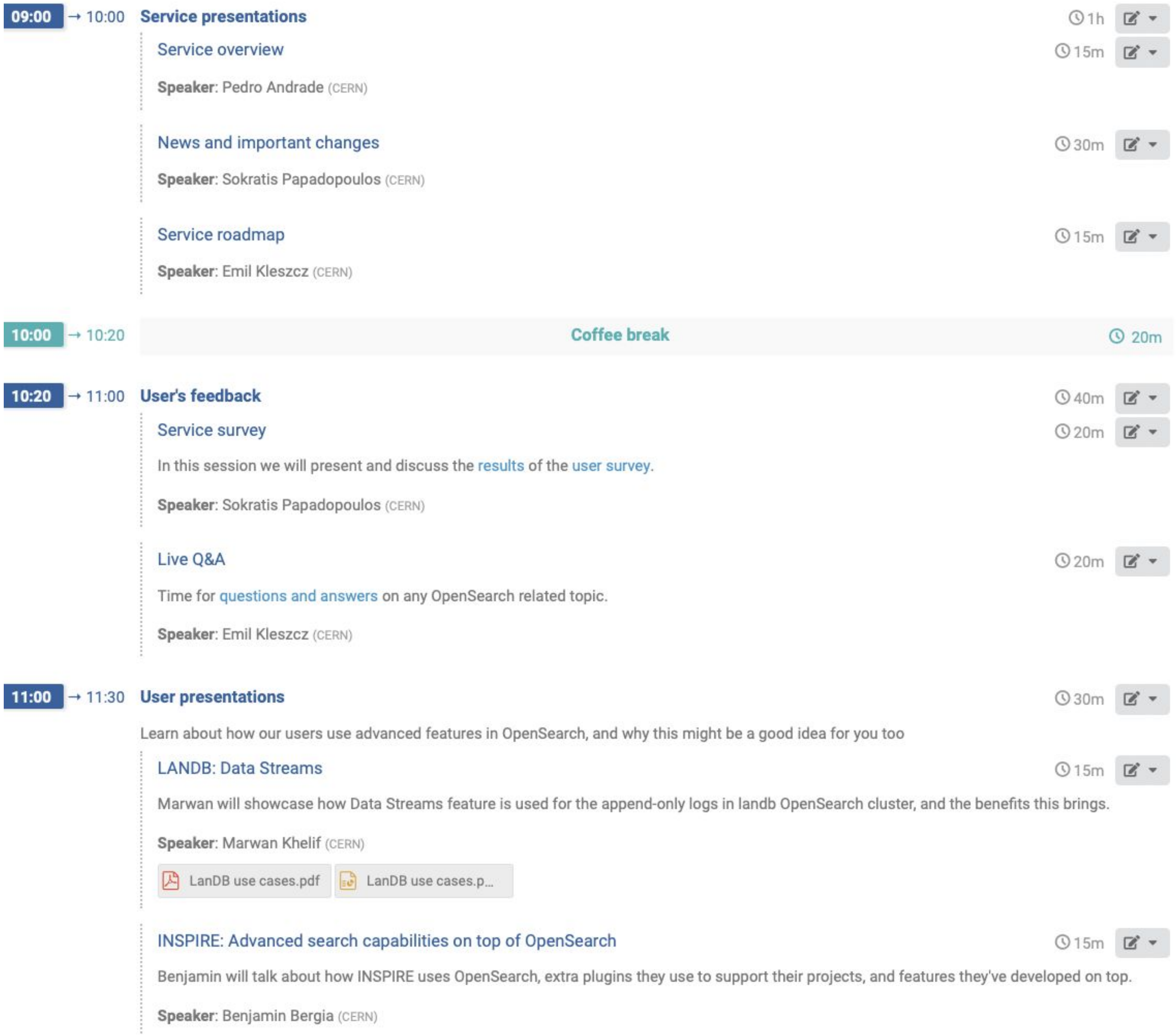
CERN

OpenSearch User Forum #1

The OpenSearch team

Welcome

Thanks for joining :)



09:00 → 10:00 **Service presentations** 1h

Service overview 15m

Speaker: Pedro Andrade (CERN)

News and important changes 30m

Speaker: Sokratis Papadopoulos (CERN)

Service roadmap 15m

Speaker: Emil Kleszcz (CERN)

10:00 → 10:20 **Coffee break** 20m

10:20 → 11:00 **User's feedback** 40m

Service survey 20m

In this session we will present and discuss the results of the user survey.

Speaker: Sokratis Papadopoulos (CERN)

Live Q&A 20m

Time for questions and answers on any OpenSearch related topic.

Speaker: Emil Kleszcz (CERN)

11:00 → 11:30 **User presentations** 30m

Learn about how our users use advanced features in OpenSearch, and why this might be a good idea for you too

LANDB: Data Streams 15m

Marwan will showcase how Data Streams feature is used for the append-only logs in landb OpenSearch cluster, and the benefits this brings.

Speaker: Marwan Khelif (CERN)

LanDB use cases.pdf LanDB use cases.p...

INSPIRE: Advanced search capabilities on top of OpenSearch 15m

Benjamin will talk about how INSPIRE uses OpenSearch, extra plugins they use to support their projects, and features they've developed on top.

Speaker: Benjamin Bergia (CERN)

Service overview

intro, history, team, mandate, architecture, and stats

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 a fork of Kibana 7.10.2 open-source codebase

Service history

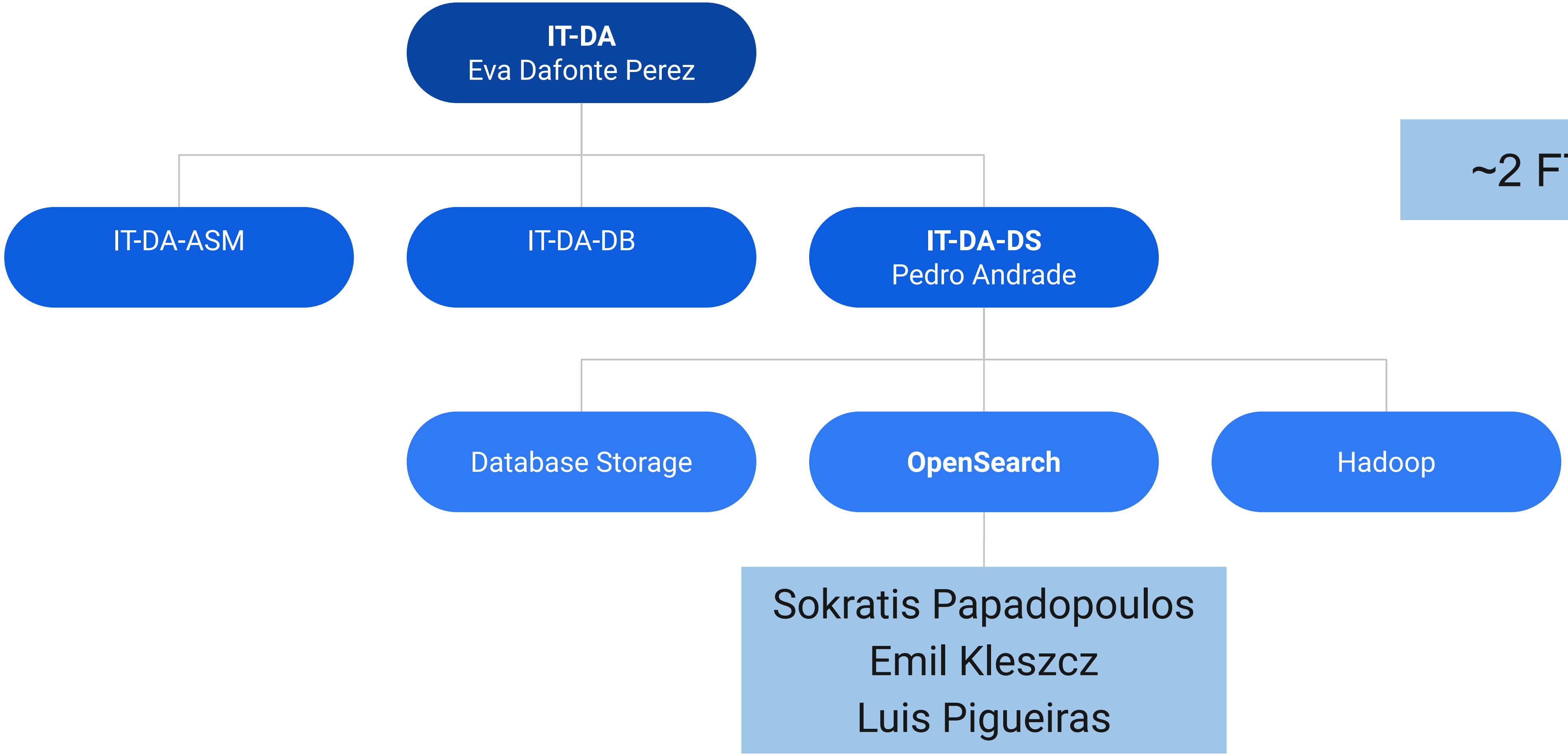
- **2016:** Creation of Elasticsearch service (thanks to Pablo and Ulrich)
- **2017:** Upgrade to Elasticsearch v5
- **2018:** Upgrade to Elasticsearch v6
- **2020:** Upgrade to Elasticsearch v7.1
- **2020:** Evaluation of OpenDistro
- **2021:** Migration towards OpenDistro (from ES6/7)
- **2022:** Evaluation of OpenSearch v1
- **2022:** Migration towards OpenSearch (from ES6/7 and OD)
- **2023**: Decommission of Elasticsearch
- **2024:** Decommission of OpenDistro *(thanks for your help!)*

