The CERN's perspective on Search wars

Sokratis Papadopoulos
it-opensearch-experts@cern.ch

Overview

- Service definition & numbers
- Legacy Elasticsearch service
- OpenDistro for Elasticsearch evaluation
- OpenSearch service
- Migration from Elasticsearch/OpenDistro to OpenSearch
- Service security configuration
- Service operations
- Roadmap

What is Elasticsearch and OpenSearch?

- **Elasticsearch** is a distributed, search and analytics engine based on Apache Lucene
- **Kibana** is the web user interface that lets you visualise your Elasticsearch data

- **OpenSearch** is a fork of Elasticsearch 7.10.2 open source codebase
- **OpenSearch Dashboards** is the fork of Kibana 7.10.2 open source codebase

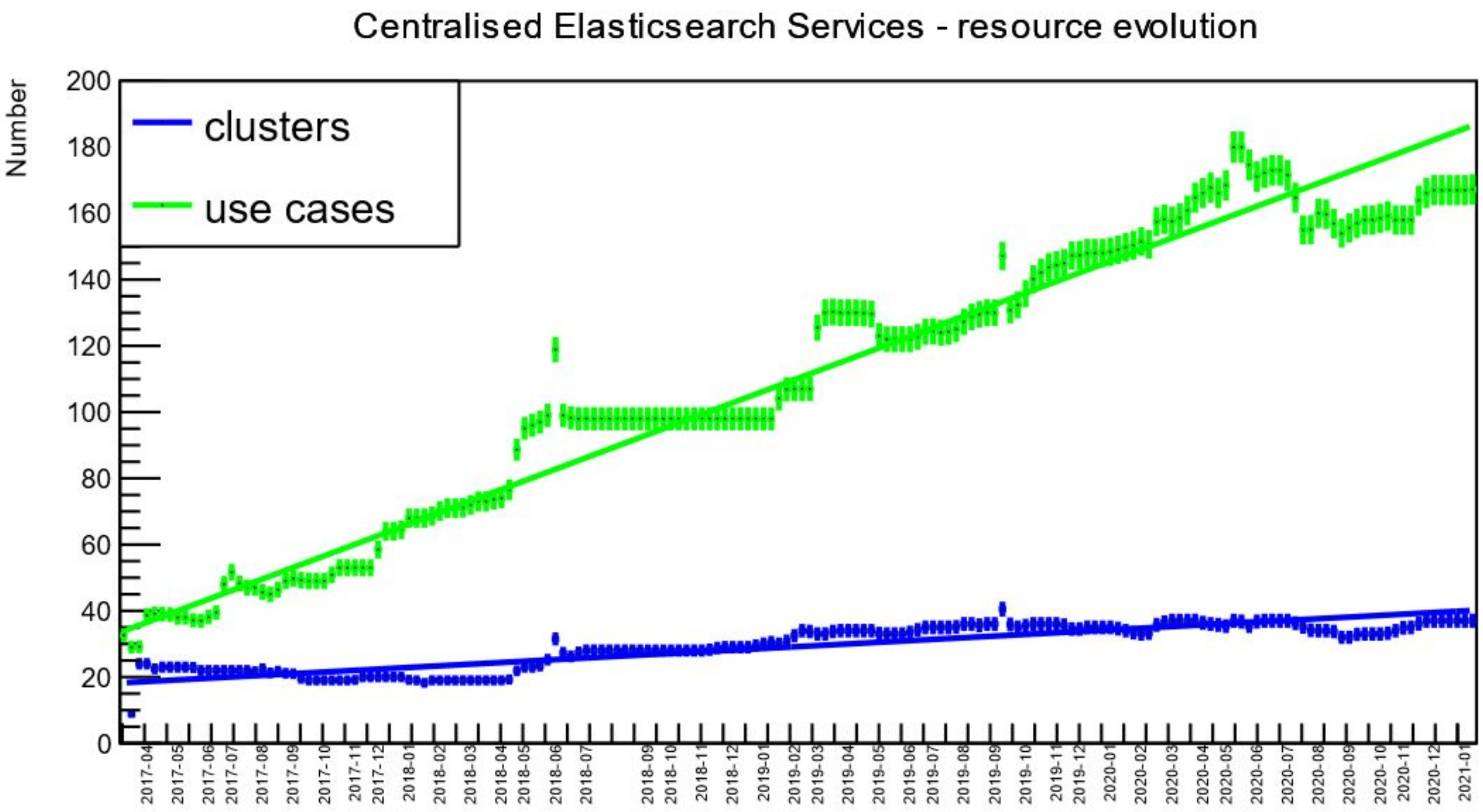
play with it here
https://playground.opensearch.org

Search timeline at CERN

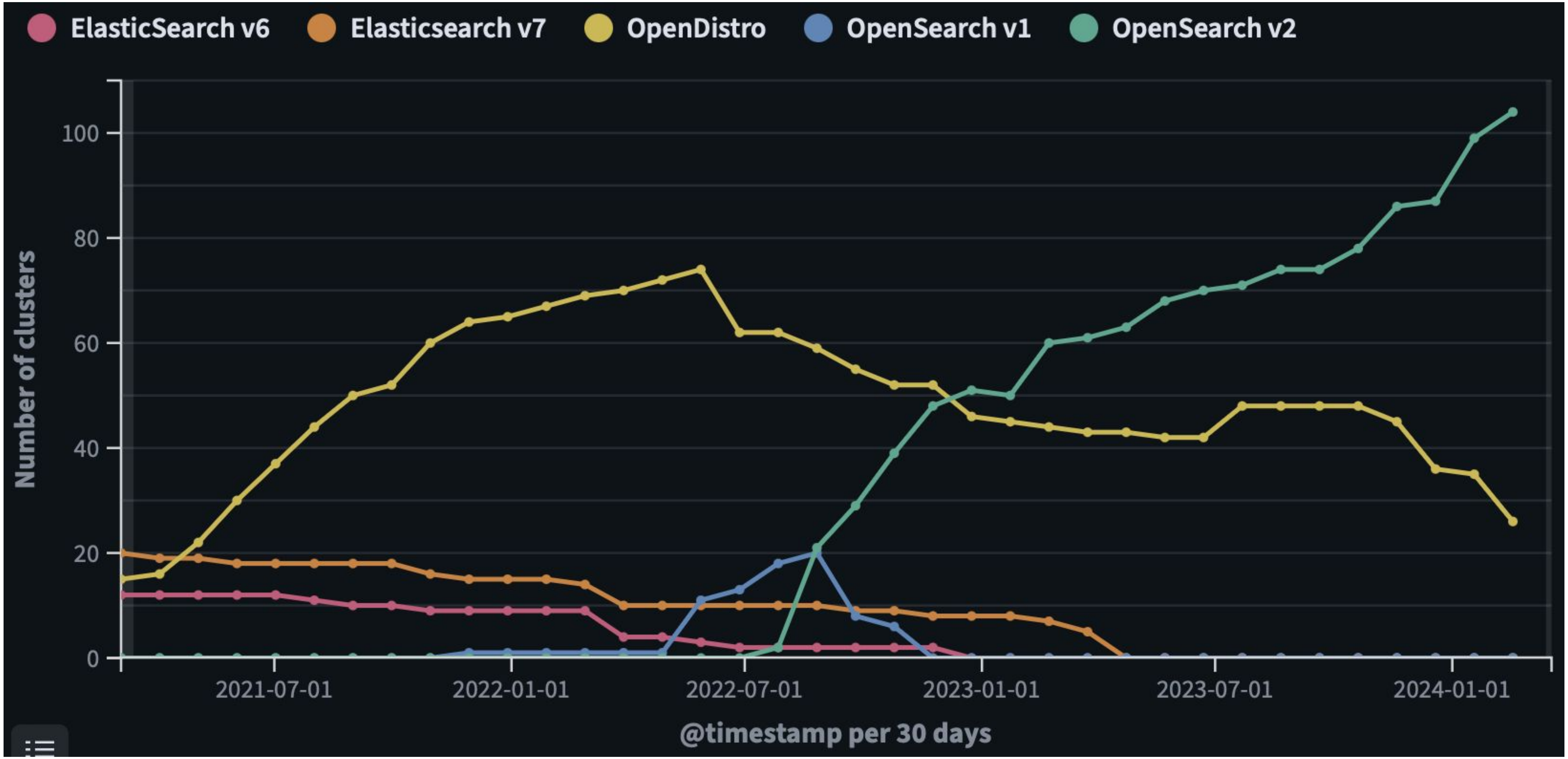
- **< 2016:** Dedicated Elasticsearch clusters all around CERN
- **2016:** Creation of Centralised Elasticsearch service, v2
- **2017:** Upgrade to Elasticsearch v5
- **2018:** Upgrade to Elasticsearch v6
- **2020-Q1:** Upgrade to Elasticsearch v7.1
- **2020-Q4:** Evaluation of OpenDistro for Elasticsearch
- **2021:** Decision to migrate towards OpenDistro
- **2022:** OpenSearch v1 is out, migration out of Elasticsearch and OpenDistro
- **2022-Q2:** Upgrade to OpenSearch v2
- **2024-Q4:** (Hopefully) all clusters on OpenSearch v2

Search timeline at CERN - Legacy service

- Centralised Elasticsearch + Kibana instances since 2016
- Approach
  - few big clusters
  - many endpoints

