

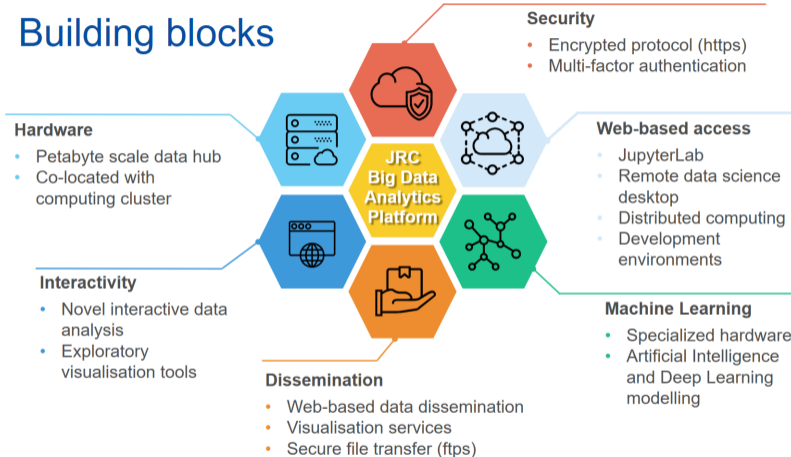
EOS at the Joint Research Centre

Unit I.3 Text and Data Mining

M. Scavazzon, A. Burger, F. Eyraud, P. Tognoli, 5th EOS Workshop, 2021

JRC Big Data Analytics Platform

Building blocks



Homepage: <https://jeodpp.jrc.ec.europa.eu/home/>

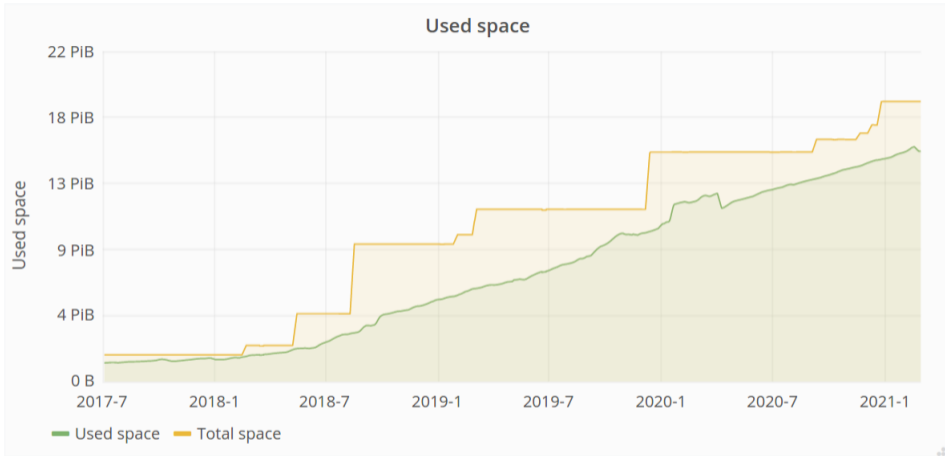
EOS @ JRC - Setup [1]

- ▶ In production since 2016
- ▶ Hardware
 - ▶ JBOD's of 24x6 TB and 24x10 TB disks (soon 60x14)
 - ▶ 1 or 2 JBOD's per FST
 - ▶ 2x10 Gbps Ethernet interconnection per FST
- ▶ Software
 - ▶ OS CentOS 7
 - ▶ MGM v4.5.15 / FST v4.5.17
 - ▶ QuarkDB 0.4.2

EOS @ JRC - Data Access

- ▶ Authentication with Kerberos
- ▶ Data Access
 - ▶ Batch processing cluster
 - ▶ Web-based remote desktops (Guacamole)
 - ▶ Jupyter notebooks
 - ▶ Desktop PCs (CIFS read-only via SAMBA gateway server)
 - ▶ NextCloud
 - ▶ FTPS
- ▶ Clients
 - ▶ fuse: 35
 - ▶ fusex: 50

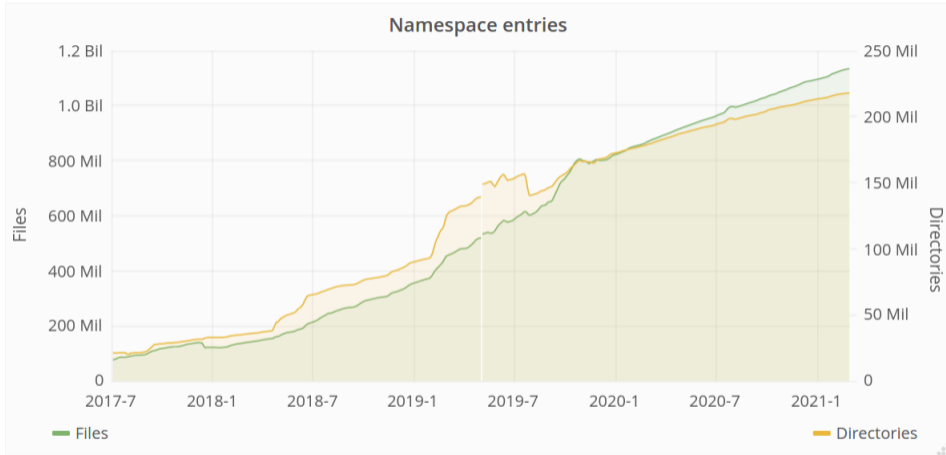
EOS @ JRC - Used Space



► Total space: 19 PiB

Replica 2

EOS @ JRC - Namespace



► Files: 1.1 B

Directories: 218 M

2020 - Changes

- ▶ Space increased by 3.5 PiB (8 new FSTs)
- ▶ Automated FST installation (Foreman + Ansible)
- ▶ More fusex clients
- ▶ Upgraded fusex (from 4.5.15 to 4.8.10)
- ▶ Started network migration into segregated VLAN. Use of geotag
- ▶ InvenioRDM deployment for the JRC Research Data Repository

2020 - Incidents

- ▶ Experienced a couple of MGM crashes.
 1. MGM did not start afterwards. Deadlock due to fusex clients
 2. Adding disks when FST is up caused deadlocks.
- ▶ Instance unresponsive at boot. Balancer needs to be disabled beforehand.
- ▶ Client crashes
 - ▶ Fuse with Guacamole remote desktops
 - ▶ Fusex with batch processing workloads

Coming soon

- ▶ MGM/FST software upgrade
- ▶ Start using 60 disks JBODs
- ▶ HTTP-based access to EOS data
- ▶ No more fuse?
- ▶ Bye Bye CentOS?

Keep in touch



EU Science Hub: ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub – Joint Research Centre



EU Science, Research and Innovation



Eu Science Hub

Thank you



©European Union 2021

Unless otherwise noted the reuse of this presentation is authorised under the CC BY 4.0 license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.