elasticsearch

Open Distro for Elasticsearch

OpenSearch

Team members



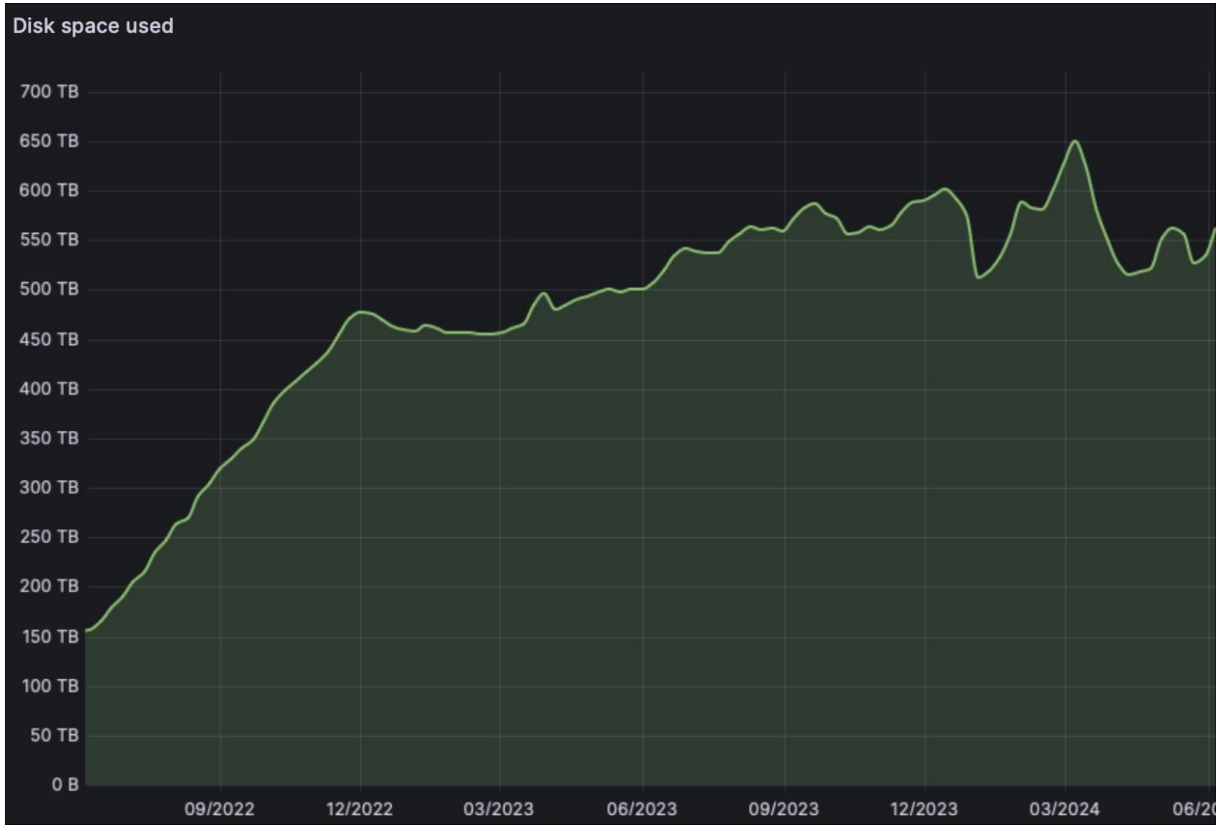
~2 FTEs

Service mandate

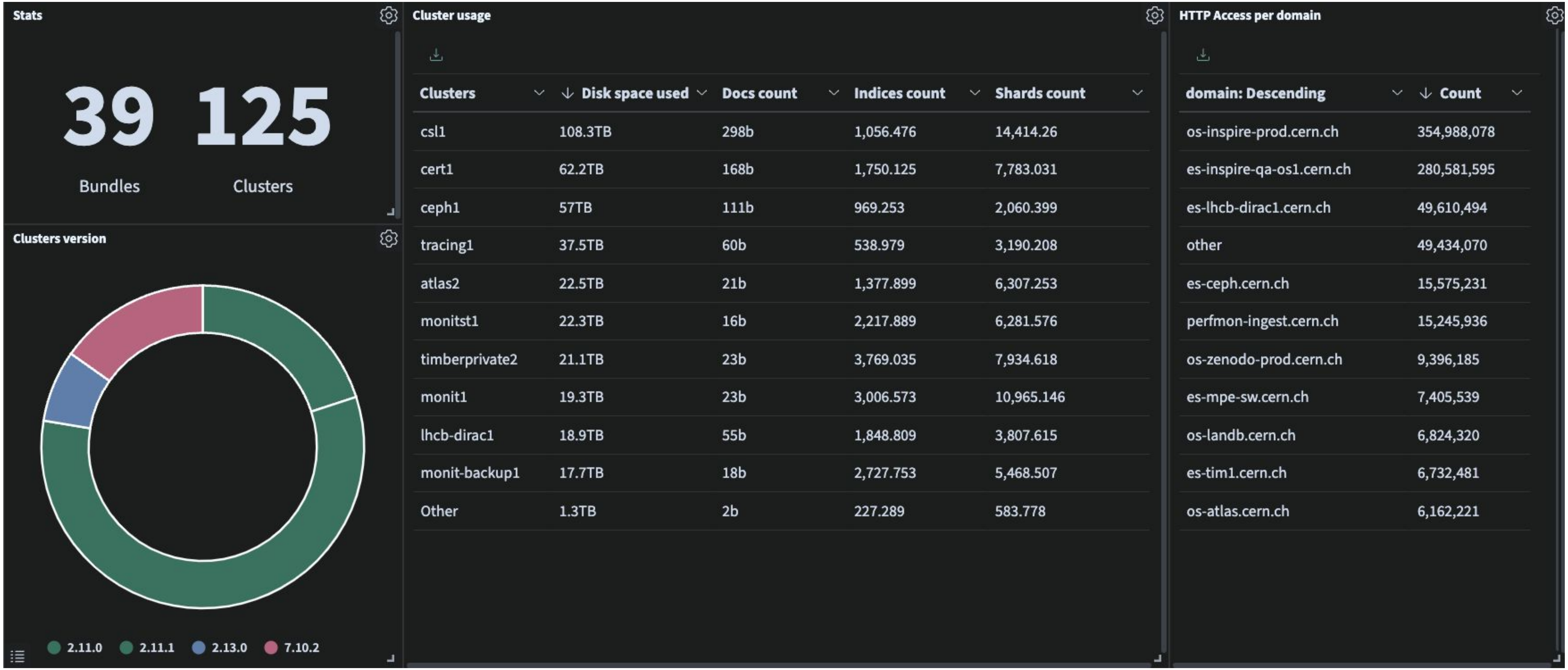
- Provide **OpenSearch as a Service** for the CERN and WLCG communities
- Infrastructure **management:** monitoring, alerting, upgrades, and security updates
- Users' **support** via ServiceNow/Mattermost during office hours
- OpenSearch usage **consultancy** on best effort basis
  - Client tools (e.g., Logstash) fall in this category
- Data **backup** strategy aligned with BC/DR
  - For cluster settings and dashboard objects: enabled on all clusters by default
  - For cluster's data: enabled per cluster upon request *(WIP)*

Service stats

- Available resources
  - **156** Ironic managed **physical machines**
    - 256 GB RAM / 64 cores / 10.5 TB SSD
    - 3 availability zones / “racks”
    - Expiring at the end of 2026
  - 400 TB of CephFS (cold) storage
  - 30 TB of S3 storage (backups)
- Service numbers
  - **108 OpenSearch + 17 OpenDistro** clusters
  - **~600 TB** of indexed data ~**1.2 trillion** docs



