您无法将任务安排为以快于 1 小时的间隔运行。

使用 cron 表达式

对于在特定时间和日期运行的任务计划，请使用 cron 表达式。例如，下面展示了如何在 AWS CLI 中配置任务计划，使其在每周日和周三中午 12:00（UTC 时间）运行。

```
cron(0 12 ? * SUN,WED *)
```

使用 rate 表达式

对定期运行的任务计划（例如每 12 小时运行一次），请使用 rate 表达式。例如，下面展示了如何在 AWS CLI 中配置任务计划，使其每 12 小时运行一次。

```
rate(12 hours)
```

> **Tip**
>
> 有关 cron 和 rate 表达式语法的更多信息，请参阅 Amazon EventBridge 用户指南。

创建 DataSync 任务计划

您可以使用 DataSync 控制台、AWS CLI、或 DataSync API 安排任务的运行频率。

使用 DataSync 控制台

以下说明介绍了在创建任务时，如何设置计划。稍后，在编辑任务时，您还可以修改计划。

在控制台中，一些计划选项允许您指定任务运行的确切时间（例如每天晚上 10:30）。如果不为这些选项指定时间，将在您创建（或更新）任务的时间运行该任务。

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。

4. 对于计划频率，执行以下操作之一：

   - 如果不想按计划运行任务，请选择未计划。
   - 选择每小时，然后选择要在一小时中的哪一分钟运行任务。
   - 选择每天，然后输入要运行任务的 UTC 时间。
   - 选择每周和一周中的某一天，然后输入要运行任务的 UTC 时间。
   - 选择周中日期，选择具体的一天或数天，然后以 HH:MM 格式输入应运行任务的 UTC 时间。
   - 选择自定义，然后选择 Cron 表达式或 Rate 表达式。输入任务计划，最小时间间隔为 1 小时。

使用 AWS CLI

您可以在 `create-task`、`update-task` 或 `start-task-execution` 命令中使用 `--schedule` 参数，为 DataSync 任务创建计划。

以下说明介绍了如何使用 `create-task` 命令执行此操作。

1. 复制以下 `create-task` 命令：

```
aws datasync create-task \
  --source-location-arn arn:aws:datasync:us-east-1:123456789012:location/
loc-12345678abcdefgh \
  --destination-location-arn arn:aws:datasync:us-east-1:123456789012:location/
loc-abcdefgh12345678 \
  --schedule '{
    "ScheduleExpression": "cron(0 12 ? * SUN,WED *)"
  }'
```

2. 对于 `--source-location-arn` 参数，指定数据传输源位置的 Amazon 资源名称（ARN）。
3. 对于 `--destination-location-arn` 参数，指定数据传输目标位置的 ARN。
4. 对于 `--schedule` 参数，为您的计划指定 cron 或 rate 表达式。

   在示例中，cron 表达式 *`cron(0 12 ? * SUN,WED *)`* 设置了在每周日和每周三中午 12:00（UTC 时间）运行的任务计划。
5. 运行 `create-task` 命令创建带有计划的任务。

暂停 DataSync 任务计划

在某些情况下，您可能需要暂停 DataSync 任务计划。例如，您可能需要暂时禁用定期传输，以修复任务中的问题，或对存储系统进行维护。

出于以下原因，DataSync 可能会自动禁用您的任务计划：

- 您的任务因相同错误反复失败。
- 任务正在使用的 AWS 区域 被禁用。

使用 DataSync 控制台

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务。
3. 选择要暂停计划的任务，然后选择编辑。
4. 在计划中，关闭启用计划。选择保存更改。

使用 AWS CLI

1. 复制以下 `update-task` 命令：

```
aws datasync update-task \
  --task-arn arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh \
  --schedule '{
    "ScheduleExpression": "cron(0 12 ? * SUN,WED *)",
    "Status": "DISABLED"
  }'
```

2. 对于 `--task-arn` 参数，指定要暂停其计划的任务 ARN。
3. 对于 `--schedule` 参数，请执行以下操作：
   - 对于 `ScheduleExpression`，为计划指定一个 cron 或 rate 表达式。

     在示例中，表达式 *`cron(0 12 ? * SUN,WED *)`* 设置了在每周日和每周三中午 12:00（UTC 时间）运行的任务计划。
   - 对于 `Status`，指定 `DISABLED` 以暂停任务计划。
4. 运行 `update-task` 命令。
5. 要恢复计划，请运行相同的 `update-task` 命令，并将 `Status` 设置为 `ENABLED`。

检查 DataSync 任务计划的状态

您可以查看 DataSync 任务计划是否已启用。

使用 DataSync 控制台

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务。
3. 在计划列中，检查任务的计划是启用还是禁用。

使用 AWS CLI

1. 复制以下 `describe-task` 命令：

```
aws datasync describe-task \
  --task-arn arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh
```

2. 对于 `--task-arn` 参数，请指定您想要了解其信息的任务 ARN。
3. 运行 `describe-task` 命令。

您将收到提供任务详细信息的响应，其中包括任务的计划。（下面的示例主要关注任务计划配置，不显示完整的 `describe-task` 响应）。

该示例显示任务的计划已手动禁用。如果 DataSync SERVICE 禁用了计划，则会显示 `DisabledReason` 的错误消息，帮助您了解任务不断失败的原因。有关更多信息，请参阅 ???。

```
{
    "TaskArn": "arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh",
    "Status": "AVAILABLE",
    "Schedule": {
        "ScheduleExpression": "cron(0 12 ? * SUN,WED *)",
        "Status": "DISABLED",
        "StatusUpdateTime": 1697736000,
        "DisabledBy": "USER",
        "DisabledReason": "Manually disabled by user."
    },
    ...
}
```

为你的 AWS DataSync 任务加标签

标签是键值对，可帮助您管理、筛选和搜索 AWS DataSync 资源。您最多可以为每个 DataSync 任务和任务执行添加 50 个标签。

例如，您可以为大型数据迁移创建任务，并使用密钥**Project**和值标记任务**Large Migration**。若要进一步组织迁移，您可以使用密钥**Transfer Date**和值**May 2021**标记任务的一次运行（后续任务执行可能会被标记**June 2021**、**July 2021**等）。

给你的 DataSync 任务加标签

只有在创建 DataSync 任务时才能为任务添加标签。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 在配置设置页面，选择添加新标签来标记您的任务。

使用 AWS CLI

1. 复制以下 `create-task` 命令：

```
aws datasync create-task \
    --source-location-arn 'arn:aws:datasync:region:account-id:location/source-
location-id' \
    --destination-location-arn 'arn:aws:datasync:region:account-
id:location/destination-location-id' \
    --tags Key=tag-key,Value=tag-value
```

2. 指定命令中的以下参数：

   - `--source-location-arn` – 指定传输中源位置的 Amazon 资源名称（ARN）。
   - `--destination-location-arn` – 请指定传输中目标位置的 ARN。
   - `--tags` – 指定要应用于任务的标签。

     对于多个标签，请用空格分隔密钥值对。

3. （可选）指定适用于您传输场景的其他参数。

   有关--options列表，请参阅 create-task命令。

4. 运行 create-task 命令。

   您会收到一条回复，表明您刚刚创建的任务。

```
{
    "TaskArn": "arn:aws:datasync:us-east-2:123456789012:task/task-
abcdef01234567890"
}
```

要查看您为此任务添加的标签，可以使用list-tags-for-resource命令。

为 DataSync 任务执行添加标签

您可以为 DataSync 任务的每一次运行添加标签。

如果您的任务已有标签，请记住以下有关任务执行中使用标签的信息：

- 如果您使用控制台启动任务，其用户创建的标签将自动应用于任务执行。但是，系统创建的以开头的标签 aws:不适用。
- 如果您使用 DataSync API 或启动任务 AWS CLI，则其标签不会自动应用于任务执行。

使用控制 DataSync 台

要在任务执行中添加、编辑或移除标签，必须通过覆盖选项启动任务。

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务。
3. 选择任务。
4. 选择开始，然后选择下列选项之一：
   - 从默认值开始-应用与您的任务关联的所有标签。
   - 从@@ 覆盖选项开始 — 允许您添加、编辑或删除此特定任务执行的标签。

使用 AWS CLI

1. 复制以下 `start-task-execution` 命令：

```
aws datasync start-task-execution \
    --task-arn 'arn:aws:datasync:region:account-id:task/task-id' \
    --tags Key=tag-key,Value=tag-value
```

2. 指定命令中的以下参数：

   - `--task-arn` – 指定要启动的任务的 ARN。
   - `--tags` – 指定要应用于此特定任务运行的标签。

     对于多个标签，请用空格分隔密钥值对。

3. （可选）指定对您有意义的其他参数。

   有关更多信息，请参阅 start-task-execution 命令。

4. 运行 `start-task-execution` 命令。

   您会收到一条回复，其中显示了您刚刚开始的任务执行情况。

```
{
    "TaskExecutionArn": "arn:aws:datasync:us-east-2:123456789012:task/task-
abcdef01234567890"
}
```

要查看您为此任务添加的标签，可以使用list-tags-for-resource命令。

启动传输数据的任务

创建 AWS DataSync传输任务后，就可以开始移动数据了。每次任务运行都称为任务执行。若要了解任务执行中会发生的事情，请参阅如何 DataSync 传输文件、对象和目录。

> **Important**
>
> 如果您计划向或从某个 Amazon S3 位置传输数据，请在开始之前先查看DataSync 会如何影响您的 S3 请求费用以及DataSync 定价页面。

开始您的任务

创建任务后，您可立即开始移动数据。

使用 DataSync 控制台

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务。
3. 选择您想要运行的任务。

   确保任务处于可用状态。您也可以选择多个任务。
4. 选择操作，然后选择下列选项之一：
   - 开始 - 运行任务（如果您选择了多个任务，则运行任务）。
   - 从覆盖选项开始 - 允许您在开始移动数据之前修改某些任务设置。准备就绪后，选择 开始。
5. 选择查看执行详细信息 以查看有关正在运行的任务执行的详细信息。

使用 AWS CLI

要启动 DataSync 任务，您只需指定您要运行的任务的 Amazon 资源名称（ARN）。以下是一个示例 `start-task-execution` 命令：

```
aws datasync start-task-execution \
    --task-arn 'arn:aws:datasync:region:account-id:task/task-id'
```

以下示例使用一些与任务默认设置不同的设置来启动任务：

```
aws datasync start-task-execution \
    --override-options VerifyMode=NONE,OverwriteMode=NEVER,PosixPermissions=NONE
```

该命令为任务执行返回一个 ARN，类似于以下示例：

```
{
    "TaskExecutionArn": "arn:aws:datasync:us-east-1:209870788375:task/
task-08de6e6697796f026/execution/exec-04ce9d516d69bd52f"
}
```

> **Note**
>
> 每个代理每次可以运行单个任务。

使用 DataSync API

你可以使用StartTaskExecution运算开始你的任务。使用DescribeTaskExecution运算获取有关正在运行的任务执行的详细信息。

启动后，您可以在 DataSync 复制数据时检查任务执行状态。如果需要，您还可以限制任务执行带宽。

任务执行状态

当您启动数据同步任务时，可能会看到这些状态。（任务状态与任务执行状态不同。）

控制台状态	API 状态	描述
排队	QUEUED	另一项任务正在执行，并使用相同的 DataSync 代理。有关更多信息，请参阅 了解任务何时排队。
正在启动	LAUNCHING	DataSync 正在初始化任务执行。此状态通常很快就会结束，但也可能需要几分钟。
已启动	LAUNCHED	DataSync 已启动任务执行。
正在准备	PREPARING	DataSync 正在确定要传输哪些数据。 准备过程可能只需要几分钟，也可能需要几个小时甚至更长时间，具体取决于两个位置的文件、对象或目录的数量，以及任务的配置方式。准备方式还取决于具体任务模式。有关更多信息，请参阅 如何 DataSync 准备数据传输。
正在传输	TRANSFERRING	DataSync 正在执行实际数据传输。
正在验证	VERIFYING	在传输结束时，DataSync 验证数据的完整性。
成功	SUCCESS	任务执行成功。

控制台状态	API 状态	描述
正在取消	CANCELLING	正在取消任务的执行。
错误	ERROR	任务执行失败。

了解任务何时排队

运行多个任务时（例如，传输大型数据集时），DataSync 可能会将任务放入队列，按序列连续运行（先进先出）。发生这种情况的一些示例包括：

- 运行使用相同 DataSync 代理的不同任务。虽然您可以将同一个代理用于多个任务，但一个代理一次只能运行一个任务。
- 任务正在执行，您使用不同的筛选条件或清单开始再次执行同一个任务。

在各个示例中，队列中的任务在前面的任务完成之前不会启动。

取消执行任务

您可以停止执行任何正在运行或排队的 DataSync 任务。

使用控制台取消执行任务

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务。
3. 为要监控的正在运行的任务选择 任务 ID。

   任务状态应为正在运行。
4. 选择历史记录以查看任务的执行情况。
5. 选择要停止的任务执行，然后选择停止。
6. 在对话框中，选择停止。

要使用 DataSync API 取消正在运行或排队的任务，请参阅CancelTaskExecution。

自动取消卡住的任务

有时，正在运行的 DataSync 任务执行可能会卡滞。

监控您的 AWS DataSync 转账

要维护 AWS DataSync 传输活动的可靠性和性能，实施监控非常重要。我们建议您收集监控数据，以便更轻松地调试出现的错误。但是 DataSync，在开始监控之前，请先制定一份包含以下问题答案的监控计划：

- 监控目的是什么？
- 您将监控哪些资源？
- 监控这些资源的频率如何？
- 您将使用哪些监控工具？
- 谁负责执行监控任务？
- 出现错误时应通知谁？

AWS 提供各种监控服务和工具 DataSync。可以配置其中一些工具为您执行监控任务，但有些工具需要手动干预。建议您尽可能实现监控任务自动化。

主题

- 了解数据传输性能计数器
- 使用 Amazon CloudWatch 指标监控数据传输
- 使用任务报告监控您的数据传输
- 使用 Amazon CloudWatch 日志监控数据传输
- 使用记录 AWS DataSync API 调用 AWS CloudTrail
- 使用 Amazon 监控事件 EventBridge
- AWS DataSync 使用手动工具进行监控

了解数据传输性能计数器

启动任务时，会 AWS DataSync 提供计数器来帮助跟踪数据传输的性能和进度。

使用以下信息，了解每个计数器代表的信息。您可以在 DataSync 控制台或DescribeTaskExecution响应中查看这些计数器。某些计数器并非在所有任务模式下都可用。

控制台	**DescribeTaskExecution**	任务模式支持	说明
–	BytesWritten	增强、基本	DataSync 实际写入目标位置的逻辑字节数。
数据吞吐量	–	增强、基本	向目标位置 DataSync 写入逻辑字节的速率。 如果你使用的是 DescribeTaskExecution，如何计算这个计数器取决于你的任务模式： • 增强模式：BytesWritten 除以 TotalDuration • 基本模式：BytesWritten 除以 TransferDuration
传输的数据	BytesTransferred	增强、基本	压缩前 DataSync 发送到网络的字节数（如果可以压缩）。 有关通过网络传输的字节数，请参阅网络吞吐量（在控制台中）或BytesCompressed （输入

控制台	DescribeTaskExecution	任务模式支持	说明
			DescribeTaskExecution）计数器。
已从目标删除	FilesDeleted	Basic	在目标位置 DataSync 实际删除的文件、对象和目录的数量。 如果您未将任务配置为在目标位置中删除不存在于源位置的数据： • 已从目标删除不会显示在控制台中。 • FilesDeleted 始终显示的值为 0。
已从目标删除	FilesDeleted，FoldersDeleted	增强	目标位置中 DataSync 实际删除的文件或对象以及目录的数量。 如果您未将任务配置为在目标位置中删除不存在于源位置的数据： • 已从目标删除不会显示在控制台中。 • FilesDeleted 并FoldersDeleted 始终显示的值为0。

控制台	**DescribeTaskExecution**	任务模式支持	说明
–	`EstimatedBytesToTransfer`	增强、基本	DataSync 预期写入目标位置的逻辑字节数。
–	`EstimatedFilesToDelete`	Basic	DataSync 预计在目标位置删除的文件、对象和目录的数量。 如果您未将任务配置为在目标位置中删除不存在于源位置的数据，则该值始终为 0。
–	`EstimatedFilesToDelete`，`EstimatedFoldersToDelete`	增强	目标位置中 DataSync 预计要删除的文件或对象以及目录的数量。 如果您未将任务配置为在目标位置中删除不存在于源位置的数据，则该值始终为 0。

控制台	**DescribeTaskExecution**	任务模式支持	说明

–	`EstimatedFilesToTransfer`	Basic	DataSync 预计通过网络传输的文件、对象和目录的数量。此值是在 DataSync 准备传输时计算的。
			
			计算方法主要取决于使用的传输模式：
			
			• 如果传输模式设置为仅传输更改的数据：计算方法是比较源位置和目标位置的内容，确定需要传输的差异。差异可能包括：
			  • 在源位置添加或修改的任何内容。
			  • 存在于两个位置，且在初始传输后于目标位置发生修改的任何内容（除非将任务设置为不覆盖目标位置的数据）。
			  • DataSync 预计要删除的项目数量（如果您将任务配置为删除目标中的数据）。

控制台	DescribeTaskExecution	任务模式支持	说明
			• 如果传输模式设置为传输所有数据：计算仅基于在源位置 DataSync 找到的项目。

控制台	**DescribeTaskExecution**	任务模式支持	说明

–	`EstimatedFilesToTransfer`，`EstimatedFoldersToTransfer`	增强	DataSync 预计通过网络传输的文件或对象以及目录的数量。此值是在 DataSync 准备传输时计算的。
			
			计算方法主要取决于使用的传输模式：
			
			• 如果传输模式设置为仅传输更改的数据：计算方法是比较源位置和目标位置的内容，确定需要传输的差异。差异可能包括：
			  • 在源位置添加或修改的任何内容。
			  • 存在于两个位置，且在初始传输后于目标位置发生修改的任何内容（除非将任务设置为不覆盖目标位置的数据）。
			• 如果传输模式设置为传输所有数据：计算仅基于在源位置 DataSync 找到的项目。

控制台	DescribeTaskExecution	任务模式支持	说明
文件吞吐量	–	增强、基本	通过网络 DataSync 传输文件、对象和目录的速率。 如果你使用的是 DescribeTaskExecution，如何计算这个计数器取决于你的任务模式： • 增强模式：FilesTransferred 除以 TotalDuration • 基本模式：FilesTransferred 除以 TransferDuration
–	FilesFailed，FoldersFailed	增强	任务执行期间 DataSync 无法准备、传输、验证和删除的文件或对象以及目录的数量。 如果出现故障，可以分别在“准备好”、“已传输”、“已跳过”和“已从目标删除”控制台计数器旁边进行查看。

控制台	DescribeTaskExecution	任务模式支持	说明
在源处列出	FilesListed.AtSource，FoldersListed.AtSource	增强	在您的源位置 DataSync 找到的文件或对象以及目录的数量。 • 对于 DataSync 清单，仅列出清单中的内容（而不是源位置的所有内容）。 • 使用包含过滤器，仅 DataSync列出与来源位置的过滤器匹配的内容。 • 使用排除过滤器，在应用过滤器之前，会 DataSync 列出源位置的所有内容。
–	FilesListed.AtDestinationForDelete，FoldersListed.AtDestinationForDelete	增强	在目标位置 DataSync 找到的文件或对象以及目录的数量。 该计数器仅适用于您将任务配置为在目标位置中删除不存在于源位置的数据。

控制台	DescribeTaskExecution	任务模式支持	说明
网络吞吐量 *	BytesCompressed	增强、基本	压缩后通过网络 DataSync 传输的物理字节数（如果可以压缩）。 除非数据不可压缩，否则此数字通常小于传输的数据BytesTransferred （在控制台中 DescribeTaskExecution）或（输入）。 * 对于增强模式，网络吞吐量不会显示在控制台中。
压缩百分比	–	增强、基本	在通过网络发送之前 DataSync 压缩的传输数据所占的百分比。 如果你使用的是 DescribeTaskExecution，你可以用计算这个计数器1 - BytesCompressed / BytesWritten 。

控制台	DescribeTaskExecution	任务模式支持	说明
准备好	FilesPrepared，FoldersPrepared	增强	比较源位置和目标位置后 DataSync 将尝试传输的文件或对象以及目录的数量。 在控制台中，此计数器还可以显示准备过程中 DataSync 跳过的对象数量。有关更多信息，请参阅 如何 DataSync 准备数据传输。 如果您将任务配置为传输所有数据，则此计数器不适用。在这种情况下， DataSync 将所有内容从源复制到目的地，而不比较位置之间的差异。

控制台	DescribeTaskExecution	任务模式支持	说明
处理速率	–	增强、基本	DataSync 读取源位置文件、对象和目录的速率。 处理速率基于多个CloudWatch 指标。确切的指标取决于使用的任务模式。 增强模式： • FilesListedSource • FilesPrepared • FilesTransferred • FilesVerified 基本模式： • FilesPreparedSource • FilesPreparedDestination • FilesTransferred • FilesVerifiedSource • FilesVerifiedDestination

控制台	DescribeTaskExecution	任务模式支持	说明
剩余	–	Basic	DataSync预计通过网络传输的剩余文件、对象和目录数量。 如果你使用的是DescribeTaskExecution，你可以通过减去FilesTransferred 来计算这个计数器。EstimatedFilesToTransfer
已跳过*	FilesSkipped	Basic	传输过程中 DataSync 跳过的文件、对象和目录的数量。
–	FilesSkipped，FoldersSkipped	增强	传输过程中 DataSync 跳过的文件或对象以及目录的数量。 仅传输已更改的数据时，跳过的项目将包含在准备好的计数器中，或者在传输所有数据时包含在已传输的计数器中。

控制台	**DescribeTaskExecution**	任务模式支持	说明
已传输	FilesTransferred	Basic	通过网络 DataSync 传输的文件、对象和目录的数量。在任务执行期间，当从源读取内容并通过网络发送内容时，该值会定期更新。 如果 DataSync 无法传输某些内容，则此值可以小于EstimatedFilesToTransfer 或EstimatedFoldersToTransfer 。在某些情况下，该值也可以大于EstimatedFilesToTransfer 或EstimatedFoldersToTransfer 。此计数器是特定于某些位置类型的实施，因此不要将其用作所传输内容的确切指示或用来监控任务执行。

控制台	DescribeTaskExecution	任务模式支持	说明
已传输	FilesTransferred，FoldersTransferred	增强	通过网络 DataSync 传输的文件或对象以及目录的数量。在任务执行期间，当从源读取内容并通过网络发送内容时，该值会定期更新。 如果 DataSync 无法传输某些内容，则此值可以小于EstimatedFilesToTransfer 或EstimatedFoldersToTransfer 。在某些情况下，该值也可以大于EstimatedFilesToTransfer 或EstimatedFoldersToTransfer 。此计数器是特定于某些位置类型的实施，因此不要将其用作所传输内容的确切指示或用来监控任务执行。

控制台	DescribeTaskExecution	任务模式支持	说明
已验证	FilesVerified	Basic	传输过程中要 DataSync 验证的文件、对象和目录的数量。 当您将任务配置为仅验证传输的数据时，在某些情况下 DataSync 不验证目录或无法传输的文件或对象。
已验证	FilesVerified，FoldersVerified	增强	传输过程中要 DataSync 验证的文件或对象以及目录的数量。

使用 Amazon CloudWatch 指标监控数据传输

Amazon CloudWatch 提供了跟踪 DataSync 传输性能和解决传输任务问题的指标。

可使用 Amazon CloudWatch 指标监控 AWS DataSync 传输性能。DataSync 指标每隔 5 分钟自动发送至 CloudWatch（无论如何配置日志记录）。这些指标保留 15 个月的时间。

要查看 DataSync 的 CloudWatch 指标，您可以使用以下工具：

- CloudWatch 控制台
- CloudWatch CLI
- CloudWatch API
- DataSync 控制台（在任务执行的详细信息页面上）

有关更多信息，请参阅 Amazon CloudWatch 用户指南。

DataSync 的 CloudWatch 指标

DataSync 指标使用 aws/datasync 命名空间并提供以下维度的指标：

- AgentId — 代理的唯一 ID（如果任务使用代理）。
- TaskId — 任务的唯一 ID。其格式为 task-*01234567890abcdef*。

aws/datasync 命名空间包括以下指标。某些指标并非适用于所有任务模式。

CloudWatch 指标	任务模式支持	描述
BytesCompressed	基本	DataSync 在压缩后通过网络传输的物理字节数（如果可以压缩）。除非数据不可压缩，否则该数字通常小于 BytesTransferred 。 单位：字节
BytesPreparedDestination	基本	DataSync 在目标位置准备的逻辑字节数。 单位：字节
BytesPreparedSource	基本	DataSync 在源位置准备的逻辑字节数。 单位：字节
BytesTransferred	基本	DataSync 在压缩之前发送到网络的字节数（如果可以压缩）。有关通过网络传输的字节数，请参阅 BytesCompressed 指标。 单位：字节
BytesVerifiedDestination	基本	DataSync 在目标位置验证的逻辑字节数。 单位：字节
BytesVerifiedSource	基本	DataSync 在源位置验证的逻辑字节数。 单位：字节
BytesWritten	增强、基本	DataSync 写入目标位置的逻辑字节数。

CloudWatch 指标	任务模式支持	描述
		单位：字节
FilesDeleted	增强、基本	DataSync 在目标位置删除的文件、对象和目录的数量。如果您未将任务配置为在目标位置中删除不存在于源位置的数据，则该值始终为 0。 单位：个
FilesListedSource	增强	DataSync 在您的源位置找到的对象的数量。 单位：个
FilesPrepared	增强	DataSync 在比较源位置和目标位置后将尝试传输的对象数量。有关更多信息，请参阅 如何 DataSync 准备数据传输。 如果您将任务配置为传输所有数据，则此指标不适用。在这种情况下，DataSync 将源中的所有内容复制到目标，而不比较位置之间的差异。 单位：个
FilesPreparedDestination	基本	DataSync 在目标位置准备的文件、对象和目录的数量。 单位：个
FilesPreparedSource	基本	DataSync 在源位置准备的文件、对象和目录的数量。 单位：个
FilesSkipped	基本	DataSync 在传输过程中跳过的文件、对象和目录的数量。 单位：个

CloudWatch 指标	任务模式支持	描述
FilesTransferred	增强、基本	DataSync 通过网络传输的文件、对象和目录的数量。在任务执行期间，当从源读取内容并通过网络发送内容时，该值会定期更新。 **Note** 如果 DataSync 传输内容失败，在 DescribeTaskExecution 响应中，此值可以小于 EstimatedFilesToTransfer 。在某些情况下，此值也可能大于 EstimatedFilesToTransfer 。此指标特定于某些位置类型实施，因此请不要将其用作传输内容的准确指示，也不要用于监控任务执行情况。 单位：个
FilesVerified	增强	DataSync 在传输过程中验证的对象的数量。 单位：个
FilesVerifiedDestination	基本	DataSync 在目标位置验证的文件、对象和目录的数量。 单位：个
FilesVerifiedSource	基本	DataSync 在源位置验证的文件、对象和目录的数量。 单位：个

使用任务报告监控您的数据传输

任务报告可提供详细信息，说明在任务执行期间， AWS DataSync 尝试传输、跳过、验证和删除的内容。有关更多信息，请参阅 如何 DataSync 传输文件、对象和目录。

任务报告以 JSON 格式生成。您可以自定义报告的详细程度：

- 仅摘要任务报告可为您提供有关任务执行的必要详细信息，例如传输了多少文件以及是否 DataSync 可以验证这些文件的数据完整性。
- 标准任务报告包括摘要和详细报告，其中列出了 DataSync尝试传输、跳过、验证和删除的每个文件、对象或文件夹。使用标准任务报告，您还可以指定报告级别，以仅显示任务执行的错误，或其成功和错误。

使用案例

下面是任务报告可以帮助您对数据传输进行监控和审计的一些情境：

- 迁移数百万个文件时，请快速识别传输 DataSync 有问题的文件。
- 验证文件的 chain-of-custody进程。

仅摘要任务报告

仅包含任务执行摘要的报告包括以下详细信息：

- AWS 账户 运行任务执行的
- 源和目标位置
- 跳过、传输、验证和删除的文件、对象和文件夹总数
- 传输的总字节（逻辑字节和物理字节）
- 任务执行是否完成、取消或遇到错误
- 开始和结束时间（包括传输的总时间）
- 任务的设置（例如带宽限制、数据完整性验证和其他 DataSync 传输选项）

标准任务报告

标准任务报告包括任务执行摘要，以及有关 DataSync 尝试传输、跳过、验证和删除的内容的详细报告。

主题

- 报告级别
- 传输报告
- 跳过报告

- 验证报告
- 删除报告

报告级别

使用标准任务报告，您可以选择以下报告级别之一：

- 仅错误
- 成功与错误（本质上是任务执行过程中所发生一切情况的列表）

例如，您可能想查看哪些文件在传输过程中成功 DataSync 跳过，哪些文件没有成功跳过。成功 DataSync 跳过的文件可能是您故意 DataSync 要排除的文件，因为它们已经存在于您的目标位置。但是，例如，跳过的错误可能表明该用户 DataSync 没有读取文件的正确权限。

传输报告

任务执行期间 DataSync 尝试传输的文件、对象和目录的列表。传输报告包括以下详细信息：

- 传输数据的路径
- 传输数据的内容（内容、元数据或两者兼而有之）
- 元数据，包括数据类型、内容大小（仅限对象和文件）等
- 项目被转移的时间
- 对象版本（当目标是启用了版本控制的 Amazon S3 存储桶时）
- 目标中的某些内容是否被覆盖
- 项目是否传输成功

> **Note**
>
> 在 S3 存储桶之间移动数据时，您在源位置中指定的前缀可能会显示在您的报告（或 Amazon CloudWatch 日志中）中，即使该前缀不作为目标位置中的对象存在。（在 DataSync 控制台中，您可能还会注意到此前缀显示为已跳过或已验证的数据。）

跳过报告

在您的源位置 DataSync 找到但未尝试传输的文件、对象和目录的列表。 DataSync 跳过数据的原因可能取决于多种因素，例如如何配置任务和存储系统权限。下面是一些示例：

- 一个文件存在于源位置和目标位置。自上次执行任务以来，源位置的文件未被修改。由于您只传输已更改的数据，因此下次运行任务时 DataSync 不会传输该文件。
- 两个位置都存在的某个对象在源位置发生变化。运行任务时，会 DataSync 跳过目标中的此对象，因为您的任务不会覆盖目标中的数据。
- DataSync 跳过源代码中使用存档存储类且未恢复的对象。您必须恢复已存档的对象 DataSync 才能读取它。
- DataSync 跳过源位置中的文件、对象或目录，因为它无法读取它。如果发生这种情况但不是预料之中的，请检查存储空间的访问权限，并确保该权限 DataSync 可以读取跳过的内容。

跳过报告包括以下详细信息：

- 已跳过数据的路径
- 跳过某件项目的时间
- 被跳过的原因
- 项目是否成功跳过

> **Note**
>
> 如果跳过报告包含成功和错误，您将任务配置为只传输已更改的数据，并且源数据已存在于目标位置，那么跳过报告可能会很大。

验证报告

在任务执行期间 DataSync 试图验证完整性的文件、对象和目录的列表。已验证数据的报告包括以下详细信息：

- 已验证数据的路径
- 项目得到验证的时间
- 验证错误的原因（如果有）
- 源校验和目标 SHA256 校验和（仅限文件）

- 项目是否成功验证

请注意有关验证报告的以下注意事项：

- 当您将任务配置为仅验证传输的数据时，在某些情况下 DataSync 不验证目录或无法传输的文件或对象。无论哪种情况，都 DataSync 不会在此报告中包含未经验证的数据。
- 如果使用增强模式，则在传输大型对象时，验证花费的时间可能会比平时更长。

删除报告

任务执行期间删除的文件、目录和对象的列表。 DataSync 仅当您将任务配置为删除目标位置中不在源中的数据时，才会生成此报告。已删除数据的报告包括以下详细信息：

- 已删除数据的路径
- 项目是否成功删除
- 删除项目的时间

示例任务报告

任务报告的详细程度由您决定。下面是一些采用以下配置的传输数据报告示例：

- 报告类型：标准
- 报告级别：成功和错误

> **Note**
>
> 报告使用 ISO-8601 标准作为时间戳格式。时间采用 UTC 格式，以纳秒为单位。这种行为不同于其他一些任务报告指标的测量方式。例如，任务执行细节（例如 `TransferDuration` 和 `VerifyDuration` ）以毫秒为单位进行测量。

增强模式任务报告使用的架构与基本模式任务报告略有不同。根据所用任务模式，以下示例有助于您了解相应报告将呈现哪些内容。

具有成功状态的传输数据报告示例

以下报告显示名为 `object1.txt` 的对象的传输取得成功。

Enhanced mode


```
{
    "TaskExecutionId": "exec-abcdefgh12345678",
    "Transferred": [{
        "RelativePath": "object1.txt",
        "SourceMetadata": {
            "Type": "Object",
            "ContentSize": 6,
            "LastModified": "2024-10-04T14:40:55Z",
            "SystemMetadata": {
                "ContentType": "binary/octet-stream",
                "ETag": "\"9b2d7e1f8054c3a2041905d0378e6f14\"",
                "ServerSideEncryption": "AES256"
            },
            "UserMetadata": {},
            "Tags": []
        },
        "Overwrite": "False",
        "DstS3VersionId": "jtqRtX3jN4J2G8k0sFSGYK1f35KqpAVP",
        "TransferTimestamp": "2024-10-04T14:48:39.748862183Z",
        "TransferType": "CONTENT_AND_METADATA",
        "TransferStatus": "SUCCESS"
    }]
}
```


Basic mode


```
{
    "TaskExecutionId": "exec-abcdefgh12345678",
    "Transferred": [{
        "RelativePath": "/object1.txt",
        "SrcMetadata": {
            "Type": "Regular",
            "ContentSize": 6,
            "Mtime": "2022-01-07T16:59:26.136114671Z",
            "Atime": "2022-01-07T16:59:26.136114671Z",
            "Uid": 0,
            "Gid": 0,
            "Mode": "0644"
        },
        "Overwrite": "False",
        "DstS3VersionId": "jtqRtX3jN4J2G8k0sFSGYK1f35KqpAVP",
```

```
        "TransferTimestamp": "2022-01-07T16:59:45.747270957Z",
        "TransferType": "CONTENT_AND_METADATA",
        "TransferStatus": "SUCCESS"
    }]
}
```

具有错误状态的传输数据报告示例

以下报告提供了何时 DataSync 无法传输名为的对象的示例`object1.txt`。

Enhanced mode

此报告显示`object1.txt`由于 AWS KMS 权限问题而 DataSync 无法访问名为的对象。（如果您遇到此类错误，请参阅 使用服务器端加密访问 S3 存储桶。）

```
{
    "TaskExecutionId": "exec-abcdefgh12345678",
    "Transferred": [{
        "RelativePath": "object1.txt",
        "SourceMetadata": {
            "Type": "Object",
            "ContentSize": 6,
            "LastModified": "2022-10-07T20:48:32Z",
            "SystemMetadata": {
                "ContentType": "binary/octet-stream",
                "ETag": "\"3a7c0b2f1d9e5c4a6f8b2e0d1c9f7a3b2\"",
                "ServerSideEncryption": "AES256"
            },
            "UserMetadata": {},
            "Tags": []
        },
        "Overwrite": "False",
        "TransferTimestamp": "2022-10-09T16:05:11.134040717Z",
        "TransferType": "CONTENT_AND_METADATA",
        "TransferStatus": "FAILED",
        "ErrorCode": "AccessDenied",
        "ErrorDetail": "User: arn:aws:sts::111222333444:assumed-role/
AWSDataSyncS3Bucket/AwsSync-loc-0b3017fc4ba4a2d8d is not authorized
 to perform: kms:GenerateDataKey on resource: arn:aws:kms:us-
east-1:111222333444:key/1111aaaa-22bb-33cc-44d-5555eeee6666 because no identity-
based policy allows the kms:GenerateDataKey action"
    }]
}
```

Basic mode

此报告显示，由于 S3 存储桶权限问题，名为 `object1.txt` 的对象未传输。（如果您遇到此类错误，请参阅 提供 DataSync 对 S3 存储桶的访问权限。）

```
{
    "TaskExecutionId": "exec-abcdefgh12345678",
    "Transferred": [{
        "RelativePath": "/object1.txt",
        "SrcMetadata": {
            "Type": "Regular",
            "ContentSize": 6,
            "Mtime": "2022-01-07T16:59:26.136114671Z",
            "Atime": "2022-01-07T16:59:26.136114671Z",
            "Uid": 0,
            "Gid": 0,
            "Mode": "0644"
        },
        "Overwrite": "False",
        "DstS3VersionId": "jtqRtX3jN4J2G8k0sFSGYK1f35KqpAVP",
        "TransferTimestamp": "2022-01-07T16:59:45.747270957Z",
        "TransferType": "CONTENT_AND_METADATA",
        "TransferStatus": "FAILED",
        "FailureReason": "S3 Get Object Failed",
        "FailureCode": 40974
    }]
}
```

限制

- 单个任务报告大小不能超过 5 MB。如果您要复制大量文件，您的任务报告可能会被拆分为多个报告。
- 在某些情况下，创建任务报告会影响数据传输的性能。例如，当网络连接延迟较高、传输的文件较小，或只复制元数据更改时，您可能会注意到这种情况。

创建 DataSync 任务报告

AWS DataSync 任务报告只能是任务执行的摘要，也可以是关于 DataSync 尝试传输、跳过、验证和删除的内容的一组详细报告。

先决条件

在能够创建任务报告之前，您必须先执行以下操作。

主题

- 为您的任务报告创建 S3 存储桶
- 允许将任务报告上传 DataSync 到您的 S3 存储桶

为您的任务报告创建 S3 存储桶

如果您还没有 S3 存储桶，请创建一个 DataSync 可以在其中上传任务报告的 S3 存储桶。报告存储在 S3 Standard 存储类中。

我们建议在存储桶执行以下操作：

- 如果您计划将数据传输到 S3 存储桶，当您禁用保留已删除文件选项时，不要使用相同的存储桶上传任务报告。否则，每次执行任务时 DataSync 都会删除之前的所有任务报告，因为这些报告不存在于您的来源位置。
- 为避免复杂的访问权限设置，请确保您的任务报告存储桶 AWS 账户 与您的 DataSync 转移任务位于同一区域中。

允许将任务报告上传 DataSync 到您的 S3 存储桶

您必须配置允许将任务报告上传 DataSync 到 S3 存储桶的 AWS Identity and Access Management (IAM) 角色。

在 DataSync 控制台中，您可以创建一个 IAM 角色，该角色在大多数情况下会自动包含将任务报告上传到存储桶的权限。请记住，从最低权限的角度来看，这个自动生成的角色可能无法满足您的需求。如果您的存储桶使用客户托管 AWS Key Management Service (AWS KMS) 密钥 (SSE-KMS) 加密，则此角色也将不起作用。在上述情况下，您可以手动创建角色，前提是该角色至少要完成以下操作：

- 防止角色受信任实体中跨服务的代理混淆问题。

  以下完整示例显示了如何使用`aws:SourceArn`和全`aws:SourceAccount`局条件上下文键来防止混淆副手问题 DataSync。

  JSON

```
{
```

```
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "datasync.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                "aws:SourceAccount": "123456789012"
                },
                "ArnLike": {
                "aws:SourceArn": "arn:aws:datasync:us-east-1:123456789012:*"
                }
            }
        }
    ]
}
```

- 允许将任务报告上传 DataSync 到您的 S3 存储桶。

  以下示例通过仅包含存储桶中特定前缀 (`reports/`) 的 `s3:PutObject` 操作来实现此目的。

  JSON

```
{
        "Version":"2012-10-17",
        "Statement": [{
            "Action": [
                "s3:PutObject"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:s3:::your-task-reports-bucket/reports/*"
        }]
}
```

- 如果您的 S3 存储桶使用客户托管的 SSE-KMS 密钥进行加密，则该密钥的策略必须包含 DataSync 用于访问该存储桶的 IAM 角色。

  有关更多信息，请参阅 使用服务器端加密访问 S3 存储桶。

创建仅限摘要的任务报告

只有在创建任务、开始任务或更新 DataSync任务时，您才能配置包含摘要的任务报告。

以下步骤说明如何在创建任务时配置仅摘要任务报告。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 向下滚动到任务报告部分。对于报告类型，选择仅摘要。
5. 对于用于报告的 S3 存储桶，请选择 DataSync 要在其中上传任务报告的 S3 存储桶。

   > **Tip**
   >
   > 如果您计划将数据传输到 S3 存储桶，当您禁用保留已删除文件选项时，不要使用相同的存储桶上传任务报告。否则，每次执行任务时 DataSync 都会删除之前的所有任务报告，因为这些报告不存在于您的来源位置。

6. 在“文件夹”中，输入将报告 DataSync 上传到 S3 存储桶时用于任务报告的前缀（例如，**`reports/`**）。

   确保前缀末尾包含适当的分隔符。分隔符通常使用正斜杠 (/)。有关更多信息，请参阅Amazon S3用户指南中的使用前缀组织对象。
7. 对于 IAM 角色，执行以下操作之一：

   - 选择“自动生成”可 DataSync自动创建具有访问 S3 存储桶所需权限的 IAM 角色。

     如果 DataSync 之前为此 S3 存储桶创建了 IAM 角色，则默认情况下会选择该角色。
   - 选择您创建的自定义 IAM 角色。

     在某些情况下，您可能需要自己创建角色。有关更多信息，请参阅 允许将任务报告上传 DataSync 到您的 S3 存储桶。

> **Important**
>
> 如果您的 S3 存储桶使用客户托管的 SSE-KMS 密钥进行加密，则该密钥的策略必须包含 DataSync 用于访问该存储桶的 IAM 角色。
> 有关更多信息，请参阅 使用服务器端加密访问 S3 存储桶。

8. 完成任务的创建，然后启动任务以开始传输数据。

传输完成后，您可以查看任务报告。

使用 AWS CLI

1. 复制以下 `create-task` AWS Command Line Interface (AWS CLI) 命令：

```
aws datasync create-task \
  --source-location-arn arn:aws:datasync:us-east-1:123456789012:location/
loc-12345678abcdefgh \
  --destination-location-arn arn:aws:datasync:us-east-1:123456789012:location/
loc-abcdefgh12345678 \
  --task-report-config '{
    "Destination":{
      "S3":{
        "Subdirectory":"reports/",
        "S3BucketArn":"arn:aws:s3:::your-task-reports-bucket",
        "BucketAccessRoleArn":"arn:aws:iam::123456789012:role/bucket-iam-role"
        }
    },
    "OutputType":"SUMMARY_ONLY"
  }'
```

2. 对于 `--source-location-arn`参数，指定传输中源位置的 Amazon 资源名称（ARN）。*us-east-1*替换为相应的 AWS 账户 编号 AWS 区域，*123456789012*替换为相应的来源位置 ID，然后*12345678abcdefgh*替换为相应的来源位置 ID。
3. 对于 `--destination-location-arn` 参数，请指定传输中目标位置的 ARN。*us-east-1*替换为相应的数字 AWS 区域，*123456789012*替换为相应的 AWS 账户 数字，然后*abcdefgh12345678*替换为相应的目标位置 ID。
4. 对于 `--task-report-config` 参数，请执行以下操作：

   - `Subdirectory`— *reports/* 替换为您的 S3 存储桶中 DataSync 要上传任务报告的前缀。

确保前缀末尾包含适当的分隔符。分隔符通常使用正斜杠 (/)。有关更多信息，请参阅Amazon S3 用户指南中的使用前缀组织对象。

- `S3BucketArn` — 指定您想要用于上传任务报告的 S3 存储桶的 ARN。

> **Tip**
>
> 如果您计划将数据传输到 S3 存储桶，当您禁用保留已删除文件选项时，不要使用相同的存储桶上传任务报告。否则，每次执行任务时 DataSync 都会删除之前的所有任务报告，因为这些报告不存在于您的来源位置。

- `BucketAccessRoleArn`— 指定允许将任务报告上传 DataSync 到您的 S3 存储桶的 IAM 角色。

  有关更多信息，请参阅 允许将任务报告上传 DataSync 到您的 S3 存储桶。

> **Important**
>
> 如果您的 S3 存储桶使用客户托管的 SSE-KMS 密钥进行加密，则该密钥的策略必须包含 DataSync 用于访问该存储桶的 IAM 角色。
> 有关更多信息，请参阅 使用服务器端加密访问 S3 存储桶。

- `OutputType` — 指定 `SUMMARY_ONLY`。

  有关更多信息，请参阅 仅摘要任务报告。

5. 运行 `create-task` 命令来创建您的任务。

   您会收到如下所示的响应，其中显示了您创建的任务的 ARN。您需要此 ARN 才能运行 `start-task-execution` 命令。

```
{
    "TaskArn": "arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh"
}
```

6. 复制以下 `start-task-execution` 命令。

```
aws datasync-task-report start-task-execution \
  --task-arn arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh
```

7. 对于 `--task-arn` 参数，请指定您正在启动的任务的 ARN。使用您在运行 `create-task` 命令时收到的 ARN。
8. 运行 `start-task-execution` 命令。

传输完成后，您可以查看任务报告。

创建标准任务报告

您可以在创建任务、启动任务或更新 DataSync 任务时配置标准任务报告。

以下步骤说明在创建任务时如何配置标准任务报告。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 向下滚动到任务报告部分。对于报告类型，选择标准报告。
5. 对于报告级别，请选择下列选项之一：

   - 仅限错误-您的任务报告仅包含 DataSync 尝试传输、跳过、验证和删除的内容的问题。
   - 成功和错误-您的任务报告包括 DataSync 成功传输、跳过、验证和删除的内容以及未成功传输的内容。
   - 自定义：允许您选择仅查看错误或仅查看任务报告特定方面的成功和错误。

     例如，您可以为已传输的文件列表选择成功和错误，但为报告的其余部分选择仅错误。
6. 如果您要传输到使用对象版本控制的 S3 存储桶，并且希望您的报告中包含每个已传输对象的新版本，请选中包含 Amazon S3 对象版本。
7. 对于用于报告的 S3 存储桶，请选择 DataSync 要在其中上传任务报告的 S3 存储桶。

> **Tip**
>
> 如果您计划将数据传输到 S3 存储桶，当您禁用保留已删除文件选项时，不要使用相同的存储桶上传任务报告。否则，每次执行任务时 DataSync 都会删除之前的所有任务报告，因为这些报告不存在于您的来源位置。

8. 在“文件夹”中，输入将报告 DataSync 上传到 S3 存储桶时用于任务报告的前缀（例如，**`reports/`**）。确保前缀末尾包含适当的分隔符。分隔符通常使用正斜杠 (/)。有关更多信息，请参阅Amazon S3 用户指南中的使用前缀组织对象。
9. 对于 IAM 角色，执行以下操作之一：
   - 选择“自动生成”可 DataSync自动创建具有访问 S3 存储桶所需权限的 IAM 角色。

     如果 DataSync 之前为此 S3 存储桶创建了 IAM 角色，则默认情况下会选择该角色。
   - 选择您创建的自定义 IAM 角色。

     在某些情况下，您可能需要自己创建角色。有关更多信息，请参阅 允许将任务报告上传 DataSync 到您的 S3 存储桶。

     > **Important**
     >
     > 如果您的 S3 存储桶使用客户托管的 SSE-KMS 密钥进行加密，则该密钥的策略必须包含 DataSync 用于访问该存储桶的 IAM 角色。
     > 有关更多信息，请参阅 使用服务器端加密访问 S3 存储桶。
10. 完成任务的创建并启动任务以开始传输数据。

传输完成后，您可以查看任务报告。

使用 AWS CLI

1. 复制以下 `create-task` 命令：

```
aws datasync create-task \
  --source-location-arn arn:aws:datasync:us-east-1:123456789012:location/
loc-12345678abcdefgh \
  --destination-location-arn arn:aws:datasync:us-east-1:123456789012:location/
loc-abcdefgh12345678 \
  --task-report-config '{
    "Destination":{
      "S3":{
        "Subdirectory":"reports/",
        "S3BucketArn":"arn:aws:s3:::your-task-reports-bucket",
        "BucketAccessRoleArn":"arn:aws:iam::123456789012:role/bucket-iam-role"
        }
    },
```

```
    "OutputType":"STANDARD",
    "ReportLevel":"level-of-detail",
    "ObjectVersionIds":"include-or-not"
  }'
```

2. 对于（`--source-location-arn` 参数），指定传输中源位置的 ARN。*us-east-1*替换为相应的 AWS 账户 编号 AWS 区域，*123456789012*替换为相应的来源位置 ID，然后*12345678abcdefgh*替换为相应的来源位置 ID。
3. 对于 `--destination-location-arn` 参数，请指定传输中目标位置的 ARN。*us-east-1*替换为相应的数字 AWS 区域，*123456789012*替换为相应的 AWS 账户 数字，然后*abcdefgh12345678*替换为相应的目标位置 ID。
4. 对于 `--task-report-config` 参数，请执行以下操作：

   - `Subdirectory`— *reports/* 替换为您的 S3 存储桶中 DataSync 要上传任务报告的前缀。确保前缀末尾包含适当的分隔符。分隔符通常使用正斜杠 (/)。有关更多信息，请参阅Amazon S3 用户指南中的使用前缀组织对象。
   - `S3BucketArn` — 指定您想要用于上传任务报告的 S3 存储桶的 ARN。

     > **Tip**
     >
     > 如果您计划将数据传输到 S3 存储桶，当您禁用保留已删除文件选项时，不要使用相同的存储桶上传任务报告。否则，每次执行任务时 DataSync 都会删除之前的所有任务报告，因为这些报告不存在于您的来源位置。

   - `BucketAccessRoleArn`— 指定允许将任务报告上传 DataSync 到您的 S3 存储桶的 IAM 角色。

     有关更多信息，请参阅 允许将任务报告上传 DataSync 到您的 S3 存储桶。

     > **Important**
     >
     > 如果您的 S3 存储桶使用客户托管的 SSE-KMS 密钥进行加密，则该密钥的策略必须包含 DataSync 用于访问该存储桶的 IAM 角色。
     > 有关更多信息，请参阅 使用服务器端加密访问 S3 存储桶。

   - `OutputType`：指定 `STANDARD` 报告。

     有关详细信息，请参阅标准任务报告任务报告类型。

- （可选）ReportLevel：指定您想要在您的报告中包含 ERRORS_ONLY（默认）还是 SUCCESSES_AND_ERRORS。
- （可选）ObjectVersionIds： 如果您要传输到使用对象版本控制的 S3 存储桶，且您不希望再报告中包含每个已传输对象的新版本，请指定 NONE。

  默认情况下，此选项设置为 INCLUDE。
- （可选）Overrides：自定义您的报告的特定 ReportLevel 方面。

  例如，您可能想查看SUCCESSES_AND_ERRORS目标位置中 DataSync删除的内容的列表，但您想要ERRORS_ONLY查看其他所有内容。在此示例中，您可以在 --task-report-config 参数中添加以下 Overrides 选项：

  ```
  "Overrides":{
    "Deleted":{
      "ReportLevel":"SUCCESSES_AND_ERRORS"
    }
  }
  ```

  如果您不使用 Overrides，则整个报告将使用您指定的 ReportLevel。

5. 运行 create-task 命令来创建您的任务。

   您会收到如下所示的响应，其中显示了您创建的任务的 ARN。您需要此 ARN 才能运行 start-task-execution 命令。

   ```
   {
       "TaskArn": "arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh"
   }
   ```

6. 复制以下 start-task-execution 命令。

   ```
   aws datasync-task-report start-task-execution \
     --task-arn arn:aws:datasync:us-east-1:123456789012:task/task-12345678abcdefgh
   ```

7. 对于 --task-arn 参数，请指定您正在运行的任务的 ARN。使用您在运行 create-task 命令时收到的 ARN。
8. 运行 start-task-execution 命令。

传输完成后，您可以查看任务报告。

查看您的 DataSync 任务报告

DataSync 为每个任务执行创建任务报告。执行完成后，您可以在 S3 存储桶中找到相关任务报告。任务报告按前缀进行组织，前缀包括您的任务及其执行 IDs 情况。

要帮助在您的 S3 存储桶中查找任务报告，请使用以下示例：

- 仅摘要任务报告 — *reports-prefix*/Summary-Reports/*task-id-folder*/*task-execution-id-folder*
- 标准任务报告 — *reports-prefix*/Detailed-Reports/*task-id-folder*/*task-execution-id-folder*

由于任务报告采用 JSON 格式，因此您可以通过多种方式查看报告：

- 使用 Amazon S3 Select 查看报告。
- 使用 Amazon Athena 和 Amazon Quick 等 AWS AWS Glue服务可视化报告。有关任务报告可视化的更多信息，请参阅 AWS 存储博客文章。

使用 Amazon CloudWatch 日志监控数据传输

您可以使用 CloudWatch 日志监控您的 AWS DataSync 传输。建议将任务配置为至少记录基本信息（如传输错误）。

DataSync 允许将日志上传到 CloudWatch 日志组

要为您的 DataSync任务配置日志记录，您需要一个 CloudWatch DataSync 有权向其发送日志的日志组。您可以通过 AWS Identity and Access Management (IAM) 角色设置此访问权限。具体的工作原理取决于任务模式。

Enhanced mode

在增强模式下， DataSync 自动将任务日志发送到名为的日志组/aws/datasync。如果您的日志组不存在 AWS 区域，则在 DataSync 创建任务时使用 IAM 服务相关角色代表您创建日志组。

Basic mode

有几种方法可以使用 “基本” 模式为 DataSync 任务设置 CloudWatch 日志组。在控制台中，您可以自动创建一个 IAM 角色，在大多数情况下，该角色包含上传日志 DataSync 所需的权限。请记住，从最低权限的角度来看，这个自动生成的角色可能无法满足您的需求。

如果您想使用现有的 CloudWatch 日志组或以编程方式创建任务，则必须自己创建 IAM 角色。

以下示例是授予这些权限的 IAM 策略。

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "DataSyncLogsToCloudWatchLogs",
            "Effect": "Allow",
            "Action": [
                "logs:PutLogEvents",
                "logs:CreateLogStream"
            ],
            "Principal": {
                "Service": "datasync.amazonaws.com"
            },
            "Condition": {
                "ArnLike": {
                    "aws:SourceArn": [
                    "arn:aws:datasync:us-east-1:444455556666:task/*"
                    ]
                },
                "StringEquals": {
                "aws:SourceAccount": "444455556666"
                }
            },
            "Resource": "arn:aws:logs:us-east-1:444455556666:log-group:*:*"
        }
    ]
}
```

该策略使用Condition语句来帮助确保只有来自指定账户的 DataSync 任务才能访问指定的 CloudWatch 日志组。我们建议在这些 Condition 语句中使用 aws:SourceArn 和 aws:SourceAccount 全局条件上下文键来防范混淆代理人问题。有关更多信息，请参阅 防止跨服务混淆代理。

要指定一个或多个任务，请*region*替换为任务 AWS 区域 所在的区域代码（例如us-west-2），然后*account-id*替换为包含任务的账户的 AWS 账户 ID。 DataSync 要指定 CloudWatch日

志组，请替换相同的值。您也可以修改该 Resource 语句以将特定日志组作为目标。有关使用 SourceArn 和 SourceAccount 的更多信息，请参阅《IAM 用户指南》中的全局条件上下文键。

要应用此策略，请将此策略语句保存到本地计算机上的文件中。然后运行以下 AWS CLI 命令以应用资源策略。要使用此示例命令，请将 *full-path-to-policy-file* 替换为包含您的策略语句的文件路径。

```
aws logs put-resource-policy --policy-name trust-datasync --policy-document
 file://full-path-to-policy-file
```

> **Note**
>
> 使用与您激活 DataSync 代理相同的 AWS 账户 命令来运行此命令。 AWS 区域

有关更多信息，请参阅 Amazon CloudWatch 日志用户指南。

为您的 DataSync 任务配置日志记录

我们建议您至少为 DataSync任务配置一定级别的日志记录。

开始前的准备工作

DataSync 需要权限才能将日志上传到 CloudWatch 日志组。有关更多信息，请参阅 DataSync 允许将日志上传到 CloudWatch 日志组。

使用控制 DataSync 台

以下说明描述了如何在创建任务时配置 CloudWatch 日志记录。您还可以在编辑任务时配置日志记录。

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务，然后选择创建任务。
3. 配置您任务的源位置和目标位置。

   有关更多信息，请参阅 我可以在哪里传输我的数据 AWS DataSync？。
4. 在配置设置页面上，选择任务模式和其他任何选项。

   您可能对以下一些选项感兴趣：

- 使用清单或筛选条件指定要传输的数据。
- 配置如何处理文件元数据和验证数据完整性。

5. 对于日志级别，选择以下选项之一：

   - 记录传输错误等基本信息：发布只包含基本信息（如传输错误）的日志。
   - 记录所有传输的对象和文件-发布所有 DataSync传输和执行数据完整性检查的文件或对象的日志。
   - 不要生成日志

6. 根据您用于创建或指定 CloudWatch 日志组的任务模式，执行以下任一操作：

   Enhanced mode

   选择 “创建任务” 时， DataSync会自动使用（或创建）名为的日志组/aws/datasync。

   Basic mode

   对于CloudWatch 日志组，通过执行以下任一操作来指定 DataSync 有权将日志上传到的日志组：

   - 选择 “自动生成” 可自动创建一个允许 DataSync向其上传日志的日志组。
   - 在当前的 AWS 区域中选择一个现有日志组。

     如果您选择现有日志组，请确保该日志组DataSync 有权将日志上传到该日志组。

7. 选择创建任务。

您已经准备好，可以开始任务。

使用 AWS CLI

1. 复制以下 `create-task` 命令：

```
aws datasync create-task \
  --source-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-id" \
  --destination-location-arn "arn:aws:datasync:us-east-1:account-id:location/location-id" \
  --task-mode "ENHANCED-or-BASIC" \
  --name "task-name" \
  --options '{"LogLevel": "log-level"}' \
```

```
  --cloudwatch-log-group-arn "arn:aws:logs:us-east-1:account-id:log-group:log-group-name:*"
```

2. 对于 `--source-location-arn`，指定源位置的 Amazon 资源名称（ARN）。
3. 对于 `--destination-location-arn`，指定目标位置的 ARN。

   如果您要跨 AWS 区域 账户转账，请确保 ARN 包含其他地区或账户 ID。
4. 对于 `--task-mode`，指定 ENHANCED 或 BASIC。
5. （推荐）对于 `--name`，请为任务指定一个您可以记住的名称。
6. 对于 `LogLevel`，指定下列选项之一：

   - BASIC：仅发布包含基本信息（如传输错误）的日志。
   - TRANSFER— 发布所有 DataSync 传输和执行数据完整性检查的文件或对象的日志。
   - NONE：不要生成日志。
7. 对于`--cloudwatch-log-group-arn`，指定 CloudWatch日志组的 ARN。

   > **Important**
   >
   > 如果 `--task-mode` 为 ENHANCED，您不需要指定此选项。有关更多信息，请参阅 DataSync 允许将日志上传到 CloudWatch 日志组。
8. 运行 `create-task` 命令。

   如果命令成功，您将收到一条响应，显示所创建任务的 ARN。例如：

