Remote_Reindex

Migration method

One of the easiest ways to migrate between clusters is to perform a remote reindex. This function extracts the document source from the source index and indexes the documents into the destination index. You can copy all documents to the destination index or just reindex a subset of the documents.

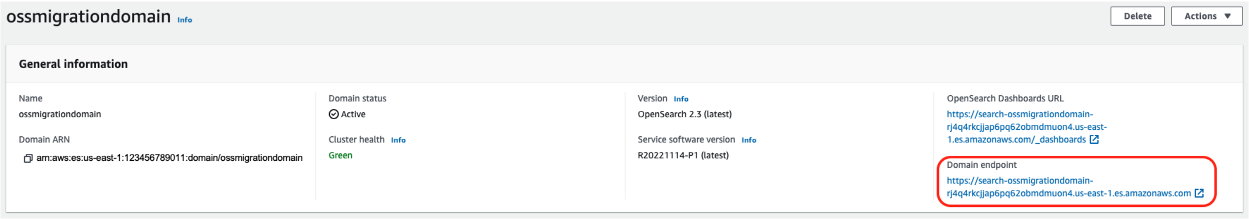
Using the remote reindex function allows you to extract from indices in an Amazon OpenSearch Service domain to an Elastic cluster. You can also migrate indices from any self-managed OpenSearch cluster to Elastic clusters.

The next section describes how you can reindex an Amazon OpenSearch Service domain remotely to an Elastic Cloud cluster. The first section documents the steps that can be followed if your OpenSearch Domain is configured to use Public Access. If your OpenSearch Domain is configured for VPC Access, it is not possible for Elastic Cloud to directly access it and, therefore, you will need to follow the steps in the second section to prepare a new Domain with Public Access.

If your cluster is deployed using self-managed open source OpenSearch, then the required information will still be the same, but you will need to find that information from the initial configuration that was deployed.

OpenSearch configured with Public Access

1.1 - The first piece of information you will need is the domain endpoint address. This can be found under General Information of the Amazon OpenSearch Service Domain in the AWS console:



ossmigrationdomain

General information

Name
ossmigrationdomain

Domain ARN
arn:aws:es:us-east-1:123456789011:domain/ossmigrationdomain

Domain status
Active

Cluster health
Green

Version
OpenSearch 2.3 (latest)

Service software version
R20221114-P1 (latest)

OpenSearch Dashboards URL
https://search-ossmigrationdomain-rj4q4rkcjjap6pq62obmdmuon4.us-east-1.es.amazonaws.com/_dashboards

Domain endpoint
https://search-ossmigrationdomain-rj4q4rkcjjap6pq62obmdmuon4.us-east-1.es.amazonaws.com

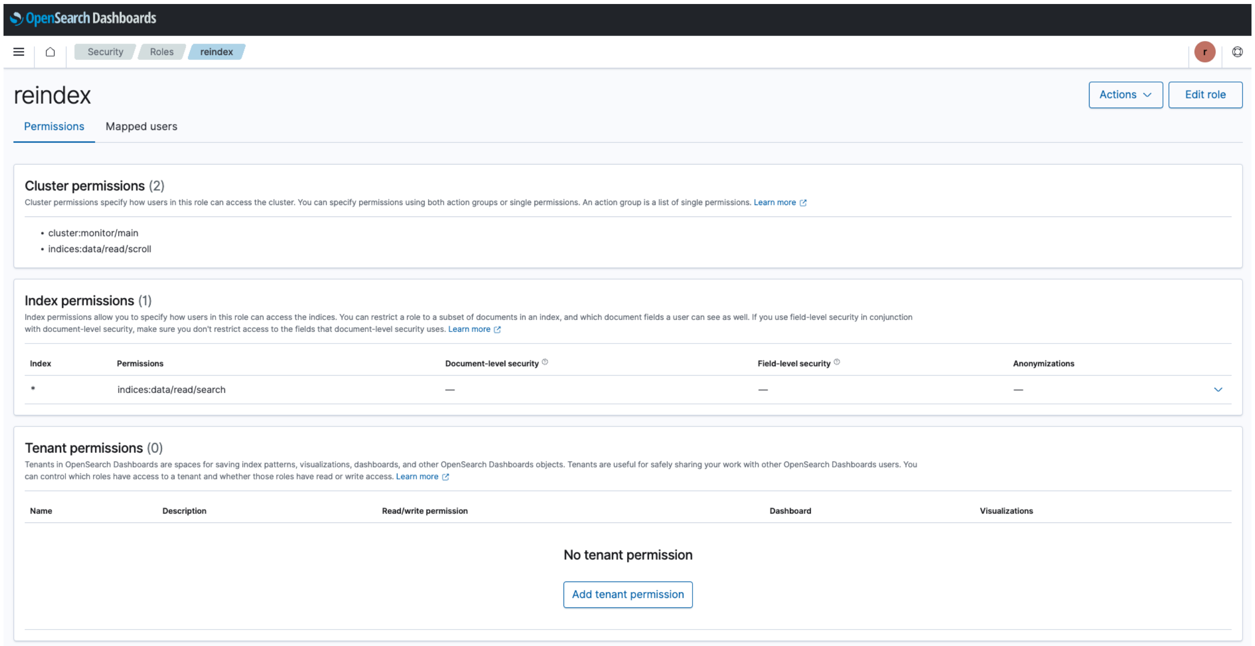
1.2 - In addition to the Domain endpoint address, you will also need the username and password of a user mapped to a role with the following permissions. In the example below, we have called this role “reindex” and mapped our user “elastic”:

Cluster Permissions:

- cluster:monitor/main
- indices:data/read/scroll

Index Permissions (for the indices you wish to reindex or “*” to include all indices):

- indices:data/read/search



reindex

Cluster permissions (2)

Cluster permissions specify how users in this role can access the cluster. You can specify permissions using both action groups or single permissions. An action group is a list of single permissions. Learn more

- cluster:monitor/main
- indices:data/read/scroll

Index permissions (1)

Index permissions allow you to specify how users in this role can access the indices. You can restrict a role to a subset of documents in an index, and which document fields a user can see as well. If you use field-level security in conjunction with document-level security, make sure you don't restrict access to the fields that document-level security uses. Learn more

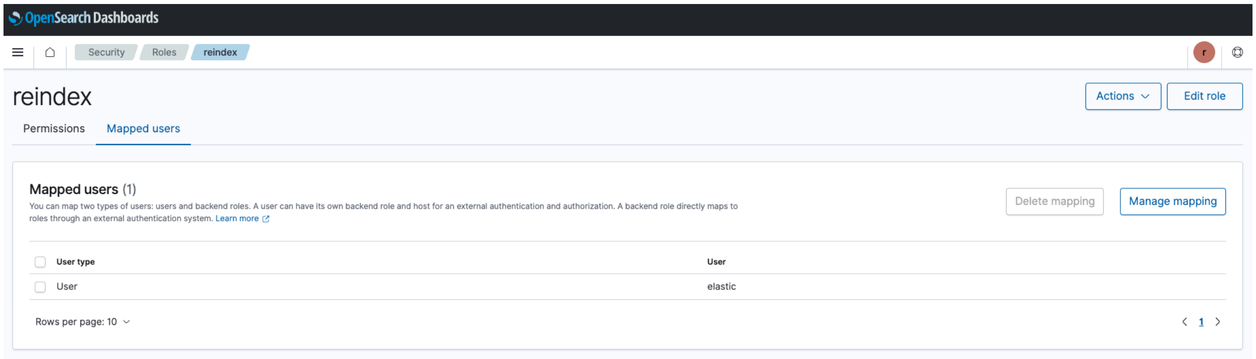
Index	Permissions	Document-level security	Field-level security	Anonymizations
*	indices:data/read/search	—	—	—

Tenant permissions (0)