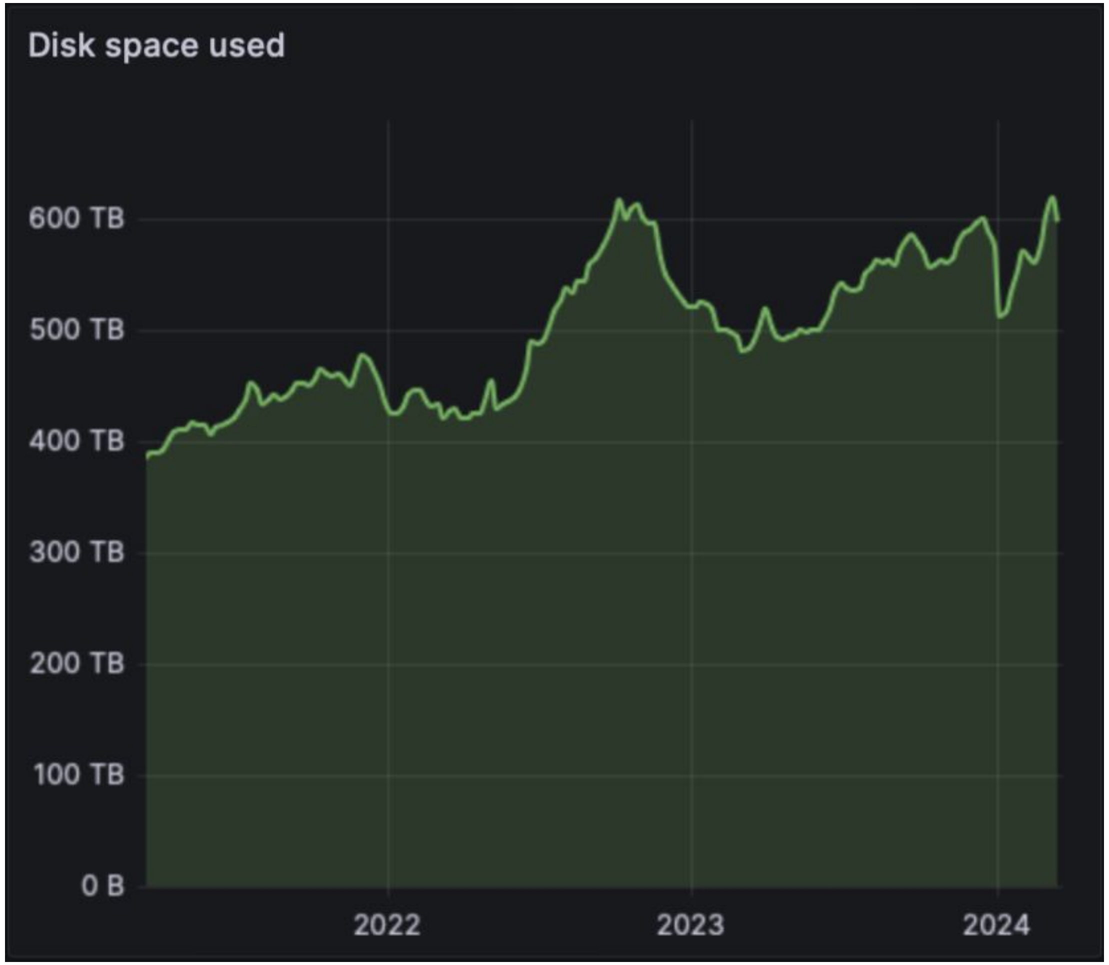


Search timeline at CERN - OpenDistro/OpenSearch

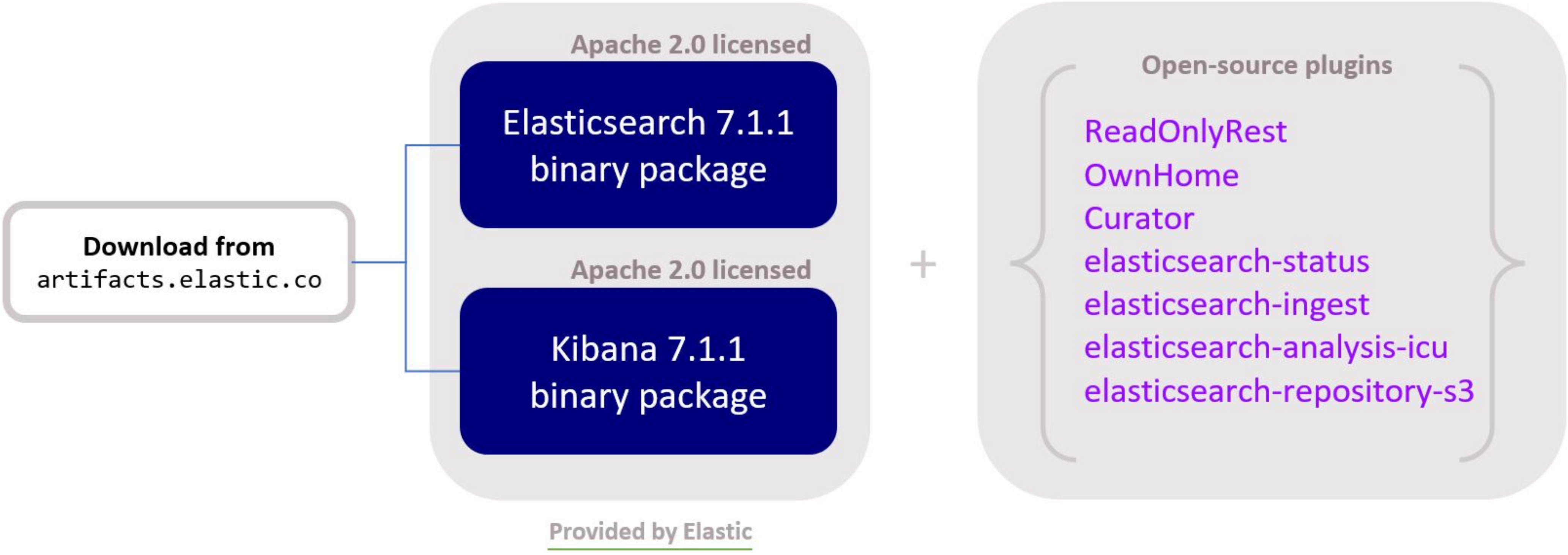


Current Elasticsearch/OpenSearch usage at CERN

- Usage at CERN & HEP
  - ALICE, ATLAS, CMS, LHCb, NA62, …
  - Beams, INSPIRE, Zenodo, CDS, …
  - IT: Monitoring, **Security**, Storage, …
- Service numbers
  - **105 OpenSearch** v2.11 + **20 OpenDistro** clusters
  - **600 TBs** indexed data ~ **1.2 trillion** docs
- Available hardware
  - 3 availability zones
  - **156** Ironic managed **physical** machines
    - 256 GB RAM - 64 cores - 10.5 TB SSD disks



Legacy service packages + plugins



Download from `artifacts.elastic.co`

Apache 2.0 licensed

Elasticsearch 7.1.1 binary package

Apache 2.0 licensed

Kibana 7.1.1 binary package

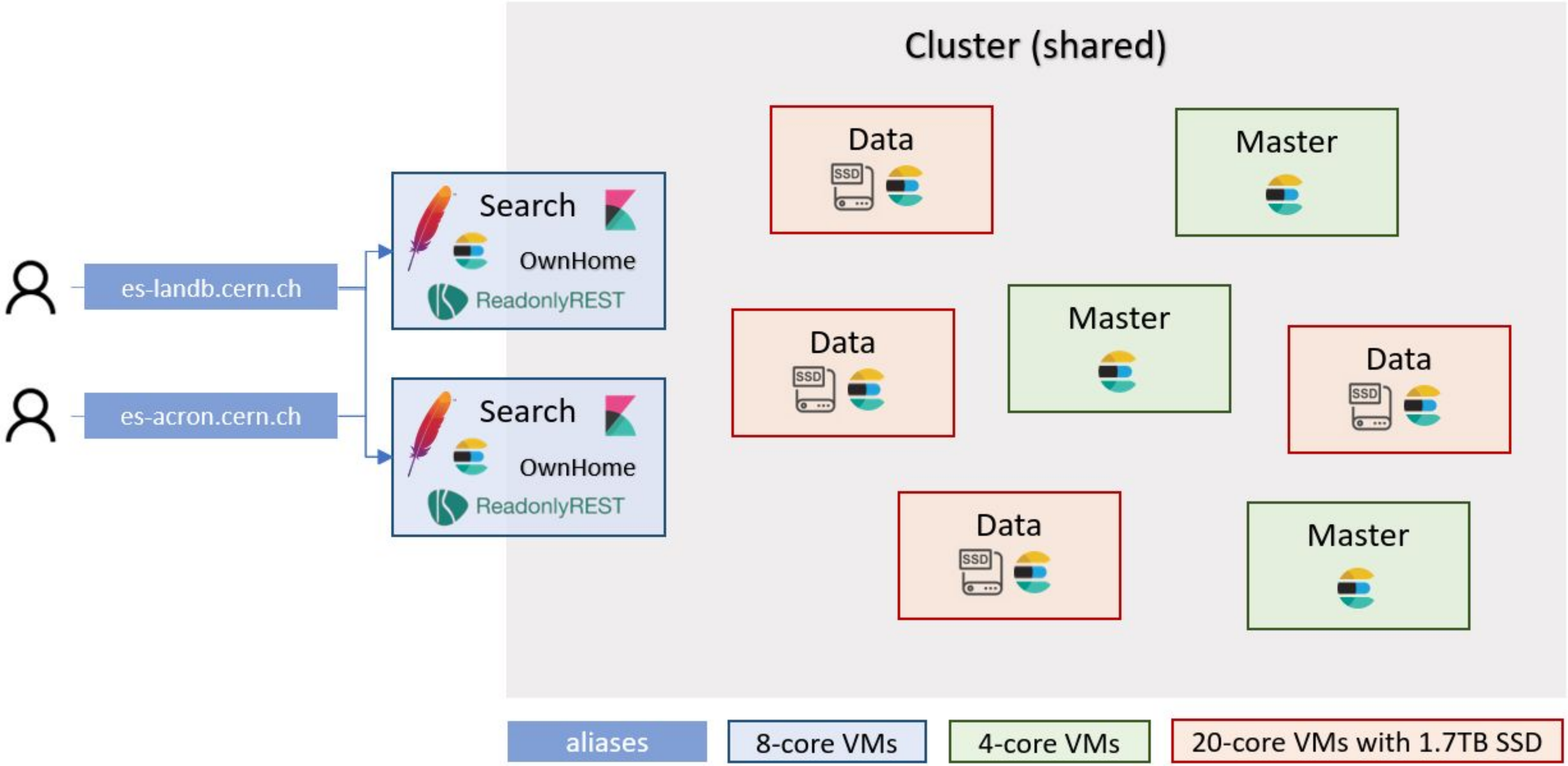
Provided by Elastic

+

Open-source plugins

- ReadOnlyRest
- OwnHome
- Curator
- elasticsearch-status
- elasticsearch-ingest
- elasticsearch-analysis-icu
- elasticsearch-repository-s3

Legacy service design



Legacy service pros & cons

Pros

- Resource sharing and optimisation
- Centrally managed and worry-free use from customers
- Fully open-source

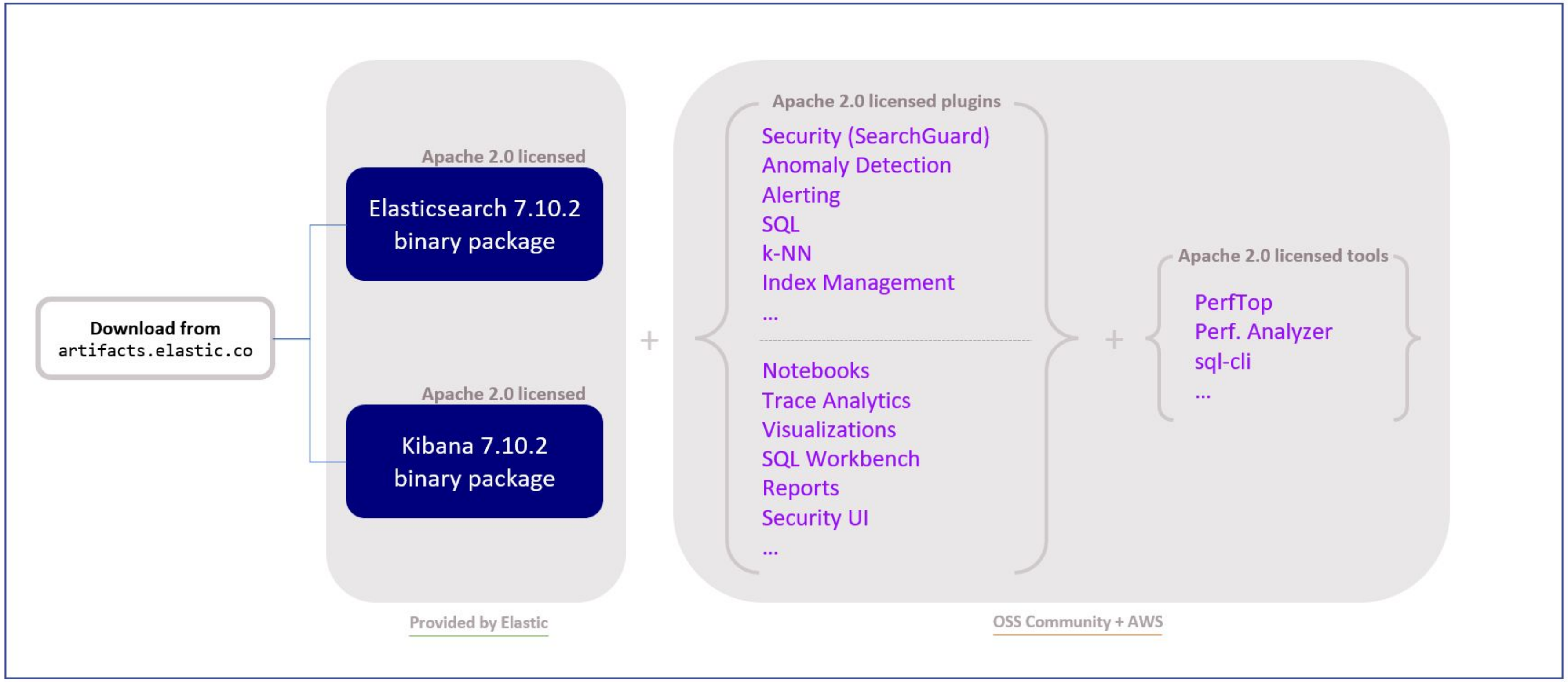
Cons

- **Maintainability**: even minor ES releases caused big issues on the external plugins
  - As a result, we started to race against EOL of versions used at CERN
- Customers isolation
  - Heavy queries of one user hurting another

