Use search integration service

StorageGRID

NetApp
May 03, 2022

Table of Contents

Use search integration service

The search integration service is one of the three StorageGRID platform services. You can enable this service to send object metadata to a destination search index whenever an object is created, deleted, or its metadata or tags are updated.

You can configure search integration by using the Tenant Manager to apply custom StorageGRID configuration XML to a bucket.

> Because the search integration service causes object metadata to be sent to a destination, its configuration XML is referred to as *metadata notification configuration XML*. This configuration XML is different than the *notification configuration XML* used to enable event notifications.

See the instructions for implementing S3 client applications for details about the following custom StorageGRID S3 REST API operations:

- DELETE Bucket metadata notification configuration request
- GET Bucket metadata notification configuration request
- PUT Bucket metadata notification configuration request

Related information

Configuration XML for search integration

Object metadata included in metadata notifications

JSON generated by search integration service

Configure search integration service

Use S3

Configuration XML for search integration

The search integration service is configured using a set of rules contained within `<MetadataNotificationConfiguration>` and `</MetadataNotificationConfiguration>` tags. Each rule specifies the objects that the rule applies to, and the destination where StorageGRID should send those objects' metadata.

Objects can be filtered on the prefix of the object name. For example, you could send metadata for objects with the prefix `images` to one destination, and metadata for objects with the prefix `videos` to another. Configurations that have overlapping prefixes are not valid, and are rejected when they are submitted. For example, a configuration that includes one rule for objects with the prefix `test` and a second rule for objects with the prefix `test2` is not allowed.

Destinations must be specified using the URN of a StorageGRID endpoint that has been created for the search integration service. These endpoints refer to an index and type defined on an Elasticsearch cluster.

```
<MetadataNotificationConfiguration>
    <Rule>
        <ID>Rule-1</ID>
        <Status>rule-status</Status>
        <Prefix>key-prefix</Prefix>
        <Destination>
           <Urn>arn:aws:es:region:account-
ID:domain/mydomain/myindex/mytype</Urn>
        </Destination>
    </Rule>
    <Rule>
        <ID>Rule-2</ID>
         ...
    </Rule>
     ...
</MetadataNotificationConfiguration>
```

The table describes the elements in the metadata notification configuration XML.

Name	Description	Required
MetadataNotificationConfiguration	Container tag for rules used to specify the objects and destination for metadata notifications. Contains one or more Rule elements.	Yes
Rule	Container tag for a rule that identifies the objects whose metadata should be added to a specified index. Rules with overlapping prefixes are rejected. Included in the MetadataNotificationConfiguration element.	Yes
ID	Unique identifier for the rule. Included in the Rule element.	No
Status	Status can be 'Enabled' or 'Disabled'. No action is taken for rules that are disabled. Included in the Rule element.	Yes

Name	Description	Required
Prefix	Objects that match the prefix are affected by the rule, and their metadata is sent to the specified destination. To match all objects, specify an empty prefix. Included in the Rule element.	Yes
Destination	Container tag for the destination of a rule. Included in the Rule element.	Yes
Urn	URN of the destination where object metadata is sent. Must be the URN of a StorageGRID endpoint with the following properties: • `es` must be the third element. • The URN must end with the index and type where the metadata is stored, in the form `domain-name/myindex/mytype`. Endpoints are configured using the Tenant Manager or Tenant Management API. They take the following form: • `arn:aws:es:region:account-ID:domain/mydomain/myindex/mytype` • `urn:mysite:es:::mydomain/myindex/mytype` The endpoint must be configured before the configuration XML is submitted, or configuration will fail with a 404 error. URN is included in the Destination element.	Yes

Use the sample metadata notification configuration XML to learn how to construct your own XML.

Metadata notification configuration that applies to all objects

In this example, object metadata for all objects is sent to the same destination.

```
<MetadataNotificationConfiguration>
    <Rule>
        <ID>Rule-1</ID>
        <Status>Enabled</Status>
        <Prefix></Prefix>
        <Destination>
           <Urn>urn:myes:es:::sgws-notifications/test1/all</Urn>
        </Destination>
    </Rule>
</MetadataNotificationConfiguration>
```

