Moving from Elasticsearch to OpenSearch at CERN

I. A bit of history
II. Motivation for change
III. Design and implementation
IV. Current state
V. What's next

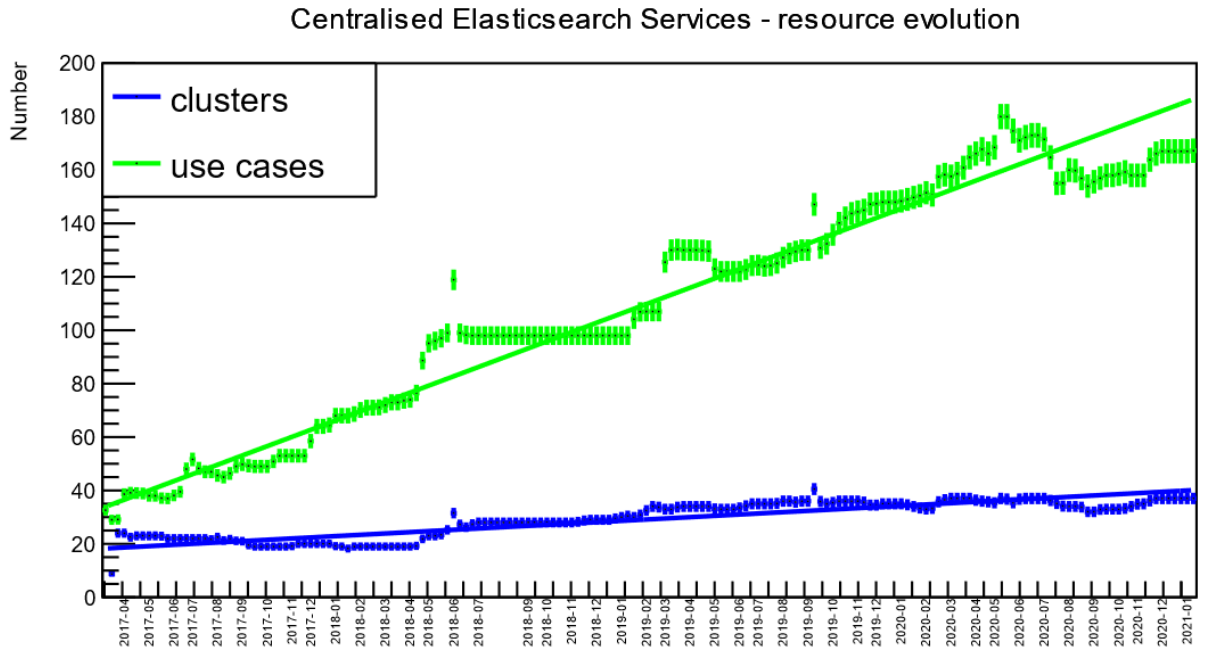
Sokratis Papadopoulos
it-elasticsearch-admin@cern.ch
HEPiX Spring 2022