Evaluation of OpenDistro for Elasticsearch

- Started in Oct 2020: a *complete* open-source Elasticsearch + Kibana product

Open Distro
for Elasticsearch



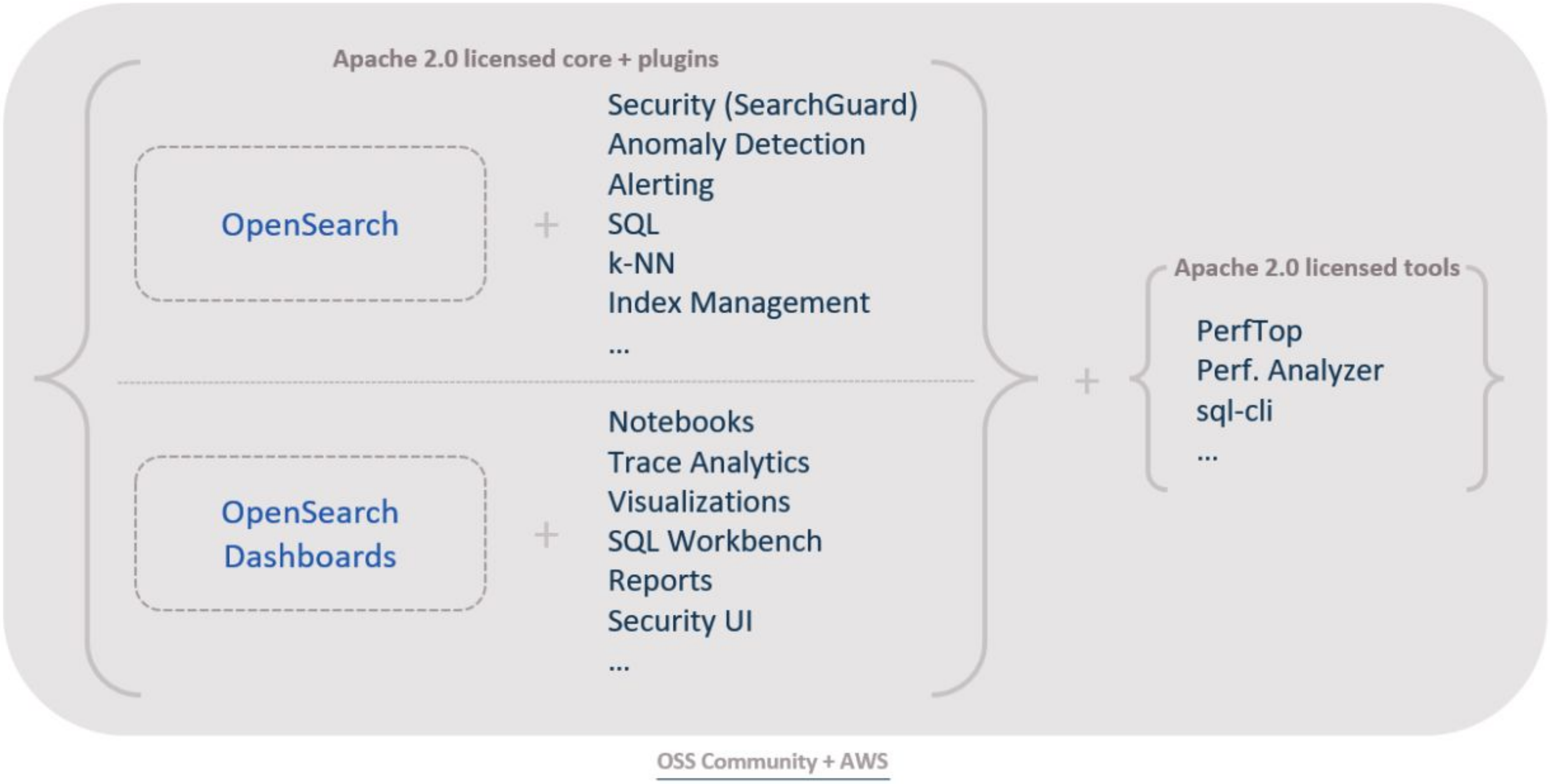
Open-source no more for Elastic

- As of v7.11 (January 2021) Elastic no longer offers an open source version of ES+Kibana
- AWS has decided to fork the latest ES+Kibana open source version (7.10.2) and OpenSearch was born
  - Essentially, OpenDistro project was re-branded as OpenSearch
  - Gathered Elastic-disappointed contributors (72 partners - incl. CERN)
  - Initial governance concerns (which are now overcome)
- All Elastic clients are burning bridges
  - Newer Elastic clients (e.g. elasticsearch-py) do not talk to OpenSearch
  - They also don’t talk to open-source Elasticsearch clusters!

Motivation for change

- **Licensing**
  - After v7.10.2, Elastic no longer provides Apache 2.0 releases
  - OpenSearch is licensed under Apache 2.0
- **Maintainability**
- **Streamlined deployment**
- **Customers isolation**
- **Features**
  - Many native plugins (alerting, index-management, etc.)
  - Fine-grained security access control

OpenSearch service packages + plugins



Apache 2.0 licensed core + plugins

OpenSearch +

Security (SearchGuard)
Anomaly Detection
Alerting
SQL
k-NN
Index Management
...

OpenSearch Dashboards +

Notebooks
Trace Analytics
Visualizations
SQL Workbench
Reports
Security UI
...

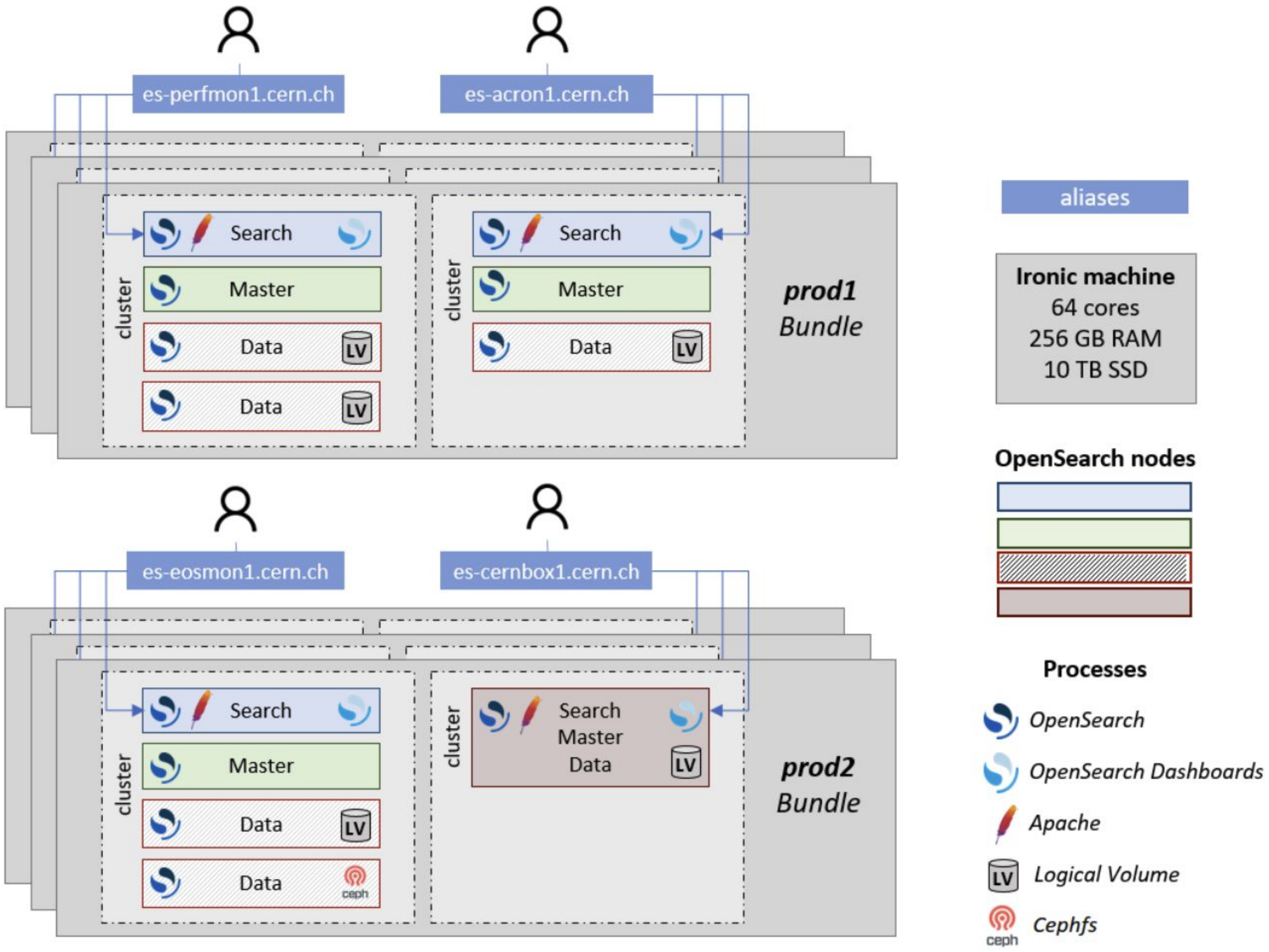
+

Apache 2.0 licensed tools

PerfTop
Perf. Analyzer
sql-cli
...

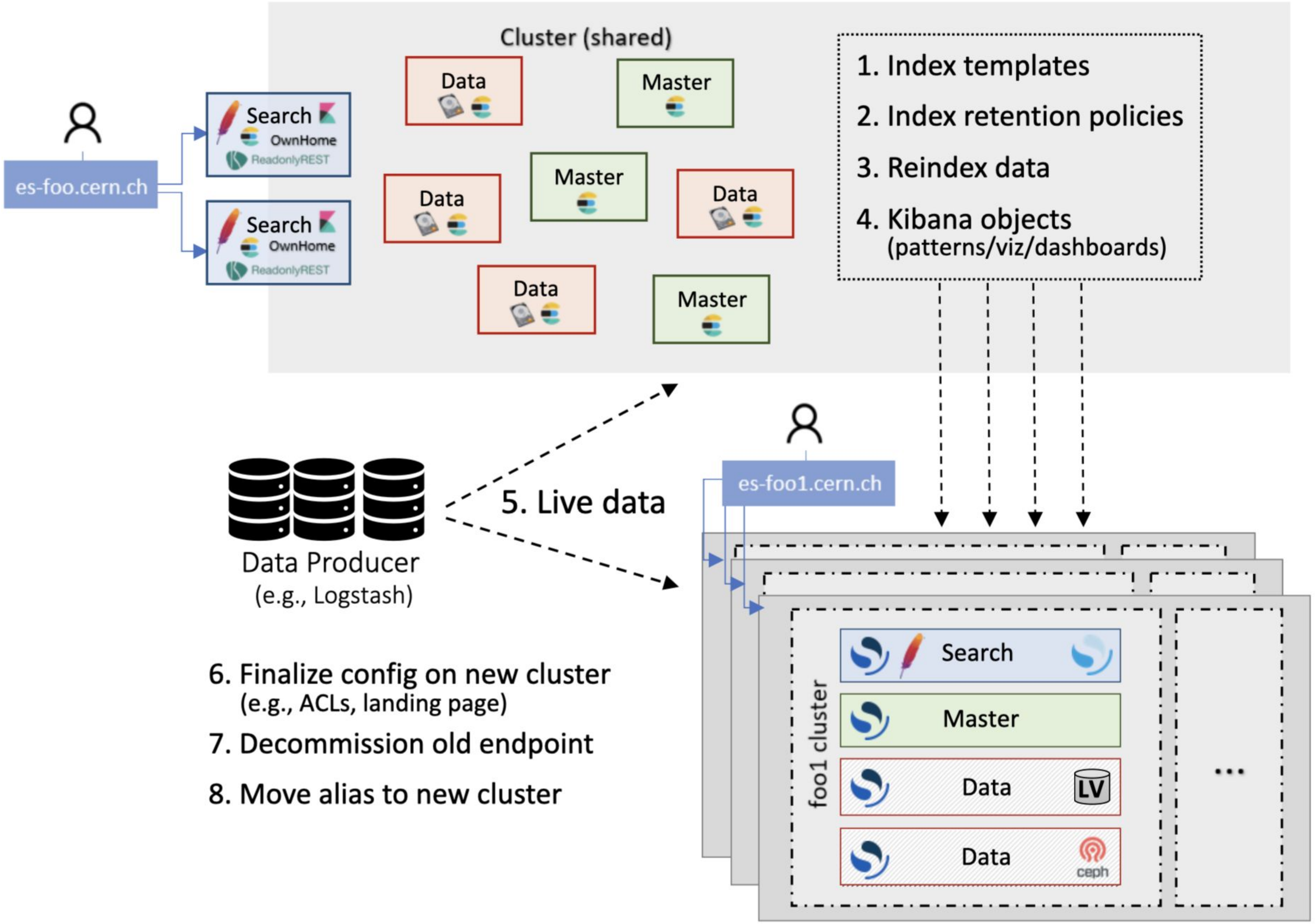
OSS Community + AWS

The OpenSearch service design



ES to OpenSearch - offline migration

- We completed Elasticsearch v6 (2022) and v7 migrations (2023)
- As the security plugins used were different, there was no possibility for *online* upgrade

