CERN site report
HEPiX Autumn 2022

31 October - 03 November, Umeå, Sweden

Jarek Polok <jaroslaw.polok@cern.ch>
On behalf of CERN IT

CERN IT department reorganised

The IT Department's vision is to reinforce its position as a trusted and efficient IT service-provider and technology partner for the CERN community, and to be recognised as a reference catalyst driving collaboration and innovation in the scientific computing environment in the member states and beyond.
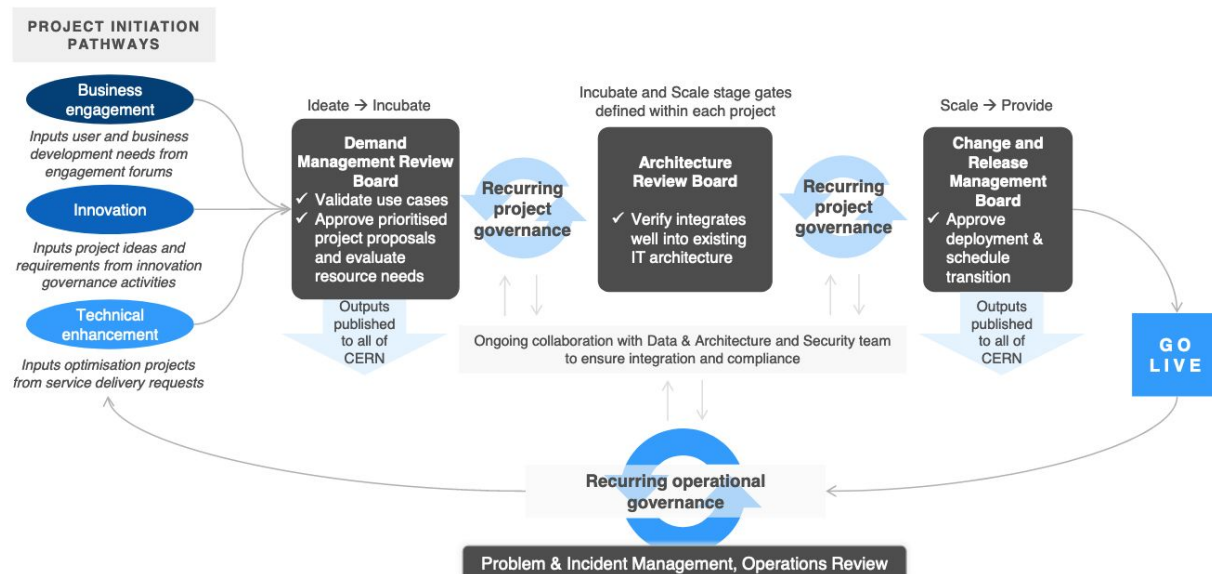
Structure, **Vision**, **Strategy**, **Projects**

CERN Data Centre(s)

Prévessin Data Centre (PCC):

- **Construction started in March of this year**
- **Internal works have started, e.g. lighting, painting**
- **Installation of ventilation ducts and piping starting now**
- **Building expected to be sealed (doors installed) by the end of the year**
- **Commissioning due to start in the summer of 2023**
- **End of commissioning foreseen for October 2023**

Hardware deliveries:

- **Quite challenging in 2022 due to overall supply chain situation**
- **Average dispatch time between order and delivery between 16 and 25 weeks**
- **Some supplies still not delivered after more than 12 months**

Other:

- **Preparing for (un)scheduled power cuts this winter**
- **OpenDCIM extensively used for planning and de(commissioning)**

