CERN Site Report

Andrei Dumitru
CERN IT Department
15th of May 2018

CERN

European Organisation for Nuclear Research
(**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire)

Founded in 1954, 22 member states today

World's largest particle physics laboratory

Located at Franco-Swiss border near Geneva
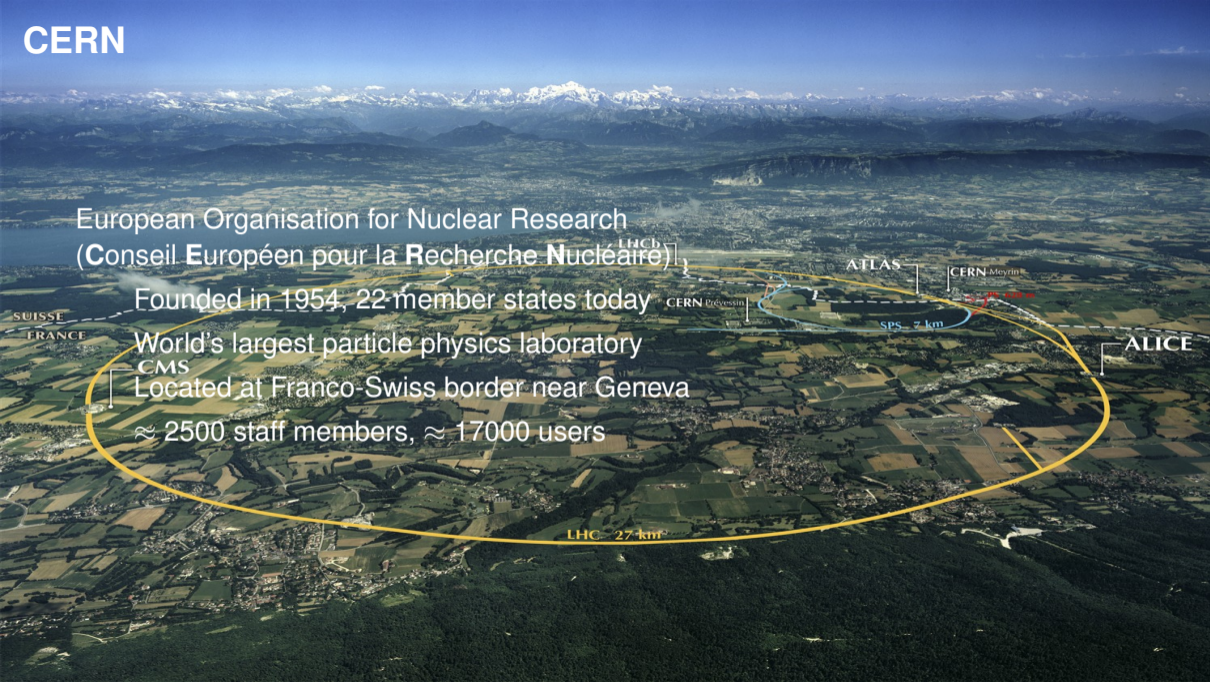
$\approx$ 2500 staff members, $\approx$ 17000 users

CERN IT Department

Enabling the laboratory to fulfill its mission

- Main data centre in Meyrin site
- Wigner data centre in Budapest, 23 ms away
- Connected via 3 dedicated 100-Gb/s links
- Service resources in both sites where possible (+ disaster recovery)

COMPUTING

Servers (Meyrin)	Cores (Meyrin)
11.5 K	**174.3 K**

Servers (Wigner)	Cores (Wigner)
3.5 K	**56.0 K**

STORAGE

Disks (Meyrin)	Tape Drives
61.9 K	**97**

Disks (Wigner)	Tape Cartridges
29.7 K	**32.2 K**

NETWORK

Routers	Star Points
246	**688**

Switches	Wifi Points
4.0 K	**1.5 K**

Batch Jobs

100 K
80 K
60 K
40 K
20 K
0
16:00 20:00 00:00 04:00 08:00 12:00

■ atlas ■ cms ■ lhcb ■ alice ■ other
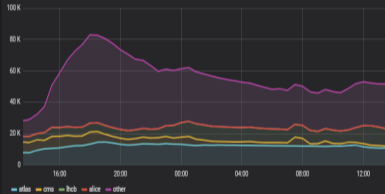
EOS Active Data Transfers

100 K
80 K
60 K
40 K
20 K
0
16:00 20:00 00:00 04:00 08:00 12:00

■ alice ■ atlas ■ cms ■ lhcb ■ public
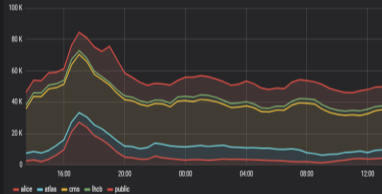
File Transfer Throughput

