**Amazon S3 Inventory**

Amazon S3 inventory is one of the tools Amazon S3 provides to help manage your storage. You can use it to audit and report on the replication and encryption status of your objects for business, compliance, and regulatory needs. You can also simplify and speed up business workflows and big data jobs using Amazon S3 inventory, which provides a scheduled alternative to the Amazon S3 synchronous List API operation.

Amazon S3 inventory provides comma-separated values (CSV) or [Apache optimized row columnar (ORC)](https://orc.apache.org/)output files that list your objects and their corresponding metadata on a daily or weekly basis for an S3 bucket or a shared prefix (that is, objects that have names that begin with a common string). For information about Amazon S3 inventory pricing, see [Amazon S3 Pricing](https://aws.amazon.com/s3/pricing/).

You can configure multiple inventory lists for a bucket. You can configure what object metadata to include in the inventory, whether to list all object versions or only current versions, where to store the inventory list file output, and whether to generate the inventory on a daily or weekly basis. You can also specify that the inventory list file be encrypted.

You can query Amazon S3 inventory using standard SQL by using Amazon Athena, Amazon Redshift Spectrum, and other tools such as [Presto](https://prestodb.io/), [Apache Hive](https://hive.apache.org/), and [Apache Spark](https://databricks.com/spark/about/). It's easy to use Athena to run queries on your inventory files. You can use Athena for Amazon S3 inventory queries in all Regions where Athena is available.

**Topics**

* [How Do I Set Up Amazon S3 Inventory?](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-how-to-set-up)
* [What's Included in an Amazon S3 Inventory?](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-contents)
* [Where Are Inventory Lists Located?](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-location)
* [How Do I Know When an Inventory Is Complete?](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-notification)
* [Querying Inventory with Amazon Athena](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-athena-query)
* [Amazon S3 Inventory REST APIs](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-related-resources)

## How Do I Set Up Amazon S3 Inventory?

This section describes how to set up an inventory, including details about the inventory source and destination buckets.

### Amazon S3 Inventory Source and Destination Buckets

The bucket that the inventory lists the objects for is called the source bucket. The bucket where the inventory list file is stored is called the destination bucket.

**Source Bucket**

The inventory lists the objects that are stored in the source bucket. You can get inventory lists for an entire bucket or filtered by (object key name) prefix.

The source bucket:

* Contains the objects that are listed in the inventory.
* Contains the configuration for the inventory.

**Destination Bucket**

Amazon S3 inventory list files are written to the destination bucket. To group all the inventory list files in a common location in the destination bucket, you can specify a destination (object key name) prefix in the inventory configuration.

The destination bucket:

* Contains the inventory file lists.
* Contains the manifest files that list all the file inventory lists that are stored in the destination bucket. For more information, see [What Is an Inventory Manifest?](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-location-manifest)
* Must have a bucket policy to give Amazon S3 permission to verify ownership of the bucket and permission to write files to the bucket.
* Must be in the same AWS Region as the source bucket.
* Can be the same as the source bucket.
* Can be owned by a different AWS account than the account that owns the source bucket.

### Setting Up Amazon S3 Inventory

Amazon S3 inventory helps you manage your storage by creating lists of the objects in an S3 bucket on a defined schedule. You can configure multiple inventory lists for a bucket. The inventory lists are published to CSV or ORC files in a destination bucket.

The easiest way to set up an inventory is by using the AWS Management Console, but you can also use the REST API, AWS CLI, or AWS SDKs. The console performs the first step of the following procedure for you: adding a bucket policy to the destination bucket.

**To set up Amazon S3 inventory for an S3 bucket**

1. **Add a bucket policy for the destination bucket.**

You must create a bucket policy on the destination bucket to grant permissions to Amazon S3 to write objects to the bucket in the defined location. For an example policy, see [Granting Permissions for Amazon S3 Inventory and Amazon S3 Analytics](https://docs.aws.amazon.com/AmazonS3/latest/dev/example-bucket-policies.html#example-bucket-policies-use-case-9).