What is Elasticsearch and Kibana

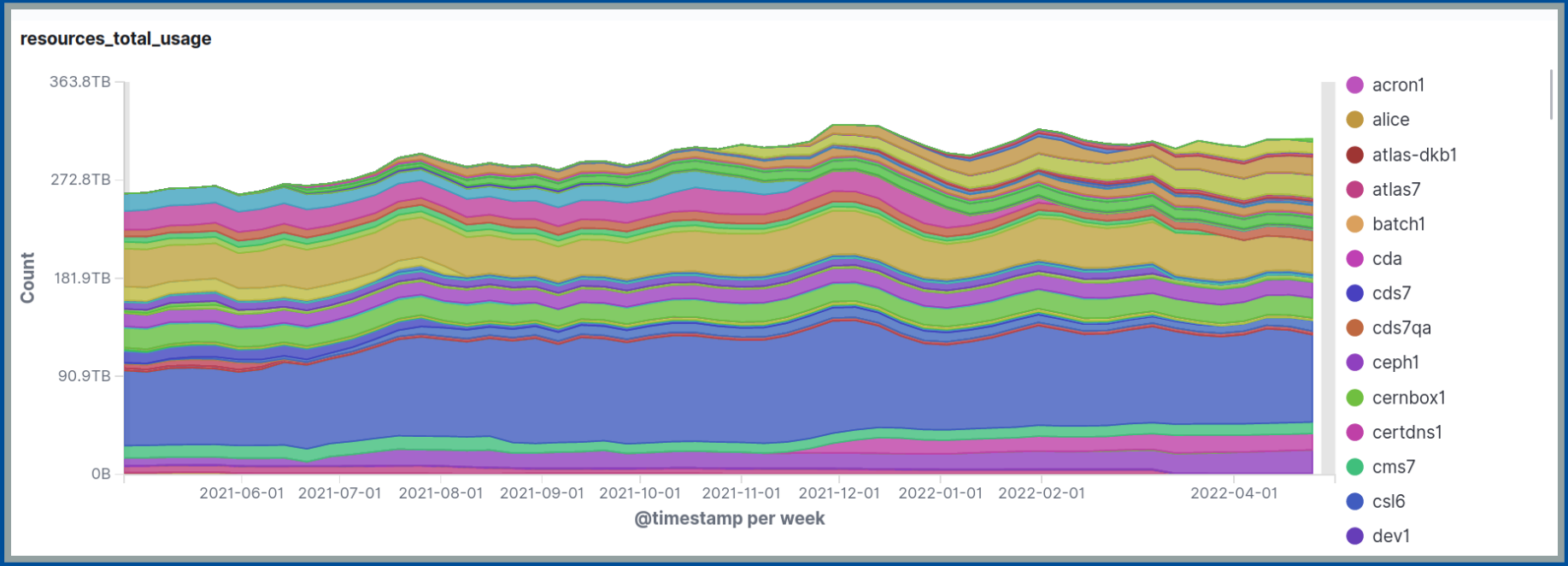
- **Elasticsearch** is a distributed, search and analytics engine
- **Kibana** is the web user inteface that lets you visualise your Elasticsearch data
- Usage at CERN & HEP:
  - ALICE, ATLAS, CMS, LHCb, NA62, ...
  - Beams, INSPIRE, ...
  - IT: Monitoring, Security, Storage, ...

A bit of history

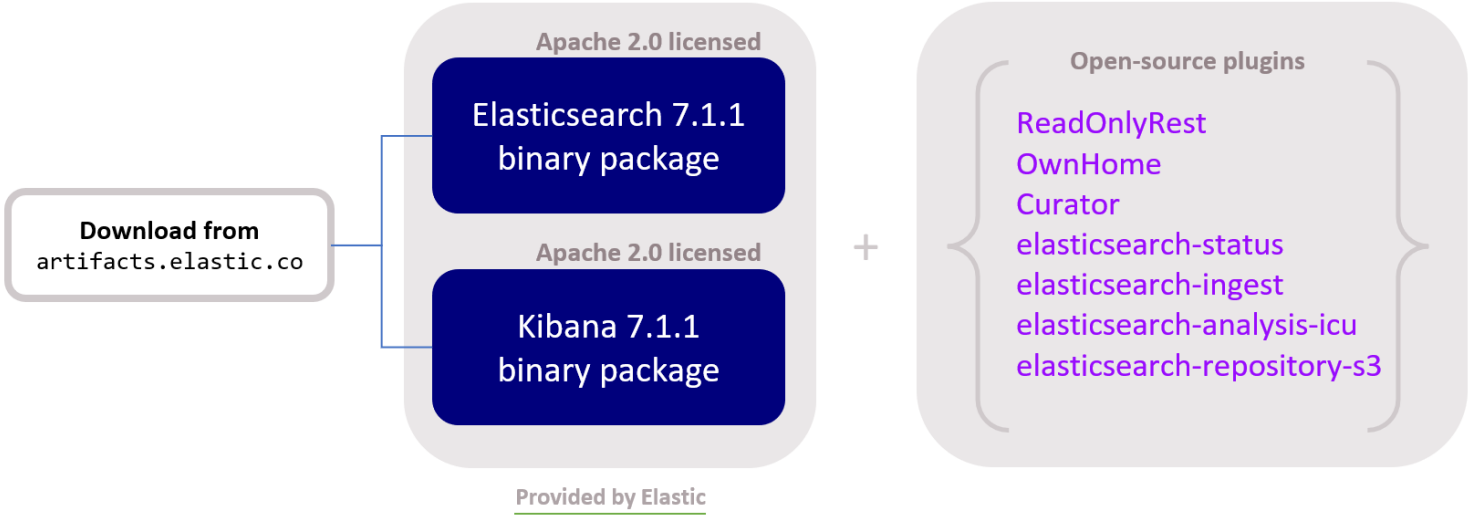
- Centralised Elasticsearch + Kibana instances since 2016