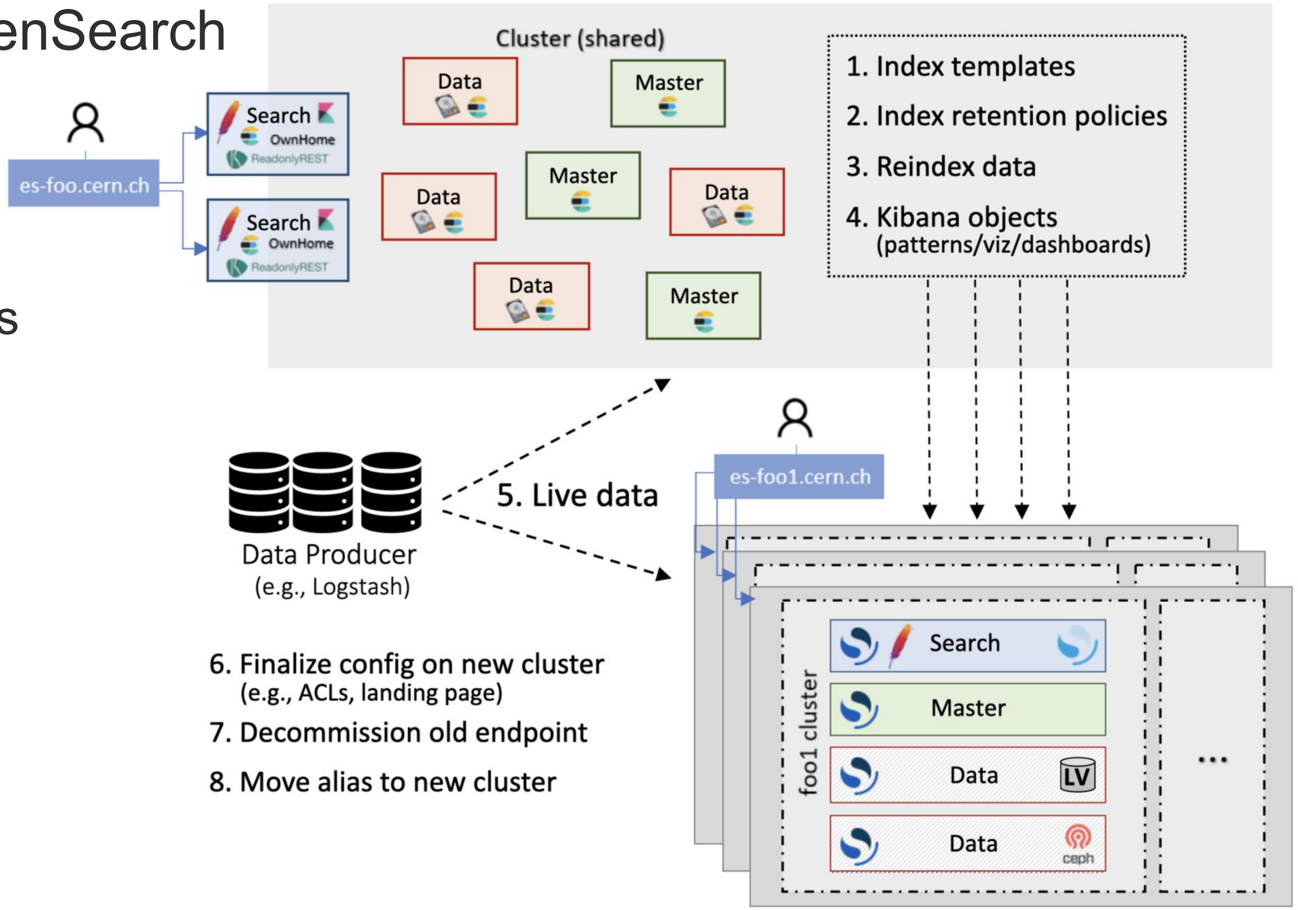


1. Index templates
2. Index retention policies
3. Reindex data
4. Kibana objects (patterns/viz/dashboards)

6. Finalize config on new cluster (e.g., ACLs, landing page)
7. Decommission old endpoint
8. Move alias to new cluster

OD to OpenSearch - offline migration

- *Offline*: Similarly to Elasticsearch-to-OpenSearch
  - Brand new OpenSearch cluster
  - Data migration
  - Pipelines redirection
  - Except that now we can also migrate ACLs
    - Internal_users
    - Roles
    - Role mappings
    - Tenants
    - …



1. Index templates
2. Index retention policies
3. Reindex data
4. Kibana objects (patterns/viz/dashboards)

5. Live data

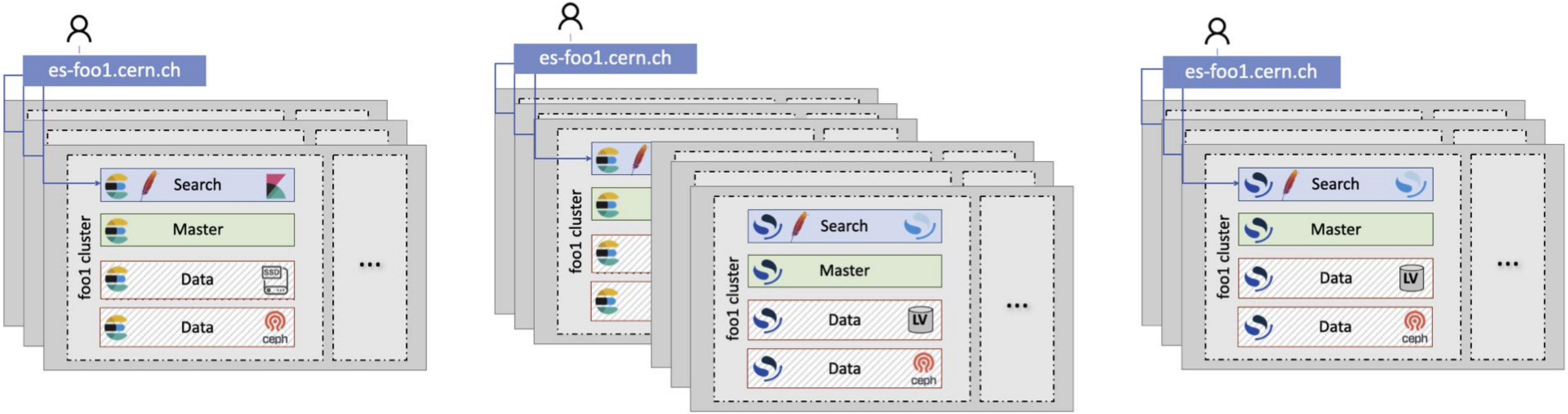
6. Finalize config on new cluster (e.g., ACLs, landing page)

7. Decommission old endpoint

8. Move alias to new cluster

OD to OpenSearch - online migration

- *Online*: In-cluster upgrade involving some **downtime**
  - Connect OpenSearch nodes, drain OpenDistro nodes, change alias hosts, stop OpenDistro nodes



Lessons learned and challenges emerged

- OpenSearch integration with CERN internal tools was quite easy
- Upstream puppet module does not support multiple instances
- Elastic burning bridges with OpenSearch
  - Some adjustments were needed on user side clients (e.g. logstash, filebeat, python, etc.)
- Users side engagement
  - Maintainers have left the organization
  - Deprioritizing migration
- Maintaining a service on 5 different major versions at a time
- Providing dedicated clusters now, users *must* respect their quotas
- OpenSearch v2 dropped support for _type field

Service security configuration - certificates

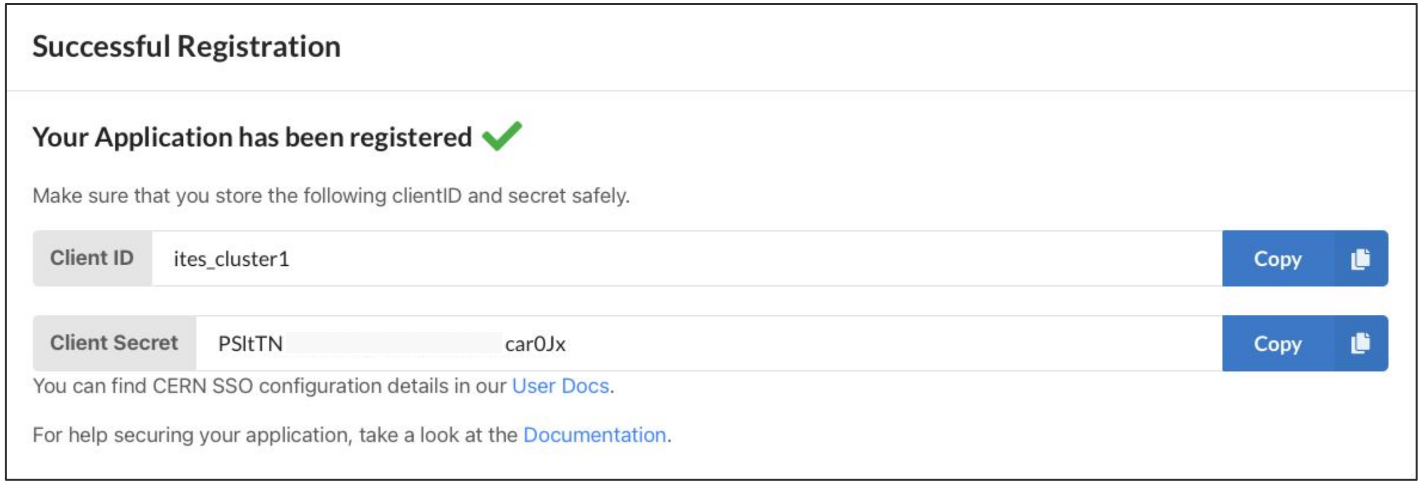
- CERN-CA-certs RPM as certificates
  - Secure transport-layer traffic (node-to-node communication)
  - Secure REST-layer traffic (communication between a client and a node within the cluster)
  - Hot reload upon new certificates
- Then, we use a CERN service account to produce a robot certificate
  - Used for superadmin API calls that require certificate authentication (e.g., certs hot reload)

```
curl -XPUT \
  --cert /etc/opensearch/admin.pem \
  --key /etc/opensearch/admin.key \
https://localhost:9200/_opendistro/_security/api/ssl/{transport,http}/reloadcerts
```

Service security configuration - AuthC/AuthZ

- **OpenID** integration for **CERN SSO**
- **LDAP** integration for OpenSearch Roles management based on CERN egroups

```
# check cluster's security configuration
GET _plugins/_security/api/securityconfig
```