18 Gbps
15 Gbps
13 Gbps
10 Gbps
8 Gbps
5 Gbps
3 Gbps
16:00 20:00 00:00 04:00 08:00 12:00

■ atlas ■ cms ■ lhcb

Cloud Virtual Machines

250
200
150
100
50
0
16:00 20:00 00:00 04:00 08:00 12:00

■ vms created per half hour
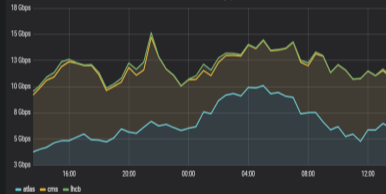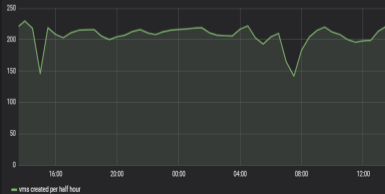
Databases Transactions

7.0 K
6.0 K
5.0 K
4.0 K
3.0 K
2.0 K
1.0 K
0
16:00 20:00 00:00 04:00 08:00 12:00

■ transactions per second

Wigner Network Links

300 Gbps
250 Gbps
200 Gbps
150 Gbps
100 Gbps
50 Gbps
0 bps
16:00 20:00 00:00 04:00 08:00 12:00

■ to cern ■ to wigner
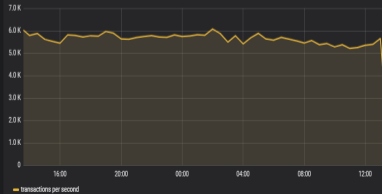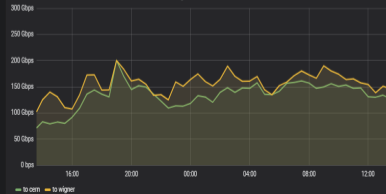
Linux

CentOS 7.5 release pending (tentative date: end of May)

- ▶ Cephfs kernel module (technical preview)
- ▶ Improved nvidia GPU support for kvm
- ▶ DFS access using SMB2/3.
- ▶ Rebase of GNOME desktop to 3.26

Kubernetes packages available via CentOS Virt SIG repositories

Migrating completely to CC7 'soon'

January 2019 (LS2):

- ▶ Start migrating 30% of lxbatch to CC7
- ▶ lxplus.cern.ch alias still pointing at SLC6 nodes
- ▶ lxplus7.cern.ch points to CC7 nodes
- ▶ lxplus6.cern.ch points to SLC6 nodes

June 2019 (LS2 + 6 months):

- ▶ Migrate 80% of the lxbatch resources to CC7
- ▶ Change lxplus.cern.ch alias to point to CC7
- ▶ Remaining SLC6 nodes in lxplus6.cern.ch

June 2020 (LS2 + 18 months):

- ▶ Migrate remaining lxbatch SLC6 resources to CC7
- ▶ Stop lxplus6.cern.ch

Before December 2020 (tbd):

- ▶ CERN end of support for SLC6

Centralising Elasticsearch

Production level service
ACL implementation based on
- ReadonlyREST (basic version)
- Kibana Ownhome

Resource consolidation status:
- O(20) clusters with O(70) use cases
- Up to 23 use cases on a single cluster
- Ongoing hardware upgrade

Supported ES version
- Migrating from 5.5.2 to 5.6.4 (version 6.2.4 in preparation)

More information: https://cern.ch/esdocs

ReadonlyREST

Batch

Bulk of batch capacity now in HTCondor

- ▶ 156k slots in HTCondor public share vs 10k left in LSF
- ▶ Will retire public LSF end of Run 2
- ▶ CMS Tier-0 processing now running on HTCondor rather than VMs managed by the CMS factory

HTCondor is helping us run on heterogenous resources e.g. disk-servers and remote clouds

- ▶ Successful test with running multi-site HTCondor on federated Kubernetes
- ▶ Looking at GPU deployment to support experiments in this area

Volunteer LHC@Home

- ▶ more than 400k tasks for Sixtrack,Theory & LHC

Number of Hosts with recent credit
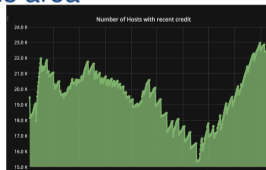
HPC at CERN

The bulk of computing at CERN is HTC (190k cores) compared with 5k for HPC
HPC clusters reserved for use cases that do not fit the standard batch High
Throughput Computing (HTC) model. (e.g. MPI applications)

- Lattice/QCD simulations (Theory)
- Accelerator physics and plasma simulations
- Computational Fluid Dynamics (CFD), field calculations and other engineering applications

Batch HPC facility using SLURM "HPC-batch"

- Cluster with ~100 16 core batch nodes with low-latency 10Gb Ethernet interconnects
- New HPC cluster resources: 2x72 nodes with Infiniband, "HPC-BE" since end of last year
- CERN CentOS 7.4 OS, CephFS used for shared storage
- Backfill of idle resources with HTCondor grid jobs in progress
- Additonally dedicated 72 nodes Infiniband cluster for Theory (Lattice/QCD)

Streaming service

Pilot service for streaming based on Apache Kafka.