October 2022: Major concrete works are now finished

talk this week: *CERN IT Hardware Procurement activities in 2020-2022*

talk this week: *Managing CERN data centers with OpenDCIM*

Cloud Infrastructure Migration Campaign

What: Move from deprecated to supported network control plane in OpenStack
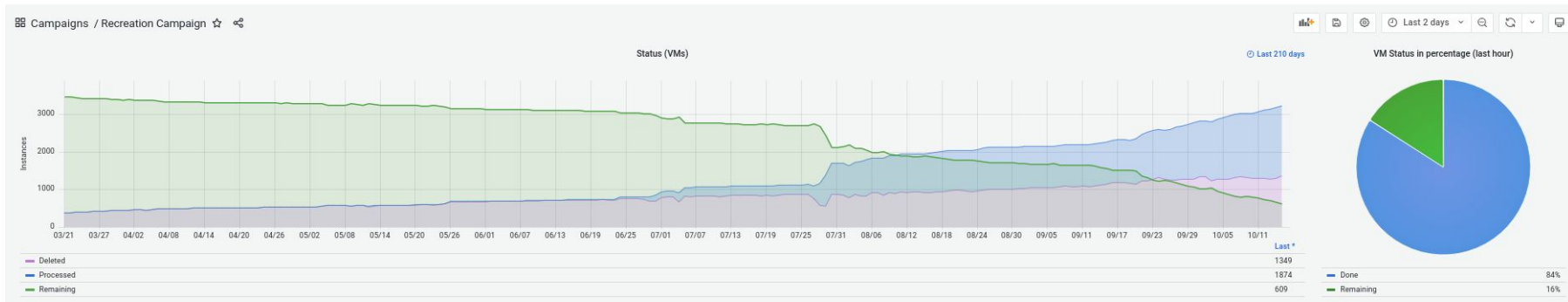
Why: Unblock upgrades and enable evolution of network features

When: Since November 2021, just finished!

4'400 VMs moved!

talk this week: *Cloud Infrastructure Update: Operations, Campaigns and Evolution*

CentOS Stream Status

Re-evaluation of initial recommendation to use CentOS Stream for production:

- **Experiences after ~1 year of running services on CentOS Stream**
- **State of the Red Hat Enterprise Linux rebuilds**

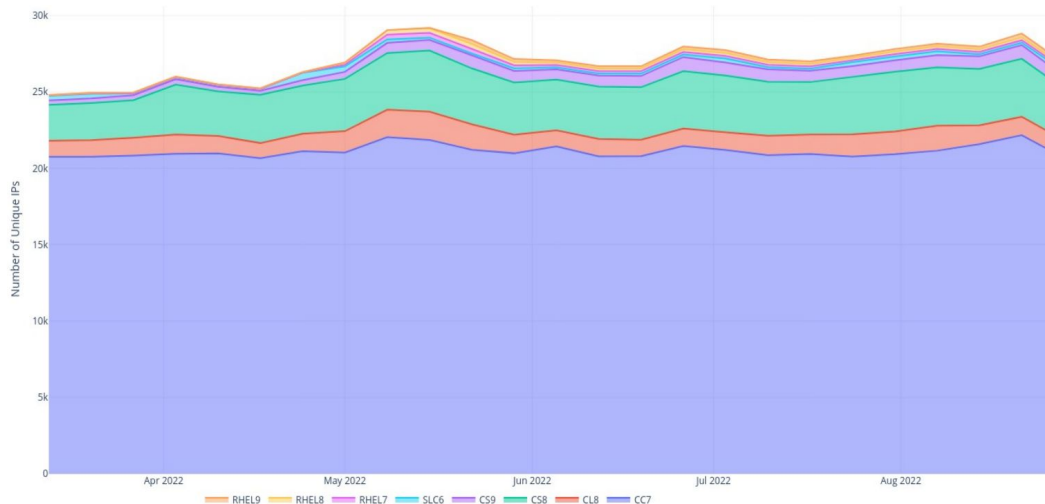**Continue to work with the HEP community and encourage feedback from the community!**

talk this week:

CERN Linux Updates and Outlook

&

BoF session:

Recent experiences with CentOS Stream

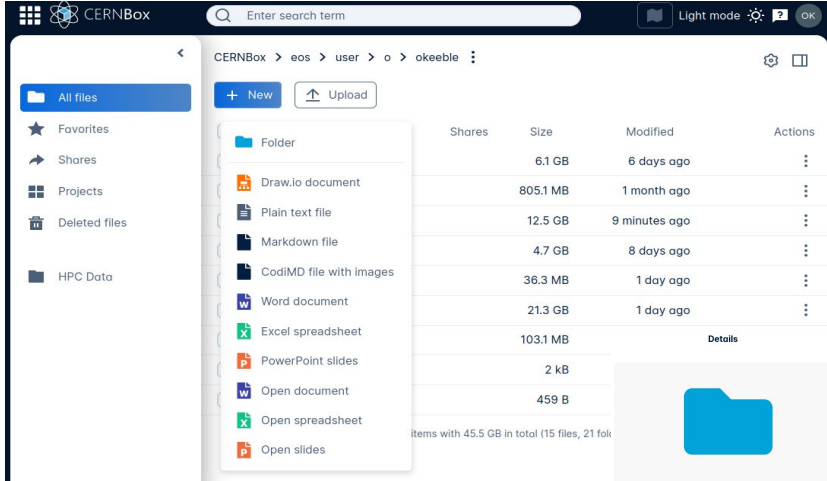|		
SLC6	329	1.2%
CC7	20,680	76.9%
CL8	1,155	4.3%
CS8	3,373	12.5%
CS9	834	3.1%
RHEL7	143	0.5%
RHEL8	147	0.5%
RHEL9	220	0.8%