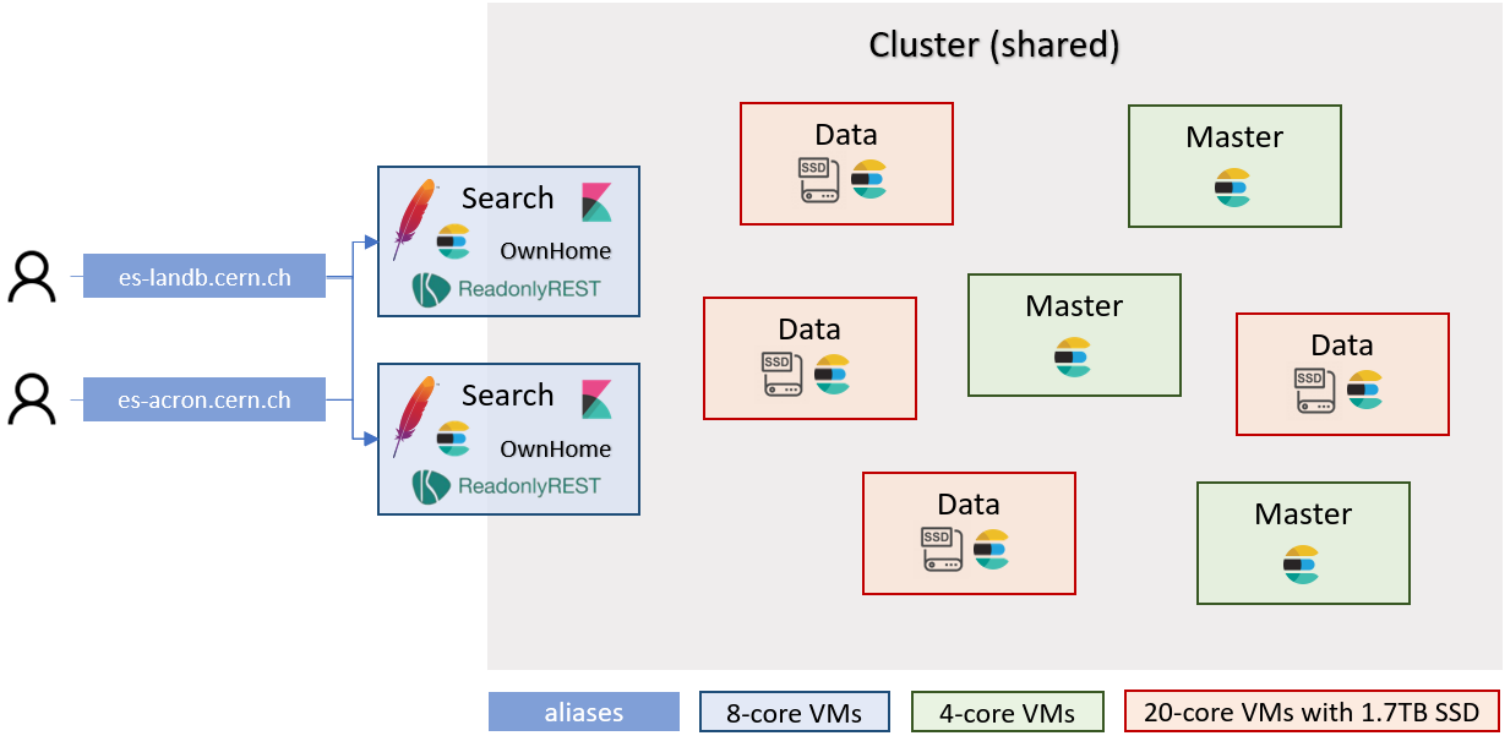
Resources used



Legacy service: packages + plugins



Legacy service: design

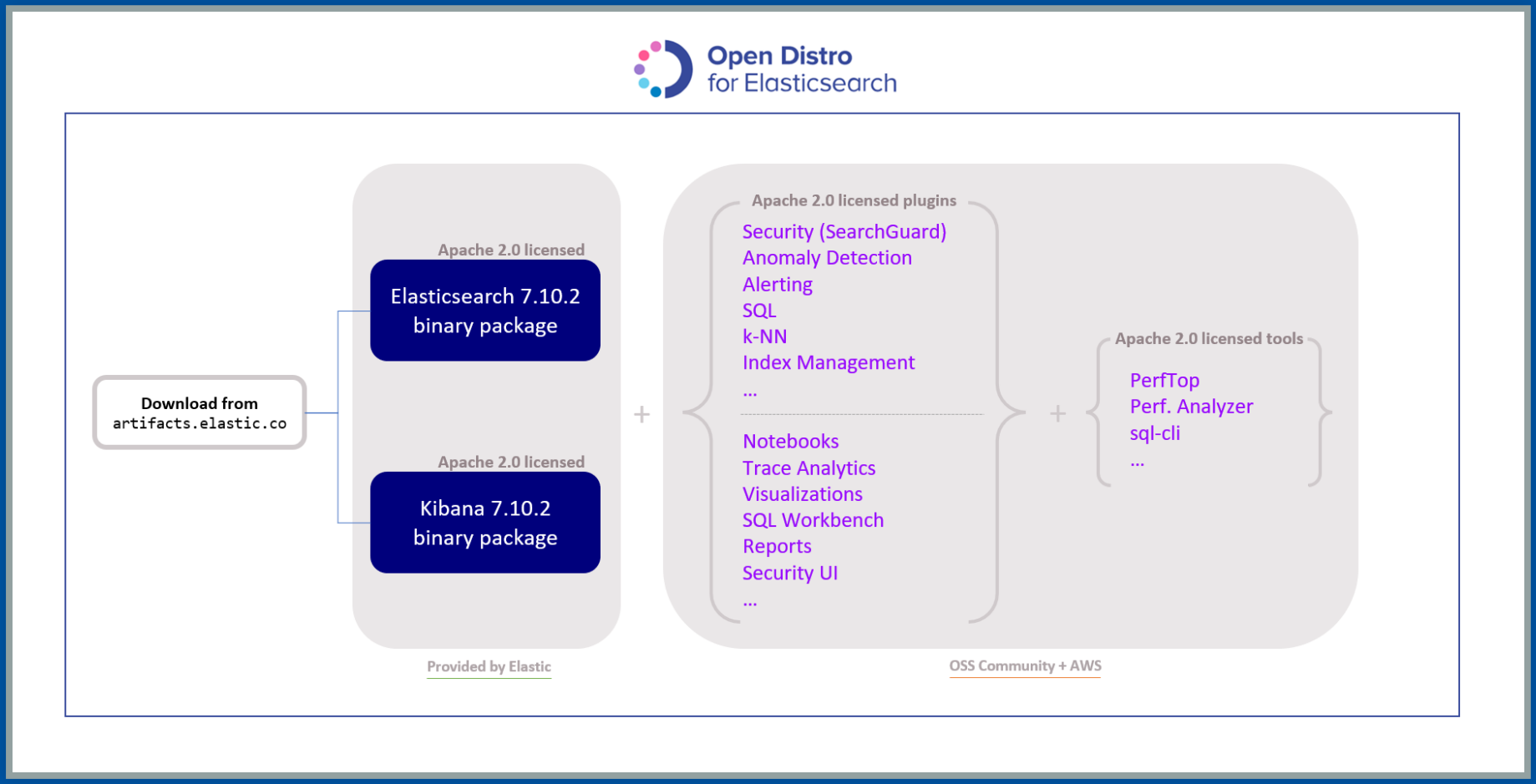


Pros and cons of the legacy design

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

Evaluation of OpenDistro for Elasticsearch