Service stats



Stats

39 Bundles

125 Clusters

Clusters version

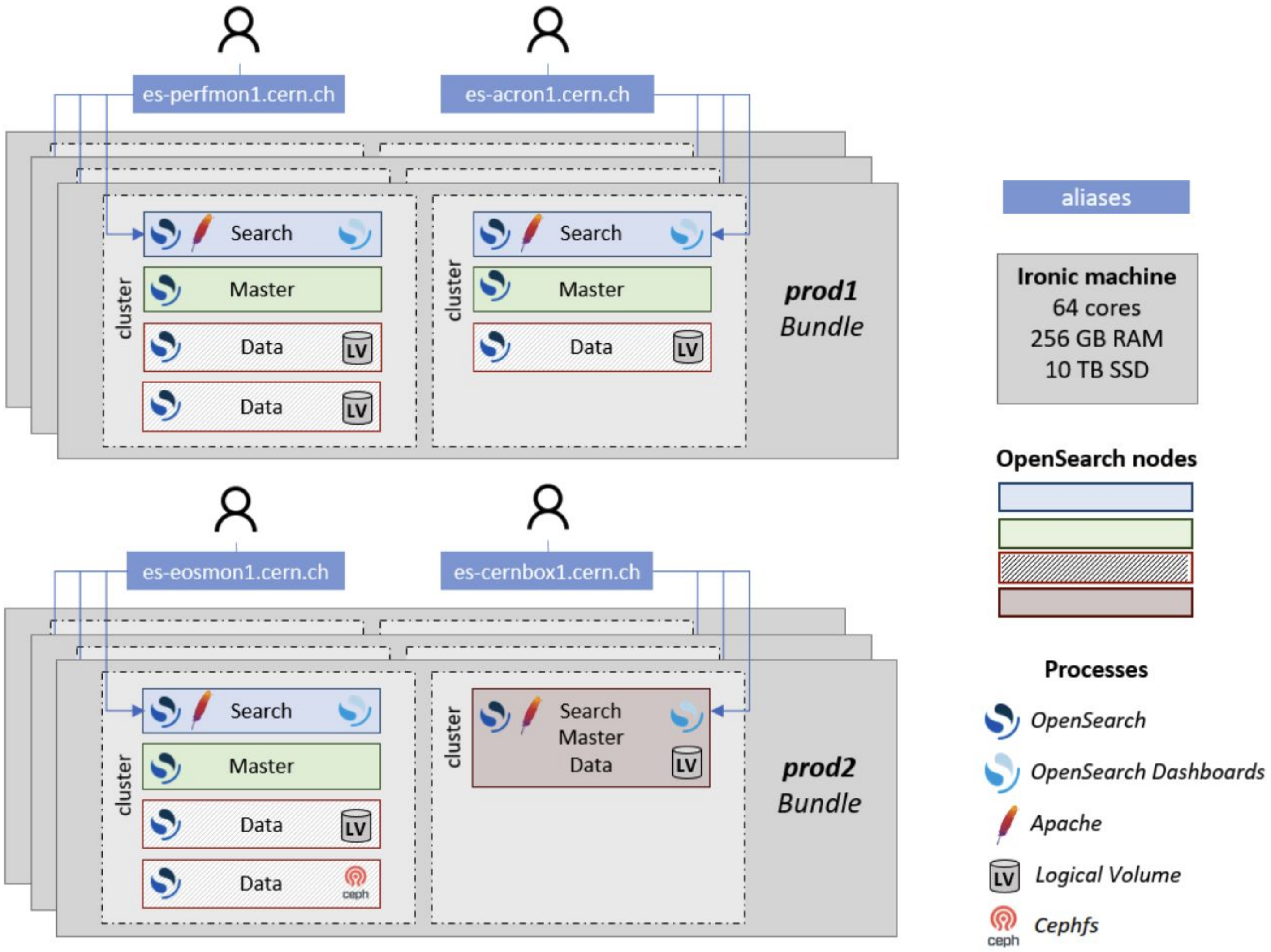
Cluster usage

Clusters	↓ Disk space used	Docs count	Indices count	Shards count
csl1	108.3TB	298b	1,056.476	14,414.26
cert1	62.2TB	168b	1,750.125	7,783.031
ceph1	57TB	111b	969.253	2,060.399
tracing1	37.5TB	60b	538.979	3,190.208
atlas2	22.5TB	21b	1,377.899	6,307.253
monitst1	22.3TB	16b	2,217.889	6,281.576
timberprivate2	21.1TB	23b	3,769.035	7,934.618
monit1	19.3TB	23b	3,006.573	10,965.146
lhcb-dirac1	18.9TB	55b	1,848.809	3,807.615
monit-backup1	17.7TB	18b	2,727.753	5,468.507
Other	1.3TB	2b	227.289	583.778

HTTP Access per domain

domain: Descending	↓ Count
os-inspire-prod.cern.ch	354,988,078
es-inspire-qa-os1.cern.ch	280,581,595
es-lhcb-dirac1.cern.ch	49,610,494
other	49,434,070
es-ceph.cern.ch	15,575,231
perfmon-ingest.cern.ch	15,245,936
os-zenodo-prod.cern.ch	9,396,185
es-mpe-sw.cern.ch	7,405,539
os-landb.cern.ch	6,824,320
es-tim1.cern.ch	6,732,481
os-atlas.cern.ch	6,162,221

Service design



Tech stack

- **Software packages**
  - OpenSearch upstream RPMs
- **Installation & Configuration**
  - AlmaLinux v9.4
  - puppet custom modules
- **Security**
  - OpenID Connect (CERN SSO) / Basic Auth
  - LDAP (CERN e-groups) / Kerberos

- **Cluster management & automation**
  - itostools repo python™
  - Few cron jobs BASH THE BOURNE-AGAIN SHELL
- **Monitoring and SLS**
  - MONIT: Collectd, SLS, GNI, remote probes
  - OpenSearch monitoring cluster (perfmon)
  - Logstash, [File|Metric]beat

Interventions

- **Gradual rollouts**
  - Staged deployment per environment: DEV > QA > PROD > PROD CRITICAL
  - Request a QA cluster to test upgrades before arriving to your PROD cluster
- **Rolling restart approach**
  - No downtime and transparent to the users
  - Exception: clusters based on cold storage due to single replica on CephFS
- **High availability**
  - Automatic master failover
  - Automatic cluster rebalancing on data node failures

Communication

- **ServiceNow**
  - Open tickets to the OpenSearch FE to report problems or ask questions.
- **Mattermost**
  - General channel for announcements (IT OpenSearch Service)
  - Dedicated channels with each team for direct support.
- **Documentation**
  - https://opensearch.docs.cern.ch
- **User Forum**
  - To discuss service status and plans and used’s feedback

News & latest changes

cluster aliases, index templates, multi-tenancy, alerting, user docs

Cluster aliases & terminology

- Added **os-cluster.cern.ch** alias for all OpenSearch clusters
  - Your es-* aliases (es-cluster.cern.ch) are deprecated
- Added **/os** suffix for direct OpenSearch communication
  - The /es suffix is deprecated

```
output {
  opensearch {
    hosts => "https://es-cluster.cern.ch:443/es"
    …
    user => "<user>"
    password => "<password>"
  }
}
```

→

```
output {
  opensearch {
    hosts => "https://os-cluster.cern.ch:443/os"
    …
    user => "<user>"
    password => "<password>"
  }
}
```

Please update your pipelines!

- **September 2nd, 2024**: Generate report and notify people still using *https://es-cluster.cern.ch/es*
- **September 16th, 2024**: Delete *es-cluster.cern.ch alias and /es path suffix*

Define index mappings and settings

- Initialize new indexes with predefined **index templates**
  - *mappings*: expected types for each document field (e.g., text/keyword, date formats, long/double)
  - *settings*: number of shards/replicas, refresh interval, etc.
- Organise your index templates with **component templates**
  - A set of settings/mappings to be reused on multiple index templates

Defining templates is crucial for optimal performance and disk optimization

…and for preventing painful reindexing operations

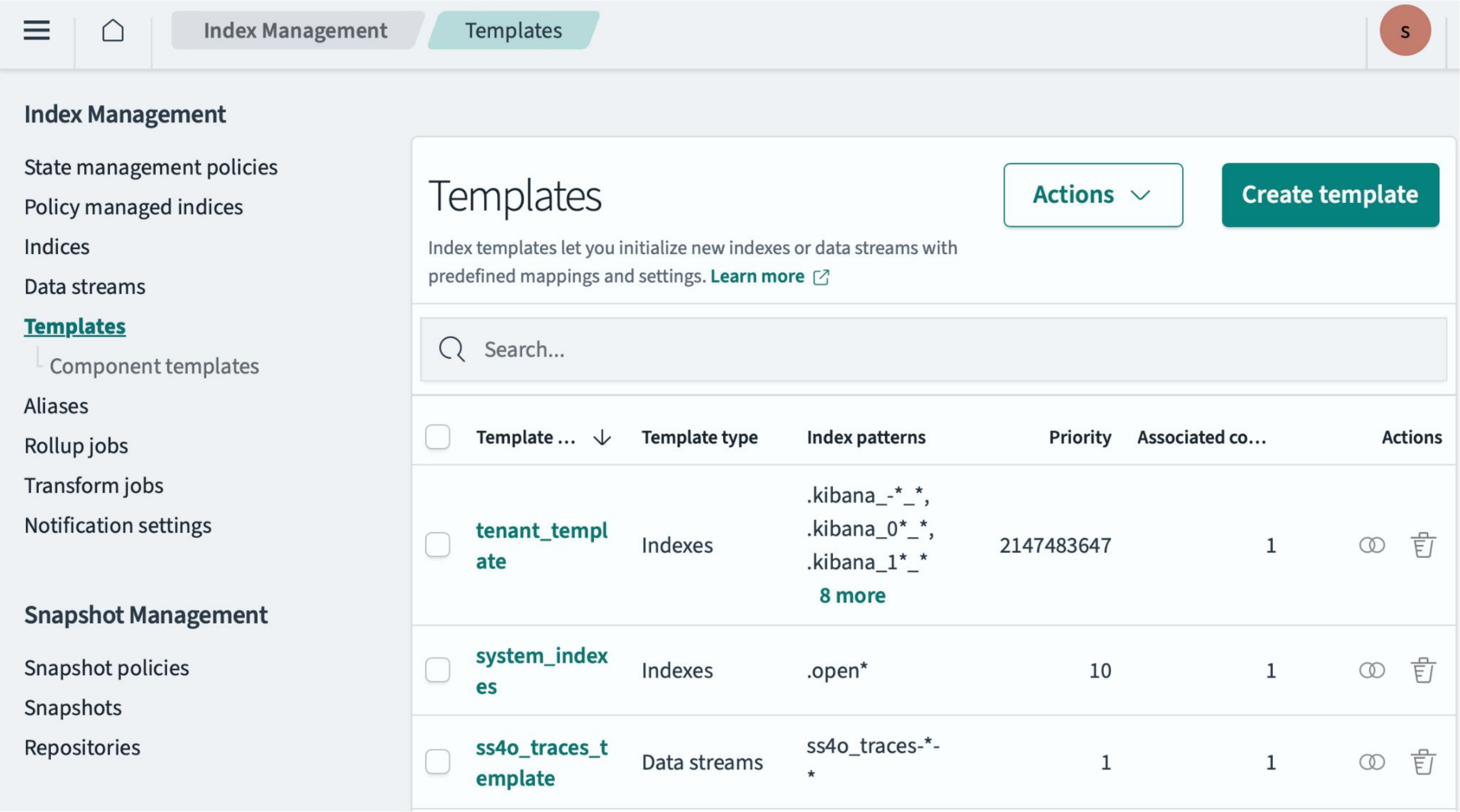
Index templates management

- The old **_template** API is *deprecated*

- Migration to the newer **_index_template** API has been completed

- Usage of the old API is monitored and users will be contacted if still using it