```
{
    "TaskArn": "arn:aws:datasync:us-east-1:111222333444:task/task-08de6e6697796f026"
}
```

您已经准备好，可以开始任务。

使用 DataSync API

您可以通过将`CloudWatchLogGroupArn`参数与以下任一操作一起使用来配置任务的 CloudWatch 日志记录：

- CreateTask

- UpdateTask

查看 DataSync 任务日志

启动任务时，您可以使用 CloudWatch 控制台或 AWS CLI （以及其他选项）查看任务执行日志。有关更多信息，请参阅 Amazon CloudWatch 日志用户指南。

DataSync 为增强模式任务提供 JSON 结构的日志。基本模式任务具有非结构化的日志。以下示例显示与基本模式日志相比，增强模式日志的验证错误显示方式。

Enhanced mode log example

```
{
    "Action": "VERIFY",
    "Source": {
        "LocationId": "loc-abcdef01234567890",
        "RelativePath": "directory1/directory2/file1.txt"
    },
    "Destination": {
        "LocationId": "loc-05ab2fdc272204a5f",
        "RelativePath": "directory1/directory2/file1.txt",
        "Metadata": {
            "Type": "Object",
            "ContentSize": 66060288,
            "LastModified": "2024-10-03T20:46:58Z",
            "S3": {
                "SystemMetadata": {
                    "ContentType": "binary/octet-stream",
                    "ETag": "\"1234abcd5678efgh9012ijkl3456mnop\"",
                    "ServerSideEncryption": "AES256"
                },
                "UserMetadata": {
                    "file-mtime": "1602647222/222919600"
                },
                "Tags": {}
            }
        }
    },
    "ErrorCode": "FileNotAtSource",
    "ErrorDetail": "Verification failed due to file being present at the destination
 but not at the source"
}
```

Basic mode log example

```
[NOTICE] Verification failed > /directory1/directory2/file1.txt
[NOTICE] /directory1/directory2/file1.txt   dstMeta: type=R mode=0755 uid=65534
 gid=65534 size=8972938 atime=1728657659/0 mtime=1728657659/0 extAttrsHash=0
[NOTICE]   dstHash: f9c2cca900301d38b0930367d8d587153154af467da0fdcf1bebc0848ec72c0d
```

使用记录 AWS DataSync API 调用 AWS CloudTrail

AWS DataSync 与 AWS CloudTrail一项服务集成，该服务提供用户、角色或角色所执行操作 AWS 服务 的记录 DataSync。 CloudTrail 将所有 API 调用捕获 DataSync为事件。捕获的调用包括来自 DataSync 控制台的调用和对 DataSync API 操作的代码调用。

如果您创建了跟踪，则可以允许将 CloudTrail 事件持续传输到 Amazon S3 存储桶，包括的事件 AWS DataSync。如果您未配置跟踪，您仍然可以在 CloudTrail 控制台的 “事件历史记录” 中查看最新的事件。使用收集的信息 CloudTrail，您可以确定向哪个请求发出 AWS DataSync、发出请求的 IP 地址、谁发出了请求、何时发出请求以及其他详细信息。

要了解更多信息 CloudTrail，请参阅AWS CloudTrail 用户指南。

处理中的 DataSync 信息 CloudTrail

CloudTrail 在您创建账户 AWS 账户 时已在您的账户上启用。当活动发生在中时 AWS DataSync，该活动会与其他 CloudTrail 事件一起记录在 AWS 服务 事件历史记录中。您可以在中查看、搜索和下载最近发生的事件 AWS 账户。有关更多信息，请参阅使用事件历史查看 CloudTrail 事件。

对于 AWS 账户中的事件的持续记录（包括 AWS DataSync 的事件），请创建跟踪记录。跟踪允许 CloudTrail 将日志文件传输到 Amazon S3 存储桶。默认情况下，在控制台中创建跟踪记录时，此跟踪记录应用于所有 AWS 区域。该跟踪将所有事件记录 AWS 区域 在同一个 AWS 分区中，并将日志文件传送到您指定的 Amazon S3 存储桶。此外，您可以配置其他 AWS 服务 ，以进一步分析和处理 CloudTrail 日志中收集的事件数据。有关更多信息，请参阅下列内容：

- 创建跟踪记录概述
- CloudTrail 支持的服务和集成
- 配置 Amazon SNS 通知 CloudTrail
- 接收来自多个区域的 CloudTrail 日志文件和接收来自多个账户的 CloudTrail 日志文件

所有 DataSync 操作都由记录 CloudTrail。（有关更多信息，请参阅 DataSync API 参考。）

例如，对CreateAgentCreateTask、和ListLocations操作的调用会在 CloudTrail 日志文件中生成条目。

每个事件或日志条目都包含有关生成请求的人员信息。身份信息有助于您确定以下内容：

- 请求是使用根证书还是 AWS Identity and Access Management (IAM) 凭证发出。
- 请求是使用角色还是联合用户的临时安全凭证发出的。
- 请求是否由其他 AWS 服务发出。

有关更多信息，请参阅《AWS CloudTrail 用户指南》中的CloudTrailuserIdentity元素。

了解 DataSync 日志文件条目

跟踪是一种配置，允许将事件作为日志文件传输到您指定的 Amazon S3 存储桶。 CloudTrail 日志文件包含一个或多个日志条目。事件代表来自任何来源的单个请求，包括有关请求的操作、操作的日期和时间、请求参数等的信息。 CloudTrail 日志文件不是公共 API 调用的有序堆栈跟踪，因此它们不会按任何特定顺序出现。

以下示例显示了演示该CreateTask操作的 CloudTrail 日志条目。

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "IAMUser",
        "principalId": "1234567890abcdef0",
        "arn": "arn:aws:iam::123456789012:user/user1",
        "accountId": "123456789012",
        "accessKeyId": "access key",
        "userName": "user1",
        "sessionContext": {
            "attributes": {
                "mfaAuthenticated": "false",
                "creationDate": "2018-12-13T14:56:46Z"
            }
        },
        "invokedBy": "signin.amazonaws.com"
    },
    "eventTime": "2018-12-13T14:57:02Z",
    "eventSource": "datasync.amazonaws.com",
    "eventName": "CreateTask",
```

```
    "awsRegion": "ap-southeast-1",
    "sourceIPAddress": "192.0.2.1",
    "userAgent": "signin.amazonaws.com",
    "requestParameters": {
        "cloudWatchLogGroupArn": "arn:aws:logs:ap-southeast-1:123456789012:log-
group:MyLogGroup",
        "name": "MyTask-NTIzMzY1",
        "tags": [],
        "destinationLocationArn": "arn:aws:datasync:ap-
southeast-1:123456789012:location/loc-abcdef01234567890",
        "options": {
            "bytesPerSecond": -1,
            "verifyMode": "POINT_IN_TIME_CONSISTENT",
            "uid": "INT_VALUE",
            "posixPermissions": "PRESERVE",
            "mtime": "PRESERVE",
            "gid": "INT_VALUE",
            "preserveDevices": "NONE",
            "preserveDeletedFiles": "REMOVE",
            "atime": "BEST_EFFORT"
        },
        "sourceLocationArn": "arn:aws:datasync:ap-southeast-1:123456789012:location/
loc-021345abcdef6789"
    },
    "responseElements": {
        "taskArn": "arn:aws:datasync:ap-southeast-1:123456789012:task/
task-1234567890abcdef0"
    },
    "requestID": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
    "eventID": "a1b2c3d4-5678-90ab-cdef-EXAMPLE22222",
    "eventType": "AwsApiCall",
    "recipientAccountId": "123456789012"
}
```

使用 Amazon 监控事件 EventBridge

Amazon EventBridge 事件描述了 DataSync 资源的变化。您可以设置规则来匹配这些事件，并将这些事件路由到一个或多个目标函数或流。尽最大努力发布事件。

DataSync 转移事件

以下 EventBridge 事件可用于 DataSync 转移。

Agent state changes

Event (事件)	描述
在线	已正确配置代理并且可供使用。这是代理的正常运行状态。
离线	代理与该 DataSync 服务部门断开联系已有五分钟或更长时间。这可能是由于多种原因造成。有关更多信息，请参阅 如果我的代理处于离线状态，我该怎么办？。

Location state changes

Event (事件)	描述
正在添加	DataSync 正在添加地点。
可用	位置已创建并且可供使用。

Task state changes

Event (事件)	描述
Available	任务已创建，可以开始了。
运行	任务正在进行中，运行正常。
不可用	任务配置不正确，无法使用。当与任务关联的代理脱机时，您可能会看到此事件。
已排队	另一项任务正在运行并使用同一个代理。DataSync 按顺序运行任务（先入先出）。

Task execution state changes

Event (事件)	描述
排队	另一个任务正在运行并使用相同的 DataSync 代理。有关更多信息，请参阅 了解任务何时排队。

Agent state changes	
正在启动	DataSync 正在初始化任务执行。此状态通常很快就会结束，但也可能需要几分钟。
正在准备	DataSync 正在确定要传输哪些数据。 这一步可能只需要几分钟，也可能需要几个小时，具体取决于两个位置的文件、对象或目录的数量，以及任务配置方式。准备可能也不适用于您的任务。有关更多信息，请参阅 如何 DataSync 准备数据传输。
正在传输	DataSync 正在执行实际的数据传输。
正在验证	DataSync 正在传输结束时执行数据完整性检查。
成功	任务执行成功。
正在取消	正在取消任务的执行。
错误	任务执行失败。

AWS DataSync 使用手动工具进行监控

您可以从控制台或命令行跟踪您的 AWS DataSync 传输。

使用 DataSync 控制台监控您的转账

您可以使用控制台监控 DataSync 传输情况，该控制台提供实时指标，例如传输的数据、数据和文件吞吐量以及数据压缩。

使用 DataSync 控制台监控您的转账

1. 开始 DataSync 任务后，选择 “查看执行详情”。
2. 查看有关您的传输的指标。

使用以下方法监控您的转账 AWS CLI

您可以使用 AWS Command Line Interface (AWS CLI) 监控您的 DataSync 传输。

复制以下 `describe-task-execution` 命令。要使用此示例命令，请将 *`user input placeholders`* 替换为您自己的信息。

```
aws datasync describe-task-execution \
  --task-execution-arn 'arn:aws:datasync:region:account-id:task/task-id/execution/task-execution-id'
```

此命令返回有关任务执行的信息，如下所示。

```
{
    "BytesCompressed": 3500,
    "BytesTransferred": 5000,
    "BytesWritten": 5000,
    "EstimatedBytesToTransfer": 5000,
    "EstimatedFilesToDelete": 10,
    "EstimatedFilesToTransfer": 100,
    "FilesDeleted": 10,
    "FilesSkipped": 0,
    "FilesTransferred": 100,
    "FilesVerified": 100,
    "Result": {
        "ErrorCode": "??????",
        "ErrorDetail": "??????",
        "PrepareDuration": 100,
        "PrepareStatus": "SUCCESS",
        "TransferDuration": 60,
        "TransferStatus": "AVAILABLE",
        "VerifyDuration": 30,
        "VerifyStatus": "SUCCESS"
    },
    "StartTime": 1532660733.39,
    "Status": "SUCCESS",
    "OverrideOptions": {
        "Atime": "BEST_EFFORT",
        "BytesPerSecond": "1000",
        "Gid": "NONE",
        "Mtime": "PRESERVE",
        "PosixPermissions": "PRESERVE",
```

```
        "PreserveDevices": "NONE",
        "PreserveDeletedFiles": "PRESERVE",
        "Uid": "NONE",
        "VerifyMode": "POINT_IN_TIME_CONSISTENT"
    },
    "TaskExecutionArn": "arn:aws:datasync:us-east-1:111222333444:task/task-
aaaabbbbccccddddf/execution/exec-1234abcd1234abcd1",
    "TaskReportConfig": {
        "Destination": {
            "S3": {
                "BucketAccessRoleArn": "arn:aws:iam::111222333444:role/my-datasync-
role",
                "S3BucketArn": "arn:aws:s3:::amzn-s3-demo-bucket/*",
                "Subdirectory": "reports"
            }
        },
        "ObjectVersionIds": "INCLUDE",
        "OutputType": "STANDARD",
        "Overrides": {
            "Deleted": {
                "ReportLevel": "ERRORS_ONLY"
            },
            "Skipped": {
                "ReportLevel": "SUCCESSES_AND_ERRORS"
            },
            "Transferred": {
                "ReportLevel": "ERRORS_ONLY"
            },
            "Verified": {
                "ReportLevel": "ERRORS_ONLY"
            }
        },
        "ReportLevel": "ERRORS_ONLY"
    }
}
```

- 如果任务执行成功，则状态的值会变为成功。有关响应元素含义的信息，请参阅 DescribeTaskExecution。
- 如果任务执行失败，则结果会发送可帮助您排查问题的错误代码。有关错误代码的信息，请参阅 TaskExecutionResultDetail。

使用 watch 实用程序监控传输

要从命令行实时监控任务执行的进度，您可以使用标准 watch 监视实用工具。任务执行持续时间值以毫秒为单位。

该watch实用程序无法识别 DataSync 别名。下面的示例显示了如何直接调用 CLI。要使用此示例命令，请将 *user input placeholders* 替换为您自己的信息。

```
# pass '-n 1' to update every second and '-d' to highlight differences
        $ watch -n 1 -d \ "aws datasync describe-task-execution --task-execution-arn
 'arn:aws:datasync:region:account-id:task/task-id/execution/task execution-id'"
```

管理 AWS DataSync 资源

了解如何管理您的 AWS DataSync 资源，例如代理、地点和任务。

管理您的 DataSync 代理

激活 DataSync 代理后，将为您 AWS 管理代理（包括软件更新）。了解详情

测试 DataSync 代理的连接和系统资源

虽然在 DataSync 代理部署和激活后对其进行 AWS 管理，但在某些情况下，您可能需要更改代理的设置或对问题进行故障排除。了解详情

更换您的 DataSync 代理

要替换 DataSync 代理，必须创建一个新代理并更新使用旧代理的所有位置。了解详情

清理 DataSync 资源

如果您 DataSync 用于测试或者只是不再需要其资源，请删除这些资源，这样您就无需为此付费。了解详情

重复使用 DataSync 代理的基础架构

从中删除代理资源后 DataSync，您仍然可以使用代理的虚拟机或 Amazon EC2 实例来激活新的代理。了解详情

管理您的 AWS DataSync 代理

激活 AWS DataSync 代理后，将为您 AWS 管理虚拟机 (VM) 设备。

代理软件更新

AWS 自动更新代理的软件，包括底层操作系统和相关 DataSync 软件包。

DataSync 仅在代理处于空闲状态时才对其进行更新。例如，您的代理在传输完成之前不会更新。

代理在更新后可能会短暂处于离线状态。例如，当 AWS 更新代理时，这种情况可能在代理激活后不久发生。

> **Important**
>
> - DataSync 自动定期修补代理以维护其安全性和稳定性。 DataSync 基本模式代理使用 Amazon Linux 2 作为其基本操作系统。 DataSync 增强模式代理使用亚马逊 Linux 2023 作为其基本操作系统。您可以在 Amazon Linux 安全中心查看检测到的常见漏洞和风险（CVE）问题的当前状态。如 Amazon Linux 安全中心所示，CVE 补丁将在发布日期后 30 天内自动应用。只要代理处于联机状态且未主动运行任务执行，就会进行修补。
> - DataSync 不支持使用云初始化指令手动更新 Amazon EC2 代理。如果以这种方式更新代理，则可能会 DataSync 遇到无法激活或使用代理的互操作性问题。

代理状态

下表描述了 DataSync 代理的状态。

座席状态	含义
在线	已正确配置代理并且可供使用。这是代理的正常运行状态。
离线	代理与该 DataSync 服务部门断开联系已有五分钟或更长时间。这可能是由于多种原因造成。有关更多信息，请参阅如果我的代理处于离线状态，我该怎么办？。

代理故障排除

在为您 AWS 管理 DataSync 代理时，在某些情况下，您可能需要再次直接使用代理。例如，如果代理离线或失去与本地存储系统的连接，您可以尝试在代理的本地控制台解决这些问题。

有关更多信息，请参阅对 DataSync 代理进行故障排除。

对代理进行维护

虽然在 AWS DataSync 代理部署和激活后对其进行 AWS 管理，但在某些情况下，您可能需要更改代理的设置或对问题进行故障排除。以下是一些示例，说明为什么要通过代理的本地控制台与其合作：

- 手动为代理分配 IP 地址。
- 检查代理的系统资源。

> **Important**
>
> 您无需使用代理的本地控制台即可获得标准 DataSync功能。

访问代理的本地控制台

访问本地控制台的方式取决于您使用的代理类型。

访问本地控制台（VMware ESXi、Linux KVM、Nutanix AHV 或微软 Hyper-V）

出于安全考虑，您无法远程连接到 DataSync 代理虚拟机 (VM) 的本地控制台。您必须从虚拟机管理程序管理界面访问本地控制台。

- 如果这是您首次登录本地控制台，请使用临时凭证登录。初始用户名为 **`admin`**，临时密码为 **`password`**。首次登录时必须更改密码。

  > **Note**
  >
  > 增强模式代理具有以下密码要求：
  >
  > - 必须至少包含 15 个字符
  > - 必须至少包含一个大写字符
  > - 必须包含至少一个小写字符
  > - 必须包含至少一个数字字符
  > - 必须包含至少一个特殊字符
  > - 密码更新时必须更改至少 50% 的字符
  > - 密码不能是字典中的单词

> **Note**
>
> 初始密码设置之后，可随时更改密码。在控制台主菜单上，输入命令提示符旁边的数字，然后运行 passwd 命令来更改密码。

访问本地控制台 (Amazon EC2)

要连接到 Amazon EC2 代理的本地控制台，必须使用 SSH。

开始之前：确保您的 EC2 实例的安全组允许通过 SSH 进行访问（TCP 端口 22）。

1. 打开一个终端，并复制以下 ssh 命令：

```
ssh -i /path/key-pair-name.pem instance-user-name@instance-public-ip-address
```

   - 对于*/path/key-pair-name*，请指定连接到您的实例所需的私有密钥的路径和文件名 (.pem)。
   - 对于 *instance-user-name*，请指定 admin。
   - 对于*instance-public-ip-address*，请指定您的实例的公有 IP 地址。
2. 运行 ssh 命令连接实例。

连接后，将显示代理本地控制台的主菜单。

配置代理的 DHCP、DNS 和 IP 设置

代理的默认网络配置是动态主机配置协议 (DHCP)。借助 DHCP，您的代理可获得自动分配的 IP 地址。在某些情况中，您可能需要手动将代理的 IP 地址分配为静态 IP 地址，如下所示。

1. 登录到代理的本地控制台。
2. 在 “AWS DataSync 激活-配置” 主菜单上**1**，输入开始配置您的网络。
3. 在 Network Configuration (网络配置) 菜单上，选择下列选项之一。

目的	请执行该操作
获取有关网络适配器的信息	

目的	请执行该操作

	输入 **1**。
	
	将显示一个适配器名称列表，并且系统会提示您输入一个适配器名称例如 **eth0**。如果您指定的适配器正在使用中，则显示该适配器的网络信息，如下例所示：
	
	`IP Preference: IPv4`
	`MAC address: 52:54:12:a4:f7:7d`
	
	`IPv4 address: 192.168.100.482`
	`Netmask: 255.255.255.0`
	`Gateway: 192.168.100.4`
	`DHCP enabled: Yes`
	
	`IPv6 address: abcd:4444:e5ee:fd0`
	`0::4daf`
	`Prefix length: 128`
	`Gateway: fe80::5021:ff:ff88:4acd`
	`DHCPV6 enabled: Yes`
	
	`DNS: abcd:4444:e5ee:fd00::1`
	`DNS: 192.168.100.4`
	
	配置静态 IP 地址（选项 3）或设置代理的默认路由适配器（选项 5）时，您可使用相同的适配器名称。

配置 DHCP	输入 **2**。 选择要使用的 IP 版本，然后将网络接口配置为使用 DHCP。

目的	请执行该操作
为代理配置静态 IP 地址	输入 **3**。 系统会提示您选择要使用的 IP 协议、IPv4 IPv6、或两者兼而有之。然后，系统会提示输入网络适配器名称以配置静态 IP 地址。 **Important** 如果您的代理已被激活，则必须将其关闭并从 DataSync 控制台重新启动，设置才能生效。
将代理的所有网络配置重置为 DHCP	输入 **4**。 选择要重置为 DHCP 的 IP 版本。所选 IP 版本的所有网络接口均设置为使用 DHCP。 **Important** 如果您的代理已经激活，则必须关闭代理并从 DataSync 控制台重新启动代理才能使设置生效。
设置代理的默认路由适配器	输入 **5**。 将显示可供网关使用的适配器，系统会提示您选择其中一个适配器例如 **`eth0`**。

目的	请执行该操作
编辑代理的域名系统 (DNS) 配置	输入 **6**。 这将显示主 DNS 和备用 DNS 服务器的可用适配器。系统将提示您提供新的 IP 地址。
查看代理的 DNS 配置	输入 **7**。 这将显示主 DNS 和备用 DNS 服务器的可用适配器。 **Note** 对于某些版本的 VMware 虚拟机管理程序，您可以在此菜单中编辑适配器配置。
查看路由表	输入 **8**。 选择 IP 版本（IPv4 IPv6、或两者）以查看代理的默认路由表。
查看代理用于数据传输的 IP 版本	输入 **9**。 代理用于数据传输的 IP 版本设置显示为 `IPv4`、`IPv6`、`IPv4 (auto)` 或 `IPv6 (auto)`。

目的	请执行该操作
编辑代理用于数据传输的 IP 协议	输入 **10**。 将显示用于数据传输的可用 IP 版本设置。可选择 IPv4、IPv6、IPv4 (auto) 或 IPv6 (auto)。有关代理用于数据传输的 IP 版本设置的更多信息，请参阅the section called “IPv6 支持”。

检查代理的系统资源

登录代理控制台时，会自动检查虚拟 CPU 内核、根卷大小和 RAM。如果有任何错误或警告，则会在控制台菜单显示屏上用横幅进行标记，提供有关这些错误或警告的详细信息。

如果主机启动时没有错误或警告，则菜单将显示白色文本。View System Resource Check（查看系统资源检查）选项将显示 (`0 Errors`)。

如果出现错误或警告，控制台菜单会在菜单顶部的横幅中分别以红色和黄色显示错误和警告的数量。例如 (`1 ERROR, 1 WARNING`)。

检查代理的系统资源

1. 登录到代理的本地控制台。
2. 在 “AWS DataSync 激活-配置” 主菜单上，输入**4**以查看系统资源检查的结果。

   控制台为每个资源显示 [OK]、[WARNING] 或 [FAIL] 消息，如下表中所述。

   对于 Amazon EC2 实例，系统资源检查会验证实例类型是否为推荐与一起 DataSync使用的实例之一。如果实例类型与该列表匹配，则单个结果将以绿色文本显示，如下所示。

   ```
   [ OK ] Instance Type Check
   ```

   如果 Amazon EC2 实例不在建议列表中，则系统资源检查将验证以下资源。

   - CPU 内核检查：至少需要四个内核。
   - 磁盘大小检查：至少需要 80 GB 的可用磁盘空间。
   - RAM 检查：

- 分配给实例的 32 GB RAM 用于处理多达 2000 万个文件、对象或目录的任务执行。
- 分配给实例的 64 GB RAM 用于处理超过 2000 万个文件、对象或目录的任务执行。

- CPU 标志检查：代理虚拟机 CPU 必须具有 SSSE3 或 SSE4 指令集标志。

如果 Amazon EC2 实例不在的推荐实例列表中 DataSync，但它有足够的资源，则系统资源检查的结果将显示四个结果，全部以绿色文本显示。

对于部署在 Hyper-V、基于 Linux 内核的虚拟机 (KVM) 和中的代理，也要验证相同的资源。VMware VMs

VMware 还会检查代理是否有支持的版本；不支持的版本会导致红色横幅错误。支持的版本包括 VMware 版本 6.5 和 6.7。

查看和管理代理的系统时间服务器配置

可查看和管理代理的系统时间服务器配置。

1. 登录到代理的本地控制台。
2. 在 “AWS DataSync 激活-配置” 主菜单上，输入 “系统时间管理” 选项（例如**5**用于 VMware 代理）。
3. 在系统时间管理菜单中，执行以下操作之一：

目的	请执行该操作
查看系统时间和服务状态	输入 **1**。 查看当前系统时间（UTC）、时间服务状态、活动时间服务器及同步状态。
同步系统时间	输入 **2**。 将显示立即同步时间服务器的提示。 在某些情况下，代理的时间可能会出现偏差。例如，可能会出现长时间的网络中断，虚拟机监控程序主机和代理无法获得时间更新，因此代理的

目的	请执行该操作
	时间与实际时间不同。当出现此类时间偏差时，操作（如快照）的声明时间与实际发生时间之间就会出现差异。
重启系统时间服务	输入 **3**。 将显示重启时间同步服务的提示。
管理时间服务器配置	输入 **4**。 查看和管理时间服务器设置。添加或删除时间服务器及服务器池，设置首选服务器以实现精确同步。

运行与代理维护相关的命令

在 DataSync 代理的本地控制台中，您可以执行一些维护任务并诊断代理的问题。

在代理的本地控制台中运行配置或诊断命令

1. 登录到代理的本地控制台。
2. 在 “AWS DataSync 激活-配置” 主菜单上，输入**5**（或 **6** VMware VM 的）命令提示符。
3. 使用以下命令与代理一起执行以下任务。

命令	说明
dig	查找有关主机的 DNS 信息。
diskclean	执行磁盘清理。
exit	返回控制台配置菜单。
h	显示可用命令的列表。
ifconfig	显示或配置网络接口。
ip	显示或配置路由、设备和隧道。

命令	说明
iptables	设置和维护 IPv4 数据包过滤和 NAT。
ip6tables	设置和维护 IPv6 数据包过滤和 NAT。
ncport	测试与特定网络 TCP 端口的连接。
nping	获取信息以排查网络故障问题。
passwd	更改用户密码。
save-iptables	永久保存 IPv4 表格防火墙规则。
save-ip6tables	永久保存 IPv6 表格防火墙规则。
save-routing-table	保存新添加的路由表条目。
sslcheck	验证 SSL 证书是否有效。
tcptraceroute	收集通往目标位置的 TCP 流量的 `traceroute` 输出。

4. 按照屏幕上的说明进行操作。

更换AWS DataSync代理

若要更换AWS DataSync代理，必须创建一个新代理，并更新使用旧代理的所有传输位置。

创建新代理

若要创建新的 DataSync 代理，请按照创建旧代理时相同的过程进行操作：

1. 在您的存储环境中 部署代理。
2. 选择代理用来与 AWS 通信的服务端点。
3. 配置网络，以便代理可以与存储和 AWS 进行通信。
4. 激活您的代理。
5. 激活后，记下代理的 Amazon 资源名称（ARN）。

   更新 DataSync 位置以使用新代理时，您需要此 ARN。

通过新代理更新您的位置

创建新代理后，您可以更新现有的 DataSync 位置，以使用此代理。大多数情况下，您还必须重新输入访问凭证才能更新位置。这是因为 DataSync 存储位置凭证的方式只有您的代理可以使用它们。

使用 DataSync 控制台

以下说明介绍如何使用 DataSync 控制台，通过新代理更新位置。

NFS

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择位置。
3. 选择要更新的位置，然后选择编辑。
4. 对于代理，选择您的新代理。

   如果您要为一个位置更换多个代理，则可以选择多个代理。
5. 选择 保存更改。

SMB

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择位置。
3. 选择要更新的位置，然后选择编辑。
4. 对于代理，选择您的新代理。

   如果您要为一个位置更换多个代理，则可以选择多个代理。
5. 对于密码，输入可挂载 SMB 文件服务器并有权访问传输中涉及文件和文件夹的用户的密码。
6. 选择 保存更改。

HDFS

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择位置。
3. 选择要更新的位置，然后选择编辑。
4. 对于代理，选择您的新代理。

如果您要为一个位置更换多个代理，则可以选择多个代理。

5. 如果你使用的是 Kerberos 身份验证，请上传您的Keytab 文件 和 Kerberos 配置文件。
6. 选择保存更改。

Object storage

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择位置。
3. 选择要更新的位置，然后选择编辑。
4. 对于代理，选择您的新代理。

   如果您要为一个位置更换多个代理，则可以选择多个代理。
5. 如果您的位置需要凭证，请输入允许 DataSync 访问对象存储桶的私有密钥。
6. 选择保存更改。

Azure Blob Storage

执行以下操作，以更新您的Microsoft Azure Blob Storage位置：

1. 打开 AWS DataSync 控制台，地址：https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择位置。
3. 选择要更新的位置，然后选择编辑。
4. 对于代理，选择您的新代理。

   如果您要为一个位置更换多个代理，则可以选择多个代理。
5. 对于 SAS 令牌，请输入允许 DataSync 访问您的 blob 存储空间的共享访问签名 (SAS) 令牌。
6. 选择 保存更改。

使用 AWS CLI

以下说明介绍如何使用AWS CLI，通过新代理更新位置。（您也可以通过使用 DataSync API 来执行此操作。）

NFS

1. 复制以下 update-location-nfs命令：

```
aws datasync update-location-nfs \
  --location-arn datasync-nfs-location-arn \
  --on-prem-config AgentArns=new-datasync-agent-arn
```

2. 对于--location-arn参数，请指定要更新的 NFS 位置 ARN。
3. 对于--on-prem-config参数AgentArns选项，请指定新代理的 ARN。

   如果您要为一个位置更换多个代理，则可以指定多个 ARN。
4. 运行update-location-nfs 命令以更新位置。

SMB

1. 复制以下update-location-smb 命令：

```
aws datasync update-location-smb \
  --location-arn datasync-smb-location-arn \
  --agent-arns new-datasync-agent-arn \
  --password smb-file-server-password
```

2. 对于--location-arn 参数，请指定要更新的 SMB 位置的 ARN。
3. 对于--agent-arns 参数，请指定新代理的 ARN。

   如果您要为一个位置更换多个代理，则可以指定多个 ARN。
4. 对于 --password 参数，指定可挂载 SMB 文件服务器并有权访问传输中涉及文件和文件夹的用户的密码。
5. 运行 update-location-smb命令以更新位置。

HDFS

1. 复制以下 update-location-hdfs命令：

```
aws datasync update-location-hdfs \
  --location-arn datasync-hdfs-location-arn \
  --agent-arns new-datasync-agent-arn \
  --kerberos-keytab keytab-file \
```

```
  --kerberos-krb5-conf krb5-conf-file
```

2. 对于--location-arn 参数，请指定要更新的 HDFS 位置的 ARN。
3. 对于--agent-arns 参数，请指定新代理的 ARN。

   如果您要为一个位置更换多个代理，则可以指定多个 ARN。
4. 如果您使用的是 Kerberos 身份验证，请纳入--kerberos-keytab和--kerberos-krb5-conf参数：

   - 对于 --kerberos-keytab 参数，请指定 Kerberos 密钥表（keytab），其中包含定义的 Kerberos 主体和加密密钥之间的映射。

     您可以通过提供文件的地址指定密钥表文件。
   - 对于--kerberos-krb5-conf 参数，请指定包含您的 Kerberos 领域配置的文件。

     您可以通过提供文件的地址指定 krb5.conf文件。

   如果您使用的是简单身份验证，则无需在命令中纳入与这些与 Kerberos 相关的参数。
5. 运行 update-location-hdfs命令以更新位置。

Object storage

1. 复制以下 update-location-object-storage命令：

```
aws datasync update-location-object-storage \
  --location-arn datasync-object-storage-location-arn \
  --agent-arns new-datasync-agent-arn \
  --secret-key bucket-secret-key
```

2. 对于--location-arn 参数，请指定要更新的对象存储位置的 ARN。
3. 对于--agent-arns 参数，请指定新代理的 ARN。

   如果您要为一个位置更换多个代理，则可以指定多个 ARN。
4. 根据对象存储位置是否需要访问凭证，执行以下操作：

   - 如果您的位置需要凭证，为 --secret-key 参数指定使 DataSync 能够访问对象存储桶的私有密钥。

- 如果您的位置需要凭证，为 `--access-key` 和 `--secret-key` 参数指定空字符串。以下是一个示例 命令：

```
aws datasync update-location-object-storage \
  --location-arn arn:aws:datasync:us-east-2:111122223333:location/
loc-abcdef01234567890 \
  --agent-arns arn:aws:datasync:us-east-2:111122223333:agent/
agent-1234567890abcdef0 \
  --access-key "" \
  --secret-key ""
```

5. 运行 `update-location-object-storage`命令以更新位置。

Azure Blob Storage

1. 复制以下 update-location-azure-blob命令：

```
aws datasync update-location-azure-blob \
  --location-arn datasync-azure-blob-storage-location-arn \
  --agent-arns new-datasync-agent-arn \
  --sas-configuration '{
      "Token": "sas-token-for-azure-blob-storage"
    }'
```

2. 对于 `--location-arn` 参数，请指定要更新的 Azure Blob Storage 位置的 ARN。
3. 对于`--agent-arns` 参数，请指定新代理的 ARN。

   如果您要为一个位置更换多个代理，则可以指定多个 ARN。
4. 对于 `--sas-configuration` 参数的 Token 选项，请指定允许 DataSync 访问您的 blob 存储空间的 SAS 令牌。
5. 运行 `update-location-azure-blob`命令以更新位置。

后续步骤

1. 删除旧代理。如果您有任何正在使用此代理运行的 DataSync 任务，请等到这些任务完成后再删除。
2. 如果需要更换多个位置的代理，请重复前面的步骤。
3. 完成后，您可以继续运行任务。

> **Note**
>
> 为计划任务更换代理：如果要为计划任务更换代理，而新代理使用的服务端点类型与旧代理不同，则必须手动启动该任务。如果不在下一次计划运行前手动运行任务，任务将失败。
> 例如，如果旧代理使用公共服务端点，而新代理使用 VPC 端点，则应使用控制台或 `StartTaskExecution` 操作，手动启动该任务。之后，您的任务将继续按计划运行。

筛选 AWS DataSync 资源

您可以使用中的和 `ListTasks` API 操作筛选您的 AWS DataSync 位置`ListLocations`和任务 AWS CLI。例如，检索最近任务的列表。

用于筛选的参数

您可以使用 API 筛选器来缩小 `ListTasks` 和 `ListLocations` 返回的资源列表的范围。例如，要检索所有 Amazon S3 位置，您可以使用筛选器名称为 `LocationType`*`S3`* 和 `Operator`*`Equals`* 的 `ListLocations`。

要筛选 API 结果，您必须指定筛选器名称、运算符和值。

- `Name` — 正使用的筛选器的名称。每个 API 调用都支持可用于其的筛选器列表（例如，`ListLocations` 的 `LocationType`）。
- `Values` — 要筛选的值。例如，您可能希望仅显示 Amazon S3 位置。
- `Operator` — 用于比较筛选值的运算符（例如，`Equals` 或 `Contains`）。

下表列出了可用的运算符。

运算符	密钥类型
`Equals`	字符串、数字
`NotEquals`	字符串、数字
`LessThan`	数字

运算符	密钥类型
LessThanOrEqual	数字
GreaterThan	数字
GreaterThanOrEqual	数字
In	字符串
Contains	字符串
NotContains	字符串
BeginsWith	字符串

按位置筛选

ListLocations 支持以下筛选器名称：

- LocationType — 按位置类型进行筛选：
  - SMB
  - NFS
  - HDFS
  - OBJECT_STORAGE
  - S3
  - OUTPOST_S3
  - FSX_WINDOWS
  - FSX_LUSTRE
  - FSX_OPENZFS_NFS
  - FSX_ONTAP_NFS
  - FSX_ONTAP_SMB
- LocationUri — 按分配给该位置的统一资源标识符（URI）筛选，该标识符由 DescribeLocation* API 调用返回（例如，Amazon S3 位置的 s3://*bucket-name*/*your-prefix*）。

- `CreationTime` — 按创建位置的时间筛选。输入格式为 `yyyy-MM-dd:mm:ss`，采用协调世界时 (UTC)。

以下 AWS CLI 示例列出了所有类型为 Amazon S3 的营业地点，这些营业地点 URI 以字符串开头，"s3://amzn-s3-demo-bucket"且创建时间为 2019-12-15 17:15:20 或之后（世界标准时间）。

```
aws datasync list-locations \
    --filters [{Name=LocationType, Values=["S3"], Operator=Equals},
 {Name=LocationUri, Values=["s3://amzn-s3-demo-bucket"], Operator=BeginsWith},
 {Name=CreationTime,Values=["2019-12-15 17:15:20"],Operator=GreaterThanOrEqual}]
```

该命令会返回类似以下内容的输出。

```
{
    "Locations": [
        {
            "LocationArn": "arn:aws:datasync:us-east-1:111122223333:location/
loc-333333333abcdef0",
            "LocationUri": "s3://amzn-s3-demo-bucket1/"
        },
        {
            "LocationArn": "arn:aws:datasync:us-east-1:123456789012:location/
loc-987654321abcdef0",
            "LocationUri": "s3://amzn-s3-demo-bucket2/"
        }
    ]
}
```

按任务筛选

`ListTasks` 支持以下筛选器名称。

- `LocationId` — 按 Amazon 资源名称（ARN）值上的源位置和目标位置筛选。
- `CreationTime` — 按创建任务的时间筛选。输入格式为 UTC 形式的 `yyyy-MM-dd:mm:ss`。

以下 AWS CLI 示例显示了筛选时的语法`LocationId`。

```
aws datasync list-tasks \
```

```
    --filters Name=LocationId,Values=arn:aws:datasync:us-east-1:your-account-
id:location/your-location-id,Operator=Contains
```

此命令的输出如下。

```
{
    "Tasks": [
        {
            "TaskArn": "arn:aws:datasync:us-east-1:your-account-id:task/your-task-id",
            "Status": "AVAILABLE",
            "Name": "amzn-s3-demo-bucket"
        }
    ]
}
```

清理您的 AWS DataSync 资源

如果您 AWS DataSync 用于测试或者不需要您创建的 AWS 资源，请将其删除，这样您就无需为不打算使用的资源付费。

> **Note**
>
> 如果您在禁用的可选区域中有 DataSync 资源，则这些资源不会自动删除。如果您再次启用该区域，这些资源仍然存在。

删除代 DataSync 理

当您从中删除代理时 AWS DataSync，该代理资源将不再与您的关联， AWS 账户 并且无法撤消。

请记住，从中删除代理 DataSync 并不会将其虚拟机 (VM) 或 Amazon EC2 实例从您的存储环境中移除。您可以删除虚拟机或实例，也可以再次用它来激活一个新代理。

先决条件

在更新或移除依赖代理的 DataSync 资源之前，请勿删除代理。如果您要更换代理，请使用新的代理更新您的传输位置。如果您不更换代理，则先使用该代理删除传输任务和位置。

删除代理

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。

2. 在左侧导航窗格中，选择代理。
3. 选择要删除的代理。
4. 选择删除，在显示的文本框中键入 **delete**，然后选择删除。
5. 如果您不打算将代理的基础设施重复用于其他 DataSync 活动，请删除代理的虚拟机或 Amazon EC2 实例，将其从存储环境中移除。

重复使用 DataSync 代理的基础架构

您可以从中 DataSync 删除代理资源，但仍可使用代理的底层虚拟机或 Amazon EC2 实例来激活新代理。

要重复使用代理的基础设施

1. 测试代理与的连接 AWS。如果网络测试通过，请转到下一步。

   继续执行下一步之前，必须先通过网络测试。
2. 从@@ 代理的虚拟机或 Amazon EC2 实例中删除代理资源， DataSync 但不要删除该代理的虚拟机或 Amazon EC2 实例。
3. 重复步骤 1， AWS 再次测试代理与的连接。如果网络测试通过，请转到下一步。
4. 从中删除代理资源大约三分钟后 DataSync，请检查代理虚拟机或 Amazon EC2 实例上的端口 80 是否已打开。如果已打开，请转到下一步。
5. 使用现有虚拟机或 Amazon EC2 实例激活新代理。

   您可以在不同的 AWS 区域 AWS 账户、和使用其他类型的服务终端节点中激活新代理。如果您使用不同类型的服务端点，则必须调整网络配置。

删除 DataSync 地点

最佳做法是移除不再需要 AWS DataSync 的地点。

使用 DataSync 控制台移除营业地点

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择位置。
3. 选择要删除的位置。
4. 选择删除。通过输入 **delete** 确认删除，然后选择删除。

删除 DataSync 任务

如果您不再需要某 AWS DataSync 项任务，则可以删除该任务及其相关 AWS 资源。

先决条件

运行任务时， DataSync 会自动创建和管理用于数据传输流量的网络接口。删除任务时，只要您具有以下权限，您也可以删除其相关的网络接口：

- `ec2:DeleteNetworkInterface`
- `ec2:DescribeNetworkInterfaces`
- `ec2:ModifyNetworkInterfaceAttribute`

这些权限可在 AWS 托管策略中找到`AWSDataSyncFullAccess`。有关更多信息，请参阅 AWS 的托管策略 AWS DataSync。

删除任务

任务删除后，其将无法恢复。

使用控制 DataSync 台

1. 打开 AWS DataSync 控制台，网址为https://console.aws.amazon.com/datasync/。
2. 在左侧导航窗格中，展开数据传输，然后选择任务。
3. 选择要删除的任务。
4. 选择 Actions (操作)，然后选择 Delete (删除)。
5. 在对话框中，选择删除。

使用 AWS CLI

1. 复制以下 `delete-task` 命令：

```
aws datasync delete-task \
      --task-arn "task-to-delete"
```

2. 对于`--task-arn`参数，指定要删除的任务的 Amazon 资源名称（ARN）（例如`arn:aws:datasync:us-east-2:123456789012:task/task-012345678abcd0123`）。

3. 运行 `delete-task` 命令。

安全性 AWS DataSync

云安全 AWS 是重中之重。作为 AWS 客户，您可以受益于专为满足大多数安全敏感型组织的要求而构建的数据中心和网络架构。

安全是双方共同承担 AWS 的责任。责任共担模式将其描述为云的 安全性和云中 的安全性：

- 云安全 — AWS 负责保护在 AWS 云中运行 AWS 服务的基础架构。 AWS 还为您提供可以安全使用的服务。作为 AWS 合规性计划的一部分，第三方审核人员将定期测试和验证安全性的有效性。要了解适用于 AWS DataSync的合规性计划，请参阅合规性计划范围内的AWS 服务。
- 云端安全-您的责任由您使用的 AWS 服务决定。您还需要对其它因素负责，包括您的数据的敏感性、您公司的要求以及适用的法律法规。

本文档可帮助您了解在使用时如何应用分担责任模型 DataSync。以下主题向您介绍如何进行配置 DataSync 以满足您的安全和合规性目标。您还将学习如何使用其他 AWS 服务来帮助您监控和保护您的 DataSync 资源。

主题

- 中的数据保护 AWS DataSync
- 中的身份和访问管理 AWS DataSync
- 的合规性验证 AWS DataSync
- AWS DataSync 中的恢复功能
- AWS DataSync 中的基础结构安全性
- 使用 Secrets Manager 保护存储位置凭证

中的数据保护 AWS DataSync

AWS DataSync 在自行管理的存储系统和存储服务之间以及 AWS 存储服务之间 AWS 安全地传输数据。您的存储数据在传输过程中的加密方式在一定程度上取决于传输所涉及的位置。

传输完成后，存储数据的系统或服务会对静态数据进行加密（不是 DataSync）。

主题

- AWS DataSync 传输中的加密
- AWS DataSync 静态加密

- 互联网络流量隐私

AWS DataSync 传输中的加密

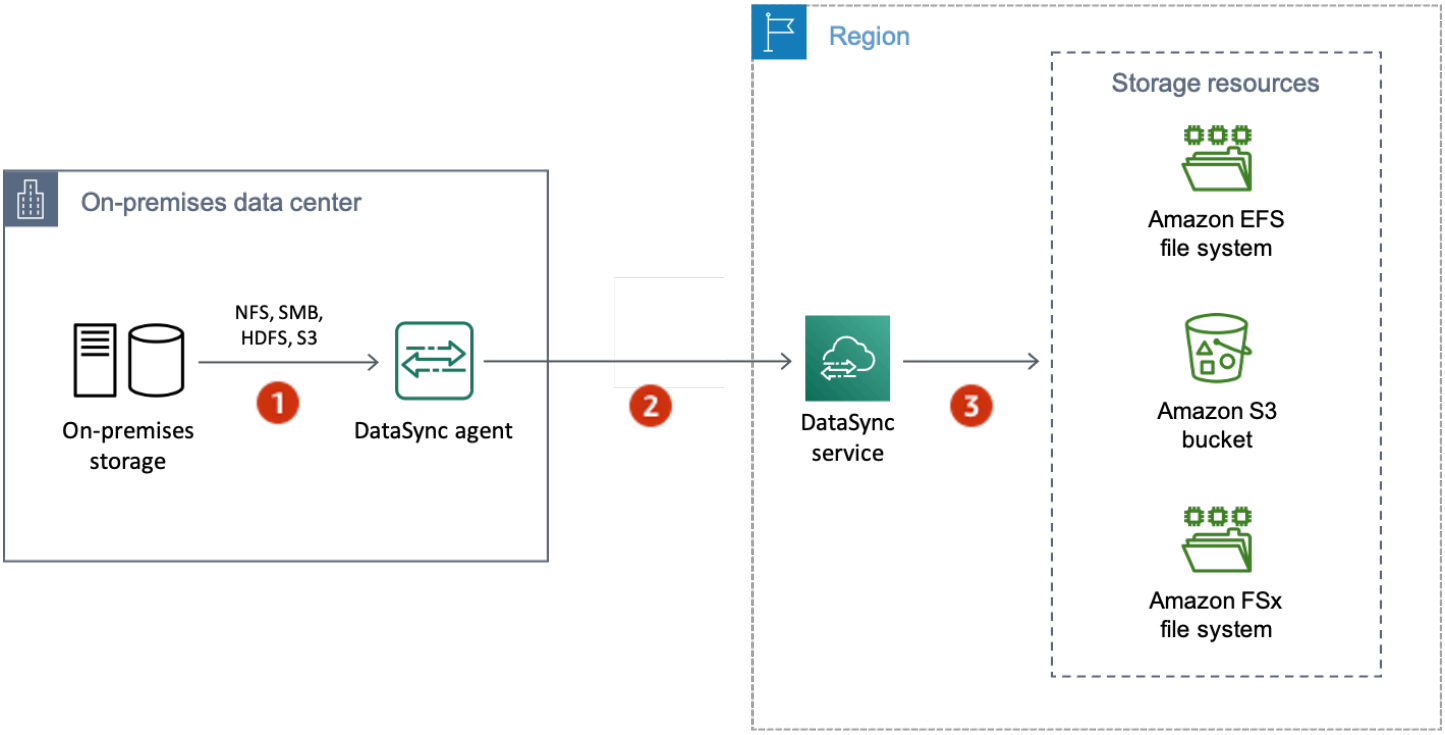
您的存储数据（包括元数据）在传输过程中会进行加密，但在整个传输过程中如何对其进行加密取决于您的源位置和目标位置。

连接某个位置时， DataSync 使用该地点的数据访问协议提供的最安全的选项。例如，使用服务器消息块 (SMB) 连接文件系统时， DataSync 使用 SMB 提供的安全功能。

传输过程中的网络连接

DataSync 需要三个网络连接才能复制数据：一个用于从源位置读取数据的连接，另一个用于在不同位置之间传输数据的连接，以及一个用于将数据写入目标位置的连接。

下图是 DataSync 用于将数据从本地存储系统传输到 AWS 存储服务的网络连接示例。要了解连接发生的位置，以及数据在通过每个连接进行传输时如何受到保护，请使用随附的表格。



参考	网络连接	说明
1	从源位置读取数据	DataSync 使用存储系统的数据访问协议（例如 SMB 或 Amazon S3

参考	网络连接	说明
		API ）进行连接。对于这种连接，除非 DataSync 不支持这些功能，否则将使用存储系统的安全功能来保护数据。例如， DataSync 目前不支持使用 NFS 文件服务器进行 Kerberos 身份验证，也不支持在 HDFS 中使用 TDE 加密时进行身份验证。
2	在不同位置之间传输数据	对于此连接， DataSync 使用相互传输层安全 (mTLS) 1.3 加密所有网络流量。
3	将数据写入目标位置	与源位置一样，使用存储系统的数据访问协议进行 DataSync 连接。除非DataSync不支持存储系统的安全功能，否则数据将再次受到保护。

了解在 DataSync 连接到以下 AWS 存储服务时，您的数据在传输过程中是如何加密的：

- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的亚马逊
- Amazon S3

TLS 密码

在不同位置之间传输数据时， DataSync 使用不同的 TLS 密码。TLS 密码取决于您的代理用来与之通信的服务端点的类型。 DataSync（有关更多信息，请参阅 为 AWS DataSync 代理选择服务端点。）

目录

- 公共或 VPC 端点
- FIPS 端点

公共或 VPC 端点

对于公有云和虚拟私有云 (VPC) 服务终端节点，请 DataSync 使用以下 TLS 密码之一：

- TLS_ECDHE_RSA_WITH_AES_256_GCM_ (ecdh SHA384 _x25519)
- TLS_ECDHE_RSA_WITH_ _ _ (ecdh_x25519CHACHA20) POLY1305 SHA256
- TLS_ECDHE_RSA_WITH_AES_128_GCM_ (ecdh SHA256 _x25519)

FIPS 端点

对于联邦信息处理标准 (FIPS) 服务端点， DataSync使用以下 TLS 密码：

- TLS_AES_128_GCM_ (secp256r1SHA256)

AWS DataSync 静态加密

由于 AWS DataSync 是传输服务，因此它通常不管理您的静态存储数据。 DataSync 支持的存储服务和系统负责保护处于该状态的数据。但是，有些与服务相关的数据是静态 DataSync管理的。

什么是加密？

唯一 DataSync 可以处理的静态数据与完成传输所需的详细信息有关。 DataSync 在 Amazon DynamoDB 中使用完全静态加密存储以下数据：

- 任务配置（例如，有关传输过程中位置的详细信息）。
- 允许您的 DataSync 代理使用位置进行身份验证的用户凭证。这些凭证使用代理的公钥进行加密。代理可以根据需要使用其私钥对这些密钥进行解密。

有关更多信息，请参阅 Amazon DynamoDB 开发人员指南 中的 静态加密。

密钥管理

您无法管理用于在 DynamoDB 中存储与运行任务相关的信息的加密密钥。 DataSync 此信息包括您的任务配置以及代理用于对存储位置进行身份验证的凭证。

哪些内容未加密？

尽管 DataSync 不能控制存储数据的静态加密方式，但我们仍然建议您使用它们所支持的最高级别的安全性来配置您的位置。例如，您可以使用 Amazon S3 托管加密密钥 (SSE-S3) 或 (KMS) 密AWS 钥 AWS Key Management Service (SSE-KMS) 加密对象。

详细了解 AWS 存储服务如何加密静态数据：

- Amazon S3
- Amazon EFS
- FSx 适用于 Windows 文件服务器的亚马逊
- 亚马逊 f FSx or Lustre
- FSx 适用于 OpenZFS 的亚马逊
- FSx 适用于 NetApp ONTAP 的亚马逊

互联网络流量隐私

我们建议将源位置和目标位置配置为各自支持的最高安全级别。连接到某个位置时 AWS DataSync ，使用存储系统使用的最安全版本的数据访问协议。此外，请考虑将子网流量限制为已知协议和服务。

DataSync 使用传输层安全 (TLS) 1.3 保护位置之间的连接（包括 AWS 账户 AWS 区域、和可用区之间）。

中的身份和访问管理 AWS DataSync

AWS 使用安全证书来识别您的身份并授予您访问 AWS 资源的权限。您可以使用 AWS Identity and Access Management (IAM) 的功能允许其他用户、服务和应用程序完全或以有限的方式使用您的 AWS 资源，而无需共享您的安全证书。

默认情况下，IAM 身份（用户、群组和角色）无权创建、查看或修改 AWS 资源。要允许用户、群组和角色访问 AWS DataSync 资源并与 DataSync 控制台和 API 进行交互，我们建议您使用 IAM 策略，授予他们使用他们所需的特定资源和 API 操作的权限。然后，将策略附加到需要访问的 IAM 身份。有关策略的基本要素的概述，请参阅的访问管理 AWS DataSync。

主题

- 的访问管理 AWS DataSync
- AWS 的托管策略 AWS DataSync

- IAM 客户托管的政策 AWS DataSync
- 将服务相关角色用于 DataSync
- 在创建期间为 DataSync 资源添加标签的权限
- 防止跨服务混淆代理

的访问管理 AWS DataSync

每个 AWS 资源都归某人所有 AWS 账户。创建或访问资源的权限由权限策略进行管理。账户管理员可以向 AWS Identity and Access Management (IAM) 身份附加权限策略。某些服务（例如 AWS Lambda）还支持为资源附加权限策略。

> **Note**
>
> 账户管理员是在 AWS 账户中具有管理员权限的用户。有关更多信息，请参阅《IAM 用户指南》中的 IAM 最佳实践。

主题

- DataSync 资源和运营
- 了解资源所有权
- 管理对 资源的访问
- 指定策略元素：操作、效果、资源和主体
- 在策略中指定条件
- 创建 VPC 端点策略

DataSync 资源和运营

在中 DataSync，主要资源是代理、位置、任务和任务执行。

这些资源具有与之关联的唯一 Amazon 资源名称 (ARNs)，如下表所示。

资源类型	ARN 格式
座席 ARN	`arn:aws:datasync:` *region*`:`*account-id* `:agent/`*agent-id*
位置 ARN	`arn:aws:datasync:` *region*`:`*account-id* `:location/` *location-id*

资源类型	ARN 格式
任务 ARN	`arn:aws:datasync:` *region*`:`*account-id* `:task/`*task-id*
任务执行 ARN	`arn:aws:datasync:` *region*`:`*account-id* `:task/`*task-id*`/execution/` *exec-id*

要授予特定 API 操作（例如创建任务）的权限，需要 DataSync 定义一组可在权限策略中指定的操作。一个 API 操作可能需要执行多个操作的权限。

了解资源所有权

资源所有者 AWS 账户 是创建资源的人。也就是说，资源所有者是委托人实体（例如，IAM 角色），负责对创建资源的请求进行身份验证。 AWS 账户 以下示例说明了该行为的工作原理：

- 如果您使用您的 AWS 账户 根账户证书创建任务，则您 AWS 账户 就是该资源的所有者（在中 DataSync，资源就是任务）。
- 如果您在中创建 IAM 角色 AWS 账户 并向该用户授予`CreateTask`操作权限，则该用户可以创建任务。但是，该用户所属的 AWS 账户是此任务资源的所有者。
- 如果您在中创建 AWS 账户 具有创建任务权限的 IAM 角色，则任何能够担任该角色的人都可以创建任务。角色 AWS 账户所属的您拥有任务资源。

管理对 资源的访问

权限策略规定谁可以访问哪些内容。下一节介绍创建权限策略时的可用选项。

> **Note**
>
> 本节讨论在的上下文中使用 IAM DataSync。这里不提供有关 IAM 服务的详细信息。有关完整的 IAM 文档，请参阅 IAM 用户指南中的什么是 IAM？。有关 IAM policy 语法和说明的信息，请参阅 IAM 用户指南中的 AWS Identity and Access Management 策略参考。

附加到 IAM 身份的策略称为基于身份的策略（IAM 策略），附加到资源的策略称为基于资源的策略。DataSync 仅支持基于身份的策略（IAM 策略）。

主题

- 基于身份的策略
- 基于资源的策略

基于身份的策略

您可以使用 IAM 策略管理 DataSync 资源访问权限。这些策略可以帮助 AWS 账户 管理员执行以下操作 DataSync：

- 授予创建和管理 DataSync资源的权限-创建 IAM 策略， AWS 账户 允许您中的 IAM 角色创建和管理 DataSync 资源，例如代理、位置和任务。
- 向其他角色 AWS 账户 或中的角色授予权限 AWS 服务 — 创建 IAM 策略，向其他角色 AWS 账户 或中的某个 IAM 角色授予权限 AWS 服务。例如：
  1. 账户 A 管理员可以创建一个 IAM 角色，然后向该角色附加授予其访问账户 A 中资源的权限策略。
  2. 账户 A 管理员向将账户 B 标识为能够担任该角色的委托人的角色附加信任策略。

     要授予担任该角色的 AWS 服务 权限，账户 A 管理员可以在信任策略中指定 AWS 服务 为委托人。
  3. 随后，账户 B 管理员可以将担任角色的权限委派给账户 B 中的任何用户。这将允许账户 B 中担任此角色的任何用户创建或访问账户 A 中的资源。

  有关使用 IAM 委派权限的更多信息，请参阅 IAM 用户指南中的访问权限管理。

以下示例策略授予对所有资源执行所有 `List*` 操作的权限。此操作为只读操作，不允许修改资源。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "AllowAllListActionsOnAllResources",
            "Effect": "Allow",
            "Action": [
                "datasync:List*"
            ],
            "Resource": "*"
        }
    ]
```

```
}
```

有关将基于身份的策略与配合使用的更多信息 DataSync，请参阅AWS 托管策略和客户托管策略。有关 IAM 身份的更多信息，请参阅 IAM 用户指南。

基于资源的策略

其他服务 [如 Simple Storage Service（Amazon S3）] 支持基于资源的权限策略。例如，您可以将策略附加到 Amazon S3 存储桶以管理对该存储桶的访问权限。但是， DataSync 不支持基于资源的策略。

指定策略元素：操作、效果、资源和主体

对于每种 DataSync 资源，该服务都定义了一组 API 操作（请参阅操作）。要授予这些 API 操作的权限，请 DataSync 定义一组可在策略中指定的操作。例如，对于 DataSync 资源，定义了以下操作：`CreateTask`、`DeleteTask`、和`DescribeTask`。执行一个 API 操作可能需要多个操作的权限。

以下是最基本的策略元素：

- 资源：在策略中，您可以使用 Amazon Resource Name（ARN）标识策略应用到的资源。对于 DataSync 资源，您可以在 IAM 策略中使用通配符（*）。有关更多信息，请参阅 DataSync 资源和运营。
- 操作 – 您可以使用操作关键字标识要允许或拒绝的资源操作。例如，根据指定的`Effect`元素，该`datasync:CreateTask`权限允许或拒绝用户执行 DataSync `CreateTask`操作的权限。
- 效果 – 您可以指定当用户请求特定操作时的效果（该效果可以是 `Allow` 或 `Deny`）。如果您没有显式授予对（`Allow`）资源的访问权限，则隐式拒绝访问。您也可显式拒绝对资源的访问，这样可确保用户无法访问该资源，即使有其他策略授予了用户访问权限的情况下也是如此。有关更多信息，请参阅 IAM 用户指南中的授权。
- 主体：在基于身份的策略（IAM 策略）中，附加了策略的用户是隐式主体。对于基于资源的策略，您可以指定要获得权限的用户、账户、服务或其他实体（仅适用于基于资源的策略）。 DataSync 不支持基于资源的策略。

有关 IAM policy 语法和描述的更多信息，请参阅 IAM 用户指南中的 AWS Identity and Access Management 策略参考。

在策略中指定条件

当您授予权限时，可使用 IAM policy 语言指定一些条件，这些条件规定在授予权限时策略何时生效。例如，您可能希望策略仅在特定日期后应用。有关使用策略语言指定条件的更多信息，请参阅 IAM 用户指南中的条件。

要表示条件，您可以使用预定义的条件键。没有特定于 DataSync 的条件键。但是，您可以根据需要使用 AWS 多种条件键。有关 AWS 宽密钥的完整列表，请参阅 IAM 用户指南中的可用密钥。

创建 VPC 端点策略

VPC 终端节点策略可帮助控制通过 DataSyncVPC服务终端节点和启用 FIPS 的 VPC 服务终端节点对 DataSync API 操作的访问。VPC 终端节点策略允许您限制通过VPC服务终端节点访问的特定 DataSync API 操作，例如`CreateTask`或`StartTaskExecution`。

端点策略指定以下信息：

- 可以执行操作的主体。
- 可执行的操作。
- 可对其执行操作的资源。

有关更多信息，请参阅使用端点策略控制对 VPC 端点的访问。

策略示例

下面是端点策略的一个示例。

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Principal": "*",
        "Action": [
          "datasync:CreateTask",
          "datasync:StartTaskExecution",
          "datasync:DescribeTask"
        ],
        "Resource": "arn:aws:datasync:us-east-1:123456789012:task/*"
      }
   ]
```

```
}
```