1. **Configure an inventory to list the objects in a source bucket and publish the list to a destination bucket.**

When you configure an inventory list for a source bucket, you specify the destination bucket where you want the list to be stored, and whether you want to generate the list daily or weekly. You can also configure what object metadata to include and whether to list all object versions or only current versions.

You can specify that the inventory list file be encrypted by using Amazon S3-managed keys (SSE-S3) or AWS KMS-managed keys (SSE-KMS). For more information about SSE-S3 and SSE-KMS, see [Protecting Data Using Server-Side Encryption](https://docs.aws.amazon.com/AmazonS3/latest/dev/serv-side-encryption.html). If you plan to use SSE-KMS encryption, see Step 3.

* + For information about how to use the console to configure an inventory list, see [How Do I Configure Amazon S3 Inventory?](http://docs.aws.amazon.com/AmazonS3/latest/user-guide/configure-inventory.html) in the Amazon Simple Storage Service Console User Guide.
  + To use the Amazon S3 API to configure an inventory list, use the [PUT Bucket inventory configuration](http://docs.aws.amazon.com/AmazonS3/latest/API/RESTBucketPUTInventoryConfig.html) REST API, or the equivalent from the AWS CLI or AWS SDKs.

1. **To encrypt the inventory list file with SSE-KMS, grant Amazon S3 permission to use the AWS KMS key.**

You can configure encryption for the inventory list file by using the AWS Management Console, REST API, AWS CLI, or AWS SDKs. Whichever way you choose, you must grant Amazon S3 permission to use the AWS KMS customer master key (CMK) to encrypt the inventory file. You grant Amazon S3 permission by modifying the key policy for the AWS KMS CMK that is being used to encrypt the inventory file. For more information, see the next section, [Grant Amazon S3 Permission to Encrypt Using Your AWS KMS Key](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-inventory.html#storage-inventory-kms-key-policy).

#### Grant Amazon S3 Permission to Encrypt Using Your AWS KMS Key

You must grant Amazon S3 permission to encrypt using your AWS KMS key with a key policy. The following procedure describes how to use the AWS Identity and Access Management (IAM) console to modify the key policy for the AWS KMS CMK that is used to encrypt the inventory file.

**To grant permissions to encrypt using your AWS KMS key**

1. Sign in to the AWS Management Console using the AWS account that owns the AWS KMS CMK, and open the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the left navigation pane, choose **Encryption keys**.
3. For **Region**, choose the appropriate AWS Region. Do not use the region selector in the navigation bar (upper-right corner).
4. Choose the alias of the CMK that you want to encrypt inventory with.
5. In the **Key Policy** section of the page, choose **Switch to policy view**.
6. Using the **Key Policy** editor, insert following key policy into the existing policy and then choose **Save Changes**. You might want to copy the policy to the end of the existing policy.
7. {
8. "Sid": "Allow Amazon S3 use of the key",
9. "Effect": "Allow",
10. "Principal": {
11. "Service": "s3.amazonaws.com"
12. },
13. "Action": [
14. "kms:GenerateDataKey\*"
15. ],
16. "Resource": "\*"

}

You can also use the AWS KMS PUT key policy API [PutKeyPolicy](http://docs.aws.amazon.com/kms/latest/APIReference/API_PutKeyPolicy.html) to copy the key policy to the CMK that is being used to encrypt the inventory file. For more information about creating and editing AWS KMS CMKs, see [Getting Started](http://docs.aws.amazon.com/kms/latest/developerguide/getting-started.html) in the AWS Key Management Service Developer Guide.

## What's Included in an Amazon S3 Inventory?

An inventory list file contains a list of the objects in the source bucket and metadata for each object. The inventory lists are stored in the destination bucket as a CSV file compressed with GZIP or as an Apache optimized row columnar (ORC) file.