Metadata notification configuration with two rules

In this example, object metadata for objects that match the prefix `/images` is sent to one destination, while object metadata for objects that match the prefix `/videos` is sent to a second destination.

```
<MetadataNotificationConfiguration>
    <Rule>
        <ID>Images-rule</ID>
        <Status>Enabled</Status>
        <Prefix>/images</Prefix>
        <Destination>
           <Urn>arn:aws:es:us-east-1:3333333:domain/es-
domain/graphics/imagetype</Urn>
        </Destination>
    </Rule>
    <Rule>
        <ID>Videos-rule</ID>
        <Status>Enabled</Status>
        <Prefix>/videos</Prefix>
        <Destination>
           <Urn>arn:aws:es:us-west-1:22222222:domain/es-
domain/graphics/videotype</Urn>
        </Destination>
    </Rule>
</MetadataNotificationConfiguration>
```

Related information

Use S3

Object metadata included in metadata notifications

JSON generated by search integration service

Configure search integration service

Configure the search integration service

The search integration service sends object metadata to a destination search index whenever an object is created, deleted, or its metadata or tags are updated.

What you'll need

- Platform services must be enabled for your tenant account by a StorageGRID administrator.
- You must have already created an S3 bucket whose contents you want to index.
- The endpoint that you intend to use as a destination for the search integration service must already exist, and you must have its URN.
- You must belong to a user group that has the Manage All Buckets or the Root Access permission, which allows you to manage the settings for all S3 buckets in your tenant account. These permissions override the permission settings in group or bucket policies when configuring the bucket using the Tenant Manager.

About this task

After you configure the search integration service for a source bucket, creating an object or updating an object's metadata or tags triggers object metadata to be sent to the destination endpoint. If you enable the search integration service for a bucket that already contains objects, metadata notifications are not automatically sent for existing objects. You must update these existing objects to ensure that their metadata is added to the destination search index.

Steps

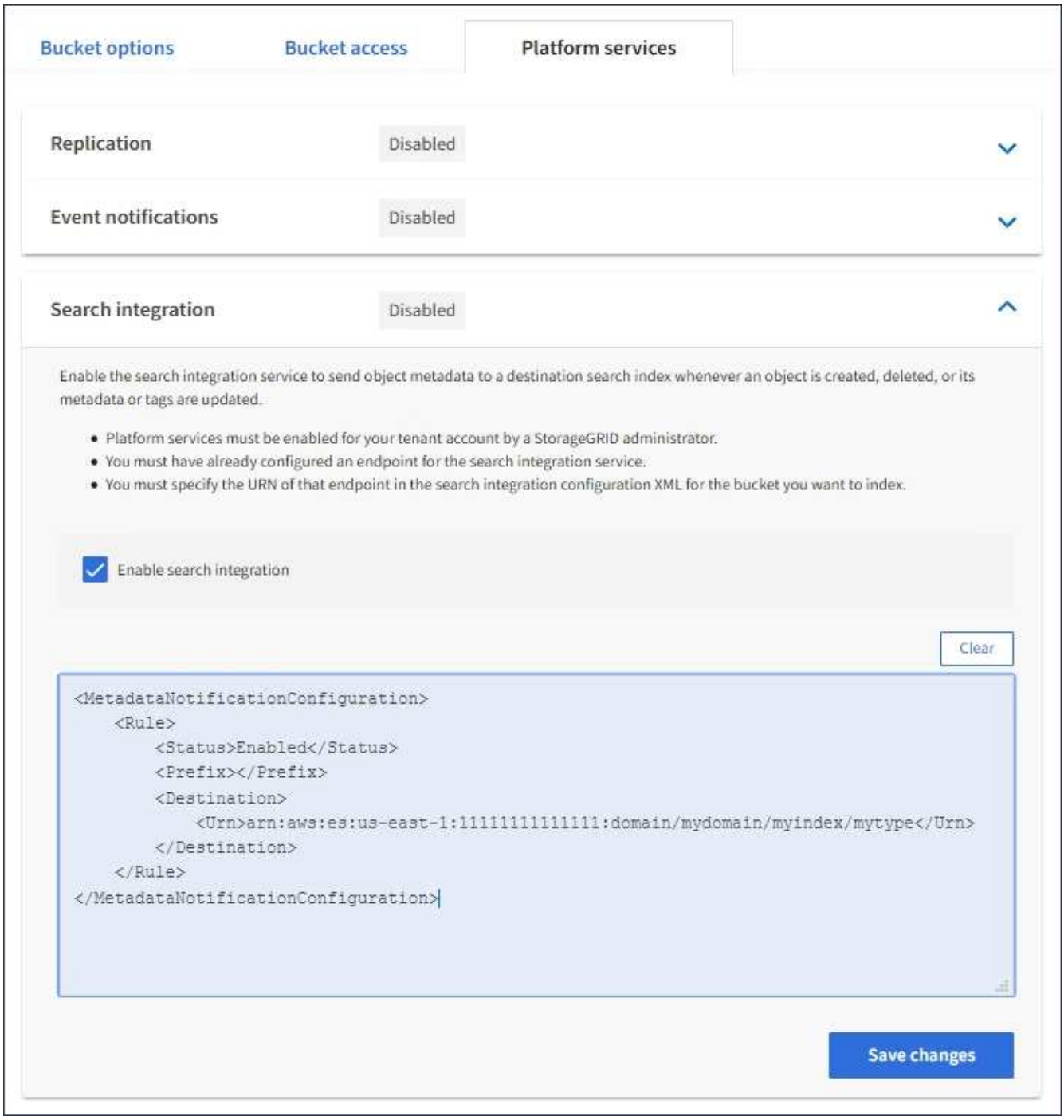
1. Use a text editor to create the metadata notification XML required to enable search integration.
   - See the information about configuration XML for search integration.
   - When configuring the XML, use the URN of a search integration endpoint as the destination.

```
<MetadataNotificationConfiguration>
 <Rule>
    <Status>Enabled</Status>
    <Prefix></Prefix>
    <Destination>
       <Urn>arn:aws:es:us-east-
1:11111111111111:domain/mydomain/myindex/mytype</Urn>
       </Destination>
 </Rule>
</MetadataNotificationConfiguration>
```