Tenants in OpenSearch Dashboards are spaces for saving index patterns, visualizations, dashboards, and other OpenSearch Dashboards objects. Tenants are useful for safely sharing your work with other OpenSearch Dashboards users. You can control which roles have access to a tenant and whether those roles have read or write access. Learn more

Name	Description	Read/write permission	Dashboard	Visualizations

No tenant permission



reindex

Mapped users (1)

You can map two types of users: users and backend roles. A user can have its own backend role and host for an external authentication and authorization. A backend role directly maps to roles through an external authentication system. Learn more

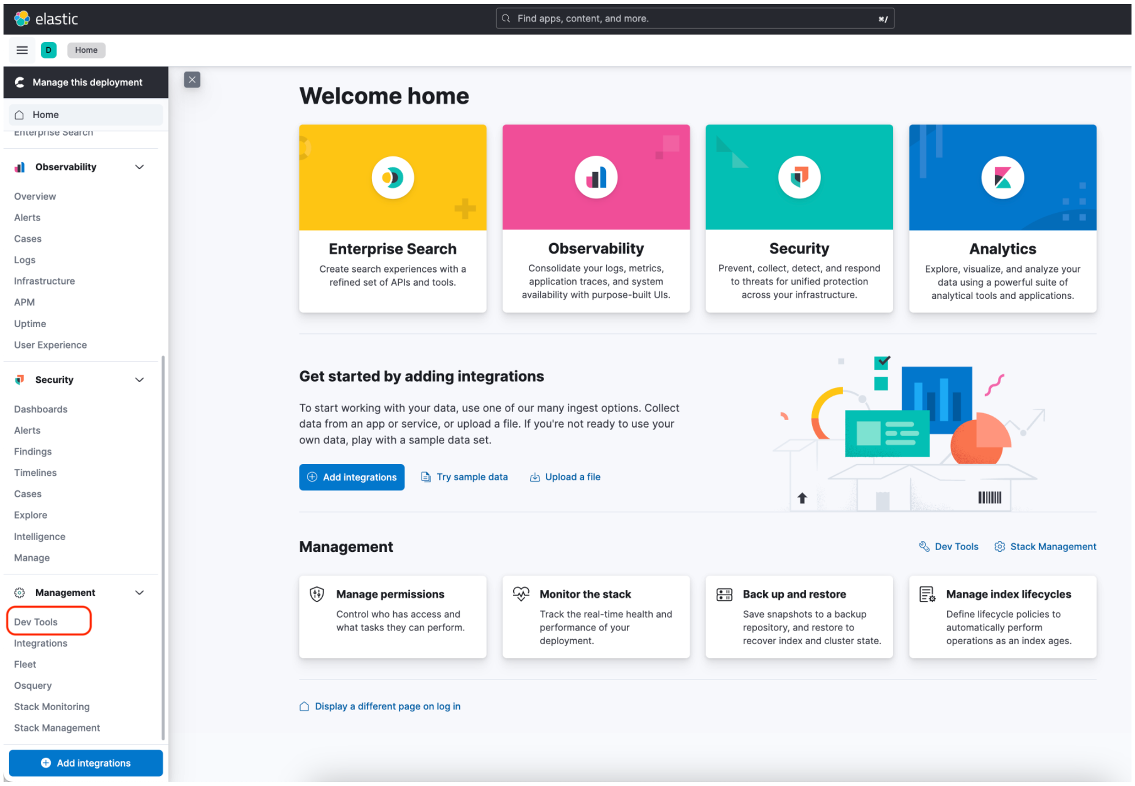
User type	User
□ User	elastic

Rows per page: 10

Now you have all the information you will require. Ensure that the target Elastic cluster has access to the OpenSearch domain endpoint through port 443.

You can now log in to your Elastic cluster.

1.3 - In Kibana, go to DevTools.



Create search experiences with a refined set of APIs and tools.

Consolidate your logs, metrics, application traces, and system availability with purpose-built UIs.

Prevent, collect, detect, and respond to threats for unified protection across your infrastructure.

Explore, visualize, and analyze your data using a powerful suite of analytical tools and applications.

To start working with your data, use one of our many ingest options. Collect data from an app or service, or upload a file. If you're not ready to use your own data, play with a sample data set.

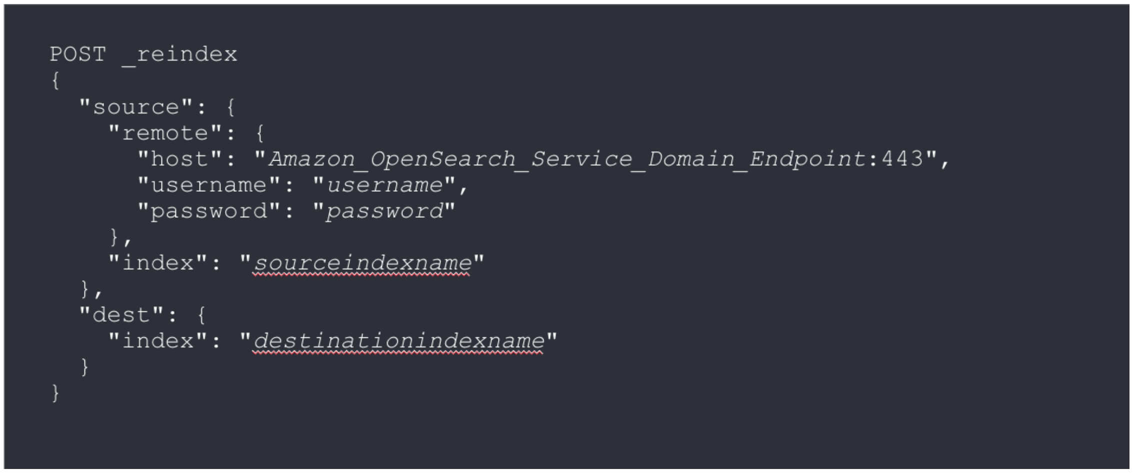
Control who has access and what tasks they can perform.

Track the real-time health and performance of your deployment.

Save snapshots to a backup repository, and restore to recover index and cluster state.

Define lifecycle policies to automatically perform operations as an index ages.

Use the following code:



```
POST _reindex
{
  "source": {
    "remote": {
      "host": "Amazon_OpenSearch_Service_Domain_Endpoint:443",
      "username": "username",
      "password": "password"
    },
    "index": "sourceindexname"
  },
  "dest": {
    "index": "destinationindexname"
  }
}
```

https://github.com/rkernutt/oss_migration/blob/main/Elastic_DEVTOOLS_reindex.json

1.3.1 - Modify “host” to be the OpenSearch Domain endpoint — note that this uses port 443 (unlike Elastic which uses port 9200).

1.3.2 - Modify the “username” and “password” with the details of the user mapped to the role we created earlier.