The inventory list contains a list of the objects in an S3 bucket and the following metadata for each listed object:

* **Bucket name** – The name of the bucket that the inventory is for.
* **Key name** – Object key name (or key) that uniquely identifies the object in the bucket. When using the CSV file format, the key name is URL-encoded and must be decoded before you can use it.
* **Version ID** – Object version ID. When you enable versioning on a bucket, Amazon S3 assigns a version number to objects that are added to the bucket. For more information, see [Object Versioning](https://docs.aws.amazon.com/AmazonS3/latest/dev/ObjectVersioning.html). (This field is not included if the list is only for the current version of objects.)
* **IsLatest** – Set to True if the object is the current version of the object. (This field is not included if the list is only for the current version of objects.)
* **Size** – Object size in bytes.
* **Last modified date** – Object creation date or the last modified date, whichever is the latest.
* **ETag** – The entity tag is a hash of the object. The ETag reflects changes only to the contents of an object, not its metadata. The ETag may or may not be an MD5 digest of the object data. Whether it is depends on how the object was created and how it is encrypted.
* **Storage class** – Storage class used for storing the object. For more information, see [Storage Classes](https://docs.aws.amazon.com/AmazonS3/latest/dev/storage-class-intro.html).
* **Multipart upload flag** – Set to True if the object was uploaded as a multipart upload. For more information, see [Multipart Upload Overview](https://docs.aws.amazon.com/AmazonS3/latest/dev/mpuoverview.html).
* **Delete marker** – Set to True, if the object is a delete marker. For more information, see [Object Versioning](https://docs.aws.amazon.com/AmazonS3/latest/dev/ObjectVersioning.html). (This field is not included if the list is only for the current version of objects.)
* **Replication status** – Set to PENDING, COMPLETED, FAILED, or REPLICA. For more information, see [Finding the Cross-Region Replication Status](https://docs.aws.amazon.com/AmazonS3/latest/dev/crr-status.html).
* **Encryption status** – Set to SSE-S3, SSE-C, SSE-KMS, or NOT-SSE. The server-side encryption status for SSE-S3, SSE-KMS, and SSE with customer-provided keys (SSE-C). A status of NOT-SSE means that the object is not encrypted with server-side encryption. For more information, see [Protecting Data Using Encryption](https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingEncryption.html).

The following is an example CSV inventory list opened in a spreadsheet application. The heading row is shown only to help clarify the example; it is not included in the actual list.


        Screenshot of an example inventory list opened in a spreadsheet
          application.
      

We recommend that you create a lifecycle policy that deletes old inventory lists. For more information, see [Object Lifecycle Management](https://docs.aws.amazon.com/AmazonS3/latest/dev/object-lifecycle-mgmt.html).

### Inventory Consistency

All of your objects might not appear in each inventory list. The inventory list provides eventual consistency for PUTs of both new objects and overwrites, and DELETEs. Inventory lists are a rolling snapshot of bucket items, which are eventually consistent (that is, the list might not include recently added or deleted objects).

To validate the state of the object before you take action on the object, we recommend that you perform a HEAD Object REST API request to retrieve metadata for the object, or check the object's properties in the Amazon S3 console. You can also check object metadata with the AWS CLI or the AWS SDKS. For more information, see [HEAD Object](http://docs.aws.amazon.com/AmazonS3/latest/API/RESTObjectHEAD.html) in the Amazon Simple Storage Service API Reference.

## Where Are Inventory Lists Located?

When an inventory list is published, the manifest files are published to the following location in the destination bucket.

*destination-prefix*/*source-bucket*/*config-ID*/*YYYY-MM-DDTHH-MMZ*/manifest.json

*destination-prefix*/*source-bucket*/*config-ID*/*YYYY-MM-DDTHH-MMZ*/manifest.checksum

*destination-prefix*/*source-bucket*/*config-ID*/hive/dt=*YYYY-MM-DD-HH-MM*/symlink.txt

* *destination-prefix* is the (object key name) prefix set in the inventory configuration, which can be used to group all the inventory list files in a common location within the destination bucket.
* *source-bucket* is the source bucket that the inventory list is for. It is added to prevent collisions when multiple inventory reports from different source buckets are sent to the same destination bucket.
* *config-ID* is added to prevent collisions with multiple inventory reports from the same source bucket that are sent to the same destination bucket.
* *YYYY-MM-DDTHH-MMZ* is the time stamp that consists of the start time and the date when the inventory report generation begins scanning the bucket; for example, 2016-11-06T21-32Z. Storage added after the time stamp will not be in the report.
* manifest.json is the manifest file.
* manifest.checksum is the MD5 of the content of the manifest.json file.
* symlink.txt is the Apache Hive-compatible manifest file.

