Creating a location for Amazon S3

PDF (sync-dg.pdf#create-s3-location) | RSS (aws-datasync-release-notes.rss)

An Amazon S3 bucket can be a source or destination location for AWS DataSync.

Remember the following when using Amazon S3 with DataSync:

- Changes to object data or metadata are equivalent to deleting and replacing an object. These changes result in additional charges in the following scenarios:
  - **When using object versioning**: Changes to object data or metadata create a new version of the object.
  - **When using storage classes that can incur additional charges for overwriting, deleting, or retrieving objects**: Changes to object data or metadata result in such charges. For smore information, see Considerations when working with Amazon S3 storage classes in DataSync (#using-storage-classes) .
- When using object versioning, running a DataSync task once might create more than one version of an Amazon S3 object.
- DataSync requires access to your Amazon S3 bucket. To do this, DataSync assumes an IAM role that includes an IAM policy and security token service trust (STS) relationship. The policy determines which actions the role can perform. Let DataSync create the role for you or specify a role you created. For more information, see Manually configuring an IAM role to access your Amazon S3 bucket (#create-role-manually) .
- In addition to the IAM policies that grant DataSync permissions, we recommend creating a multipart upload bucket policy for your S3 buckets to help control your storage costs. For more information, see the blog post Amazon S3 Lifecycle Management Update - Support for Multipart Uploads and Delete Markers (http://aws.amazon.com/blogs/aws/s3-lifecycle-management-update-support-for-multipart-uploads-and-delete-markers/) .

To create an Amazon S3 location

1. Open the AWS DataSync console at https://console.aws.amazon.com/datasync/ (https://console.aws.amazon.com/datasync/) .
2. Go to the **Locations** page and select **Create location**.
3. For **Location type**, choose **Amazon S3**.
4. For **S3 bucket**, choose the bucket that you want to use as a location. (When creating your DataSync task later, you specify whether this location is a source or destination location.)

   If your S3 bucket is located on AWS Outposts, you must specify an Amazon S3 access point. For more information, see Managing data access with Amazon S3 access points (https://docs.aws.amazon.com/AmazonS3/latest/userguide/access-points.html) in the *Amazon S3 User Guide*.

5. For **S3 storage class**, choose a storage class that you want to transfer objects into.

   Some storage classes can affect your Amazon S3 costs. For more information, see Considerations when working with Amazon S3 storage classes in DataSync (#using-storage-classes) . For Amazon S3 on AWS Outposts, DataSync by default uses the S3 Outposts storage class.

6. For **Agents**, specify the Amazon Resource Name (ARN) of the DataSync agent on your AWS Outposts.

   For more information, see Deploy your agent on AWS Outposts (./deploy-agents.html#outposts-agent) .

7. For **Folder**, enter a prefix in the S3 bucket that DataSync reads from or writes to (depending on whether the bucket is a source or destination location).

   > **Note**
   >
   > The prefix can't begin with a slash (for example, `/photos`) or include consecutive slashes, such as `photos//2006/January`.

8. For **IAM role**, do one of the following:

   - Choose **Autogenerate** for DataSync to automatically create an IAM role with the permissions required to access the S3 bucket.

     If DataSync previously created an IAM role for this S3 bucket, that role is chosen by default.

   - Select a custom IAM role you created. For more information, see Manually configuring an IAM role to access your Amazon S3 bucket (#create-role-manually)

9. (Optional) Select **Add tag** to tag your Amazon S3 location.

   A *tag* is a key-value pair that helps you manage, filter, and search for your locations.

10. Choose **Create location**.

    Once created, the location displays on the **Locations** page.

Considerations when working with Amazon S3 storage classes in DataSync

DataSync can transfer objects directly into the Amazon S3 storage class that you choose. For more information about Amazon S3 storage classes, see Amazon S3 storage classes (http://aws.amazon.com/s3/storage-classes/) . Some storage classes have behaviors that can affect your Amazon S3 storage costs. For more information, see Amazon S3 pricing (http://aws.amazon.com/s3/pricing/) .