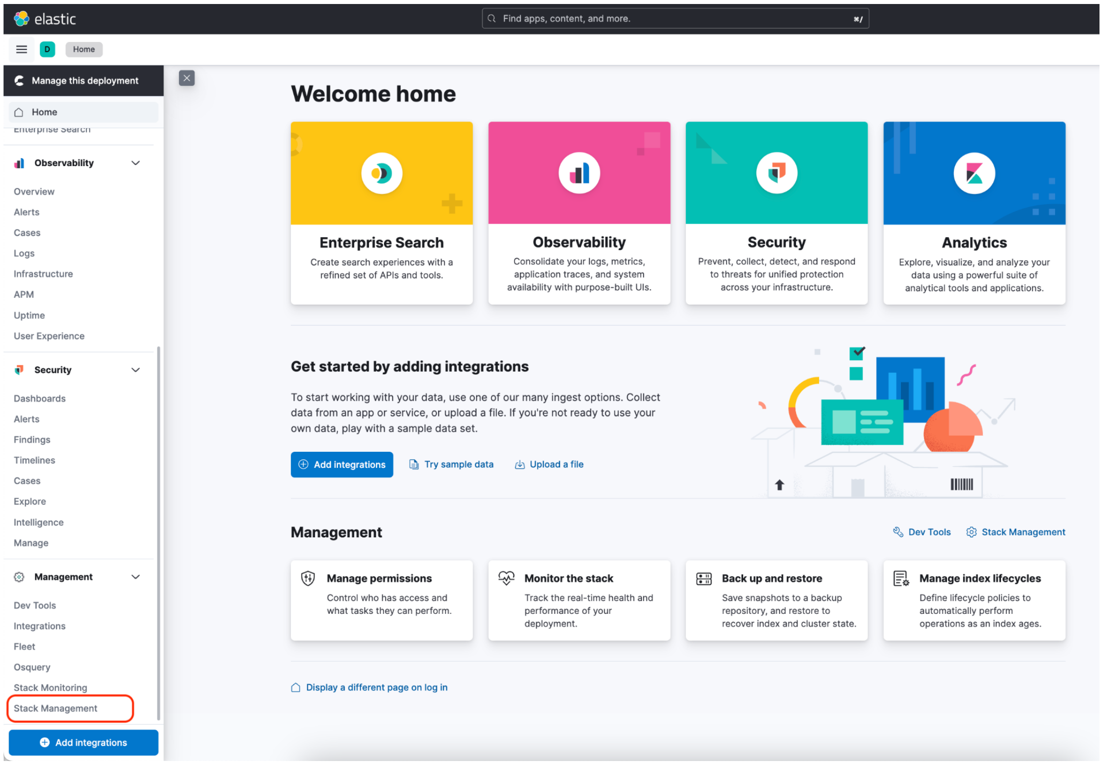
1.3.3 - For the first “index,” enter the source index name within the OpenSearch domain that you wish to reindex.

1.3.4 - Under “dest”:

- Enter the index name that you wish to create in your Elastic cluster with the ingested data.

1.4 - Once your reindex has completed, you can create a Data View and start to visualize the data.

1.4.1 - In Kibana, go to the Management menu and select “Stack Management.”



Welcome home

Enterprise Search
Create search experiences with a refined set of APIs and tools.

Observability
Consolidate your logs, metrics, application traces, and system availability with purpose-built UIs.

Security
Prevent, collect, detect, and respond to threats for unified protection across your infrastructure.

Analytics
Explore, visualize, and analyze your data using a powerful suite of analytical tools and applications.

Get started by adding integrations

To start working with your data, use one of our many ingest options. Collect data from an app or service, or upload a file. If you're not ready to use your own data, play with a sample data set.

Management

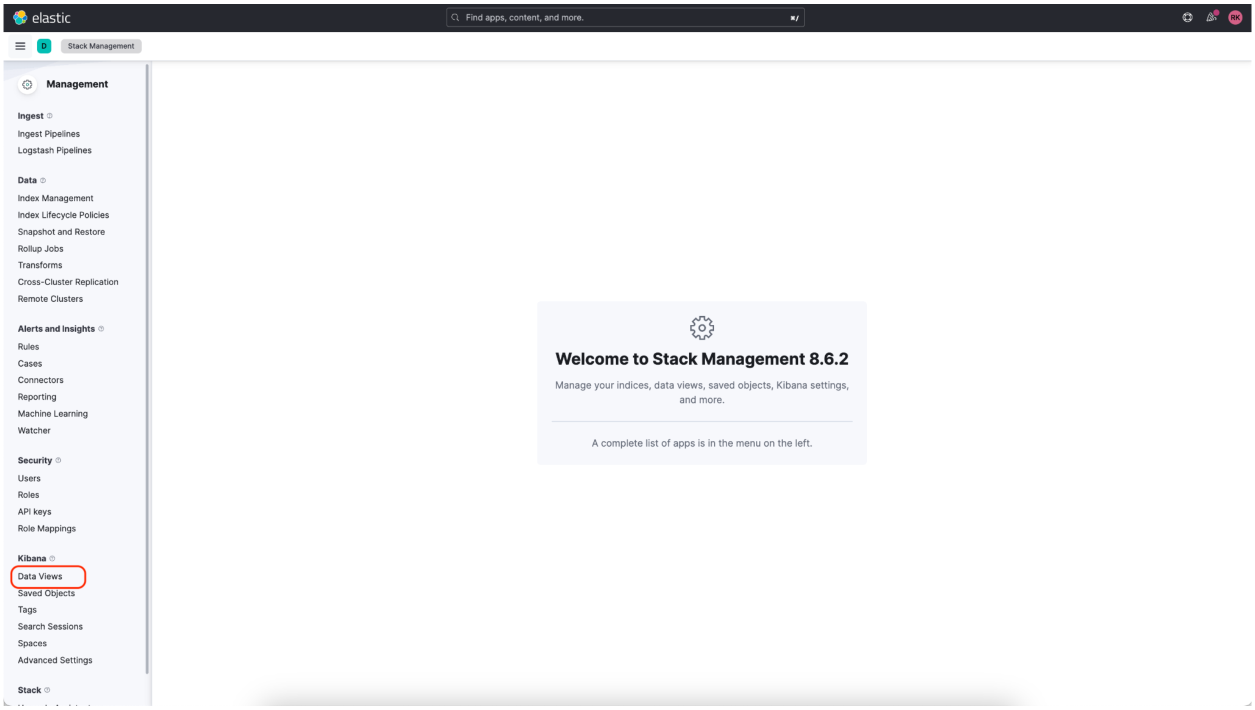
Manage permissions
Control who has access and what tasks they can perform.

Monitor the stack
Track the real-time health and performance of your deployment.

Back up and restore
Save snapshots to a backup repository, and restore to recover index and cluster state.

Manage index lifecycles
Define lifecycle policies to automatically perform operations as an index ages.

1.4.2 - Under the Kibana section, select “Data Views.”

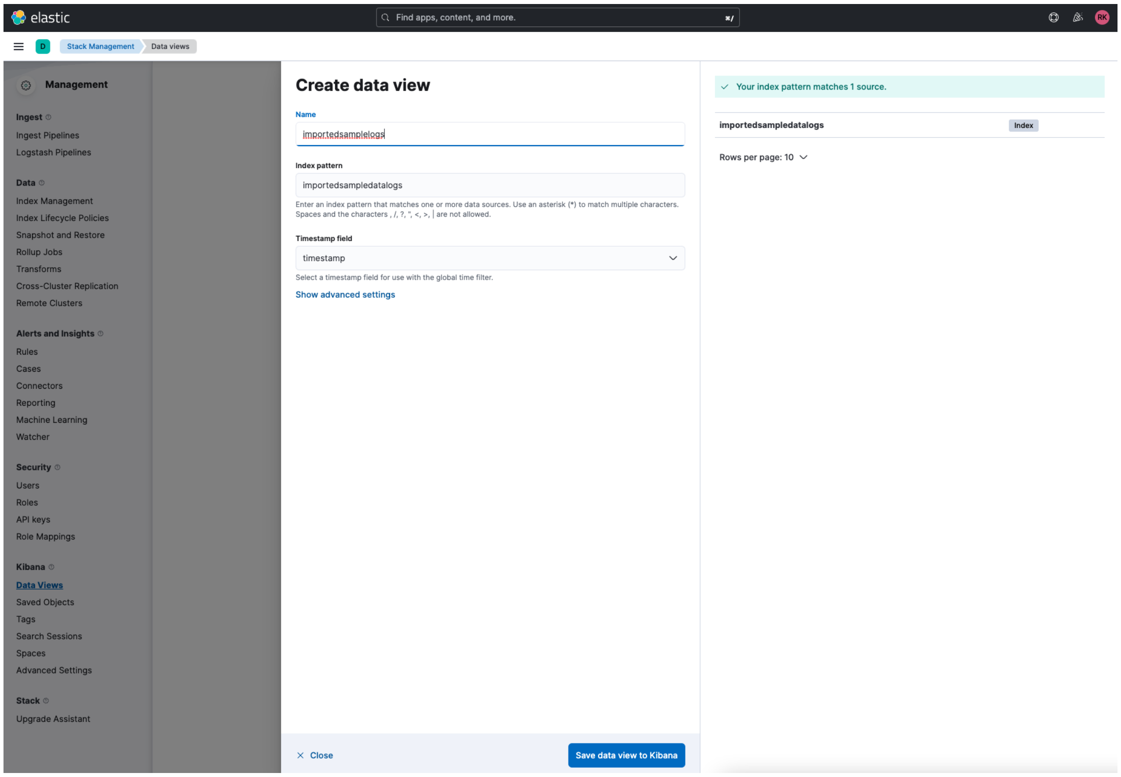


Welcome to Stack Management 8.6.2

Manage your indices, data views, saved objects, Kibana settings, and more.