- ▶ Already used by the CERN security team for their intrusion detection system

Development

- ▶ Web interface for self-service operations and provisioning of shared Kafka clusters
- ▶ Work on using Kafka for Internet of Things use cases in collaboration with network experts (working on LoRa network)

Hadoop / Spark / SWAN

Spark and Hadoop clusters for data analysis

- ▶ Consolidate and in prod for storing and processing data for monitoring, security, beams

Spark for physics data processing

- ▶ Use of Apache Spark to read and process physics data in ROOT format (with spark-root)
- ▶ Can read data using XRootD protocol (notably allows to read root data directly from EOS)
- ▶ Open sourced: https://github.com/cerndb/hadoop-xrootd

Integrated Spark with web notebooks service: SWAN_Spark

- ▶ SWAN for Web Analysis with notebooks as front-end + Spark and ROOT software
- ▶ Run data at scale using Spark

Hadoop / Spark / SWAN

We have deployed a new Hadoop cluster for Beams accelerator logging system.

- ▶ Use of Apache Hadoop Vanilla+CERN patches of interest to be more agile (patches, new version, etc.)
- ▶ Currently using free version of vendor distribution. Interest by HEP sites for collaboration?

Development

- ▶ Integration of Spark with Kubernetes on OpenStack. Interest by HEP sites for collaboration?

Hadoop / Spark / SWAN

This Thursday

Evolution of the Hadoop and Spark platform for HEP
by Zbigniew Baranowski

Application server

Java web applications

- ▶ Standardized on Kubernetes as foundation for future web deployments
- ▶ Building a pilot based on OpenStack/Magnum

Investigation on new Java application server

Single-Sign-On

- ▶ Integration of Keycloak service provider with CERN SSO

Oracle Application Express (APEX)

- ▶ Refresh of management tools (web interface, user credential management)
- ▶ Migration of legacy middletier (Oracle HTTP Server) to supported ORDS Java app

Oracle

Contract renewed until 2023.

Further database consolidation efforts coupled with hardware changes
- SSD/Flash based disk systems

Improving HA setups and working on better isolation for Oracle services based in the CERN Technical Network.

Evaluation of Cloud services for DB site recovery.

Big upgrades during LS2 to version 18c or 19c.
- Increase security across all the different DB service players

Database on Demand

Evolution of the amount of MySQL, PostgreSQL, and InfluxDB instances in the DBOD service
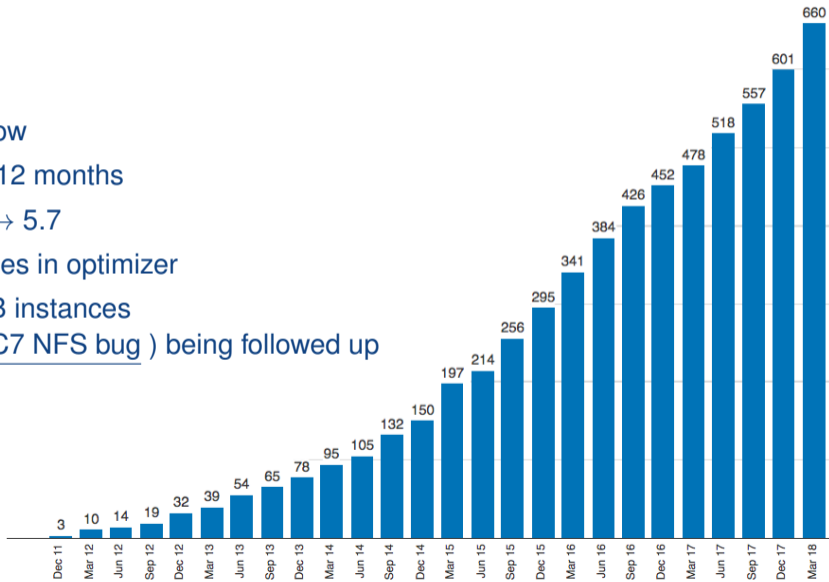
Over 650 instances now