AWS 的托管策略 AWS DataSync

要向用户、群组和角色添加权限，使用 AWS 托管策略比自己编写策略要容易得多。创建仅为团队提供所需权限的 IAM 客户管理型策略需要时间和专业知识。要快速入门，您可以使用我们的 AWS 托管策略。这些策略涵盖常见使用案例，可在您的 AWS 账户中使用。有关 AWS 托管策略的更多信息，请参阅 IAM 用户指南中的AWS 托管策略。

AWS 服务 维护和更新 AWS 托管策略。您无法更改 AWS 托管策略中的权限。服务偶尔会向 AWS 托管式策略添加额外权限以支持新特征。此类更新会影响附加策略的所有身份（用户、组和角色）。当启动新特征或新操作可用时，服务最有可能会更新 AWS 托管式策略。服务不会从 AWS 托管策略中移除权限，因此策略更新不会破坏您的现有权限。

此外，还 AWS 支持跨多个服务的工作职能的托管策略。例如，`ReadOnlyAccess` AWS 托管策略提供对所有资源 AWS 服务 和资源的只读访问权限。当服务启动一项新功能时， AWS 会为新操作和资源添加只读权限。有关工作职能策略的列表和说明，请参阅 IAM 用户指南中的适用于工作职能的AWS 托管式策略。

AWS 托管策略：AWSDataSyncReadOnlyAccess

您可以将 `AWSDataSyncReadOnlyAccess` 策略附加到 IAM 身份。此策略授予的只读权限 DataSync。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AWSDataSyncReadOnlyAccess。

AWS 托管策略：AWSDataSyncFullAccess

您可以将 `AWSDataSyncFullAccess` 策略附加到 IAM 身份。此策略授予对服务的管理权限 DataSync ，并且是 AWS 管理控制台 访问该服务所必需的。 `AWSDataSyncFullAccess`提供对 DataSync API 操作以及与相关资源（例如 Amazon S3 存储桶、Amazon EFS 文件系统、 AWS KMS 密钥和 Secrets Manager 密钥）交互的操作的完全访问权限。该策略还向 Amazon 授予权限 CloudWatch，包括创建日志组以及创建或更新资源策略。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AWSDataSyncFullAccess。

AWS 托管策略：AWSDataSyncServiceRolePolicy

您不可以将 AWSDataSyncServiceRolePolicy 策略附加得到 IAM 身份。此策略附加到允许代表您执行操作 DataSync 的服务相关角色。有关更多信息，请参阅 将服务相关角色用于 DataSync。

此策略授予管理权限，允许服务相关角色使用增强模式为 DataSync 任务创建 Amazon CloudWatch 日志。

策略更新

更改	描述	日期
AWSDataSyncFullAccess – 更改	DataSync 修改了以下内容的权限声明AWSDataSyncFullAccess ： 更新后的语句从用于创建 Secret DataSync s Manager 密钥的权限中删除了标记条件。	2025 年 5 月 13 日
AWSDataSyncFullAccess – 更改	DataSync 向以下内容添加了新权限AWSDataSyncFullAccess ： • secretsmanager:CreateSecret • secretsmanager:PutSecretValue • secretsmanager:DeleteSecret • secretsmanager:UpdateSecret 这些权限允许 DataSync 创建、编辑和删除 AWS Secrets Manager 密钥。	2025 年 5 月 7 日

更改	描述	日期
AWSDataSyncFullAccess – 更改	DataSync 向以下内容添加了新权限AWSDataSyncFullAccess ： • secretsmanager:ListSecrets • kms:ListAliases • kms:DescribeKey 这些权限允许 DataSync 检索有关您的 AWS Secrets Manager 机密和密 AWS KMS 钥的元数据，包括与您的密钥关联的任何别名。	2025 年 4 月 23 日
AWSDataSyncServiceRolePolicy – 更改	DataSync 为 DataSync服务相关角色AWSServiceRoleForDataSync 使用的AWSDataSyncServiceRolePolicy 策略添加了新权限： • secretsmanager:DescribeSecret • secretsmanager:GetSecretValue 这些权限允许 DataSync 读取由管理的密钥的元数据和值 AWS Secrets Manager。	2025 年 4 月 15 日

更改	描述	日期
AWSDataSyncServiceRolePolicy：新策略	DataSync 添加了 DataSync 服务相关角色`AWSServiceRoleForDataSync` 使用的策略。这项新的托管策略会自动为您的使用增强模式的 DataSync 任务创建 Amazon CloudWatch 日志。	2024 年 10 月 30 日
AWSDataSyncFullAccess – 更改	DataSync 为以下内容添加了新的权限`AWSDataSyncFullAccess` ： • `iam:CreateServiceLinkedRole` 此权限允许您 DataSync 创建与服务相关的角色。	2024 年 10 月 30 日
AWSDataSyncFullAccess – 更改	DataSync 为以下内容添加了新的权限`AWSDataSyncFullAccess` ： • `ec2:DescribeRegions` 此权限允许您在创建 DataSync 任务时选择选择加入区域，以便在两者之间 AWS 区域进行转移。	2024 年 7 月 22 日

更改	描述	日期
AWSDataSyncFullAccess – 更改	DataSync 为以下内容添加了新的权限AWSDataSyncFullAccess ： • s3:ListBucketVersions 此权限允许您选择DataSync 清单的特定版本。	2024 年 2 月 16 日

更改	描述	日期
AWSDataSyncFullAccess – 更改	DataSync 向以下内容添加了新权限AWSDataSyncFullAccess ： • ec2:DescribeVpcEndpoints • elasticfilesystem:DescribeAccessPoints • fsx:DescribeStorageVirtualMachines • outposts:ListOutposts • s3:GetBucketLocation • s3-outposts:ListAccessPoints • s3-outposts:ListRegionalBuckets 这些权限可以帮助你在 Outposts 上为亚马逊 EFS、Amazon f NetApp o FSx r ONTAP、Amazon S3 和 S3 创建 DataSync 代理和位置。	2023 年 5 月 2 日
DataSync 开始跟踪更改	DataSync 开始跟踪其 AWS 托管策略的更改。	2021 年 3 月 1 日

IAM 客户托管的政策 AWS DataSync

除了 AWS 托管策略外，您还可以为需要这些权限的 (IAM) 身份创建自己的基于身份的策略， AWS DataSync 并将其附加到需要这些权限的 AWS Identity and Access Management (IAM) 身份。这些被称为客户管理型策略，是您在自己的 AWS 账户中管理的独立策略。

> **Important**
>
> 在开始之前，我们建议您了解管理 DataSync 资源访问权限的基本概念和选项。有关更多信息，请参阅 的访问管理 AWS DataSync。

在创建客户托管策略时，您需要包括有关可在某些 AWS 资源上使用的 DataSync 操作的声明。以下策略示例包含两个语句 (请注意每个语句中的 `Action` 和 `Resource` 元素)：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "AllowsSpecifiedActionsOnAllTasks",
            "Effect": "Allow",
            "Action": "datasync:DescribeTask",
            "Resource": "arn:aws:datasync:us-east-1:111122223333:task/*"
        },
        {
            "Sid": "ListAllTasks",
            "Effect": "Allow",
            "Action": "datasync:ListTasks",
            "Resource": "*"
        }
    ]
}
```

此策略的语句执行以下操作：

- 第一条语句通过使用通配符 (*) 指定 Amazon 资源名称 (ARN) 来授予对某些转移任务资源执行 `datasync:DescribeTask` 操作的权限。

- 第二条语句通过仅指定一个通配符（*）来授予对所有任务执行 datasync:ListTasks 操作的权限。

客户管理型策略示例

以下示例客户托管策略授予各种 DataSync操作的权限。如果您使用的是 AWS Command Line Interface (AWS CLI) 或 AWS SDK，则这些策略将起作用。要在控制台中使用这些策略，还必须使用托管策略 AWSDataSyncFullAccess。

主题

- 示例 1：创建允许 DataSync 访问您的 Amazon S3 存储桶的信任关系
- 示例 2：允许读 DataSync 取和写入您的 Amazon S3 存储桶
- 示例 3： DataSync 允许将日志上传到 CloudWatch 日志组

示例 1：创建允许 DataSync 访问您的 Amazon S3 存储桶的信任关系

以下是允许担任 IAM 角色 DataSync 的信任策略示例。此角色 DataSync 允许访问 Amazon S3 存储桶。为防范跨服务混淆代理问题，我们建议在此策略中使用 aws:SourceArn 和 aws:SourceAccount 全局条件上下文键。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "datasync.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                "aws:SourceAccount": "111111111111"
                },
                "ArnLike": {
                "aws:SourceArn": "arn:aws:datasync:us-east-1:111111111111:*"
                }
            }
```

```
        }
    ]
}
```

示例 2：允许读 DataSync 取和写入您的 Amazon S3 存储桶

以下示例策略授予 DataSync 读取和向用作目标位置的 S3 存储桶写入数据的最低权限。

> **Note**
>
> `aws:ResourceAccount` 的值应为拥有策略中指定 Amazon S3 存储桶的账户 ID。

```
{
 "Version": "2012-10-17",
 "Statement": [
     {
         "Action": [
             "s3:GetBucketLocation",
             "s3:ListBucket",
             "s3:ListBucketMultipartUploads"
         ],
         "Effect": "Allow",
         "Resource": "arn:aws:s3:::amzn-s3-demo-bucket",
         "Condition": {
             "StringEquals": {
                 "aws:ResourceAccount": "123456789012"
             }
         }
     },
     {
         "Action": [
             "s3:AbortMultipartUpload",
             "s3:DeleteObject",
             "s3:GetObject",
             "s3:GetObjectTagging",
             "s3:GetObjectVersion",
             "s3:GetObjectVersionTagging",
             "s3:ListMultipartUploadParts",
             "s3:PutObject",
             "s3:PutObjectTagging"
```

```
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/*",
            "Condition": {
                "StringEquals": {
                    "aws:ResourceAccount": "123456789012"
                }
            }
        }
    ]
}
```

示例 3： DataSync 允许将日志上传到 CloudWatch 日志组

DataSync 需要权限才能将日志上传到您的 Amazon CloudWatch 日志组。您可以使用 CloudWatch 日志组来监控和调试任务。

有关授予这类权限的示例 IAM policy，请参阅 DataSync 允许将日志上传到 CloudWatch 日志组。

将服务相关角色用于 DataSync

AWS DataSync 使用 AWS Identity and Access Management (IAM) 服务相关角色。服务相关角色是一种与之直接关联的 IAM 角色的独特类型。 DataSync服务相关角色由服务预定义 DataSync ，包括该服务代表您调用其他 AWS 服务所需的所有权限。

主题

- 将角色用于 DataSync

将角色用于 DataSync

AWS DataSync 使用 AWS Identity and Access Management (IAM) 服务相关角色。服务相关角色是一种与之直接关联的 IAM 角色的独特类型。 DataSync服务相关角色由服务预定义 DataSync ，包括该服务代表您调用其他 AWS 服务所需的所有权限。

服务相关角色使设置变得 DataSync 更加容易，因为您不必手动添加必要的权限。 DataSync 定义其服务相关角色的权限，除非另有定义，否则 DataSync 只能担任其角色。定义的权限包括信任策略和权限策略，以及不能附加到任何其他 IAM 实体的权限策略。

只有在首先删除相关资源后，您才能删除服务关联角色。这样可以保护您的 DataSync 资源，因为您不会无意中删除访问资源的权限。

有关支持服务相关角色的其他服务的信息，请参阅与 IAM 配合使用的AWS 服务，并在服务相关角色列中查找标有 “是” 的服务。选择是和链接，查看该服务的服务关联角色文档。

的服务相关角色权限 DataSync

DataSync 使用名为的服务相关角色 AWSServiceRoleForDataSync— DataSync 允许执行传输任务执行的基本操作，包括从中 AWS Secrets Manager读取密钥以及创建 CloudWatch 日志组和事件。

AWSServiceRoleForDataSync 服务相关角色信任以下服务来代入该角色：

- `datasync.amazonaws.com`

服务相关角色使用名为的 AWS 托管策略 AWSDataSyncServiceRolePolicy，该策略 DataSync 允许对指定资源完成以下操作：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [{
            "Sid": "DataSyncCloudWatchLogCreateAccess",
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream"
            ],
            "Resource": [
                "arn:*:logs:*:*:log-group:/aws/datasync*"
            ]
        },
        {
            "Sid": "DataSyncCloudWatchLogStreamUpdateAccess",
            "Effect": "Allow",
            "Action": [
                "logs:PutLogEvents"
            ],
            "Resource": [
                "arn:*:logs:*:*:log-group:/aws/datasync*:log-stream:*"
            ]
        },
        {
            "Sid": "DataSyncSecretsManagerReadAccess",
```

```
            "Effect": "Allow",
            "Action": [
                "secretsmanager:DescribeSecret",
                "secretsmanager:GetSecretValue"
            ],
            "Resource": [
                "arn:*:secretsmanager:*:*:secret:aws-datasync!*"
            ],
            "Condition": {
                "StringEquals": {
                    "secretsmanager:ResourceTag/
aws:secretsmanager:owningService": "aws-datasync",
                    "aws:ResourceAccount": "${aws:PrincipalAccount}"
                }
            }
        }
    ]
}
```

您必须配置使用户、组或角色能够创建、编辑或删除服务相关角色的权限。有关更多信息，请参阅《IAM 用户指南》中的服务相关角色权限。

为创建服务相关角色 DataSync

您无需手动创建服务关联角色。当您在 AWS 管理控制台、或 AWS API 中创建 DataSync 任务时，DataSync会为您创建服务相关角色。 AWS CLI

在 AWS CLI 或 AWS API 中，您可以使用服务名称创建服务相关角色。`datasync.amazonaws.com`有关更多信息，请参阅 IAM 用户指南 中的创建服务相关角色。如果您删除了此服务相关角色，可以使用同样的过程再次创建角色。

如果您删除该服务关联角色，然后需要再次创建，您可以使用相同流程在账户中重新创建此角色。创建 DataSync 任务时， DataSync 会再次为您创建服务相关角色。

如果删除该服务相关角色，您可以使用相同的 IAM 过程再次创建该角色。

编辑的服务相关角色 DataSync

DataSync 不允许您编辑 AWSServiceRoleForDataSync 服务相关角色。创建服务关联角色后，您将无法更改角色的名称，因为可能有多种实体引用该角色。但是可以使用 IAM 编辑角色描述。有关更多信息，请参阅《IAM 用户指南》中的编辑服务关联角色。

删除的服务相关角色 DataSync

如果不再需要使用某个需要服务关联角色的功能或服务，我们建议您删除该角色。这样就没有未被主动监控或维护的未使用实体。但是，您必须先清除您的服务相关角色，然后才能手动删除它。

清除服务相关角色

必须先删除服务相关角色使用的所有资源，然后才能使用 IAM 删除该角色。

> **Note**
>
> 如果您尝试删除资源时 DataSync 服务正在使用该角色，则删除可能会失败。如果发生这种情况，请等待几分钟后重试。

要删除使用的 DataSync 资源 AWSService RoleForDataSync

1. 删除任务使用的 DataSync 代理（如果有）。
2. 删除任务的位置。
3. 删除任务。

手动删除服务相关角色

使用 IAM 控制台 AWS CLI、或 AWS API 删除 AWSServiceRoleForDataSync服务相关角色。有关更多信息，请参阅《IAM 用户指南》中的删除服务关联角色。

DataSync 服务相关角色支持的区域

DataSync 支持在提供服务的所有区域中使用服务相关角色。有关更多信息，请参阅 AWS 区域和端点。

在创建期间为 DataSync 资源添加标签的权限

某些创建资源 AWS DataSync 的 API 操作允许您在创建资源时指定标签。您可以使用资源标签来实现基于属性的访问权限控制（ABAC）。有关更多信息，请参阅 ABAC 有什么用 AWS？ 在 IAM 用户指南中。

为使用户能够在创建时为资源添加标签，他们必须具有使用创建该资源的操作（如 `datasync:CreateAgent` 或 `datasync:CreateTask`）的权限。如果在资源创建操作中指定了标签，则用户还必须具有使用 `datasync:TagResource` 操作的显式权限。

仅当用户在资源创建操作中应用了标签时，系统才会评估 datasync:TagResource 操作。因此，如果未在此请求中指定任何标签，则拥有创建资源权限 (假定没有标记条件) 的用户无需具备使用 datasync:TagResource 操作的权限。

但是，如果用户不具备使用 datasync:TagResource 操作的权限而又试图创建带标签的资源，则请求将失败。

IAM policy 语句示例

使用以下 IAM 策略声明示例，向创建 DataSync 资源的用户授予TagResource权限。

以下语句允许用户在创建 DataSync 代理时标记代理。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "datasync:TagResource",
            "Resource": "arn:aws:datasync:us-east-1:444455556666:agent/*"
        }
    ]
}
```

以下语句允许用户在创建 DataSync 营业地点时为其添加标签。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "datasync:TagResource",
            "Resource": "arn:aws:datasync:us-east-1:111122223333:location/*"
        }
    ]
}
```

以下语句允许用户在创建 DataSync 任务时对任务进行标记。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "datasync:TagResource",
            "Resource": "arn:aws:datasync:us-east-1:444455556666:task/*"
        }
    ]
}
```

防止跨服务混淆代理

混淆代理问题是一个安全性问题，即不具有某操作执行权限的实体可能会迫使具有更高权限的实体执行该操作。在中 AWS，跨服务模仿可能会导致混乱的副手问题。一个服务（呼叫服务）调用另一项服务（所谓的服务）时，可能会发生跨服务模拟。可以操纵调用服务，使用其权限以在其他情况下该服务不应有访问权限的方式对另一个客户的资源进行操作。为防止这种情况， AWS 提供可帮助您保护所有服务的数据的工具，而这些服务中的服务主体有权限访问账户中的资源。

我们建议在资源策略中使用aws:SourceArn和aws:SourceAccount全局条件上下文密钥来限制为资源 AWS DataSync 提供其他服务的权限。如果同时使用全局条件上下文密钥和包含账户 ID 的 aws:SourceArn 值，则 aws:SourceAccount 值和 aws:SourceArn 值中的账户在同一策略语句中使用时，必须使用相同的账户 ID。如果您只希望将一个资源与跨服务访问相关联，请使用 aws:SourceArn。如果您想该账户中的任何资源与跨服务使用相关联，请使用 aws:SourceAccount。

的值aws:SourceArn必须包含允许担任 IAM 角色 DataSync 的地点 ARN。 DataSync

防范混淆代理问题最有效的方法是使用 aws:SourceArn 键和资源的完整 ARN。如果您不知道此完整 ARN，或者正在指定多个资源，请针对未知部分使用通配符字符 (*)。以下是一些如何执行此操作的示例 DataSync：

- 要将信任策略限制在现有 DataSync 位置，请在策略中包含完整的地点 ARN。 DataSync 只有在处理该特定位置时才会担任 IAM 角色。

- 在为创建 Amazon S3 营业地点时 DataSync，您不知道该营业地点的 ARN。在这种情况下，请对 `aws:SourceArn` 键使用以下格式：`arn:aws:datasync:us-east-2:123456789012:*`。此格式用于验证分区 (aws)、账户 ID 和区域。

以下完整示例显示了如何在信任策略中使用`aws:SourceArn`和`aws:SourceAccount`全局条件上下文密钥来防止出现混淆的代理问题 DataSync。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "datasync.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                "aws:SourceAccount": "123456789012"
                },
                "ArnLike": {
                "aws:SourceArn": "arn:aws:datasync:us-east-2:123456789012:*"
                }
            }
        }
    ]
}
```

有关显示如何使用`aws:SourceArn`和`aws:SourceAccount`全局条件上下文密钥的更多策略示例 DataSync，请参阅以下主题：

- 创建允许 DataSync 访问您的 Amazon S3 存储桶的信任关系
- 配置 IAM 角色以访问您的 Amazon S3 存储桶

的合规性验证 AWS DataSync

要了解某个 AWS 服务是否在特定合规性计划范围内，请参阅合规性计划范围内的 AWS 服务，然后选择您感兴趣的合规性计划。有关常规信息，请参阅 AWS 合规性计划、、。

您可以使用 AWS Artifact 下载第三方审计报告。有关更多信息，请参阅在 AWS Artifact 中下载报告、。

您在使用 AWS 服务 时的合规性责任由您的数据的敏感性、您公司的合规性目标以及适用的法律法规决定。有关您在使用 AWS 服务时的合规责任的更多信息，请参阅 AWS 安全性文档。

AWS DataSync 中的恢复功能

AWS全球基础架构围绕AWS区域和可用区构建。AWS区域提供多个在物理上独立且隔离的可用区，这些可用区通过延迟低、吞吐量高且冗余性高的网络连接在一起。利用可用区，您可以设计和操作在可用区之间无中断地自动实现故障转移的应用程序和数据库。与传统的单个或多个数据中心基础架构相比，可用区具有更高的可用性、容错性和可扩展性。

> **Note**
>
> 如果在运行 DataSync 任务时，您要将数据迁入或从中迁出的可用区发生故障，则该任务也将失败。

有关 AWS 区域和可用区的更多信息，请参阅 AWS 全球基础设施。

AWS DataSync 中的基础结构安全性

作为一项托管式服务，AWS DataSync 受 AWS 全球网络安全保护。有关 AWS 安全服务以及 AWS 如何保护基础架构的信息，请参阅 AWS 云安全。要按照基础设施安全最佳实践设计您的 AWS 环境，请参阅《安全性支柱 AWS Well-Architected Framework》中的 基础设施保护。

您可以使用 AWS 发布的 API 调用，通过网络访问 DataSync。客户端必须支持以下内容：

- 传输层安全性协议（TLS）。我们要求使用 TLS 1.2，建议使用 TLS 1.3。
- 具有完全向前保密（PFS）的密码套件，例如 DHE（临时 Diffie-Hellman）或 ECDHE（临时椭圆曲线 Diffie-Hellman）。大多数现代系统（如 Java 7 及更高版本）都支持这些模式。

使用 Secrets Manager 保护存储位置凭证

DataSync 使用位置来访问位于本地、其他云端或内部的存储资源 AWS。某些位置类型要求您提供凭据（例如，访问密钥和私有密钥、用户名和密码、Kerberos keytab、SAS 令牌等），以便在存储系统中进行身份验证。当您创建需要凭据进行身份验证的 DataSync 位置时，可以使用 AWS Secrets Manager (Secrets Manager) 来存储您的凭据的密钥。有以下选项可用：

- ManagedSecretConfig: DataSync-托管密钥配置。使用使用默认密钥加密的 DataSync 服务管理密钥将密钥存储在 Secrets Manager 中。
- CmkSecretConfig: DataSync-使用客户管理密钥 (CMK) 配置的托管密钥。使用使用您管理的密钥加密的 DataSync 服务管理密钥将密 AWS KMS 钥存储在 Secrets Manager 中。
- CustomSecretConfig：客户管理的密钥配置。使用您创建和管理的密钥和密钥将密钥存储在 Secrets Manager 中。 DataSync 使用您提供的 IAM 角色访问此密钥。

在所有情况下，Secrets Manager 密钥都存储在您的账户中，允许您根据需要独立于 DataSync 服务更新密钥。由创建和管理的密钥 DataSync 具有前缀`aws-datasync`。

只有当您在以外的服务之外创建机密 DataSync 或从服务管理的密钥发出 API 调用时，才需要支付使用密钥的费用。 DataSync

使用通过默认密钥加密的服务托管密钥

创建 DataSync 位置时，您只需提供密钥字符串即可。 DataSync在 Secrets Manager 中创建密钥资源来存储你提供的密钥，并使用你账户的默认 Secrets Manager KMS 密钥对密钥进行加密。你可以直接在 Secrets Manager 中更改密钥值，也可以使用 DataSync 控制台 AWS CLI、或 SDK 更新位置。当您删除位置资源或将其更新为使用自定义密钥时， DataSync会自动删除该密钥资源。

> **Note**
>
> 要在 Secrets Manager 中创建、修改和删除机密资源， DataSync 必须具有相应的权限。有关更多信息，请参阅适用于 DataSync 的AWS 托管策略。

使用使用自定义 AWS KMS 密钥加密的服务管理密钥

创建 DataSync 位置时，您需要提供密钥的密钥和 ARN。 AWS KMS DataSync 自动在 Secrets Manager 中创建密钥资源来存储您提供的密钥，并使用您的密 AWS KMS 钥对其进行加密。你可以直

接在 Secrets Manager 中更改密钥值，也可以使用 DataSync 控制台 AWS CLI、或 SDK 更新位置。当您删除位置资源或将其更新为使用自定义密钥时， DataSync 会自动删除该密钥资源。

> **Note**
>
> 您的 AWS KMS 密钥必须使用密ENCRYPT_DECRYPT钥类型的对称加密。有关更多信息，请参阅《AWS Secrets Manager 用户指南》中的选择 AWS Key Management Service 密钥。

要在 Secrets Manager 中创建、修改和删除机密资源， DataSync 必须具有相应的权限。有关更多信息，请参阅 AWS 托管策略：AWSDataSyncFullAccess。

除了使用正确的 DataSync 托管策略外，您还需要以下权限：

```
{
    "Sid": "DataSyncKmsPermissions",
    "Effect": "Allow",
    "Action": [
        "kms:Encrypt",
        "kms:GenerateDataKey",
        "kms:Decrypt"
    ],
    "Resource": "your-kms-key-arn",
    "Condition": {
        "StringLike": {
            "kms:ViaService": "secretsmanager.*.amazonaws.com"
        }
    }
}
```

*your-kms-key-arn*替换为您的 KMS 密钥 ARN。

要检索和解密密钥值，请 DataSync 使用服务关联角色 (SLR) 访问您的密钥。 AWS KMS 为确保 DataSync 可以使用您的 KMS 密钥，请在密钥的政策声明中添加以下内容：

```
{
    "Sid": "Allow DataSync to use the key for decryption",
    "Effect": "Allow",
    "Principal": {
            "AWS": "arn:aws:iam::111122223333:role/aws-service-role/
datasync.amazonaws.com/AWSServiceRoleForDataSync"
```

```
    },
    "Action": "kms:Decrypt",
    "Resource": "*"
}
```

*111122223333*用您的 AWS 账户 身份证替换。

使用您管理的密钥

在创建 DataSync 位置之前，请在 Secrets Manager 中创建一个密钥。密钥值只能包含明文形式的密钥字符串本身。创建 DataSync 位置时，您需要提供密钥的 ARN 以及用于访问您的密钥和用于加密您的密 AWS KMS 钥的 IAM 角色。 DataSync 要创建具有适当权限的 IAM 角色，请执行以下操作：

1. 使用 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。
2. 在左侧导航窗格的访问管理下，选择角色，然后选择创建角色。
3. 在选择可信实体页面中，在可信实体类型下选择AWS 服务。
4. 对于 “用例”，请DataSync从下拉列表中进行选择。选择下一步。
5. 在添加权限页面上，选择下一步。输入角色的名称，然后选择创建角色。
6. 在角色页面上，搜索您刚刚创建的角色并选择其名称。
7. 在角色的详细信息页面上，选择权限选项卡。选择添加权限，然后选择创建内联策略。
8. 选择 JSON 选项卡，然后将以下权限添加到策略编辑器中。

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret"
            ],
            "Resource": "arn:aws:secretsmanager:us-
east-1:111122223333:secret:your-secret-name"
        }
    ]
}
```

*your-secret-name*替换为你的 Secrets Manager 密钥的名称。

9. 选择下一步。输入策略名称，然后选择创建策略。
10. （推荐）为防止出现跨服务混淆代理问题，请执行以下操作：

    a. 在角色的详细信息页面上，选择信任关系选项卡。选择编辑信任策略。

    b. 使用以下示例更新信任策略，其中包括 `aws:SourceArn` 和 `aws:SourceAccount` 全局条件上下文键：

    JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "datasync.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                "aws:SourceAccount": "111122223333"
                },
                "ArnLike": {
                "aws:SourceArn": "arn:aws:datasync:us-
east-1:111122223333:*"
                }
            }
        }
    ]
}
```

    c. 选择更新策略。

您可以在创建位置时指定此角色。如果您的密钥使用客户管理的密 AWS KMS 钥进行加密，则您还需要更新密钥的策略，以允许您在上一个过程中创建的角色进行访问。要更新政策，请在 AWS KMS 密钥的政策声明中添加以下内容：

```
{
    "Sid": "Allow DataSync use of the key",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam:111122223333:role/your-role-name”
    },
    "Action": "kms:Decrypt",
    "Resource": "*"
}
```

*111122223333*替换为您的 AWS 账户 ID，*your-role-name*以及您在上一个过程中创建的 IAM 角色的名称。

> **Note**
>
> 当你在 Secret AWS 账户 s Manager 中存储密钥时，会产生费用。有关定价的信息，请参阅 AWS Secrets Manager 定价。

AWS DataSync 限额

了解使用 AWS DataSync 时的资源配额和限制。

存储系统、文件和对象限制

下表介绍了 DataSync 在处理存储系统、文件和对象时的限制。

描述	限制
文件路径最大总长度	4,096 个字节
文件路径组成部分（文件名、目录或子目录）的最大长度	255 个字节
Windows 域最大长度	253 个字符
服务器主机名最大长度	255 个字符
Amazon S3 对象名称最大长度	1,024 个 UTF-8 字符

DataSync 配额

下表介绍了特定 AWS 账户 和区域中 DataSync 资源的配额。

资源	限额	可调整
您可以创建的最大任务数量	100	是
（增强模式任务）每次任务执行时，DataSync 可以处理的源对象和目标对象的最大数量 有关更多信息，请参阅 如何 DataSync 传输文件、对象和目录。	几乎无限制	不适用
（基本模式任务）在本地存储、自我管理存储或其他云存储和 AWS 存储服务之间，每次任务执	5000 万	是

资源	限额	可调整
行时，DataSync 可以处理的源和目标文件、对象或目录的最大数量 有关更多信息，请参阅 如何 DataSync 传输文件、对象和目录。	**Important** 请记住有关此配额的以下事项： • 如果您传输带有前缀的 Amazon S3 对象，则前缀将被视为目录并计入配额。例如，DataSync 会将 `s3://bucket/foo/bar.txt` 视为两个目录（`./` 和 `./foo/`）和一个对象 (`bar.txt`)。 • 如果您的任务要处理超过 2,000 万个文件、对象或目录，则确保为 DataSync 代理分配至少 64GB 的 RAM。有关更多信息，请参阅 DataSync 传输的代理要求。	**Tip** 您可以使用“包含”和“排除”筛选器创建以特定目录为重点的任务，而不是请求增加。有关更多信息，请参阅 筛选由 DataSync 传输的数据。

资源	限额	可调整
（基本模式任务）在 AWS 存储服务之间，每次任务执行时，DataSync 可以处理的源和目标文件、对象和目录的最大数量 有关更多信息，请参阅 如何 DataSync 传输文件、对象和目录。	2500 万 **Important** 如果您传输带有前缀的 Amazon S3 对象，则前缀将被视为目录并计入配额。例如，DataSync 会将 `s3://bucket/foo/bar.txt` 视为两个目录（`./` 和 `./foo/`）和一个对象 (`bar.txt`)。	是 **Tip** 您可以使用“包含”和“排除”筛选器创建以特定目录为重点的任务，而不是请求增加。有关更多信息，请参阅 筛选由 DataSync 传输的数据。

资源	限额	可调整
每个任务的最大吞吐量（适用于使用 DataSync 代理的传输）	10 Gbps	否
每个任务的最大吞吐量（适用于不使用 DataSync 代理的传输）	5 Gbps	否
任务筛选器中可包含的最大字符数	102,400 个字符 **Note** 如果您使用的是 DataSync 控制台，则此限制包括“包含”和“排除”模式中组合的所有字符。	否
单个任务排队执行的最大次数	50	不可以
并发增强模式任务执行的最大次数	120	否
任务执行历史记录的最长保留天数	30	否
增强模式任务清单文件的最大大小	20GB	否

请求提高限额

您可以请求增加一些 DataSync 限额。增加限额不会立即获得批准，可能需要几天才能生效。

要请求提高限额

1. 访问 https://console.aws.amazon.com/servicequotas/，打开服务配额控制台。
2. 在导航窗格中，选择 AWS 服务，然后选择 AWS DataSync。
3. 选择要增加的配额，然后选择请求增加账户级配额。
4. 输入您想要的配额总量，然后选择请求。

   如果您需要增加一个不同的限额，请单独填写一份申请。

疑难解答 AWS DataSync 问题

使用以下信息对 AWS DataSync 问题和错误进行故障排除。

主题

- 对 DataSync 代理问题进行故障排除
- 对 DataSync 位置问题进行故障排除
- 对 DataSync 任务问题进行故障排除
- 解决数据验证的问题
- 使用 S3 存储成本高于预期进行故障排除 DataSync

对 DataSync 代理问题进行故障排除

使用以下信息来帮助您解决与 AWS DataSync 代理有关的问题。其中一些问题可能包括：

- 连接到 Amazon EC2 代理的本地控制台时出错
- 未能检索代理的激活密钥
- 使用 VPC 服务端点激活代理时出现问题
- 发现代理处于离线状态

如何连接到 Amazon EC2 代理的本地控制台？

要连接到 Amazon EC2 代理的本地控制台，必须使用 SSH。请确保您的 EC2 实例安全组允许通过 SSH 进行访问（TCP 端口 22）。

在终端中，使用以下 ssh 命令连接到该实例：

```
ssh -i /path/key-pair-name.pem instance-user-name@instance-public-ip-address
```

- 对于*/path/key-pair-name*，请指定连接到您的实例所需的私有密钥的路径和文件名 (.pem)。
- 对于 *instance-user-name*，请指定 admin。
- 对于*instance-public-ip-address*，请指定您的实例的公有 IP 地址。

无法检索代理激活密钥错误是什么意思？

激活 DataSync 代理时，代理会连接到您指定的服务端点以请求激活密钥。此错误可能意味着您的网络安全设置阻止了连接。

要采取的操作

如果您正在使用虚拟私有云（VPC）服务端点，请验证您的安全组设置是否允许代理连接到 VPC 端点。有关所需端口的信息，请参阅VPC 或 FIPS VPC 服务端点的网络要求。

如果您使用的是公共或联邦信息处理标准（FIPS）端点，请检查您的防火墙和路由器设置，以允许您的代理连接到该端点。有关信息，请参阅公共或 FIPS 服务端点的网络要求。

我仍然无法使用 VPC 服务端点激活代理

如果您在使用 VPC 服务终端节点激活 DataSync 代理时仍然遇到问题，请参阅 我不知道我的代理出了什么问题。有人能帮我吗？

如果我的代理处于离线状态，我该怎么办？

您的 DataSync 代理可能出于多种原因而处于离线状态，但您也许可以将其恢复在线。在删除代理并创建新代理之前，请仔细阅读以下清单，以帮助您了解可能发生的情况。

- 联系您的备份团队 — 如果您的代理由于其虚拟机 (VM) 已从快照或备份中恢复而处于离线状态，则可能需要更换代理。
- 检查代理的虚拟机或 Amazon EC2 实例是否已关闭 — 根据您使用的代理类型，如果虚拟机或 EC2 实例已关闭，请尝试将其重新打开。再次开启后，测试您的代理与的网络连接 AWS。
- 验证您的代理是否满足最低硬件要求 — 您的代理可能处于离线状态，因为自代理激活以来，其虚拟机或 EC2 实例配置被意外更改。例如，如果您的虚拟机不再具有所需的最低内存或空间，则代理可能会显示为离线。有关更多信息，请参阅 对 AWS DataSync 代理人的要求。
- 等待代理相关的软件更新完成 — 您的代理可能会在AWS提供的软件更新后短暂脱机。如果您认为这是代理离线的原因，请稍等片刻，然后检查代理是否恢复联机。
- 检查您的 VPC 服务终端节点设置 — 如果您的离线代理使用公共服务终端节点，并且也在您为其创建 VPC 服务终端节点的同一 VPC 中 DataSync，则可能需要禁用该 VPC 终端节点的私有 DNS 支持。

如果这些似乎都不是代理离线的原因，则可能需要更换代理。

我不知道我的代理出了什么问题。有人能帮我吗？

您可以允许 AWS 支持 访问您的 DataSync 代理并帮助解决与代理相关的问题。您必须通过代理的本地控制台启用此访问权限。

向您的代理提供 支持 访问权限

1. 登录到代理的本地控制台。
2. 在提示符处，输入**5**以打开命令提示符（fo VMware VMs r，使用**6**）。
3. 输入 **h** 以打开 AVAILABLE COMMANDS 窗口。
4. 在 AVAILABLE COMMANDS（可用命令）窗口中，输入以下内容连接 支持：

   ```
   open-support-channel
   ```

   如果您使用带有 VPC 端点的代理，则必须为支持通道提供 VPC 端点 IP 地址，如下所示：

   ```
   open-support-channel vpc-ip-address
   ```

   您的防火墙必须允许入站 TCP 端口 22 以启动针对 AWS的支持通道。当您连接到时 支持，会为您 DataSync 分配一个支持号码。请记下您的支持编号。

   > **Note**
   >
   > 信道号不是 TCP/UDP 端口号。相反，它建立到服务器的 SSH (TCP 22) 连接，并为该连接提供支持通道。

5. 建立支持渠道后，请向提供您的支持服务号码， 支持 以便他们可以提供故障排除帮助。
6. 在支持会话完成后，按下 **Enter** 以结束。
7. 输入**exit**以注销 DataSync 本地控制台。
8. 按照提示操作退出本地控制台。

对 DataSync 位置问题进行故障排除

使用以下信息来帮助您解决 AWS DataSync 营业地点问题。其中一些问题可能包括：

- NFS 位置的权限和挂载错误
- 文件所有权问题

- 访问使用 Kerberos 身份验证的 SMB 位置时出现问题
- 对象存储的权限与访问问题，例如 Amazon S3 和 Microsoft Azure Blob 位置

我的任务失败，出现 NFS 权限被拒绝错误

如果用 `root_squash` 或 `all_squash` 配置 NFS 服务器，并且您的文件没有所有读取访问权限，则可能会显示“权限被拒绝”错误消息。

要采取的操作

要修复此问题，请用 `no_root_squash` 配置 NFS 导出，或者确保要传输的所有文件的权限允许所有用户进行读取访问。

DataSync 要访问目录，还必须启用所有执行访问权限。要确保可以挂载目录，首先连接到与代理具有相同网络配置的任何计算机。然后运行以下 CLI 命令：

```
mount -t nfs -o nfsvers=<your-nfs-server-version> <your-nfs-server-name>:<nfs-export-path-you-specified> <new-test-folder-on-your-computer>
```

如果问题仍未解决，请联系AWS 支持 中心。

我的任务失败，出现 NFS 挂载错误

在运行涉及 NFS 文件服务器位置的 DataSync 任务时，您可能会看到以下错误：

Task failed to access location loc-1111222233334444a: x40016: mount.nfs: Connection timed out

要采取的操作

在错误得到解决之前，请执行以下操作：

1. 确保您在 DataSync所在位置指定的 NFS 文件服务器和导出文件有效。如果无效，请删除您的位置和任务，然后创建使用有效 NFS 文件服务器和导出的新位置和任务。有关更多信息，请参阅 使用控制 DataSync 台。
2. 检查代理和 NFS 文件服务器之间的防火墙配置。有关更多信息，请参阅 本地、自我管理和其他云存储的网络要求。
3. 确保代理可以访问 NFS 文件服务器并挂载导出。有关更多信息，请参阅 提供 DataSync 对 NFS 文件服务器的访问权限。

4. 如果您仍然看到错误，请使用打开支持频道 支持。有关更多信息，请参阅 我不知道我的代理出了什么问题。有人能帮我吗？。

我的任务失败，出现 Amazon EFS 挂载错误

在运行涉及 Amazon EFS 位置的 DataSync 任务时，您可能会看到以下错误：

Task failed to access location loc-1111222233334444a: x40016: Failed to connect to EFS mount target with IP: 10.10.1.0.

如果您使用自己的位置配置的 Amazon EFS 文件系统的挂载路径被更新或删除，则可能会发生这种情况。 DataSync 不知道文件系统中的这些变化。

要采取的操作

删除您的位置和任务，并使用新的挂载路径创建新的 Amazon EFS 位置。

NFS 传输未保留文件所有权

传输完成后，您可能会注意到 DataSync 目标位置的文件具有不同的用户 IDs (UIDs) 或组 IDs (GIDs)，而不是源位置的相同文件。例如，目标位置中文件的 UID 可能是 65534、99 或 nobody。

如果传输中涉及的文件系统使用 NFS 版本 4 ID 映射（该功能 DataSync 不支持），则可能会发生这种情况。

要采取的操作

您有几种方法可以解决此问题：

- 为使用 NFS 版本 3（而不是版本 4）的文件系统创建一个新位置。
- 在文件系统上禁用 NFS 版本 4 ID 映射。

重试传输。无论哪种方法都应解决问题。

我的任务无法访问使用 Kerberos 的 SMB 位置

DataSync 使用 Kerberos 身份验证的 SMB 位置出现错误通常与您的位置与 Kerber os 配置不匹配有关。也可能存在网络问题。

无法访问位置

以下错误表示 SMB 位置或 Kerberos 设置可能存在配置问题：

```
Task failed to access location
```

验证以下各项：

- 您为所在位置指定的 SMB 文件服务器是一个域名。对于 Kerberos，无法指定文件服务器的 IP 地址。
- 您为所在位置指定的 Kerberos 主体与用于创建 Kerberos 密钥表（keytab）文件的主体相匹配。主体名称区分大小写。
- 自创建 keytab 文件以来，Kerberos 主体的映射用户密码未出现更改。如果密码更改（由于密码轮换或其他原因），则任务执行可能会失败，并出现以下错误消息：

  Task failed to access location loc-1111222233334444a: x40015: kinit: Preauthentication failed while getting initial credentials

无法连接到 KDC 领域

以下错误表示存在网络问题：

```
kinit: Cannot contact any KDC for realm 'MYDOMAIN.ORG' while getting initial
 credentials"
```

验证以下各项：

- 您提供的 Kerberos 配置文件 (`krb5.conf`) 包含有关您的 K DataSync erberos 领域的正确信息。有关示例 `krb5.conf` 文件，请参阅 Kerberos 身份验证的先决条件。
- Kerberos 密钥分配中心（KDC）的服务器端口已打开。KDC 端口通常为 TCP 端口 88。
- 网络上的 DNS 配置。

我的任务因 input/output 错误而失败

如果您的存储系统未能通过 DataSync 代理 I/O 请求，则会收到一条 input/output 错误消息。常见原因包括服务器磁盘故障、防火墙配置更改或网络路由器故障。

如果错误涉及 NFS 文件服务器或 Hadoop Distributed File System（HDFS）集群，请跟随以下步骤来解决错误。

采取的操作（NFS）

首先检查 NFS 文件服务器的日志和指标，以确定问题是否始于 NFS 服务器。如果是，请解决该问题。

接下来，检查网络配置是否尚未更改。要检查 NFS 文件服务器的配置是否正确以及是否 DataSync 可以访问它，请执行以下操作：

1. 在与代理相同的网络子网上设置另一个 NFS 客户端。
2. 在客户端上装载您的共享。
3. 验证客户端可以成功地读取和写入共享。

采取的操作（HDFS）

在解决错误之前，请执行以下操作：

1. 确保您的 HDFS 集群允许您的 DataSync 代理与集群 NameNode 和 DataNode 端口通信。

   在大多数集群中，您可以在以下配置文件中找到集群使用的端口号：

   - 要找到 NameNode 端口，请在`fs.default`或`fs.default.name`属性下的`core-site.xml`文件中查找（取决于 Hadoop 发行版）。
   - 要找到端 DataNode 口，请在`dfs.datanode.address`属性下的`hdfs-site.xml`文件中查找。
2. 在您的 `hdfs-site.xml` 文件中，验证您的 `dfs.data.transfer.protection` 属性是否只有一个值。例如：

```
<property>
   <name>dfs.data.transfer.protection</name>
   <value>privacy</value>
</property>
```

错误：`FsS3UnableToConnectToEndpoint`

DataSync 无法连接到你的 Amazon S3 地点。这可能意味着无法访问该位置的 S3 存储桶，或者该位置的配置不正确。

在解决问题之前，请执行以下操作：

- 检查是否 DataSync 可以访问您的 S3 存储桶。
- 使用 DataSync 控制台或 DescribeLocationS3 操作确保位置配置正确。

错误：FsS3HeadBucketFailed

DataSync 无法访问您要传入或传出的 S3 存储桶。使用 Amazon S3 HeadBucket操作检查是否 DataSync有权访问存储桶。如果您需要调整权限，请参阅 提供 DataSync 对 S3 存储桶的访问权限。

任务因 Unable to list Azure Blobs on the volume root 错误而失败

如果您的 DataSync 转移任务因`Unable to list Azure Blobs on the volume root`错误而失败，则可能是您的共享访问签名 (SAS) 令牌或Azure存储帐户的网络存在问题。

要采取的操作

尝试以下操作并再次运行任务，直到问题得到解决：

- 确保您的 SAS 令牌具有访问您的 Microsoft Azure Blob Storage 的相应权限。
- 如果您在中运行 DataSync 代理Azure，请将您的存储帐户配置为允许从代理所在的虚拟网络进行访问。
- 如果您在 Amazon EC2 上运行代理，请将 Azure 存储防火墙配置为允许从代理的公有 IP 地址进行访问。

有关如何配置 Azure 存储账户网络的信息，请参阅Azure Blob Storage 文档。

错误：FsAzureBlobVolRootListBlobsFailed

DataSync 用于访问您的共享访问签名 (SAS) 令牌Microsoft Azure Blob Storage没有 “列表” 权限。

要解决问题，请使用具有列表权限的令牌更新您的位置，然后尝试重新运行任务。

错误：SrcLocHitAccess

DataSync 无法访问您的来源位置。检查是否 DataSync 有权访问该地点，然后尝试再次运行您的任务。

错误：SyncTaskErrorLocationNotAdded

DataSync 无法访问您的位置。检查是否 DataSync 有权访问该地点，然后尝试再次运行您的任务。

错误：S3 location creation failed with (InvalidRequestException) when calling the CreateLocationS3 operation

此错误可能与 IAM 权限、Amazon S3 存储桶策略、 AWS KMS 权限或其他权限问题有关。如果遇到此错误，请使用以下信息进行故障排除：

- 《Amazon Simple Storage Service 用户指南》中的排查 Amazon S3 中的拒绝访问（403 禁止）错误。
- 如何解决来自 Amazon S3 的 403 访问被拒绝错误？ on AWS re:Post

S3 源位置的任务失败并显示 HeadObject 或 GetObjectTagging 错误

与 HeadObject 或 GetObjectTagging 相关的错误

如果您要 IDs 从 S3 存储桶传输具有特定版本的对象，则可能会看到与HeadObject或相关的错误GetObjectTagging。例如，以下是与 GetObjectTagging 相关的错误：

```
[WARN] Failed to read metadata for file /picture1.png (versionId: 111111): S3 Get
 Object Tagging Failed
[ERROR] S3 Exception: op=GetObjectTagging photos/picture1.png, code=403, type=15,
 exception=AccessDenied,
msg=Access Denied req-hdrs: content-type=application/xml, x-amz-api-version=2006-03-01
 rsp-hdrs: content-type=application/xml,
date=Wed, 07 Feb 2024 20:16:14 GMT, server=AmazonS3, transfer-encoding=chunked,
x-amz-id-2=IOWQ4fDEXAMPLEQM+ey7N9WgVhSnQ6JEXAMPLEZb7hSQDASK+Jd1vEXAMPLEa3Km, x-amz-
request-id=79104EXAMPLEB723
```

如果您看到其中任何一个错误，请验证用于访问您的 DataSync S3 源位置的 IAM 角色是否具有以下权限：

- s3:GetObjectVersion
- s3:GetObjectVersionTagging

如果您需要更新角色，使其具有这些权限，请参阅 为创建 IAM 角色 DataSync 以访问您的 Amazon S3 位置。

对 DataSync 任务问题进行故障排除

使用以下信息来帮助您解决与 AWS DataSync 任务和任务执行有关的问题。这些问题可能包括任务设置问题、任务执行卡滞以及数据无法按预期传输。

错误：SyncOption 值无效。选项： TransferMode，PreserveDeletedFiles，值：全部，移除。

当您创建或编辑 DataSync 任务并选择 “传输所有数据” 选项并取消选择 “保留已删除文件” 选项时，就会发生此错误。

当您传输所有数据时， DataSync 不会扫描您的目标位置，也不知道要删除什么。

任务执行失败并EniNotFound出现错误

如果您删除虚拟私有云（VPC）中任务的网络接口之一，则会发生此错误。如果您的任务已安排或已排队，则如果缺少传输数据所需的网络接口，则该任务将失败。

要采取的操作

要解决此问题，您具有以下选项：

- 手动重启任务。执行此操作时， DataSync 将创建运行任务所需的所有缺失网络接口。
- 如果您需要清理 VPC 中的资源，请确保不要删除与您仍在使用的 DataSync 任务相关的网络接口。

  要查看分配给任务的网络接口，请执行以下操作之一：

  - 使用 DescribeTask 操作。您可以在 `SourceNetworkInterfaceArns` 和 `DestinationNetworkInterfaceArns` 响应元素中查看网络接口。
  - 在 Amazon EC2 控制台中，搜索您的任务 ID（例如 `task-f012345678abcdef0`）以查找其网络接口。
- 考虑不要自动运行任务。这可能包括禁用任务队列或调度（通过 DataSync 或自定义自动化）。

任务执行失败，出现无法分配内存错误

当您的 DataSync 任务因无法分配内存错误而失败时，这可能意味着一些不同的含义。

要采取的操作

请尝试以下操作，直到问题不再出现为止：

- 如果传输涉及代理，请确保代理符合虚拟机（VM）或 Amazon EC2 实例的要求。
- 使用筛选器将传输分成多个任务。您尝试传输的文件或对象可能超出了一项 DataSync 任务所能处理的范围。
- 如果问题依然存在，请联系 支持。

`Input/Output error`对于 ONTAP 文件 FSx 系统，任务失败

当您的 DataSync 任务在使用 for ONTAP 文件系统传输数据`Input/Output error`时失败时，可能是由于以下一个或多个问题所致。 FSx

FSx 适用于 ONTAP 的卷已达到其最大文件容量

卷上可用的索引节点或文件指针数量用尽时，就会出现此错误。

要采取的行动

首先，查看该卷的最大文件容量。然后，通过增加索引节点数量或增大存储容量来增加卷的文件容量。有关更多信息，请参阅《f or ONTAP 用户指南》FSx 中的增加卷的最大文件容量。

FSx 适用于 ONTAP 的卷已用完可用存储容量

如果卷上没有可用存储容量，就会出现此错误。

要采取的行动

首先，确定卷的可用存储容量？ 然后增加卷的存储容量 有关更多信息，请参阅《f or ONTAP 用户指南》中的增加卷FSx 的存储容量。

> **Note**
>
> 要在需要时自动增加卷的存储容量，请参阅《FSx 适用于 ONTAP 的用户指南》中的 “使用卷自动调整大小”。

f FSx or ONTAP 目录已达到每个目录中可以存储的最大文件数

当达到每个目录中可存储的最大文件数时，就会出现此错误。

要采取的操作

增加最大目录大小以支持更大的目录。有关更多信息，请参阅AWS 规范性指南中的 FSx 使用 ONTAP 最大目录大小的最佳实践。

DataSync 任务执行产生了过多的读写并发性，占用了文件系统吞吐容量的很大一部分

当 DataSync 任务执行消耗了文件系统的可用吞吐量过多时，就会发生此错误。

要采取的行动

首先，使用以下方法确定任务执行是否消耗了过多的文件系统吞吐能力：

- 使用可用 CloudWatch 指标监控文件系统的性能。有关更多信息，请参阅《适用于 ONTAP 的用户指南》FSx 中的监控文件系统指标。
- 在 Amazon FSx 控制台中监控文件系统是否存在文件服务器性能警告。有关更多信息，请参阅《适用于 ONTAP 的用户指南》FSx 中的性能警告和建议。

然后，通过执行以下操作之一，确保任务未用尽文件系统的所有可用吞吐能力：

- 将任务执行的带宽限制设置 FSx 为小于 ONTAP 文件系统的预配置吞吐量容量。有关更多信息，请参阅 为 AWS DataSync 任务设置带宽限制。
- 提高文件系统的预置吞吐能力。有关更多信息，请参阅 ONTAP 用户指南中的FSx 更新吞吐量容量。

任务失Connection Reset by peer败并显示 ONTAP 文件系统的Host is down消息 FSx

如果您的 DataSync 任务在使用 for ONTAP 文件系统传输数据时失败并Host is down显示Connection Reset by peer或消息，则可能是由于以下一个或多个问题所致： FSx

- 在任务执行期间，文件系统的 SMB 服务器重启或以其他方式断开连接。
- 任务执行期间，文件系统从主服务器（和 IP 地址）故障转移到辅助服务器。 DataSync 不支持在任务执行期间故障转移到辅助 IP 地址。

  FSx 对于 ONTAP 文件系统在以下事件期间故障切换到辅助服务器和 IP 地址：

  - 主服务器不可用。

- 主服务器的可用区变为不可用（对于多可用区文件系统而言）。
- 在用户启动的吞吐能力更改期间。
- 在文件系统的定期维护时段内。

有关更多信息，请参阅《FSx 适用于 ONTAP 的用户指南》 FSx 中的 ONTAP 故障转移流程。

要采取的操作

重新启动任务。

任务执行处于启动状态，但似乎什么也没发生

您的 DataSync 任务可能会停留在 “启动” 状态，这通常是因为代理已关闭电源或已断开网络连接。

要采取的操作

确保您的代理的状态为在线。如果代理处于离线状态，请确保其已开机。

如果代理已打开电源并且任务仍停滞在正在启动状态，则很可能是您的代理和 AWS之间出现网络连接问题。有关如何测试网络连接的信息，请参阅验证您的代理与 DataSync 服务的连接。

如果您仍遇到此问题，请参阅 我不知道我的代理出了什么问题。有人能帮我吗？。

任务执行似乎卡在准备状态

DataSync 传输任务处于 “准备中” 状态的时间取决于传输源和目标中的数据量以及这些存储系统的性能。

任务启动时， DataSync 执行递归目录列表以发现源和目标中的所有文件、对象、目录和元数据。DataSync 使用这些列表来识别存储系统之间的差异并确定要复制的内容。此过程可能耗时数分钟甚至数小时。

要采取的操作

您无需做任何事情。只需继续等待直至任务状态变为正在传输。如果状态仍未变更，请联系 AWS 支持中心。

任务执行在传输完成之前停止

如果您的 DataSync 任务执行提前停止，则您的任务配置中可能包含已禁用的 AWS 账户。 AWS 区域

要采取的操作

执行以下操作以重新运行任务：

1. 检查任务区域的选择加入状态，并确保它们已启用。
2. 再次启动任务。

从 Google 云存储桶传输时，任务执行失败

由于使用 Amazon S3 API DataSync 与 Google 云存储进行通信，因此如果您尝试复制对象标签，则存在一个限制，可能会导致 DataSync 传输失败。您的 CloudWatch 日志中会显示以下与该问题相关的消息：

[警告] 无法读取文件的元数据/*your-bucket*/*your-object*: S3 获取对象标记失败：继续不加标记

为防止出现这种情况，请在配置传输任务设置时取消选择复制对象标签选项。

任务执行的时间戳不匹配

在查看 DataSync 控制台或 Amazon CloudWatch 日志时，您可能会注意到 DataSync 任务执行的开始和结束时间与您在其他监控工具中看到的时间戳不匹配。这是因为控制台和 CloudWatch 日志会考虑任务执行在启动或排队状态下所花费的时间，而其他一些工具则没有。

在比较 DataSync 控制台或 CloudWatch 日志与以下位置之间的执行时间戳时，您可能会注意到这种差异：

- 传输中涉及的文件系统的日志
- DataSync 写入到 Amazon S3 对象的最后修改日期
- 来自 DataSync 代理的网络流量
- 亚马逊 EventBridge 活动

任务执行失败并出现 **NoMem** 错误

您要传输的数据集可能太大 DataSync。如果您看到此错误，请联系 AWS 支持 中心。

任务执行失败并出现 **FsNfsIdMappingEnabled** 错误

DataSync 不支持 NFSv4 ID 映射。要解决此问题，请配置要使用 NFSv3的 NFS 位置。

对象无法传输到 Azure Blob Storage，并出现 `user metadata key` 错误

从 S3 存储桶传输到 Azure Blob Storage 时，可能会出现以下错误：

```
[ERROR] Failed to transfer file /user-metadata/file1: Azure Blob user metadata key must
 be a CSharp identifier
```

这表示 */user-metadata/file1* 包含的用户元数据没有使用有效的 C# 标识符。有关更多信息，请参阅 Microsoft 文档。

目标位置存在 `/.aws-datasync` 文件夹

DataSync `/.aws-datasync`在目标位置创建一个名为的文件夹，以帮助简化数据传输。

虽然 DataSync 通常会在传输后删除此文件夹，但在某些情况下可能不会发生这种情况。

要采取的操作

只要没有正在执行的任务将数据复制到该位置，您就可以随时删除此文件夹。

无法使用 SMB 在两个位置之间传输符号链接

当任务执行完毕后，您会看到以下错误：

```
Transfer and verification completed. Selected files transferred except for files
 skipped due to errors. If no skipped files are listed in Cloud Watch Logs, please
 contact AWS Support for further assistance.
```

在 SMB 存储系统（例如 SMB 文件服务器和 Ama FSx zon for Windows 文件服务器文件系统）之间传输时，您可能会在日志中看到以下警告和错误： CloudWatch

```
[WARN] Failed to read metadata for file /appraiser/symlink: No data available
[ERROR] Failed to read metadata for directory /appraiser/symlink: No data available
```

要采取的操作

DataSync 不支持在这些位置类型之间传输符号链接（或硬链接）。有关更多信息，请参阅 由复制的链接和目录 AWS DataSync。

任务报告错误

在尝试使用任务报告监控您的 DataSync传输时，您可能会遇到以下错误之一。

错误消息	解决办法
文件路径超过了 4096 个字符的最大长度。无法写入任务报告	无。 DataSync 无法传输路径超过 4,096 字节的文件。 有关更多信息，请参阅 存储系统、文件和对象限制。
由于存储桶或 IAM 角色无效，无法将任务报告上传到 S3	检查 DataSync IAM 角色是否具有将任务报告上传到您的 S3 存储桶的权限。
生成任何任务报告之前出现执行错误	检查您的CloudWatch 日志，确定任务执行失败的原因。

解决数据验证的问题

默认情况下，会在传输结束时 AWS DataSync 验证数据的完整性。使用以下信息来帮助您诊断常见的验证错误和警告，例如在 DataSync 完成数据验证之前文件被修改或删除。

对于验证问题，很多时候，除了看到的任务执行错误之外，查看CloudWatch 日志（或任务报告）也会有所帮助。 DataSync为增强模式任务提供 JSON 结构化日志，而基本模式任务具有非结构化日志。

文件内容不匹配

当任务执行完毕后，您会看到以下错误：