2. In the Tenant Manager select **STORAGE (S3)** > **Buckets**.
3. Select the name of the source bucket.

   The bucket details page appears.

4. Select **Platform services** > **Search integration**
5. Select the **Enable search integration** check box.
6. Paste the metadata notification configuration into the text box, and select **Save changes**.



Platform services must be enabled for each tenant account by a StorageGRID administrator using the Grid Manager or Management API. Contact your StorageGRID administrator if an error occurs when you save the configuration XML.

7. Verify that the search integration service is configured correctly:
   a. Add an object to the source bucket that meets the requirements for triggering a metadata notification as specified in the configuration XML.

      In the example shown earlier, all objects added to the bucket trigger a metadata notification.

   b. Confirm that a JSON document that contains the object's metadata and tags was added to the search index specified in the endpoint.

After you finish

As necessary, you can disable search integration for a bucket using either of the following methods:

- Select **STORAGE (S3)** > **Buckets** and unselect the **Enable search integration** check box.
- If you are using the S3 API directly, use a DELETE Bucket metadata notification request. See the instructions for implementing S3 client applications.

Related information

Understand search integration service

Configuration XML for search integration

Use S3

Create platform services endpoint

JSON generated by search integration service

When you enable the search integration service for a bucket, a JSON document is generated and sent to the destination endpoint each time object metadata or tags are added, updated, or deleted.

This example shows an example of the JSON that could be generated when an object with the key `SGWS/Tagging.txt` is created in a bucket named `test`. The `test` bucket is not versioned, so the `versionId` tag is empty.

```
{
  "bucket": "test",
  "key": "SGWS/Tagging.txt",
  "versionId": "",
  "accountId": "86928401983529626822",
  "size": 38,
  "md5": "3d6c7634a85436eee06d43415012855",
  "region":"us-east-1"
  "metadata": {
    "age": "25"
  },
  "tags": {
    "color": "yellow"
  }
}
```

Object metadata included in metadata notifications

The table lists all the fields that are included in the JSON document that is sent to the destination endpoint when search integration is enabled.

The document name includes the bucket name, object name, and version ID if present.

Type	Item name and description
Bucket and object information	`bucket`: Name of the bucket
	`key`: Object key name
	`versionID`: Object version, for objects in versioned buckets
	`region`: Bucket region, for example `us-east-1`
System metadata	`size`: Object size (in bytes) as visible to an HTTP client
	`md5`: Object hash
User metadata	`metadata`: All user metadata for the object, as key-value pairs `key:value`
Tags	`tags`: All object tags defined for the object, as key-value pairs `key:value`

> For tags and user metadata, StorageGRID passes dates and numbers to Elasticsearch as strings or as S3 event notifications. To configure Elasticsearch to interpret these strings as dates or numbers, follow the Elasticsearch instructions for dynamic field mapping and for mapping date formats. You must enable the dynamic field mappings on the index before you configure the search integration service. After a document is indexed, you cannot edit the document's field types in the index.