Following, you can find some considerations for how each Amazon S3 storage class works with DataSync.

Amazon S3 storage class	**Considerations**
S3 Standard	Choose S3 Standard to store your frequently accessed files redundantly in multiple Availability Zones that are geographically separated. This is the default if you don't specify a storage class.

S3 Intelligent-Tiering	Choose S3 Intelligent-Tiering to optimize storage costs by automatically moving data to the most cost-effective storage access tier. You pay a monthly charge per object stored in the S3 Intelligent-Tiering storage class. This Amazon S3 charge includes monitoring data access patterns and moving objects between tiers.
S3 Standard-IA	Choose S3 Standard-IA to store your infrequently accessed files redundantly in multiple Availability Zones that are geographically separated. Objects stored in the S3 Standard-IA storage class can incur additional charges for overwriting, deleting, or retrieving. Consider how often these objects change, how long you plan to keep these objects, and how often you need to access them. Changes to object data or metadata are equivalent to deleting an object and creating a new one to replace it. This results in additional charges for objects stored in the S3 Standard-IA storage class. Objects less than 128 KB are smaller than the minimum capacity charge per object in the S3 Standard-IA storage class. These objects are stored in the S3 Standard storage class.

S3 One Zone-IA	Choose S3 One Zone-IA to store your infrequently accessed files in a single Availability Zone. Objects stored in the S3 One Zone-IA storage class can incur additional charges for overwriting, deleting, or retrieving. Consider how often these objects change, how long you plan to keep these objects, and how often you need to access them. Changes to object data or metadata are equivalent to deleting an object and creating a new one to replace it. This results in additional charges for objects stored in the S3 One Zone-IA storage class. Objects less than 128 KB are smaller than the minimum capacity charge per object in the S3 One Zone-IA storage class. These objects are stored in the S3 Standard storage class.
S3 Glacier Flexible Retrieval	Choose S3 Glacier Flexible Retrieval for more active archives. Objects stored in S3 Glacier Flexible Retrieval can incur additional charges for overwriting, deleting, or retrieving. Consider how often these objects change, how long you plan to keep these objects, and how often you need to access them. Changes to object data or metadata are equivalent to deleting an object and creating a new one to replace it. This results in additional charges for objects stored in the S3 Glacier Flexible Retrieval storage class. Objects less than 40 KB are smaller than the minimum capacity charge per object in the S3 Glacier Flexible Retrieval storage class. These objects are stored in the S3 Standard storage class. You must restore objects archived in this storage class before DataSync can read them. For information, see Working with archived objects (https://docs.aws.amazon.com/AmazonS3/latest/userguide/archived-objects.html) in the *Amazon S3 User Guide*. When using S3 Glacier Flexible Retrieval, choose **Verify only the data transferred** to compare data and metadata checksums at the end of the transfer. You can't use the **Verify all data in the destination** option for this storage class because it requires retrieving all existing objects from the destination.

S3 Glacier Deep Archive	Choose S3 Glacier Deep Archive to archive your files for long-term data retention and digital preservation where data is accessed once or twice a year. Objects stored in S3 Glacier Deep Archive can incur additional charges for overwriting, deleting, or retrieving. Consider how often these objects change, how long you plan to keep these objects, and how often you need to access them. Changes to object data or metadata are equivalent to deleting an object and creating a new one to replace it. This results in additional charges for objects stored in the S3 Glacier Deep Archive storage class. Objects less than 40 KB are smaller than the minimum capacity charge per object in the S3 Glacier Deep Archive storage class. These objects are stored in the S3 Standard storage class. You must restore objects archived in this storage class before DataSync can read them. For information, see Working with archived objects (https://docs.aws.amazon.com/AmazonS3/latest/userguide/archived-objects.html) in the *Amazon S3 User Guide*. When using S3 Glacier Deep Archive, choose **Verify only the data transferred** to compare data and metadata checksums at the end of the transfer. **Verify all data in the destination** isn't an available option for this storage class, because it requires retrieving all existing objects from the destination.
S3 Outposts	The storage class for Amazon S3 on Outposts.