```
Transfer and verification completed. Verification detected mismatches. Files with
 mismatches are listed in Cloud Watch Logs
```

在 CloudWatch 日志中，您可能会注意到源位置和目标位置之间存在差异的内容验证失败。如果在传输过程中修改了文件，就有可能出现这种情况。

例如，以下日志显示了 `file1.txt` 具有不同 `mtime`、`srcHash` 和 `dstHash` 值：

基本模式的日志示例

```
[NOTICE] Verification failed <> /directory1/directory2/file1.txt
[NOTICE] /directory1/directory2/file1.txt   srcMeta: type=R mode=0755 uid=65534
 gid=65534 size=534528 atime=1633100003/684349800 mtime=1602647222/222919600
 extAttrsHash=0
[NOTICE]   srcHash: 0c506c26bd1e43bd3ac346734f1a9c16c4ad100d1b43c2903772ca894fd24e44
[NOTICE] /directory1/directory2/file1.txt   dstMeta: type=R mode=0755 uid=65534
 gid=65534 size=511001 atime=1633100003/684349800 mtime=1633106855/859227500
 extAttrsHash=0
[NOTICE]   dstHash: dbd798929f11a7c0201e97f7a61191a83b4e010a449dfc79fbb8233801067c46
```

中 DataSync，mtime表示准备前最后一次写入文件的时间。验证传输时， DataSync 比较来源和目标位置之间的mtime值。如果两个位置的文件的 mtime 不相同，就会产生这样的验证失败。srcHash 和 dstHash 之间存在差异表明两个位置的文件内容不匹配。

要采取的操作

执行以下操作：

1. 使用 epoch 时间转换器，确定源文件或目标文件或对象最近是否遭到修改。这有助于确定最新版本。
2. 为避免再次出现此错误，请将任务安排在维护时段内运行，此时源和目标处于不活动状态。

文件的 SMB 元数据不匹配

当任务执行完毕后，您会看到以下错误：

```
Transfer and verification completed. Verification detected mismatches. Files with
 mismatches are listed in Cloud Watch Logs
```

在支持服务器消息块（SMB）协议的存储系统之间传输时，如果源和目标的文件扩展 SMB 属性不匹配，则可能会看到此错误。

例如，以下日志显示两个位置的 file1.txt 具有不同的 extAttrsHash 值，这表明文件内容相同，但未在目标位置设置扩展属性：

基本模式的日志示例

```
[NOTICE] Verification failed <> /directory1/directory2/file1.txt
```

```
[NOTICE] /directory1/directory2/file1.txt   srcMeta: type=R mode=0755 uid=65534
 gid=65534 size=1469752 atime=1631354985/174924200 mtime=1536995541/986211400
 extAttrsHash=2272191894
[NOTICE]   srcHash: 38571d42b646ac8f4034b7518636b37dd0899c6fc03cdaa8369be6e81a1a2bb5
[NOTICE] /directory1/directory2/file1.txt   dstMeta: type=R mode=0755 uid=65534
 gid=65534 size=1469752 atime=1631354985/174924200 mtime=1536995541/986211400
 extAttrsHash=3051150340
[NOTICE]   dstHash: 38571d42b646ac8f4034b7518636b37dd0899c6fc03cdaa8369be6e81a1a2bb5
```

您可能还会看到扩展属性的以下相关错误消息：

```
[ERROR] Deferred error: WriteFileExtAttr2 failed to setextattrlist(filename="/
directory1/directory2/file1.txt"): Input/output error
```

要采取的操作

当没有足够的权限将访问控制列表 (ACLs) 复制到目标时，通常会发生此错误。要解决此问题，请根据目标类型查看以下配置指南：

- Windows 文件服务器文件系统@@ 所需的权限 FSx
- 使用 SMB FSx 的 ONTAP 文件系统@@ 所需的权限

要传输的文件已不在源位置

当任务执行完毕后，您会看到以下错误：

```
Transfer and verification completed. Selected files transferred except for files
 skipped due to errors. If no skipped files are listed in Cloud Watch Logs, please
 contact AWS Support for further assistance.
```

在日志中，您可能会看到表明文件不在源位置的错误消息。如果文件（例如`file1.dll`和`file2.dll`）在准备之后但在 DataSync 传输之前被删除，则会发生这种情况：

基本模式的日志示例

```
[ERROR] Failed to open source file /file1.dll: No such file or directory
[ERROR] Failed to open source file /file2.dll: No such file or directory
```

要采取的操作

为避免出现这些情况，请将任务安排在源位置不活动时运行。

例如，您可以在用户和应用程序未积极操作该位置时，在维护时段内运行任务。

在某些情况下，您可能看不到此错误的相关日志。如果发生这种情况，请联系 AWS 支持 中心。

DataSync 无法验证目标数据

当任务执行完毕后，您会看到以下错误：

```
Transfer and verification completed. Verification detected mismatches. Files with
 mismatches are listed in Cloud Watch Logs
```

在日志中，您可能会注意到 DataSync 无法验证目标位置中的某些文件夹或文件。这些错误可能如下所示：

基本模式的日志示例

```
[ERROR] Failed to read metadata for destination file /directory1/directory2/
file1.txt: No such file or directory
```

对于文件，您可能会看到以下验证失败：

基本模式的日志示例

```
[NOTICE] Verification failed <> /directory1/directory2/file1.txt
[NOTICE] /directory1/directory2/file1.txt   srcMeta: type=R mode=0755 uid=65534
 gid=65534 size=61533 atime=1633099987/747713800 mtime=1536995631/894267700
 extAttrsHash=232104771
[NOTICE]   srcHash: 1426fe40f669a7d36cca1b5329983df31a9aeff8eb9fe3ac885f26de2f8fff6b
[NOTICE] /directory1/directory2/file1.txt   dstMeta: type=R mode=0755 uid=65534
 gid=65534 size=0 atime=0/0 mtime=0/0 extAttrsHash=0
[NOTICE]   dstHash: 0000000000000000000000000000000000000000000000000000000000000000
```

要采取的操作

这些日志表明，目标数据是在传输之后但在验证之前遭到删除。（在同一时间段内，将数据上传到源位置时的日志看起来很相似。）

为避免出现这些情况，请将任务安排在目标位置不活动时运行。

例如，您可以在用户和应用程序未积极操作该位置时，在维护时段内运行任务。

DataSync 无法读取对象元数据

当任务执行完毕后，您会看到以下错误：

```
Transfer and verification completed. Selected files transferred except for files
 skipped due to errors. If no skipped files are listed in Cloud Watch Logs, please
 contact AWS Support for further assistance.
```

在您的日志中，您 DataSync 可能会注意到由`file1.png`于 Amazon S3 `HeadObject` 请求失败而无法读取。 DataSync 在任务准备和验证期间向 S3 地点@@ 发出`HeadObject`请求。

基本模式的日志示例

```
[WARN] Failed to read metadata for file /file1.png: S3 Head Object Failed
```

要采取的操作

要解决此问题，请验证是否 DataSync 具有使用您的 S3 存储桶的适当权限级别：

- 确保 DataSync 用于访问您的 Amazon S3 营业地点的 IAM 角色允许该`s3:GetObject`权限。有关更多信息，请参阅 所需的权限。
- 如果您的 S3 存储桶使用服务器端加密，请确保 DataSync 允许服务器端加密访问该存储桶中的对象。有关更多信息，请参阅 使用服务器端加密访问 S3 存储桶。

对象的系统定义元数据不匹配

在 S3 存储桶之间的增强模式任务执行完毕后，您会看到以下错误：

```
Verification failed due to a difference in metadata
```

您可能会在日志中注意到，对象的 Amazon S3 系统定义元数据不匹配。在这一特定示例中，源对象没有 `Content-Type` 元数据，但目标对象有。之所以发生这种情况，是因为目标 S3 存储桶在将`"ContentType": "application/octet-stream"`元数据 DataSync 传输到那里时会自动将元数据应用于该对象。

增强模式的日志示例


```
{
    "Action": "VERIFY",
    "Source": {
        "LocationId": "loc-0b3017fc4ba4a2d8d",
        "RelativePath": "encoding/content-null",
        "Metadata": {
            "Type": "Object",
            "ContentSize": 24,
            "LastModified": "2024-12-23T15:48:15Z",
            "S3": {
                "SystemMetadata": {
                    "ETag": "\"68b9c323bb846841ee491481f576ed4a\""
                },
                "UserMetadata": {},
                "Tags": {}
            }
        }
    },
    "Destination": {
        "LocationId": "loc-abcdef01234567890",
        "RelativePath": "encoding/content-null",
        "Metadata": {
            "Type": "Object",
            "ContentSize": 24,
            "LastModified": "2024-12-23T16:00:03Z",
            "S3": {
                "SystemMetadata": {
                    "ContentType": "application/octet-stream",
                    "ETag": "\"68b9c323bb846841ee491481f576ed4a\""
                },
                "UserMetadata": {
                    "file-mtime": "1734968895000"
                },
                "Tags": {}
            }
        }
    },
    "TransferType": "CONTENT_AND_METADATA",
    "ErrorCode": "MetadataDiffers",
    "ErrorDetail": "Verification failed due to a difference in metadata"
}
```

要采取的操作

为避免出现此错误，请将源位置对象更新为包含 `Content-Type` 元数据属性。

了解数据验证时间

DataSync 验证包括文件内容的 SHA256 校验和以及不同位置之间文件元数据的精确比较。验证所需的时间取决于多种因素，包括所涉及的文件或对象数量、存储系统中的数据大小以及存储系统性能。

要采取的操作

考虑到可能影响验证时间的因素，您无需执行任何操作。但是，如果任务执行看起来卡在了验证状态，请联系 AWS 支持 中心。

使用 S3 存储成本高于预期进行故障排除 DataSync

如果您的 Amazon S3 存储成本高于您在 AWS DataSync 转移后的预期，则可能是由于以下一个或多个原因造成的：

- 向 S3 存储桶传入或传出 S3 存储桶时，会产生与发出的 S3 API 请求相关的费用。 DataSync
- DataSync 使用 Amazon S3 分段上传功能将对象上传到 S3 存储桶。这种方法可能会导致对未成功完成的上传收取意外的存储费用。
- DataSync 在控制台上启用 “复制对象标签” 或`ObjectTags`设置为 “复制” 时，会从源对象和目标对象复制对象标签`PRESERVE`。复制这些对象标签可能会产生 S3 API 请求费用。
- 可能对您的 S3 存储桶启用了对象版本控制。对象版本控制导致 Amazon S3 存储同名对象的多个副本。

要采取的操作

在这些情况下，您可以执行以下步骤：

- 请务必了解 S3 请求是如何 DataSync 使用的，以及它们可能如何影响您的存储成本。有关更多信息，请参阅 使用时评估 S3 请求成本 DataSync。
- 如果问题与分段上传有关，请配置 S3 存储桶分段上传策略以清除未完成的分段上传来降低存储成本。有关更多信息，请参阅 AWS 博客文章 S3 生命周期管理更新-支持分段上传和删除标记。
- 如果问题与复制对象标签有关，并且您不需要对象标签，请在创建、启动或更新任务`None`时清除 DataSync 控制台中的 “复制对象标签” 复选框或`ObjectTags`将其设置为。
- 如果问题与对象版本控制有关，请在 S3 存储桶上禁用对象版本控制。

如果需要其他帮助，请联系 AWS 支持 中心。

AWS DataSync 教程

这些教程将引导您了解使用 AWS DataSync的真实场景。

主题

- 教程：将数据从本地存储传输到 Amazon S3 AWS 账户
- 教程：在 Amazon S3 存储桶之间传输数据 AWS 账户

教程：将数据从本地存储传输到 Amazon S3 AWS 账户

AWS DataSync 与本地存储一起使用时，您通常会将数据传输到与您的 DataSync 代理同属 AWS 账户的 AWS 存储服务。但是，在某些情况下，您可能需要将数据传输到与其他账户关联的 Amazon S3 存储桶。

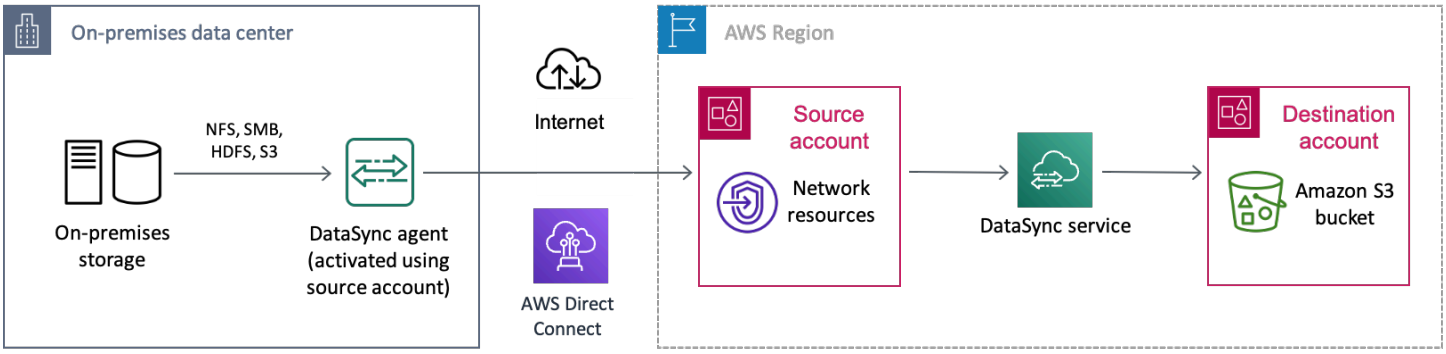
> **Important**
>
> 只有当 Amazon S3 是传输位置之一时，才可以使用本教程中的方法进行数据 DataSync 传输。 AWS 账户

概述

需要在不同组织之间传输数据的情况并不少见 AWS 账户，尤其是在您有单独的团队管理组织资源的情况下。以下是使用跨账户转账的 DataSync 样子：

- 源账户： AWS 账户 用于管理网络资源。这是您激活 DataSync代理时使用的帐户。
- 目标账户： AWS 账户 用于管理需要向其传输数据的 S3 存储桶。

下图阐明了此类方案。



先决条件：所需的源账户权限

对于您的来源 AWS 账户，这种跨账户转账需要考虑两组权限：

- 允许@@ 用户使用的用户权限 DataSync（可能是您或您的存储管理员）。这些权限允许您创建 DataSync 位置和任务。
- DataSync 允许将数据传输 DataSync 到目标账户存储桶的@@ 服务权限。

User permissions

在您的源账户中，至少向用于创建 DataSync 营业地点和任务的 IAM 角色添加以下权限。有关如何为角色添加权限的信息，请参阅创建或修改 IAM 角色。

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "SourceUserRolePermissions",
            "Effect": "Allow",
            "Action": [
                "datasync:CreateLocationS3",
                "datasync:CreateTask",
                "datasync:DescribeLocation*",
                "datasync:DescribeTaskExecution",
                "datasync:ListLocations",
                "datasync:ListTaskExecutions",
                "datasync:DescribeTask",
                "datasync:CancelTaskExecution",
                "datasync:ListTasks",
                "datasync:StartTaskExecution",
                "s3:GetBucketLocation",
```

```
                "s3:ListAllMyBuckets"
            ],
            "Resource": "*"
        },
        {
            "Sid": "IAMPermissions",
            "Effect": "Allow",
            "Action": [
                "iam:CreateRole",
                "iam:ListRoles",
                "iam:CreatePolicy"
            ],
            "Resource": "arn:aws:iam::111122223333:role/DataSync-*"
        },
        {
            "Sid": "IAMAttachRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:AttachRolePolicy"
            ],
            "Resource": "arn:aws:iam::111122223333:role/DataSync-*",
            "Condition": {
            "ArnLike": {
                "iam:PolicyARN": [
                   "arn:aws:iam::111122223333:policy/DataSync-*",
                   "arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess",
                   "arn:aws:iam::aws:policy/service-role/AWSDataSyncFullAccess"
                ]
            }
          }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": [
                        "datasync.amazonaws.com"
                    ]
                }
            }
```

```
        }
    ]
}
```

> **Tip**
>
> 要设置您的用户权限，请考虑使用AWSDataSyncFullAccess。这是一项 AWS 托管策略，可为用户提供对其依赖项的完全访问权限 DataSync 和最少访问权限。

DataSync service permissions

该 DataSync 服务需要您的源账户中的以下权限才能将数据传输到目标账户存储桶。

在本教程的后面部分，您将在为创建 IAM 角色时添加这些权限 DataSync。您还可以在目标存储桶策略和创建 DataSync 目标位置时指定此角色 (*source-datasync-role*)。

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Action": [
        "s3:GetBucketLocation",
        "s3:ListBucket",
        "s3:ListBucketMultipartUploads"
      ],
      "Effect": "Allow",
      "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket"
    },
    {
      "Action": [
        "s3:AbortMultipartUpload",
        "s3:DeleteObject",
        "s3:GetObject",
        "s3:ListMultipartUploadParts",
        "s3:PutObject",
        "s3:GetObjectTagging",
        "s3:PutObjectTagging"
      ],
      "Effect": "Allow",
```

```
      "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket/*"
    }
  ]
}
```

先决条件：所需的目标账户权限

在您的目标账户中，您的用户权限必须允许您更新目标存储桶的策略并禁用其访问控制列表（ACLs）。有关这些特定权限的更多信息，请参阅 Amazon S3 用户指南https://docs.aws.amazon.com/AmazonS3/latest/userguide/。

第 1 步：在您的源账户中，创建 DataSync 代理

首先，您必须创建一个可以从本地存储系统读取数据并与 DataSync 服务通信的 DataSync 代理。此过程包括在您的本地存储环境中部署代理以及在源 AWS 账户中激活代理。

> **Note**
>
> 本教程中的步骤适用于您使用的任何类型的代理和服务端点。

创建代 DataSync 理

1. 在本地存储环境中部署 DataSync 代理。
2. 选择代理将用来与之通信的服务端点 AWS。
3. 在您的源账户中激活代理。

步骤 2：在您的源账户中，创建用于访问目标存储桶 DataSync 的 IAM 角色

在您的源账户中，您需要一个 IAM 角色来授予向目标账户存储桶传输数据的权限。 DataSync

由于您要跨账户转账，因此必须手动创建角色。（使用同一个账户进行转账时，DataSync可以在控制台中为你创建此角色。）

创建 I DataSync AM 角色

创建以 DataSync 作为可信实体的 IAM 角色。

创建 IAM 角色

1. 使用您的源账户登录。 AWS 管理控制台
2. 使用 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。
3. 在左侧导航窗格的访问管理下，选择角色，然后选择创建角色。
4. 在选择可信实体页面中，为可信实体类型选择 AWS 服务。
5. 对于 “用例”，DataSync在下拉列表中进行选择，然后选择DataSync。选择下一步。
6. 在添加权限页面上，选择下一步。
7. 输入角色名称，然后选择创建角色。

有关更多信息，请参阅 IAM 用户指南中的为 AWS 服务 （ 控制台 ）创建角色。

向 DataSync IAM 角色添加权限

您刚刚创建的 IAM 角色需要允许将数据传输 DataSync 到目标账户中的 S3 存储桶的权限。

为您的 IAM 角色添加权限

1. 在 IAM 控制台的角色页面上，搜索您刚刚创建的角色并选择其名称。
2. 在角色的详情页面上，选择权限选项卡。选择添加权限，然后选择创建内联策略。
3. 选择 JSON 选项卡并执行以下操作：

    a. 将以下 JSON 粘贴到策略编辑器中：

> **Note**
>
> `aws:ResourceAccount` 的值应为拥有策略中指定 Amazon S3 存储桶的账户 ID。

JSON

```
{
 "Version":"2012-10-17",
 "Statement": [
     {
         "Action": [
             "s3:GetBucketLocation",
```

```
            "s3:ListBucket",
            "s3:ListBucketMultipartUploads"
        ],
        "Effect": "Allow",
        "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket",
        "Condition": {
            "StringEquals": {
             "aws:ResourceAccount": "123456789012"
            }
        }
    },
    {
        "Action": [
            "s3:AbortMultipartUpload",
            "s3:DeleteObject",
            "s3:GetObject",
            "s3:GetObjectTagging",
            "s3:GetObjectVersion",
            "s3:GetObjectVersionTagging",
            "s3:ListMultipartUploadParts",
            "s3:PutObject",
            "s3:PutObjectTagging"
          ],
        "Effect": "Allow",
        "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket/*",
        "Condition": {
            "StringEquals": {
                "aws:ResourceAccount": "123456789012"
            }
        }
    }
 ]
}
```

b. 将 *amzn-s3-demo-destination-bucket* 的每个实例替换为目标账户中 S3 存储桶的名称。

4. 选择下一步。为您的策略输入名称，然后选择创建策略。

第 3 步：在您的目标账户中，更新 S3 存储桶策略

在您的目标账户中，修改目标 S3 存储桶策略以包含您在源账户中创建的 DataSync IAM 角色。

开始之前：请确保您拥有目标账户所需的权限。

更新目标 S3 存储桶策略

1. 在中 AWS 管理控制台，切换到您的目标账户。
2. 打开 Amazon S3 控制台，网址为 https://console.aws.amazon.com/s3/。
3. 在左侧导航窗格中，选择存储桶。
4. 在 存储桶列表中，选择要向其传输数据的 S3 存储桶。
5. 在存储桶的详情页面上，选择权限选项卡。
6. 在 存储桶策略下，选择 编辑并执行以下操作来修改您的 S3 存储桶策略：

   a. 更新编辑器中的内容，以包含以下策略声明：

      JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "DataSyncCreateS3LocationAndTaskAccess",
      "Effect": "Allow",
      "Principal": {
      "AWS": "arn:aws:iam::111122223333:role/source-datasync-role"
      },
      "Action": [
        "s3:GetBucketLocation",
        "s3:ListBucket",
        "s3:ListBucketMultipartUploads",
        "s3:AbortMultipartUpload",
        "s3:DeleteObject",
        "s3:GetObject",
        "s3:ListMultipartUploadParts",
        "s3:PutObject",
        "s3:GetObjectTagging",
        "s3:PutObjectTagging"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-destination-bucket",
        "arn:aws:s3:::amzn-s3-demo-destination-bucket/*"
      ]
```

```
    }
  ]
}
```

b. 将 *source-account* 的每个实例替换为您源账户的 AWS 账户 ID。

c. *source-datasync-role*替换为您在源账户 DataSync 中为其创建的 IAM 角色。

d. 将 *amzn-s3-demo-destination-bucket* 的每个实例替换为目标账户中 S3 存储桶的名称。

7. 选择保存更改。

步骤 4：在您的目标账户中， ACLs 对您的 S3 存储桶禁用

您复制到 S3 存储桶的所有数据都属于您的目标账户，这一点至关重要。为确保该账户拥有数据，请禁用存储桶的访问控制列表 (ACLs)。有关更多信息，请参阅 Amazon S3 用户指南中的控制对象所有权和禁用 ACLs 存储桶。

为目标存储桶禁用 ACLs

1. 仍使用目标账户登录 S3 控制台时，请选择要向其传输数据的 S3 存储桶。
2. 在存储桶的详情页面上，选择权限选项卡。
3. 在对象所有权下方，请选择编辑。
4. 如果尚未选中，请选择ACLs 禁用（推荐）选项。
5. 选择保存更改。

第 5 步：在您的源账户中，为您的本地存储创建 DataSync 源位置

在您的源帐户中，为要从中传输数据的本地存储系统创建DataSync 源位置。此位置使用您在源账户中激活的代理。

步骤 6：在您的源账户中，为您的 S3 存储桶创建 DataSync 目标位置

仍然是在源账户中，为要向其传输数据的 S3 存储桶创建位置。

开始之前：请确保您拥有源账户所需的权限。

由于您无法使用 DataSync 控制台界面创建跨账户营业地点，因此这些说明要求您运行`create-location-s3`命令来创建目标营业地点。我们建议使用直接从控制台启动的基于浏览器 AWS

CloudShell、经过预先验证的 shell 来运行该命令。 CloudShell 允许你在`create-location-s3`不下载或安装 AWS CLI 命令行工具的情况下运行命令。

> **Note**
>
> 要使用除之外的命令行工具完成以下步骤 CloudShell，请确保您的AWS CLI 个人资料使用相同的 IAM 角色，其中包含在源账户 DataSync 中使用的所需用户权限。

要创建 DataSync 目标位置，请使用以下方法： CloudShell

1. 在您的源账户中，执行以下任一操作以 CloudShell 从控制台启动：
   - 选择控制台导航栏上的 CloudShell 图标。它位于搜索框的右侧。
   - 使用控制台导航栏上的搜索框进行搜索，CloudShell然后选择CloudShell选项。
2. 复制以下命令：

```
aws datasync create-location-s3 \
  --s3-bucket-arn arn:aws:s3:::amzn-s3-demo-destination-bucket \
  --s3-config '{
    "BucketAccessRoleArn":"arn:aws:iam::source-user-account:role/source-datasync-role"
  }'
```

3. 将 *amzn-s3-demo-destination-bucket* 替换为目标账户中 S3 存储桶的名称。
4. 将 *source-user-account* 替换为您源账户的 AWS 账户 ID。
5. *source-datasync-role*替换为您在源账户中创建的 DataSync IAM 角色。
6. 在中运行命令 CloudShell。

   如果命令返回类似于以下 DataSync 位置的地点 ARN，则表示您成功创建了该地点：

```
{
  "LocationArn": "arn:aws:datasync:us-east-2:123456789012:location/loc-abcdef01234567890"
}
```

7. 在左侧导航窗格中，展开数据传输，然后选择位置。

在您的源账户中，您可以看到刚刚为目标账户存储桶创建的 S3 位置。

第 7 步：在您的源账户中，创建并启动 DataSync 任务

在开始传输数据的 DataSync 任务之前，让我们回顾一下你到目前为止所做的事情：

- 在您的来源账户中，您创建了 DataSync 代理。代理可以从您的本地存储系统读取数据并与 DataSync 服务通信。
- 在您的源账户中，您创建了一个 IAM 角色，该角色允许将数据传输 DataSync 到目标账户中的 S3 存储桶。
- 在您的目标账户中，您配置了 S3 存储桶，以便 DataSync 可以向其传输数据。
- 在您的来源账户中，您创建了转账的 DataSync 来源和目标地点。

创建并启动 DataSync 任务

1. 在源账户中仍在使用 DataSync 控制台的同时，展开左侧导航窗格中的数据传输，然后选择任务和创建任务。
2. 在配置源位置页面上，选择选择现有位置。选择要从中复制数据的源位置（您的本地存储），然后选择下一步。
3. 在配置目标位置页面上，选择选择现有位置。选择要向其复制数据的目标位置（目标账户中的 S3 存储桶），然后选择下一步。
4. 在配置设置页面上，为任务命名。根据需要配置其他设置，例如指定 Amazon CloudWatch 日志组。选择下一步。
5. 在审核页面上，检查您的设置，然后选择创建任务。
6. 在任务的详细信息页面上，选择开始，然后选择以下选项之一：
   - 如需运行任务而不进行修改，请选择从默认值开始。
   - 如需修改任务再运行，请选择使用覆盖选项开始。

任务完成后，请检查目标账户中的 S3 存储桶。您应该会看到从源位置移出的数据。

相关资源

有关本教程中操作的更多信息，请参阅以下主题：

- 为 AWS 服务（控制台）创建角色
- 修改角色信任策略（控制台）

- 使用 Amazon S3 控制台添加存储桶策略
- 使用创建 S3 位置 AWS CLI

教程：在 Amazon S3 存储桶之间传输数据 AWS 账户

使用 AWS DataSync，您可以在属于不同 AWS 账户的 Amazon S3 存储桶之间传输数据。

> **Important**
>
> AWS 账户 使用本教程中的方法进行数据传输仅适用于 Amazon S3。此外，本教程还可以帮助您在不同 AWS 区域的 S3 存储桶之间传输数据。

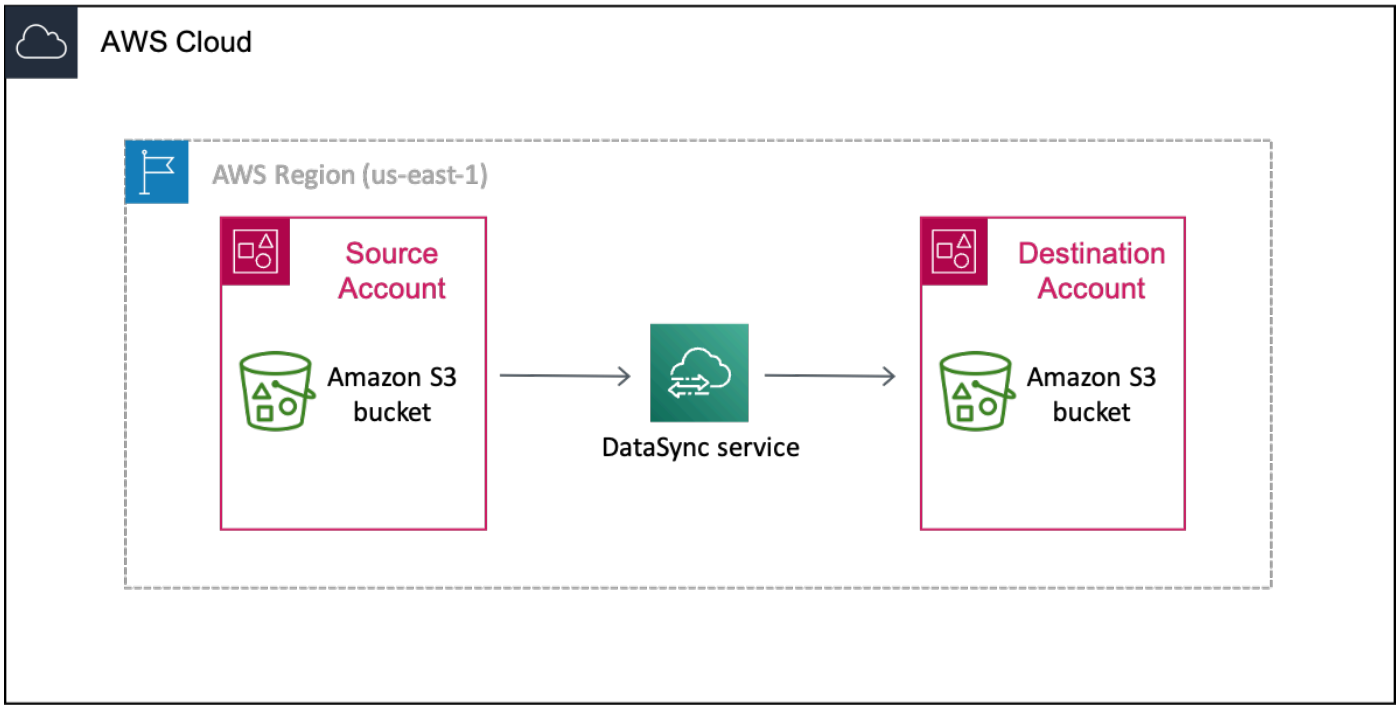
概述

在两者之间传输数据的情况并不少见 AWS 账户，尤其是在您有单独的团队管理组织资源的情况下。以下是使用跨账户转账的 DataSync 样子：

- 源账户： AWS 账户 用于管理您需要从中传输数据的 S3 存储桶。
- 目标账户： AWS 账户 用于管理需要向其传输数据的 S3 存储桶。

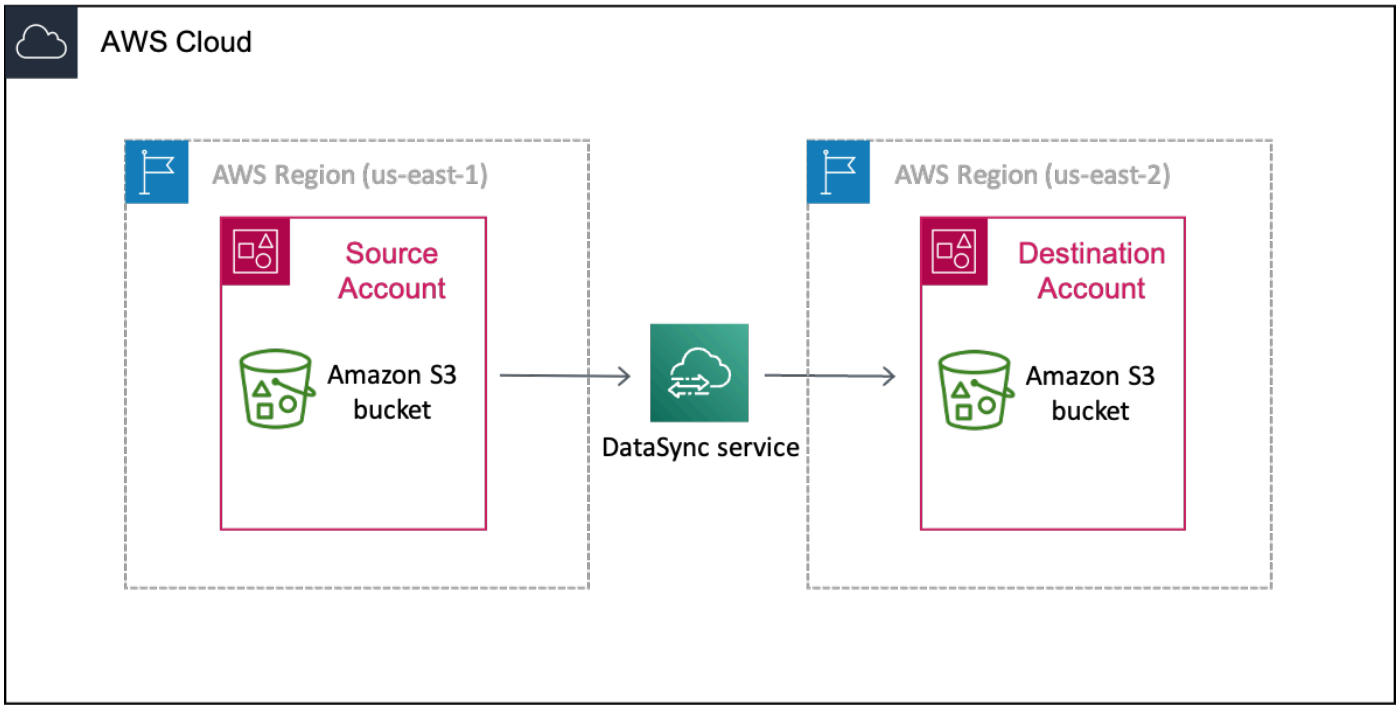
Transfers across accounts

下图说明了您将数据从一个 S3 存储桶传输到位于另一个不同的 AWS 账户中的 S3 存储桶中的场景。



Transfers across accounts and Regions

下图说明了您将数据从 S3 存储桶传输到位于不同 AWS 账户 和区域的另一个 S3 存储桶的场景。



先决条件：所需的源账户权限

对于您的来源 AWS 账户，这种跨账户转账需要考虑两组权限：

- 允许@@ 用户使用的用户权限 DataSync（可能是您或您的存储管理员）。这些权限允许您创建 DataSync 位置和任务。
- DataSync 允许将数据传输 DataSync 到目标账户存储桶的@@ 服务权限。

源账户的用户权限

在您的源账户中，至少向用于创建 DataSync 营业地点和任务的 IAM 角色添加以下权限。有关如何为角色添加权限的信息，请参阅创建或修改 IAM 角色。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "SourceUserRolePermissions",
            "Effect": "Allow",
            "Action": [
                "datasync:CreateLocationS3",
                "datasync:CreateTask",
                "datasync:DescribeLocation*",
                "datasync:DescribeTaskExecution",
                "datasync:ListLocations",
                "datasync:ListTaskExecutions",
                "datasync:DescribeTask",
                "datasync:CancelTaskExecution",
                "datasync:ListTasks",
                "datasync:StartTaskExecution",
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets"
            ],
            "Resource": "*"
        },
        {
            "Sid": "IAMPermissions",
            "Effect": "Allow",
            "Action": [
```

```
                "iam:CreateRole",
                "iam:ListRoles",
                "iam:CreatePolicy"
            ],
            "Resource": "arn:aws:iam::111122223333:role/DataSync-*"
        },
        {
            "Sid": "IAMAttachRolePermissions",
            "Effect": "Allow",
            "Action": [
               "iam:AttachRolePolicy"
            ],
            "Resource": "arn:aws:iam::111122223333:role/DataSync-*",
            "Condition": {
               "ArnLike": {
                 "iam:PolicyARN": [
                    "arn:aws:iam::111122223333:policy/DataSync-*",
                    "arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess",
                    "arn:aws:iam::aws:policy/service-role/AWSDataSyncFullAccess"
                 ]
               }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": [
                        "datasync.amazonaws.com"
                    ]
                }
            }
        }
    ]
}
```

> **Tip**
>
> 要设置您的用户权限，请考虑使用AWSDataSyncFullAccess。这是一个 AWS 托管策略，为用户提供对其依赖项的完全访问权限 DataSync 和最少访问权限。

DataSync 您的源账户的服务权限

该 DataSync 服务需要您的源账户中的以下权限才能将数据传输到目标账户存储桶。

在本教程的后面部分，您将在为创建 IAM 角色时添加这些权限 DataSync。您还可以在目标存储桶策略和创建 DataSync目标位置时指定此角色 (*source-datasync-role*)。

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Action": [
        "s3:GetBucketLocation",
        "s3:ListBucket",
        "s3:ListBucketMultipartUploads"
      ],
      "Effect": "Allow",
      "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket"
    },
    {
      "Action": [
        "s3:AbortMultipartUpload",
        "s3:DeleteObject",
        "s3:GetObject",
        "s3:ListMultipartUploadParts",
        "s3:PutObject",
        "s3:GetObjectTagging",
        "s3:PutObjectTagging"
      ],
      "Effect": "Allow",
      "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket/*"
    }
  ]
```

```
}
```

先决条件：所需的目标账户权限

在您的目标账户中，您的用户权限必须允许您更新目标存储桶的策略并禁用其访问控制列表（ACLs）。有关这些特定权限的更多信息，请参阅 Amazon S3 用户指南https://docs.aws.amazon.com/AmazonS3/latest/userguide/。

步骤 1：在您的源账户中，创建用于访问目标存储桶 DataSync 的 IAM 角色

在您的源中 AWS 账户，您需要一个 IAM 角色来授予向目标账户存储桶传输数据的权限。 DataSync

由于您要跨账户转账，因此必须手动创建角色。（使用同一个账户进行转账时，DataSync可以在控制台中为你创建此角色。）

创建 I DataSync AM 角色

创建以 DataSync 作为可信实体的 IAM 角色。

1. 使用您的源账户登录。 AWS 管理控制台
2. 使用 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。
3. 在左侧导航窗格的访问管理下，选择角色，然后选择创建角色。
4. 在选择可信实体页面中，为可信实体类型选择 AWS 服务。
5. 对于“用例”，DataSync在下拉列表中进行选择，然后选择DataSync。选择下一步。
6. 在添加权限页面上，选择下一步。
7. 输入角色名称，然后选择创建角色。

有关更多信息，请参阅 IAM 用户指南中的为 AWS 服务（控制台）创建角色。

向 DataSync IAM 角色添加权限

您刚刚创建的 IAM 角色需要允许将数据传输 DataSync 到目标账户中的 S3 存储桶的权限。

1. 在 IAM 控制台的角色页面上，搜索您刚刚创建的角色并选择其名称。
2. 在角色的详情页面上，选择权限选项卡。选择添加权限，然后选择创建内联策略。
3. 选择 JSON 选项卡并执行以下操作：

    a. 将以下 JSON 粘贴到策略编辑器中：

> **Note**
>
> `aws:ResourceAccount` 的值应为拥有策略中指定 Amazon S3 存储桶的账户 ID。

JSON

```
{
 "Version":"2012-10-17",
 "Statement": [
     {
         "Action": [
             "s3:GetBucketLocation",
             "s3:ListBucket",
             "s3:ListBucketMultipartUploads"
         ],
         "Effect": "Allow",
         "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket",
         "Condition": {
             "StringEquals": {
              "aws:ResourceAccount": "123456789012"
             }
         }
     },
     {
         "Action": [
             "s3:AbortMultipartUpload",
             "s3:DeleteObject",
             "s3:GetObject",
             "s3:GetObjectTagging",
             "s3:GetObjectVersion",
             "s3:GetObjectVersionTagging",
             "s3:ListMultipartUploadParts",
             "s3:PutObject",
             "s3:PutObjectTagging"
           ],
         "Effect": "Allow",
         "Resource": "arn:aws:s3:::amzn-s3-demo-destination-bucket/*",
         "Condition": {
             "StringEquals": {
                 "aws:ResourceAccount": "123456789012"
```

```
            }
        }
    }
 ]
}
```

b. 将 *amzn-s3-demo-destination-bucket* 的每个实例替换为目标账户中 S3 存储桶的名称。

4. 选择下一步。为您的策略输入名称，然后选择创建策略。

第 2 步：在您的目标账户中，更新 S3 存储桶策略

在您的目标账户中，修改目标 S3 存储桶策略以包含您在源账户中创建的 DataSync IAM 角色。

开始之前：请确保您拥有目标账户所需的权限。

更新目标 S3 存储桶策略

1. 在中 AWS 管理控制台，切换到您的目标账户。
2. 打开 Amazon S3 控制台，网址为 https://console.aws.amazon.com/s3/。
3. 在左侧导航窗格中，选择存储桶。
4. 在 存储桶列表中，选择要向其传输数据的 S3 存储桶。
5. 在存储桶的详情页面上，选择权限选项卡。
6. 在 存储桶策略下，选择 编辑并执行以下操作来修改您的 S3 存储桶策略：

   a. 更新编辑器中的内容，以包含以下策略声明：

      JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "DataSyncCreateS3LocationAndTaskAccess",
      "Effect": "Allow",
      "Principal": {
      "AWS": "arn:aws:iam::111122223333:role/source-datasync-role"
      },
      "Action": [
```

```
        "s3:GetBucketLocation",
        "s3:ListBucket",
        "s3:ListBucketMultipartUploads",
        "s3:AbortMultipartUpload",
        "s3:DeleteObject",
        "s3:GetObject",
        "s3:ListMultipartUploadParts",
        "s3:PutObject",
        "s3:GetObjectTagging",
        "s3:PutObjectTagging"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-destination-bucket",
        "arn:aws:s3:::amzn-s3-demo-destination-bucket/*"
      ]
    }
  ]
}
```

b. 将的*source-account*每个实例替换为您的源账户的 AWS 账户 ID。

c. *source-datasync-role*替换为您在源账户 DataSync 中为其创建的 IAM 角色。

d. 将 *amzn-s3-demo-destination-bucket* 的每个实例替换为目标账户中 S3 存储桶的名称。

7. 选择保存更改。

步骤 3：在您的目标账户中，ACLs 对您的 S3 存储桶禁用

您传输到 S3 存储桶的所有数据都属于您的目标账户，这一点至关重要。为确保该账户拥有数据，请禁用存储桶的访问控制列表 (ACLs)。有关更多信息，请参阅 Amazon S3 用户指南中的控制对象所有权和禁用 ACLs 存储桶。

开始之前：请确保您拥有目标账户所需的权限。

禁用您的目标 S3 存储桶 ACLs

1. 仍使用目标账户登录 S3 控制台时，请选择要向其传输数据的 S3 存储桶。
2. 在存储桶的详情页面上，选择权限选项卡。
3. 在对象所有权下方，请选择编辑。
4. 如果尚未选中，请选择ACLs 禁用（推荐）选项。

5. 选择保存更改。

第 4 步：在您的来源账户中，创建您的 DataSync 营业地点

在您的源账户中，为您的源和目标 S3 存储桶创建 DataSync 位置。

开始之前：请确保您拥有源账户所需的权限。

创建您的 DataSync 来源位置

- 在您的源账户中，为您要从中传输数据的 S3 存储桶创建位置。

创建您的 DataSync 目的地位置

仍然是在源账户中，为要向其传输数据的 S3 存储桶创建位置。

由于您无法使用 DataSync 控制台界面创建跨账户营业地点，因此这些说明要求您运行`create-location-s3`命令来创建目标营业地点。我们建议使用直接从控制台启动的基于浏览器 AWS CloudShell、经过预先验证的 shell 来运行该命令。 CloudShell允许你在`create-location-s3`不下载或安装 AWS CLI 命令行工具的情况下运行命令。

> **Note**
>
> 要使用除之外的命令行工具完成以下步骤 CloudShell，请确保您的AWS CLI 个人资料使用相同的 IAM 角色，其中包含在源账户 DataSync 中使用的所需用户权限。

使用创建 DataSync 目标位置 CloudShell

1. 在您的源账户中，执行以下任一操作以 CloudShell 从控制台启动：
   - 选择控制台导航栏上的 CloudShell 图标。它位于搜索框的右侧。
   - 使用控制台导航栏上的搜索框进行搜索，CloudShell然后选择CloudShell选项。
2. 复制以下 `create-location-s3` 命令：

```
aws datasync create-location-s3 \
  --s3-bucket-arn arn:aws:s3:::amzn-s3-demo-destination-bucket \
  --region amzn-s3-demo-destination-bucket-region \
  --s3-config '{
```

```
    "BucketAccessRoleArn":"arn:aws:iam::source-account-id:role/source-datasync-
role"
  }'
```

3. 将 *amzn-s3-demo-destination-bucket* 替换为目标账户中 S3 存储桶的名称。
4. 如果您的目标存储桶所在区域与源存储桶不同，请将 *amzn-s3-demo-destination-bucket-region* 替换为目标存储桶所在区域（例如 *us-east-2*）。如果您的存储桶位于同一区域，请移除此选项。
5. *source-account-id*替换为来源 AWS 账户 ID。
6. *source-datasync-role*替换为您在源账户中创建的 DataSyncIAM 角色。
7. 在中运行命令 CloudShell。

   如果命令返回类似于以下 DataSync 位置的地点 ARN，则表示您成功创建了该地点：