The inventory lists are published daily or weekly to the following location in the destination bucket.

*destination-prefix*/*source-bucket*/data/*example-file-name.csv.gz*

...

*destination-prefix*/*source-bucket*/data/*example-file-name-1.csv.gz*

* *destination-prefix* is the (object key name) prefix set in the inventory configuration. It can be used to group all the inventory list files in a common location in the destination bucket.
* *source-bucket* is the source bucket that the inventory list is for. It is added to prevent collisions when multiple inventory reports from different source buckets are sent to the same destination bucket.
* *example-file-name*.csv.gz is one of the CSV inventory files. ORC inventory names end with the file name extension .orc.

### What Is an Inventory Manifest?

The manifest files manifest.json and symlink.txt describe where the inventory files are located. Whenever a new inventory list is delivered, it is accompanied by a new set of manifest files.

Each manifest contained in the manifest.json file provides metadata and other basic information about an inventory. This information includes the following:

* Source bucket name
* Destination bucket name
* Version of the inventory
* Creation time stamp in the epoch date format that consists of the start time and the date when the inventory report generation begins scanning the bucket
* Format and schema of the inventory files
* Actual list of the inventory files that are in the destination bucket

Whenever a manifest.json file is written, it is accompanied by a manifest.checksum file that is the MD5 of the content of manifest.json file.

The following is an example of a manifest in a manifest.json file for a CSV-formatted inventory.

{

"sourceBucket": "example-source-bucket",

"destinationBucket": "example-inventory-destination-bucket",

"version": "2016-11-30",

"creationTimestamp" : "1514944800000",

"fileFormat": "CSV",

"fileSchema": "Bucket, Key, VersionId, IsLatest, IsDeleteMarker, Size, LastModifiedDate, ETag, StorageClass, MultipartUploaded, ReplicationStatus",

"files": [

{

"key": "Inventory/example-source-bucket/2016-11-06T21-32Z/files/939c6d46-85a9-4ba8-87bd-9db705a579ce.csv.gz",

"size": 2147483647,

"MD5checksum": "f11166069f1990abeb9c97ace9cdfabc",

"inventoriedRecord": 58050695

}

]

}

The following is an example of a manifest in a manifest.json file for an ORC-formatted inventory.

{

"sourceBucket": "example-source-bucket",

"destinationBucket": "arn:aws:s3:::example-destination-bucket",

"version": "2016-11-30",

"creationTimestamp" : "1514944800000",

"fileFormat": "ORC",

"fileSchema": "struct<bucket:string,key:string,version\_id:string,is\_latest:boolean,is\_delete\_marker:boolean,size:bigint,last\_modified\_date:timestamp,e\_tag:string,storage\_class:string,is\_multipart\_uploaded:boolean,replication\_status:string,encryption\_status:string>",

"files": [

{

"key": "inventory/example-source-bucket/data/d794c570-95bb-4271-9128-26023c8b4900.orc",

"size": 56291,

"MD5checksum": "5925f4e78e1695c2d020b9f6eexample"

}

]

}

The symlink.txt file is an Apache Hive-compatible manifest file that allows Hive to automatically discover inventory files and their associated data files. The Hive-compatible manifest works with any Hive-compatible service, including Athena, AWS Glue, and Amazon Redshift Spectrum. It also works with Hive-compatible applications, including [Presto](https://prestodb.io/), [Apache Hive](https://hive.apache.org/), [Apache Spark](https://databricks.com/spark/about/), and many others.

## How Do I Know When an Inventory Is Complete?

You can set up an Amazon S3 event notification to receive notice when the manifest checksum file is created, which indicates that an inventory list has been added to the destination bucket. The manifest is an up-to-date list of all the inventory lists at the destination location.

Amazon S3 can publish events to an Amazon Simple Notification Service (Amazon SNS) topic, an Amazon Simple Queue Service (Amazon SQS) queue, or an AWS Lambda function. For more information, see[Configuring Amazon S3 Event Notifications](https://docs.aws.amazon.com/AmazonS3/latest/dev/NotificationHowTo.html).