You can now manage your templates in the UI or with the proper APIs



Templates

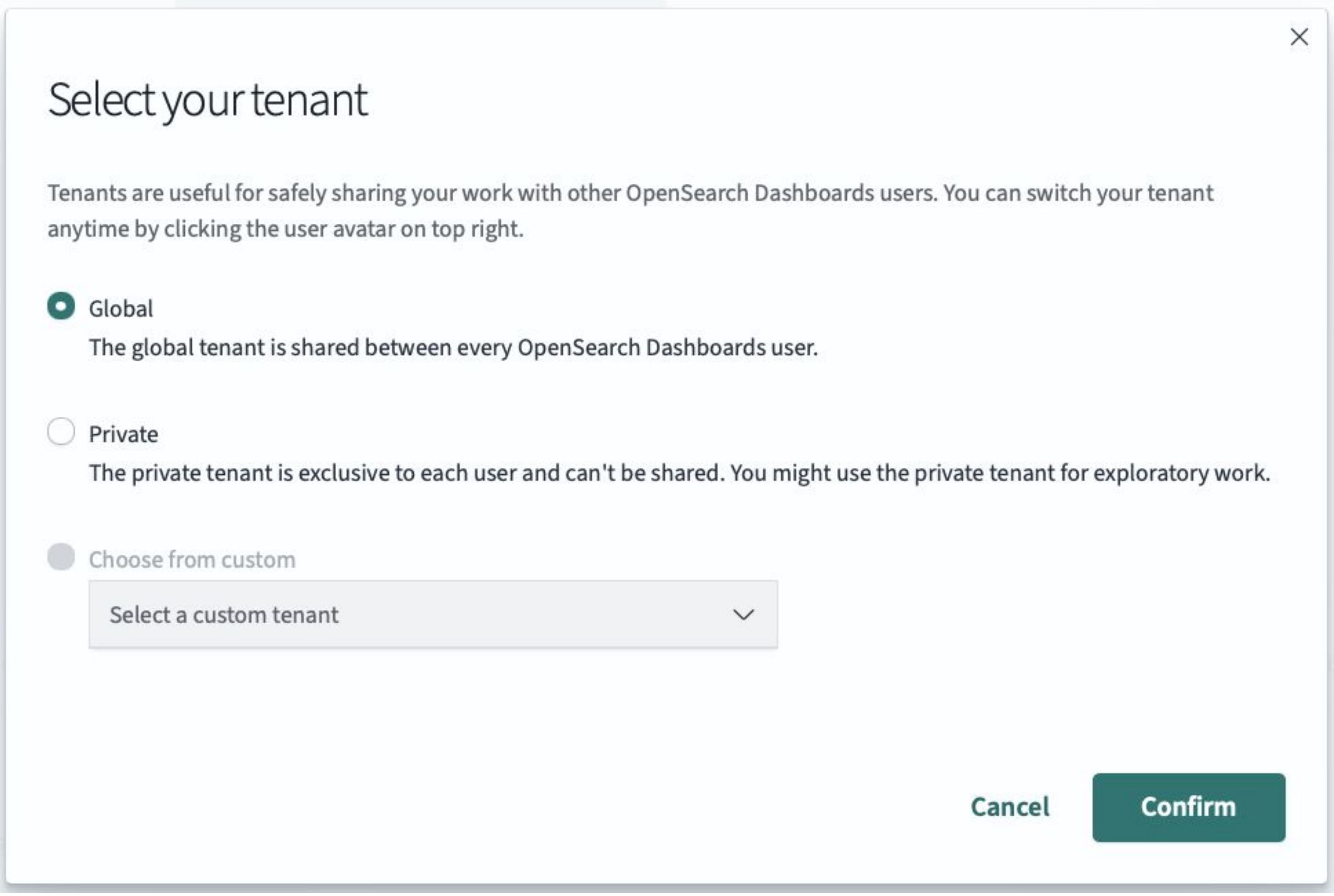
Index templates let you initialize new indexes or data streams with predefined mappings and settings. Learn more

□	Template ... ↓	Template type	Index patterns	Priority	Associated co...	Actions
□	tenant_template	Indexes	.kibana_-*_*, .kibana_0*_*, .kibana_1*_* 8 more	2147483647	1	
□	system_indexes	Indexes	.open*	10	1	
□	ss4o_traces_template	Data streams	ss4o_traces-*-*	1	1	

Multi-tenancy feature is now optional

- Tenants are spaces for saving objects like visualisations, dashboards, etc.
  - Isolate different teams' exploration in the same cluster
- Multi-tenancy is now disabled by default, unless already used
  - Benefit: pop-up for selecting tenant will no longer appear

We can enable it upon request



Select your tenant

Tenants are useful for safely sharing your work with other OpenSearch Dashboards users. You can switch your tenant anytime by clicking the user avatar on top right.

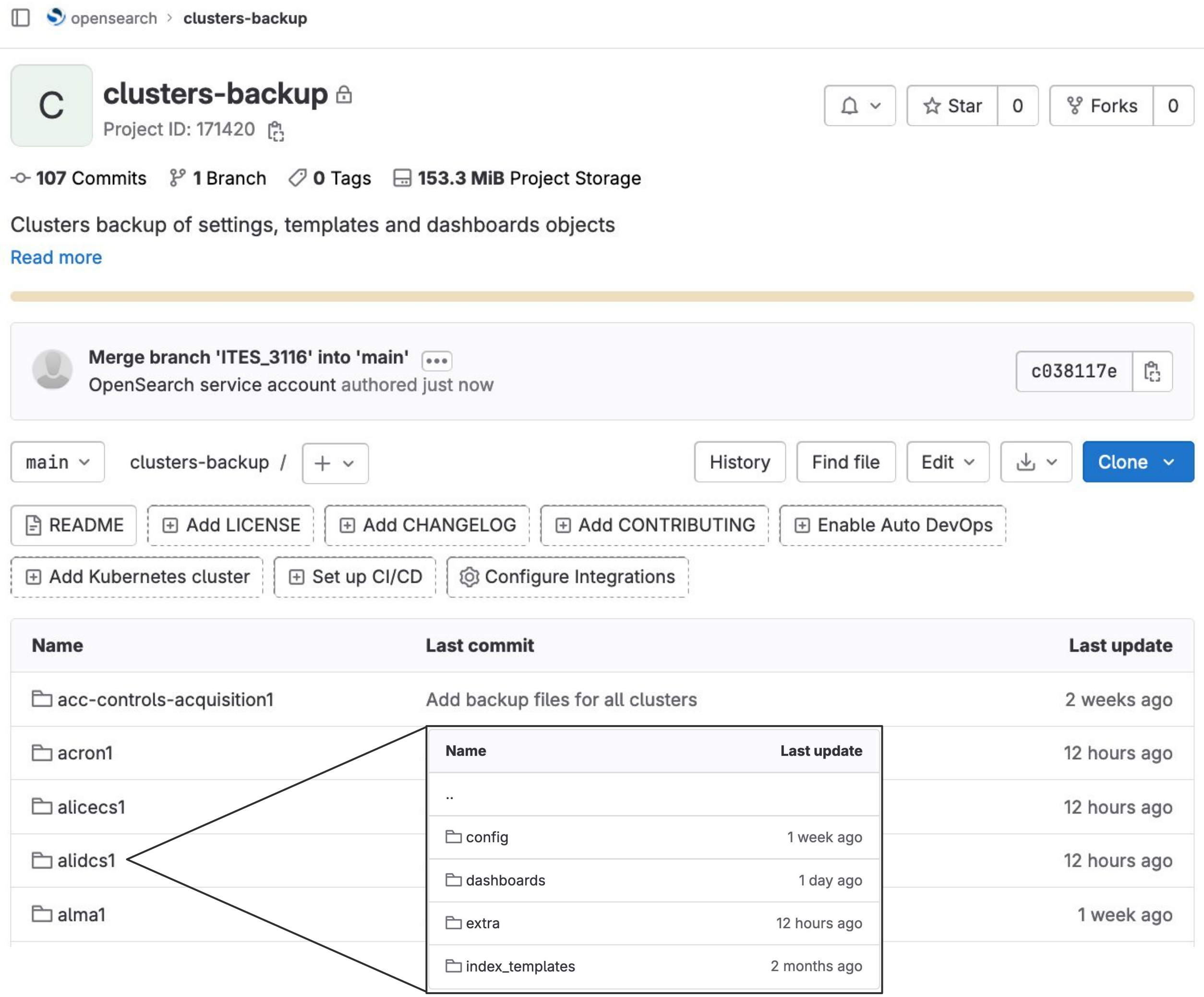
● Global
The global tenant is shared between every OpenSearch Dashboards user.

○ Private
The private tenant is exclusive to each user and can't be shared. You might use the private tenant for exploratory work.