A complete list of apps is in the menu on the left.

1.4.3 - Now select “Create data view” and choose the index you defined under “dest” 3.4.



Create data view

Name: importedsamplelogs

Index pattern: importedsampledatalogs

Enter an index pattern that matches one or more data sources. Use an asterisk (*) to match multiple characters. Spaces and the characters , /, ?, ", <, >, | are not allowed.

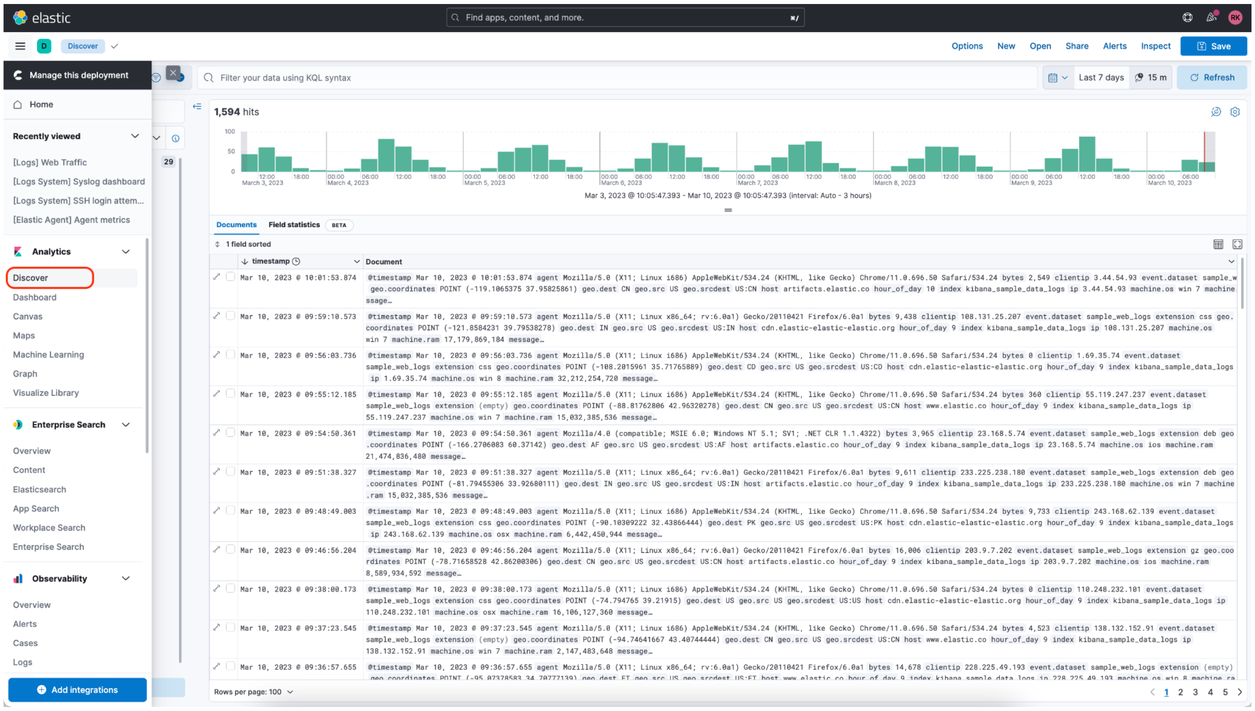
Timestamp field: timestamp

Select a timestamp field for use with the global time filter.

✓ Your index pattern matches 1 source.

importedsampledatalogs	Index

1.4.4 - Once you have created your data view, you can now use Discover to analyze your data and start to work with it.



Mar 3, 2023 @ 10:05:47.393 - Mar 10, 2023 @ 10:05:47.393 (interval: Auto - 3 hours)

```
Mar 10, 2023 @ 10:01:53.874  @timestamp Mar 10, 2023 @ 10:01:53.874 agent Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes 2,549 clientip 3.44.54.93 event.dataset sample_w
                             geo.coordinates POINT (-119.1065375 37.95825861) geo.dest CN geo.src US geo.srcdest US:CN host artifacts.elastic.co hour_of_day 10 index kibana_sample_data_logs ip 3.44.54.93 machine.os win 7 machine
                             ssage…
Mar 10, 2023 @ 09:59:10.573  @timestamp Mar 10, 2023 @ 09:59:10.573 agent Mozilla/5.0 (X11; Linux x86_64; rv:6.0a1) Gecko/20110421 Firefox/6.0a1 bytes 9,438 clientip 108.131.25.207 event.dataset sample_web_logs extension css geo.
                             coordinates POINT (-121.8584231 39.79538278) geo.dest IN geo.src US geo.srcdest US:IN host cdn.elastic-elastic-elastic.org hour_of_day 9 index kibana_sample_data_logs ip 108.131.25.207 machine.os
                             win 7 machine.ram 17,179,869,184 message…
Mar 10, 2023 @ 09:56:03.736  @timestamp Mar 10, 2023 @ 09:56:03.736 agent Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes 0 clientip 1.69.35.74 event.dataset
                             sample_web_logs extension css geo.coordinates POINT (-108.2015961 35.71765889) geo.dest CD geo.src US geo.srcdest US:CD host cdn.elastic-elastic-elastic.org hour_of_day 9 index kibana_sample_data_logs
                             ip 1.69.35.74 machine.os win 8 machine.ram 32,212,254,720 message…
Mar 10, 2023 @ 09:55:12.185  @timestamp Mar 10, 2023 @ 09:55:12.185 agent Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes 360 clientip 55.119.247.237 event.dataset
                             sample_web_logs extension (empty) geo.coordinates POINT (-88.81762806 42.96320278) geo.dest CN geo.src US geo.srcdest US:CN host www.elastic.co hour_of_day 9 index kibana_sample_data_logs ip
                             55.119.247.237 machine.os win 7 machine.ram 15,032,385,536 message…
Mar 10, 2023 @ 09:54:50.361  @timestamp Mar 10, 2023 @ 09:54:50.361 agent Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322) bytes 3,965 clientip 23.168.5.74 event.dataset sample_web_logs extension deb geo
                             .coordinates POINT (-166.2706083 60.37142) geo.dest AF geo.src US geo.srcdest US:AF host artifacts.elastic.co hour_of_day 9 index kibana_sample_data_logs ip 23.168.5.74 machine.os ios machine.ram
                             21,474,836,480 message…
Mar 10, 2023 @ 09:51:38.327  @timestamp Mar 10, 2023 @ 09:51:38.327 agent Mozilla/5.0 (X11; Linux x86_64; rv:6.0a1) Gecko/20110421 Firefox/6.0a1 bytes 9,611 clientip 233.225.238.180 event.dataset sample_web_logs extension deb geo
                             .coordinates POINT (-81.79455306 33.92680111) geo.dest IN geo.src US geo.srcdest US:IN host artifacts.elastic.co hour_of_day 9 index kibana_sample_data_logs ip 233.225.238.180 machine.os win 7 machine
                             .ram 15,032,385,536 message…
Mar 10, 2023 @ 09:48:49.003  @timestamp Mar 10, 2023 @ 09:48:49.003 agent Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes 9,733 clientip 243.168.62.139 event.dataset
                             sample_web_logs extension css geo.coordinates POINT (-90.10309222 32.43866444) geo.dest PK geo.src US geo.srcdest US:PK host cdn.elastic-elastic-elastic.org hour_of_day 9 index kibana_sample_data_logs
                             ip 243.168.62.139 machine.os osx machine.ram 6,442,450,944 message…
Mar 10, 2023 @ 09:46:56.204  @timestamp Mar 10, 2023 @ 09:46:56.204 agent Mozilla/5.0 (X11; Linux x86_64; rv:6.0a1) Gecko/20110421 Firefox/6.0a1 bytes 16,006 clientip 203.9.7.202 event.dataset sample_web_logs extension gz geo.coo
                             rdinates POINT (-78.71658528 42.86200306) geo.dest CN geo.src US geo.srcdest US:CN host artifacts.elastic.co hour_of_day 9 index kibana_sample_data_logs ip 203.9.7.202 machine.os ios machine.ram
                             8,589,934,592 message…
Mar 10, 2023 @ 09:38:00.173  @timestamp Mar 10, 2023 @ 09:38:00.173 agent Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes 0 clientip 110.248.232.101 event.dataset
                             sample_web_logs extension css geo.coordinates POINT (-74.794765 39.21915) geo.dest US geo.src US geo.srcdest US:US host cdn.elastic-elastic-elastic.org hour_of_day 9 index kibana_sample_data_logs ip
                             110.248.232.101 machine.os osx machine.ram 16,106,127,360 message…
Mar 10, 2023 @ 09:37:23.545  @timestamp Mar 10, 2023 @ 09:37:23.545 agent Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes 4,523 clientip 138.132.152.91 event.dataset
                             sample_web_logs extension (empty) geo.coordinates POINT (-94.74641667 43.40744444) geo.dest CN geo.src US geo.srcdest US:CN host www.elastic.co hour_of_day 9 index kibana_sample_data_logs ip
                             138.132.152.91 machine.os win 7 machine.ram 2,147,483,648 message…
Mar 10, 2023 @ 09:36:57.655  @timestamp Mar 10, 2023 @ 09:36:57.655 agent Mozilla/5.0 (X11; Linux x86_64; rv:6.0a1) Gecko/20110421 Firefox/6.0a1 bytes 14,678 clientip 228.225.49.193 event.dataset sample_web_logs extension (empty)
                             geo.coordinates POINT (-95.07378583 34.70777139) geo.dest ET geo.src US geo.srcdest US:ET host www.elastic.co hour_of_day 9 index kibana_sample_data_logs ip 228.225.49.193 machine.os win 8 machine.ra
```