The following notification configuration defines that all manifest.checksum files newly added to the destination bucket are processed by the AWS Lambda cloud-function-list-write.

<NotificationConfiguration>

<QueueConfiguration>

<Id>1</Id>

<Filter>

<S3Key>

<FilterRule>

<Name>prefix</Name>

<Value>*destination-prefix*/*source-bucket*</Value>

</FilterRule>

<FilterRule>

<Name>suffix</Name>

<Value>checksum</Value>

</FilterRule>

</S3Key>

</Filter>

<Cloudcode>arn:aws:lambda:us-west-2:222233334444:cloud-function-list-write</Cloudcode>

<Event>s3:ObjectCreated:\*</Event>

</QueueConfiguration>

</NotificationConfiguration>

For more information, see [Using AWS Lambda with Amazon S3](http://docs.aws.amazon.com/lambda/latest/dg/with-s3.html) in the AWS Lambda Developer Guide.

## Querying Inventory with Amazon Athena

You can query Amazon S3 inventory using standard SQL by using Amazon Athena in all Regions where Athena is available. To check for AWS Region availability, see the [AWS Region Table](https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services/).

Athena can query Amazon S3 inventory files in ORC or CSV format. When you use Athena to query inventory, we recommend that you use ORC-formatted inventory files instead of CSV. ORC provides faster query performance and lower query costs. ORC is a self-describing type-aware columnar file format designed for [Apache Hadoop](http://hadoop.apache.org/). The columnar format lets the reader read, decompress, and process only the columns that are required for the current query. The ORC format for Amazon S3 inventory is available in all AWS Regions.

**To get started using Athena to query Amazon S3 inventory**

1. Create an Athena table. For information about creating a table, see [Getting Started](http://docs.aws.amazon.com/athena/latest/ug/getting-started.html) in the Amazon Athena User Guide.

The following sample query includes all optional fields in the inventory report. Drop any optional field that you did not choose for your inventory so that the query corresponds to the fields chosen for your inventory. Also, you must use your bucket name and the location. The location points to your inventory destination path; for example, s3://destination-prefix/source-bucket/config-ID/hive.

CREATE EXTERNAL TABLE *your-table-name*(

`bucket` string,

key string,

version\_id string,

is\_latest boolean,

is\_delete\_marker boolean,

size bigint,

last\_modified\_date timestamp,

e\_tag string,

storage\_class string,

is\_multipart\_uploaded boolean,

replication\_status string,

encryption\_status string

)

PARTITIONED BY (dt string)

ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.orc.OrcSerde'

STORED AS INPUTFORMAT 'org.apache.hadoop.hive.ql.io.SymlinkTextInputFormat'

OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.IgnoreKeyTextOutputFormat'

LOCATION 's3://*destination-prefix/source-bucket/config-ID*/hive';

1. To add new inventory lists to your table, use the following MSCK REPAIR TABLE command.

MSCK REPAIR TABLE *your-table-name*;

1. After performing the first two steps, you can run ad hoc queries on your inventory, as shown in the following example.

SELECT encryption\_status, count(\*) FROM *your-table-name* GROUP BY encryption\_status;

For more information about using Athena, see [Amazon Athena User Guide](http://docs.aws.amazon.com/athena/latest/ug/).

## Amazon S3 Inventory REST APIs

The following are the REST operations used for Amazon S3 inventory.

* [DELETE Bucket inventory configuration](http://docs.aws.amazon.com/AmazonS3/latest/API/RESTBucketDELETEInventoryConfiguration.html)
* [GET Bucket inventory configuration](http://docs.aws.amazon.com/AmazonS3/latest/API/RESTBucketGETInventoryConfig.html)
* [List Bucket Inventory Configuration](http://docs.aws.amazon.com/AmazonS3/latest/API/RESTBucketListInventoryConfigs.html)
* [PUT Bucket inventory configuration](http://docs.aws.amazon.com/AmazonS3/latest/API/RESTBucketPUTInventoryConfig.html)