Successful Registration

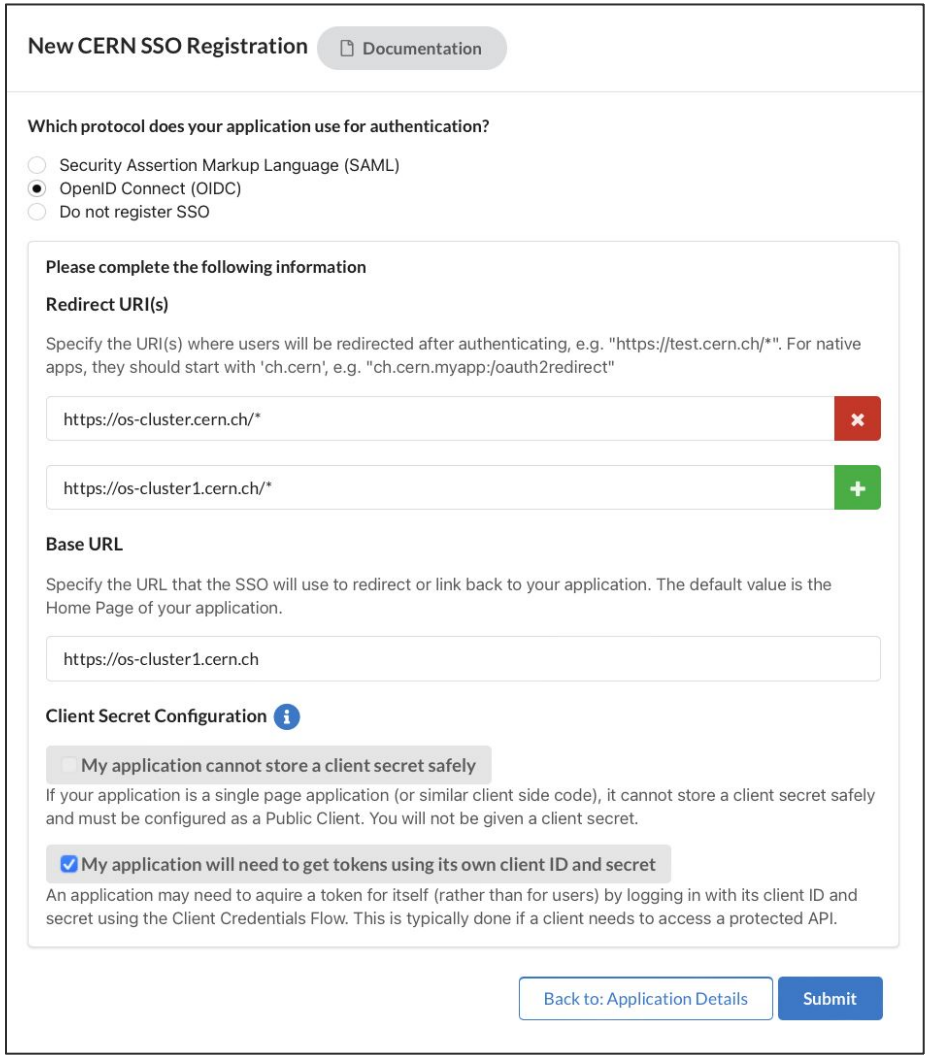
Your Application has been registered ✓

Make sure that you store the following clientID and secret safely.

Client ID	ites_cluster1	Copy
Client Secret	PSItTN [redacted] car0Jx	Copy

You can find CERN SSO configuration details in our User Docs.

For help securing your application, take a look at the Documentation.



New CERN SSO Registration

Which protocol does your application use for authentication?

○ Security Assertion Markup Language (SAML)
● OpenID Connect (OIDC)
○ Do not register SSO

Please complete the following information

Redirect URI(s)

Specify the URI(s) where users will be redirected after authenticating, e.g. "https://test.cern.ch/*". For native apps, they should start with 'ch.cern', e.g. "ch.cern.myapp:/oauth2redirect"

https://os-cluster.cern.ch/*

https://os-cluster1.cern.ch/*

Base URL

Specify the URL that the SSO will use to redirect or link back to your application. The default value is the Home Page of your application.

https://os-cluster1.cern.ch

Client Secret Configuration

□ My application cannot store a client secret safely

If your application is a single page application (or similar client side code), it cannot store a client secret safely and must be configured as a Public Client. You will not be given a client secret.

☑ My application will need to get tokens using its own client ID and secret

An application may need to aquire a token for itself (rather than for users) by logging in with its client ID and secret using the Client Credentials Flow. This is typically done if a client needs to access a protected API.

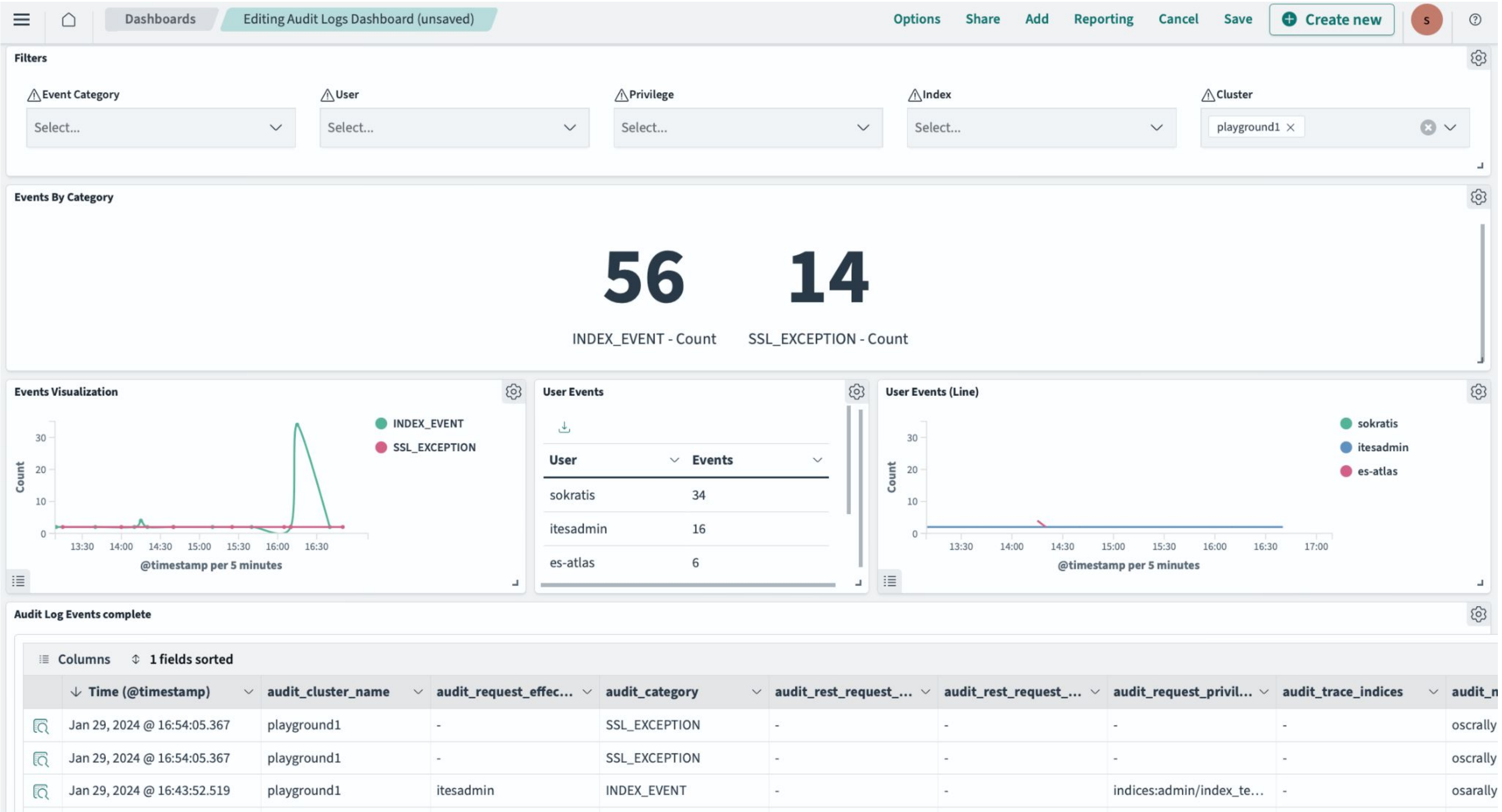
Service security configuration - AuthC/AuthZ

- **Kerberos** integration for CERN users authorised API calls
- OpenSearch **internal users** for data ingestion and queries
- We have **delegated** the ACLs to our “customers”
  - Each cluster is managed by a CERN e-group, who has full access to it
  - Some take advantage of document and field level security

Service security configuration - audit logs

- All clusters send their audit logs to our centralised cluster, where Audit Logs Dashboard show your cluster’s activity

- We have monitors on them to notify us of interesting cases



Event Category: Select...
User: Select...
Privilege: Select...
Index: Select...
Cluster: playground1

56 INDEX_EVENT - Count

14 SSL_EXCEPTION - Count

User	Events
sokratis	34
itesadmin	16
es-atlas	6

Time (@timestamp)	audit_cluster_name	audit_request_effec...	audit_category	audit_rest_request_...	audit_rest_request_...	audit_request_privil...	audit_trace_indices	audit_n
Jan 29, 2024 @ 16:54:05.367	playground1	-	SSL_EXCEPTION	-	-	-	-	oscrally
Jan 29, 2024 @ 16:54:05.367	playground1	-	SSL_EXCEPTION	-	-	-	-	oscrally
Jan 29, 2024 @ 16:43:52.519	playground1	itesadmin	INDEX_EVENT	-	-	indices:admin/index_te...	-	osarally

Service operations - monitoring

- OpenSearch clusters do some of their own monitoring using the OpenSearch Alerting plugin
  - For example, periodically run *GET _cluster/health*, *GET _cat/shards*, etc.
  - 7 monitors (cluster health, big shards, disk watermarks, etc.) reporting to one or multiple channels
    - Mattermost, SNOW, Email
  - Alarms that are related to cluster usage end up on customer’s channel
  - Alarms that are related to infrastructure problems end up on our channel
- Centralised monitoring on our internal perfmon cluster
  - Collecting all OpenSearch, Apache, Audit logs + more
  - Created centralised alarms on top of those logs (e.g. Apache errors, Requests with missing privileges)

Service operations - monitoring

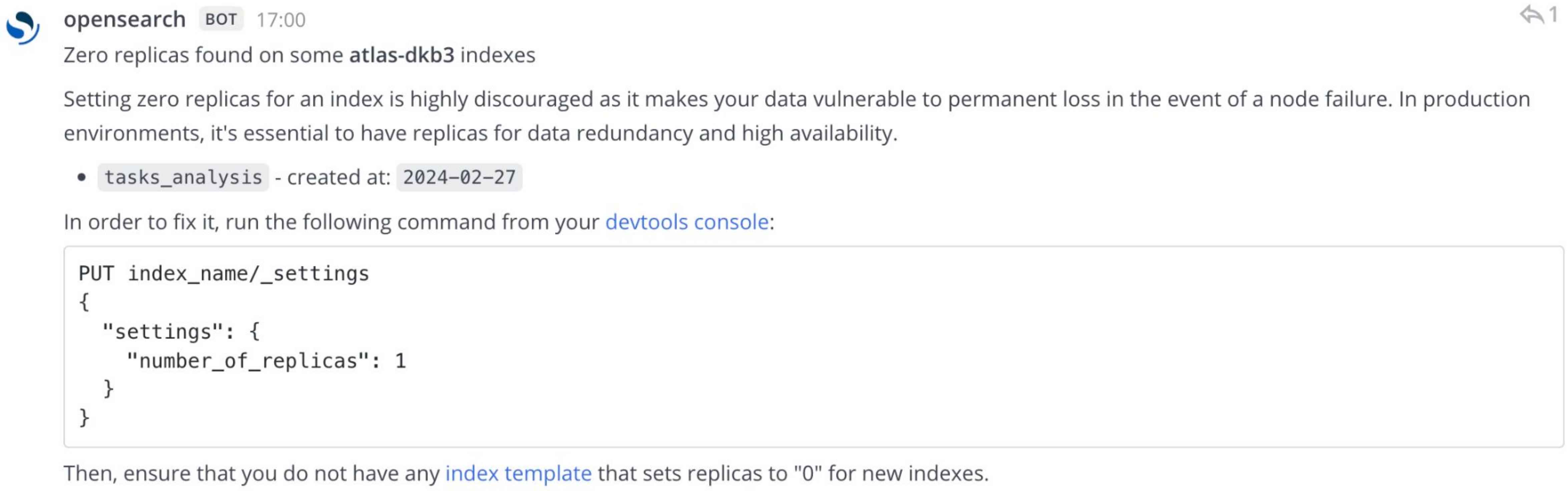
- In central monitoring instance, perfmon, **document-level security** ensures that our customers can only see their cluster(s) logs
- Index patterns examples:
  - perfmon_cluster_health-2024.03
  - perfmon_cluster_allocation-2024.03.20
  - perfmon_apacheaccess-2024.03.20

Security role config

```
{
 "perfmon_cert-soc-admins": {
  "cluster_permissions": [],
  "index_permissions": [
   {
    "index_patterns": [ "perfmon*" ],
    "dls": """{ "terms":{"cluster":["cert1", "csl1", "certdns1" ] } }""",
    "fls": [],
    "masked_fields": [],
    "allowed_actions": [ "read" ]
   }
  ],
  "tenant_permissions": []
 }
}
```

Service operations - monitoring

- Example calls to action from triggered alarms



opensearch BOT 17:00

Zero replicas found on some **atlas-dkb3** indexes

Setting zero replicas for an index is highly discouraged as it makes your data vulnerable to permanent loss in the event of a node failure. In production environments, it's essential to have replicas for data redundancy and high availability.