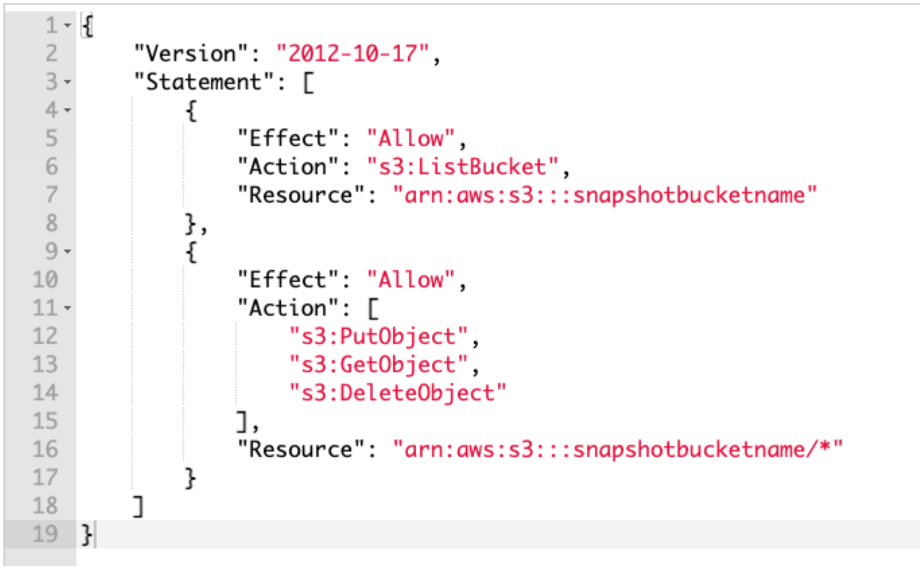
Now that these steps have all been completed, you can continue to build Dashboards and visualize your data in all the ways you require.

OpenSearch Domain with VPC Access

If your OpenSearch Domain is configured to have VPC Access, then your Elastic Cluster will not be able to communicate with the Domain endpoint. At the time of writing this blog, you are not able to change the Network settings of the Domain from VPC Access to Public Access. Therefore, the following steps will allow you to create a manual snapshot of the OpenSearch Domain and restore the snapshot to another Domain with Public Access. Once this is complete, you can follow the steps from the previous section to perform a remote reindex.

2.1 - Create a S3 bucket.

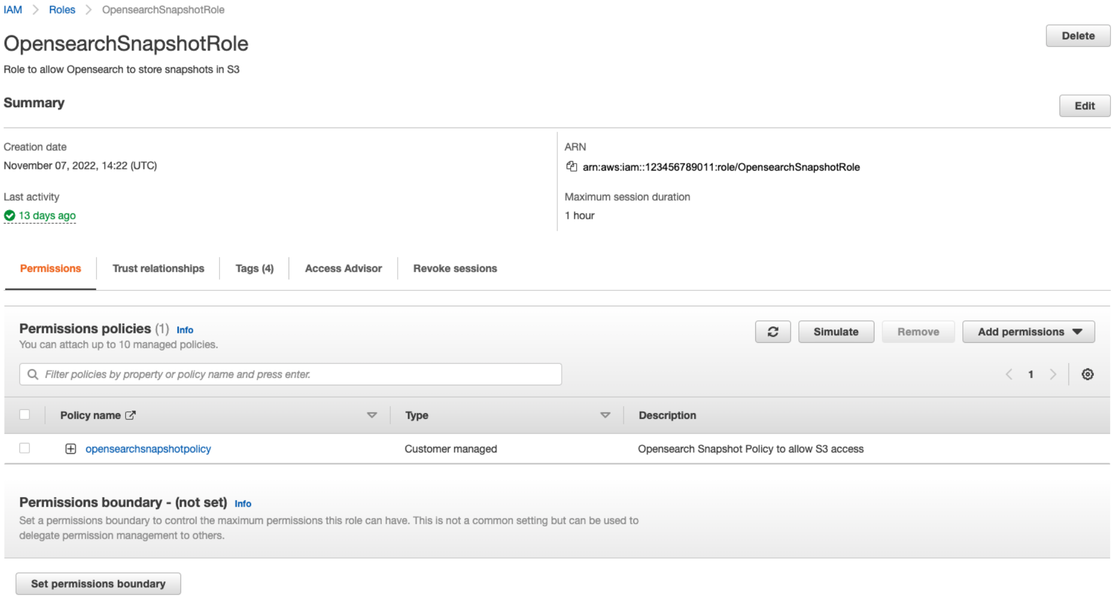
2.2 - Create the following IAM roles and policies:
"*Opensearchsnapshotpolicy*" — a policy that allows S3 access to OpenSearch to perform snapshot operations.



```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:ListBucket",
            "Resource": "arn:aws:s3:::snapshotbucketname"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": "arn:aws:s3:::snapshotbucketname/*"
        }
    ]
}
```

https://github.com/rkernutt/oss_migration/blob/main/AWSIAM_OpenSearchsnapshotpolicy.json

“*OpensearchSnapshotRole*” — A role that grants permission to OpenSearch from “*Opensearchsnapshotpolicy*” (above) and also edits the trust relationships.