```
{
  "LocationArn": "arn:aws:datasync:us-east-2:123456789012:location/loc-
abcdef01234567890"
}
```

8. 在左侧导航窗格中，展开数据传输，然后选择位置。
9. 如果您在将该位置创建在了其他区域，请在导航窗格中选择对应的区域。

在您的源账户中，您可以看到刚刚为目标账户存储桶创建的 S3 位置。

第 5 步：在您的源账户中，创建并启动 DataSync 任务

在开始传输数据的 DataSync 任务之前，让我们回顾一下你到目前为止所做的事情：

- 在您的源账户中，您创建了一个 IAM 角色，允许将数据传输 DataSync 到目标账户中的 S3 存储桶。
- 在您的目标账户中，您配置了 S3 存储桶，以便 DataSync 可以向其传输数据。
- 在您的来源账户中，您创建了转账的 DataSync 来源和目标地点。

创建并开始您的 DataSync 任务

1. 在源账户中仍在使用 DataSync 控制台的同时，展开左侧导航窗格中的数据传输，然后选择任务和创建任务。
2. 如果目标账户中的存储桶与源账户中的存储桶位于不同的区域，请在顶部导航窗格中选择目标存储桶所在的区域。

> **Important**
>
> 为避免网络连接错误，您必须在与目标位置相同的区域创建 DataSync任务。

3. 在配置源位置页面上，执行以下操作：
   a. 选择选择现有地点。
   b. （对于跨区域传输）在区域下拉列表中，选择源存储桶所在的区域。
   c. 对于现有位置，选择要从中传输数据的 S3 存储桶的源位置，然后选择下一步。
4. 在配置目标位置页面上，执行以下操作：
   a. 选择选择现有地点。
   b. 对于现有位置，选择要将数据传输到的 S3 存储桶的目标位置，然后选择下一步。
5. 在配置设置页面上，选择任务模式。

> **Tip**
>
> 建议使用增强模式。有关更多信息，请参阅 为数据传输选择任务模式。

6. 为任务命名并配置其他设置，例如指定 Amazon CloudWatch 日志组。选择下一步。
7. 在审核页面上，检查您的设置，然后选择创建任务。
8. 在任务的详细信息页面上，选择开始，然后选择以下选项之一：
   - 如需运行任务而不进行修改，请选择从默认值开始。
   - 如需修改任务再运行，请选择使用覆盖选项开始。

任务完成后，请检查目标账户中的 S3 存储桶。您应该会看到从源账户存储桶中移出的数据。

问题排查

如果在尝试完成跨账户传输时遇到问题，请参考以下信息。

连接错误

在具有基本模式任务的不同 AWS 账户 区域的 S3 存储桶之间传输时，启动任务时可能会出现网络连接错误。 DataSync 要解决此问题，请使用增强模式任务。或者，请在与目标位置相同的区域创建基本模式任务，然后尝试运行该任务。

相关信息：使用服务器端加密的 S3 存储桶进行跨账户传输

如果您正在尝试使用服务器端加密的 S3 存储桶进行此类传输，请参阅 AWS 存储博客了解相关说明。

使用执行大型数据迁移 AWS DataSync

大规模的数据迁移可能涉及传输大量数据，这些数据包含数百万个不同格式的文件或对象。 AWS DataSync 通过管理调度、监控、加密和数据验证来简化这些复杂的传输。

什么是大规模数据迁移？

大规模数据迁移通常涉及将分布在不同来源的 TB 级或更多数据传输到新的目标存储环境（在此情况下为 AWS）。这种迁移需要在组织内部精心规划和协调，以确保成功迁移数据，同时最大限度地减少业务中断。

DataSync 可以简化这些迁移，这些迁移本质上通常很复杂。 DataSync 用于迁移的一些好处包括：

- 自动管理数据传输过程以及高性能和安全数据传输所需的基础架构。
- End-to-end 安全性，包括加密和数据完整性验证，以帮助确保您的数据安全、完好无损且随时可用。
- 专用网络协议和并行多线程架构，可加快迁移速度。

大规模数据迁移的关键阶段

通常可以将大规模迁移分为以下几个阶段：

- （第 1 阶段）规划数据迁移：在此阶段，您正在尝试了解迁移的原因，以及正在使用什么样的数据。规划活动包括：
  - 了解迁移的原因
  - 组建一支团队，协助您处理迁移的各方面事务。
  - 识别数据位置、格式和使用模式
  - 评估可用硬件资源和网络需求（如果从本地数据中心迁移）
  - 使用概念验证 (POC) 测试 DataSync 来估算迁移时间表、规划切换窗口，并了解需要如何配置 DataSync
- （第 2 阶段）实施大规模数据迁移：此时，您正在验证迁移计划并开始迁移。实施活动包括：
  - 验证迁移计划
  - 执行阶段割接，包括监控和验证数据传输是否符合预期
  - 在割接的间期根据需要进行优化和调整

- 完成后，清理未使用的资源

其他资源

AWS 规范性指导包含以下资源，可以帮助您规划和实施大规模迁移。使用本指南来了解 DataSync 如何在常见迁移流程和活动的背景下工作。

- 向云端的大规模迁移 AWS
- AWS 大型迁移的策略和最佳实践
- 在@@ AWS 大规模迁移中迁移共享文件系统 — 此资源包括一个 SFS-Discovery-Workbook，您可以下载该工作簿并使用它来规划文件共享级别的迁移。

第 1 阶段：规划大规模数据迁移

迁移大型数据集时，规划至关重要。你必须了解要迁移的数据、迁移的动机，以及 AWS DataSync 如何帮助你将数据放在你想要的地方。

主题

- 收集迁移要求
- 运行 DataSync 概念验证
- 估算迁移时间表

收集迁移要求

在大规模数据迁移的第一步中，需要在整个组织中收集各种信息。

这些信息有助于建立迁移过程，对于大规模迁移，这一过程可能包括从源存储到目标存储的多次传输和割接操作（多波完成）步骤。

了解迁移的原因

在开始迁移到之前 AWS，您需要清楚地了解迁移数据的原因。这有助于应对常见的迁移挑战，例如满足截止日期要求、管理资源以及跨团队协调。

如果需要帮助确定迁移的动机，请回答以下问题：

- 您是否正在释放本地存储空间？

- 您的硬件是否在硬件支持合同的期限内？
- 这是为了退出数据中心吗？
- 迁移时间表如何？
- 是否要从其他云存储传输数据？
- 是迁移部分数据集还是完整数据集？
- 这是为了归档数据吗？
- 应用程序或用户是否需要定期访问这些数据？

解决后勤问题

解决有关存储环境、迁移与组织的一些基本后勤问题：

1. 粗略了解当前的数据存储基础架构。
2. 验证是否需要代DataSync 理。例如，如果从本地存储进行传输，则需要代理。
3. 如果需要代理，请确保了解代理要求：
   - 代理可以作为 VMware ESXi、Linux 基于内核的虚拟机（KVM）和 Microsoft Hyper-V 虚拟机监控程序上的虚拟机（VM）运行。您还可以在 AWS中将代理部署为 Amazon EC2 实例。
   - 大规模迁移通常占用大量内存。确保代理具有足够的 RAM。
4. 确定领导层、网络、存储和 IT 部门中需要参与迁移的主要利益相关人。这可能包括：
   - 寻找一位专注于项目及其结果的专职领导者。
   - 确定迁移数据所有权和分类的负责人。
   - 确定谁管理您的来源，以及谁最终将管理您要迁移到的 AWS 存储服务。
   - 找出谁将为您的数据创建和管理任何其他流程 AWS。
5. 建立跨部门沟通渠道。
6. 制定突发事件回滚计划。
7. 记录完整的迁移过程，包括多波完成、验证和割接步骤。将其作为整个迁移的运行手册。在规划和实施迁移时需要更新此流程。

查看要迁移的数据

与存储和应用程序团队合作，分析要迁移的数据特征。此信息可帮助您确定可以执行的迁移策略 DataSync。

目录

- 确定数据使用模式
- 识别数据结构和布局
- 记录共享和文件夹
- 分析文件大小

确定数据使用模式

- 对于经常修改的常用数据，可规划多波增量传输，以避免业务运营中断。
- 对于可能视为存档的只读数据，可能无需规划多波完成。
- 如果有多种数据使用模式，需规划独立完成这些不同数据集的迁移波次。例如，存档数据可能是一个波次，其余波次专门用于迁移活动数据。

识别数据结构和布局

- 确定数据是按时间段（年、月、日）还是其他模式组织的。
- 使用此组织结构规划迁移波次。例如，您可能在一个波次中迁移一整年的存档数据。

记录共享和文件夹

- 创建共享和文件夹清单（包括列明每个共享和文件夹的文件或对象数量）。
- 使用活动数据集识别共享和文件夹。这些共享和文件夹在迁移期间可能需要增量传输。
- 查看配DataSync 额。这可以帮助您在配置时计划如何对数据集进行分区 DataSync。

分析文件大小

- 大文件（MB 或 GB）的传输数据吞吐量预计会比小文件（KB）更高。
- 如果您正在处理大量较小的文件，则预计存储系统上的元数据操作会更多，数据吞吐量会降低。DataSync在比较和验证源位置和目标位置时执行这些操作。

确定存储要求

要选择兼容的 AWS 存储服务来迁移数据，您需要评估源存储系统的特性和性能。

这些信息还可以帮助安排传输时间，以最大限度地降低迁移期间对业务运营的影响。

目录

- 确定源存储支持
- 查看元数据保留要求
- 从源存储收集性能指标
- 选择目标 AWS 存储服务

确定源存储支持

DataSync 可以与各种存储系统配合使用，这些存储系统允许通过 NFS、SMB、HDFS 和 S3 兼容的对象存储客户端进行访问。

如果您要从其他云存储迁移，请确认该提供商是否 DataSync 可以使用。有关受支持的源位置的列表，请参阅我可以在哪里传输我的数据 AWS DataSync？。

查看元数据保留要求

DataSync 可以在传输过程中保留您的文件或对象元数据。如何保留元数据取决于您的传输位置以及这些位置是否使用类似类型的元数据。

DataSync 在某些情况下，需要额外的权限才能保留文件元数据，例如 NTFS 自由访问列表 () DACLs。

有关更多信息，请参阅 了解如何 DataSync 处理文件和对象元数据。

从源存储收集性能指标

测量源存储在平均和峰值工作负载期间的基准 IOPS 与磁盘吞吐量。传输数据会增加源存储系统和目标存储系统的 I/O 开销。

将此性能数据与存储系统的规格参数进行比较，确定可用的性能资源。

选择目标 AWS 存储服务

此时，您可能已经知道哪种 AWS 存储服务对您的数据有意义。如果不知道，数据使用模式和存储性能是决策时需要考虑的两个方面。例如，如果您有存档数据，则可以考虑使用 Amazon S3，而对于活动数据，则可以考虑使用 Amazon FSx 或 Amazon EFS。

为了帮助您为数据选择正确的对象或基于文件的存储，请参阅选择 AWS 存储服务。

确定网络要求

要使用迁移数据 DataSync，必须在源存储、代理和之间建立网络连接 AWS。此外还需要规划足够的网络带宽和基础设施。

与网络工程师和存储管理员合作，收集以下网络需求。

目录

- 评估可用网络带宽
- 考虑将您的网络连接至的选项 AWS
- 选择代理通信的服务端点
- 规划足够的网络基础设施

评估可用网络带宽

可用网络带宽会影响传输速度和总迁移时间。如果要从本地存储系统进行传输，请执行以下操作：

- 与网络团队合作确定平均和峰值带宽利用率。
- 确定何时可以传输数据，避免日常运营中断。这可以提供迁移波次和割接何时发生的信息。

您可以控制带宽 DataSync 使用量。有关更多信息，请参阅 为 AWS DataSync 任务设置带宽限制。

由于来自其他云存储的传输通常发生在公共互联网上，因此这种传输的带宽限制和注意事项通常较少。

考虑将您的网络连接至的选项 AWS

在为 DataSync 传输建立网络连接时，请考虑以下选项：

- Direct Connect-查看使用 Direct Connect 的架构和路由示例 DataSync。您可以使用亚马逊监控 Direct Connect 活动 CloudWatch。
- VPN：AWS Site-to-Site VPN 提供每条隧道高达 1.25 Gbps 的吞吐量。
- 公共互联网：请联系互联网服务提供商获取网络使用量数据。

选择代理通信的服务端点

DataSync 代理使用服务端点与 DataSync 服务通信。所用端点类型取决于网络与 AWS连接的方式。

规划足够的网络基础设施

对于您创建的每项传输任务，都会 DataSync 自动生成和管理用于数据传输的网络基础架构。这种基础设施称为网络接口或弹性网络接口，它们是 Amazon 虚拟私有云（VPC）中代表虚拟网卡的逻辑网络组件。有关更多信息，请参阅《Amazon EC2 用户指南》。

每个网络接口在目标 VPC 子网中使用一个 IP 地址。要确保有足够的网络基础设施进行迁移，请执行以下操作：

- 记下 DataSync 将为您的 DataSync目标位置创建的网络接口的数量。
- 确保您的子网有足够的 IP 地址来 DataSync完成您的任务。例如，使用代理的任务需要四个 IP 地址。如果为迁移创建了四个任务，则意味着子网中需要 16 个可用的 IP 地址。

运行 DataSync 概念验证

通过运行概念验证 (POC) AWS DataSync 可以帮助您验证数据迁移计划的以下方面：

- 验证源位置和目标位置之间的网络连接。
- 验证您的初始 DataSync 任务配置。
- 测量数据传输性能。
- 估算迁移时间表。
- 与参与迁移的主要利益相关者一起定义成功的标准。

开始使用概念验证

1. 创建您的 DataSync 代理：
    1. 部署代理。
    2. 选择代理的服务端点。
    3. 激活您的代理。
    4. 验证代理的网络连接。
2. 选择一小部分数据代表要迁移的数据。

    例如，如果源存储混合了大文件和小文件，则在 POC 中传输的数据子集中应反映这一点。这使您可以初步了解存储系统、网络和的性能 DataSync。
3. 为您的本地或其他云存储系统创建 DataSync 源位置。

4. 为您的AWS 存储服务创建 DataSync 目标位置。
5. 使用仅@@ DataSync 传输您的数据子集的筛选器创建传输任务。
6. 开始你的 DataSync 任务。
7. 通过监控以下内容来收集传输性能指标：

   - 任务执行的数据与文件吞吐量。您可以通过 DataSync 控制台或操作来执行此DescribeTaskExecution操作。如果使用的是 `DescribeTaskExecution`，以下是计算这些指标的方法：
     - 数据吞吐量：`BytesWritten` 除以 `TransferDuration`
     - 文件吞吐量：`FilesTransferred` 除以 `TransferDuration`
   - 源存储和目标存储的利用率。与存储管理员密切合作以获取此信息。
   - 网络使用量。
8. 验证传输到目标位置的数据：

   - 查看 CloudWatch 日志中是否存在任务执行错误。
   - 验证目标位置是否保留了相应权限和元数据。
   - 确认应用程序和用户可以按预期访问目标数据。
   - 解决遇到的任何问题。有关更多信息，请参阅 疑难解答 AWS DataSync 问题。
9. 再运行几次任务，以了解准备、传输和验证数据需要 DataSync 多长时间。（有关更多信息，请参阅 任务执行状态。）

   如果您多次运行任务，则 DataSync 默认情况下会执行增量传输，并且仅复制与上一次任务运行相比更改的数据。

   虽然增量传输的传输时间可能会更短， DataSync 但通过扫描和比较您的位置以确定要传输的内容，始终以相同的方式准备传输。可使用这些准备时间来估算迁移的割接时间表。
10. 如果需要，请根据 POC 期间了解的信息更新迁移计划。

估算迁移时间表

使用到目前为止收集的信息，您可以使用 AWS DataSync估算迁移要花费的时间。

估算数据传输时间表

您可以根据您在收集迁移需求期间收集的以下信息以及概念 DataSync 验证 (POC) 来估算传输数据所需的时间 DataSync ：

- 可用网络带宽
- 源存储和目标存储的利用率指标
- 来自 DataSync PO C 的绩效指标

估算数据传输时间表

1. 将 POC 的数据和文件吞吐量与可用的网络带宽进行比较。
2. 如果您的吞吐量低于可用带宽（例如，网络带宽为 10 Gbps 的吞吐量为 300 MiB/s ），请考虑将数据集分为多个任务，以最大限度地提高带宽使用率。

   DataSync 有几个用于对数据集进行分区的选项。有关更多信息，请参阅 通过数据分区加速迁移。
3. 使用以下公式计算传输天数，该公式提供了理论上的最短传输时间：

   ```
   (DATA_SIZE * 8 bits per byte)/(CIRCUIT * NETWORK_UTILIZATION percentage * 3600
    seconds per hour * AVAILABLE_HOURS) = Number of days
   ```

   使用此公式时，请将以下内容替换为您自己的值：

   - DATA_SIZE：要迁移的数据量（以字节表示）。
   - CIRCUIT：可用网络带宽（以每秒位数表示）。
   - NETWORK_UTILIZATION: 网络使用百分比。
   - AVAILABLE_HOURS：每天可用的运行小时数。

   例如，可以按以下方式计算涉及 100 TB 数据、1 Gbps 互联网连接、80% 网络利用率和每天 24 小时可用性的迁移：

   ```
   (100,000,000,000,000 bytes * 8) / (1,000,000,000 bps * 0.80 * 3600 *
   24) = 11.57 days
   ```

   在这种情况下，若不考虑到实际情况，迁移过程可能耗时近 12 天。
4. 请根据实际情况，调整计算的传输时长：

   - 网络性能波动
   - 存储性能差异
   - 迁移波次之间的停机时间

估算割接时间表

如果要迁移活动数据集，可能需要割接，以免中断业务运营。

不要低估割接需要的时间。在大规模迁移中，割接活动占到总迁移时间的 30% 的情况并不少见。

1. 评估是否需要进行多波割接，以减少为增量更改而扫描的数据量。

   为实现这一点，一种策略是对根据共享、文件夹或存储系统分区的数据集进行割接。
2. 查看在 POC 期间准备、传输和验证数据通常 DataSync 需要多长时间。

   请特别注意任务执行的准备时长。要查找此信息，请运行该DescribeTaskExecution操作，然后检查持续时间（以毫秒PrepareDuration为单位）的值。
3. 通过测量并行任务之间的时间差，估算割接可能需要的时间。

   有关并行任务的更多信息，请参阅通过数据分区加速迁移。
4. 使用割接估算值来安排割接活动。当源数据无法修改时，这些时间本质上就是维护时段。

后续步骤

在估算好时间表之后，就可以开始实施迁移了。

第 2 阶段：实施大规模数据迁移

利用您在规划期间收集的信息，就可以开始使用迁移 AWS DataSync 到新的存储系统了。如果还没有开始，建议查看大规模迁移的AWS Prescriptive Guidance 资源。

主题

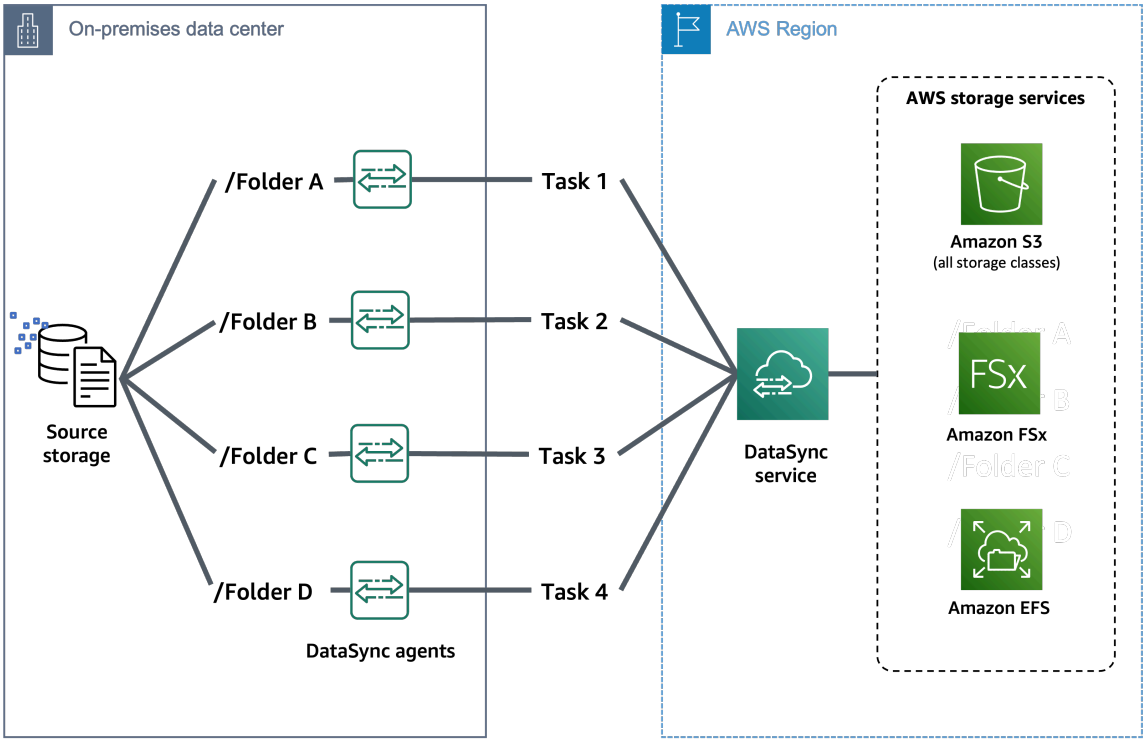
- 通过数据分区加速迁移
- 正在运行您的 DataSync 转移任务
- 监控传输

通过数据分区加速迁移

对于大规模迁移，我们建议使用多个 DataSync任务对数据集进行分区。将源数据划分到多个任务（可能还有代理）中，可以并行处理传输并缩短迁移时间。

分区还可以帮助您保持在 DataSync 配额范围内，并简化对任务的监控和调试。

下图显示了如何使用多个 DataSync 任务和代理从同一个源存储位置传输数据。在这种情况下，各项任务都专注于源位置的特定文件夹。有关这些方法的更多信息和示例，请参阅如何使用横向 AWS DataSync 扩展架构加速数据传输。



按文件夹或前缀对数据集进行分区

创建 DataSync 源位置时，您可以指定从中 DataSync 读取的文件夹、目录或前缀。例如，如果要迁移具有顶级目录的文件共享，则可创建多个位置来指定不同的目录路径。然后，您可以在迁移期间使用这些位置运行多个 DataSync任务。

使用筛选条件对数据集进行分区

您可以应用筛选条件，在传输中包含或排除源位置的数据。在大规模迁移的背景下，筛选条件有助于将任务范围限定到数据集的特定部分。

例如，如果要迁移按年份组织的存档数据，则可创建一个包含筛选条件来匹配特定年份或多个年份。此外还可以在每次运行任务时修改筛选条件以匹配不同的年份。

使用清单对数据集进行分区

清单是您要传输的文件或对象 DataSync 的列表。有了清单，就 DataSync 不必读取源位置的所有内容就能确定要传输的内容。

您可以根据源存储的清单创建清单，也可以通过事件驱动的方法创建清单（例如，参见AWS DataSync 使用数亿个对象实现）。也可以在每次启动任务时使用不同的清单，这样就可以在同一个任务中传输不同的数据集。

正在运行您的 DataSync 转移任务

在每个迁移波次中，数据传输通常遵循相同的常规过程：

1. 对数据进行初始全量传输。
2. 验证目标中的数据。
3. 对自首次传输以来可能发生变化的所有数据运行增量传输。
4. 将操作割接到目标位置。
5. 查看割接结果。

运行任务

您可能需要在工作时间运行 DataSync 转移任务，以最大限度地缩短总体迁移时间。在这些情况下，通常会先进行全量传输，然后根据用户和应用程序对源位置的更改进行增量传输。

为避免在工作时间出现与网络相关的问题，可以限制任务使用的带宽量。有关更多信息，请参阅 为 AWS DataSync 任务设置带宽限制。

1. 运行初始全量传输：
    a. 启动您的 DataSync 任务（如果您正在并行运行任务，则启动任务）。
    b. 监控任务执行的进度和性能。
    c. 验证数据是否按照预期的方式传输（例如，保留文件元数据）。
2. 运行增量传输：
    a. 安排任务定期运行。
    b. 监控任务执行情况，并在遇到错误时修复错误。

执行割接

完成初始和增量传输后，就可以启动将操作割接到目标位置的过程。

1. 启动计划的维护时段。

2. 将源存储系统更新为应用程序和用户只读模式。
3. 运行最终的增量传输，在源位置和目标位置之间复制剩余增量。
4. 进行全面的数据验证（例如，通过查看 CloudWatch日志和任务报告）。
5. 将应用程序和用户切换至目标位置的新环境。
6. 测试应用程序的功能，并确保用户可访问目标位置的数据。
7. 安排回顾会议，与迁移团队一起回顾传输情况。提出以下调查性的示例问题：
   - 割接成功了吗？ 如果没成功，问题出在哪里？
   - 是否使用了所有可用的带宽？
   - 源存储和目标存储是否得到充分利用？
   - 能否通过额外的任务获得更大的数据吞吐量？
   - 是否需要计划更长的维护时段？
8. 如果需要，请在开始下一波迁移之前更新迁移计划。

监控传输

AWS DataSync 提供了多个监控选项，可帮助您验证和调试传输。

使用 CloudWatch 指标监控您的转账

您可以使用 DataSync 任务执行中的指标创建自定义 CloudWatch 仪表板。有关更多信息，请参阅 使用 Amazon CloudWatch 指标监控数据传输。

使用任务报告监控您的传输

如果要传输数百万个文件或对象，可考虑使用任务报告。任务报告提供有关在任务执行期间 DataSync 尝试传输、跳过、验证和删除的内容的详细信息。有关更多信息，请参阅 使用任务报告监控您的数据传输。

您还可以使用 Amazon Athena 和 Amazon Quick 等 AWS AWS Glue服务将任务报告可视化。有关更多信息，请参阅 AWS 存储博客。

使用 CloudWatch 日志监控您的转账

建议将任务配置为至少记录基本信息和传输错误。有关更多信息，请参阅使用 Amazon CloudWatch 日志监控数据传输。

AWS DataSync API

除了 AWS 管理控制台 和之外 AWS CLI，您还可以使用 AWS DataSync API 对进行配置和管理 DataSync AWS SDKs。

主题

- 操作
- 数据类型
- 常见错误类型
- 常见参数

操作

支持以下操作：

- CancelTaskExecution
- CreateAgent
- CreateLocationAzureBlob
- CreateLocationEfs
- CreateLocationFsxLustre
- CreateLocationFsxOntap
- CreateLocationFsxOpenZfs
- CreateLocationFsxWindows
- CreateLocationHdfs
- CreateLocationNfs
- CreateLocationObjectStorage
- CreateLocationS3
- CreateLocationSmb
- CreateTask
- DeleteAgent
- DeleteLocation
- DeleteTask

- DescribeAgent
- DescribeLocationAzureBlob
- DescribeLocationEfs
- DescribeLocationFsxLustre
- DescribeLocationFsxOntap
- DescribeLocationFsxOpenZfs
- DescribeLocationFsxWindows
- DescribeLocationHdfs
- DescribeLocationNfs
- DescribeLocationObjectStorage
- DescribeLocationS3
- DescribeLocationSmb
- DescribeTask
- DescribeTaskExecution
- ListAgents
- ListLocations
- ListTagsForResource
- ListTaskExecutions
- ListTasks
- StartTaskExecution
- TagResource
- UntagResource
- UpdateAgent
- UpdateLocationAzureBlob
- UpdateLocationEfs
- UpdateLocationFsxLustre
- UpdateLocationFsxOntap
- UpdateLocationFsxOpenZfs
- UpdateLocationFsxWindows
- UpdateLocationHdfs

- UpdateLocationNfs
- UpdateLocationObjectStorage
- UpdateLocationS3
- UpdateLocationSmb
- UpdateTask
- UpdateTaskExecution

CancelTaskExecution

停止正在执行的 AWS DataSync 任务。部分文件传输突然中断。传输到目标位置的文件内容可能不完整或与源文件不一致。

但是，若您对相同的任务开启新的任务执行并让其完成，则目标位置的文件内容将完整且一致。这适用于中断任务执行的其他意外故障。在所有这些情况下，在开始执行下一个任务时 DataSync成功完成传输。

请求语法

```
{
   "TaskExecutionArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

TaskExecutionArn

要停止执行任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}/execution/exec-[0-9a-f]{17}$`

是否必需：是

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateAgent

激活您在存储环境中部署的 AWS DataSync 代理。激活过程会将代理与您的关联 AWS 账户。

如果您尚未部署代理，请参阅我需要 DataSync代理吗？

请求语法

```
{
   "ActivationKey": "string",
   "AgentName": "string",
   "SecurityGroupArns": [ "string" ],
   "SubnetArns": [ "string" ],
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ],
   "VpcEndpointId": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

ActivationKey

指定 DataSync 代理的激活密钥。如果您没有激活密钥，请参阅激活您的代理。

类型：字符串

长度约束：最大长度为 29。

模式：`[A-Z0-9]{5}(-[A-Z0-9]{5}){4}`

是否必需：是

AgentName

指定代理名称。我们建议您指定一个您可以记住的名称。

类型：字符串

长度限制：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

必需：否

SecurityGroupArns

指定允许代理与 VPC 服务端点之间流量的安全组的 Amazon 资源名称（ARN）。您只能指定一个 ARN。

类型：字符串数组

数组成员：固定数量为 1 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

必需：否

SubnetArns

指定 VPC 服务端点所在子网的 ARN。您只能指定一个 ARN。

类型：字符串数组

数组成员：固定数量为 1 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:subnet/subnet-[a-f0-9]+$`

必需：否

Tags

指定可帮助您对 AWS 资源进行分类、筛选和搜索的标签。我们建议为您的代理创建至少一个标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

VpcEndpointId

指定您正在使用的 VPC 服务端点的 ID。例如，VPC 端点 ID 类似如下内容：vpce-01234d5aff67890e1。

> **Important**
>
> 您使用的 VPC 服务终端节点必须包含 DataSync 服务名称（例如com.amazonaws.us-east-2.datasync）。

类型：字符串

模式：^vpce-[0-9a-f]{17}$

必需：否

响应语法

```
{
   "AgentArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AgentArn

您刚激活的代理的 ARN。使用ListAgents操作返回 AWS 账户 和中的代理列表 AWS 区域。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

以下示例激活 DataSync 代理。

```
{
    "ActivationKey": "AAAAA-1AAAA-BB1CC-33333-EEEEE",
    "AgentName": "MyAgent",
    "Tags": [{
        "Key": "Job",
        "Value": "TransferJob-1"
    }]
}
```

示例响应

响应返回已激活代理的 ARN。

```
{
    "AgentArn": "arn:aws:datasync:us-east-2:111222333444:agent/agent-0b0addbeef44baca3"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationAzureBlob

为微软 Azure Blob 存储容器创建传输位置。 AWS DataSync 可以将此位置用作传输源或目的地。无论是否有连接到您的集装箱的DataSync 代理，您都可以进行转账。

在开始之前，请务必知道如何 DataSync 访问 Azure Blob 存储空间以及如何使用访问层和 b lo b 类型。

请求语法

```
{
   "AccessTier": "string",
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "BlobType": "string",
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "ContainerUrl": "string",
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "SasConfiguration": {
      "Token": "string"
   },
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AccessTier

指定您希望将对象或文件传输到的访问层。这仅适用于使用该位置作为传输目标的情况。有关更多信息，请参阅访问层。

类型：字符串

有效值：`HOT | COOL | ARCHIVE`

必需：否

AgentArns

（可选）指定可以与 Azure Blob 存储容器连接的 DataSync 代理的亚马逊资源名称 (ARN)。如果设置的是无代理跨云传输，则无需为此参数指定值。

您可以指定多个代理。有关更多信息，请参阅使用多个代理进行传输。

> **Note**
>
> 首次创建存储位置时，务必确保正确配置此参数。创建代理后，就无法在存储位置添加或删除代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

AuthenticationType

指定用于访问你的 Azure Blob 存储空间的身份验证方法 DataSync 。 DataSync 可以使用共享访问签名 (SAS) 访问 blob 存储。

类型：字符串

有效值：`SAS | NONE`

是否必需：是

BlobType

指定在将对象或文件传输到 Azure Blob Storage 时想要的 blob 类型。目前， DataSync 仅支持将数据作为块 blob 移动到 Azure Blob 存储。有关 blob 类型的更多信息，请参阅 Azure Blob Storage 文档。

类型：字符串

有效值：BLOCK

必需：否

CmkSecretConfig

指定由客户 DataSync AWS KMS key管理的密钥的配置信息，其中包括 DataSync 用于访问特定 AzureBlob存储位置的身份验证令牌。

当您在CreateLocationAzureBlob请求中包含此参数时，您仅提供 KMS 密钥 ARN。DataSync 使用此 KMS 密钥和您为其指定的身份验证令牌SasConfiguration来创建存储位置访问凭证的 DataSync托管密钥。

确保该用户 DataSync 有权访问您指定的 KMS 密钥。有关更多信息，请参阅使用使用自定义 AWS KMS key密钥加密的服务托管密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 SasConfiguration）或 CustomSecretConfig（不带 SasConfiguration）为 CreateLocationAzureBlob 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CmkSecretConfig 对象

必需：否

ContainerUrl

指定传输中涉及的 Azure Blob Storage 容器的 URL。

类型：字符串

长度约束：最大长度为 325。

模式：`^https:\/\/[A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252}\/[a-z0-9](-?[a-z0-9]){2,62}$`

是否必需：是

CustomSecretConfig

指定客户管理的 Secrets Manager 密钥的配置信息，其中 AzureBlob 存储位置的身份验证令牌以纯文本形式存储在 Secrets Manager 中。此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。有关更多信息，请参阅使用您管理的密钥。

> **Note**
>
> 可使用 `CmkSecretConfig`（带有 `SasConfiguration`）或 `CustomSecretConfig`（不带 `SasConfiguration`）为 `CreateLocationAzureBlob` 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CustomSecretConfig 对象

必需：否

SasConfiguration

指定允许 DataSync 访问你的 Azure Blob 存储空间的 SAS 配置。

> **Note**
>
> 如果您使用提供身份验证令牌`SasConfiguration`，但不使用`CmkSecretConfig`或提供密钥配置详细信息`CustomSecretConfig`，则使用您 AWS 账户的密钥管理器密钥 DataSync 存储令牌。

类型：AzureBlobSasConfiguration 对象

必需：否

Subdirectory

如果要将传输限制到容器中的虚拟目录（例如 `/my/images`），请指定路径段。

类型：字符串

长度限制：最大长度为 1024。

模式：`^[\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}\p{C}]*$`

必需：否

Tags

指定帮助您分类、筛选和搜索 AWS 资源的标签。我们建议至少为您的传输位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您创建的 Azure Blob Storage 传输位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationEfs

为 Amazon EFS 文件系统创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目标。

在开始之前，请务必了解如何 DataSync访问 Amazon EFS 文件系统。

请求语法

```
{
   "AccessPointArn": "string",
   "Ec2Config": {
      "SecurityGroupArns": [ "string" ],
      "SubnetArn": "string"
   },
   "EfsFilesystemArn": "string",
   "FileSystemAccessRoleArn": "string",
   "InTransitEncryption": "string",
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AccessPointArn

指定用于挂载 Amazon EFS 文件系统的接入点的 DataSync 亚马逊资源名称 (ARN)。

有关更多信息，请参阅访问受限文件系统。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):elasticfilesystem:[a-z\-0-9]+:[0-9]{12}:access-point/fsap-[0-9a-f]{8,40}$`

必需：否

Ec2Config

指定用于连接您的 Amazon EFS 文件系统的其中一个挂载目标的子网和安全组 DataSync 。

类型：Ec2Config 对象

是否必需：是

EfsFilesystemArn

指定 Amazon EFS 文件系统的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):elasticfilesystem:[a-z\-0-9]+:[0-9]{12}:file-system/fs-[0-9a-f]{8,40}$`

是否必需：是

FileSystemAccessRoleArn

指定允许 DataSync 访问您的 Amazon EFS 文件系统的 AWS Identity and Access Management (IAM) 角色。

有关创建此角色的信息，请参阅创建用于文件系统访问 DataSync的 IAM 角色。

类型：字符串

长度约束：最大长度为 2048。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):iam::[0-9]{12}:role/.*$`

必需：否

InTransitEncryption

指定在传输层安全 (TLS) 1.2 加密与您的 Amazon EFS 文件系统之间传输数据时，是否要使用传输层安全 (TLS) 1.2 加密。 DataSync

如果您使用 AccessPointArn 指定访问点或使用 FileSystemAccessRoleArn 指定 IAM 角色，则必须将此参数设置为 TLS1_2。

类型：字符串

有效值：NONE | TLS1_2

必需：否

Subdirectory

指定 Amazon EFS 文件系统的挂载路径。这是在文件系统上 DataSync 读取或写入数据的地方（取决于这是源位置还是目标位置）。

默认情况下， DataSync 使用根目录（如果使用提供接入点，则使用接入点AccessPointArn）。您也可以使用正斜杠包含子目录（例如，/path/to/folder）。

类型：字符串

长度约束：最大长度为 4096。

模式：^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$

必需：否

Tags

指定表示要添加到资源的标签的键值对。该值可为空字符串。该值可帮助您管理、筛选和搜索资源。我们建议您为自己的位置创建名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
```

```
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

创建的 Amazon EFS 文件系统位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

以下示例为 Amazon EFS 文件系统创建位置。

```
{
    "Ec2Config": {
        "SubnetArn": "arn:aws:ec2:us-east-2:11122233344:subnet/
subnet-1234567890abcdef1",
        "SecurityGroupArns": [
            "arn:aws:ec2:us-east-2:11122233344:security-group/sg-1234567890abcdef2"
        ]
    },
    "EfsFilesystemArn": "arn:aws:elasticfilesystem:us-east-2:111222333444:file-system/
fs-021345abcdef6789",
    "Subdirectory": "/mount/path",
    "Tags": [{
        "Key": "Name",
        "Value": "ElasticFileSystem-1"
    }]
}
```

请求示例：为受限制的 Amazon EFS 文件系统创建位置

以下示例系为限制访问的 Amazon EFS 文件系统创建位置。在这种情况下，您可能需要在请求中指定 `AccessPointArn`、`FileSystemAccessRoleArn`、和 `InTransitEncryption` 的值。

```
{
    "AccessPointArn": "arn:aws:elasticfilesystem:us-east-2:111222333444:access-point/
fsap-1234567890abcdef0",
    "Ec2Config": {
        "SubnetArn": "arn:aws:ec2:us-east-2:111222333444:subnet/
subnet-1234567890abcdef1",
        "SecurityGroupArns": [
            "arn:aws:ec2:us-east-2:111222333444:security-group/sg-1234567890abcdef2"
        ]
    },
    "FileSystemAccessRoleArn": "arn:aws:iam::111222333444:role/
AwsDataSyncFullAccessNew",
    "InTransitEncryption": "TLS1_2",
    "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/loc-
abcdef01234567890",
    "LocationUri": "efs://us-east-2.fs-021345abcdef6789/",
    "Subdirectory": "/mount/path",
    "Tags": [{
        "Key": "Name",
        "Value": "ElasticFileSystem-1"
    }]
```

```
}
```

示例响应

响应将返回 Amazon EFS 文件系统的 ARN。

```
{
  "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-12abcdef012345678"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationFsxLustre

为 Amazon FSx for Lustre 文件系统创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。

在开始之前，请务必了解 Lustre 文件 FSx 系统的 DataSync访问方式。

请求语法

```
{
   "FsxFilesystemArn": "string",
   "SecurityGroupArns": [ "string" ],
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

FsxFilesystemArn

指定 for Lustre 文件系统的亚马逊资源名称 (ARN)。 FSx

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):fsx:[a-z\-0-9]+:[0-9]{12}:file-system/fs-[0-9a-f]+$`

是否必需：是

SecurityGroupArns

指定最多五个安全组的 Amazon 资源名称 (ARNs)，这些安全组提供 FSx 对您的 for Lustre 文件系统的访问权限。

安全组必须能够访问文件系统的端口。文件系统还必须允许来自安全组的访问。有关文件系统访问的信息，请参阅 Amazon FSx for Lustre 用户指南。

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

是否必需：是

Subdirectory

为您 FSx 的 for Lustre 文件系统指定装载路径。该路径可包含子目录。

当该位置用作源时，会从装载路径 DataSync 读取数据。当该位置用作目标时， DataSync 会将数据写入装载路径。如果不包含此参数，则 DataSync 使用文件系统的根目录 (/)。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$`

必需：否

Tags

指定帮助您分类、筛选和搜索 AWS 资源的标签。我们建议至少为您的位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
   "LocationArn": "string"
```

```
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您创建的 for Lustre 文件系统位置 FSx 的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationFsxOntap

为 Amazon FSx for NetApp ONTAP 文件系统创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。

在开始之前，请务必了解 ONTAP 文件 FSx 系统的 DataSync访问方式。

请求语法

```
{
   "Protocol": {
      "NFS": {
         "MountOptions": {
            "Version": "string"
         }
      },
      "SMB": {
         "CmkSecretConfig": {
            "KmsKeyArn": "string",
            "SecretArn": "string"
         },
         "CustomSecretConfig": {
            "SecretAccessRoleArn": "string",
            "SecretArn": "string"
         },
         "Domain": "string",
         "ManagedSecretConfig": {
            "SecretArn": "string"
         },
         "MountOptions": {
            "Version": "string"
         },
         "Password": "string",
         "User": "string"
      }
   },
   "SecurityGroupArns": [ "string" ],
   "StorageVirtualMachineArn": "string",
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
```

```
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

Protocol

指定用于访问您的 AWS DataSync Amazon FSx 文件系统的数据传输协议。

类型：FsxProtocol 对象

是否必需：是

SecurityGroupArns

指定提供对文件系统首选子网访问权限的 Amazon EC2 安全组。

安全组必须允许以下端口的出站流量（取决于您使用的协议）：

- 网络文件系统（NFS）：TCP 端口 111、635 和 2049
- 服务器消息块（SMB）：TCP 端口 445

文件系统的安全组还必须允许同一端口上的入站流量。

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

是否必需：是

StorageVirtualMachineArn

指定要将数据复制到或从中复制数据的文件系统中的存储虚拟机（SVM）的 ARN。

类型：字符串

长度约束：最大长度为 162。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):fsx:[a-z\-0-9]+:[0-9]{12}:storage-virtual-machine/fs-[0-9a-f]+/svm-[0-9a-f]{17,}$`

是否必需：是

Subdirectory

指定要传入或传出的 SVM 中的文件共享路径。

您可以指定连接路径（也称为挂载点）、qtree 路径（用于 NFS 文件共享）或共享名称（用于 SMB 文件共享）。例如，您的挂载路径可能是 `/vol1`、`/vol1/tree1` 或 `/share1`。

> **Note**
>
> 不要在 SVM 的根卷中指定连接路径。有关更多信息，请参阅《Amazon FSx FSx for ONTAP 用户指南》中的管理 NetApp ONTAP 存储虚拟机。

类型：字符串

长度限制：长度上限为 255。

模式：`^[^\u0000\u0085\u2028\u2029\r\n]{1,255}$`

必需：否

Tags

指定帮助您分类、筛选和搜索 AWS 资源的标签。我们建议至少为您的位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
```

```
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

指定您创建的 ONTAP 文件系统位置的 ARN。 FSx

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationFsxOpenZfs

为亚马逊 FSx 版 OpenZFS 文件系统创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。

在开始之前，请务必了解 OpenZFS 文件 FSx 系统的 DataSync访问方式。

> **Note**
>
> `CreateLocationFsxOpenZfs` 操作不支持与 SMB 相关的请求参数。

请求语法

```
{
   "FsxFilesystemArn": "string",
   "Protocol": {
      "NFS": {
         "MountOptions": {
            "Version": "string"
         }
      },
      "SMB": {
         "CmkSecretConfig": {
            "KmsKeyArn": "string",
            "SecretArn": "string"
         },
         "CustomSecretConfig": {
            "SecretAccessRoleArn": "string",
            "SecretArn": "string"
         },
         "Domain": "string",
         "ManagedSecretConfig": {
            "SecretArn": "string"
         },
         "MountOptions": {
            "Version": "string"
         },
         "Password": "string",
         "User": "string"
      }
   },
```

```
   "SecurityGroupArns": [ "string" ],
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

FsxFilesystemArn

FSx 适用于 OpenZFS 的文件系统的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):fsx:[a-z\-0-9]+:[0-9]{12}:file-system/fs-[0-9a-f]+$`

是否必需：是

Protocol

AWS DataSync 用于访问您的文件系统的协议类型。

类型：FsxProtocol 对象

是否必需：是

SecurityGroupArns

用于为 ARNs OpenZFS 文件系统配置的安全组。 FSx

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

是否必需：是

Subdirectory

位置路径中必须以开头的子目录。`/fsx` DataSync 使用此子目录读取或写入数据（取决于文件系统是源位置还是目标位置）。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[^\u0000\u0085\u2028\u2029\r\n]{1,4096}$`

必需：否

Tags

表示要添加到资源的标签的键值对。该值可为空字符串。该值可帮助您管理、筛选和搜索资源。我们建议您为自己的位置创建名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您创建的 OpenZFS 文件系统位置的 ARN。 FSx

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3

- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationFsxWindows

FSx 为亚马逊 Windows 文件服务器文件系统创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。

在开始之前，请务必了解 Windows 文件服务器文件系统的 DataSync访问 FSx 方式。

请求语法

```
{
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "Domain": "string",
   "FsxFilesystemArn": "string",
   "Password": "string",
   "SecurityGroupArns": [ "string" ],
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ],
   "User": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

CmkSecretConfig

指定由客户 DataSync AWS KMS key管理的密钥的配置信息，其中包括 DataSync 用于访问特定 FSx Windows 存储位置的密码。

当您在CreateLocationFsxWindows请求中包含此参数时，您仅提供 KMS 密钥 ARN。DataSync 使用此 KMS 密钥和Password您为指定的密钥来创建存储位置访问凭证的 DataSync托管密钥。

确保该用户 DataSync 有权访问您指定的 KMS 密钥。有关更多信息，请参阅使用使用自定义 AWS KMS key密钥加密的服务托管密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 Password）或 CustomSecretConfig（不带 Password）为 CreateLocationFsxWindows 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的 Secr FSx ets Manager 密钥的配置信息，其中 Windows 文件服务器存储位置的密码以纯文本形式存储在 Secrets Manager 中。此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。有关更多信息，请参阅使用您管理的密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 Password）或 CustomSecretConfig（不带 Password）为 CreateLocationFsxWindows 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CustomSecretConfig 对象

必需：否

Domain

指定 Windows 文件服务器文件系统所属 FSx 的 Windows 域的名称。

如果您的环境中有多个 Active Directory 域，则配置此参数可确保这些域 DataSync 连接到正确的文件系统。

类型：字符串

长度约束：最大长度为 253。

模式：`^[A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252}$`

必需：否

FsxFilesystemArn

为 Windows 文件服务器文件系统指定亚马逊资源名称 (ARN)。 FSx

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):fsx:[a-z\-0-9]+:[0-9]{12}:file-system/fs-[0-9a-f]+$`

是否必需：是

Password

指定有权装载和访问您 FSx 的 Windows 文件服务器文件系统中的文件、文件夹和文件元数据的用户的密码。

类型：字符串

长度约束：最大长度为 104。

模式：`^.{0,104}$`

必需：否

SecurityGroupArns

指定提供对文件系统首选子网访问权限的 Amazon EC2 安全组的 ARN。

您指定的安全组必须能够与文件系统的安全组通信。有关为文件系统访问配置安全组的信息，请参阅《Amazon f FSx or Windows 文件服务器用户指南》。

> **Note**
>
> 如果您选择的安全组不允许从其自身进行连接，请执行以下操作之一：
>
> - 将安全组配置为允许其在自身内部进行通信。

- 选择可与挂载目标的安全组进行通信的其他安全组。

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

是否必需：是

Subdirectory

使用正斜杠指定文件系统的挂载路径。这是 DataSync 读取或写入数据的地方（取决于这是源位置还是目标位置）。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$`

必需：否

Tags

指定帮助您分类、筛选和搜索 AWS 资源的标签。我们建议至少为您的位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

User

指定有权装载和访问您 FSx 的 Windows 文件服务器文件系统中的文件、文件夹和文件元数据的用户。

有关选择具有适当访问权限级别的用户进行传输的信息，请参阅 Windows 文件服务器位置 FSx 所需的权限。

类型：字符串

长度约束：最大长度为 104。

模式：^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$

必需：是

响应语法

```
{
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您创建的 Windows 文件服务器文件系统位置的 ARN。 FSx

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationHdfs

为 Hadoop 分布式文件系统 (HDFS) 创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。

在开始之前，请务必了解如何 DataSync访问 HDFS 集群。

请求语法

```
{
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "BlockSize": number,
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "KerberosKeytab": blob,
   "KerberosKrb5Conf": blob,
   "KerberosPrincipal": "string",
   "KmsKeyProviderUri": "string",
   "NameNodes": [
      {
         "Hostname": "string",
         "Port": number
      }
   ],
   "QopConfiguration": {
      "DataTransferProtection": "string",
      "RpcProtection": "string"
   },
   "ReplicationFactor": number,
   "SimpleUser": "string",
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
```

```
    ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArns

可以连接到您的 HDFS 集群的 DataSync 代理的 Amazon 资源名称 (ARNs)。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

是否必需：是

AuthenticationType

用于确定用户身份的身份验证的类型。

类型：字符串

有效值：`SIMPLE | KERBEROS`

是否必需：是

BlockSize

要写入 HDFS 集群的数据块大小。块大小必须是 512 字节的倍数。默认块大小为 128 兆字节（MiB）。

类型：整数

有效范围：最小值为 1048576。最大值为 1073741824。

必需：否

CmkSecretConfig

指定由客户 DataSync管理的密钥的配置信息，其中包括用于访问特定 Hadoop 分布式文件系统 (HDFS) 存储位置的 Kerberos 密钥表。 DataSync AWS KMS key

当您在CreateLocationHdfs请求中包含此参数时，您仅提供 KMS 密钥 ARN。 DataSync 使用此 KMS 密钥和KerberosKeytab您为指定的密钥来创建存储位置访问凭证的 DataSync托管密钥。

确保该用户 DataSync 有权访问您指定的 KMS 密钥。有关更多信息，请参阅使用使用自定义 AWS KMS key密钥加密的服务托管密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 KerberosKeytab）或 CustomSecretConfig（不带 KerberosKeytab）为 CreateLocationHdfs 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的 Secrets Manager 密钥的配置信息，其中 HDFS 存储位置的 Kerberos 密钥表以二进制形式存储在 Secrets Manager 中。此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。有关更多信息，请参阅使用您管理的密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 KerberosKeytab）或 CustomSecretConfig（不带 KerberosKeytab）为 CreateLocationHdfs 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CustomSecretConfig 对象

必需：否

KerberosKeytab

Kerberos 密钥表（keytab），其中包含定义的 Kerberos 主体和加密密钥之间的映射。您可以通过提供文件的地址从该文件加载密钥表。

> **Note**
>
> 如果为 `AuthenticationType` 指定了 `KERBEROS`，则此参数为必填项。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 65536。

必需：否

KerberosKrb5Conf

包含 Kerberos 配置信息的 `krb5.conf` 文件。您可以通过提供文件的地址来加载 `krb5.conf` 文件。如果您使用的是 AWS CLI，它将为您执行 base64 编码。否则，应提供 base64 编码文本。

> **Note**
>
> 如果为 `AuthenticationType` 指定了 `KERBEROS`，则此参数为必填项。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 131072。

必需：否

KerberosPrincipal

有权访问 HDFS 集群上的文件和文件夹的 Kerberos 主体。

> **Note**
>
> 如果为 `AuthenticationType` 指定了 `KERBEROS`，则此参数为必填项。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：^.+$

必需：否

KmsKeyProviderUri

HDFS 集群的密钥管理服务器（KMS）的 URI。

类型：字符串

长度限制：最小长度为 1。最大长度为 255。

模式：^kms:\/\/http[s]?@(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-]*[A-Za-z0-9])(;(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-]*[A-Za-z0-9]))*:[0-9]{1,5}\/kms$

必需：否

NameNodes

管理 HDFS 命名空间的。 NameNode NameNode 执行诸如打开、关闭和重命名文件和目录之类的操作。 NameNode 包含用于将数据块映射到的信息 DataNodes。您只能使用一个 NameNode。

类型：HdfsNameNode 对象数组

数组成员：最少 1 个物品。

是否必需：是

QopConfiguration

保护质量（QOP）配置指定在 Hadoop Distributed File System（HDFS）集群上配置的远程程序调用（RPC）和数据传输保护设置。如果未指定 QopConfiguration，则 RpcProtection 和 DataTransferProtection 默认值为 PRIVACY。如果设置 RpcProtection 或 DataTransferProtection，则其他参数采用相同的值。

类型：QopConfiguration 对象

必需：否

ReplicationFactor

写入 HDFS 集群时要将数据复制到的数量。 DataNodes 默认情况下，数据将复制到三个 DataNodes。

类型：整数

有效范围：最小值为 1。最大值为 512。

必需：否

SimpleUser

用于在主机操作系统上标识客户端的用户名。

> **Note**
>
> 如果为 AuthenticationType 指定了 SIMPLE，则此参数为必填项。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：^[_.A-Za-z0-9][-_.A-Za-z0-9]*$

必需：否

Subdirectory

HDFS 集群中的子目录。此子目录用于从 HDFS 集群读取数据或向其写入数据。如果未指定子目录，则默认为 /。

类型：字符串

长度约束：最大长度为 4096。

模式：^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$

必需：否

Tags

表示要添加到位置的标签的键值对。该值可为空字符串。建议您使用标签来命名资源。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您创建的源 HDFS 集群位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationNfs

为网络文件系统 (NFS) 文件服务器创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。

在开始之前，请务必了解如何 DataSync访问 NFS 文件服务器。

请求语法

```
{
   "MountOptions": {
      "Version": "string"
   },
   "OnPremConfig": {
      "AgentArns": [ "string" ]
   },
   "ServerHostname": "string",
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

MountOptions

指定 DataSync 可用于装载 NFS 文件服务器的选项。

类型：NfsMountOptions 对象

必需：否

OnPremConfig

指定可以连接到您的 NFS 文件服务器的 DataSync 代理的 Amazon 资源名称 (ARN)。

您可以指定多个代理。有关更多信息，请参阅使用多个 DataSync 代理。

类型：OnPremConfig 对象

是否必需：是

ServerHostname

指定 DataSync 代理所连接的 NFS 文件服务器的 DNS 名称IPv4 或 IP 地址（或 IPv6）。

类型：字符串

长度限制：长度上限为 255。

模式：`^(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-:]*[A-Za-z0-9])$`

是否必需：是

Subdirectory

在 NFS 文件服务器中指定 DataSync 要装载的导出路径。

此路径（或路径的子目录）是传入或 DataSync 传出数据的地方。有关为配置导出的信息 DataSync，请参阅访问 NFS 文件服务器。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]+$`

是否必需：是

Tags

指定可帮助您对 AWS 资源进行分类、筛选和搜索的标签。我们建议至少为您的位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
```

```
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您为 NFS 文件服务器创建的传输位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例

以下示例为 NFS 文件服务器创建 DataSync 传输位置。

示例请求

```
{
  "MountOptions": {
     "Version": : "NFS4_0"
     },
  "OnPremConfig": {
    "AgentArn": [ "arn:aws:datasync:us-east-2:111222333444:agent/
agent-0b0addbeef44b3nfs" ]
          },

           "ServerHostname": "MyServer@amazon.com",
           "Subdirectory": "/MyFolder",
           "Tags": [
              {
                "Key": "Name",
                "Value": "FileSystem-1"
              }
           ]
}
```

示例

响应会返回 NFS 位置的 ARN。

示例响应

```
{
  "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-07db7abfc326c50aa"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2

- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationObjectStorage

为对象存储系统创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。无论有没有DataSync代理人，您都可以进行转账。

在开始之前，请务必了解使用对象存储系统的先决条件。 DataSync

请求语法

```
{
   "AccessKey": "string",
   "AgentArns": [ "string" ],
   "BucketName": "string",
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "SecretKey": "string",
   "ServerCertificate": blob,
   "ServerHostname": "string",
   "ServerPort": number,
   "ServerProtocol": "string",
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AccessKey

如果需要凭证来向对象存储服务器进行身份验证，则指定访问密钥（例如，用户名）。

类型：字符串

长度限制：最小长度为 0。最大长度为 200。

模式：`^.*$`

必需：否

AgentArns

（可选）指定可以与您的对象存储系统连接的 DataSync 代理的 Amazon 资源名称 (ARNs)。如果设置的是无代理跨云传输，则无需为此参数指定值。

> **Note**
>
> 首次创建存储位置时，务必确保正确配置此参数。创建代理后，就无法在存储位置添加或删除代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

BucketName

指定传输中涉及的对象桶的名称。

类型：字符串

长度限制：最小长度为 3。最大长度为 63。

模式：`^[a-zA-Z0-9_\-\+\.\(\)\$\p{Zs}]+$`

是否必需：是

CmkSecretConfig

指定由客户 DataSync AWS KMS key管理的密钥（包括 DataSync 用于访问特定对象存储位置的密钥）的配置信息。SecretKey

当您在CreateLocationObjectStorage请求中包含此参数时，您仅提供 KMS 密钥 ARN。DataSync 使用此 KMS 密钥和您为SecretKey参数指定的值来创建存储位置访问凭证的 DataSync 托管密钥。

确保该用户 DataSync 有权访问您指定的 KMS 密钥。有关更多信息，请参阅使用使用自定义 AWS KMS key密钥加密的服务管理密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 SecretKey）或 CustomSecretConfig（不带 SecretKey）为 CreateLocationObjectStorage 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的 Secrets Manager 密钥的配置信息，其中特定对象存储位置的密钥以纯文本形式存储在 Secrets Manager 中。此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。有关更多信息，请参阅使用您管理的密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 SecretKey）或 CustomSecretConfig（不带 SecretKey）为 CreateLocationObjectStorage 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CustomSecretConfig 对象

必需：否

SecretKey

如果需要凭证来向对象存储服务器进行身份验证，则指定私有密钥（例如，密码）。

> **Note**
>
> 如果您使用提供密钥`SecretKey`，但未使用`CmkSecretConfig`或提供密钥配置详细信息`CustomSecretConfig`，则使用您 AWS 账户的 Secrets Manager 密钥 DataSync 存储令牌。

类型：字符串

长度限制：最小长度为 0。最大长度为 200。

模式：`^.*$`

必需：否

ServerCertificate

如果您的对象存储系统使用私有或自签名证书颁发机构 (CA)，则指定一个证书链， DataSync 以便在对象存储系统中进行身份验证。您必须指定具有完整证书链的单个 `.pem` 文件（例如，`file:///home/user/.ssh/object_storage_certificates.pem`）。

证书链可能包括：

- 对象存储系统的证书
- 所有中间证书（如果有）
- 签名 CA 的根证书

您可以将证书串联到一个 `.pem` 文件中（在进行 base64 编码前，文件最大长度可以为 32768 字节）。以下示例 `cat` 命令创建了一个包含三个证书的 `object_storage_certificates.pem` 文件：

```
cat object_server_certificate.pem intermediate_certificate.pem
ca_root_certificate.pem > object_storage_certificates.pem
```

要使用此参数，请配置 `ServerProtocol` 为 `HTTPS`。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 32768。

必需：否

ServerHostname

指定 DataSync 代理所连接的对象存储服务器的域名IPv4 或 IP 地址（或 IPv6）。

类型：字符串

长度限制：长度上限为 255。

模式：`^(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-:]*[A-Za-z0-9])$`

是否必需：是

ServerPort

指定对象存储服务器接受入站网络流量的端口（例如，端口 443）。

类型：整数

有效范围：最小值为 1。最大值为 65536。

必需：否

ServerProtocol

指定对象存储服务器通信时使用的协议。如果未指定，则默认值为 HTTPS。

类型：字符串

有效值：`HTTPS | HTTP`

必需：否

Subdirectory

指定对象存储服务器的对象前缀。如果这是源位置，则 DataSync 仅复制带有此前缀的对象。如果这是目标位置，则 DataSync 写入带有此前缀的所有对象。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$`

必需：否

Tags

指定表示要添加到资源的标签的键值对。标签可帮助您管理、筛选和搜索资源。我们建议为您的位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

指定您创建的对象存储系统位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationS3

为 Amazon S3 存储桶创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目标。

> **Important**
>
> 开始之前，请务必阅读以下主题：
>
> - Amazon S3 位置的存储类考虑因素
> - 使用时评估 S3 请求成本 DataSync

有关更多信息，请参阅配置与 Amazon S3 的传输。

请求语法

```
{
   "AgentArns": [ "string" ],
   "S3BucketArn": "string",
   "S3Config": {
      "BucketAccessRoleArn": "string"
   },
   "S3StorageClass": "string",
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArns

（仅限 Outposts 上的 Amazon S3）指定 Outpost 上 DataSync 代理的亚马逊资源名称 (ARN)。

有关更多信息，请参阅在上部署 DataSync 代理 AWS Outposts。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

S3BucketArn

指定要用作位置的 S3 存储桶的 ARN。（稍后创建 DataSync 任务时，您可以指定此位置是传输源还是目的地。）

如果您的 S3 存储桶位于 AWS Outposts 资源上，则必须指定 Amazon S3 接入点。有关更多信息，请参阅 Amazon S3 用户指南中的使用 Amazon S3 接入点管理数据访问。

类型：字符串

长度限制：最大长度为 268。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3:[a-z\-0-9]*:[0-9]{12}:accesspoint[/:][a-zA-Z0-9\-.]{1,63}$|^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3-outposts:[a-z\-0-9]+:[0-9]{12}:outpost[/:][a-zA-Z0-9\-]{1,63}[/:]accesspoint[/:][a-zA-Z0-9\-]{1,63}$|^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3:::[a-zA-Z0-9.\-_]{1,255}$`

是否必需：是

S3Config

指定用于访问您的 S3 存储桶的 AWS Identity and Access Management (IAM) 角色的 DataSync 亚马逊资源名称 (ARN)。

有关更多信息，请参阅提供 DataSync 对 S3 存储桶的访问权限。

类型：S3Config 对象

是否必需：是

S3StorageClass

指定当 Amazon S3 作为传输目标时，您希望对象使用的存储类别。

对于中的存储桶 AWS 区域，存储类默认为。STANDARD对于开启的存储桶 AWS Outposts，存储类默认为。OUTPOSTS

有关更多信息，请参阅 Amazon S3 传输的存储类别注意事项。

类型：字符串

有效值：STANDARD | STANDARD_IA | ONEZONE_IA | INTELLIGENT_TIERING | GLACIER | DEEP_ARCHIVE | OUTPOSTS | GLACIER_INSTANT_RETRIEVAL

必需：否

Subdirectory

在 S3 存储桶中指定 DataSync 读取或写入的前缀（取决于存储桶是源位置还是目标位置）。

> **Note**
>
> DataSync 无法传输前缀以斜杠 (/) 开头或包含///./、或/../模式的对象。例如：
>
> - /photos
> - photos//2006/January
> - photos/./2006/February
> - photos/../2006/March

类型：字符串

长度约束：最大长度为 4096。

模式：^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$

必需：否

Tags

指定可帮助您对 AWS 资源进行分类、筛选和搜索的标签。我们建议至少为您的传输位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

响应语法

```
{
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您创建的 S3 位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateLocationSmb

为服务器消息块 (SMB) 文件服务器创建传输位置。 AWS DataSync 可以将此位置用作传输数据的源或目的地。

在开始之前，请务必了解如何 DataSync 访问 SMB 文件服务器。有关更多信息，请参阅[提供 DataSync 对 SMB 文件服务器的访问权限](#)。

请求语法

```
{
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "DnsIpAddresses": [ "string" ],
   "Domain": "string",
   "KerberosKeytab": blob,
   "KerberosKrb5Conf": blob,
   "KerberosPrincipal": "string",
   "MountOptions": {
      "Version": "string"
   },
   "Password": "string",
   "ServerHostname": "string",
   "Subdirectory": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ],
   "User": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArns

指定 DataSync 可以连接到 SMB 文件服务器的一个或多个代理。使用代理的 Amazon 资源名称 (ARN) 指定代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

是否必需：是

AuthenticationType

指定用于连接您的 DataSync SMB 文件服务器的身份验证协议。 DataSync 支持`NTLM`（默认）和`KERBEROS`身份验证。

有关更多信息，请参阅提供 DataSync 对 SMB 文件服务器的访问权限。

类型：字符串

有效值：`NTLM | KERBEROS`

必需：否

CmkSecretConfig

指定用于访问特定 SMB 存储位置的 DataSync托管密钥的配置信息，可以是`Password`或`KerberosKeytab`NTLM（分别是（默认 DataSync ）和`KERBEROS`身份验证类型），该密钥由客户管理。 AWS KMS key

当您在`CreateLocationSmbRequest`请求中包含此参数时，您仅提供 KMS 密钥 ARN。DataSync 将此 KMS 密钥与`KerberosKeytab`您指定的`Password`或一起使用来创建存储位置访问凭证的 DataSync托管密钥。

确保该用户 DataSync 有权访问您指定的 KMS 密钥。有关更多信息，请参阅使用使用自定义 AWS KMS key密钥加密的服务管理密钥。

> **Note**
>
> 您可以使用CmkSecretConfig（使用Password或KerberosKeytab）或CustomSecretConfig（不带任何Password和KerberosKeytab）为CreateLocationSmbRequest请求提供凭证。不要为同一个请求同时提供两个CmkSecretConfig和CustomSecretConfig参数。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的 Secrets Manager 密钥的配置信息，其中 SMB 存储位置凭据以纯文本（对于Password）或二进制（用于）存储在 Secrets Manager 中。KerberosKeytab此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。有关更多信息，请参阅使用您管理的密钥。

> **Note**
>
> 可使用 CmkSecretConfig（带有 SasConfiguration）或 CustomSecretConfig（不带 SasConfiguration）为 CreateLocationSmbRequest 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CustomSecretConfig 对象

必需：否

DnsIpAddresses

为您的 SMB 文件服务器所属的 DNS 服务器指定 IPv4 或 IPv6 地址。此参数仅在 AuthenticationType 设置为 KERBEROS 时适用。

如果您的环境中有多个域，则配置此参数 DataSync 可确保连接到正确的 SMB 文件服务器。

类型：字符串数组

数组成员：最多 2 项。

长度限制：最小长度为 7。最大长度为 39。

模式：`\A((25[0-5]|2[0-4]\d|[0-1]?\d?\d)(\.(25[0-5]|2[0-4]\d|[0-1]?\d?\d)){3}|([0-9a-fA-F]{1,4}:){7,7}[0-9a-fA-F]{1,4}|([0-9a-fA-F]{1,4}:){1,7}:|([0-9a-fA-F]{1,4}:){1,6}:[0-9a-fA-F]{1,4}|([0-9a-fA-F]{1,4}:){1,5}(:[0-9a-fA-F]{1,4}){1,2}|([0-9a-fA-F]{1,4}:){1,4}(:[0-9a-fA-F]{1,4}){1,3}|([0-9a-fA-F]{1,4}:){1,3}(:[0-9a-fA-F]{1,4}){1,4}|([0-9a-fA-F]{1,4}:){1,2}(:[0-9a-fA-F]{1,4}){1,5}|[0-9a-fA-F]{1,4}:((:[0-9a-fA-F]{1,4}){1,6}))\z`

必需：否

Domain

指定您的 SMB 文件服务器所属的 Windows 域名称。此参数仅在 `AuthenticationType` 设置为 `NTLM` 时适用。

如果您的环境中有多个域，则配置此参数可确保 DataSync 连接到正确的文件服务器。

类型：字符串

长度约束：最大长度为 253。

模式：`^[A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252}$`

必需：否

KerberosKeytab

指定 Kerberos 密钥表（keytab）文件，其中包括 Kerberos 主体和加密密钥之间的映射。

为避免出现任务执行错误，请确保用于创建 keytab 文件的 Kerberos 主体与为 `KerberosPrincipal` 指定的内容完全匹配。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 65536。

必需：否

KerberosKrb5Conf

指定用于定义 Kerberos 领域配置的 Kerberos 配置文件（`krb5.conf`）。

该文件必须采用 base64 编码。如果您使用的是 AWS CLI，则编码已为您完成。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 131072。

必需：否

KerberosPrincipal

指定 Kerberos 主体，即 Kerberos 领域中有权访问 SMB 文件服务器中文件、文件夹和文件元数据的身份。

Kerberos 主体可能看起来像 `HOST/kerberosuser@MYDOMAIN.ORG`。

主体名称区分大小写。如果您为此参数指定的主体与您用于创建 keytab 文件的主体不完全匹配，则 DataSync 任务执行将失败。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：`^.+$`

必需：否

MountOptions

指定用于访问您的 SMB 文件服务器的 DataSync SMB 协议的版本。

类型：SmbMountOptions 对象

必需：否

Password

指定可以挂载 SMB 文件服务器并有权限访问传输中涉及的文件和文件夹的用户的密码。此参数仅在 `AuthenticationType` 设置为 `NTLM` 时适用。

类型：字符串

长度约束：最大长度为 104。

模式：`^.{0,104}$`

必需：否

ServerHostname

指定您的 DataSync 代理所连接的 SMB 文件服务器的域名IPv4 或 IP 地址（或 IPv6）。

> **Note**
>
> 如果使用的是 Kerberos 身份验证，则必须指定域名。

类型：字符串

长度限制：长度上限为 255。

模式：`^(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-:]*[A-Za-z0-9])$`

是否必需：是

Subdirectory

指定 SMB 文件服务器导出的共享的名称，该共享 DataSync将在其中读取或写入数据。您可以在共享路径中包含子目录（例如，`/path/to/subdirectory`）。确保网络中的其他 SMB 客户端也可以挂载此路径。

要复制子目录中的所有数据， DataSync 必须能够装载 SMB 共享并访问其所有数据。有关更多信息，请参阅提供 DataSync 对 SMB 文件服务器的访问权限。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$`

是否必需：是

Tags

指定帮助您分类、筛选和搜索 AWS 资源的标签。我们建议至少为您的位置创建一个名称标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

User

指定可以挂载和访问 SMB 文件服务器中文件、文件夹和文件元数据的用户。此参数仅在 AuthenticationType 设置为 NTLM 时适用。

有关选择具有适当访问权限级别的用户进行传输的信息，请参阅提供对 SMB 文件服务器的 DataSync 访问权限。

类型：字符串

长度约束：最大长度为 104。

模式：^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$

必需：否

响应语法

```
{
   "LocationArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

LocationArn

您创建的 SMB 位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

以下示例为 SMB 文件服务器创建一个位置。

```
{
   "AgentArns":[
      "arn:aws:datasync:us-east-2:111222333444:agent/agent-0b0addbeef44b3nfs",
      "arn:aws:datasync:us-east-2:111222333444:agent/agent-2345noo35nnee1123ovo3"
   ],
   "Domain":"AMAZON",
   "MountOptions":{
      "Version":"SMB3"
   },
   "Password":"string",
   "ServerHostname":"MyServer.amazon.com",
   "Subdirectory":"share",
   "Tags":[
      {
         "Key":"department",
         "Value":"finance"
      }
   ],
   "User":"user-1"
}
```

示例响应

响应会返回您的 SMB 文件服务器的位置 ARN。

```
{
    "LocationArn": "arn:aws:datasync:us-east-1:111222333444:location/
loc-0f01451b140b2af49"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版 SDK
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

CreateTask

配置任务，在其中定义 AWS DataSync 传输数据的位置和方式。

任务包括源位置、目标位置和传输选项（如带宽限制、时间计划等）。

> **Important**
>
> 如果您计划在 Amazon S3 地点之间传输数据，请在开始之前查看 DataSync 如何影响您的 S3 请求费用和DataSync 定价页面。