○ Choose from custom

Backup of cluster configuration

- A central clusters-backup repo daily backs up all clusters' information
  - Cluster settings, index templates, index policies, alerting, etc.
- Dedicated *endpoint-cluster-settings* repos are archived



opensearch › clusters-backup

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

main · clusters-backup /

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

Alerting - Default monitors

- **Each OpenSearch cluster** comes by default with a number of **default monitors**
  - Cluster health → alerts sent to service experts
  - Cluster usage → alerts sent to **you** (Mattermost / SNOW / Email)

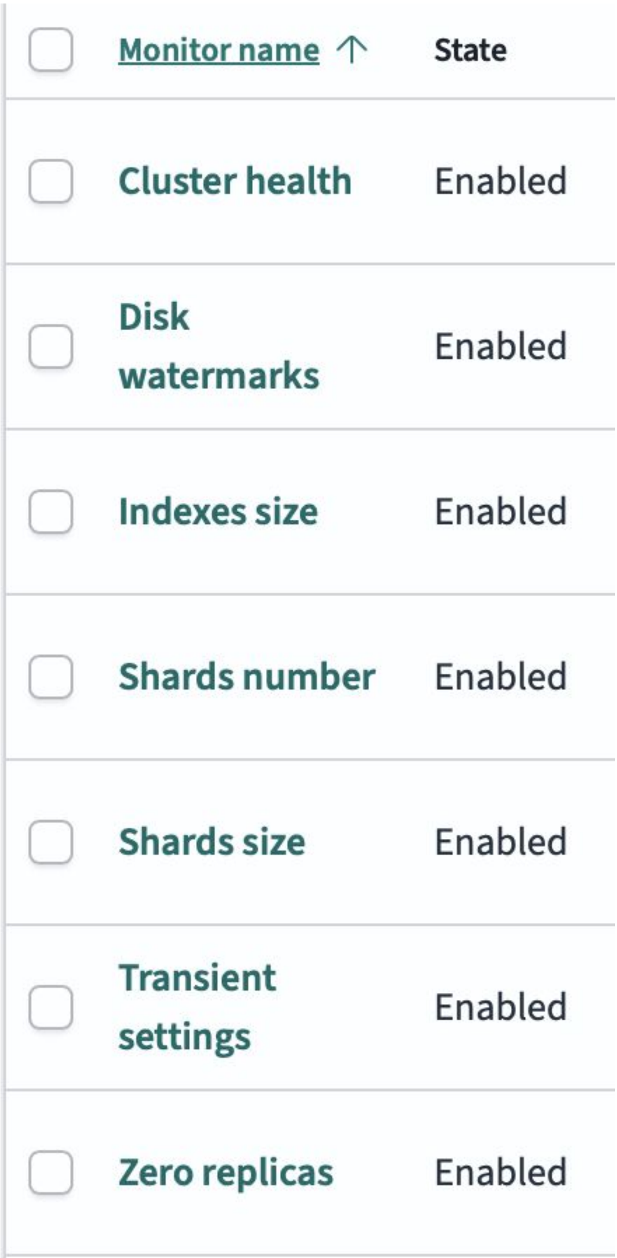
Feel free to create more monitors based on your use-case using OpenSearch Alerting

Alerting - Default monitors

1. Cluster health
   Monitors if cluster gets to “yellow” or “red” status
2. Disk watermarks
   Cluster is getting full, unexpectedly large or unrotated indexes
3. Indexes size
   Reports an index with zero documents
4. Shards number
   Stored shards are too many for your cluster capacity
5. Shards size
   Too big shards
6. Transient settings
   Reports if transient cluster settings are used
7. Zero replicas
   Ensure that all indexes have at least one replica
8. Indexes with no retention policy
   Review indexes that are set to live forever

extra

running externally

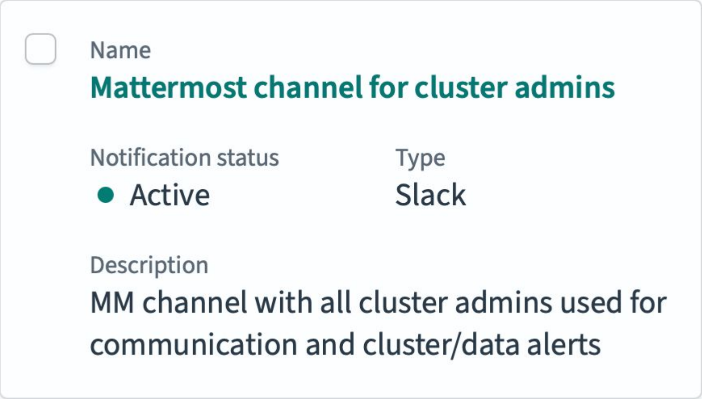


□	Monitor name ↑	State
□	Cluster health	Enabled
□	Disk watermarks	Enabled
□	Indexes size	Enabled
□	Shards number	Enabled
□	Shards size	Enabled
□	Transient settings	Enabled
□	Zero replicas	Enabled

Alerting - Default notification channels

- Time-sensitive alarms run every minute/hour and alert immediately
- Non-urgent alarms will be sent ***every Monday at 14:00***
  - You will keep getting them until the reported problem is fixed

Request to be <u>excluded</u> from a monitor, if not relevant for you

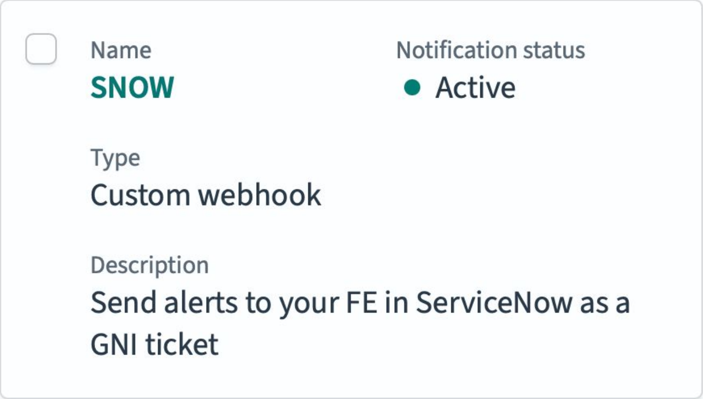


☐ Name
Mattermost channel for cluster admins

Notification status
● Active

Type
Slack

Description
MM channel with all cluster admins used for communication and cluster/data alerts

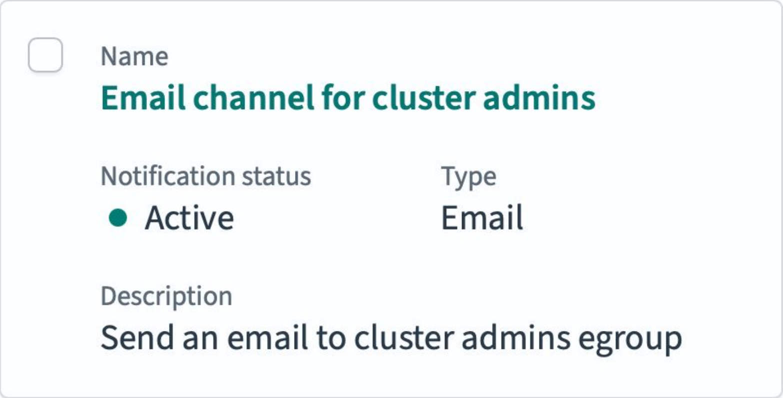


☐ Name
SNOW

Notification status
● Active

Type
Custom webhook

Description
Send alerts to your FE in ServiceNow as a GNI ticket



☐ Name
Email channel for cluster admins

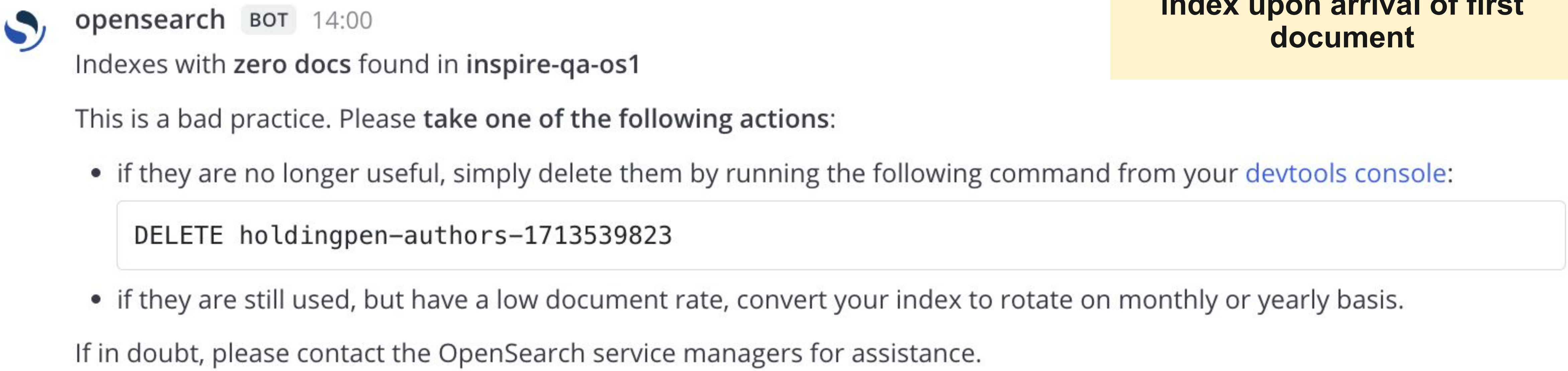
Notification status
● Active

Type
Email

Description
Send an email to cluster admins egroup

Alerting - Examples

Let OpenSearch create the index upon arrival of first document



opensearch BOT 14:00

Indexes with **zero docs** found in **inspire-qa-os1**

This is a bad practice. Please **take one of the following actions**:

- if they are no longer useful, simply delete them by running the following command from your devtools console:

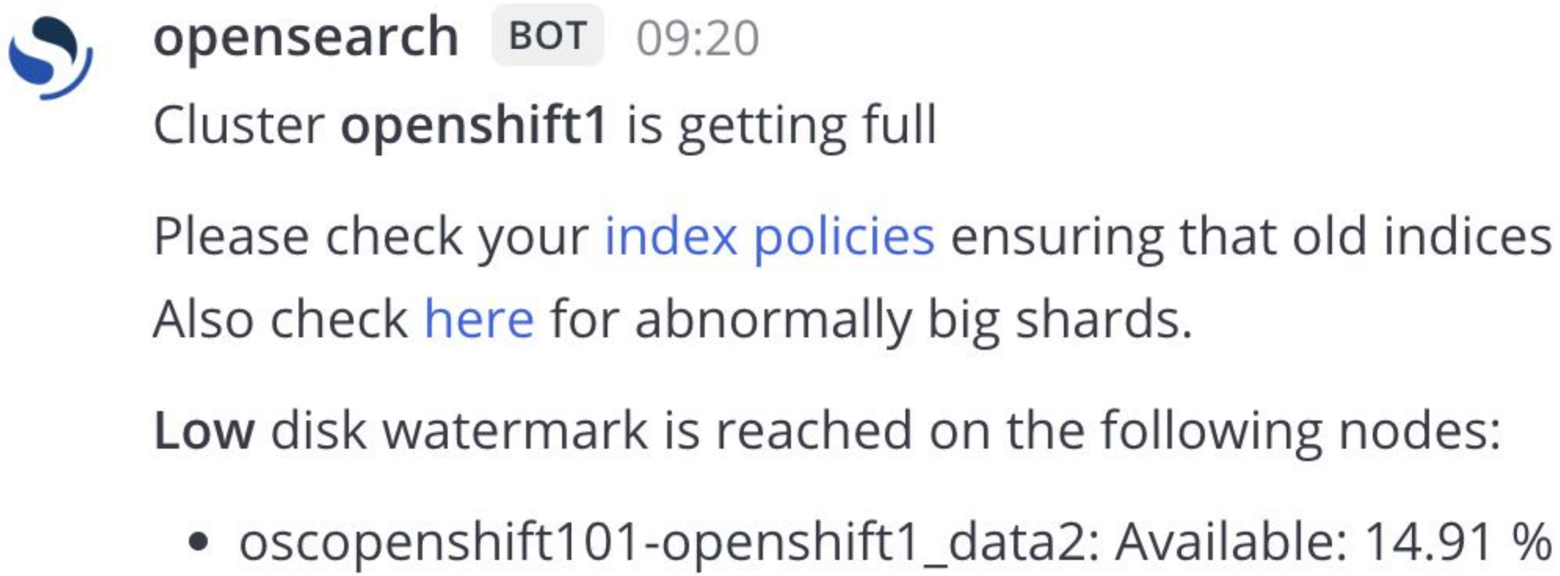
```
DELETE holdingpen-authors-1713539823
```

- if they are still used, but have a low document rate, convert your index to rotate on monthly or yearly basis.

If in doubt, please contact the OpenSearch service managers for assistance.

Alerting - Examples

Cleanup or request for quota extend



opensearch BOT 09:20

Cluster **openshift1** is getting full

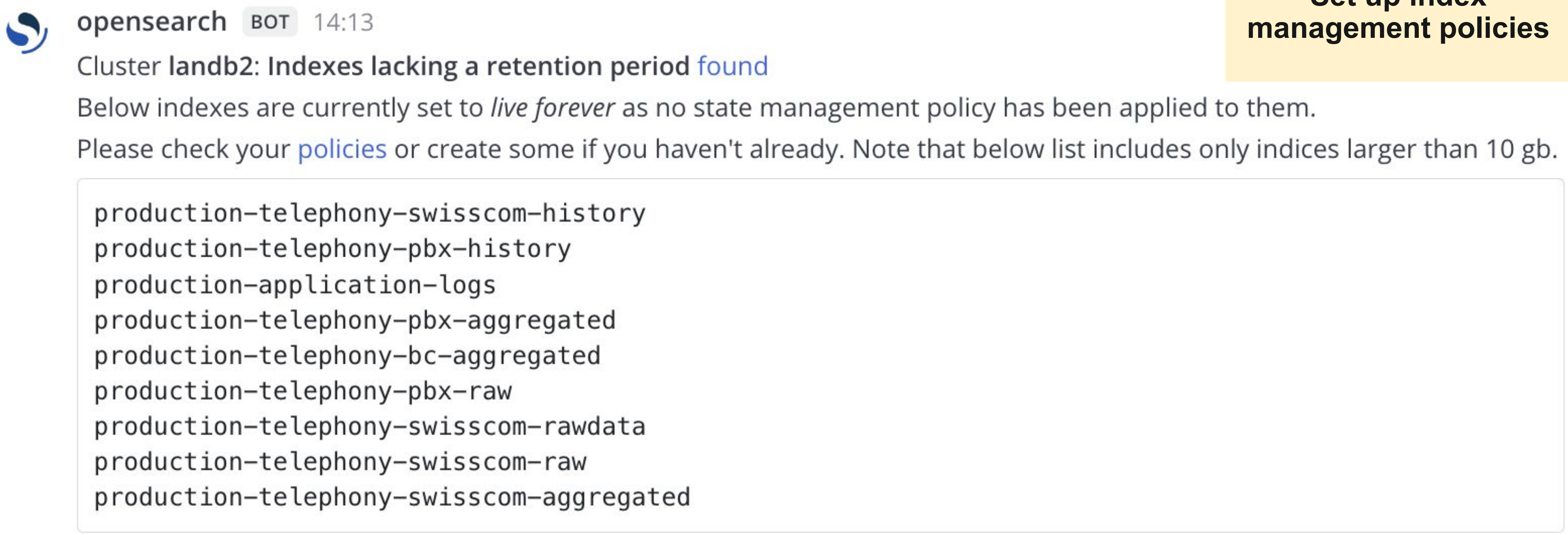
Please check your index policies ensuring that old indices have been deleted.
Also check here for abnormally big shards.

Low disk watermark is reached on the following nodes:

- oscopenshift101-openshift1_data2: Available: 14.91 %

Alerting - Examples

Set up index management policies



opensearch BOT 14:13

Cluster **landb2: Indexes lacking a retention period** found

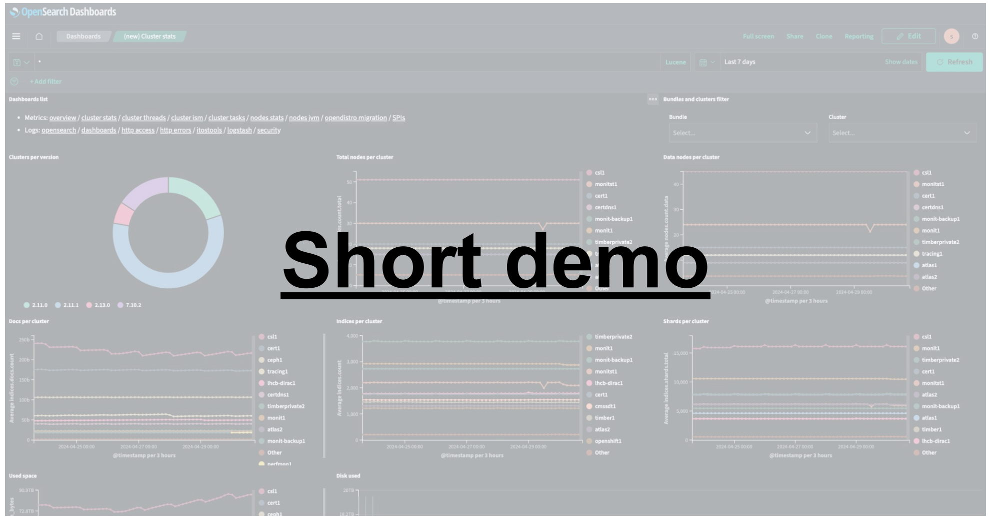
Below indexes are currently set to *live forever* as no state management policy has been applied to them.