OpensearchSnapshotRole

Role to allow Opensearch to store snapshots in S3

Summary

Creation date
November 07, 2022, 14:22 (UTC)

Last activity
13 days ago

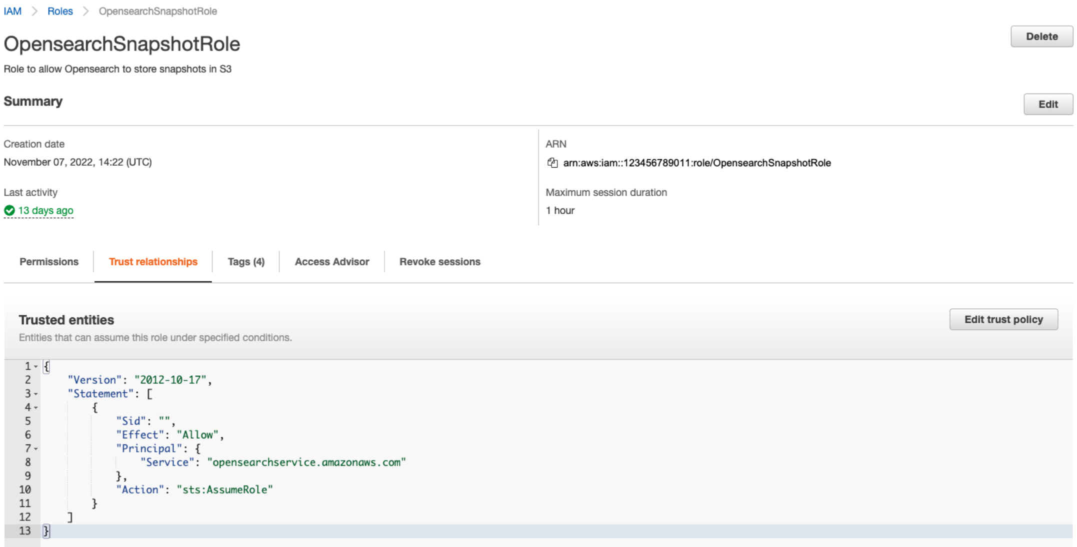
ARN
arn:aws:iam::123456789011:role/OpensearchSnapshotRole

Maximum session duration
1 hour

Permissions policies (1) Info
You can attach up to 10 managed policies.

□	Policy name	Type	Description
□	⊞ opensearchsnapshotpolicy	Customer managed	Opensearch Snapshot Policy to allow S3 access

Permissions boundary - (not set) Info
Set a permissions boundary to control the maximum permissions this role can have. This is not a common setting but can be used to delegate permission management to others.



OpensearchSnapshotRole

Role to allow Opensearch to store snapshots in S3

Summary

Creation date
November 07, 2022, 14:22 (UTC)

Last activity
13 days ago

ARN
arn:aws:iam::123456789011:role/OpensearchSnapshotRole

Maximum session duration
1 hour

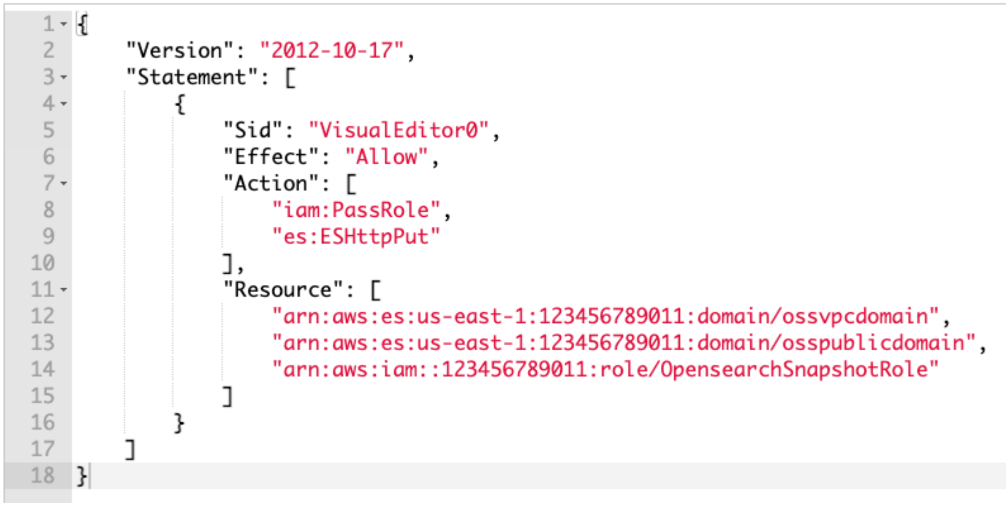
Trusted entities

Entities that can assume this role under specified conditions.

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

https://github.com/rkernutt/oss_migration/blob/main/AWSIAM_OpenSearchsnapshotroletrustedentity.json

“*OpensearchPassRolePolicy*” — A policy that allows a role to be passed to OpenSearch domains to perform snapshots.

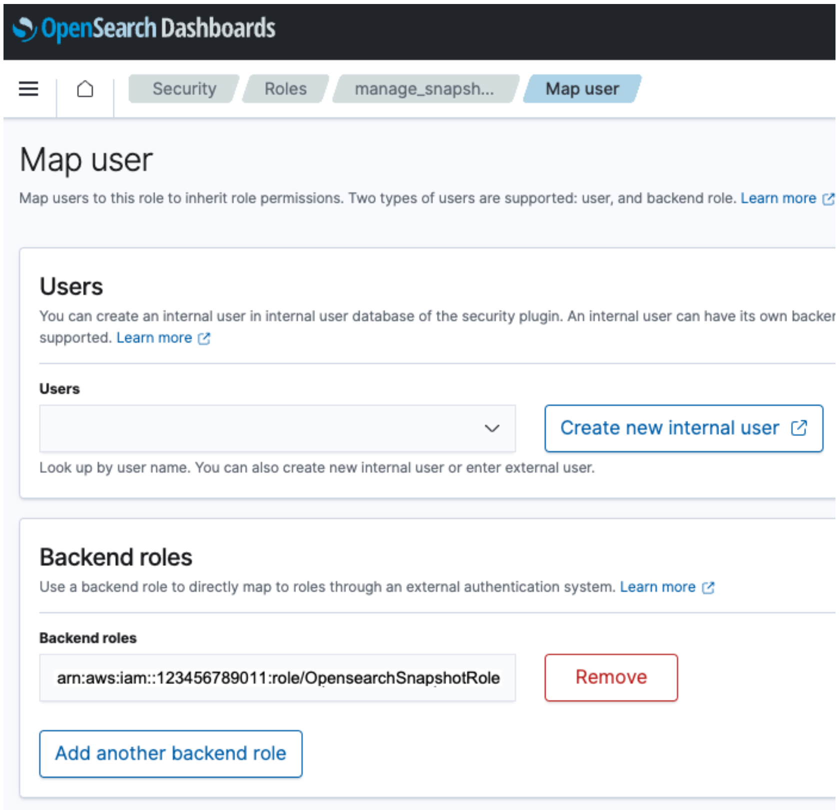


```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole",
                "es:ESHttpPut"
            ],
            "Resource": [
                "arn:aws:es:us-east-1:123456789011:domain/ossvpcdomain",
                "arn:aws:es:us-east-1:123456789011:domain/osspublicdomain",
                "arn:aws:iam::123456789011:role/OpensearchSnapshotRole"
            ]
        }
    ]
}
```

https://github.com/rkernutt/oss_migration/blob/main/AWSIAM_OpenSearchPassRolePolicy.json

2.3 - Assign this to your IAM user permissions.