请求语法

```
{
   "CloudWatchLogGroupArn": "string",
   "DestinationLocationArn": "string",
   "Excludes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "Includes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "ManifestConfig": {
      "Action": "string",
      "Format": "string",
      "Source": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "ManifestObjectPath": "string",
            "ManifestObjectVersionId": "string",
            "S3BucketArn": "string"
         }
      }
   },
   "Name": "string",
```

```
   "Options": {
      "Atime": "string",
      "BytesPerSecond": number,
      "Gid": "string",
      "LogLevel": "string",
      "Mtime": "string",
      "ObjectTags": "string",
      "OverwriteMode": "string",
      "PosixPermissions": "string",
      "PreserveDeletedFiles": "string",
      "PreserveDevices": "string",
      "SecurityDescriptorCopyFlags": "string",
      "TaskQueueing": "string",
      "TransferMode": "string",
      "Uid": "string",
      "VerifyMode": "string"
   },
   "Schedule": {
      "ScheduleExpression": "string",
      "Status": "string"
   },
   "SourceLocationArn": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ],
   "TaskMode": "string",
   "TaskReportConfig": {
      "Destination": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "S3BucketArn": "string",
            "Subdirectory": "string"
         }
      },
      "ObjectVersionIds": "string",
      "OutputType": "string",
      "Overrides": {
         "Deleted": {
            "ReportLevel": "string"
         },
         "Skipped": {
```

```
            "ReportLevel": "string"
         },
         "Transferred": {
            "ReportLevel": "string"
         },
         "Verified": {
            "ReportLevel": "string"
         }
      },
      "ReportLevel": "string"
   }
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

CloudWatchLogGroupArn

指定用于监控任务的亚马逊 CloudWatch 日志组的亚马逊资源名称 (ARN)。

对于增强模式任务，您无需指定任何内容。 DataSync自动将日志发送到名为的 CloudWatch 日志组`/aws/datasync`。

有关更多信息，请参阅使用 CloudWatch 日志监控数据传输。

类型：字符串

长度约束：最大长度为 562。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):logs:[a-z\-0-9]+:[0-9]{12}:log-group:([^:\*]*)(:\*)?$`

必需：否

DestinationLocationArn

指定传输目标位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

Excludes

指定排除过滤器，用于定义源位置中您不想 DataSync 传输的文件、对象和文件夹。有关更多信息和示例，请参阅使用过滤器指定 DataSync 传输的内容。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

必需：否

Includes

指定包含过滤器，用于定义源位置中 DataSync 要传输的文件、对象和文件夹。有关更多信息和示例，请参阅使用过滤器指定 DataSync 传输的内容。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

必需：否

ManifestConfig

配置清单，这是您要 DataSync 传输的文件或对象的列表。有关更多信息和配置示例，请参阅使用清单指定 DataSync 传输的内容。

使用此参数时，您的来电者身份（您使用的 DataSync 角色）必须具有`iam:PassRole`权限。该AWSDataSyncFullAccess策略包含此权限。

类型：ManifestConfig 对象

必需：否

Name

指定任务的名称。

类型：字符串

长度限制：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

必需：否

Options

指定任务的设置，例如保留文件元数据、验证数据完整性等选项。

类型：Options 对象

必需：否

Schedule

指定任务运行的时间计划。有关更多信息，请参阅计划任务。

类型：TaskSchedule 对象

必需：否

SourceLocationArn

指定传输源位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

Tags

指定要应用于任务的标签。

标签是键值对，可帮助您管理、筛选和搜索 DataSync 资源。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

TaskMode

为数据传输指定以下任务模式之一：

- ENHANCED：传输几乎无限的对象数量，传输性能高于基本模式。增强模式任务通过并行列示、准备、传输和验证数据来优化数据传输过程。增强模式目前可用于 Amazon S3 位置之间的传输、Azure Blob 和 Amazon S3 之间的无代理传输，以及其他云和 Amazon S3 之间的无代理传输。

  > **Note**
  >
  > 要创建增强模式任务，用于调用 `CreateTask` 操作的 IAM 角色必须拥有 `iam:CreateServiceLinkedRole` 权限。

- BASIC（默认）-在 AWS 存储和所有其他支持的 DataSync 位置之间传输文件或对象。基本模式任务在数据集中文件、对象和目录数量方面存在配额限制。基本模式按顺序准备、传输和验证数据，因此对于大多数工作负载来说，基本模式的处理速度慢于增强模式。

有关更多信息，请参阅了解任务模式差异。

类型：字符串

有效值：`BASIC | ENHANCED`

必需：否

TaskReportConfig

指定您想要如何配置任务报告，该报告提供有关您的 DataSync 转移的详细信息。有关更多信息，请参阅使用任务报告监控您的 DataSync转账。

使用此参数时，您的来电者身份（您使用的 DataSync 角色）必须具有`iam:PassRole`权限。该AWSDataSyncFullAccess策略包含此权限。

类型：TaskReportConfig 对象

必需：否

响应语法

```
{
```

```
   "TaskArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

TaskArn

任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

增强模式任务的请求示例

以下示例创建了一个使用增强模式的 DataSync 任务。

与创建基本模式任务时不同，您无需指定 Amazon CloudWatch 日志组。使用增强模式任务，DataSync 会自动将任务日志发送到名为的日志组/aws/datasync。如果您的日志组不存在 AWS 区域，则在 DataSync 创建任务时代表您创建该日志组。

```
{
  "SourceLocationArn": "arn:aws:datasync:us-
east-1:111222333444:location/1111aaaa2222bbbb3",
  "DestinationLocationArn": "arn:aws:datasync:us-
east-1:111222333444:location/0000zzzz1111yyyy2",
  "Name": "My Enhanced mode task",
  "TaskMode": "ENHANCED",
  "Options": {
    "TransferMode": "CHANGED",
    "VerifyMode": "ONLY_FILES_TRANSFERRED",
    "ObjectTags": "PRESERVE",
    "LogLevel": "TRANSFER"
  }
}
```

基本模式任务的请求示例

以下示例创建了一个使用 “基本” 模式的 DataSync 任务。

```
{
    "SourceLocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-1111aaaa2222bbbb3",
    "DestinationLocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-0000zzzz1111yyyy2",
    "Name": "My Basic mode task",
    "TaskMode": "BASIC",
    "Options": {
        "Atime": "BEST_EFFORT",
        "Gid": "NONE",
        "Mtime": "PRESERVE",
        "PosixPermissions": "PRESERVE",
        "PreserveDevices": "NONE",
        "PreserveDeletedFiles": "PRESERVE",
        "Uid": "NONE",
        "VerifyMode": "ONLY_FILES_TRANSFERRED"
    },
    "Schedule": {
        "ScheduleExpression": "0 12 ? * SUN,WED *"
    },
```

```
    "CloudWatchLogGroupArn": "arn:aws:logs:us-east-2:111222333444:log-group:/log-group-
name:*",
    "Tags": [
        {
            "Key": "Name",
            "Value": "Migration-wave-1"
        }
    ]
}
```

示例响应

以下响应包括所创建任务的 ARN。

```
{
  "TaskArn": "arn:aws:datasync:us-east-2:111222333444:task/task-08de6e6697796f026"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DeleteAgent

从您的中移除 AWS DataSync 代理资源 AWS 账户。

请记住，此操作（无法撤消）不会将代理的虚拟机 (VM) 或 Amazon EC2 实例从存储环境中移除。接下来的步骤中，您可以从存储环境中删除 VM 或实例，也可以再次用它来激活一个新代理。

请求语法

```
{
   "AgentArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArn

要删除的代理的 Amazon 资源名称（ARN）。使用该ListAgents操作返回您账户的代理列表，然后 AWS 区域。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$

是否必需：是

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DeleteLocation

从中删除转移位置资源 AWS DataSync。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

要删除位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DeleteTask

从中删除转移任务资源 AWS DataSync。

请求语法

```
{
   "TaskArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

TaskArn

指定您要删除的任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$`

是否必需：是

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeAgent

返回有关 AWS DataSync 代理的信息，例如其名称、服务端点类型和状态。

请求语法

```
{
   "AgentArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArn

指定您想要获取相关信息的 DataSync 代理的 Amazon 资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "AgentArn": "string",
   "CreationTime": number,
   "EndpointType": "string",
   "LastConnectionTime": number,
   "Name": "string",
   "Platform": {
      "Version": "string"
   },
   "PrivateLinkConfig": {
```

```
      "PrivateLinkEndpoint": "string",
      "SecurityGroupArns": [ "string" ],
      "SubnetArns": [ "string" ],
      "VpcEndpointId": "string"
   },
   "Status": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AgentArn

座席的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

CreationTime

激活代理的时间。

类型：时间戳

EndpointType

您的代理连接到的服务端点的类型。

类型：字符串

有效值：`PUBLIC | PRIVATE_LINK | FIPS | FIPS_PRIVATE_LINK`

LastConnectionTime

代理上次与 DataSync服务通信的时间。

类型：时间戳

Name

代理名称。

类型：字符串

长度限制：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

Platform

有关代理的平台相关详细信息，例如版本号。

类型：Platform 对象

PrivateLinkConfig

代理在连接 VPC 服务端点时使用的网络配置。

类型：PrivateLinkConfig 对象

Status

代理的状态。

- 如果状态为 ONLINE，则表示已正确配置代理并且可供使用。
- 如果状态为OFFLINE，则表示代理已与 DataSync 该代理失去联系五分钟或更长时间。这可能是由于多种原因造成。有关更多信息，请参阅如何解决代理离线问题？

类型：字符串

有效值：`ONLINE | OFFLINE`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

以下示例返回请求中指定的代理信息。

```
{
  "AgentArn": "arn:aws:datasync:us-east-2:111122223333:agent/agent-1234567890abcdef0"
}
```

示例响应

以下示例响应描述了使用公共服务端点的代理。

```
{
    "AgentArn": "arn:aws:datasync:us-east-2:111122223333:agent/
agent-1234567890abcdef0",
    "Name": "Data center migration agent",
    "Status": "ONLINE",
    "LastConnectionTime": "2022-10-17T17:21:35.540000+00:00",
    "CreationTime": "2022-10-05T20:52:29.499000+00:00",
    "EndpointType": "PUBLIC",
    "Platform": {
        "Version": "2"
    }
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK

- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationAzureBlob

提供有关如何配置 Microsoft Azure Blob 存储的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定 Azure Blob 存储传输位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "AccessTier": "string",
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "BlobType": "string",
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CreationTime": number,
```

```
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "LocationArn": "string",
   "LocationUri": "string",
   "ManagedSecretConfig": {
      "SecretArn": "string"
   }
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AccessTier

您希望将对象或文件传输到的访问层级。这仅适用于使用该位置作为传输目标的情况。有关更多信息，请参阅访问层。

类型：字符串

有效值：`HOT | COOL | ARCHIVE`

AgentArns

可以与你 ARNs 的 Azure Blob 存储容器连接的 DataSync 代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

AuthenticationType

DataSync 用于访问你的 Azure Blob 存储空间的身份验证方法。 DataSync 可以使用共享访问签名 (SAS) 访问 blob 存储。

类型：字符串

有效值：SAS | NONE

BlobType

在将对象或文件传输到 Azure Blob Storage 时想要的 blob 类型。目前， DataSync 仅支持将数据作为块 blob 移动到 Azure Blob 存储。有关 blob 类型的更多信息，请参阅 Azure Blob Storage 文档。

类型：字符串

有效值：BLOCK

CmkSecretConfig

描述由客户 DataSync AWS KMS key管理的密钥的配置信息，例如 DataSync 用于访问特定存储位置的身份验证令牌。

类型：CmkSecretConfig 对象

CreationTime

您创建的 Azure Blob Storage 传输位置的时间。

类型：时间戳

CustomSecretConfig

描述客户管理的密钥的配置信息，例如用于访问特定存储位置的 DataSync 身份验证令牌，以及提供对密钥的访问权限的客户管理 AWS Identity and Access Management (IAM) 角色。

类型：CustomSecretConfig 对象

LocationArn

Azure Blob Storage 传输位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$

LocationUri

传输中涉及的 Azure Blob Storage 容器的 URL。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

ManagedSecretConfig

描述 DataSync托管密钥的配置信息，例如用于访问特定存储位置的 DataSync 身份验证令牌。DataSync 使用默认 AWS托管的 KMS 密钥在中 AWS Secrets Manager加密此密钥。

类型：ManagedSecretConfig 对象

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK

- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationEfs

提供有关如何配置 Amazon EFS 文件系统的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

您想获取其相关信息的 Amazon EFS 文件系统位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "AccessPointArn": "string",
   "CreationTime": number,
   "Ec2Config": {
      "SecurityGroupArns": [ "string" ],
      "SubnetArn": "string"
   },
   "FileSystemAccessRoleArn": "string",
   "InTransitEncryption": "string",
   "LocationArn": "string",
```

```
   "LocationUri": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AccessPointArn

DataSync 用于访问 Amazon EFS 文件系统的接入点的 ARN。

有关更多信息，请参阅访问受限文件系统。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):elasticfilesystem:[a-z\-0-9]+:[0-9]{12}:access-point/fsap-[0-9a-f]{8,40}$`

CreationTime

创建位置的时间。

类型：时间戳

Ec2Config

AWS DataSync 用于连接您的 Amazon EFS 文件系统的挂载目标之一的子网和安全组。

类型：Ec2Config 对象

FileSystemAccessRoleArn

允许 DataSync 访问您的 Amazon EFS 文件系统的 AWS Identity and Access Management (IAM) 角色。

有关更多信息，请参阅创建用于访问文件系统 DataSync的 IAM 角色。

类型：字符串

长度约束：最大长度为 2048。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):iam::[0-9]{12}:role/.*$`

InTransitEncryption

指明在向 Amazon EFS 文件系统传输数据或从中传输数据时是否 DataSync 使用传输层安全 (TLS) 加密。

类型：字符串

有效值：`NONE | TLS1_2`

LocationArn

Amazon EFS 文件系统位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

Amazon EFS 文件系统位置的 URL。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

以下示例说明如何获取有关 Amazon EFS 文件系统位置的信息。

```
{
  "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-12abcdef012345678"
}
```

示例响应

以下示例返回有关 Amazon EFS 文件系统的位置详情。

```
{
    "CreationTime": 1653319021.353,
    "Ec2Config": {
        "SubnetArn": "arn:aws:ec2:us-east-2:111222333444:subnet/
subnet-1234567890abcdef1",
        "SecurityGroupArns": [
            "arn:aws:ec2:us-east-2:111222333444:security-group/sg-1234567890abcdef2"
        ]
    },
    "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/loc-
abcdef01234567890",
    "LocationUri": "efs://us-east-2.fs-021345abcdef6789/"
}
```

响应示例：描述受限的 Amazon EFS 文件系统的位置

以下示例返回有关访问受限的 Amazon EFS 文件系统的位置详情，包括 `AccessPointArn`、`FileSystemAccessRoleArn` 和 `InTransitEncryption` 元素。

```
{
    "CreationTime": 1653319021.353,
    "AccessPointArn": "arn:aws:elasticfilesystem:us-east-2:111222333444:access-point/
fsap-1234567890abcdef0",
    "Ec2Config": {
```

```
        "SubnetArn": "arn:aws:ec2:us-east-2:111222333444:subnet/
subnet-1234567890abcdef1",
        "SecurityGroupArns": [
            "arn:aws:ec2:us-east-2:111222333444:security-group/sg-1234567890abcdef2"
        ]
    },
    "FileSystemAccessRoleArn": "arn:aws:iam::111222333444:role/
AwsDataSyncFullAccessNew",
    "InTransitEncryption": "TLS1_2",
    "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/loc-
abcdef01234567890",
    "LocationUri": "efs://us-east-2.fs-021345abcdef6789/",
    "Subdirectory": "/mount/path",
    "Tags": [{
        "Key": "Name",
        "Value": "ElasticFileSystem-1"
    }]
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationFsxLustre

提供有关如何配置 Amazon FSx for Lustre 文件系统的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

要描述的 for Lustre 地点的 FSx 亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "CreationTime": number,
   "LocationArn": "string",
   "LocationUri": "string",
   "SecurityGroupArns": [ "string" ]
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

CreationTime

for Lustre 位置 FSx 的创建时间。

类型：时间戳

LocationArn

所描述的 for Lustre 地点 FSx 的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

所描述的 f FSx or Lustre 位置的 URI。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

SecurityGroupArns

为 Lustre 文件系统配置的安全组的 Amazon 资源名称 (ARNs)。 FSx

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationFsxOntap

提供有关如何配置 Amazon FSx for NetApp ONTAP 文件系统的 AWS DataSync 传输位置的详细信息。

> **Note**
>
> 如果您的位置使用 SMB，则 `DescribeLocationFsxOntap` 操作实际上不会返回 `Password`。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定您想要了解的 ONTAP 文件系统位置 FSx 的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "CreationTime": number,
```

```
   "FsxFilesystemArn": "string",
   "LocationArn": "string",
   "LocationUri": "string",
   "Protocol": {
      "NFS": {
         "MountOptions": {
            "Version": "string"
         }
      },
      "SMB": {
         "CmkSecretConfig": {
            "KmsKeyArn": "string",
            "SecretArn": "string"
         },
         "CustomSecretConfig": {
            "SecretAccessRoleArn": "string",
            "SecretArn": "string"
         },
         "Domain": "string",
         "ManagedSecretConfig": {
            "SecretArn": "string"
         },
         "MountOptions": {
            "Version": "string"
         },
         "Password": "string",
         "User": "string"
      }
   },
   "SecurityGroupArns": [ "string" ],
   "StorageVirtualMachineArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

CreationTime

创建位置的时间。

类型：时间戳

FsxFilesystemArn

FSx 适用于 ONTAP 文件系统的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):fsx:[a-z\-0-9]+:[0-9]{12}:file-system/fs-[0-9a-f]+$`

LocationArn

FSx 适用于 ONTAP 文件系统位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

ONTAP 文件系统位置的统一资源标识符 (URI)。 FSx

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

Protocol

指定用于访问您的 AWS DataSync Amazon FSx 文件系统的数据传输协议。

类型：FsxProtocol 对象

SecurityGroupArns

FSx 用于访问您的 DataSync for ONTAP 文件系统的安全组。

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

StorageVirtualMachineArn

您的 for ONTAP 文件系统上 FSx 用于向其中复制数据或从中复制数据的存储虚拟机 (SVM) 的 ARN。

类型：字符串

长度约束：最大长度为 162。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):fsx:[a-z\-0-9]+:[0-9]{12}:storage-virtual-machine/fs-[0-9a-f]+/svm-[0-9a-f]{17,}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK

- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationFsxOpenZfs

提供有关如何配置 Amazon FSx for OpenZFS 文件系统的 AWS DataSync 传输位置的详细信息。

> **Note**
>
> DescribeLocationFsxOpenZfs 操作不支持与 SMB 相关的响应元素。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

要描述的 openZFS 位置的亚马逊资源名称 (ARN)。 FSx

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$

必需：是

响应语法

```
{
   "CreationTime": number,
   "LocationArn": "string",
   "LocationUri": "string",
```

```
   "Protocol": {
      "NFS": {
         "MountOptions": {
            "Version": "string"
         }
      },
      "SMB": {
         "CmkSecretConfig": {
            "KmsKeyArn": "string",
            "SecretArn": "string"
         },
         "CustomSecretConfig": {
            "SecretAccessRoleArn": "string",
            "SecretArn": "string"
         },
         "Domain": "string",
         "ManagedSecretConfig": {
            "SecretArn": "string"
         },
         "MountOptions": {
            "Version": "string"
         },
         "Password": "string",
         "User": "string"
      }
   },
   "SecurityGroupArns": [ "string" ]
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

CreationTime

FSx 为 OpenZFS 创建位置的时间。

类型：时间戳

LocationArn

所描述的 FSx 适用于 OpenZFS 的位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

所描述的 OpenZFS 位置 FSx 的统一资源标识符 (URI)。

示例：`fsxz://us-west-2.fs-1234567890abcdef02/fsx/folderA/folder`

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

Protocol

AWS DataSync 用于访问您的文件系统的协议类型。

类型：FsxProtocol 对象

SecurityGroupArns

为 ARNs OpenZFS 文件系统配置的安全组中的 FSx 一个。

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationFsxWindows

提供有关如何配置 Amazon for Window FSx s 文件服务器文件系统的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定 FSx 适用于 Windows 文件服务器的位置的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CreationTime": number,
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
```

```
    },
    "Domain": "string",
    "LocationArn": "string",
    "LocationUri": "string",
    "ManagedSecretConfig": {
       "SecretArn": "string"
    },
    "SecurityGroupArns": [ "string" ],
    "User": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

CmkSecretConfig

描述由客户 DataSync AWS KMS key管理的密钥（例如 DataSync 用于访问特定存储位置的密钥）的配置信息。Password

类型：CmkSecretConfig 对象

CreationTime

Windows 文件服务器位置 FSx 的创建时间。

类型：时间戳

CustomSecretConfig

描述客户管理的密钥的配置信息，例如用于访问特定存储位置的密钥，以及提供对密钥的访问权限的客户托管 AWS Identity and Access Management (IAM) 角色。Password DataSync

类型：CustomSecretConfig 对象

Domain

FSx 适用于 Windows 文件服务器的文件系统所属的 Microsoft Active Directory 域的名称。

类型：字符串

长度约束：最大长度为 253。

模式：^[A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252}$

LocationArn

FSx 适用于 Windows 文件服务器的位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$

LocationUri

FSx 适用于 Windows 文件服务器位置的统一资源标识符 (URI)。

类型：字符串

长度约束：最大长度为 4360。

模式：^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$

ManagedSecretConfig

描述 DataSync托管密钥的配置信息，例如 DataSync 用于访问特定存储位置的。Password DataSync 使用默认 AWS托管的 KMS 密钥在中 AWS Secrets Manager加密此密钥。

类型：ManagedSecretConfig 对象

SecurityGroupArns

提供对文件系统首选子网访问权限的 Amazon EC2 安全组的 ARN。

有关为文件系统访问配置安全组的信息，请参阅《Amazon f FSx or Windows 文件服务器用户指南》。

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$

User

有权装载和访问 Window FSx s 文件服务器文件系统的用户。

类型：字符串

长度约束：最大长度为 104。

模式：^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包

- AWS 适用于 Ruby V3 的 SDK

DescribeLocationHdfs

提供有关如何配置 Hadoop 分布式文件系统 (HDFS) AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定 HDFS 位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "BlockSize": number,
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CreationTime": number,
```

```
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "KerberosPrincipal": "string",
   "KmsKeyProviderUri": "string",
   "LocationArn": "string",
   "LocationUri": "string",
   "ManagedSecretConfig": {
      "SecretArn": "string"
   },
   "NameNodes": [
      {
         "Hostname": "string",
         "Port": number
      }
   ],
   "QopConfiguration": {
      "DataTransferProtection": "string",
      "RpcProtection": "string"
   },
   "ReplicationFactor": number,
   "SimpleUser": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AgentArns

可以连接您 ARNs 的 HDFS 集群的 DataSync 代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

AuthenticationType

用于确定用户身份的身份验证的类型。

类型：字符串

有效值：SIMPLE | KERBEROS

BlockSize

要写入 HDFS 集群的数据块大小。

类型：整数

有效范围：最小值为 1048576。最大值为 1073741824。

CmkSecretConfig

描述由客户 DataSync AWS KMS key管理的密钥（例如 DataSync 用于访问特定存储位置的密钥）的配置信息。KerberosKeytab

类型：CmkSecretConfig 对象

CreationTime

创建 HDFS 位置的时间。

类型：时间戳

CustomSecretConfig

描述客户管理的密钥的配置信息，例如用于访问特定存储位置的密钥，以及提供对密钥的访问权限的客户托管 AWS Identity and Access Management (IAM) 角色。KerberosKeytab DataSync

类型：CustomSecretConfig 对象

KerberosPrincipal

有权访问 HDFS 集群上的文件和文件夹的 Kerberos 主体。如果 AuthenticationType 定义为 KERBEROS，则使用此参数。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：^.+$

KmsKeyProviderUri

HDFS 集群的密钥管理服务器（KMS）的 URI。

类型：字符串

长度限制：最小长度为 1。最大长度为 255。

模式：`^kms:\/\/http[s]?@(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-]*[A-Za-z0-9])(;(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-]*[A-Za-z0-9]))*:[0-9]{1,5}\/kms$`

LocationArn

HDFS 位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

HDFS 位置的 URI。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

ManagedSecretConfig

描述 DataSync托管密钥的配置信息，例如 DataSync 用于访问特定存储位置的。`KerberosKeytab` DataSync 使用默认 AWS托管的 KMS 密钥在中 AWS Secrets Manager加密此密钥。

类型：ManagedSecretConfig 对象

NameNodes

管理 HDFS 命名空间的。 NameNode

类型：HdfsNameNode 对象数组

数组成员：最少 1 个物品。

QopConfiguration

保护质量（QOP）配置，用于指定在 HDFS 集群上配置的远程程序调用（RPC）和数据传输保护设置。

类型：QopConfiguration 对象

ReplicationFactor

写入 HDFS 集群时要将数据复制到的数量。 DataNodes

类型：整数

有效范围：最小值为 1。最大值为 512。

SimpleUser

用于在主机操作系统上标识客户端的用户名。如果 AuthenticationType 定义为 SIMPLE，则使用此参数。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：^[_.A-Za-z0-9][-_.A-Za-z0-9]*$

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationNfs

提供有关如何配置网络文件系统 (NFS) 文件服务器的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定您要获取信息的 NFS 位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "CreationTime": number,
   "LocationArn": "string",
   "LocationUri": "string",
   "MountOptions": {
      "Version": "string"
   },
   "OnPremConfig": {
      "AgentArns": [ "string" ]
   }
```

```
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

CreationTime

NFS 位置的创建时间。

类型：时间戳

LocationArn

NFS 位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

NFS 位置的 URI。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

MountOptions

DataSync 用于装载 NFS 文件服务器的装载选项。

类型：NfsMountOptions 对象

OnPremConfig

可以连接到您的网络文件系统 (NFS) 文件服务器的 AWS DataSync 代理。

类型：OnPremConfig 对象

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例

以下示例返回示例请求中指定的 NFS 位置的信息。

示例请求

```
{
  "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-07db7abfc326c50aa"
}
```

示例

此示例说明了的一种用法 DescribeLocationNfs。

示例响应

```
{
   "CreationTime": 1532660733.39,
   "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-07db7abfc326c50aa",
   "LocationUri": "hostname.amazon.com",
   "OnPremConfig": {
      "AgentArns": [ "arn:aws:datasync:us-east-2:111222333444:agent/
agent-0b0addbeef44b3nfs" ]
```

```
    }
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationObjectStorage

提供有关如何配置对象存储系统的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定对象存储系统位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "AccessKey": "string",
   "AgentArns": [ "string" ],
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CreationTime": number,
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
```

```
      "SecretArn": "string"
   },
   "LocationArn": "string",
   "LocationUri": "string",
   "ManagedSecretConfig": {
      "SecretArn": "string"
   },
   "ServerCertificate": blob,
   "ServerPort": number,
   "ServerProtocol": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AccessKey

向对象存储系统进行身份验证所需的访问密钥（例如，用户名）。

类型：字符串

长度限制：最小长度为 0。最大长度为 200。

模式：`^.*$`

AgentArns

可以与您的对象存储系统连接的 DataSync 代理。 ARNs

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

CmkSecretConfig

描述 DataSync托管密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组凭证，DataSync 以及客户 AWS KMS key管理的密钥。

类型：CmkSecretConfig 对象

CreationTime

创建位置的时间。

类型：时间戳

CustomSecretConfig

描述客户管理的密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组证书，以及提供密钥访问权限的客户管理 AWS Identity and Access Management (IAM) 角色。 DataSync

类型：CustomSecretConfig 对象

LocationArn

对象存储系统位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

对象存储系统位置的 URI。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

ManagedSecretConfig

描述 DataSync托管密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组凭证。DataSync DataSync 使用默认 AWS托管的 KMS 密钥在中 AWS Secrets Manager加密此密钥。

类型：ManagedSecretConfig 对象

ServerCertificate

如果您的对象存储系统使用私有或自签名证书颁发机构 (CA)，则用于 DataSync 向该系统进行身份验证的证书链。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 32768。

ServerPort

对象存储服务器接受入站网络流量的端口（例如，端口 443）。

类型：整数

有效范围：最小值为 1。最大值为 65536。

ServerProtocol

对象存储系统通信时使用的协议。

类型：字符串

有效值：`HTTPS | HTTP`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK

- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeLocationS3

提供有关如何配置 S3 存储桶的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定 Amazon S3 位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "AgentArns": [ "string" ],
   "CreationTime": number,
   "LocationArn": "string",
   "LocationUri": "string",
   "S3Config": {
      "BucketAccessRoleArn": "string"
   },
   "S3StorageClass": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AgentArns

在 ARNs Outposts 上使用与 Amazon S3 配合使用时，部署在你的 Outpost 上的 DataSync 代理。

有关更多信息，请参阅在上部署 DataSync 代理 AWS Outposts。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

CreationTime

创建 Amazon S3 位置的时间。

类型：时间戳

LocationArn

Amazon S3 位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

所描述的 Amazon S3 位置的 URL。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

S3Config

指定用于访问您的 S3 存储桶的 AWS Identity and Access Management (IAM) 角色的 DataSync 亚马逊资源名称 (ARN)。

有关更多信息，请参阅提供 DataSync 对 S3 存储桶的访问权限。

类型：S3Config 对象

S3StorageClass

当 Amazon S3 作为目标位置时，您应该为对象选择此存储类别。

有些存储类别的行为会影响您的 Amazon S3 存储成本。有关更多信息，请参阅 Amazon S3 传输的存储类别注意事项。

类型：字符串

有效值：`STANDARD | STANDARD_IA | ONEZONE_IA | INTELLIGENT_TIERING | GLACIER | DEEP_ARCHIVE | OUTPOSTS | GLACIER_INSTANT_RETRIEVAL`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例

以下示例会返回示例请求中指定的 Amazon S3 位置信息。

示例请求

```
{
  "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-07db7abfc326c50s3"
}
```

示例

此示例说明了 DescribeLocation S3 的一种用法。

示例响应

```
{
   "CreationTime": 1532660733.39,
   "LocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-07db7abfc326c50s3",
   "LocationUri": "s3://amzn-s3-demo-bucket",
   "S3Config": {
      "BucketAccessRoleArn": "arn:aws:iam::111222333444:role/amzn-s3-demo-bucket-
access-role",
   }
    "S3StorageClass": "STANDARD"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包

- AWS 适用于 Ruby V3 的 SDK

DescribeLocationSmb

提供有关如何配置服务器消息块 (SMB) 文件服务器的 AWS DataSync 传输位置的详细信息。

请求语法

```
{
   "LocationArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定您想要了解的 SMB 位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：是

响应语法

```
{
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CreationTime": number,
   "CustomSecretConfig": {
```

```
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "DnsIpAddresses": [ "string" ],
   "Domain": "string",
   "KerberosPrincipal": "string",
   "LocationArn": "string",
   "LocationUri": "string",
   "ManagedSecretConfig": {
      "SecretArn": "string"
   },
   "MountOptions": {
      "Version": "string"
   },
   "User": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

AgentArns

可以与您的 SMB 文件服务器连接的 DataSync 代理。 ARNs

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

AuthenticationType

DataSync 用于连接您的 SMB 文件服务器的身份验证协议。

类型：字符串

有效值：`NTLM | KERBEROS`

CmkSecretConfig

描述由客户 DataSync AWS KMS key管理的密钥（例如 DataSync 用于访问特定存储位置的Password或KerberosKeytab）的配置信息。

类型：CmkSecretConfig 对象

CreationTime

创建 SMB 位置的时间。

类型：时间戳

CustomSecretConfig

描述客户管理的密钥的配置信息，例如用于访问特定存储位置的Password或KerberosKeytab，并 DataSync 使用客户托管 AWS Identity and Access Management (IAM) 角色提供对密钥的访问权限。

类型：CustomSecretConfig 对象

DnsIpAddresses

您的 SMB 文件服务器所属的 DNS 服务器的 IPv4 或 IPv6 地址。此元素仅在 AuthenticationType 设置为 KERBEROS 时适用。

类型：字符串数组

数组成员：最多 2 项。

长度限制：最小长度为 7。最大长度为 39。

模式：\A((25[0-5]|2[0-4]\d|[0-1]?\d?\d)(\.(25[0-5]|2[0-4]\d|[0-1]?\d?\d)){3}|([0-9a-fA-F]{1,4}:){7,7}[0-9a-fA-F]{1,4}|([0-9a-fA-F]{1,4}:){1,7}:|([0-9a-fA-F]{1,4}:){1,6}:[0-9a-fA-F]{1,4}|([0-9a-fA-F]{1,4}:){1,5}(:[0-9a-fA-F]{1,4}){1,2}|([0-9a-fA-F]{1,4}:){1,4}(:[0-9a-fA-F]{1,4}){1,3}|([0-9a-fA-F]{1,4}:){1,3}(:[0-9a-fA-F]{1,4}){1,4}|([0-9a-fA-F]{1,4}:){1,2}(:[0-9a-fA-F]{1,4}){1,5}|[0-9a-fA-F]{1,4}:((:[0-9a-fA-F]{1,4}){1,6}))\z

Domain

SMB 文件服务器所属的 Windows 域的名称。此元素仅在 AuthenticationType 设置为 NTLM 时适用。

类型：字符串

长度约束：最大长度为 253。

模式：`^[A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252}$`

KerberosPrincipal

有权访问 SMB 文件服务器中文件、文件夹和文件元数据的 Kerberos 主体。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：`^.+$`

LocationArn

SMB 位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

LocationUri

SMB 位置的 URI。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

ManagedSecretConfig

描述 DataSync托管密钥的配置信息，例如 DataSync 用于访问特定存储位置的`Password`或`KerberosKeytab`。 DataSync 使用默认 AWS托管的 KMS 密钥在中 AWS Secrets Manager加密此密钥。

类型：ManagedSecretConfig 对象

MountOptions

DataSync 用于访问您的 SMB 文件服务器的 SMB 协议版本。

类型：SmbMountOptions 对象

User

可以挂载和访问 SMB 文件服务器中文件、文件夹和文件元数据的用户。此元素仅在 AuthenticationType 设置为 NTLM 时适用。

类型：字符串

长度约束：最大长度为 104。

模式：^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例

此示例说明了的一种用法 DescribeLocationSmb。

示例请求

```
{
```

```
  "arn:aws:datasync:us-east-1:111222333444:location/loc-0f01451b140b2af49"
}
```

示例

此示例说明了的一种用法 DescribeLocationSmb。

示例响应

```
{
   "AgentArns":[
      "arn:aws:datasync:us-east-2:111222333444:agent/agent-0bc3b3dc9bbc15145",
      "arn:aws:datasync:us-east-2:111222333444:agent/agent-04b3fe3d261a18c8f"
   ],
   "CreationTime":"1532660733.39",
   "Domain":"AMAZON",
   "LocationArn":"arn:aws:datasync:us-east-1:111222333444:location/
loc-0f01451b140b2af49",
   "LocationUri":"smb://hostname.amazon.com/share",
   "MountOptions":{
      "Version":"SMB3"
   },
   "User":"user-1"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeTask

提供任务的相关信息。任务定义了 AWS DataSync 传输数据的位置和方式。

请求语法

```
{
   "TaskArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

TaskArn

指定您想要了解的传输任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$`

必需：是

响应语法

```
{
   "CloudWatchLogGroupArn": "string",
   "CreationTime": number,
   "CurrentTaskExecutionArn": "string",
   "DestinationLocationArn": "string",
   "DestinationNetworkInterfaceArns": [ "string" ],
   "ErrorCode": "string",
   "ErrorDetail": "string",
   "Excludes": [
      {
         "FilterType": "string",
         "Value": "string"
```

```
      }
   ],
   "Includes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "ManifestConfig": {
      "Action": "string",
      "Format": "string",
      "Source": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "ManifestObjectPath": "string",
            "ManifestObjectVersionId": "string",
            "S3BucketArn": "string"
         }
      }
   },
   "Name": "string",
   "Options": {
      "Atime": "string",
      "BytesPerSecond": number,
      "Gid": "string",
      "LogLevel": "string",
      "Mtime": "string",
      "ObjectTags": "string",
      "OverwriteMode": "string",
      "PosixPermissions": "string",
      "PreserveDeletedFiles": "string",
      "PreserveDevices": "string",
      "SecurityDescriptorCopyFlags": "string",
      "TaskQueueing": "string",
      "TransferMode": "string",
      "Uid": "string",
      "VerifyMode": "string"
   },
   "Schedule": {
      "ScheduleExpression": "string",
      "Status": "string"
   },
   "ScheduleDetails": {
      "DisabledBy": "string",
```

```
      "DisabledReason": "string",
      "StatusUpdateTime": number
   },
   "SourceLocationArn": "string",
   "SourceNetworkInterfaceArns": [ "string" ],
   "Status": "string",
   "TaskArn": "string",
   "TaskMode": "string",
   "TaskReportConfig": {
      "Destination": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "S3BucketArn": "string",
            "Subdirectory": "string"
         }
      },
      "ObjectVersionIds": "string",
      "OutputType": "string",
      "Overrides": {
         "Deleted": {
            "ReportLevel": "string"
         },
         "Skipped": {
            "ReportLevel": "string"
         },
         "Transferred": {
            "ReportLevel": "string"
         },
         "Verified": {
            "ReportLevel": "string"
         }
      },
      "ReportLevel": "string"
   }
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

CloudWatchLogGroupArn

用于监控您的任务的亚马逊 CloudWatch 日志组的亚马逊资源名称 (ARN)。

有关更多信息，请参阅使用 CloudWatch 日志监控数据传输。

类型：字符串

长度约束：最大长度为 562。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):logs:[a-z\-0-9]+:[0-9]{12}:log-group:([^:\*]*)(:\*)?$`

CreationTime

任务的创建时间。

类型：时间戳

CurrentTaskExecutionArn

最近一次执行的任务的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}/execution/exec-[0-9a-f]{17}$`

DestinationLocationArn

传输目标位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

DestinationNetworkInterfaceArns

为您的目标位置 DataSync 创建的网络接口。 ARNs

类型：字符串数组

长度限制：最大长度为 128。

模式：`^arn:aws[\-a-z]{0,}:ec2:[a-z\-0-9]*:[0-9]{12}:network-interface/eni-[0-9a-f]+$`

ErrorCode

如果任务出现问题，您可以使用错误代码来帮助排除故障。有关更多信息，请参阅 DataSync 转移问题疑难解答。

类型：字符串

ErrorDetail

如果任务出现问题，您可以使用错误详细信息来帮助排除故障。有关更多信息，请参阅 DataSync 转移问题疑难解答。

类型：字符串

Excludes

排除过滤器，用于定义源位置中您不想 DataSync 传输的文件、对象和文件夹。有关更多信息和示例，请参阅使用过滤器指定 DataSync 传输的内容。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

Includes

包括过滤器，用于定义源位置中 DataSync 要传输的文件、对象和文件夹。有关更多信息和示例，请参阅使用过滤器指定 DataSync 传输的内容。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

ManifestConfig

清单的配置，其中列出了您 DataSync 要传输的文件或对象。有关更多信息，请参阅使用清单指定 DataSync 传输的内容。

类型：ManifestConfig 对象

Name

任务的名称。

类型：字符串

长度限制：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

Options

任务的设置。例如，保留哪些文件元数据、如何在传输结束时验证数据完整性，以及带宽限制等选项。

类型：Options 对象

Schedule

任务运行的时间计划。有关更多信息，请参阅计划任务。

类型：TaskSchedule 对象

ScheduleDetails

有关任务计划的详细信息。

类型：TaskScheduleDetails 对象

SourceLocationArn

传输源位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

SourceNetworkInterfaceArns

为您的源位置 DataSync 创建的网络接口。 ARNs

类型：字符串数组

长度限制：最大长度为 128。

模式：^arn:aws[\-a-z]{0,}:ec2:[a-z\-0-9]*:[0-9]{12}:network-interface/eni-[0-9a-f]+$

Status

任务的状态。有关每种状态含义的信息，请参阅任务状态。

类型：字符串

有效值：AVAILABLE | CREATING | QUEUED | RUNNING | UNAVAILABLE

TaskArn

任务的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$

TaskMode

当前使用的任务模式。有关更多信息，请参阅为数据传输选择任务模式。

类型：字符串

有效值：BASIC | ENHANCED

TaskReportConfig

任务报告的配置，其中提供有关您的 DataSync 转移的详细信息。有关更多信息，请参阅使用任务报告监控您的 DataSync转账。

类型：TaskReportConfig 对象

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

以下示例指定要获取相关信息的 DataSync 任务的 ARN。

```
{
  "TaskArn": "arn:aws:datasync:us-east-2:111222333444:task/task-08de6e6697796f026"
}
```

示例响应

下面的示例介绍一个 DescribeTask 响应。

```
{
  "TaskArn": "arn:aws:datasync:us-east-2:111222333444:task/task-08de6e6697796f026",
  "Name": "MyTask",
  "TaskMode": "BASIC",
  "Status": "RUNNING",
  "SourceLocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-1111aaaa2222bbbb3",
  "DestinationLocationArn": "arn:aws:datasync:us-east-2:111222333444:location/
loc-0000zzzz1111yyyy2",
  "CurrentTaskExecutionArn": "arn:aws:datasync:us-east-2:111222333444:task/
task-08de6e6697796f026/execution/exec-04ce9d516d69bd52f",
  "CreationTime": 1532660733.39,
  "Options": {
    "Atime": "BEST_EFFORT",
    "BytesPerSecond": 1000,
    "Gid": "NONE",
    "Mtime": "PRESERVE",
    "PosixPermissions": "PRESERVE",
    "PreserveDevices": "NONE",
    "PreserveDeletedFiles": "PRESERVE",
    "Uid": "NONE",
```

```
    "VerifyMode": "POINT_IN_TIME_CONSISTENT"
  },
  "CloudWatchLogGroupArn": "arn:aws:logs:us-east-2:111222333444:log-group:/log-group-
name:*"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

DescribeTaskExecution

提供有关 AWS DataSync 任务执行的信息。您可以使用此操作来帮助监控正在进行的数据传输进度或检查传输结果。

> **Note**
>
> 某些 DescribeTaskExecution 响应元素仅与特定任务模式相关。有关信息，请参阅了解任务模式差异和了解数据传输性能计数器。

请求语法

```
{
   "TaskExecutionArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

TaskExecutionArn

指定您想要获取信息的任务执行的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}/execution/exec-[0-9a-f]{17}$

必需：是

响应语法

```
{
   "BytesCompressed": number,
```

```
   "BytesTransferred": number,
   "BytesWritten": number,
   "EndTime": number,
   "EstimatedBytesToTransfer": number,
   "EstimatedFilesToDelete": number,
   "EstimatedFilesToTransfer": number,
   "EstimatedFoldersToDelete": number,
   "EstimatedFoldersToTransfer": number,
   "Excludes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "FilesDeleted": number,
   "FilesFailed": {
      "Delete": number,
      "Prepare": number,
      "Transfer": number,
      "Verify": number
   },
   "FilesListed": {
      "AtDestinationForDelete": number,
      "AtSource": number
   },
   "FilesPrepared": number,
   "FilesSkipped": number,
   "FilesTransferred": number,
   "FilesVerified": number,
   "FoldersDeleted": number,
   "FoldersFailed": {
      "Delete": number,
      "List": number,
      "Prepare": number,
      "Transfer": number,
      "Verify": number
   },
   "FoldersListed": {
      "AtDestinationForDelete": number,
      "AtSource": number
   },
   "FoldersPrepared": number,
   "FoldersSkipped": number,
   "FoldersTransferred": number,
```

```
   "FoldersVerified": number,
   "Includes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "LaunchTime": number,
   "ManifestConfig": {
      "Action": "string",
      "Format": "string",
      "Source": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "ManifestObjectPath": "string",
            "ManifestObjectVersionId": "string",
            "S3BucketArn": "string"
         }
      }
   },
   "Options": {
      "Atime": "string",
      "BytesPerSecond": number,
      "Gid": "string",
      "LogLevel": "string",
      "Mtime": "string",
      "ObjectTags": "string",
      "OverwriteMode": "string",
      "PosixPermissions": "string",
      "PreserveDeletedFiles": "string",
      "PreserveDevices": "string",
      "SecurityDescriptorCopyFlags": "string",
      "TaskQueueing": "string",
      "TransferMode": "string",
      "Uid": "string",
      "VerifyMode": "string"
   },
   "ReportResult": {
      "ErrorCode": "string",
      "ErrorDetail": "string",
      "Status": "string"
   },
   "Result": {
      "ErrorCode": "string",
```

```
      "ErrorDetail": "string",
      "PrepareDuration": number,
      "PrepareStatus": "string",
      "TotalDuration": number,
      "TransferDuration": number,
      "TransferStatus": "string",
      "VerifyDuration": number,
      "VerifyStatus": "string"
   },
   "StartTime": number,
   "Status": "string",
   "TaskExecutionArn": "string",
   "TaskMode": "string",
   "TaskReportConfig": {
      "Destination": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "S3BucketArn": "string",
            "Subdirectory": "string"
         }
      },
      "ObjectVersionIds": "string",
      "OutputType": "string",
      "Overrides": {
         "Deleted": {
            "ReportLevel": "string"
         },
         "Skipped": {
            "ReportLevel": "string"
         },
         "Transferred": {
            "ReportLevel": "string"
         },
         "Verified": {
            "ReportLevel": "string"
         }
      },
      "ReportLevel": "string"
   }
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

BytesCompressed

压缩后通过网络 DataSync 传输的物理字节数（如果可以压缩）。除非数据不可压缩，否则该数字通常小于 BytesTransferred。

类型：长整型

BytesTransferred

压缩前 DataSync 发送到网络的字节数（如果可以压缩）。有关通过网络传输的字节数，请参阅 BytesCompressed。

类型：长整型

BytesWritten

DataSync 实际写入目标位置的逻辑字节数。

类型：长整型

EndTime

传输任务结束的时间。

类型：时间戳

EstimatedBytesToTransfer

DataSync 预期写入目标位置的逻辑字节数。

类型：长整型

EstimatedFilesToDelete

DataSync 预计在目标位置删除的文件、对象和目录的数量。如果您未将任务配置为在目标位置中删除不存在于源位置的数据，则该值始终为 0。

> **Note**
>
> 对于增强模式任务，此计数器仅包含文件或对象。目录已计入EstimatedFoldersToDelete。

类型：长整型

EstimatedFilesToTransfer

DataSync 预计通过网络传输的文件、对象和目录的数量。此值是在 DataSync准备传输时计算的。

此值的计算方式主要取决于任务的传输模式配置：

- 如果 TranserMode 设置为 CHANGED：计算方法是比较源位置和目标位置的内容，确定需要传输的差异。差异可能包括：
  - 在源位置添加或修改的任何内容。
  - 在两个位置均存在且在初始传输后在目的地修改的任何内容（除非设置OverwriteMode为NEVER）。
  - （仅限基本任务模式）DataSync 预计要删除的项目数（如果设置PreserveDeletedFiles为REMOVE）。
- 如果设置TranserMode为 ALL-计算仅基于在来源位置 DataSync 找到的项目。

> **Note**
>
> 对于增强模式任务，此计数器仅包含文件或对象。目录已计入EstimatedFoldersToTransfer。

类型：长整型

EstimatedFoldersToDelete

目标位置中 DataSync 预计要删除的目录数量。如果您未将任务配置为在目标位置中删除不存在于源位置的数据，则该值始终为 0。

> **Note**
>
> 仅适用于增强模式任务。

类型：长整型

EstimatedFoldersToTransfer

DataSync 预计通过网络传输的目录数量。此值是在 DataSync准备要传输的目录时计算得出的。

此值的计算方式主要取决于任务的传输模式配置：

- 如果 TranserMode 设置为 CHANGED：计算方法是比较源位置和目标位置的内容，确定需要传输的差异。差异可能包括：
  - 在源位置添加或修改的任何内容。
  - 在两个位置均存在且在初始传输后在目的地修改的任何内容（除非设置OverwriteMode为NEVER）。
- 如果设置TranserMode为 ALL-计算仅基于在来源位置 DataSync 找到的项目。

> **Note**
>
> 仅适用于增强模式任务。

类型：长整型

Excludes

在传输过程中排除特定数据的筛选规则列表。有关更多信息和示例，请参阅筛选传输的数据 DataSync。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

FilesDeleted

在目标位置 DataSync 实际删除的文件、对象和目录的数量。如果您未将任务配置为在目标位置中删除不存在于源位置的数据，则该值始终为 0。

> **Note**
>
> 对于增强模式任务，此计数器仅包含文件或对象。目录已计入FoldersDeleted。

类型：长整型

FilesFailed

在任务执行期间 DataSync 无法准备、传输、验证和删除的文件或对象的数量。

> **Note**
>
> 仅适用于增强模式任务。

类型：TaskExecutionFilesFailedDetail 对象

FilesListed

在您的位置 DataSync 找到的文件或对象的数量。

> **Note**
>
> 仅适用于增强模式任务。

类型：TaskExecutionFilesListedDetail 对象

FilesPrepared

比较源位置和目标位置后 DataSync 将尝试传输的文件或对象的数量。

> **Note**
>
> 仅适用于增强模式任务。

如果您将任务配置为传输所有数据，则此计数器不适用。在这种情况下， DataSync 将所有内容从源复制到目的地，而不比较位置之间的差异。

类型：长整型

FilesSkipped

传输过程中 DataSync 跳过的文件、对象和目录的数量。

> **Note**
>
> 对于增强模式任务，此计数器仅包含文件或对象。目录已计入FoldersSkipped。

类型：长整型

FilesTransferred

DataSync 实际通过网络传输的文件、对象和目录的数量。在任务执行期间，当从源读取内容并通过网络发送内容时，该值会定期更新。

如果 DataSync 无法传输某些内容，则此值可以小于EstimatedFilesToTransfer。在某些情况下，此值也可能大于 EstimatedFilesToTransfer。此元素是特定于某些位置类型的实施，因此请勿将其用作传输内容的确切指示或监控任务执行。

> **Note**
>
> 对于增强模式任务，此计数器仅包含文件或对象。目录已计入FoldersTransferred。

类型：长整型

FilesVerified

传输过程中要 DataSync 验证的文件、对象和目录的数量。

> **Note**
>
> 当您将任务配置为仅验证传输的数据时，在某些情况下 DataSync 不验证目录或无法传输的文件。
> 对于增强模式任务，此计数器仅包含文件或对象。目录已计入FoldersVerified。

类型：长整型

FoldersDeleted

目标位置中 DataSync 实际删除的目录数量。如果您未将任务配置为在目标位置中删除不存在于源位置的数据，则该值始终为 0。

> **Note**
>
> 仅适用于增强模式任务。

类型：长整型

FoldersFailed

在任务执行期间 DataSync 未能列出、准备、传输、验证和删除的目录数量。

> **Note**
>
> 仅适用于增强模式任务。

类型：TaskExecutionFoldersFailedDetail 对象

FoldersListed

在您的位置 DataSync 找到的目录数量。

> **Note**
>
> 仅适用于增强模式任务。

类型：TaskExecutionFoldersListedDetail 对象

FoldersPrepared

比较源位置和目标位置后 DataSync 将尝试传输的目录数量。

> **Note**
>
> 仅适用于增强模式任务。

如果您将任务配置为传输所有数据，则此计数器不适用。在这种情况下， DataSync 将所有内容从源复制到目的地，而不比较位置之间的差异。

类型：长整型

FoldersSkipped

传输过程中 DataSync 跳过的目录数量。

> **Note**
>
> 仅适用于增强模式任务。

类型：长整型

FoldersTransferred

DataSync 实际通过网络传输的目录数量。在任务执行期间，当从源读取内容并通过网络发送内容时，该值会定期更新。

如果 DataSync 无法传输某些内容，则此值可以小于`EstimatedFoldersToTransfer`。在某些情况下，此值也可能大于 `EstimatedFoldersToTransfer`。

> **Note**
>
> 仅适用于增强模式任务。

类型：长整型

FoldersVerified

在传输过程中进行 DataSync 验证的目录数量。

> **Note**
>
> 仅适用于增强模式任务。

类型：长整型

Includes

在传输期间包含特定数据的筛选规则列表。有关更多信息和示例，请参阅筛选传输的数据 DataSync。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

LaunchTime

任务执行实际开始的时间。对于非排队任务，`LaunchTime` 和 `StartTime` 通常相同。对于排队任务，`LaunchTime` 通常晚于 `StartTime`，因为先前排队的任务必须完成运行后，才能开始新的任务。

类型：时间戳

ManifestConfig

列出了要传输的文件或对象的清单配置。有关更多信息，请参阅使用清单指定 DataSync 传输的内容。

类型：ManifestConfig 对象

Options

指明您的传输任务是如何配置的。这些选项包括在传输过程中如何 DataSync 处理文件、对象及其关联的元数据。您还可以指定如何验证数据完整性、为任务设置带宽限制等其他选项。

每个选项都有一个默认值。除非需要，否则无需在拨打电话之前配置任何选项StartTaskExecution。

您还可以在每次执行任务时覆盖任务选项。例如，您可能希望为单次执行调整 LogLevel。

类型：Options 对象

ReportResult

指明是否为您的转移 DataSync 生成了完整的任务报告。

类型：ReportResult 对象

Result

任务执行的结果。

类型：TaskExecutionResultDetail 对象

StartTime

DataSync 发送请求以开始执行任务的时间。对于非排队任务，LaunchTime 和 StartTime 通常相同。对于排队任务，LaunchTime 通常晚于 StartTime，因为先前排队的任务必须完成运行后，才能开始新的任务。

类型：时间戳

Status

任务执行的状态。

有关任务执行状态的详细信息，请参阅任务执行状态。

类型：字符串

有效值：QUEUED | CANCELLING | LAUNCHING | PREPARING | TRANSFERRING | VERIFYING | SUCCESS | ERROR

TaskExecutionArn

您想要了解相关信息的任务执行的 ARN。TaskExecutionArn 是分层的，TaskArn 包括已执行的任务。

例如，带有 ARN arn:aws:datasync:us-east-1:111222333444:task/task-0208075f79cedf4a2/execution/exec-08ef1e88ec491019b 的 TaskExecution 值使用 ARN arn:aws:datasync:us-east-1:111222333444:task/task-0208075f79cedf4a2 执行任务。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}/execution/exec-[0-9a-f]{17}$

TaskMode

当前使用的任务模式。有关更多信息，请参阅为数据传输选择任务模式。

类型：字符串

有效值：BASIC | ENHANCED

TaskReportConfig

任务报告的配置，其中提供有关 DataSync 转移的详细信息。有关更多信息，请参阅创建任务报告。

类型：TaskReportConfig 对象

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

此示例说明了一个 DescribeTaskExecution 请求。

```
{
    "TaskExecutionArn": "arn:aws:datasync:us-east-1:111222333444:task/task-
aaaabbbbccccddddf/execution/exec-1234abcd1234abcd1"
}
```

执行增强模式任务的示例响应 1

以下示例描述了使用增强模式的 DataSync 任务执行。该任务执行还通过使用排除和包含筛选条件来传输特定数据。

```
{
    "TaskExecutionArn": "arn:aws:datasync:us-east-1:111222333444:task/task-
aaaabbbbccccddddf/execution/exec-1234abcd1234abcd1",
    "Status": "SUCCESS",
    "Options": {
        "VerifyMode": "ONLY_FILES_TRANSFERRED",
        "OverwriteMode": "ALWAYS",
        "Atime": "BEST_EFFORT",
        "Mtime": "PRESERVE",
        "Uid": "NONE",
        "Gid": "NONE",
        "PreserveDeletedFiles": "PRESERVE",
        "PreserveDevices": "NONE",
        "PosixPermissions": "NONE",
        "BytesPerSecond": -1,
        "TaskQueueing": "ENABLED",
        "LogLevel": "BASIC",
        "TransferMode": "CHANGED",
        "SecurityDescriptorCopyFlags": "NONE",
        "ObjectTags": "PRESERVE"
    },
```

```
    "Excludes": [{
        "FilterType": "SIMPLE_PATTERN",
        "Value": "/archive-files"
    }],
    "Includes": [{
        "FilterType": "SIMPLE_PATTERN",
        "Value": "/files"
    }],
    "StartTime": "2024-10-16T11:19:56.844000-04:00",
    "EstimatedFilesToTransfer": 7,
    "EstimatedFoldersToTransfer": 2,
    "EstimatedBytesToTransfer": 30,
    "FilesTransferred": 7,
    "FoldersTransferred": 2,
    "BytesWritten": 30,
    "BytesTransferred": 30,
    "BytesCompressed": 30,
    "Result": {
        "PrepareDuration": 0,
        "PrepareStatus": "SUCCESS",
        "TotalDuration": 3310,
        "TransferDuration": 0,
        "TransferStatus": "SUCCESS",
        "VerifyDuration": 0,
        "VerifyStatus": "SUCCESS"
    },
    "FilesDeleted": 0,
    "FilesSkipped": 0,
    "FilesVerified": 7,
    "EstimatedFilesToDelete": 0,
    "TaskMode": "ENHANCED",
    "FilesPrepared": 7,
    "FilesListed": {
        "AtSource": 7,
        "AtDestinationForDelete": 0
    },
    "FilesFailed": {
        "Prepare": 0,
        "Transfer": 0,
        "Verify": 0,
        "Delete": 0
    },
    "FoldersDeleted": 0,
    "FoldersSkipped": 0,
```

```
    "FoldersVerified": 2,
    "FoldersPrepared": 2,
    "FoldersListed": {
        "AtSource": 2,
        "AtDestinationForDelete": 0
    },
    "FoldersFailed": {
        "List": 0,
        "Prepare": 0,
        "Transfer": 0,
        "Verify": 0,
        "Delete": 0
    }
}
```

执行增强模式任务的示例响应 2

以下示例描述了另一个使用增强模式的 DataSync 任务执行。在这种情况下，该任务执行通过使用清单而不是筛选条件来传输特定数据。

```
{
    "TaskExecutionArn": "arn:aws:datasync:us-east-1:111222333444:task/task-
aaaabbbbccccddddf/execution/exec-1234abcd1234abcd1",
    "Status": "SUCCESS",
    "Options": {
        "VerifyMode": "ONLY_FILES_TRANSFERRED",
        "OverwriteMode": "ALWAYS",
        "Atime": "BEST_EFFORT",
        "Mtime": "PRESERVE",
        "Uid": "NONE",
        "Gid": "NONE",
        "PreserveDeletedFiles": "PRESERVE",
        "PreserveDevices": "NONE",
        "PosixPermissions": "NONE",
        "BytesPerSecond": -1,
        "TaskQueueing": "ENABLED",
        "LogLevel": "TRANSFER",
        "TransferMode": "CHANGED",
        "SecurityDescriptorCopyFlags": "NONE",
        "ObjectTags": "PRESERVE"
    },
    "Excludes": [],
    "Includes": [],
```

```
    "ManifestConfig": {
        "Action": "TRANSFER",
        "Format": "CSV",
        "S3AccessRoleArn": "arn:aws:iam::111222333444:role/service-role/
DataSyncS3ManifestAccess",
        "S3Bucket": "arn:aws:s3:::manifests-datasync",
        "VersionId": "Ixs7NQzEOj8BkL9r4ywX2FtDh_cPf3mG",
        "Source": {
            "S3": {
                "ManifestObjectPath": "manifest-folder/manifest-versioned-files",
                "BucketAccessRoleArn": "arn:aws:iam::111222333444:role/my-manifest-
role/DataSyncS3ManifestAccess",
                "S3BucketArn": "arn:aws:s3:::manifests-datasync",
                "ManifestObjectVersionId": "Ixs7NQzEOj8BkL9r4ywX2FtDh_cPf3mG"
            }
        }
    },
    "StartTime": "2024-10-16T09:29:56.757000-04:00",
    "EstimatedFilesToTransfer": 1,
    "EstimatedFoldersToTransfer": 0,
    "EstimatedBytesToTransfer": 6,
    "FilesTransferred": 1,
    "FoldersTransferred": 1,
    "BytesWritten": 6,
    "BytesTransferred": 6,
    "BytesCompressed": 6,
    "Result": {
        "PrepareDuration": 0,
        "PrepareStatus": "SUCCESS",
        "TotalDuration": 3089,
        "TransferDuration": 0,
        "TransferStatus": "SUCCESS",
        "VerifyDuration": 0,
        "VerifyStatus": "SUCCESS"
    },
    "TaskReportConfig": {
        "Destination": {
            "S3": {
                "Subdirectory": "reports/",
                "S3BucketArn": "arn:aws:s3:::my-task-report",
                "BucketAccessRoleArn": "arn:aws:iam::111222333444:role/my-task-report-
role/DataSyncTaskReportS3BucketAccess"
            }
        },
```