- `tasks_analysis` - created at: `2024-02-27`

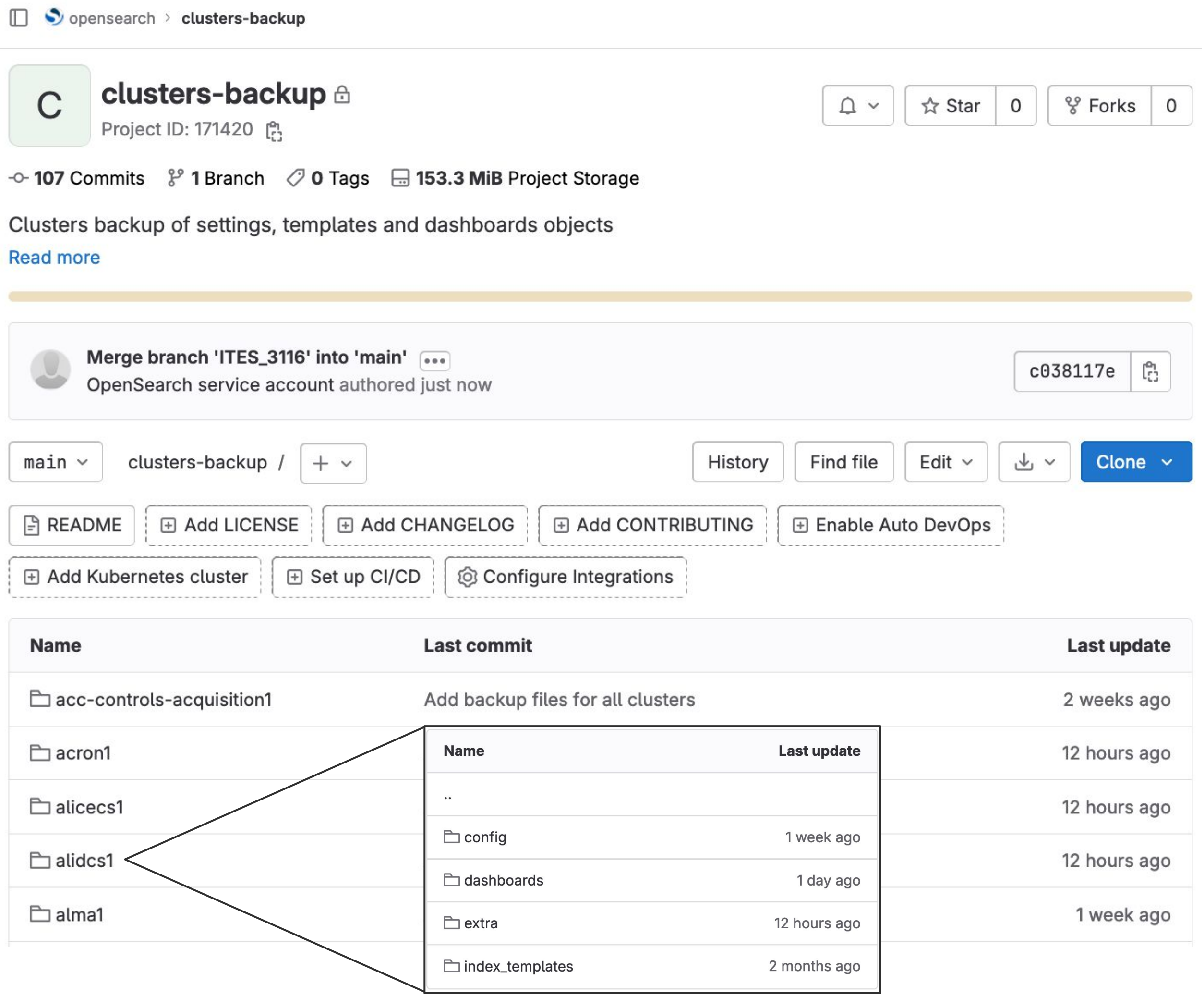
In order to fix it, run the following command from your devtools console:

```
PUT index_name/_settings
{
  "settings": {
    "number_of_replicas": 1
  }
}
```

Then, ensure that you do not have any index template that sets replicas to "0" for new indexes.

Service operations - cluster configuration backup

- A central clusters-backup repo now daily backs up all clusters' information
  - Cluster settings, index templates, index policies, alerting, etc.
- A daily script "downloads" this information and stores them in the gitlab repo



clusters-backup
Project ID: 171420

107 Commits 1 Branch 0 Tags 153.3 MiB Project Storage

Clusters backup of settings, templates and dashboards objects
Read more

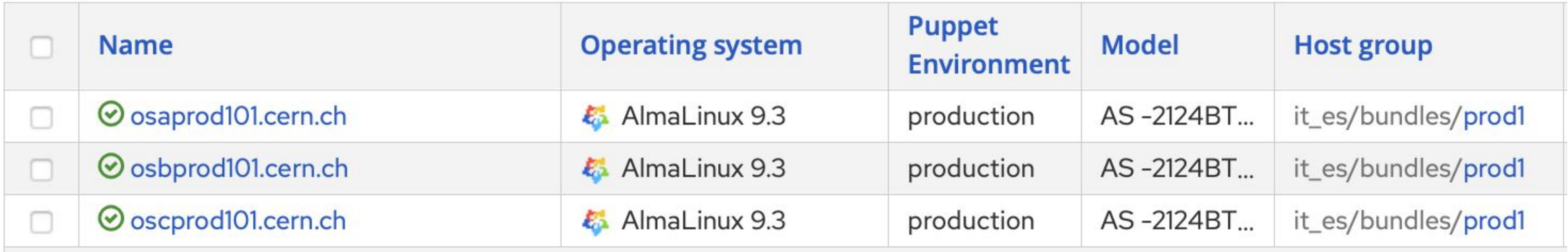
Merge branch 'ITES_3116' into 'main'
OpenSearch service account authored just now
c038117e

Name	Last commit	Last update
acc-controls-acquisition1	Add backup files for all clusters	2 weeks ago
acron1		12 hours ago
alicecs1		12 hours ago
alidcs1		12 hours ago
alma1		1 week ago

Name	Last update
..	
config	1 week ago
dashboards	1 day ago
extra	12 hours ago
index_templates	2 months ago

Service operations - deployment

FOREMAN



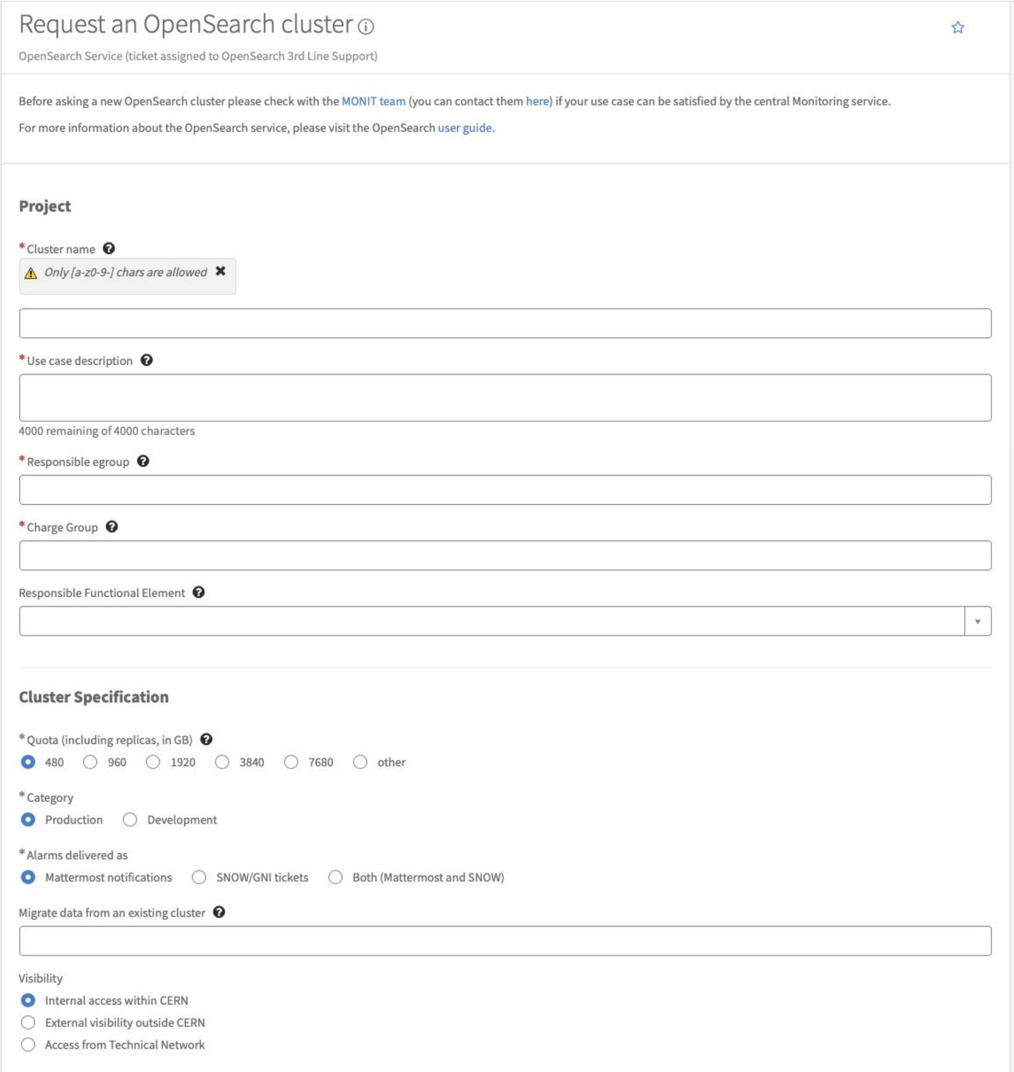
□	Name	Operating system	Puppet Environment	Model	Host group
□	✓ osaprod101.cern.ch	AlmaLinux 9.3	production	AS -2124BT...	it_es/bundles/prod1
□	✓ osbprod101.cern.ch	AlmaLinux 9.3	production	AS -2124BT...	it_es/bundles/prod1
□	✓ oscprod101.cern.ch	AlmaLinux 9.3	production	AS -2124BT...	it_es/bundles/prod1

puppet

- Puppet **modules** for everything, as **building blocks**
  - Re-used across different teams/services
  - E.g. apache, CERN SSO, firewall
- Puppet **hostgroups** to build a service gluing these building together

Service operations - cluster bootstrapping

1. Customers complete a SNOW form
   a. Cluster name & description
   b. Superadmins egroup
   c. Charge group
   d. Quota
   e. Environment (prod/dev)
   f. Alarms destination (mattermost/snow)
   g. Visibility (internal/external)



Request an OpenSearch cluster

OpenSearch Service (ticket assigned to OpenSearch 3rd Line Support)

Before asking a new OpenSearch cluster please check with the MONIT team (you can contact them here) if your use case can be satisfied by the central Monitoring service.

For more information about the OpenSearch service, please visit the OpenSearch user guide.