2.4 - In the OpenSearch Dashboard, go to the menu under OpenSearch Plugins. Go to Security > Roles, and then edit the “manage_snapshots” role and to Mapped users add the ARN of the OpensearchSnapshotRole to the Backend roles:



Map user

Map users to this role to inherit role permissions. Two types of users are supported: user, and backend role. Learn more

Users

You can create an internal user in internal user database of the security plugin. An internal user can have its own backer supported. Learn more

Users

Look up by user name. You can also create new internal user or enter external user.

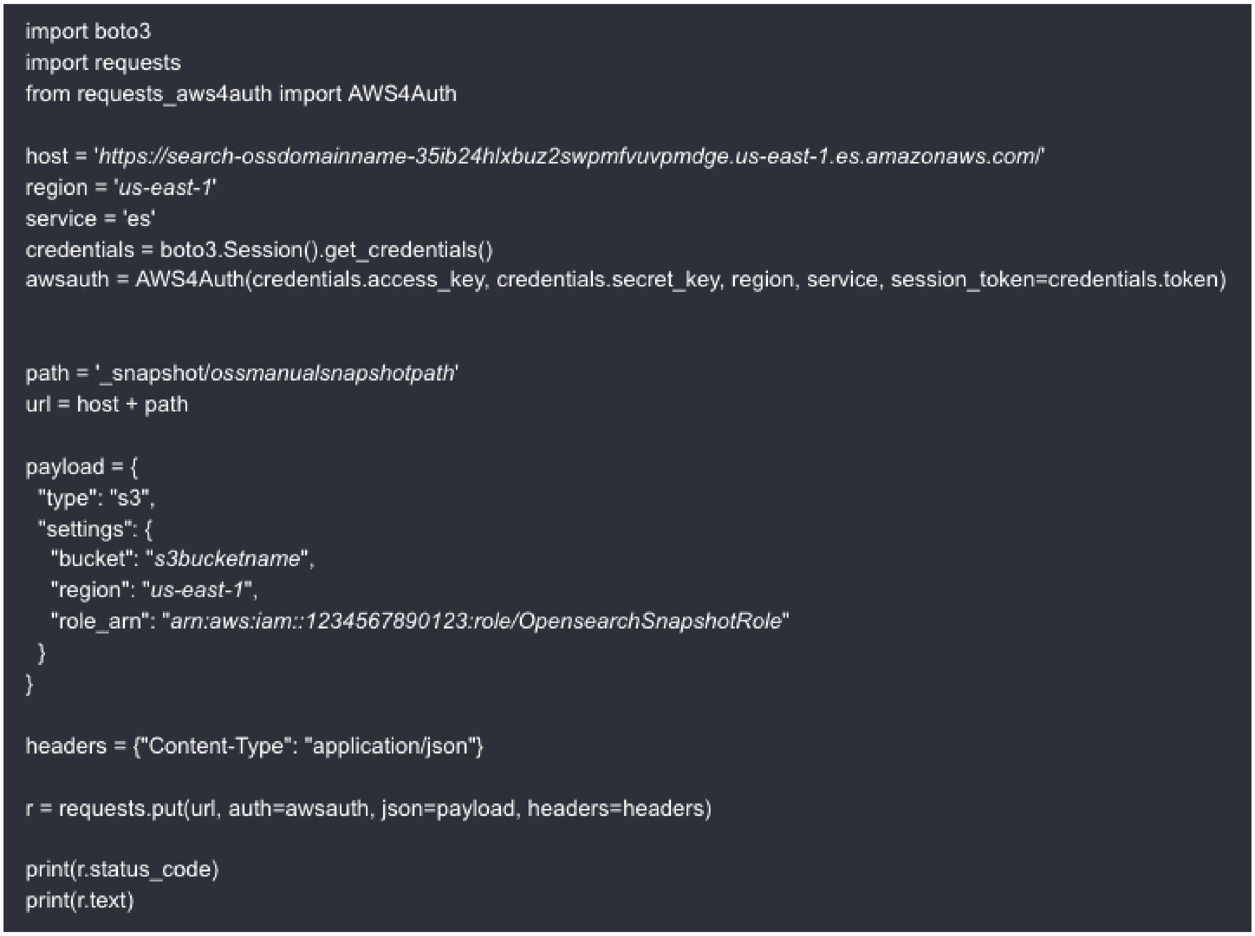
Backend roles

Use a backend role to directly map to roles through an external authentication system. Learn more

Backend roles: arn:aws:iam::123456789011:role/OpensearchSnapshotRole

2.5 - You will now need an EC2 instance with awscli, python, pip, boto3, requests, and requests_aws4auth installed.

2.6 - Register a snapshot repository:
Use python script -



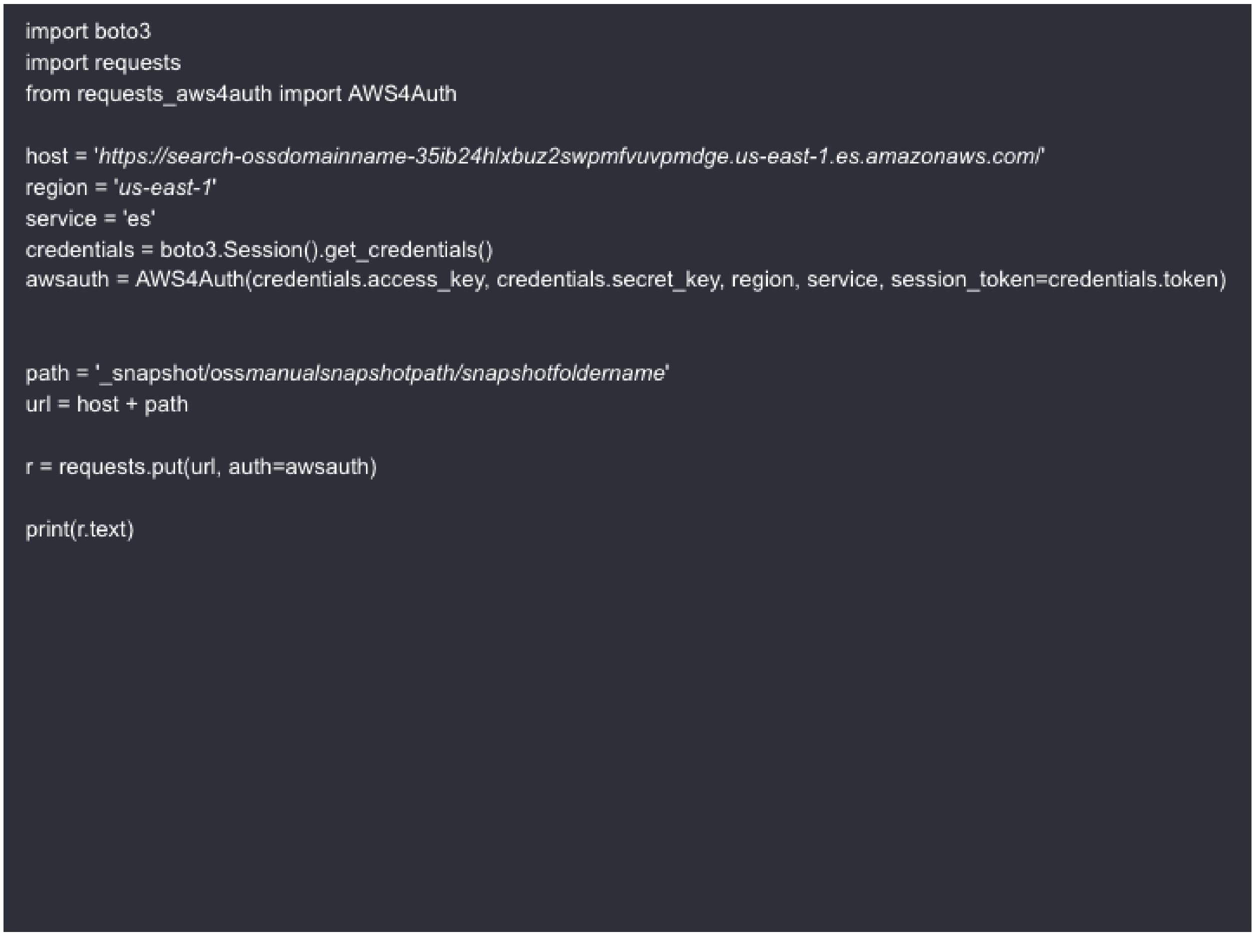
```
import boto3
import requests
from requests_aws4auth import AWS4Auth

host = 'https://search-ossdomainname-35ib24hlxbuz2swpmfvuvpmdge.us-east-1.es.amazonaws.com/'
region = 'us-east-1'
service = 'es'
credentials = boto3.Session().get_credentials()
awsauth = AWS4Auth(credentials.access_key, credentials.secret_key, region, service, session_token=credentials.token)


path = '_snapshot/ossmanualsnapshotpath'
url = host + path

payload = {
  "type": "s3",
  "settings": {
    "bucket": "s3bucketname",
    "region": "us-east-1",
    "role_arn": "arn:aws:iam::1234567890123:role/OpensearchSnapshotRole"
  }
}

headers = {"Content-Type": "application/json"}

r = requests.put(url, auth=awsauth, json=payload, headers=headers)

print(r.status_code)
print(r.text)
```