```
        "OutputType": "STANDARD",
        "ReportLevel": "SUCCESSES_AND_ERRORS",
        "ObjectVersionIds": "INCLUDE"
    },
    "FilesDeleted": 0,
    "FilesSkipped": 0,
    "FilesVerified": 1,
    "ReportResult": {
        "Status": "SUCCESS"
    },
    "EstimatedFilesToDelete": 0,
    "TaskMode": "ENHANCED",
    "FilesPrepared": 1,
    "FilesListed": {
        "AtSource": 1,
        "AtDestinationForDelete": 0
    },
    "FilesFailed": {
        "Prepare": 0,
        "Transfer": 0,
        "Verify": 0,
        "Delete": 0
    },
    "FoldersDeleted": 0,
    "FoldersSkipped": 0,
    "FoldersVerified": 0,
    "FoldersPrepared": 0,
    "FoldersListed": {
        "AtSource": 0,
        "AtDestinationForDelete": 0
    },
    "FoldersFailed": {
        "List": 0,
        "Prepare": 0,
        "Transfer": 0,
        "Verify": 0,
        "Delete": 0
    }
}
```

执行基本模式任务的示例响应

以下示例描述了使用 “基本” 模式执行的 DataSync 任务。

```
{
    "TaskExecutionArn": "arn:aws:datasync:us-east-1:111222333444:task/task-
aaaabbbbccccddddf/execution/exec-1234abcd1234abcd1",
    "BytesCompressed": 3500,
    "BytesTransferred": 5000,
    "BytesWritten": 5000,
    "EstimatedBytesToTransfer": 5000,
    "EstimatedFilesToDelete": 10,
    "EstimatedFilesToTransfer": 100,
    "FilesDeleted": 10,
    "FilesSkipped": 0,
    "FilesTransferred": 100,
    "FilesVerified": 100,
    "Result": {
        "PrepareDuration": 100,
        "PrepareStatus": "SUCCESS",
        "TransferDuration": 60,
        "TransferStatus": "SUCCESS",
        "VerifyDuration": 30,
        "VerifyStatus": "SUCCESS"
    },
    "StartTime": "2024-10-16T11:19:56.844000-04:00",
    "Status": "SUCCESS",
    "OverrideOptions": {
        "Atime": "BEST_EFFORT",
        "BytesPerSecond": "1000",
        "Gid": "NONE",
        "Mtime": "PRESERVE",
        "PosixPermissions": "PRESERVE",
        "PreserveDeletedFiles": "PRESERVE",
        "Uid": "NONE",
        "VerifyMode": "POINT_IN_TIME_CONSISTENT"
    },
    "TaskReportConfig": {
        "Destination": {
            "S3": {
                "BucketAccessRoleArn": "arn:aws:iam::111222333444:role/my-datasync-
role",
                "S3BucketArn": "arn:aws:s3:::my-task-reports-bucket/*",
                "Subdirectory": "reports"
            }
        },
        "ObjectVersionIds": "INCLUDE",
```

```
        "OutputType": "STANDARD",
        "Overrides": {
            "Deleted": {
                "ReportLevel": "ERRORS_ONLY"
            },
            "Skipped": {
                "ReportLevel": "SUCCESSES_AND_ERRORS"
            },
            "Transferred": {
                "ReportLevel": "ERRORS_ONLY"
            },
            "Verified": {
                "ReportLevel": "ERRORS_ONLY"
            }
        },
        "ReportLevel": "ERRORS_ONLY"
    }
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

ListAgents

返回属于请求中 AWS 区域 指定的 AWS 账户 中的的 AWS DataSync 代理列表。

通过分页，您可以减少响应时返回的代理数量。如果您在响应中收到截断的代理列表，则响应中包含一个标记，您可以在下一个请求中指定该标记，以获取下一页的代理。

`ListAgents` 是最终一致的。这意味着运行该操作的结果可能无法反映出您刚刚创建或删除了一个代理。例如，如果您使用创建代理CreateAgent然后立即运行`ListAgents`，则该代理可能不会立即显示在列表中。在这种情况下，您可以随时使用确认代理是否已创建（或删除）DescribeAgent。

请求语法

```
{
   "MaxResults": number,
   "NextToken": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

MaxResults

指定要在响应中列出的 DataSync 代理的最大数量。默认情况下，响应最多显示 100 个代理。

类型：整数

有效范围：最小值为 0。最大值为 100。

必需：否

NextToken

指定一个不透明的字符串，用于指示响应中下一个结果列表的起始位置。

类型：字符串

长度约束：最大长度为 65535。

模式：[a-zA-Z0-9=_-]+

必需：否

响应语法

```
{
   "Agents": [
      {
         "AgentArn": "string",
         "Name": "string",
         "Platform": {
            "Version": "string"
         },
         "Status": "string"
      }
   ],
   "NextToken": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

Agents

请求 AWS 账户 中 AWS 区域 指定的您的 DataSync 代理列表。该列表按代理商的 Amazon 资源名称 (ARNs) 排序。

类型：AgentListEntry 对象数组

NextToken

不透明字符串，用于指示响应中下一个结果列表的起始位置。

类型：字符串

长度约束：最大长度为 65535。

模式：[a-zA-Z0-9=_-]+

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

ListLocations

返回源和目标位置列表。

如果您的位置多于响应中返回的位置（也就是说，响应仅返回截断的代理列表），则响应会包含一个令牌，您可以在下次请求中指定该令牌来获取下一页位置。

请求语法

```
{
   "Filters": [
      {
         "Name": "string",
         "Operator": "string",
         "Values": [ "string" ]
      }
   ],
   "MaxResults": number,
   "NextToken": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

Filters

您可以使用 API 筛选器来缩小 ListLocations 返回的资源的列表范围。例如，要检索特定源位置上的所有任务，您可以结合筛选器名称 LocationType S3 和 Operator Equals 使用 ListLocations。

类型：LocationFilter 对象数组

必需：否

MaxResults

要返回的位置的最大数。

类型：整数

有效范围：最小值为 0。最大值为 100。

必需：否

NextToken

一个不透明的字符串，用于指示开始下一个位置列表的位置。

类型：字符串

长度约束：最大长度为 65535。

模式：[a-zA-Z0-9=_-]+

必需：否

响应语法

```
{
   "Locations": [
      {
         "LocationArn": "string",
         "LocationUri": "string"
      }
   ],
   "NextToken": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

Locations

一个包含位置列表的数组。

类型：LocationListEntry 对象数组

NextToken

一个不透明的字符串，用于指示返回下一个位置列表的位置。

类型：字符串

长度约束：最大长度为 65535。

模式：`[a-zA-Z0-9=_-]+`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版 SDK
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

ListTagsForResource

返回与 AWS 资源关联的所有标签。

请求语法

```
{
   "MaxResults": number,
   "NextToken": "string",
   "ResourceArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

MaxResults

指定您要在响应中显示多少结果。

类型：整数

有效范围：最小值为 0。最大值为 100。

必需：否

NextToken

指定一个不透明的字符串，用于指示响应中下一个结果列表的起始位置。

类型：字符串

长度约束：最大长度为 65535。

模式：[a-zA-Z0-9=_-]+

必需：否

ResourceArn

指定要对其提供标签信息的资源的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:(((agent|task|location)/(agent|task|loc)-[a-z0-9]{17}(/execution/exec-[a-f0-9]{17})?)|(system/storage-system-[a-f0-9]{8}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{12}(/job/discovery-job-[a-f0-9]{8}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{12})?))$`

必需：是

响应语法

```
{
   "NextToken": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

NextToken

不透明字符串，用于指示响应中下一个结果列表的起始位置。

类型：字符串

长度约束：最大长度为 65535。

模式：`[a-zA-Z0-9=_-]+`

Tags

应用于指定资源的一个标签数组。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 55 项。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

ListTaskExecutions

返回 AWS DataSync 传输任务的执行列表。

请求语法

```
{
   "MaxResults": number,
   "NextToken": "string",
   "TaskArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

MaxResults

指定您想要在响应中显示多少个结果。

类型：整数

有效范围：最小值为 0。最大值为 100。

必需：否

NextToken

指定一个不透明的字符串，用于指示响应中下一个结果列表的起始位置。

类型：字符串

长度约束：最大长度为 65535。

模式：[a-zA-Z0-9=_-]+

必需：否

TaskArn

指定您要获取其执行信息的任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$`

必需：否

响应语法

```
{
   "NextToken": "string",
   "TaskExecutions": [
      {
         "Status": "string",
         "TaskExecutionArn": "string",
         "TaskMode": "string"
      }
   ]
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

NextToken

不透明字符串，用于指示响应中下一个结果列表的起始位置。

类型：字符串

长度约束：最大长度为 65535。

模式：`[a-zA-Z0-9=_-]+`

TaskExecutions

任务的执行列表。

类型：TaskExecutionListEntry 对象数组

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版 SDK
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

ListTasks

返回您创建的 AWS DataSync 任务列表。

请求语法

```
{
   "Filters": [
      {
         "Name": "string",
         "Operator": "string",
         "Values": [ "string" ]
      }
   ],
   "MaxResults": number,
   "NextToken": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

Filters

您可以使用 API 筛选器来缩小 ListTasks 返回的资源的列表范围。例如，要检索特定源位置上的所有任务，您可以结合该位置的 ARN 使用筛选器名称为 LocationId 和 Operator Equals 的 ListTasks。

类型：TaskFilter 对象数组

必需：否

MaxResults

要返回的任务的最大数。

类型：整数

有效范围：最小值为 0。最大值为 100。

必需：否

NextToken

一个不透明的字符串，用于指示开始下一个任务列表的位置。

类型：字符串

长度约束：最大长度为 65535。

模式：[a-zA-Z0-9=_-]+

必需：否

响应语法

```
{
   "NextToken": "string",
   "Tasks": [
      {
         "Name": "string",
         "Status": "string",
         "TaskArn": "string",
         "TaskMode": "string"
      }
   ]
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

NextToken

一个不透明的字符串，用于指示开始返回下一个任务列表的位置。

类型：字符串

长度约束：最大长度为 65535。

模式：[a-zA-Z0-9=_-]+

Tasks

返回的所有任务的列表。

类型：TaskListEntry 对象数组

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

StartTaskExecution

启动 AWS DataSync 传输任务。对于每个任务，您一次只能运行一个任务执行。

任务执行分为几个步骤。有关更多信息，请参阅任务执行状态。

> **Important**
>
> 如果您计划在 Amazon S3 地点之间传输数据，请在开始之前查看 DataSync 如何影响您的 S3 请求费用和DataSync 定价页面。

请求语法

```
{
   "Excludes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "Includes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "ManifestConfig": {
      "Action": "string",
      "Format": "string",
      "Source": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "ManifestObjectPath": "string",
            "ManifestObjectVersionId": "string",
            "S3BucketArn": "string"
         }
      }
   },
   "OverrideOptions": {
      "Atime": "string",
      "BytesPerSecond": number,
```

```
      "Gid": "string",
      "LogLevel": "string",
      "Mtime": "string",
      "ObjectTags": "string",
      "OverwriteMode": "string",
      "PosixPermissions": "string",
      "PreserveDeletedFiles": "string",
      "PreserveDevices": "string",
      "SecurityDescriptorCopyFlags": "string",
      "TaskQueueing": "string",
      "TransferMode": "string",
      "Uid": "string",
      "VerifyMode": "string"
   },
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ],
   "TaskArn": "string",
   "TaskReportConfig": {
      "Destination": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "S3BucketArn": "string",
            "Subdirectory": "string"
         }
      },
      "ObjectVersionIds": "string",
      "OutputType": "string",
      "Overrides": {
         "Deleted": {
            "ReportLevel": "string"
         },
         "Skipped": {
            "ReportLevel": "string"
         },
         "Transferred": {
            "ReportLevel": "string"
         },
         "Verified": {
            "ReportLevel": "string"
         }
```

```
        },
        "ReportLevel": "string"
    }
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

Excludes

指定筛选规则列表，用于确定从任务中排除哪些文件。该列表包含一个由要排除的模式组成的筛选字符串。模式由“|”（即竖线）分隔，例如 `"/folder1|/folder2"`。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

必需：否

Includes

指定筛选规则列表，用于确定运行任务时要包含哪些文件。该模式应包含一个由要包含的模式组成的筛选字符串。模式由“|”（即竖线）分隔，例如 `"/folder1|/folder2"`。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

必需：否

ManifestConfig

配置清单，这是您要 DataSync 传输的文件或对象的列表。有关更多信息和配置示例，请参阅使用清单指定 DataSync 传输的内容。

使用此参数时，您的来电者身份（您使用的 DataSync 角色）必须具有`iam:PassRole`权限。该AWSDataSyncFullAccess策略包含此权限。

要移除清单配置，请将此参数指定为空值。

类型：ManifestConfig 对象

必需：否

OverrideOptions

指明您的传输任务是如何配置的。这些选项包括在传输过程中如何 DataSync 处理文件、对象及其关联的元数据。您还可以指定如何验证数据完整性、为任务设置带宽限制等其他选项。

每个选项都有一个默认值。除非需要，否则无需在拨打电话之前配置任何选项StartTaskExecution。

您还可以在每次执行任务时覆盖任务选项。例如，您可能希望为单次执行调整 LogLevel。

类型：Options 对象

必需：否

Tags

指定要应用于代表任务执行的 Amazon 资源名称（ARN）的标签。

标签是键值对，可帮助您管理、筛选和搜索 DataSync 资源。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

必需：否

TaskArn

指定您要启动的任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$`

是否必需：是

TaskReportConfig

指定您想要如何配置任务报告，该报告提供有关您的 DataSync 转移的详细信息。有关更多信息，请参阅使用任务报告监控您的 DataSync转账。

使用此参数时，您的来电者身份（您使用的 DataSync 角色）必须具有`iam:PassRole`权限。该AWSDataSyncFullAccess策略包含此权限。

要移除任务报告配置，请将此参数指定为空。

类型：TaskReportConfig 对象

必需：否

响应语法

```
{
   "TaskExecutionArn": "string"
}
```

响应元素

如果此操作成功，则该服务将会发送回 HTTP 200 响应。

服务以 JSON 格式返回以下数据。

TaskExecutionArn

运行任务执行的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}/execution/exec-[0-9a-f]{17}$`

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

示例

示例请求

以下示例使用指定任务的默认选项开启任务执行。

```
{
    "OverrideOptions": {
        "Atime": "BEST_EFFORT",
        "BytesPerSecond": 1000,
        "Gid": "NONE",
        "Mtime": "PRESERVE",
        "PosixPermissions": "PRESERVE",
        "PreserveDevices": "NONE",
        "PreserveDeletedFiles": "PRESERVE",
        "Uid": "NONE",
        "VerifyMode": "POINT_IN_TIME_CONSISTENT"
    },
    "TaskArn": "arn:aws:datasync:us-east-2:111222333444:task/task-08de6e6697796f026"
}
```

示例响应

此示例说明了的一种用法 StartTaskExecution。

```
{
  "TaskExecutionArn": "arn:aws:datasync:us-east-2:111222333444:task/
task-08de6e6697796f026/execution/exec-04ce9d516d69bd52f"
}
```

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

TagResource

将标签应用于 AWS 资源。标签是键值对，可帮助您管理、筛选和搜索资源。

其中包括 AWS DataSync 资源，例如地点、任务和任务执行。

请求语法

```
{
   "ResourceArn": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

ResourceArn

指定应用标签的资源的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:(((agent|task|location)/(agent|task|loc)-[a-z0-9]{17}(/execution/exec-[a-f0-9]{17})?)|(system/storage-system-[a-f0-9]{8}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{12}(/job/discovery-job-[a-f0-9]{8}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{12})?))$`

是否必需：是

Tags

指定要应用于资源的标签。

类型：TagListEntry 对象数组

数组成员：最少 0 个物品。最多 50 项。

是否必需：是

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包

- AWS 适用于 Ruby V3 的 SDK

UntagResource

从 AWS 资源中移除标签。

请求语法

```
{
   "Keys": [ "string" ],
   "ResourceArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

Keys

指定您要删除的标签中的密钥。

类型：字符串数组

数组成员：最少 1 个物品。最多 50 项。

长度限制：最小长度为 1。最大长度为 256。

模式：^[a-zA-Z0-9\s+=._:/-]+$

是否必需：是

ResourceArn

指定要从中删除标签的资源的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:(((agent|task|location)/(agent|task|loc)-[a-z0-9]{17}(/execution/exec-[a-f0-9]{17})?)|(system/storage-system-[a-f0-9]{8}-

```
[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{12}(/job/discovery-job-[a-f0-9]{8}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{4}-[a-f0-9]{12})?))$
```

是否必需：是

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包

- AWS 适用于 Ruby V3 的 SDK

UpdateAgent

更新 AWS DataSync 代理的名称。

请求语法

```
{
   "AgentArn": "string",
   "Name": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArn

要更新的代理的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

是否必需：是

Name

要用于配置代理的名称。

类型：字符串

长度限制：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationAzureBlob

修改你正在使用的微软 Azure Blob 存储传输位置的以下配置。 AWS DataSync

有关更多信息，请参阅使用 Azure Blob 存储配置 DataSync 传输。

请求语法

```
{
   "AccessTier": "string",
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "BlobType": "string",
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "LocationArn": "string",
   "SasConfiguration": {
      "Token": "string"
   },
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AccessTier

指定您希望将对象或文件传输到的访问层。这仅适用于使用该位置作为传输目标的情况。有关更多信息，请参阅访问层。

类型：字符串

有效值：HOT | COOL | ARCHIVE

必需：否

AgentArns

（可选）指定可以与 Azure Blob 存储容器连接的 DataSync 代理的亚马逊资源名称 (ARN)。如果设置的是无代理跨云传输，则无需为此参数指定值。

您可以指定多个代理。有关更多信息，请参阅使用多个代理进行传输。

> **Note**
>
> 最初创建代理后，就无法在存储位置添加或删除代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

AuthenticationType

指定用于访问 Azure Blob 存储空间的身份验证方法 DataSync 。 DataSync 可以使用共享访问签名 (SAS) 访问 blob 存储。

类型：字符串

有效值：`SAS | NONE`

必需：否

BlobType

指定在将对象或文件传输到 Azure Blob Storage 时想要的 blob 类型。目前， DataSync 仅支持将数据作为块 blob 移动到 Azure Blob 存储。有关 blob 类型的更多信息，请参阅 Azure Blob Storage 文档。

类型：字符串

有效值：BLOCK

必需：否

CmkSecretConfig

指定 DataSync托管密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组凭证，DataSync 以及客户 AWS KMS key管理的密钥。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组证书，以及提供密钥访问权限的客户管理 AWS Identity and Access Management (IAM) 角色。 DataSync

类型：CustomSecretConfig 对象

必需：否

LocationArn

指定您要更新的 Azure Blob Storage 传输位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

SasConfiguration

指定允许 DataSync 访问你的 Azure Blob 存储空间的 SAS 配置。

类型：AzureBlobSasConfiguration 对象

必需：否

Subdirectory

如果要将传输限制到容器中的虚拟目录（例如 `/my/images`），请指定路径段。

类型：字符串

长度限制：最大长度为 1024。

模式：`^[\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}\p{C}]*$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版 SDK
- AWS 适用于 Kotlin 的 SDK

- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationEfs

修改您正在使用的 Amazon EFS 传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅使用 Amazon EFS 配置 DataSync传输。

请求语法

```
{
   "AccessPointArn": "string",
   "FileSystemAccessRoleArn": "string",
   "InTransitEncryption": "string",
   "LocationArn": "string",
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AccessPointArn

指定用于挂载 Amazon EFS 文件系统的接入点的 DataSync 亚马逊资源名称 (ARN)。

有关更多信息，请参阅访问受限 Amazon EFS 文件系统。

类型：字符串

长度限制：最大长度为 128。

模式：`(^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):elasticfilesystem:[a-z\-0-9]+:[0-9]{12}:access-point/fsap-[0-9a-f]{8,40}$)|(^$)`

必需：否

FileSystemAccessRoleArn

指定允许 DataSync 访问您的 Amazon EFS 文件系统的 AWS Identity and Access Management (IAM) 角色。

有关创建此角色的信息，请参阅创建 DataSync IAM 角色以访问 Amazon EFS 文件系统。

类型：字符串

长度约束：最大长度为 2048。

模式：`(^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):iam::[0-9]{12}:role/.*$)|(^$)`

必需：否

InTransitEncryption

指定在传输层安全 (TLS) 1.2 加密与您的 Amazon EFS 文件系统之间传输数据时，是否要使用传输层安全 (TLS) 1.2 加密。 DataSync

如果您使用 `AccessPointArn` 指定访问点或使用 `FileSystemAccessRoleArn` 指定 IAM 角色，则必须将此参数设置为 `TLS1_2`。

类型：字符串

有效值：`NONE | TLS1_2`

必需：否

LocationArn

指定要更新的 Amazon EFS 传输位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

Subdirectory

指定 Amazon EFS 文件系统的挂载路径。这是在文件系统上 DataSync 读取或写入数据的地方（取决于这是源位置还是目标位置）。

默认情况下， DataSync 使用根目录（如果使用提供接入点，则使用接入点`AccessPointArn`）。您也可以使用正斜杠包含子目录（例如，`/path/to/folder`）。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK

- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationFsxLustre

修改您正在使用的 Amazon FSx for Lustre 传输地点的以下配置参数。 AWS DataSync

有关更多信息，请参阅为 Lustre 配置 DataSync传输。 FSx

请求语法

```
{
   "LocationArn": "string",
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定您要更新的 for Lustre 传输地点 FSx 的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

Subdirectory

为您 FSx 的 for Lustre 文件系统指定装载路径。该路径可包含子目录。

当该位置用作源时，会从装载路径 DataSync 读取数据。当该位置用作目标时， DataSync 会将数据写入装载路径。如果不包含此参数，则 DataSync 使用文件系统的根目录 (/)。

类型：字符串

长度约束：最大长度为 4096。

模式：^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationFsxOntap

修改您正在使用的 Amazon f NetApp o FSx r ONTAP 传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅为 ONTAP 配置 DataSync传输。 FSx

请求语法

```
{
   "LocationArn": "string",
   "Protocol": {
      "NFS": {
         "MountOptions": {
            "Version": "string"
         }
      },
      "SMB": {
         "CmkSecretConfig": {
            "KmsKeyArn": "string",
            "SecretArn": "string"
         },
         "CustomSecretConfig": {
            "SecretAccessRoleArn": "string",
            "SecretArn": "string"
         },
         "Domain": "string",
         "MountOptions": {
            "Version": "string"
         },
         "Password": "string",
         "User": "string"
      }
   },
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定您要更新的 FSx 适用于 ONTAP 的传输位置的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

Protocol

指定用于访问您的 DataSync Amazon FSx 文件系统的数据传输协议。

类型：FsxUpdateProtocol 对象

必需：否

Subdirectory

指定要传入或传出数据的存储虚拟机（SVM）中的文件共享路径。

您可以指定连接路径（也称为挂载点）、qtree 路径（用于 NFS 文件共享）或共享名称（用于 SMB 文件共享）。例如，您的挂载路径可能是 `/vol1`、`/vol1/tree1` 或 `/share1`。

> **Note**
>
> 不要在 SVM 的根卷中指定连接路径。有关更多信息，请参阅《Amazon FSx FSx for ONTAP 用户指南》中的管理 NetApp ONTAP 存储虚拟机。

类型：字符串

长度限制：长度上限为 255。

模式：`^[^\u0000\u0085\u2028\u2029\r\n]{1,255}$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationFsxOpenZfs

修改您正在使用的 Amazon FSx for OpenZFS 传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅为 OpenZ DataSync FS FSx 配置传输。

> **Note**
>
> UpdateLocationFsxOpenZfs 操作不支持与 SMB 相关的请求参数。

请求语法

```
{
   "LocationArn": "string",
   "Protocol": {
      "NFS": {
         "MountOptions": {
            "Version": "string"
         }
      },
      "SMB": {
         "CmkSecretConfig": {
            "KmsKeyArn": "string",
            "SecretArn": "string"
         },
         "CustomSecretConfig": {
            "SecretAccessRoleArn": "string",
            "SecretArn": "string"
         },
         "Domain": "string",
         "ManagedSecretConfig": {
            "SecretArn": "string"
         },
         "MountOptions": {
            "Version": "string"
         },
         "Password": "string",
         "User": "string"
      }
   },
   "Subdirectory": "string"
```

```
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定您要更新的 FSx 适用于 OpenZFS 的传输位置的亚马逊资源名称 (ARN)。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

Protocol

指定用于访问您的 AWS DataSync Amazon FSx 文件系统的数据传输协议。

类型：FsxProtocol 对象

必需：否

Subdirectory

指定位置路径中必须以开头的子目录。`/fsx` DataSync 使用此子目录读取或写入数据（取决于文件系统是源位置还是目标位置）。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationFsxWindows

修改你正在使用的 Amazon f FSx or Windows 文件服务器传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅为 Windows 文件服务器配置 DataSync传输。 FSx

请求语法

```
{
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "Domain": "string",
   "LocationArn": "string",
   "Password": "string",
   "Subdirectory": "string",
   "User": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

CmkSecretConfig

指定 DataSync托管密钥的配置信息，例如用于访问特定传输位置和客户 AWS KMS key管理的 DataSync 凭据的一组Password或一组凭据。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的密钥的配置信息，例如用于访问特定传输位置的一组Password或一组证书，以及提供该 DataSync 密钥访问权限的客户管理 AWS Identity and Access Management (IAM) 角色。

类型：CustomSecretConfig 对象

必需：否

Domain

指定你 FSx 的 Windows 文件服务器文件系统所属的 Windows 域的名称。

如果您的环境中有多个 Active Directory 域，则配置此参数可确保这些域 DataSync 连接到正确的文件系统。

类型：字符串

长度约束：最大长度为 253。

模式：`^([A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252})?$`

必需：否

LocationArn

指定要更新的 FSx 适用于 Windows 文件服务器的传输位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

Password

指定有权装载和访问您 FSx 的 Windows 文件服务器文件系统中的文件、文件夹和文件元数据的用户的密码。

类型：字符串

长度约束：最大长度为 104。

模式：`^.{0,104}$`

必需：否

Subdirectory

使用正斜杠为您的文件系统指定挂载路径。 DataSync 使用此子目录读取或写入数据（取决于文件系统是源位置还是目标位置）。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$`

必需：否

User

指定有权装载和访问您 FSx 的 Windows 文件服务器文件系统中的文件、文件夹和文件元数据的用户。

有关选择具有适当访问权限级别的用户进行传输的信息，请参阅 Windows 文件服务器位置 FSx 所需的权限。

类型：字符串

长度约束：最大长度为 104。

模式：`^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationHdfs

修改您正在使用的 Hadoop 分布式文件系统 (HDFS) 传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅使用 HDFS 群集配置 DataSync传输。

请求语法

```
{
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "BlockSize": number,
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "KerberosKeytab": blob,
   "KerberosKrb5Conf": blob,
   "KerberosPrincipal": "string",
   "KmsKeyProviderUri": "string",
   "LocationArn": "string",
   "NameNodes": [
      {
         "Hostname": "string",
         "Port": number
      }
   ],
   "QopConfiguration": {
      "DataTransferProtection": "string",
      "RpcProtection": "string"
   },
   "ReplicationFactor": number,
   "SimpleUser": "string",
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArns

可以连接到您的 HDFS 集群的 DataSync 代理的 Amazon 资源名称 (ARNs)。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

AuthenticationType

用于确定用户身份的身份验证的类型。

类型：字符串

有效值：`SIMPLE | KERBEROS`

必需：否

BlockSize

要写入 HDFS 集群的数据块大小。

类型：整数

有效范围：最小值为 1048576。最大值为 1073741824。

必需：否

CmkSecretConfig

指定 DataSync托管密钥的配置信息，例如用于访问特定传输位置和客户 AWS KMS key管理的 DataSync 凭据的一组`KerberosKeytab`或一组凭据。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的密钥的配置信息，例如用于访问特定传输位置的一组KerberosKeytab或一组证书，以及提供该 DataSync 密钥访问权限的客户管理 AWS Identity and Access Management (IAM) 角色。

类型：CustomSecretConfig 对象

必需：否

KerberosKeytab

Kerberos 密钥表（keytab），其中包含定义的 Kerberos 主体和加密密钥之间的映射。您可以通过提供文件的地址从该文件加载密钥表。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 65536。

必需：否

KerberosKrb5Conf

包含 Kerberos 配置信息的 krb5.conf 文件。您可以通过提供文件的地址来加载 krb5.conf 文件。如果您使用的是 AWS CLI，它将为您执行 base64 编码。否则，应提供 base64 编码文本。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 131072。

必需：否

KerberosPrincipal

有权访问 HDFS 集群上的文件和文件夹的 Kerberos 主体。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：^.+$

必需：否

KmsKeyProviderUri

HDFS 集群的密钥管理服务器（KMS）的 URI。

类型：字符串

长度限制：最小长度为 1。最大长度为 255。

模式：`^kms:\/\/http[s]?@(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-]*[A-Za-z0-9])(;(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-]*[A-Za-z0-9]))*:[0-9]{1,5}\/kms$`

必需：否

LocationArn

来源 HDFS 集群位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

NameNodes

管理 HDFS 命名空间的。 NameNode NameNode 执行诸如打开、关闭和重命名文件和目录之类的操作。 NameNode 包含用于将数据块映射到的信息 DataNodes。您只能使用一个 NameNode。

类型：HdfsNameNode 对象数组

数组成员：最少 1 个物品。

必需：否

QopConfiguration

保护质量（QOP）配置指定在 Hadoop Distributed File System（HDFS）集群上配置的远程程序调用（RPC）和数据传输隐私设置。

类型：QopConfiguration 对象

必需：否

ReplicationFactor

写入 HDFS 集群时要将数据复制到的数量。 DataNodes

类型：整数

有效范围：最小值为 1。最大值为 512。

必需：否

SimpleUser

用于在主机操作系统上标识客户端的用户名。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：`^[_.A-Za-z0-9][-_.A-Za-z0-9]*$`

必需：否

Subdirectory

HDFS 集群中的子目录。此子目录用于从 HDFS 集群读取数据或向其写入数据。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationNfs

修改您正在使用的网络文件系统 (NFS) 传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅使用 NFS 文件服务器配置传输。

请求语法

```
{
   "LocationArn": "string",
   "MountOptions": {
      "Version": "string"
   },
   "OnPremConfig": {
      "AgentArns": [ "string" ]
   },
   "ServerHostname": "string",
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定您要更新的 NFS 传输位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

MountOptions

指定 DataSync 如何使用 NFS 协议访问某个位置。

类型：NfsMountOptions 对象

必需：否

OnPremConfig

可以连接到您的网络文件系统 (NFS) 文件服务器的 AWS DataSync 代理。

类型：OnPremConfig 对象

必需：否

ServerHostname

指定 DataSync 代理所连接的 NFS 文件服务器的 DNS 名称IPv4 或 IP 地址（或 IPv6）。

类型：字符串

长度限制：长度上限为 255。

模式：`^(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-:]*[A-Za-z0-9])$`

必需：否

Subdirectory

在 NFS 文件服务器中指定 DataSync 要装载的导出路径。

此路径（或路径的子目录）是传入或 DataSync 传出数据的地方。有关为配置导出的信息 DataSync，请参阅访问 NFS 文件服务器。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]+$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationObjectStorage

修改您正在使用的对象存储传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅使用对象存储系统配置 DataSync传输。

请求语法

```
{
   "AccessKey": "string",
   "AgentArns": [ "string" ],
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "LocationArn": "string",
   "SecretKey": "string",
   "ServerCertificate": blob,
   "ServerHostname": "string",
   "ServerPort": number,
   "ServerProtocol": "string",
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AccessKey

如果需要凭证来向对象存储服务器进行身份验证，则指定访问密钥（例如，用户名）。

类型：字符串

长度限制：最小长度为 0。最大长度为 200。

模式：^.*$

必需：否

AgentArns

（可选）指定可以与您的对象存储系统连接的 DataSync 代理的 Amazon 资源名称 (ARNs)。如果设置的是无代理跨云传输，则无需为此参数指定值。

> **Note**
>
> 最初创建代理后，就无法在存储位置添加或删除代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

CmkSecretConfig

指定 DataSync托管密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组凭证，DataSync 以及客户 AWS KMS key管理的密钥。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组证书，以及提供密钥访问权限的客户管理 AWS Identity and Access Management (IAM) 角色。 DataSync

类型：CustomSecretConfig 对象

必需：否

LocationArn

指定您要更新的对象存储系统位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

SecretKey

如果需要凭证来向对象存储服务器进行身份验证，则指定私有密钥（例如，密码）。

> **Note**
>
> 如果您使用提供密钥`SecretKey`，但未使用`CmkSecretConfig`或提供密钥配置详细信息`CustomSecretConfig`，则使用您 AWS 账户的 Secrets Manager 密钥 DataSync 存储令牌。

类型：字符串

长度限制：最小长度为 0。最大长度为 200。

模式：`^.*$`

必需：否

ServerCertificate

如果您的对象存储系统使用私有或自签名证书颁发机构 (CA)，则指定一个证书链， DataSync 以便在对象存储系统中进行身份验证。您必须指定具有完整证书链的单个 `.pem` 文件（例如，`file:///home/user/.ssh/object_storage_certificates.pem`）。

证书链可能包括：

- 对象存储系统的证书
- 所有中间证书（如果有）
- 签名 CA 的根证书

您可以将证书串联到一个 `.pem` 文件中（在进行 base64 编码前，文件最大长度可以为 32768 字节）。以下示例 `cat` 命令创建了一个包含三个证书的 `object_storage_certificates.pem` 文件：

```
cat object_server_certificate.pem intermediate_certificate.pem
ca_root_certificate.pem > object_storage_certificates.pem
```

要使用此参数，请配置 ServerProtocol 为 HTTPS。

更新此参数不会干扰您正在进行的任务。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 32768。

必需：否

ServerHostname

指定 DataSync 代理所连接的对象存储服务器的域名IPv4 或 IP 地址（或 IPv6）。

类型：字符串

长度限制：长度上限为 255。

模式：`^(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-:]*[A-Za-z0-9])$`

必需：否

ServerPort

指定对象存储服务器接受入站网络流量的端口（例如，端口 443）。

类型：整数

有效范围：最小值为 1。最大值为 65536。

必需：否

ServerProtocol

指定对象存储服务器通信时使用的协议。

类型：字符串

有效值：`HTTPS | HTTP`

必需：否

Subdirectory

指定对象存储服务器的对象前缀。如果这是源位置，则 DataSync 仅复制带有此前缀的对象。如果这是目标位置，则 DataSync写入带有此前缀的所有对象。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK

- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateLocationS3

修改您正在使用的 Amazon S3 传输位置的以下配置参数。 AWS DataSync

> **Important**
>
> 开始之前，请务必阅读以下主题：
>
> - Amazon S3 位置的存储类考虑因素
> - 使用时评估 S3 请求成本 DataSync

请求语法

```
{
   "LocationArn": "string",
   "S3Config": {
      "BucketAccessRoleArn": "string"
   },
   "S3StorageClass": "string",
   "Subdirectory": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

LocationArn

指定要更新的 Amazon S3 传输位置的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

S3Config

指定用于访问您的 S3 存储桶的 AWS Identity and Access Management (IAM) 角色的 DataSync 亚马逊资源名称 (ARN)。

有关更多信息，请参阅提供 DataSync 对 S3 存储桶的访问权限。

类型：S3Config 对象

必需：否

S3StorageClass

指定当 Amazon S3 作为传输目标时，您希望对象使用的存储类别。

对于中的存储桶 AWS 区域，存储类默认为。STANDARD对于开启的存储桶 AWS Outposts，存储类默认为。OUTPOSTS

有关更多信息，请参阅 Amazon S3 传输的存储类别注意事项。

类型：字符串

有效值：STANDARD | STANDARD_IA | ONEZONE_IA | INTELLIGENT_TIERING | GLACIER | DEEP_ARCHIVE | OUTPOSTS | GLACIER_INSTANT_RETRIEVAL

必需：否

Subdirectory

在 S3 存储桶中指定 DataSync 读取或写入的前缀（取决于存储桶是源位置还是目标位置）。

> **Note**
>
> DataSync 无法传输前缀以斜杠 (/) 开头或包含///./、或/../模式的对象。例如：
>
> - /photos
> - photos//2006/January
> - photos/./2006/February
> - photos/../2006/March

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包

- AWS 适用于 Ruby V3 的 SDK

UpdateLocationSmb

修改您正在使用的服务器消息块 (SMB) 传输位置的以下配置参数。 AWS DataSync

有关更多信息，请参阅使用 SMB 文件服务器配置 DataSync传输。

请求语法

```
{
   "AgentArns": [ "string" ],
   "AuthenticationType": "string",
   "CmkSecretConfig": {
      "KmsKeyArn": "string",
      "SecretArn": "string"
   },
   "CustomSecretConfig": {
      "SecretAccessRoleArn": "string",
      "SecretArn": "string"
   },
   "DnsIpAddresses": [ "string" ],
   "Domain": "string",
   "KerberosKeytab": blob,
   "KerberosKrb5Conf": blob,
   "KerberosPrincipal": "string",
   "LocationArn": "string",
   "MountOptions": {
      "Version": "string"
   },
   "Password": "string",
   "ServerHostname": "string",
   "Subdirectory": "string",
   "User": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

AgentArns

指定 DataSync 可以连接到 SMB 文件服务器的一个或多个代理。使用代理的 Amazon 资源名称 (ARN) 指定代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

AuthenticationType

指定用于连接您的 DataSync SMB 文件服务器的身份验证协议。 DataSync 支持`NTLM`（默认）和`KERBEROS`身份验证。

有关更多信息，请参阅提供 DataSync 对 SMB 文件服务器的访问权限。

类型：字符串

有效值：`NTLM | KERBEROS`

必需：否

CmkSecretConfig

指定 DataSync托管密钥的配置信息，例如用于访问特定传输位置的一组`PasswordKerberosKeytab`或或一组凭据以及客户 AWS KMS key管理的凭据。 DataSync

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的密钥的配置信息，例如用于访问特定传输位置的一组`PasswordKerberosKeytab`或或一组证书，以及提供该 DataSync 密钥访问权限的客户管理 AWS Identity and Access Management (IAM) 角色。

类型：CustomSecretConfig 对象

必需：否

DnsIpAddresses

指定 SMB 文件服务器所属的 DNS 服务器的 IP 地址（IPv4 或 IPv6）。此参数仅在 `AuthenticationType` 设置为 `KERBEROS` 时适用。

如果您的环境中有多个域，则配置此参数 DataSync 可确保连接到正确的 SMB 文件服务器。

类型：字符串数组

数组成员：最多 2 项。

长度限制：最小长度为 7。最大长度为 39。

模式：`\A((25[0-5]|2[0-4]\d|[0-1]?\d?\d)(\.(25[0-5]|2[0-4]\d|[0-1]?\d?\d)){3}|([0-9a-fA-F]{1,4}:){7,7}[0-9a-fA-F]{1,4}|([0-9a-fA-F]{1,4}:){1,7}:|([0-9a-fA-F]{1,4}:){1,6}:[0-9a-fA-F]{1,4}|([0-9a-fA-F]{1,4}:){1,5}(:[0-9a-fA-F]{1,4}){1,2}|([0-9a-fA-F]{1,4}:){1,4}(:[0-9a-fA-F]{1,4}){1,3}|([0-9a-fA-F]{1,4}:){1,3}(:[0-9a-fA-F]{1,4}){1,4}|([0-9a-fA-F]{1,4}:){1,2}(:[0-9a-fA-F]{1,4}){1,5}|[0-9a-fA-F]{1,4}:((:[0-9a-fA-F]{1,4}){1,6}))\z`

必需：否

Domain

指定您的 SMB 文件服务器所属的 Windows 域名称。此参数仅在 `AuthenticationType` 设置为 `NTLM` 时适用。

如果您的环境中有多个域，则配置此参数可确保 DataSync 连接到正确的文件服务器。

类型：字符串

长度约束：最大长度为 253。

模式：`^[A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252}$`

必需：否

KerberosKeytab

指定 Kerberos 密钥表（keytab）文件，其中包括 Kerberos 主体和加密密钥之间的映射。

为避免出现任务执行错误，请确保用于创建 keytab 文件的 Kerberos 主体与为 `KerberosPrincipal` 指定的内容完全匹配。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 65536。

必需：否

KerberosKrb5Conf

指定用于定义 Kerberos 领域配置的 Kerberos 配置文件（`krb5.conf`）。

该文件必须采用 base64 编码。如果您使用的是 AWS CLI，则编码已为您完成。

类型：Base64 编码的二进制数据对象

长度约束：最大长度为 131072。

必需：否

KerberosPrincipal

指定 Kerberos 主体，即 Kerberos 领域中有权访问 SMB 文件服务器中文件、文件夹和文件元数据的身份。

Kerberos 主体可能看起来像 `HOST/kerberosuser@MYDOMAIN.ORG`。

主体名称区分大小写。如果您为此参数指定的主体与您用于创建 keytab 文件的主体不完全匹配，则 DataSync 任务执行将失败。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：`^.+$`

必需：否

LocationArn

指定要更新的 SMB 位置的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

是否必需：是

MountOptions

指定用于访问 SMB 文件服务器的服务器消息块 (SMB) 协议的版本。 AWS DataSync

类型：SmbMountOptions 对象

必需：否

Password

指定可以挂载 SMB 文件服务器并有权限访问传输中涉及的文件和文件夹的用户的密码。此参数仅在 `AuthenticationType` 设置为 `NTLM` 时适用。

类型：字符串

长度约束：最大长度为 104。

模式：`^.{0,104}$`

必需：否

ServerHostname

指定您的 DataSync 代理所连接的 SMB 文件服务器的域名IPv4 或 IP 地址（或 IPv6）。

> **Note**
>
> 如果使用的是 Kerberos 身份验证，则必须指定域名。

类型：字符串

长度限制：长度上限为 255。

模式：`^(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-:]*[A-Za-z0-9])$`

必需：否

Subdirectory

指定 SMB 文件服务器导出的共享的名称，该共享 DataSync将在其中读取或写入数据。您可以在共享路径中包含子目录（例如，`/path/to/subdirectory`）。确保网络中的其他 SMB 客户端也可以挂载此路径。

要复制指定子目录中的所有数据，DataSync 必须能够装载 SMB 共享并访问其所有数据。有关更多信息，请参阅提供 DataSync 对 SMB 文件服务器的访问权限。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\$\p{Zs}]+$`

必需：否

User

指定可以挂载 SMB 文件服务器并有权限访问传输中涉及的文件和文件夹的用户名。此参数仅在 `AuthenticationType` 设置为 `NTLM` 时适用。

有关选择具有适当访问权限级别的用户进行传输的信息，请参阅提供对 SMB 文件服务器的 DataSync 访问权限。

类型：字符串

长度约束：最大长度为 104。

模式：`^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$`

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateTask

更新任务的配置，该配置定义了 AWS DataSync 传输数据的位置和方式。

请求语法

```
{
   "CloudWatchLogGroupArn": "string",
   "Excludes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "Includes": [
      {
         "FilterType": "string",
         "Value": "string"
      }
   ],
   "ManifestConfig": {
      "Action": "string",
      "Format": "string",
      "Source": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "ManifestObjectPath": "string",
            "ManifestObjectVersionId": "string",
            "S3BucketArn": "string"
         }
      }
   },
   "Name": "string",
   "Options": {
      "Atime": "string",
      "BytesPerSecond": number,
      "Gid": "string",
      "LogLevel": "string",
      "Mtime": "string",
      "ObjectTags": "string",
      "OverwriteMode": "string",
      "PosixPermissions": "string",
      "PreserveDeletedFiles": "string",
```

```
      "PreserveDevices": "string",
      "SecurityDescriptorCopyFlags": "string",
      "TaskQueueing": "string",
      "TransferMode": "string",
      "Uid": "string",
      "VerifyMode": "string"
   },
   "Schedule": {
      "ScheduleExpression": "string",
      "Status": "string"
   },
   "TaskArn": "string",
   "TaskReportConfig": {
      "Destination": {
         "S3": {
            "BucketAccessRoleArn": "string",
            "S3BucketArn": "string",
            "Subdirectory": "string"
         }
      },
      "ObjectVersionIds": "string",
      "OutputType": "string",
      "Overrides": {
         "Deleted": {
            "ReportLevel": "string"
         },
         "Skipped": {
            "ReportLevel": "string"
         },
         "Transferred": {
            "ReportLevel": "string"
         },
         "Verified": {
            "ReportLevel": "string"
         }
      },
      "ReportLevel": "string"
   }
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

CloudWatchLogGroupArn

指定用于监控任务的亚马逊 CloudWatch 日志组的亚马逊资源名称 (ARN)。

对于增强模式任务，必须使用 /aws/datasync 作为日志组的名称。例如：

arn:aws:logs:us-east-1:111222333444:log-group:/aws/datasync:*

有关更多信息，请参阅使用 CloudWatch 日志监控数据传输。

类型：字符串

长度约束：最大长度为 562。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):logs:[a-z\-0-9]+:[0-9]{12}:log-group:([^:*]*)(:*)?$

必需：否

Excludes

指定排除过滤器，用于定义源位置中您不想 DataSync 传输的文件、对象和文件夹。有关更多信息和示例，请参阅使用过滤器指定 DataSync 传输的内容。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

必需：否

Includes

指定包含筛选器定义源位置中 DataSync 要传输的文件、对象和文件夹。有关更多信息和示例，请参阅使用过滤器指定 DataSync 传输的内容。

类型：FilterRule 对象数组

数组成员：最少 0 个物品。最多 1 项。

必需：否

ManifestConfig

配置清单，这是您要 DataSync 传输的文件或对象的列表。有关更多信息和配置示例，请参阅使用清单指定 DataSync 传输的内容。

使用此参数时，您的来电者身份（与之 DataSync 一起使用的 IAM 角色）必须具有`iam:PassRole`权限。该AWSDataSyncFullAccess策略包含此权限。

要移除清单配置，请将此参数指定为空。

类型：ManifestConfig 对象

必需：否

Name

指定任务的名称。

类型：字符串

长度限制：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

必需：否

Options

指明您的传输任务是如何配置的。这些选项包括在传输过程中如何 DataSync 处理文件、对象及其关联的元数据。您还可以指定如何验证数据完整性、为任务设置带宽限制等其他选项。

每个选项都有一个默认值。除非需要，否则无需在拨打电话之前配置任何选项StartTaskExecution。

您还可以在每次执行任务时覆盖任务选项。例如，您可能希望为单次执行调整 `LogLevel`。

类型：Options 对象

必需：否

Schedule

指定任务运行的时间计划。有关更多信息，请参阅计划任务。

类型：TaskSchedule 对象

必需：否

TaskArn

指定您要更新的任务的 ARN。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$`

是否必需：是

TaskReportConfig

指定您想要如何配置任务报告，该报告提供有关您的 DataSync 转移的详细信息。有关更多信息，请参阅使用任务报告监控您的 DataSync转账。

使用此参数时，您的来电者身份（与之 DataSync 一起使用的 IAM 角色）必须具有`iam:PassRole`权限。该AWSDataSyncFullAccess策略包含此权限。

要移除任务报告配置，请将此参数指定为空。

类型：TaskReportConfig 对象

必需：否

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版软件开发工具包
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

UpdateTaskExecution

更新正在运行的 AWS DataSync 任务执行的配置。

> **Note**
>
> 当前，您唯一通过UpdateTaskExecution修改的Option是 BytesPerSecond ，限制正在运行或排队的任务执行的带宽。

请求语法

```
{
   "Options": {
      "Atime": "string",
      "BytesPerSecond": number,
      "Gid": "string",
      "LogLevel": "string",
      "Mtime": "string",
      "ObjectTags": "string",
      "OverwriteMode": "string",
      "PosixPermissions": "string",
      "PreserveDeletedFiles": "string",
      "PreserveDevices": "string",
      "SecurityDescriptorCopyFlags": "string",
      "TaskQueueing": "string",
      "TransferMode": "string",
      "Uid": "string",
      "VerifyMode": "string"
   },
   "TaskExecutionArn": "string"
}
```

请求参数

有关所有操作的通用参数的信息，请参阅常用参数。

请求接受采用 JSON 格式的以下数据。

Options

指明您的传输任务是如何配置的。这些选项包括在传输过程中如何 DataSync 处理文件、对象及其关联的元数据。您还可以指定如何验证数据完整性、为任务设置带宽限制等其他选项。

每个选项都有一个默认值。除非需要，否则无需在拨打电话之前配置任何选项StartTaskExecution。

您还可以在每次执行任务时覆盖任务选项。例如，您可能希望为单次执行调整 LogLevel。

类型：Options 对象

是否必需：是

TaskExecutionArn

指定要更新的任务执行的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}/execution/exec-[0-9a-f]{17}$

是否必需：是

响应元素

如果此操作成功，则该服务会发送回带有空 HTTP 正文的 HTTP 200 响应。

错误

有关所有操作的常见错误信息，请参阅 常见错误类型。

InternalException

当 AWS DataSync 服务中发生错误时，会引发此异常。

HTTP 状态代码：500

InvalidRequestException

当客户端提交格式错误的请求时，会引发此异常。

HTTP 状态代码：400

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 命令行界面 V2
- AWS 适用于.NET 的 SDK V4
- AWS 适用于 C++ 的 SDK
- AWS 适用于 Go v2 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS JavaScript V3 版 SDK
- AWS 适用于 Kotlin 的 SDK
- AWS 适用于 PHP 的 SDK V3
- AWS Python 软件开发工具包
- AWS 适用于 Ruby V3 的 SDK

数据类型

支持以下数据类型：

- AgentListEntry
- AzureBlobSasConfiguration
- CmkSecretConfig
- CustomSecretConfig
- Ec2Config
- FilterRule
- FsxProtocol
- FsxProtocolNfs
- FsxProtocolSmb

- FsxUpdateProtocol
- FsxUpdateProtocolSmb
- HdfsNameNode
- LocationFilter
- LocationListEntry
- ManagedSecretConfig
- ManifestConfig
- NfsMountOptions
- OnPremConfig
- Options
- Platform
- PrivateLinkConfig
- QopConfiguration
- ReportDestination
- ReportDestinationS3
- ReportOverride
- ReportOverrides
- ReportResult
- S3Config
- S3ManifestConfig
- SmbMountOptions
- SourceManifestConfig
- TagListEntry
- TaskExecutionFilesFailedDetail
- TaskExecutionFilesListedDetail
- TaskExecutionFoldersFailedDetail
- TaskExecutionFoldersListedDetail
- TaskExecutionListEntry
- TaskExecutionResultDetail
- TaskFilter

- TaskListEntry
- TaskReportConfig
- TaskSchedule
- TaskScheduleDetails

AgentListEntry

表示调用ListAgents操作时 AWS DataSync 代理列表（或数组）中的单个条目。

内容

AgentArn

代理的亚马逊资源名称 (ARN)。 DataSync

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：否

Name

代理的名称。

类型：字符串

长度约束：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

必需：否

Platform

有关代理的平台相关详细信息，例如版本号。

类型：Platform 对象

必需：否

Status

代理的状态。

- 如果状态为 `ONLINE`，则表示已正确配置代理并且可供使用。
- 如果状态为`OFFLINE`，则表示代理已与 DataSync 该代理失去联系五分钟或更长时间。这可能是由于多种原因造成。有关更多信息，请参阅如何解决代理离线问题？

类型：字符串

有效值：`ONLINE | OFFLINE`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

AzureBlobSasConfiguration

允许 AWS DataSync 访问 Microsoft Azure Blob Storage 的共享访问签名 (SAS) 配置。

有关更多信息，请参阅用于访问 Azure Blob Storage 的 SAS 令牌。

内容

Token

指定提供访问 Azure Blob Storage 权限的 SAS 令牌。

该令牌是 SAS URI 字符串的一部分，位于存储资源 URI 和问号之后。令牌如下所示：

```
sp=r&st=2023-12-20T14:54:52Z&se=2023-12-20T22:54:52Z&spr=https&sv=2021-06-08&s
%2FXTI9E%2F%2Fmq171%2BZU178wcwqU%3D
```

类型：字符串

长度限制：长度下限为 1。最大长度为 255。

模式：`^.+$`

必需：是

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

CmkSecretConfig

指定由客户 DataSync管理的密钥的配置信息，例如身份验证令牌、密钥、密码或 DataSync 用于访问特定存储位置的 Kerberos 密钥表。 AWS KMS key

> **Note**
>
> 可使用 `CmkSecretConfig` 或 `CustomSecretConfig` 为 `CreateLocation` 请求提供凭证。不要为同一个请求提供这两个参数。

内容

KmsKeyArn

为客户管理的密钥指定用于加密存储 AWS KMS key 的 DataSync 托管密钥的 DataSync ARN。`SecretArn` DataSync 将此密钥提供给 AWS Secrets Manager。

类型：字符串

长度约束：最大长度为 2048。

模式：`^(arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):kms:[a-z\-0-9]+:[0-9]{12}:key/.*|)$`

必需：否

SecretArn

指定用于访问特定存储位置的 DataSync托管 AWS Secrets Manager 密钥的 ARN。此属性由生成 DataSync 且为只读属性。 DataSync 使用您为其指定的 KMS 密钥加密此密钥。`KmsKeyArn`

类型：字符串

长度约束：最大长度为 2048。

模式：`^(arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):secretsmanager:[a-z\-0-9]+:[0-9]{12}:secret:.*|)$`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

CustomSecretConfig

指定客户管理的 Secrets Manager 密钥的配置信息，其中存储位置凭据以纯文本（用于身份验证令牌、密钥或密码）或二进制（对于 Kerberos 密钥表）存储在 Secrets Manager 中。此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。

> **Note**
>
> 可使用 `CmkSecretConfig` 或 `CustomSecretConfig` 为 `CreateLocation` 请求提供凭证。不要为同一个请求提供这两个参数。

内容

SecretAccessRoleArn

为用于访问所指定密钥的 AWS Identity and Access Management 角色指定 ARN。 DataSync `SecretArn`

类型：字符串

长度约束：最大长度为 2048。

模式：`^(arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):iam::[0-9]{12}:role/[a-zA-Z0-9+=,.@_-]+|)$`

必需：否

SecretArn

指定密钥的 ARN。 AWS Secrets Manager

类型：字符串

长度约束：最大长度为 2048。

模式：`^(arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):secretsmanager:[a-z\-0-9]+:[0-9]{12}:secret:.*|)$`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

Ec2Config

AWS DataSync 用于连接您的 Amazon EFS 文件系统的挂载目标之一的子网和安全组。

内容

SecurityGroupArns

指定与 Amazon EFS 文件系统的挂载目标关联的安全组的 Amazon 资源名称 (ARNs)。

类型：字符串数组

数组成员：最少 1 个物品。最多 5 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

是否必需：是

SubnetArn

指定子网的 ARN，在该子网中 DataSync 创建用于在传输期间管理流量的网络接口。

子网必须位于：

- 与 Amazon EFS 文件系统位于同一虚拟私有云（VPC）。
- 至少与 Amazon EFS 文件系统的一个挂载目标位于同一个可用区。

> **Note**
>
> 您无需指定包含文件系统挂载目标的子网。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:subnet/subnet-[a-f0-9]+$`

必需：是

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

FilterRule

指定在将文件从源传输到目标时要包括或排除哪些文件、文件夹和对象。

内容

FilterType

要应用的筛选规则的类型。AWS DataSync 仅支持 SIMPLE_PATTERN 规则类型。

类型：字符串

长度限制：最大长度为 128。

模式：`^[A-Z0-9_]+$`

有效值：`SIMPLE_PATTERN`

必需：否

Value

由要包含或排除的模式组成的单个筛选条件字符串。模式用 "|"（即管道）分隔，例如：`/folder1|/folder2`

类型：字符串

长度约束：最大长度为 102400。

模式：`^[^\x00]+$`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

FsxProtocol

指定 AWS DataSync 用于访问您的 Amazon FSx 文件系统的数据传输协议。

内容

NFS

指定 DataSync 用于访问 FSx for OpenZFS 文件系统或 FSx for ONTAP 文件系统的存储虚拟机 (SVM) 的网络文件系统 (NFS) 协议配置。

类型：FsxProtocolNfs 对象

必需：否

SMB

指定 DataSync 用来访问 FSx for ONTAP 文件系统 SVM 的服务器消息块（SMB）协议配置。

类型：FsxProtocolSmb 对象

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

FsxProtocolNfs

指定 DataSync 用于访问 FSx for OpenZFS 文件系统或 FSx for ONTAP 文件系统的存储虚拟机 (SVM) 的网络文件系统 (NFS) 协议配置。

内容

MountOptions

指定 DataSync 如何使用 NFS 协议访问某个位置。

类型：NfsMountOptions 对象

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

FsxProtocolSmb

指定用于访问您的 Amazon FSx for NetApp ONTAP 文件系统的存储虚拟机 (SVM) 的服务器消息块 (SMB) 协议配置。 AWS DataSync 有关更多信息，请参阅提供 DataSync 对 ONTAP 文件系统的访问权限。 FSx

内容

User

指定可以挂载和访问 SVM 中文件、文件夹和元数据的用户。

有关如何为传输选择具有适当访问级别的用户，请参阅使用 SMB 协议。

类型：字符串

长度约束：最大长度为 104。

模式：`^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$`

是否必需：是

CmkSecretConfig

指定由客户 DataSync管理的密钥的配置信息，其中包括 FSx 用于访问特定的 ONTAP 存储位置（使用 SMB）的密码。 DataSync AWS KMS key

当您在`CreateLocationFsxOntap`请求中包含此参数时，您仅提供 KMS 密钥 ARN。 DataSync 使用此 KMS 密钥和`Password`您为指定的密钥来创建存储位置访问凭证的 DataSync托管密钥。

确保该用户 DataSync 有权访问您指定的 KMS 密钥。有关更多信息，请参阅使用使用自定义 AWS KMS key密钥加密的服务托管密钥。

> **Note**
>
> 可使用 `CmkSecretConfig`（带有 `Password`）或 `CustomSecretConfig`（不带 `Password`）为 `CreateLocationFsxOntap` 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的 Secr FSx ets Manager 密钥的配置信息，其中 ONTAP 存储位置（使用 SMB）的密码以纯文本形式存储在 Secrets Manager 中。此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。有关更多信息，请参阅使用您管理的密钥。

> **Note**
>
> 可使用 `CmkSecretConfig`（带有 `Password`）或 `CustomSecretConfig`（不带 `Password`）为 `CreateLocationFsxOntap` 请求提供凭证。不要为同一个请求提供这两个参数。

类型：CustomSecretConfig 对象

必需：否

Domain

指定存储虚拟机（SVM）所属的 Windows 域的名称。

如果您的环境中有多个域，则配置此设置 DataSync 可确保连接到正确的 SVM。

如果您的环境中有多个 Active Directory 域，则配置此参数可确保 DataSync 它们连接到正确的 SVM。

类型：字符串

长度约束：最大长度为 253。

模式：`^[A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252}$`

必需：否

ManagedSecretConfig

描述 DataSync托管密钥的配置信息，例如 DataSync 用于访问特定存储位置的。`Password` DataSync 使用默认 AWS托管的 KMS 密钥在中 AWS Secrets Manager加密此密钥。

> **Note**
>
> 请勿在`CreateLocation`请求时提供此信息。 `ManagedSecretConfig`是一个 ReadOnly 属性，仅在`DescribeLocation`响应中填充。

类型：ManagedSecretConfig 对象

必需：否

MountOptions

指定用于访问 SMB 文件服务器的服务器消息块 (SMB) 协议的版本。 AWS DataSync

类型：SmbMountOptions 对象

必需：否

Password

指定有权访问您的 SVM 的用户的密码。

类型：字符串

长度约束：最大长度为 104。

模式：`^.{0,104}$`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

FsxUpdateProtocol

指定 AWS DataSync 用于访问您的 Amazon FSx 文件系统的数据传输协议。

> **Note**
>
> FSx for ONTAP 位置的网络文件系统（NFS）协议配置无法更新。DataSync 目前仅支持此位置类型的 NFS 版本 3。