Storage services

CERNBox:

- [**New web interface**](#) with major improvements
- Integration with MS Office 365 and Markdown (CodiMD) for collaborative editing
- New backup orchestrator deployed (based on RESTIC) for EOS home directories and projects, writing to S3, now extended to also back up CephFS shares

    talk this week: *CERNBox: sync, share and science*

Ceph:

- Major production clusters upgraded to [**Ceph Pacific**](#)
- Capacity expansion of 2nd region Object Storage cluster to 27 PB
- Consolidated storage offer and introduction of Storage Availability zones for block storage
- Ongoing investigation of Ceph replication features for Business Continuity / Disaster Recovery plans

Storage services

EOS

DFS:

- **Review of use-cases, migration of user areas to EOS+SAMBA complete: almost 46000 accounts**

CERN Tape Archive

CTA:

- **Nominal performance/efficiency as designed**
  - **Archival at 10 GB/s per LHC experiment consistently reached**
- **CASTOR has been decommissioned**
  - **official EOL 31/10/2022**

CASTOR
CERN Advanced STORage manager

talk this week: *Storage for the LHC: operations during Run 3*

New CERN Record
Data written to CTA from 1/08/2022 to 31/08/2022

20.95 Petabytes in August 2022

15.9 Petabytes in November 2018

7.3 Petabytes in October 2015

4.61 Petabytes in November 2012

Database and Analytics

Start of Next Oracle LTS (long term support) version beta test - Oracle 23c

- **23c will be the version supported for the HL-LHC run. Gradually we want to test the upgrade impact on all critical databases in order to prepare for HL-LHC run**

More than 600 [DBoD](#) databases (MySQL, pgSQL, InfluxDB) upgraded to recent supported releases

- **Thanks to automation, large part of Database on Demand landscape was upgraded to higher versions without significant impact to the service**

Oracle instances consolidation in preparation for licensing schema change in April 2023:

- **As the end of campus license contract nears Database service is consolidating Oracle instances on powerful high frequency/low core count systems to limit the financial impact of Oracle licensing. Consolidation includes (but is not limited to):**
- **Review and optimization of existing DataGuard and Active DataGuard configurations**
- **Utilization of Oracle RAC One Node as replacement of full Oracle RAC**
- **Decommissioning of legacy infrastructure (e.g. GoldenGate replication)**

Database and Analytics

ElasticSearch service:

- **Migrating all production clusters to OpenSearch (running v2.2.1)**

- **New OpenSearch clusters running on physical machines with SSDs and Ceph storage**

- **Plan is to migrate all legacy ElasticSearch 6/7 clusters by Q1 2023**

Hadoop service:

- **Testing a dedicated HBase deployment for NXCALS to improve performance/robustness**

- **Evaluated Apache Ozone as a block storage alternative to HDFS**

SWAN service:

- **The interactive data analysis service is gradually being migrated to a Kubernetes deployment. The Kubernetes installation is part of the production service and automatically serving 60% of users regularly. Some GPUs are available to users thanks to this new deployment**
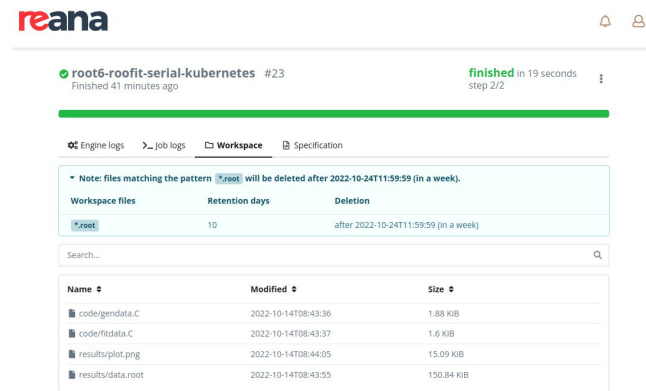
Applications and Reusable Workflows