Project

* Cluster name

⚠ *Only [a-z0-9-] chars are allowed* ✖

* Use case description

4000 remaining of 4000 characters

* Responsible egroup

* Charge Group

Responsible Functional Element

Cluster Specification

* Quota (including replicas, in GB)

● 480 ○ 960 ○ 1920 ○ 3840 ○ 7680 ○ other

* Category

● Production ○ Development

* Alarms delivered as

● Mattermost notifications ○ SNOW/GNI tickets ○ Both (Mattermost and SNOW)

Migrate data from an existing cluster

Visibility

● Internal access within CERN
○ External visibility outside CERN
○ Access from Technical Network

Service operations - cluster bootstrapping

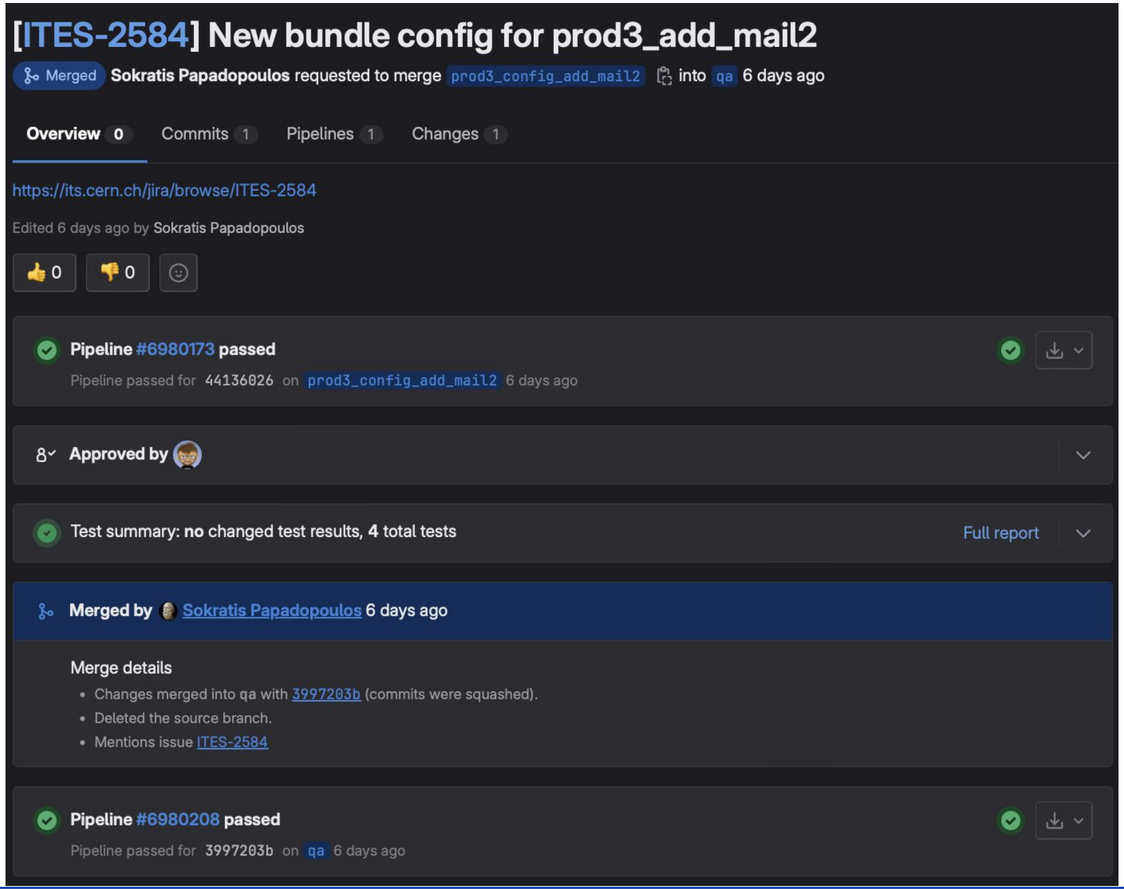
2. Script raises a MR with suggested config

```
cluster_init.py   --create-yaml
                  --bundle prod1
                  --cluster foo1
                  --superadmins it-security
                  --account-to 'Computer Security'
                  --node-type master_kibana_160
                  --ticket RQF2153756
                  --mattermost opensearch-for-cert
                  --alias os-foo
```

```
hostgroup: it_es > bundles > prod1
  * prod1.yaml file
   ---
   hg_it_es::bundles::clusters:
     cluster1:
       [ config cluster1 ]
     cluster2:
       [ config cluster2 ]
     foo1:
       alias:
       - os-foo
       metadata:
         account_to: Computer Security
         mattermost: opensearch-for-cert
         superadmins: it-security
         ticket: RQF2153756
       nodes:
       - type: master_kibana_160
       port: 9220
```

Service operations - cluster bootstrapping

3. A member of the team reviews it and merges it in QA



[ITES-2584] New bundle config for prod3_add_mail2

Merged **Sokratis Papadopoulos** requested to merge `prod3_config_add_mail2` into `qa` 6 days ago

https://its.cern.ch/jira/browse/ITES-2584

Edited 6 days ago by Sokratis Papadopoulos

Pipeline #6980173 passed
Pipeline passed for 44136026 on prod3_config_add_mail2 6 days ago

Approved by

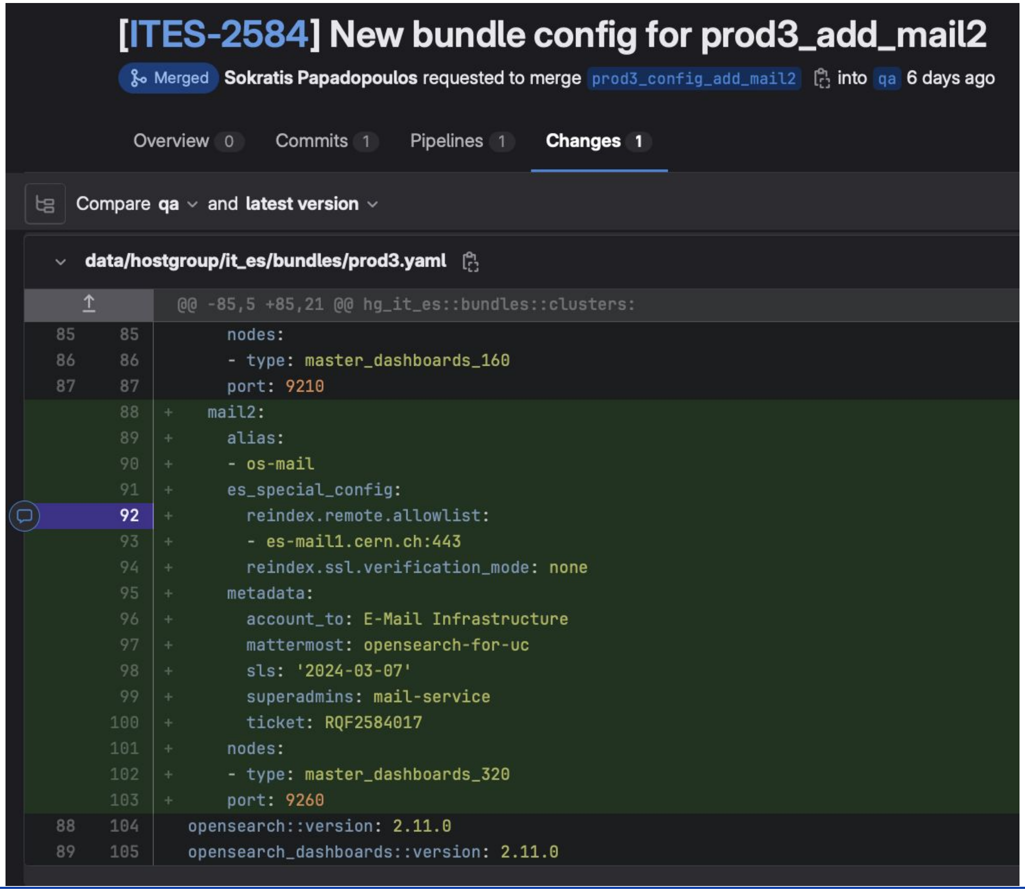
Test summary: **no changed test results, 4 total tests** Full report

Merged by Sokratis Papadopoulos 6 days ago

Merge details

- Changes merged into qa with 3997203b (commits were squashed).
- Deleted the source branch.
- Mentions issue ITES-2584

Pipeline #6980208 passed
Pipeline passed for 3997203b on qa 6 days ago



[ITES-2584] New bundle config for prod3_add_mail2

Merged **Sokratis Papadopoulos** requested to merge `prod3_config_add_mail2` into `qa` 6 days ago

Compare **qa** and **latest version**

data/hostgroup/it_es/bundles/prod3.yaml

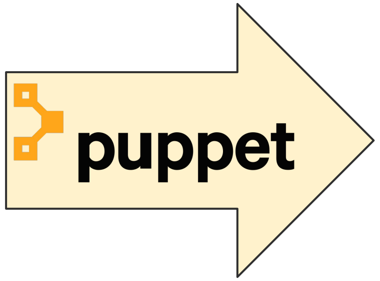
```
@@ -85,5 +85,21 @@ hg_it_es::bundles::clusters:
      nodes:
      - type: master_dashboards_160
      port: 9210
+   mail2:
+     alias:
+     - os-mail
+     es_special_config:
+       reindex.remote.allowlist:
+       - es-mail1.cern.ch:443
+       reindex.ssl.verification_mode: none
+     metadata:
+       account_to: E-Mail Infrastructure
+       mattermost: opensearch-for-uc
+       sls: '2024-03-07'
+       superadmins: mail-service
+       ticket: RQF2584017
+     nodes:
+     - type: master_dashboards_320
+     port: 9260
    opensearch::version: 2.11.0
    opensearch_dashboards::version: 2.11.0
```

Service operations - cluster bootstrapping

4. Puppet propagates the configuration to all servers in the hostgroup

```
hostgroup: it_es > bundles > prod1
  * prod1.yaml file
   ---
   hg_it_es::bundles::clusters:
     cluster1:
       [ config cluster1 ]
     cluster2:
       [ config cluster2 ]
     foo1:
       alias:
       - os-foo
       metadata:
          account_to: Computer Security
         mattermost: opensearch-for-cert
         superadmins: it-security
         ticket: RQF2153756
       nodes:
       - type: master_kibana_160
       port: 9220
```

hostgroup-it_es



module-opensearch

module-opensearch_dashboards

+ 50 modules…

- Iterate over defined clusters and for each host in the specified hostgroup…
  - Configure/Update certificate files
  - Configure apache files
  - Configure opensearch.yml
  - Configure opensearch-dashboards.yml
  - Configure cluster security & ACL files
  - Configure jvm.options for all processes
  - Create logical volume for data nodes
  - Configure kerberos files
  - Configure beats to send all logs to central logstash instance
  - …

Service operations - cluster bootstrapping

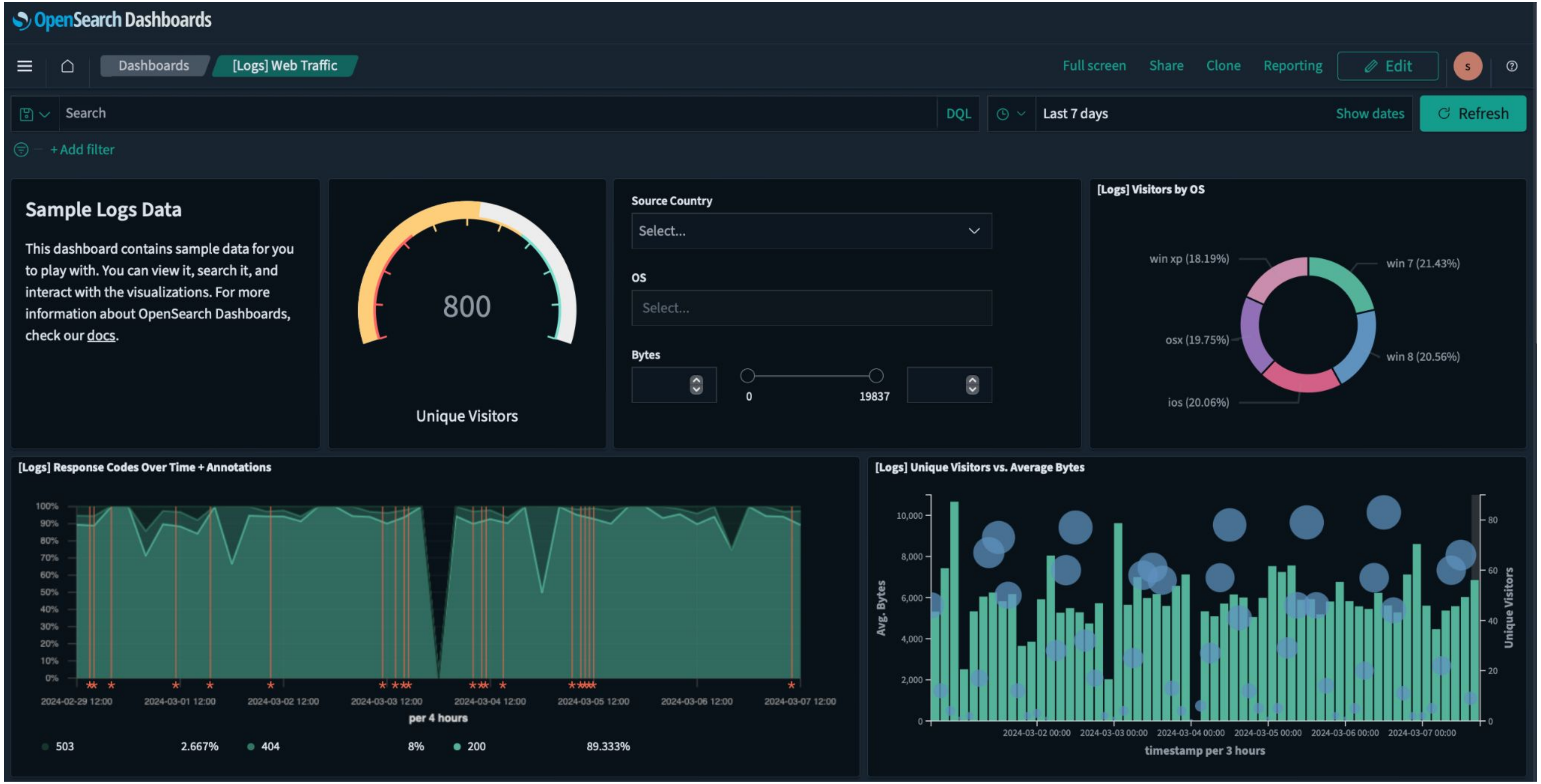
5. Script is now finishing the job, bringing the cluster up

```
cluster_init.py --up
                --cluster foo1
```