Please check your policies or create some if you haven't already. Note that below list includes only indices larger than 10 gb.

```
production-telephony-swisscom-history
production-telephony-pbx-history
production-application-logs
production-telephony-pbx-aggregated
production-telephony-bc-aggregated
production-telephony-pbx-raw
production-telephony-swisscom-rawdata
production-telephony-swisscom-raw
production-telephony-swisscom-aggregated
```

Monitoring

- Our perfmon cluster collects logs & metrics centrally for all clusters in the service
  - OpenSearch, Dashboards, Apache, Security audit, Cluster stats (e.g. disk used/free)



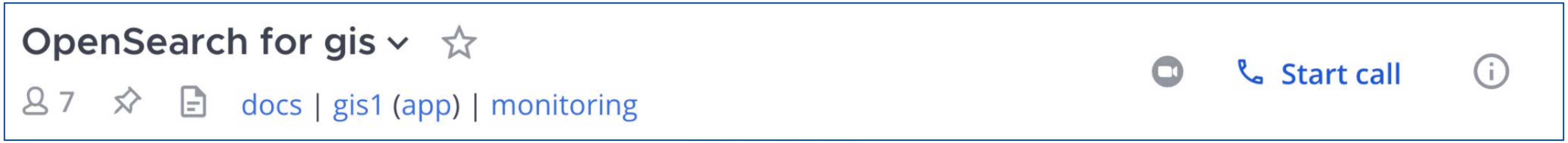
- Metrics: overview / cluster stats / cluster threads / cluster ism / cluster tasks / nodes stats / nodes jvm / opendistro migration / SPIs
- Logs: opensearch / dashboards / http access / http errors / itostools / logstash / security

Short demo

You have access to all your cluster(s) logs, so feel free to explore them

User docs & troubleshooting

- Link found on your Mattermost channel header



- *Cluster administration* section will give you plenty of information for your everyday tasks

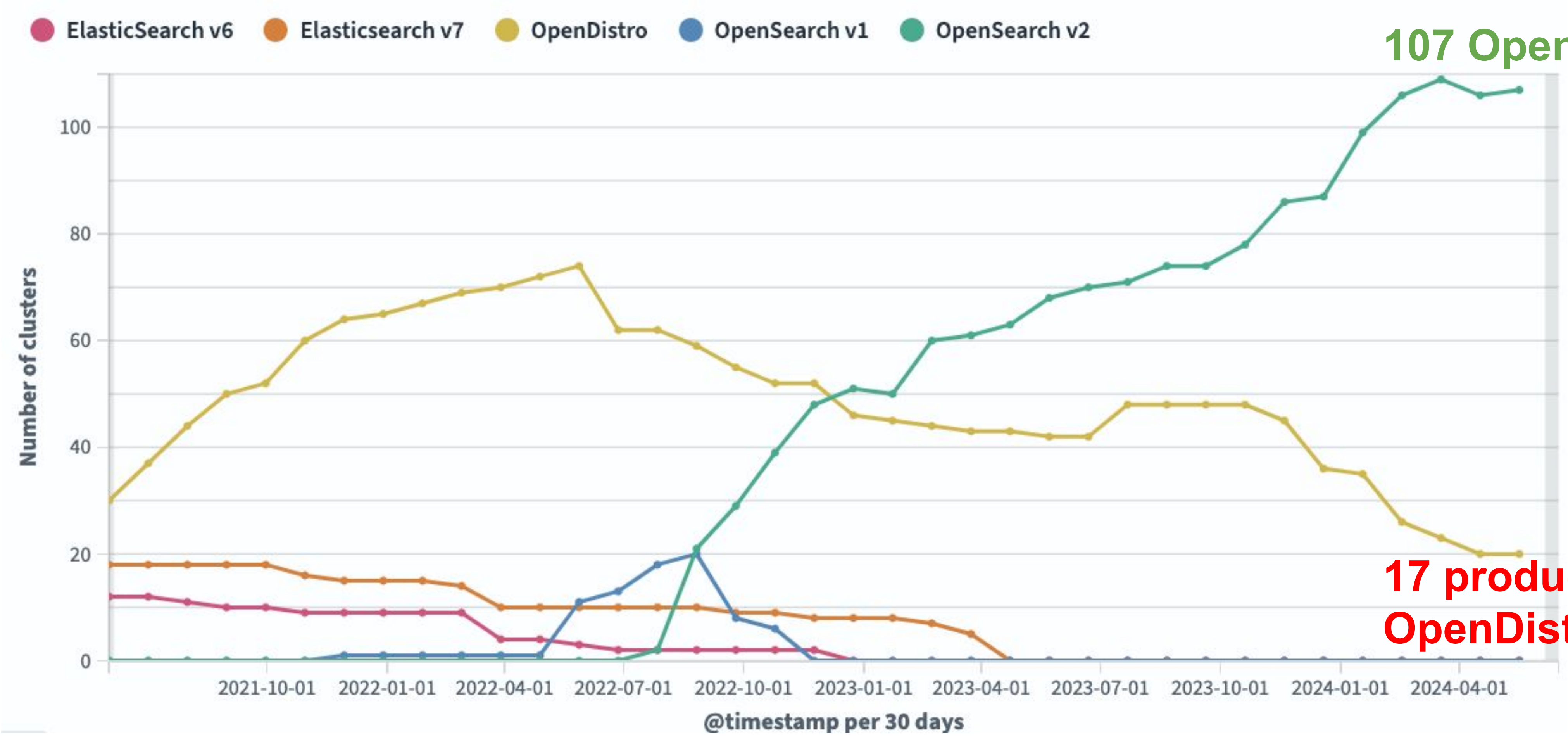
User documentation
https://opensearch.docs.cern.ch

Roadmap

migrations, upgrades, backups, data streams, K8s deployment, and more

OpenDistro to OpenSearch migration

2024-Q4



107 OpenSearch

17 production OpenDistro left

OpenSearch upgrades

- Upgrade campaign from v.2.11 to **v2.14** with rolling restarts
  - Couple of annoying bug fixes
    - index pattern selection, monitor definition text
  - New match-only text field for large-datasets (optimized storage)
  - Apache Spark integration enhanced
  - Plethora of VectorDB related enhancements
  - Performance improvements
  - More details: https://opensearch.org/blog/explore-opensearch-2-14

2024-Q3

Play with the latest versions here
https://opensearch-playground.cern.ch

- Upgrade campaign from v2.14 to **2.17**
  - Roadmap: https://github.com/orgs/opensearch-project/projects/1

2024-Q4

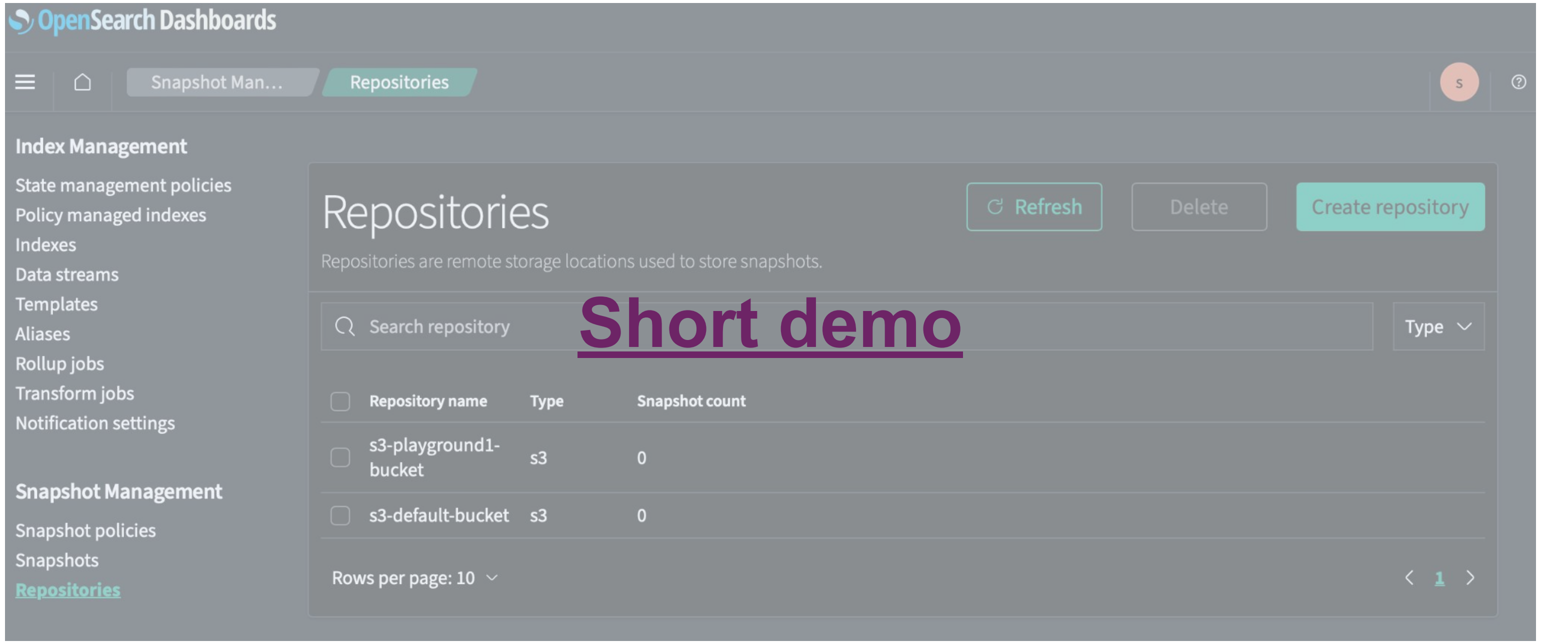
Backup of cluster data (opt in)

2024-Q2

- OpenSearch Snapshots feature is used for disaster recovery
- Snapshots are stored in S3 cluster in Prevessin
  - Physical separation in regards with OS clusters running in bldg. 513
- Notifications of snapshots such as creation, failure or deletion can be sent the CERN standard channels (Mattermost/SNOW/Email)
- Snapshots of your data are incremental, typically configured daily
  - First snapshot is slow to complete (FULL) and sequential ones are fast (INCR)
- Currently exploring searchable snapshots feature to replace Ceph cold storage
- OpenSearch User docs with more details: https://opensearch.docs.cern.ch/cluster_admin/snapshots/

Backup of cluster data - demo

2024-Q2



Repositories

Repositories are remote storage locations used to store snapshots.

Repository name	Type	Snapshot count
s3-playground1-bucket	s3	0
s3-default-bucket	s3	0

Short demo

If you want your data to be backed up, please raise a request

Data Streams

Data streams handle time-series data by managing it in sequential, time-based indices.