REANA - Reusable Analyses:

- **New user-configurable workspace file retention rules**
- **Example: retain large temporary "*.root" files for ten days only**
- **New support for Rucio data access**
- **Exposure of workflow engine logs and job duration in the web interface**
- **REANA 0.9.0 official release scheduled for November 2022**

JEEDY:

- **New cronjob infrastructure based on Argo Workflow**
- **Wide adoption of GitOps and ArgoCD for internal infrastructure components**
- **efiles.cern.ch managed fully via GitOps**
- **More apps and standard approach are coming**
- **Enabling secret management via teigi**

Identity and Access Management

- **2FA rollout in progress**
- **SSO migration**
- **Old SSO to be decommissioned summer 2023 (software end of life)**
- **8,500 applications on new SSO**
- **Some service incidents now that high density traffic events protected by new SSO:**

  architecture being reviewed urgently

- **Contact and collaboration with other high traffic Keycloak users would be very welcome!**

talk this week: *Enforcing Two-Factor Authentication at CERN: A Technical Report on Our Experiences with User Migration*
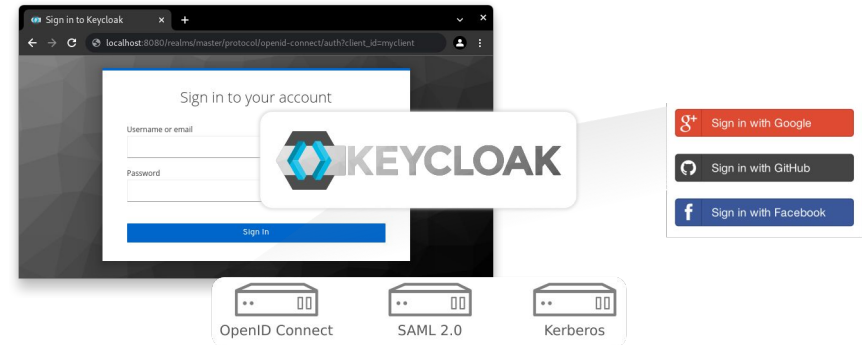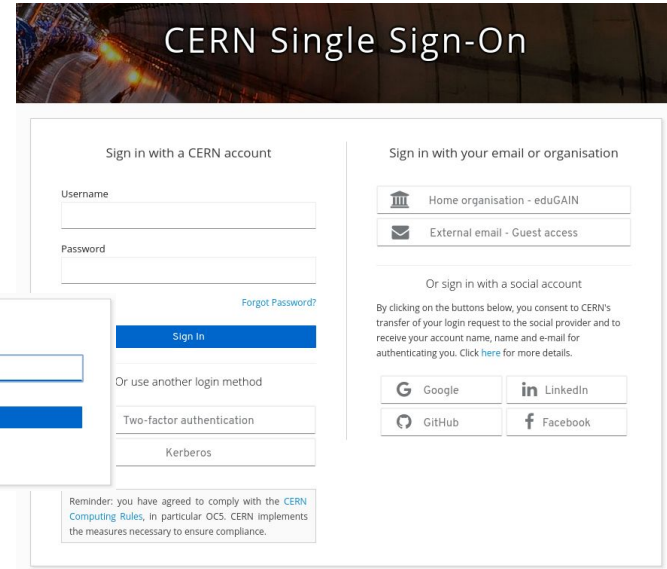
Platforms Infrastructure

PaaS: migration from Openshift3 to OKD4 complete

- **Openshift3 EOL June 2022**
- **Used the Openshift Migration Toolkit for Containers**
- **Documentation sites now served by GitLab Pages: MkDocs, and other static page generators are supported**

Kubernetes / Magnum:

- **Kubernetes 1.24 and 1.25**
- **Support for dual stack networking (ipv4/ipv6)**
- **New CVMFS CSI driver with automount support**
- **Extended support for GPUs, including Multi-Instance-GPU (MIG) for A100**
- **Support for lazy image pulling (estargz) in both builds and deployments**

talk this week: _Kubernetes operators for web hosting at CERN: experience_

GitLab:

- **Infrastructure migration from OpenShift 3 to Kubernetes complete**
- **From unsupported custom-made infrastructure to supported Cloud Native Hybrid one**

Collaborative Services

Indico: Version 3.2 (June at CERN, public release in August)

- **New batch of privacy-enhancing features**
- **Updated minutes editor (+ markdown support)**
- **Event series management interface**
- **CAPTCHAS in reg. forms and user registration**

Transcription / Translation as a Service:

- **Ongoing project for a new ASR system based on technology from MLLP (Universitat Politècnica de València)**
- **Better performance than commercial alternatives, with CERN's vocabulary and accents**
- **Integration with Zoom, webcast/recording services (both live and offline)**
- **Web APIs and interface for ad-hoc usage**
- **Pilot being set up, available before end of year**

coming soon!

Digital Repositories

Zenodo:

- **REANA badge to launch analysis directly from Zenodo integrated**
- **US National Institute of Health have funded Zenodo to enhance access to biomedical research**
- **New counter measures have significantly reduced spam**

Digital Memory:

- **Consolidating Long-term Preservation platform into a production service**
- **Supports lots of different sources of content:**
  - **repositories (CDS, Zenodo), Indico, Gitlab, CERNBox, CodiMD, etc**
- **Integrated with Archivematica to support the OAIS model**
  - **e.g. CASTOR software archived in the platform**

InvenioRDM:

- **New common platform for Zenodo and CERN Document Server**
- **New features: Community members support, review workflow, custom fields and administration interface**
- **Larger production sites launched: TU Graz, Caltech Data**

Email and Engineering Software Services

Email Service: Following [MALT closure](), move to [Exchange Online]():

- **Upgrade of the current Email platform**
- **Same functionality, same supported clients, renewed Outlook for the web**
- **Ongoing migration**
- **Migrating emails, calendar items, contacts and tasks**
- **From August 2022 to mid-2023**

Engineering Software Service:

- **New CAD software made available for 2D design: [BricsCAD Pro]()**
- **Alternative to AutoCAD**
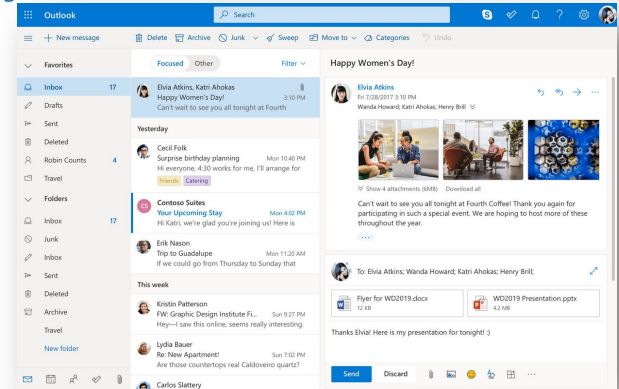
CERN presentations this HEPiX

- ***Kubernetes operators for web hosting at CERN: experience*** - Alexandre Lossent
- ***CERN Linux Updates and Outlook*** - Alex Iribarren
- ***Scalable Machine Learning at CERN with Kubeflow*** - Dejan Golubovic
- ***Automation of (Remediation) procedures for Batch Services in CERN with StackStorm*** - Ankur Singh
- ***Cloud Infrastructure Update: Operations, Campaigns, and Evolution*** - Domingo Rivera Barros
- ***Transferring data from ALICE at 2.4Tbps with DWDM*** - Daniele Pomponi
- ***Enforcing Two-Factor Authentication at CERN: A Technical Report on Our Experiences with User Migration*** - Adeel Ahmad
- ***Computer Security Landscape Update*** - Christos Arvanitis
- ***CERN IT Hardware Procurement activities in 2020-2022*** - Luca Atzori
- ***Managing CERN data centers with OpenDCIM*** - Jarek Polok
- ***CERNBox: sync, share and science*** - Diogo Castro
- ***Storage for the LHC: operations during Run 3*** - Cedric Caffy

home.cern