内容

NFS

指定 DataSync 用于访问 FSx for OpenZFS 文件系统或 FSx for ONTAP 文件系统的存储虚拟机 (SVM) 的网络文件系统 (NFS) 协议配置。

类型：FsxProtocolNfs 对象

必需：否

SMB

指定 DataSync 用来访问 FSx for ONTAP 文件系统的存储虚拟机（SVM）的服务器消息块（SMB）协议配置。

类型：FsxUpdateProtocolSmb 对象

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

FsxUpdateProtocolSmb

指定用于访问您的 Amazon FSx for NetApp ONTAP 文件系统的存储虚拟机 (SVM) 的服务器消息块 (SMB) 协议配置。 AWS DataSync 有关更多信息，请参阅提供 DataSync 对 ONTAP 文件系统的访问权限。 FSx

内容

CmkSecretConfig

指定 DataSync托管密钥的配置信息，例如用于访问特定传输位置和客户 AWS KMS key管理的 DataSync 凭据的一组Password或一组凭据。

类型：CmkSecretConfig 对象

必需：否

CustomSecretConfig

指定客户管理的密钥的配置信息，例如用于访问特定传输位置的一组Password或一组凭据。 DataSync 此配置包括密钥 ARN，以及提供密钥访问权限的 IAM 角色的 ARN。

类型：CustomSecretConfig 对象

必需：否

Domain

指定存储虚拟机（SVM）所属的 Windows 域的名称。

如果您的环境中有多个 Active Directory 域，则配置此参数可确保 DataSync 它们连接到正确的 SVM。

类型：字符串

长度约束：最大长度为 253。

模式：^([A-Za-z0-9]((\.|-+)?[A-Za-z0-9]){0,252})?$

必需：否

MountOptions

指定用于访问 SMB 文件服务器的服务器消息块 (SMB) 协议的版本。 AWS DataSync

类型：SmbMountOptions 对象

必需：否

Password

指定有权访问您的 SVM 的用户的密码。

类型：字符串

长度约束：最大长度为 104。

模式：^.{0,104}$

必需：否

User

指定可以挂载和访问 SVM 中文件、文件夹和元数据的用户。

有关如何为传输选择具有适当访问级别的用户，请参阅使用 SMB 协议。

类型：字符串

长度约束：最大长度为 104。

模式：^[^\x22\x5B\x5D/\\:;|=,+*?\x3C\x3E]{1,104}$

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

HdfsNameNode

Hadoop Distributed File System（HDFS）的 NameNode。NameNode 管理文件系统的命名空间。NameNode 执行打开、关闭和重命名文件和目录等操作。NameNode 包含将数据块映射到 DataNodes 的信息。

内容

Hostname

HDFS 集群中 NameNode 的主机名。它是 NameNode 的 IP 地址或域名服务（DNS）名称。在本地安装的代理使用此主机名与网络中的 NameNode 通信。

类型：字符串

长度限制：长度下限为 1。最大长度为 255。

模式：`^(([a-zA-Z0-9\-]*[a-zA-Z0-9])\.)*([A-Za-z0-9\-]*[A-Za-z0-9])$`

必需：是

Port

NameNode 用来侦听客户端请求的端口。

类型：整数

有效范围：最小值为 1。最大值为 65536。

必需：是

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

LocationFilter

缩小 ListLocations 返回的资源列表范围。例如，要查看您的所有 Amazon S3 地点，请使用 "Name": "LocationType"、"Operator": "Equals" 和 "Values": "S3" 创建筛选器。

有关更多信息，请参阅筛选资源。

内容

Name

正在使用的筛选器名称。每个 API 调用都支持可用于其的筛选器列表（例如，ListLocations 的 LocationType）。

类型：字符串

有效值：LocationUri | LocationType | CreationTime

必需：是

Operator

用于比较筛选值的运算符（例如，Equals 或 Contains）。

类型：字符串

有效值：Equals | NotEquals | In | LessThanOrEqual | LessThan | GreaterThanOrEqual | GreaterThan | Contains | NotContains | BeginsWith

必需：是

Values

要筛选的值。例如，您可能希望仅显示 Amazon S3 位置。

类型：字符串数组

长度约束：最小长度为 1。最大长度为 255。

模式：^[0-9a-zA-Z_\ \-\:*\.\\/\?-]*$

必需：是

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

LocationListEntry

表示位置列表中的单个条目。 `LocationListEntry`返回一个包含调用该ListLocations操作时位置列表的数组。

内容

LocationArn

位置的 Amazon 资源名称（ARN）。对于网络文件系统 (NFS) 或 Amazon EFS，位置为导出路径。对于 Amazon S3，位置是要挂载的并用作位置根的前缀路径。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:location/loc-[0-9a-z]{17}$`

必需：否

LocationUri

表示营业地点 URIs 的列表。 `LocationUri`返回一个包含调用该ListLocations操作时位置列表的数组。

格式：`TYPE://GLOBAL_ID/SUBDIR`。

TYPE 指定位置的类型（例如，`nfs` 或 `s3`）。

GLOBAL_ID 是支持该位置的资源的全局唯一标识符。EFS 的一个例子是 `us-east-2.fs-abcd1234`。Amazon S3 的一个例子是存储桶名称，例如 `myBucket`。NFS 的示例是符合 DNS 的有效 IPv4 或 IPv6 地址或主机名。

SUBDIR 是一个有效的文件系统路径，按照 *nix 惯例，由正斜杠分隔。对于 NFS 和 Amazon EFS，这是挂载位置的导出路径。对于 Amazon S3，它是要挂载的并用作位置根的前缀路径。

类型：字符串

长度约束：最大长度为 4360。

模式：`^(efs|nfs|s3|smb|hdfs|fsx[a-z0-9-]+)://[a-zA-Z0-9.:/\-]+$`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

ManagedSecretConfig

指定 DataSync托管密钥的配置信息，例如用于访问特定传输位置的身份验证令牌或一组凭据。DataSync DataSync 使用默认 AWS托管的 KMS 密钥在中 AWS Secrets Manager加密此密钥。

内容

SecretArn

指定密钥的 ARN。 AWS Secrets Manager

类型：字符串

长度约束：最大长度为 2048。

模式：`^(arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):secretsmanager:[a-z\-0-9]+:[0-9]{12}:secret:.*|)$`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

ManifestConfig

配置清单，即您希望 AWS DataSync 传输的文件或对象的列表。有关详细信息和配置示例，请参阅使用清单指定 DataSync 传输的内容。

内容

Action

指定 DataSync 使用清单的用途。

类型：字符串

有效值：TRANSFER

必需：否

Format

指定清单的文件格式。有关更多信息，请参阅创建清单。

类型：字符串

有效值：CSV

必需：否

Source

指定您希望 DataSync 使用的清单及其托管位置。

> **Note**
>
> 如果您在 2024 年 2 月 7 日或之后配置新清单，则必须指定此参数。否则会出现 400 状态代码和 ValidationException 错误，说明缺少 DataSync 的 IAM 角色，无法访问托管清单的 S3 存储桶。有关更多信息，请参阅向 DataSync 提供对清单的访问权限。

类型：SourceManifestConfig 对象

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

NfsMountOptions

指定 DataSync 如何使用 NFS 协议访问某个位置。

内容

Version

指定您希望 DataSync 在挂载 NFS 共享时使用的 NFS 版本。如果服务器拒绝使用指定的版本，则任务将失败。

可以指定以下选项：

- `AUTOMATIC`（默认）：DataSync 选择 NFS 版本 4.1。
- `NFS3`：允许在服务器上进行异步写入的无状态协议版本。
- `NFSv4_0`：支持委托和伪文件系统的有状态、防火墙友好的协议版本。
- `NFSv4_1`：支持会话、目录委托和并行数据处理的有状态协议版本。NFS 4.1 版还包括 4.0 版中提供的所有功能。

> **Note**
>
> DataSync 目前仅支持 Amazon FSx for NetApp ONTAP 位置的 NFS 版本 3。

类型：字符串

有效值：`AUTOMATIC | NFS3 | NFS4_0 | NFS4_1`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

OnPremConfig

可以连接到您的网络文件系统 (NFS) 文件服务器的 AWS DataSync 代理。

内容

AgentArns

可以连接到您的 NFS 文件服务器的 DataSync 代理的 Amazon 资源名称 (ARNs)。

您可以指定多个代理。有关更多信息，请参阅使用多个 DataSync 代理。

类型：字符串数组

数组成员：最少 1 个物品。最多 8 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:agent/agent-[0-9a-z]{17}$`

必需：是

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

Options

指明您的传输任务是如何配置的。这些选项包括在传输过程中如何 DataSync 处理文件、对象及其关联的元数据。您还可以指定如何验证数据完整性、为任务设置带宽限制等其他选项。

每个选项都有一个默认值。除非需要，否则无需在拨打电话之前配置任何选项StartTaskExecution。

您还可以在每次执行任务时覆盖任务选项。例如，您可能希望为单次执行调整 LogLevel。

内容

Atime

指定是否保留表示上次文件被读取或写入的时间的元数据。

> **Note**
>
> 的行为在各个平台上Atime并不完全标准，因此 DataSync 只能在尽力而为的基础上做到这一点。

- BEST_EFFORT（默认）- DataSync 尝试保留所有源文件的原始Atime属性（即任务执行PREPARING步骤之前的版本）。建议使用该选项。
- NONE：忽略 Atime。

> **Note**
>
> 如果 Atime 被设置为 BEST_EFFORT，则 Mtime 必须被设置为 PRESERVE。
> 如果 Atime 设置为 NONE，则 Mtime 也必须为 NONE。

类型：字符串

有效值：NONE | BEST_EFFORT

必需：否

BytesPerSecond

限制 DataSync 任务使用的带宽。例如，如果您 DataSync 想使用最大 1 MB 的空间，请将此值设置为 1048576 (=1024*1024)。

类型：长整型

有效范围：最小值为 -1。

必需：否

Gid

指定文件拥有者的 POSIX 组 ID (GID)。

- INT_VALUE（默认）：保留用户 ID（UID）和 GID 的整数值，建议使用。
- NONE：忽略 UID 和 GID。

有关更多信息，请参阅了解如何 DataSync 处理文件和对象元数据。

类型：字符串

有效值：NONE | INT_VALUE | NAME | BOTH

必需：否

LogLevel

指定 DataSync 发布到 Amazon CloudWatch 日志组的日志类型。要指定日志组，请参阅CloudWatchLogGroupArn。

- BASIC：仅发布包含基本信息（如传输错误）的日志。
- TRANSFER-发布 DataSync 任务传输的所有文件或对象的日志，并对其执行数据完整性检查。
- OFF：未发布任何日志。

类型：字符串

有效值：OFF | BASIC | TRANSFER

必需：否

Mtime

指定是否保留表示上次文件在任务执行的 PREPARING 步骤之前被读取或写入时的元数据。当您需要多次运行任务时，必须使用此选项。

- PRESERVE（默认）：保留原始 Mtime，建议使用。
- NONE：忽略 Mtime。

> **Note**
>
> 如果 Mtime 被设置为 PRESERVE，则 Atime 必须被设置为 BEST_EFFORT。
> 如果 Mtime 被设置为 NONE，则 Atime 也必须被设置为 NONE。

类型：字符串

有效值：NONE | PRESERVE

必需：否

ObjectTags

指定在PRESERVE对象存储系统之间传输时是否 DataSync 要对标签进行对象化（默认行为）。如果您希望 DataSync 任务忽略对象标签，请指定该NONE值。

类型：字符串

有效值：PRESERVE | NONE

必需：否

OverwriteMode

指定是 DataSync 应该修改还是保留目标位置的数据。

- ALWAYS（默认）-当源数据（包括元数据）发生变化时， DataSync 修改目标位置的数据。

  如果 DataSync 覆盖了对象，则某些 Amazon S3 存储类别（例如，检索或提前删除）可能会产生额外费用。有关更多信息，请参阅 Amazon S3 传输的存储类别注意事项。
- NEVER-即使源数据已更改，也 DataSync 不会覆盖目标位置的数据。您可以使用此选项防止覆盖对目标位置中文件或对象所做的更改。

类型：字符串

有效值：ALWAYS | NEVER

必需：否

PosixPermissions

用于确定哪些用户或组可以出于特定目的（例如读取、写入或执行文件）访问文件。

有关更多信息，请参阅[了解如何 DataSync 处理文件和对象元数据]。

- PRESERVE（默认）：保留 POSIX 风格权限，建议使用。
- NONE：忽略 POSIX 风格权限。

> **Note**
>
> DataSync 可以保留源位置的现有权限。

类型：字符串

有效值：NONE | PRESERVE

必需：否

PreserveDeletedFiles

指定是否应保留源中不存在的目标位置的文件。此选项可能会影响您的 Amazon S3 存储成本。如果您的任务删除对象，则您可能需要为某些存储类别支付最低存储期限费用。有关详细信息，请参阅中[使用 Amazon S3 存储类时的注意事项 DataSync]。

- PRESERVE（默认）：忽略此类目标文件，建议使用。
- REMOVE：删除源中不存在的目标文件。

> **Note**
>
> 如果您将此参数设置为 REMOVE，则无法将 TransferMode 设置为 ALL。当您传输所有数据时， DataSync 不会扫描您的目标位置，也不知道要删除什么。

类型：字符串

有效值：PRESERVE | REMOVE

必需：否

PreserveDevices

指定是否 DataSync 应在源位置保留块和字符设备的元数据，并在目标位置使用该设备名称和元数据重新创建文件。 DataSync 仅复制此类设备的名称和元数据。

> **Note**
>
> DataSync 无法复制这些设备的实际内容，因为它们是非终端设备且不返回 end-of-file (EOF) 标记。

- NONE（默认）：忽略特殊设备（推荐）。
- PRESERVE：保留字符和块设备元数据。Amazon EFS 目前不支持此选项。

类型：字符串

有效值：NONE | PRESERVE

必需：否

SecurityDescriptorCopyFlags

确定将 SMB 安全描述符的哪些组件是从源复制到目标对象的。

此值仅用于 Windows 文件服务器位置的 SMB 和 Amazon FSx 之间的传输，或者 Windows 文件服务器位置 FSx 的两个位置之间的传输。有关更多信息，请参阅了解如何 DataSync 处理文件和对象元数据。

- OWNER_DACL（默认）-对于每个复制的对象， DataSync 复制以下元数据：
  - 对象所有者。
  - NTFS 自由访问控制列表 (DACLs)，用于确定是否授予对对象的访问权限。

    DataSync 不会使用此选项复制 NTFS 系统访问控制列表 (SACLs)。
- OWNER_DACL_SACL-对于每个复制的对象， DataSync 复制以下元数据：
  - 对象所有者。
  - NTFS 自由访问控制列表 (DACLs)，用于确定是否授予对对象的访问权限。
  - SACLs，管理员使用它们来记录访问安全对象的尝试。

    复制 SACLs 需要向用于访问您的 SMB 位置的 DataSync Windows 用户授予其他权限。有关选择具有适当权限的用户的信息，请参阅 SMB、FSx Windows 文件服务器或 FSx ONTAP 所需的权限（取决于传输中的位置类型）。
- NONE：不会复制任何 SMB 安全描述符组件。目标对象归用于访问目标位置的用户所有。DACLs 并 SACLs 根据目标服务器的配置进行设置。

类型：字符串

有效值：NONE | OWNER_DACL | OWNER_DACL_SACL

必需：否

TaskQueueing

指定在运行多个任务的特定情形下，是否应该将传输任务放入一个队列。默认情况下，这是 ENABLED。

类型：字符串

有效值：ENABLED | DISABLED

必需：否

TransferMode

指定是仅 DataSync 传输初始副本后不同位置的数据（包括元数据），还是每次运行任务时传输所有数据。如果计划进行定期传输，则可能仅需要传输自上次执行任务以来发生更改的内容。

- CHANGED（默认）-初次完全传输后，仅 DataSync复制源位置和目标位置不同的数据和元数据。
- ALL- DataSync 将源中的所有内容复制到目的地，而不比较位置之间的差异。

类型：字符串

有效值：CHANGED | ALL

必需：否

Uid

指定文件拥有者的 POSIX 用户 ID (UID)。

- INT_VALUE（默认）：保留 UID 和组 ID（GID）的整数值，推荐使用。
- NONE：忽略 UID 和 GID。

有关更多信息，请参阅复制的元数据 DataSync。

类型：字符串

有效值：NONE | INT_VALUE | NAME | BOTH

必需：否

VerifyMode

指定在传输结束时是否以及如何 DataSync 检查数据的完整性。

- ONLY_FILES_TRANSFERRED（推荐）- DataSync 计算源位置传输的数据（包括元数据）的校验和。在传输结束时，将 DataSync 此校验和与在目的地根据该数据计算出的校验和进行比较。

  > **Note**
  >
  > 此为增强模式任务的默认选项。

  在传输到 S3 Glacier Flexible Retrieval 或 S3 Glacier Deep Archive 存储类时，推荐使用此选项。有关更多信息，请参阅 Amazon S3 位置的存储类考虑因素。
- POINT_IN_TIME_CONSISTENT-传输结束时， DataSync 检查整个源和目的地，以验证两个位置是否完全同步。

  > **Note**
  >
  > 此为基本模式任务的默认选项，目前增强模式任务不支持该选项。

  如果您使用清单，则 DataSync 仅扫描和验证清单中列出的内容。

  在传输到 S3 Glacier Flexible Retrieval 或 S3 Glacier Deep Archive 存储类时，不支持此选项。有关更多信息，请参阅 Amazon S3 位置的存储类考虑因素。
- NONE-仅在传输期间 DataSync 执行数据完整性检查。与其他选项不同，在传输结束时没有额外验证。

类型：字符串

有效值：`POINT_IN_TIME_CONSISTENT | ONLY_FILES_TRANSFERRED | NONE`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

Platform

有关 AWS DataSync 代理的平台相关详细信息，例如版本号。

内容

Version

DataSync 代理的版本。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

PrivateLinkConfig

指定您的 AWS DataSync 代理如何 AWS 使用虚拟私有云 (VPC) 服务端点进行连接。使用 VPC 端点的代理无法通过公共互联网访问。

内容

PrivateLinkEndpoint

指定您的代理所提供AWS PrivateLink的 VPC 端点 。

类型：字符串

长度限制：最小长度为 7。最大长度为 15。

模式：`\A(25[0-5]|2[0-4]\d|[0-1]?\d?\d)(\.(25[0-5]|2[0-4]\d|[0-1]?\d?\d)){3}\z`

必需：否

SecurityGroupArns

指定提供对您的 VPC 终端节点 DataSync 访问权限的安全组的 Amazon 资源名称 (ARN)。您只能指定一个 ARN。

类型：字符串数组

数组成员：固定数量为 1 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:security-group/sg-[a-f0-9]+$`

必需：否

SubnetArns

指定您的 VPC 端点 所在子网的 ARN。您只能指定一个 ARN。

类型：字符串数组

数组成员：固定数量为 1 项。

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):ec2:[a-z\-0-9]*:[0-9]{12}:subnet/subnet-[a-f0-9]+$`

必需：否

VpcEndpointId

指定您的代理连接到的 VPC 端点的 ID。

类型：字符串

模式：`^vpce-[0-9a-f]{17}$`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

QopConfiguration

保护质量（QOP）配置指定在 Hadoop Distributed File System（HDFS）集群上配置的远程程序调用（RPC）和数据传输隐私设置。

内容

DataTransferProtection

在 HDFS 集群上配置的数据传输保护设置。此设置对应于 Hadoop 集群上的 `hdfs-site.xml` 文件中的 `dfs.data.transfer.protection` 设置。

类型：字符串

有效值：`DISABLED | AUTHENTICATION | INTEGRITY | PRIVACY`

必需：否

RpcProtection

在 HDFS 集群上配置的 RPC 保护设置。此设置对应于您在 Hadoop 集群上的 `core-site.xml` 文件中的 `hadoop.rpc.protection` 设置。

类型：字符串

有效值：`DISABLED | AUTHENTICATION | INTEGRITY | PRIVACY`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

ReportDestination

指定 DataSync 上传任务报告的位置。

内容

S3

指定 DataSync 上传任务报告的 Amazon S3 存储桶位置。

类型：ReportDestinationS3 对象

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

ReportDestinationS3

指定用于 DataSync 上传任务报告的 Amazon S3 存储桶。

内容

BucketAccessRoleArn

指定允许将任务报告上传 DataSync 到您的 S3 存储桶的 IAM 策略的 Amazon 资源名称 (ARN)。有关更多信息，请参阅允许将任务报告上传 DataSync 到 Amazon S3 存储桶。

类型：字符串

长度约束：最大长度为 2048。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):iam::[0-9]{12}:role/.*$`

是否必需：是

S3BucketArn

指定 DataSync 上传报告的 S3 存储桶的 ARN。

类型：字符串

长度限制：最大长度为 268。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3:[a-z\-0-9]*:[0-9]{12}:accesspoint[/:][a-zA-Z0-9\-.]{1,63}$|^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3-outposts:[a-z\-0-9]+:[0-9]{12}:outpost[/:][a-zA-Z0-9\-]{1,63}[/:]accesspoint[/:][a-zA-Z0-9\-]{1,63}$|^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3:::[a-zA-Z0-9.\-_]{1,255}$`

是否必需：是

Subdirectory

为您的报告指定存储桶前缀。

类型：字符串

长度约束：最大长度为 4096。

模式：^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

ReportOverride

指定 DataSync 任务报告特定方面的详细程度。

内容

ReportLevel

指定您的任务报告是仅包含错误还是同时包含成功和错误。

例如，您的报告可能主要只包含传输中不顺利的部分（ERRORS_ONLY）。同时，您要验证您的任务筛选器是否正常运行。在这种情况下，您可以获得 DataSync 成功跳过的文件列表，并了解是否传输了您未传输的内容（SUCCESSES_AND_ERRORS）。

类型：字符串

有效值：ERRORS_ONLY | SUCCESSES_AND_ERRORS

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

ReportOverrides

DataSync 任务报告各个方面的详细信息级别。

内容

Deleted

指定 DataSync 尝试在目标位置删除的文件、对象和目录的报告级别。这仅适用于您将任务配置为在目标位置中删除不存在于源位置的数据。

类型：ReportOverride 对象

必需：否

Skipped

指定 DataSync 在传输过程中尝试跳过的文件、对象和目录的报告级别。

类型：ReportOverride 对象

必需：否

Transferred

指定 DataSync 尝试传输的文件、对象和目录的报告级别。

类型：ReportOverride 对象

必需：否

Verified

指定 DataSync 在传输结束时尝试验证的文件、对象和目录的报告级别。

类型：ReportOverride 对象

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK

- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

ReportResult

指出 DataSync 是否为传输创建了完整的任务报告。

内容

ErrorCode

指出在 DataSync 无法创建完整报告时与该错误关联的代码。

类型：字符串

必需：否

ErrorDetail

提供有关创建报告时遇到的问题的详细信息。

类型：字符串

必需：否

Status

指出 DataSync 是仍在处理您的报告、创建报告还是无法创建完整的报告。

类型：字符串

有效值：`PENDING | SUCCESS | ERROR`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

S3Config

指定用于访问您的 S3 存储桶的 AWS Identity and Access Management (IAM) 角色的 DataSync 亚马逊资源名称 (ARN)。

有关更多信息，请参阅提供 DataSync 对 S3 存储桶的访问权限。

内容

BucketAccessRoleArn

指定用于访问您的 S3 存储桶的 IAM 角色的 ARN。 DataSync

类型：字符串

长度约束：最大长度为 2048。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):iam::[0-9]{12}:role/.*$`

必需：是

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

S3ManifestConfig

指定托管要使用的清单的 S3 存储桶。 AWS DataSync 有关更多信息和配置示例，请参阅使用清单指定 DataSync 传输的内容。

内容

BucketAccessRoleArn

指定允许 DataSync 访问清单的 AWS Identity and Access Management (IAM) 角色。有关更多信息，请参阅提供对清单的 DataSync 访问权限。

类型：字符串

长度约束：最大长度为 2048。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):iam::[0-9]{12}:role/.*$`

是否必需：是

ManifestObjectPath

指定清单的 Amazon S3 对象键。其中可包含前缀（例如，`prefix/my-manifest.csv`）。

类型：字符串

长度约束：最大长度为 4096。

模式：`^[a-zA-Z0-9_\-\+\./\(\)\p{Zs}]*$`

是否必需：是

S3BucketArn

指定要在其中托管清单的 S3 存储桶的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 268。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3:[a-z\-0-9]*:[0-9]{12}:accesspoint[/:][a-zA-Z0-9\-.]{1,63}$|^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3-outposts:[a-z\-0-9]+:[0-9]`

```
{12}:outpost[/:][a-zA-Z0-9\-]{1,63}[/:]accesspoint[/:][a-zA-Z0-9\-]
{1,63}$|^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):s3:::[a-
zA-Z0-9.\-_]{1,255}$
```

是否必需：是

ManifestObjectVersionId

指定要使用的清单的对象版本 ID。 DataSync 如果您未设置此项，则 DataSync 使用该对象的最新版本。

类型：字符串

长度约束：最小长度为 1。最大长度为 100。

模式：`^.+$`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

SmbMountOptions

指定 AWS DataSync 用于访问服务器消息块（SMB）文件服务器的 SMB 协议版本。

内容

Version

默认情况下，DataSync 会根据与 SMB 文件服务器的协商自动选择 SMB 协议版本。您也可以将 DataSync 配置为使用特定的 SMB 版本，但我们建议仅在 DataSync 无法自动与 SMB 文件服务器协商的情况下才这样做。

以下是用于配置 SMB 版本的选项：

- `AUTOMATIC`（默认）：DataSync 和 SMB 文件服务器协商它们在 2.1 和 3.1.1 之间相互支持的最高版本 SMB。

  这是建议的选项。如果您改为选择文件服务器不支持的特定版本，则可能会出现 `Operation Not Supported` 错误。
- `SMB3`：将协议协商限制为仅 SMB 版本 3.0.2。
- `SMB2`：将协议协商限制为仅 SMB 版本 2.1。
- `SMB2_0`：将协议协商限制为仅 SMB 版本 2.0。
- `SMB1`：将协议协商限制为仅 SMB 版本 1.0。

> **Note**
>
> 创建 Amazon FSx for NetApp ONTAP 位置时，该 `SMB1` 选项不可用。

类型：字符串

有效值：`AUTOMATIC | SMB2 | SMB3 | SMB1 | SMB2_0`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK

- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

SourceManifestConfig

指定您希望 AWS DataSync 使用的清单及其托管位置。有关详细信息和配置示例，请参阅使用清单指定 DataSync 传输的内容。

内容

S3

指定在其中托管清单的 S3 存储桶。

类型：S3ManifestConfig 对象

必需：是

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

TagListEntry

代表已应用于 AWS 资源的单个标签的键值对。

内容

Key

AWS 资源标签的键。

类型：字符串

长度限制：最小长度为 1。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:/-]+$`

必需：是

Value

AWS 资源标签的值。

类型：字符串

长度约束：最小长度为 0。最大长度为 256。

模式：`^[a-zA-Z0-9\s+=._:@/-]+$`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

TaskExecutionFilesFailedDetail

任务执行期间 DataSync 无法准备、传输、验证和删除的文件或对象的数量。

> **Note**
>
> 仅适用于增强模式任务。

内容

Delete

在任务执行期间 DataSync 未能删除的文件或对象的数量。

类型：长整型

必需：否

Prepare

任务执行期间 DataSync 未能准备的文件或对象的数量。

类型：长整型

必需：否

Transfer

任务执行期间 DataSync 未能传输的文件或对象的数量。

类型：长整型

必需：否

Verify

任务执行期间 DataSync 未能验证的文件或对象的数量。

类型：长整型

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

TaskExecutionFilesListedDetail

在您的位置 DataSync 找到的文件或对象的数量。

> **Note**
>
> 仅适用于增强模式任务。

内容

AtDestinationForDelete

在目标位置 DataSync 找到的文件或对象的数量。该计数器仅适用于您将任务配置为在目标位置中删除不存在于源位置的数据。

类型：长整型

必需：否

AtSource

在您的源位置 DataSync 找到的文件或对象的数量。

- 对于 DataSync清单，仅列出清单中的内容（而不是源位置的所有内容）。
- 使用包含过滤器，仅 DataSync 列出与来源位置的过滤器匹配的内容。
- 使用排除过滤器，在应用过滤器之前，会 DataSync 列出源位置的所有内容。

类型：长整型

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

TaskExecutionFoldersFailedDetail

在任务执行期间 DataSync 未能列出、准备、传输、验证和删除的目录数量。

> **Note**
>
> 仅适用于增强模式任务。

内容

Delete

在任务执行期间删除 DataSync 失败的目录数量。

类型：长整型

必需：否

List

任务执行期间 DataSync 未能列出的目录数量。

类型：长整型

必需：否

Prepare

任务执行期间 DataSync 无法准备的目录数量。

类型：长整型

必需：否

Transfer

任务执行期间 DataSync 无法传输的目录数量。

类型：长整型

必需：否

Verify

任务执行期间 DataSync 未能验证的目录数量。

类型：长整型

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

TaskExecutionFoldersListedDetail

在您的位置 DataSync 找到的目录数量。

> **Note**
>
> 仅适用于增强模式任务。

内容

AtDestinationForDelete

在您的目标位置 DataSync 找到的目录数量。该计数器仅适用于您将任务配置为在目标位置中删除不存在于源位置的数据。

类型：长整型

必需：否

AtSource

在您的源位置 DataSync 找到的目录数量。

- 对于 DataSync清单，仅列出清单中的内容（而不是源位置的所有内容）。
- 使用包含过滤器，仅 DataSync 列出与来源位置的过滤器匹配的内容。
- 使用排除过滤器，在应用过滤器之前，会 DataSync 列出源位置的所有内容。

类型：长整型

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

TaskExecutionListEntry

表示随ListTaskExecutions操作返回的 AWS DataSync 任务执行列表中的单个条目。

内容

Status

任务执行的状态。有关更多信息，请参阅任务执行状态。

类型：字符串

有效值：`QUEUED | CANCELLING | LAUNCHING | PREPARING | TRANSFERRING | VERIFYING | SUCCESS | ERROR`

必需：否

TaskExecutionArn

执行任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：`^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}/execution/exec-[0-9a-f]{17}$`

必需：否

TaskMode

当前使用的任务模式。有关更多信息，请参阅为数据传输选择任务模式。

类型：字符串

有效值：`BASIC | ENHANCED`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

TaskExecutionResultDetail

提供有关 AWS DataSync 任务执行结果的详细信息。

内容

ErrorCode

DataSync 在任务执行过程中遇到的错误。您可以使用此信息来帮助排查问题。

类型：字符串

必需：否

ErrorDetail

DataSync 在任务执行过程中遇到的错误的详细描述。您可以使用此信息来帮助排查问题。

类型：字符串

必需：否

PrepareDuration

任务执行处于 PREPARING 步骤的时间（毫秒）。有关更多信息，请参阅任务执行状态。

对于增强模式任务，此值始终为 0。有关更多信息，请参阅 DataSync 如何准备数据传输。

类型：长整型

有效范围：最小值为 0。

必需：否

PrepareStatus

任务执行 PREPARING 步骤的状态。有关更多信息，请参阅任务执行状态。

类型：字符串

有效值：PENDING | SUCCESS | ERROR

必需：否

TotalDuration

任务执行的运行时间（毫秒）。

类型：长整型

有效范围：最小值为 0。

必需：否

TransferDuration

任务执行处于 TRANSFERRING 步骤的时间（毫秒）。有关更多信息，请参阅任务执行状态。

对于增强模式任务，此值始终为 0。有关更多信息，请参阅 DataSync 如何传输数据。

类型：长整型

有效范围：最小值为 0。

必需：否

TransferStatus

任务执行 TRANSFERRING 步骤的状态。有关更多信息，请参阅任务执行状态。

类型：字符串

有效值：PENDING | SUCCESS | ERROR

必需：否

VerifyDuration

任务执行处于 VERIFYING 步骤的时间（毫秒）。有关更多信息，请参阅任务执行状态。

对于增强模式任务，此值始终为 0。有关更多信息，请参阅 DataSync 如何验证数据的完整性。

类型：长整型

有效范围：最小值为 0。

必需：否

VerifyStatus

任务执行 VERIFYING 步骤的状态。有关更多信息，请参阅任务执行状态。

类型：字符串

有效值：PENDING | SUCCESS | ERROR

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

TaskFilter

您可以使用 API 筛选器来缩小 ListTasks 返回的资源的列表范围。例如，要检索源位置上的所有任务，您可以结合该位置的 ARN 使用筛选器名称为 LocationId 和 Operator Equals 的 ListTasks。

有关更多信息，请参阅筛选 DataSync 资源。

内容

Name

正在使用的筛选器名称。每个 API 调用都支持其可用的筛选器列表。例如，ListTasks 的 LocationId。

类型：字符串

有效值：LocationId | CreationTime

必需：是

Operator

用于比较筛选值的运算符（例如，Equals 或 Contains）。

类型：字符串

有效值：Equals | NotEquals | In | LessThanOrEqual | LessThan | GreaterThanOrEqual | GreaterThan | Contains | NotContains | BeginsWith

必需：是

Values

要筛选的值。例如，您可能只想显示在特定目标位置的任务。

类型：字符串数组

长度约束：最小长度为 1。最大长度为 255。

模式：^[0-9a-zA-Z_\ \-\:*\.\\/\?-]*$

必需：是

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

TaskListEntry

表示任务列表中的单个条目。 TaskListEntry返回一个数组，其中包含调用ListTasks操作时的任务列表。任务包括要同步的源文件系统和目标文件系统，以及用于这些任务的选项。

内容

Name

任务的名称。

类型：字符串

长度约束：最小长度为 0。最大长度为 256。

模式：^[a-zA-Z0-9\s+=._:@/-]+$

必需：否

Status

任务的状态。

类型：字符串

有效值：AVAILABLE | CREATING | QUEUED | RUNNING | UNAVAILABLE

必需：否

TaskArn

任务的 Amazon 资源名称（ARN）。

类型：字符串

长度限制：最大长度为 128。

模式：^arn:(aws|aws-cn|aws-us-gov|aws-eusc|aws-iso|aws-iso-b):datasync:[a-z\-0-9]+:[0-9]{12}:task/task-[0-9a-f]{17}$

必需：否

TaskMode

当前使用的任务模式。有关更多信息，请参阅为数据传输选择任务模式。

类型：字符串

有效值：`BASIC | ENHANCED`

必需：否

另请参阅

有关以特定语言之一使用此 API 的更多信息 AWS SDKs，请参阅以下内容：

- AWS 适用于 C++ 的 SDK
- AWS 适用于 Java 的 SDK V2
- AWS 适用于 Ruby V3 的 SDK

TaskReportConfig

指定您想要如何配置任务报告，该报告可提供有关 AWS DataSync 传输的详细信息。

有关详细信息，请参阅任务报告。

内容

Destination

指定 DataSync 上传任务报告的 Amazon S3 存储桶位置。有关详细信息，请参阅任务报告。

类型：ReportDestination 对象

必需：否

ObjectVersionIds

指定您的任务报告是否包含传输到 S3 存储桶的每个对象的新版本。这仅当您在存储桶上启用版本控制时才适用。注意，将其设置为 INCLUDE 可能会延长任务执行的持续时间。

类型：字符串

有效值：INCLUDE | NONE

必需：否

OutputType

指定您想要的任务报告类型：

- SUMMARY_ONLY：提供有关您的任务的必要详细信息，包括传输的文件、对象和目录的数量以及传输时长。
- STANDARD：提供有关您的任务的完整详细信息，包括已传输、跳过、已验证等的文件、对象和目录的完整列表。

类型：字符串

有效值：SUMMARY_ONLY | STANDARD

必需：否

Overrides

为任务报告的各个方面自定义报告级别。例如，您的报告通常可能只包含错误，但您可以指定仅针对 DataSync 尝试在目标位置删除的文件提供成功和错误列表。

类型：ReportOverrides 对象

必需：否

ReportLevel

指定您是希望任务报告仅包含传输中出现的问题，还是列出成功和未成功传输的内容。

- ERRORS_ONLY：显示 DataSync 无法传输、跳过、验证和删除的内容的报告。
- SUCCESSES_AND_ERRORS：显示 DataSync 能够和无法传输、跳过、验证和删除的内容的报告。

类型：字符串

有效值：ERRORS_ONLY | SUCCESSES_AND_ERRORS

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

TaskSchedule

将 AWS DataSync 任务配置为按计划运行（最少间隔 1 小时）。

内容

ScheduleExpression

使用 cron 或 rate 表达式指定任务计划。

对于在特定时间和日期运行的任务计划，请使用 cron 表达式。例如，以下 cron 表达式创建了一个任务计划，在每月第一个星期三上午 8 点运行：

```
cron(0 8 * * 3#1)
```

对定期运行的任务计划，请使用 rate 表达式。例如，以下 rate 表达式创建了每 12 小时运行一次的任务计划：

```
rate(12 hours)
```

有关 cron 和 rate 表达式语法的信息，请参阅 Amazon EventBridge 用户指南。

类型：字符串

长度约束：最大长度为 256。

模式：`^[a-zA-Z0-9\ \_\*\?\,\|\^\-\/\#\s\(\)\+]*$`

必需：是

Status

指定启用还是禁用任务计划。您的计划默认为启用，但在某些情况下，可能需要禁用。例如，您可能需要暂停定期传输，以修复任务中的问题，或对存储系统进行维护。

如果任务因相同错误反复失败，DataSync 可能会自动禁用您的计划。有关更多信息，请参阅 TaskScheduleDetails。

类型：字符串

有效值：`ENABLED | DISABLED`

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

TaskScheduleDetails

提供有关 AWS DataSync 任务计划的信息。

内容

DisabledBy

表明任务计划被禁用的方式。

- USER：您的计划通过使用 UpdateTask 操作或 DataSync 控制台手动禁用。
- SERVICE：您的计划被 DataSync 自动禁用，因为任务因重复出现相同错误而失败。

类型：字符串

有效值：USER | SERVICE

必需：否

DisabledReason

如果任务计划被禁用，提供原因。

如果您的计划被 USER 禁用，您会看到一条 Manually disabled by user. 消息。

如果您的计划被 SERVICE 禁用，您会看到一条错误消息，帮助您了解任务不断失败的原因。有关解决 DataSync 错误的信息，请参阅解决 DataSync 传输问题。

类型：字符串

长度限制：最大长度为 8192。

模式：^[\w\s.,'?!:;\/=|<>()-]*$

必需：否

StatusUpdateTime

表明任务计划的状态上次发生更改的时间。例如，如果 DataSync 因重复出错而自动禁用计划，您可以查看计划何时被禁用。

类型：时间戳

必需：否

另请参阅

有关在特定语言的 AWS SDK 中使用此 API 的更多信息，请参阅以下内容：

- 适用于 C++ 的 AWS SDK
- 适用于 Java V2 的 AWS SDK
- 适用于 Ruby V3 的 AWS SDK

常见错误类型

本节列出了此 AWS 服务可能返回的常见错误类型。并非所有服务都会返回此处列出的所有错误类型。对于特定于此服务的 API 操作的错误，请参阅该 API 操作的主题。

AccessDeniedException

您无权执行此操作。验证您的 IAM 策略是否包含所需的权限。

HTTP 状态代码：403

ExpiredTokenException

请求中包含的安全令牌已过期。申请新的安全令牌并重试。

HTTP 状态代码：403

IncompleteSignature

请求签名不符合 AWS 标准。确认您使用的是有效的 AWS 凭证，并且您的请求格式是否正确。如果您使用的是 SDK，请确保它是最新的。

HTTP 状态代码：403

InternalFailure

由于内部服务器问题，该请求现在无法处理。请稍后重试。如果问题仍然存在，请联系 Supp AWS ort。

HTTP 状态代码：500

MalformedHttpRequestException

无法处理请求正文。当无法使用指定的内容编码算法对请求正文进行解压缩时，通常会发生这种情况。验证内容编码标头是否与使用的压缩格式相匹配。

HTTP 状态代码：400

NotAuthorized

您无权执行此操作。验证您的 IAM 策略是否包含所需的权限。

HTTP 状态代码：401

OptInRequired

您的 AWS 账户需要订阅此服务。确认您已在账户中启用该服务。

HTTP 状态代码：403

RequestAbortedException

在返回响应之前，请求已中止。这通常发生在客户端关闭连接时。

HTTP 状态代码：400

RequestEntityTooLargeException

请求实体太大。请缩小请求正文的大小，然后重试。

HTTP 状态代码：413

RequestTimeoutException

请求超时。服务器未在预期的时间范围内收到完整的请求。请重试。

HTTP 状态代码：408

ServiceUnavailable

服务暂时不可用。请稍后重试。

HTTP 状态代码：503

ThrottlingException

您的请求率太高。收到此异常的 AWS SDKs 自动重试请求。请降低请求频率。

HTTP 状态代码：400

UnknownOperationException

无法识别该操作或操作。确认操作名称拼写正确，并且您使用的 API 版本是否支持该名称。

HTTP 状态代码：404

UnrecognizedClientException

您提供的 X.509 证书或 AWS 访问密钥 ID 不在我们的记录中。确认您使用的是有效的凭证，并且这些凭证尚未过期。

HTTP 状态代码：403

ValidationError

输入不符合要求的格式或限制。检查是否包含所有必需的参数以及值是否有效。

HTTP 状态代码：400

常见参数

以下列表包含所有操作用于使用查询字符串对 Signature Version 4 请求进行签名的参数。任何特定于操作的参数都列在该操作的主题中。有关签名版本 4 的更多信息，请参阅 I AM 用户指南中的签署 AWS API 请求。

X-Amz-Algorithm

您用于创建请求签名的哈希算法。

条件：当您在查询字符串中而不是 HTTP 授权标头中包括身份验证信息时，请指定此参数。

类型：字符串

有效值：AWS4-HMAC-SHA256

必需：条件

X-Amz-Credential

凭证范围值，该值是一个字符串，其中包含您的访问密钥、日期、您要定位的区域、您请求的服务以及终止字符串（“aws4_request”）。值采用以下格式表示：access_key/YYYYMMDD/region/service/aws4_request。

有关更多信息，请参阅 I AM 用户指南中的创建已签名的 AWS API 请求。

条件：当您在查询字符串中而不是 HTTP 授权标头中包括身份验证信息时，请指定此参数。

类型：字符串

必需：条件

X-Amz-Date

用于创建签名的日期。格式必须为 ISO 8601 基本格式 (YYYYMMDD'T'HHMMSS'Z')。例如，以下日期时间是有效 X-Amz-Date值：`20120325T120000Z`。

条件： X-Amz-Date 对所有请求而言是可选的；它可以用于覆盖对请求签名所使用的日期。如果日期标题以 ISO 8601 基本格式指定，则 X-Amz-Date不是必填项。使用 X-Amz-Date时，它总是会覆盖 Date 标题的值。有关更多信息，请参阅 I AM 用户指南中的 AWS API 请求签名元素。

类型：字符串

必需：条件

X-Amz-Security-Token

通过调用 AWS Security Token Service (AWS STS) 获得的临时安全令牌。有关支持来自 AWS STS的临时安全凭证的服务列表，请参阅《IAM 用户指南》中的使用 IAM 的AWS 服务。

条件：如果您使用的是中的临时安全证书 AWS STS，则必须包含安全令牌。

类型：字符串

必需：条件

X-Amz-Signature

指定从要签名的字符串和派生的签名密钥计算的十六进制编码签名。

条件：当您在查询字符串中而不是 HTTP 授权标头中包括身份验证信息时，请指定此参数。

类型：字符串

必需：条件

X-Amz-SignedHeaders

指定作为规范请求的一部分包含的所有 HTTP 标头。有关指定签名标头的更多信息，请参阅 I AM 用户指南中的创建已签名的 AWS API 请求。

条件：当您在查询字符串中而不是 HTTP 授权标头中包括身份验证信息时，请指定此参数。

类型：字符串

必需：条件

文档历史记录

下表描述了 AWS DataSync 文档的重要补充。我们还经常更新文档来处理您发送给我们的反馈意见。

如需有关文档更新的通知，您可以订阅 RSS 源。

变更	说明	日期
增加了对 VPC 端点策略和启用 FIPS 的 VPC 端点的支持	现在，可以使用 VPC 端点策略，并在启用 FIPS 的 AWS 区域选择 FIPS VPC 端点。	2025 年 9 月 30 日
为 IPv6 地址添加了 Support	现在，您可以使用 IPv4 和 IPv6 地址与支持的数据源进行连接 DataSync 和数据传输。	2025 年 7 月 16 日
Discovery 不再可用	自 2025 年 5 月 20 日起，Discovery 不再是一项可用的 DataSync 功能。	2025 年 5 月 20 日
AWS 托管策略更新-更新现有策略	该AWSDataSyncFullAccess 策略更新了权限声明，这些声明从用于创建 Secret DataSync s Manager 密钥的权限中删除了标记条件。	2025 年 5 月 13 日
AWS 托管策略更新-更新现有策略	该AWSDataSyncFullAccess 策略具有 DataSync 允许使用的新权限 AWS Secrets Manager。	2025 年 5 月 7 日
AWS 托管策略更新-更新现有策略	该AWSDataSyncFullAccess 策略具有 DataSync 允许使用的新权限 AWS Secrets Manager。	2025 年 4 月 23 日
AWS 托管策略更新-更新现有策略	该AWSDataSyncFullAccess 策略具有 DataSync	2025 年 4 月 23 日

	允许使用 AWS Secrets Manager 和的新权限 AWS Key Management Service。	
AWS 托管策略更新-更新现有策略	`AWSDataSyncServiceRolePolicy` 具有 DataSync 允许使用的新权限 AWS Secrets Manager。	2025 年 4 月 15 日
执行大规模数据迁移	了解如何使用 DataSync 将文件或对象从本地或其他云存储传输到云存储来 AWS规划大规模数据迁移。	2025 年 2 月 19 日
支持 SMB 位置使用 Kerberos	DataSync 现在可以在连接到服务器消息块 (SMB) 文件服务器时使用 Kerberos 身份验证。	2025 年 1 月 28 日
New AWS 区域	AWS DataSync 可用于墨西哥（中部）地区的数据传输。	2025 年 1 月 14 日
New AWS 区域	AWS DataSync 适用于亚太地区（泰国）地区的数据传输。	2025 年 1 月 7 日
更新 AWS 存储位置	您现在可以更新您的亚马逊 S3、亚马逊 EFS 和亚马逊 FSx 转账地点。	2024 年 12 月 18 日
不再支持 Snowball Edge	自 2024 年 11 月 12 日起，DataSync 不再支持 AWS Snowball Edge。	2024 年 11 月 13 日
新的 AWS 托管策略	名为的 DataSync 服务相关角色`AWSServiceRoleForDataSync` 使用名为的新托管策略。`AWSDataSyncServiceRolePolicy`	2024 年 10 月 30 日

增强模式简介	使用增强模式，您可以在 Amazon S3 位置之间传输几乎无限的对象数量。	2024 年 10 月 30 日
AWS 托管策略更新-更新现有策略	该AWSDataSyncFullAccess 策略对与之配合使用的服务有了新的权限 DataSync。	2024 年 10 月 30 日
支持 Azure 存储通用 v1 账户	DataSync 在转入或转出Azure 存储设备时，可以使用存储通用 v1 帐户。Microsoft Azure Blob Storage	2024 年 10 月 4 日
新增用于配置任务计划的方法	您可以使用比率表达式来配置 DataSync 任务计划。	2024 年 8 月 22 日
New AWS 区域	AWS DataSync 适用于亚太地区（马来西亚）地区的数据传输。	2024 年 8 月 21 日
支持包含选择加入区域的无代理跨区域传输	当至少有一个 AWS 存储位置处于可选 AWS 区域状态时，您不再需要 DataSync 代理在存储服务之间进行传输。	2024 年 7 月 24 日
AWS 托管策略更新-更新现有策略	该AWSDataSyncFullAccess 策略对与之配合使用的服务有了新的权限 DataSync。	2024 年 7 月 22 日
更新了 S3 跨账户教程	移除了此类传输不再需要的一些源账户用户权限。	2024 年 6 月 10 日
新增任务执行状态	CANCELLING 状态指示任务执行被取消。	2024 年 5 月 15 日
新增用于暂停任务计划的选项	当需要对问题进行故障排除或执行存储系统维护时，可以禁用 AWS DataSync 任务计划。	2024 年 4 月 24 日

更新了 FIPS 端点的 TLS 密码	AWS DataSync 将 TLS_AES_128_GCM_SHA256 (secp256r1) 密码用于联邦信息处理标准 (FIPS) 服务端点。	2024 年 4 月 22 日
AWS 托管策略更新-更新现有策略	该AWSDataSyncFullAccess 策略对与之配合使用的服务有了新的权限 DataSync。	2024 年 2 月 16 日
使用清单传输特定文件或对象	AWS DataSync 可以使用清单传输文件或对象的列表。	2024 年 2 月 7 日
New AWS 区域	AWS DataSync 现在可用于加拿大西部（卡尔加里）地区的数据传输。	2023 年 12 月 20 日
支持使用其他云提供商进行传输	AWS DataSync 现在可以在 AWS 存储服务和IBM Cloud Object Storage或之间传输数据 Seagate Lyve Cloud。	2023 年 11 月 7 日
支持使用 Alibaba Cloud Object Storage Service 进行传输	AWS DataSync 现在可以在 AWS 存储服务与之间传输数据 Alibaba Cloud Object Storage Service。	2023 年 9 月 25 日
支持任务报告	使用任务报告监控您的 AWS DataSync 转移。	2023 年 8 月 30 日
New AWS 区域	AWS DataSync 现在可以在以色列（特拉维夫）地区进行数据传输。	2023 年 8 月 23 日

支持使用其他云提供商进行传输	AWS DataSync 现在可以在 AWS 存储服务和其他几个云提供商（例如 Wasabi Cloud StorageDigitalOcean Spaces、和Oracle Cloud Infrastructure Object Storage）之间传输数据。	2023 年 8 月 8 日
Microsoft Azure Blob Storage 支持正式发布	AWS DataSync 现在可以将对象传入和传出Microsoft Azure Blob Storage了。	2023 年 7 月 25 日
TLS 1.3 支持	在存储位置之间传输时，AWS DataSync 现在使用传输层安全 (TLS) 1.3 对所有网络流量进行加密。	2023 年 6 月 28 日
新的 DataSync 发现指标	AWS DataSync 数据查找服务 现在可以告诉你存储资源集群、存储虚拟机 LUNs (SVM) 或卷中有多少（逻辑单元号）。	2023 年 6 月 28 日
New AWS 区域	AWS DataSync 现已可用于亚太地区（墨尔本）地区的数据传输。	2023 年 5 月 24 日
在 Snowball Edge 上支持兼容 S3 的存储	您可以使用 AWS DataSync 在兼容 Amazon S3 的存储 AWS Snowball Edge 和 AWS 存储服务之间传输数据。	2023 年 5 月 18 日
AWS 托管策略更新-更新现有策略	该`AWSDataSyncFullAccess` 策略对与之配合使用的服务具有新的权限 DataSync。	2023 年 5 月 2 日
的正式上市 AWS DataSync 数据查找服务	使用 DataSync Discovery 来帮助加快向的迁移 AWS。	2023 年 4 月 25 日

Microsoft Azure Blob Storage 支持的公开预览版	AWS DataSync 现在可以从中传输对象Microsoft Azure Blob Storage。	2023 年 3 月 29 日
新的 IAM 策略	要支持 DataSync Discovery 功能，请 DataSync 使用名为`AWSServiceRoleForDataSyncDiscovery` 的服务相关角色。	2023 年 3 月 21 日
New AWS 区域	AWS DataSync 现已在以下地区推出 AWS 区域：亚太地区（海得拉巴）、欧洲（西班牙）和欧洲（苏黎世）。	2023 年 2 月 6 日
在任务执行中使用标签	现在，您可以为 AWS DataSync 任务执行添加标签。	2022 年 12 月 16 日
支持 S3 Glacier 即时检索	您现在可以将对象直接传输到 S3 Glacier 即时检索存储类中。	2022 年 12 月 16 日
复制对象系统元数据	AWS DataSync 现在可以在对象存储系统和 Amazon S3 之间传输时复制系统元数据。	2022 年 12 月 16 日
New AWS 区域	AWS DataSync 现已在中国（北京）和中国（宁夏）地区推出。	2022 年 12 月 14 日
New AWS 区域	AWS DataSync 现已在中东（阿联酋）地区推出。	2022 年 11 月 16 日
支持带有对象存储位置的自签名证书	AWS DataSync 可以连接到使用自签名或私有证书的对象存储位置。	2022 年 10 月 25 日

获取数据压缩信息	AWS DataSync 可以提供应用压缩后通过网络传输的物理字节数。	2022 年 10 月 25 日
的公开预览版 AWS DataSync 数据查找服务	使用 DataSync Discovery 来帮助加快向的迁移 AWS。	2022 年 9 月 21 日
用于将数据迁入或迁出 Google Cloud Storage 的新选项	通过在 Google Cloud 中部署 AWS DataSync 代理，您可以将数据传输至或传输出 Google Cloud Storage。	2022 年 7 月 21 日
支持 Amazon FSx 的 NetApp ONTAP 文件系统	AWS DataSync 现在可以在 ONTAP 文件系统之间来回 FSx 传输文件和文件夹。	2022 年 6 月 28 日
Amazon EFS 位置的新安全选项	AWS DataSync 可以使用 TLS、接入点和 IAM 角色访问 Amazon EFS 文件系统。	2022 年 5 月 31 日
将数据迁入或迁出 Google Cloud Storage 和 Azure 文件	使用 AWS DataSync，你可以将数据传输到 Google 云存储和 Azure 文件或从中传输数据。有关更多信息，请参阅为对象存储创建位置和为 SMB 创建位置。	2022 年 5 月 24 日
新 AWS DataSync 任务设置	使用复制对象标签选项，您可以指定在对象存储系统之间传输时是否保留对象标签。	2022 年 5 月 5 日
New AWS 区域	AWS DataSync 现已在亚太地区（雅加达）地区推出。	2022 年 4 月 19 日
支持亚马逊 FSx OpenZFS 文件系统	AWS DataSync 现在可以在 OpenZFS 文件系统之间来 FSx 回传输文件和文件夹。	2022 年 4 月 5 日

支持 Amazon f FSx or Lustre 文件系统	AWS DataSync 现在可以在 Lustre 文件系统之间来 FSx 回传输文件和文件夹。	2021 年 12 月 10 日
支持 Hadoop Distributed File System（HDFS）	AWS DataSync 现在支持在 HDFS 群集之间传输文件和文件夹。	2021 年 11 月 3 日
New AWS 区域	AWS DataSync 现已在亚太地区（大阪）地区推出。	2021 年 7 月 28 日
AWS 存储服务之间的全自动传输	AWS DataSync 现在，只需在 DataSync 控制台中单击几下，即可在 Amazon S3、Amazon EFS 或 FSx Windows 文件服务器之间传输文件或对象。	2020 年 11 月 9 日
调整正在运行的任务使用的网络带宽	AWS DataSync 现在，客户可以调整正在运行的 DataSync 任务所使用的网络带宽。当任务跨越多日时，这有助于最大限度地减少对其他用户或应用程序的影响。	2020 年 11 月 9 日
增强了对本地 DataSync 虚拟机 (VM) 功能的支持	AWS DataSync 代理虚拟机主机控制台现在支持增强功能，包括从本地控制台激活代理。	2020 年 10 月 19 日
AWS DataSync 现在可以来回传输数据 AWS Outposts	DataSync 现在支持在 Amazon S3 上往返传输对象 AWS Outposts。	2020 年 9 月 30 日
支持 API 筛选	AWS DataSync 现在支持筛选`ListTasks` 和 `ListLocations` API 调用，使您能够使用过滤器（例如数据传输的源或目标）轻松检索数据传输任务的配置。	2020 年 8 月 18 日

支持从自行管理的对象存储中复制数据	AWS DataSync 现在支持自行管理的对象存储与 Amazon S3、Amazon Elastic File System 或 FSx Windows 文件服务器之间的数据传输。	2020 年 7 月 27 日
支持基于 Linux 基于内核的虚拟机（KVM）和 Microsoft Hyper-V 管理程序	AWS DataSync 除了现有的 VMware 和 Amazon EC2 选项外，现在还能够在 KVM 和 Microsoft Hyper-V 虚拟化平台上部署本地代理。	2020 年 7 月 1 日
AWS DataSync 现在可以自动配置您的 Amazon CloudWatch 日志配置	使用时 DataSync，您现在可以选择自动生成发布数据传输 CloudWatch日志所需的日志组和资源策略，从而简化任务创建和监控设置。	2020 年 7 月 1 日
AWS DataSync 现在可以来回传输数据 AWS Snowball Edge	DataSync 现在支持与边缘计算和数据传输设备中 AWS Snow Family 最小的成员进行文件传输。 AWS Snowball Edge Snowball Edge 便携、坚固耐用且安全 – 小巧轻便，可放入背包，并能抵御恶劣的环境。	2020 年 6 月 17 日
New AWS 区域	AWS DataSync 现已在非洲（开普敦）地区和欧洲（米兰）地区推出。	2020 年 6 月 16 日
带有文件级日志记录的增强型监控功能	现在，您可以为在 NFS 服务器、SMB 服务器、Amazon S3 存储桶、Amazon EFS 文件系统之间复制的文件和 FSx 对象以及 Windows 文件服务器文件系统启用详细日志记录。	2020 年 4 月 24 日

支持在您的 SMB 共享和亚马逊 Windows 文件 FSx 服务器之间复制数据	现在，您可以在 SMB 共享和 FSx Windows 文件服务器之间复制数据。	2020 年 1 月 24 日
支持计划任务	您现在可以手动运行任务，也可以根据指定计划安排任务运行。	2019 年 11 月 20 日
New AWS 区域	AWS DataSync 现已在亚太地区（香港）地区、亚太地区（孟买）地区、欧洲（斯德哥尔摩）区域、南美洲（圣保罗）地区 AWS GovCloud 和（美国东部）地区推出。	2019 年 11 月 20 日
New AWS 区域	AWS DataSync 现已在加拿大（中部）区域、欧洲（伦敦）地区和欧洲（巴黎）地区推出。	2019 年 10 月 2 日
支持 Amazon S3 存储类别	现在，您可以直接将对象传输到 Amazon S3 存储类。	2019 年 9 月 24 日
New AWS 区域	AWS DataSync 现已在中东（巴林）地区推出。	2019 年 8 月 28 日
支持在服务器消息块 (SMB) 共享和 Amazon S3 或 Amazon EFS 之间复制数据。	您现在可以在 SMB 文件共享与 Amazon S3 或 Amazon EFS 之间复制数据。	2019 年 8 月 22 日
支持使用虚拟私有云（VPC）端点	现在，您可以在代理和专用网络之间创建私有连接， AWS 并在私有网络中运行任务。这样做会增加通过网络复制数据时的安全性。	2019 年 8 月 5 日
支持美国联邦信息处理标准 (FIPS) 端点	现在，您可以使用 FIPS 端点创建代理并运行任务。	2019 年 8 月 5 日

New AWS 区域	AWS DataSync 现已在 AWS GovCloud（美国西部）地区推出。	2019 年 6 月 11 日
支持筛选	现在，当您将数据从源传输到目标位置时，您可以应用筛选条件以仅传输源位置中的文件子集。	2019 年 5 月 22 日
的首次发布 AWS DataSync	该 AWS DataSync 服务的正式发布。	2018 年 11 月 26 日

AWS 术语表

有关最新的 AWS 术语，请参阅 AWS 词汇表 参考中的 AWS 词汇表。

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。