- Oct 2020: Evaluating OpenDistro, stewarded by AWS
- A “complete” open source Elasticsearch + Kibana product



Open source no more for Elastic

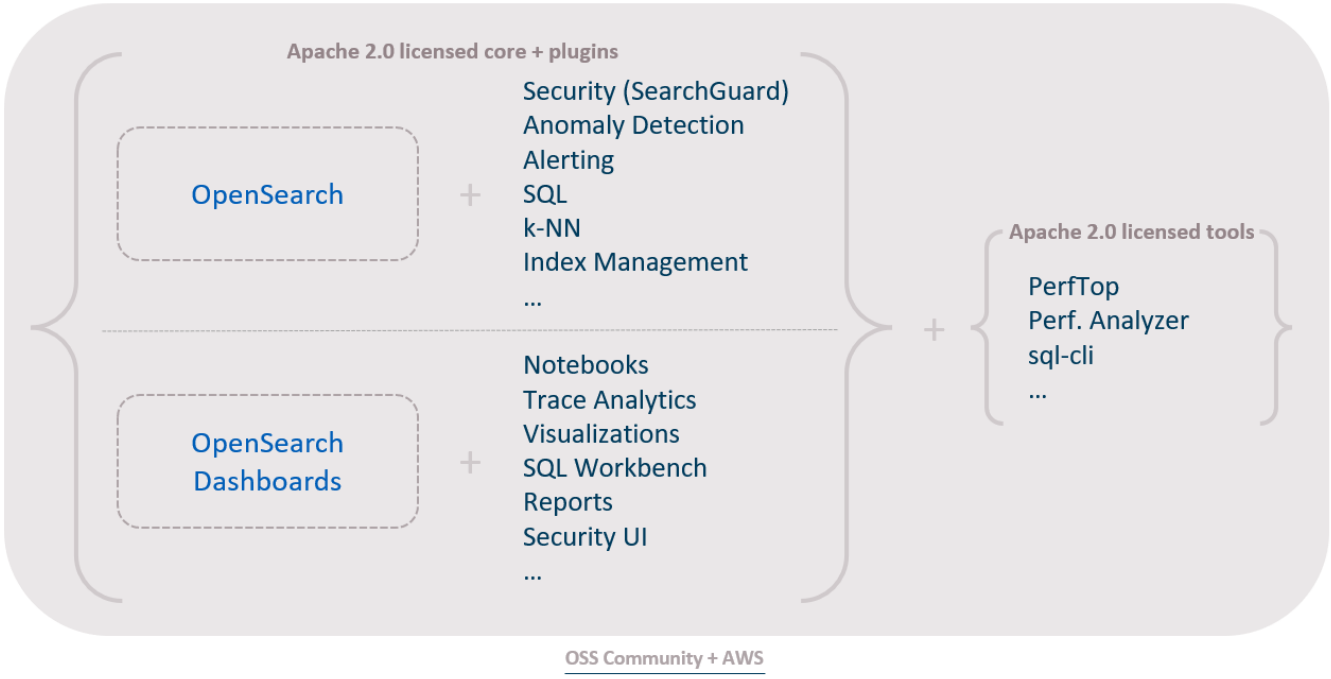
- As of v7.11 (January 2021) Elastic no longer offers an open source version of ES+Kibana
  - Still unclear whether we can work under their new (SSPL/Elastic) license
- AWS has decided to fork the latest ES+Kibana open source version (7.10.2) and OpenSearch was born
  - OpenDistro project is re-branded as OpenSearch
  - Gathered Elastic-disappointed contributors
  - Governance concerns