https://github.com/rkernutt/oss_migration/blob/main/registersnaprepository.py

2.6.1 - Modify host to be domain endpoint (put “/“ on the end) of OS domain.
2.6.2 - Modify region of OS domain.
2.6.3 - Modify path to be path for snapshots.
2.6.4 - Under payload:

- Modify bucket
- Modify region
- Modify role_arn to the role created

2.7 - Take a snapshot:
Use python script -



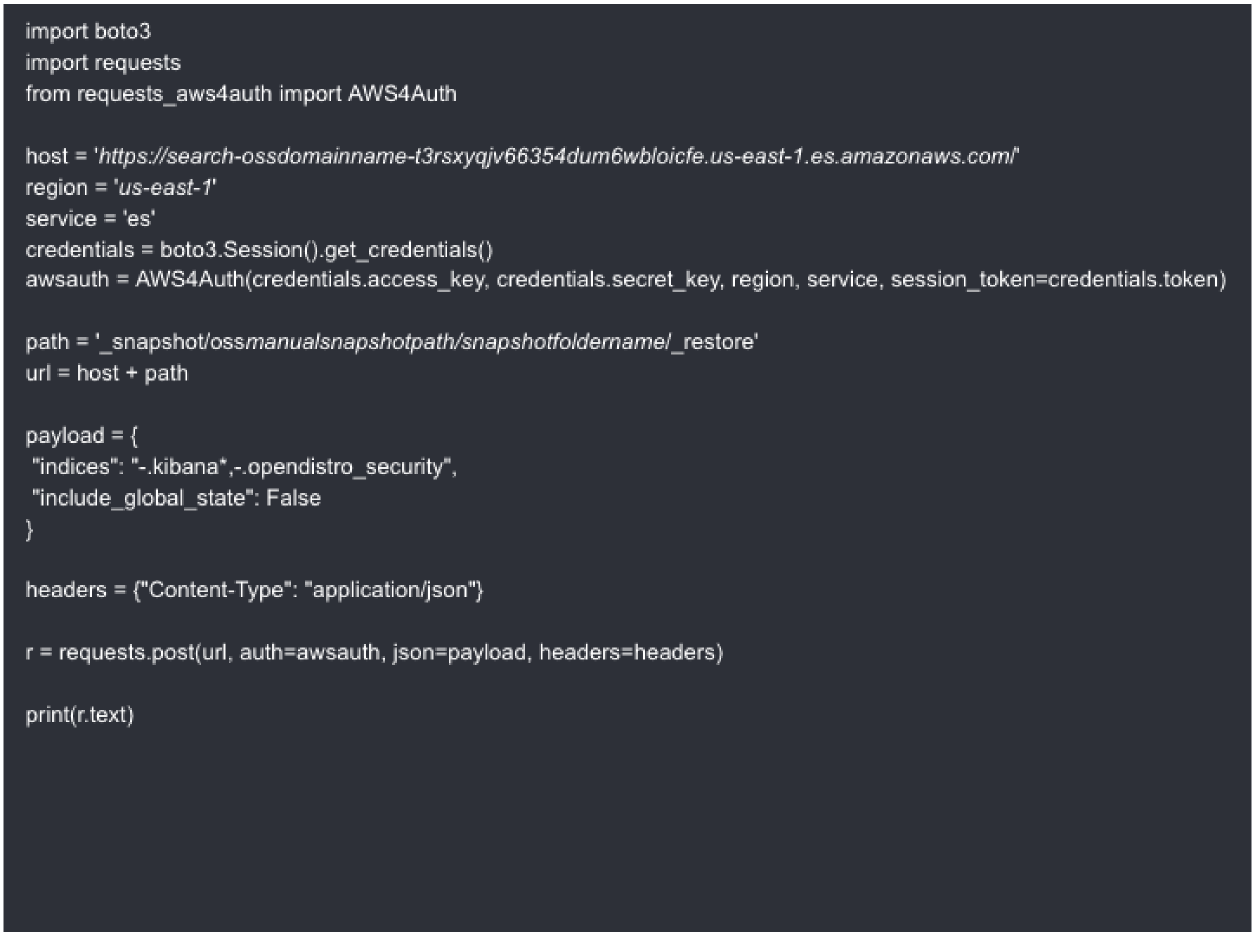
```
import boto3
import requests
from requests_aws4auth import AWS4Auth

host = 'https://search-ossdomainname-35ib24hlxbuz2swpmfvuvpmdge.us-east-1.es.amazonaws.com/'
region = 'us-east-1'
service = 'es'
credentials = boto3.Session().get_credentials()
awsauth = AWS4Auth(credentials.access_key, credentials.secret_key, region, service, session_token=credentials.token)


path = '_snapshot/ossmanualsnapshotpath/snapshotfoldername'
url = host + path

r = requests.put(url, auth=awsauth)

print(r.text)
```

https://github.com/rkernutt/oss_migration/blob/main/takesnapshot.py

2.7.1 - Modify host to be domain endpoint (put / on the end) of OS domain.
2.7.2 - Modify region of OS domain.
2.7.3 - **Modify path to be path and name for snapshot.**

2.8 - Create a new OpenSearch Domain configured with Public Access.

2.9 - Now restore the snapshot to the new OpenSearch Domain:
Use python script -



```
import boto3
import requests
from requests_aws4auth import AWS4Auth

host = 'https://search-ossdomainname-t3rsxyqjv66354dum6wbloicfe.us-east-1.es.amazonaws.com/'
region = 'us-east-1'
service = 'es'
credentials = boto3.Session().get_credentials()
awsauth = AWS4Auth(credentials.access_key, credentials.secret_key, region, service, session_token=credentials.token)

path = '_snapshot/ossmanualsnapshotpath/snapshotfoldername/_restore'
url = host + path

payload = {
 "indices": "-.kibana*,-.opendistro_security",
 "include_global_state": False
}

headers = {"Content-Type": "application/json"}

r = requests.post(url, auth=awsauth, json=payload, headers=headers)

print(r.text)
```

https://github.com/rkernutt/oss_migration/blob/main/restoresnap.py

2.9.1 - Modify host to be domain endpoint (put / on the end) of OS domain you need to restore to.
2.9.2 - Modify region of OS domain.
2.9.3 - Modify path to be path and name for snapshot you took, “/_restore” will need to be added to the end of the path.

Now if you go to Index Management under the OpenSearch Plugins menu and review the Indices, you will see the restored Indices listed. You will need these Index names when you perform the remote reindex.