2024-Q3

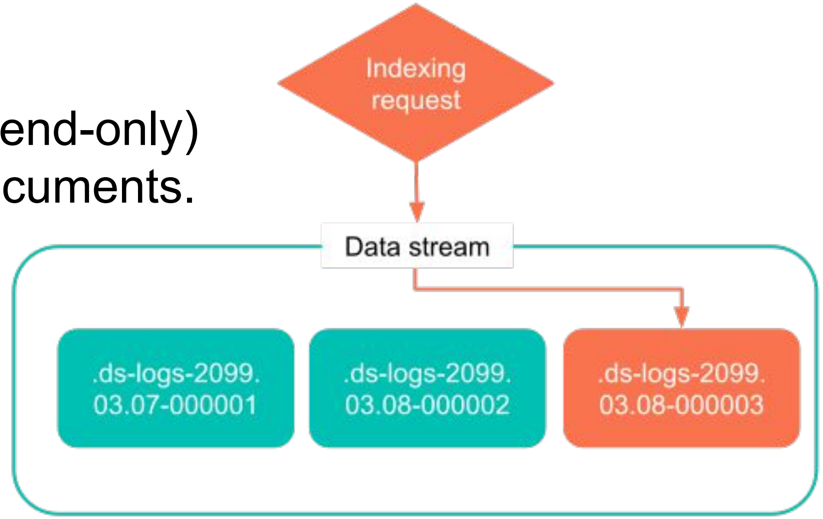
Benefits:

- **Efficient storage:** optimize storage by using time-based indices, reducing storage costs.
- **Improved performance:** better performance for write-heavy workloads - optimized for sequential writes.
- **Simplified management:** automated index management, e.g. rollover.
- **Scalability:** scalable to accommodate large volumes of data, crucial for high-throughput log data.

Ideal For:

- Log analysis, time-series data, continuous data ingestion scenarios (append-only)
- where the # of docs grows rapidly and you don't need to update older documents.

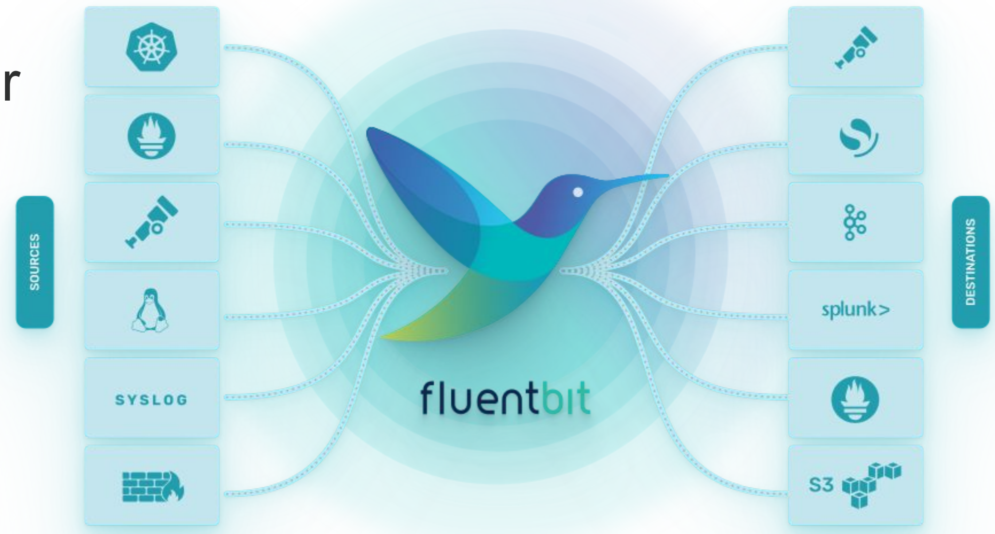
For more details, visit our User docs: OpenSearch Data Ingestion.

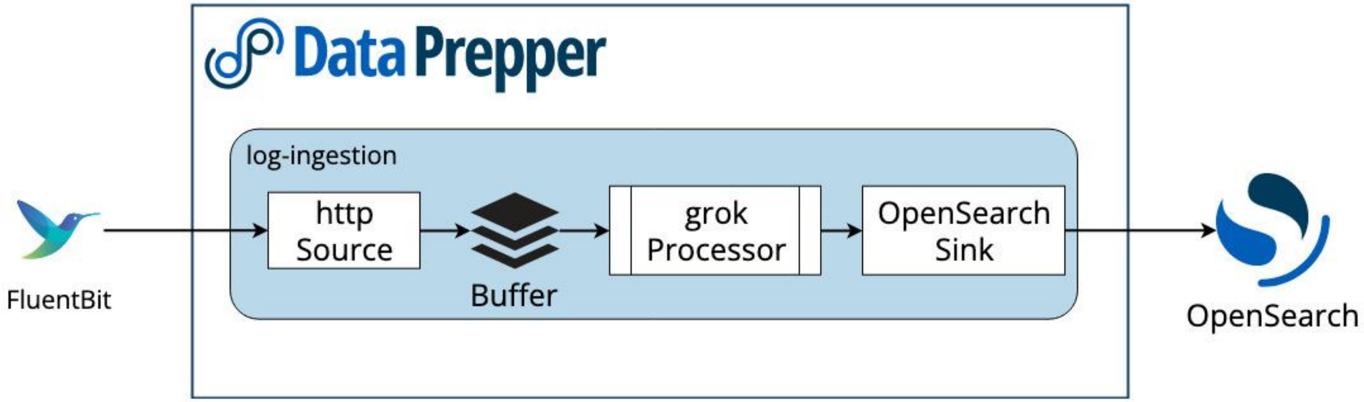

Indexing request
Data stream
.ds-logs-2099.03.07-000001
.ds-logs-2099.03.08-000002
.ds-logs-2099.03.08-000003

Logstash alternatives

2024-Q3

- Replacement of Logstash (ELK) planned for this year
  - Only supported via a plugin in OpenSearch
- **Why?**
  - Mitigate potential licence risk (dual Apache/Elastic)
  - Resource-intensive (Java-based)
  - Complex to manage for growing infrastructure
  - IT MONIT replaces Flume by Fluent Bit
- **Alternatives**
  - Fluent Bit / Fluentd (Apache License 2.0)
  - DataPrepper (MIT License)



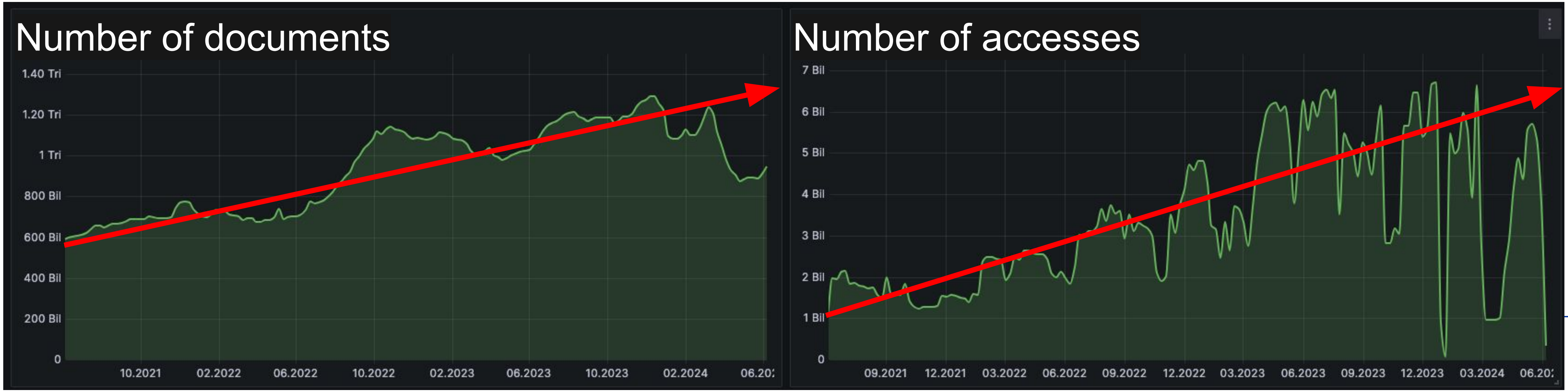


Long-term roadmap (beyond 2024)

- Exploring OpenSearch deployment on **Kubernetes**
  - Begin with a summer project
  - To provide more flexibility, scalability and resource optimization (cost-effectiveness)

- Explore **VectorDB** capabilities
  - Example: a chatbot for CERN internal knowledge bases such as accGPT, itGPT
  - Some update on the developments of AI/LLM at CERN: https://indico.cern.ch/event/1397765/

- Increase OpenSearch **community** engagement and handle service **growing needs**


Number of documents

Number of accesses

Service survey

Let’s explore your answers

https://os-perfmon.cern.ch/dashboards/app/observability-notebooks#/7x9h7Y8Bxqh9I1-2a6QM?view=output_only

Live Q&A

Time for questions and answers

https://liveqna.web.cern.ch/event/3V6N5V

Summary

- 17 **OpenDistro prod** clusters left to migrate to OpenSearch
- Please start using the new **cluster aliases** in all your pipelines
- Define **index templates** for your indices to optimise data management
- Visualise your cluster **logs and metrics** on perfmon (central monitoring)
- Request for data **snapshots** or **multi-tenancy** if interested
- Take action on **alarms** landing on your Mattermost channel/SNOW
- Upgrades, Data Streams, Kubernetes and Vector DBs on the horizon…

home.cern