Why move?

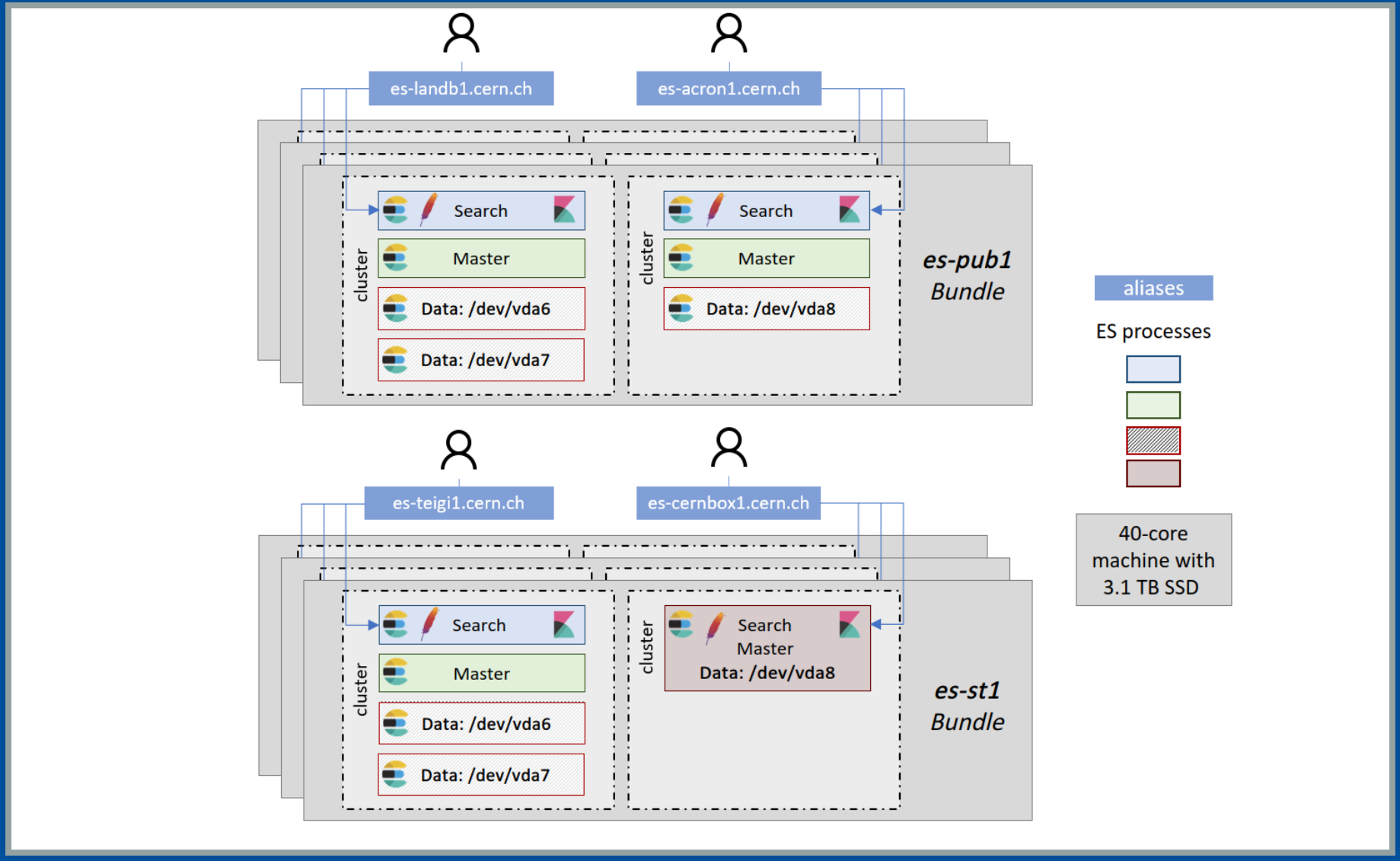
- Licensing
- Maintainability
- Streamlined deployment
- Features