- Run securityadmin.sh to load ACLs to the cluster
- Register the cluster aliases on Load Balancing service
- Create OpenID registration on CERN Application portal
- Allow access of cluster superadmins on centralised monitoring
- Create Monitor objects
- …

Service operations - cluster bootstrapping

Cluster is up!

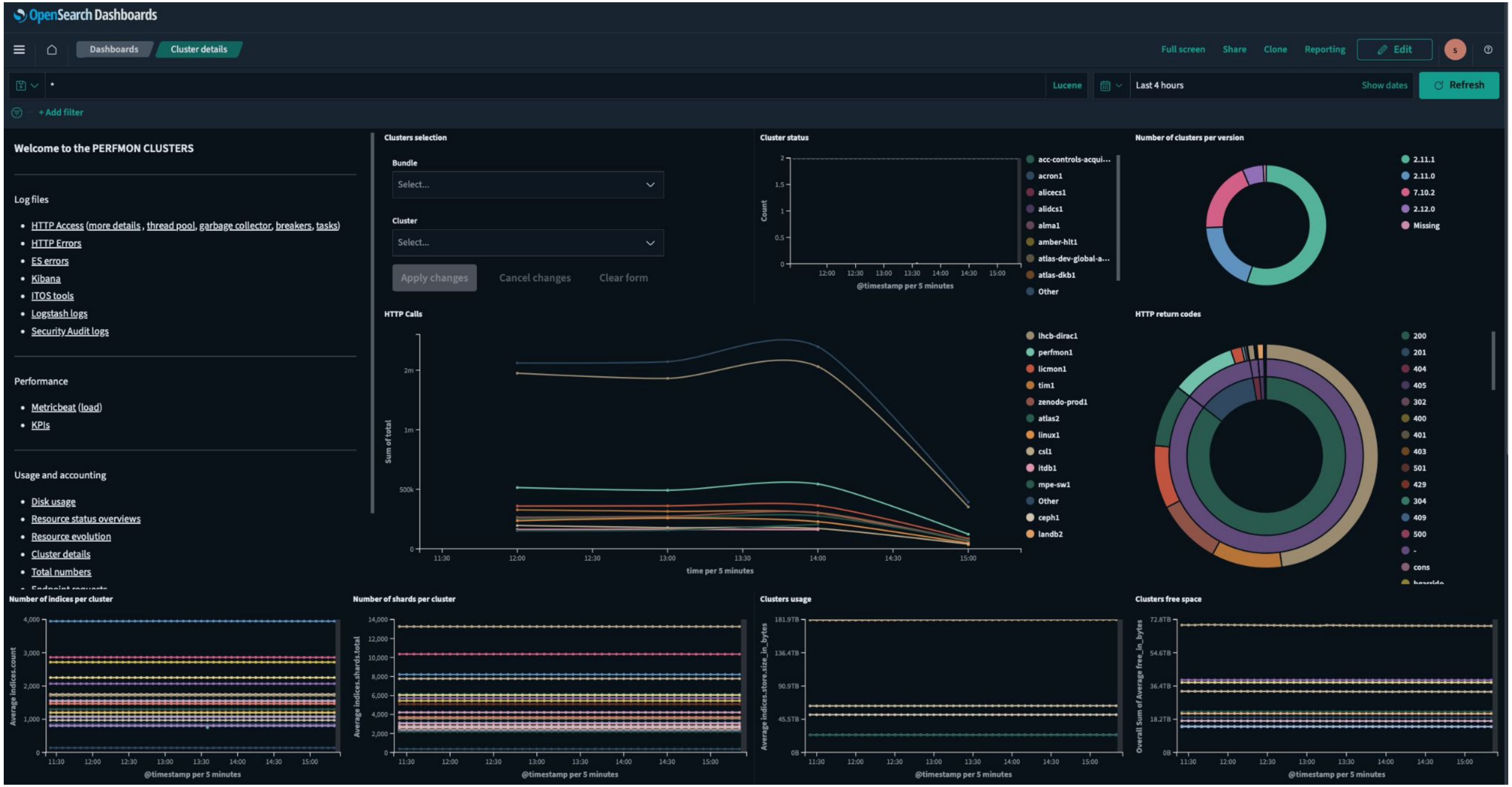


Sample Logs Data

This dashboard contains sample data for you to play with. You can view it, search it, and interact with the visualizations. For more information about OpenSearch Dashboards, check our docs.

Service operations - cluster bootstrapping

And monitoring data on the new cluster start to flow on our centralised monitoring cluster



Welcome to the PERFMON CLUSTERS

Log files

- HTTP Access (more details , thread pool, garbage collector, breakers, tasks)
- HTTP Errors
- ES errors
- Kibana
- ITOS tools
- Logstash logs
- Security Audit logs

Performance

- Metricbeat (load)
- KPIs

Usage and accounting

- Disk usage
- Resource status overviews
- Resource evolution
- Cluster details
- Total numbers

Service operations - management

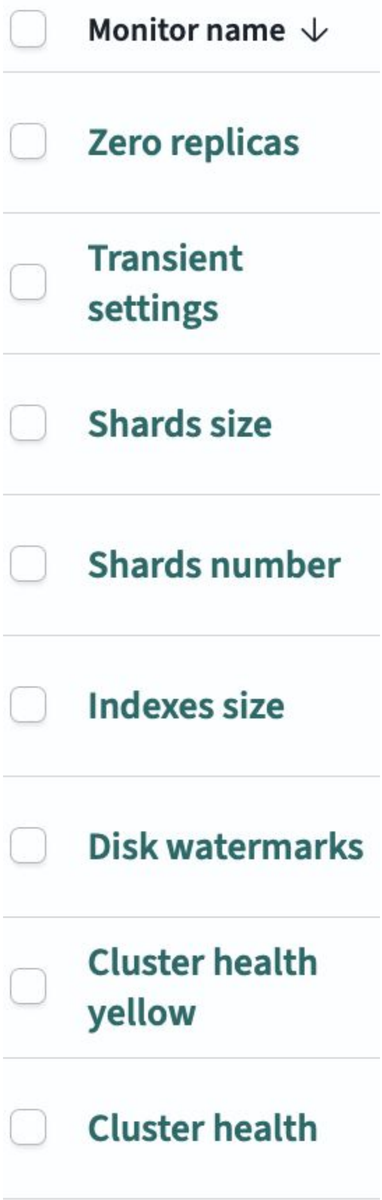
- We have a list of python (and some bash) scripts to ease cluster management
- Cluster restarts (used for upgrades)
  - Following closely latest releases
  - Upgrade process is straightforward and almost completely transparent
    `itos_restart_clusters --cluster playground`
    - For each host in the cluster's hostgroup
      - Restarts the cluster's opensearch processes (*systemctl restart opensearch-playground1**)
      - Waits for the processes to come back and cluster get back to green

Service operations - management

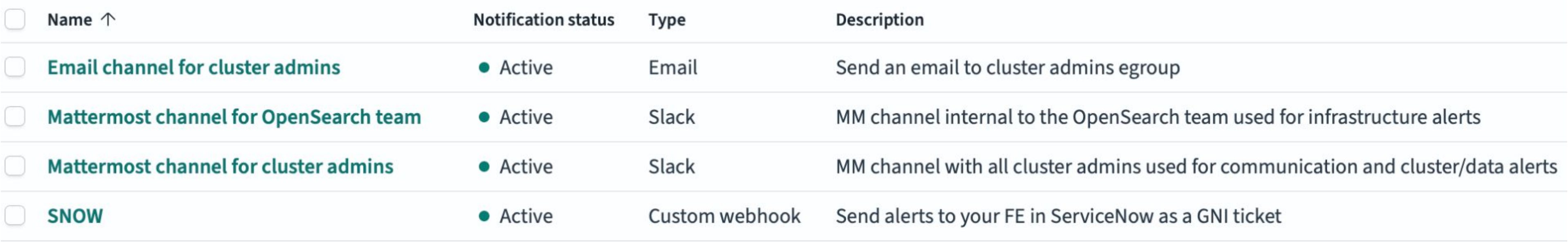
- Cluster moves to another bundle
  `itos_bundle_cluster_move --cluster playground --new-bundle prod2`
  - Opens firewall between the two different bundles
  - bootstrap new nodes on new bundle
  - drain old nodes of old bundle
  - stop old nodes
- Cluster configuration backup
  `itos_backup_config`
  - Iterate all clusters in the service
  - do some API calls and store the responses in a gitlab repo

Service operations - management

- Create/Update default monitors configuration on all OpenSearch clusters
  `itos_alerting`
  - Iterate all clusters in the service
  - Ensure that Mattermost, SNOW and Email notification channels are created
  - Ensure all 7 default Monitors are configured properly according to the cluster's values



□	Monitor name ↓
□	Zero replicas
□	Transient settings
□	Shards size
□	Shards number
□	Indexes size
□	Disk watermarks
□	Cluster health yellow
□	Cluster health



□	Name ↑	Notification status	Type	Description
□	Email channel for cluster admins	● Active	Email	Send an email to cluster admins egroup
□	Mattermost channel for OpenSearch team	● Active	Slack	MM channel internal to the OpenSearch team used for infrastructure alerts
□	Mattermost channel for cluster admins	● Active	Slack	MM channel with all cluster admins used for communication and cluster/data alerts
□	SNOW	● Active	Custom webhook	Send alerts to your FE in ServiceNow as a GNI ticket

Bonus: OpenSearch Security Analytics

- SIEM solution for OpenSearch: tools and features to detect, investigate and resolve issues, reducing MTTD, MTTR and MTTC
- Security rules engine
  - execute over 2200 sigma rules against your security logs or define your own
  - identify unusual activities based on MITRE ATT&CK knowledge base
- Log types
  - use pre-defined mappings (Windows server logs, DNS logs, System logs) or build your own
- Findings & Alerts
  - Get notifications when a potential risk (finding) is detected
- Playground & documentation

Roadmap for 2024

- Complete **OpenDistro migration** to OpenSearch by Q4 2024
- Perform OpenSearch upgrades
- Explore **Snapshots** for disaster recovery
- Engage more with the OpenSearch **community**
- Explore Data streams for append-only logs
- Experiment with **VectorDB** capabilities
- Explore OpenSearch deployment on Kubernetes (summer student project)

Summary

- A service with growing interest over the last 8 years (currently operated with 1.7 FTEs)
- OpenSearch brought **significant changes** both internally and on user side
- The new OpenSearch service is deployed with puppet on bare metal machines
- **20 OpenDistro** clusters are left to migrate to OpenSearch by **Q4 2024**
- Plethora of opportunities to further **enhance** the service
- Further reading on OpenSearch service architecture: CHEP 2023 paper

home.cern