Manually configuring an IAM role to access your Amazon S3 bucket

When you use the DataSync console to create an Amazon S3 location, DataSync automatically creates an IAM role that has the required permissions for you. If you want to create the IAM role manually, use the following procedure.

To manually configure an IAM role to access your Amazon S3 bucket

1. Open the IAM console at https://console.aws.amazon.com/iam/ (https://console.aws.amazon.com/iam/) .
2. In the left navigation pane, choose **Roles**, and then choose **Create role** to open the **Create role** page.
3. In the **Select type of trusted entity** section, make sure that **AWS service** is selected.
4. Under **Choose the service that will use this role**, choose **DataSync** or manually configure it (see the following example).

   To prevent the cross-service confused deputy problem (./cross-service-confused-deputy-prevention.html) , we recommend using the `aws:SourceArn` and `aws:SourceAccount` global condition context keys in the policy.

   ```
   {
       "Version": "2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
               "Principal": {
                   "Service": "datasync.amazonaws.com"
               },
               "Action": "sts:AssumeRole",
               "Condition": {
                   "StringEquals": {
                       "aws:SourceAccount": "123456789012"
                   },
                   "StringLike": {
                       "aws:SourceArn": "arn:aws:datasync:us-east-2:123456789012:*"
                   }
               }
           }
       ]
   }
   ```

5. Under **Select your use case**, choose **DataSync - S3 Location**.
6. Choose **Next: Permissions**.
7. For Amazon S3 buckets in AWS Regions, choose **AmazonS3FullAccess**. You can also manually configure a more restrictive policy. For an example of such a policy, see the following.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "s3:GetBucketLocation",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads"
            ],
            "Effect": "Allow",
            "Resource": "YourS3BucketArn"
        },
        {
            "Action": [
                "s3:AbortMultipartUpload",
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:ListMultipartUploadParts",
                "s3:GetObjectTagging",
                "s3:PutObjectTagging",
                "s3:PutObject"
              ],
            "Effect": "Allow",
            "Resource": "YourS3BucketArn/*"
        }
    ]
}
```

For Amazon S3 buckets on Outposts, use the following policy.

```

{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "s3-outposts:ListBucket",
                "s3-outposts:ListBucketMultipartUploads"
            ],
            "Effect": "Allow",
            "Resource": [
                "s3OutpostsBucketArn",
                "s3OutpostsAccessPointArn"
            ],
            "Condition": {
                "StringLike": {
                    "s3-outposts:DataAccessPointArn":
"s3OutpostsAccessPointArn"
                }
            }
        },
        {
            "Action": [
                "s3-outposts:AbortMultipartUpload",
                "s3-outposts:DeleteObject",
                "s3-outposts:GetObject",
                "s3-outposts:ListMultipartUploadParts",
                "s3-outposts:GetObjectTagging",
                "s3-outposts:PutObjectTagging"
            ],
            "Effect": "Allow",
            "Resource": [
                "s3OutpostsBucketArn/*",
                "s3OutpostsAccessPointArn"
            ],
            "Condition": {
                "StringLike": {
                    "s3-outposts:DataAccessPointArn":
"s3OutpostsAccessPointArn"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3-outposts:GetAccessPoint"
            ],
            "Resource": "s3OutpostsAccessPointArn"
        }
    ]
}
```

8. (Optional) **Choose Next: Tags** to create tags for the role.
9. Choose **Next: Review**, choose the role name, and then choose **Create role**.

10. Open the AWS DataSync console at https://console.aws.amazon.com/datasync/ (https://console.aws.amazon.com/datasync/) .
11. Choose the refresh button on the right side of the IAM role list, and then choose the role that you just created.