OpenSearch project

Full control over core + plugins. Currently on v1.3



The OpenDistro/OpenSearch architecture

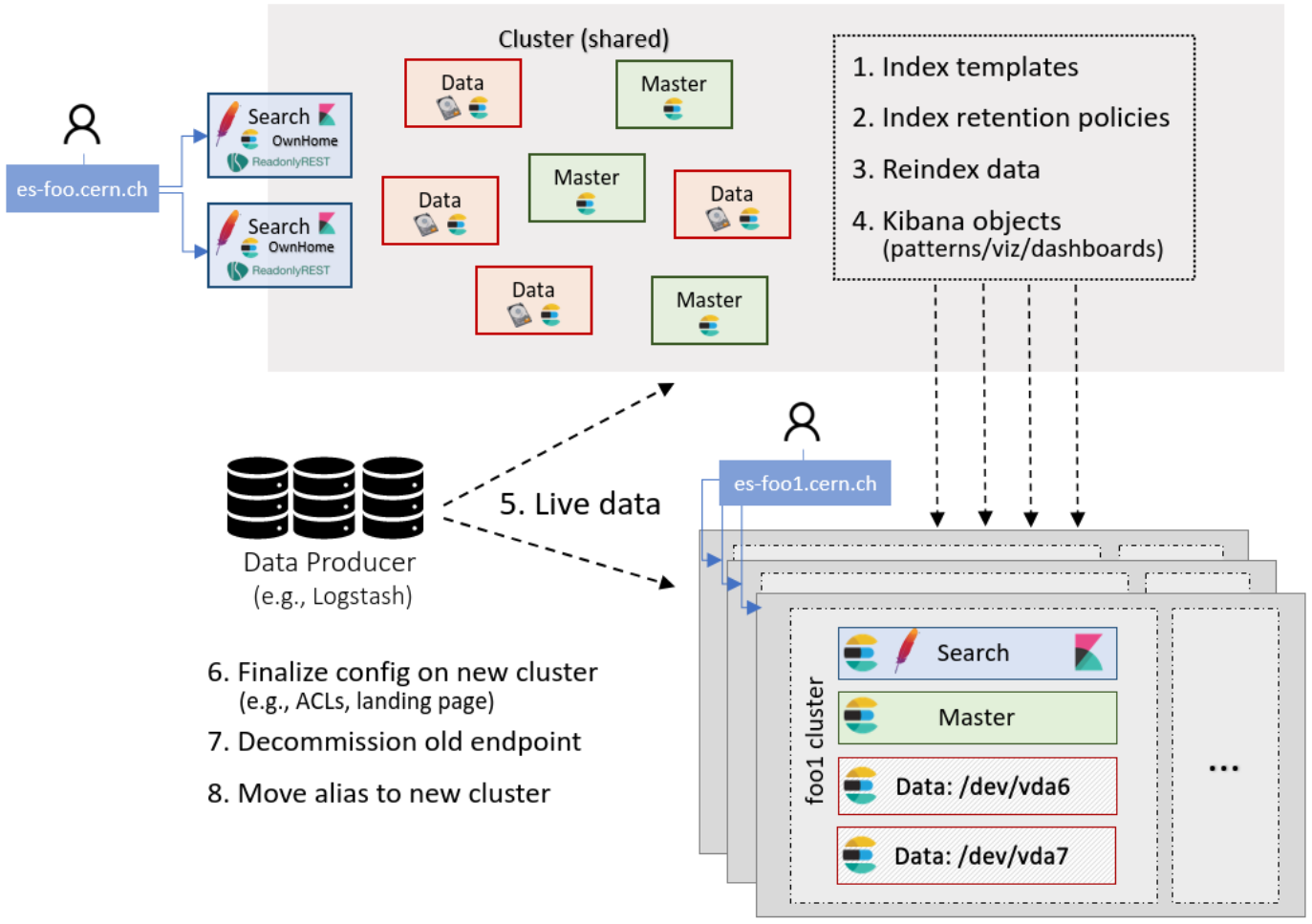


Security model differences

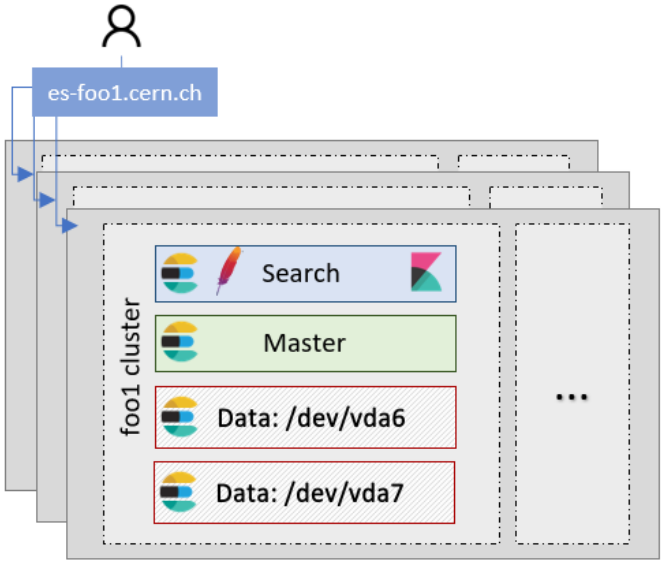
***ReadOnlyRest (Legacy) shared cluster**	**SearchGuard (OpenSearch) dedicated cluster**
external plugin	native plugin
no UI	native, integrated UI
ACL managed centrally	ACL created by the endusers
index-level access control	index, document, and field-level control
index prefix per team	no prefixes required
unencrypted none-to-node	encrypted node-to-node
no default audit logging	audit logging

* CHEP 2018 presentation for the legacy security model

Elasticsearch to OpenDistro migration



OpenDistro to OpenSearch upgrade



Rolling restart upgrade: →

- Stop all Elasticsearch processes
- Remove Elasticsearch rpms and config
- Remount from `/var/lib/elasticsearch/foo1_data` to `/var/lib/opensearch/foo1_data`
- Install OpenSearch rpms and load config
- Start OpenSearch processes

Our current state at CERN

Version	Service instances	EOL
OpenSearch 1.3	1	?
OpenDistro 1.13.2	70	May 2022
Elasticsearch 7.1.1	86	November 2020 (!)
Elasticsearch 6.8.0	4	February 2022

What's next

- Complete ES6+ES7 migration to OpenDistro
- Start upgrading from OpenDistro to OpenSearch
- Participate in the OpenSearch community
- Explore build-in plugins offered within OpenSearch
- Moving to physical hardware: managed by Ironic

Summary

- OpenDistro/OpenSearch brought significant changes both internally and on user side
- Elasticsearch to OpenDistro to OpenSearch migration is half-way there, it has been pretty smooth so far
- Soon to be moving from VMs to physical machines

Backup

Past publications

- HEPiX Autumn 2020: Anomaly Detection for the Centralised Elasticsearch Service at CERN
- CHEP 2019: Large Elasticsearch cluster management
- CHEP 2018: Securing and sharing Elasticsearch resources with Read-onlyREST


CERN