▶ +41% in the last 12 months

MySQL upgrade 5.6 → 5.7

▶ Significant changes in optimizer

More than 70 InfluxDB instances

Linux kernel issue (CC7 NFS bug) being followed up

Bar chart values by date:
Dec 11: 3, Mar 12: 10, Jun 12: 14, Sep 12: 19, Dec 12: 32, Mar 13: 39, Jun 13: 54, Sep 13: 65, Dec 13: 78, Mar 14: 95, Jun 14: 105, Sep 14: 132, Dec 14: 150, Mar 15: 197, Jun 15: 214, Sep 15: 256, Dec 15: 295, Mar 16: 341, Jun 16: 384, Sep 16: 426, Dec 16: 452, Mar 17: 478, Jun 17: 518, Sep 17: 557, Dec 17: 601, Mar 18: 660

Storage

Data recorded by CASTOR in 2017

- ► ~46PB of data (55M files), peak of over 12PB in October (mostly LHC RAW data)
- ► Currently over 250PB archived on tape

EOS installations

- ► Exceed 250PB of raw storage, over 2 billion files (+3 billion in the testing instance), more than 100% increase
- ► Sustained rates (across 12 months): above 30 GB/s reading, 7GB/s writing
- ► 400% increase in number of files in EOSUSER (the back-end of CERNBOX)

Preparation of 2018 data taking

- ► All servers running CC7
- ► Initial tests of BEER performed (see presentation on Wednesday)
- ► Working on upgrading tape infrastructure for end of Run-2 and preparing for Run-3

Storage

CVMFS

- ▶ 300% growth in 2017 and 10 new repositories (conditions databases, nightly builds, small experiments)
- ▶ Work on further reducing the publication time as required by the experiments (enable full publication of their nightly builds)

CephFS services is gaining usage as

- ▶ Replacement/evolution of the FILER service
- ▶ to serve the new HPC infrastructure for BE and TH departments

CERNBox

- ▶ Possibility to view and edit Microsoft files in the browser
- ▶ Integrated in the SWAN system, allowing data analysis from the browser

Storage

About to start an evaluation of OnlyOffice

- ▶ Initially available only for IT via CERNBox and as a desktop application
- ▶ To be exposed to a larger CERN community during the summer

The foundation of the AFS replacement is in place:

- ▶ Goal: upgrade EOSUSER. No AFS replacement before LS2.
- ▶ Development: focus on improving the stability of the fuse and the new catalogue (tested above the 4B entries)
- ▶ Deployment: testing the deployment model (migration of some users from the old to the new infrastructure)

S3 service since end 2017 to provide disk storage for IT services (ElasticSearch disk backups)

Storage: Disk, Tapes and more

This Wednesday

WLCG Archival Storage group report
   by Vladimir Bahyl

Next generation of large-scale storage services at CERN
   by Jakub Moscicki

This Thursday

BoF session: Tape storage
by Vladimir Bahyl

IT Facilities

Changed the monitoring of IPMI to remote out of band

- ► Before - local sensor running on each node
- ► Now - remote out-of-band IPMI queries to the physical hosts
- ► Enables IPMI monitoring for servers running Windows
- ► Entirely based on standard tools, running on small Kubernetes clusters
- ► Same architecture being re-used for PDU monitoring

ServiceNow license extended until the end of 2022 and upgraded to release Jakarta

Completed a tender for for intelligent PDUs and two large tenders for CPU and storage.

Fixed-telephony evolution - TONE

Telephony Open-source Network Evolution (TONE)

Service in production

- ▶ Kamailio for the front-end & Asterisk for the back-end

External connectivity:

- ▶ All operators moved from PSTN to SIP
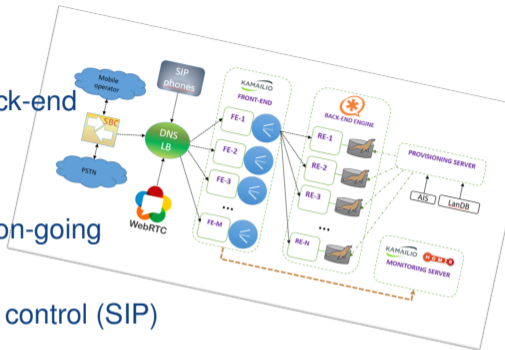- ▶ Direct SIP interconnection with HEP institutes on-going

Special services migration:

- ▶ Lifts (3G), red phones (analogue2SIP), access control (SIP)

Call centers:

- ▶ Automatic Switchboard to handle outgoing calls in production
- ▶ Call centers will be migrated in 2019

End users (2019):

- ▶ IP phones → SIP
- ▶ Analogue office phones → WebRTC softphones

IoT - LoRaWAN

LoRaWAN network deployment on-going

- ▶ Ideal for large amount of battery powered sensors/meters
- ▶ The Things Network installed on premises (private network)
- ▶ Migration to TTN v3 during summer (open-source MIT license)
- ▶ Several cheap LoRa gateways available on the market
  - ▶ LORIX One chosen for deployment at CERN

Pilot after summer

Networking

IPv6 peerings have been configured with KREOnet (Korean NREN)
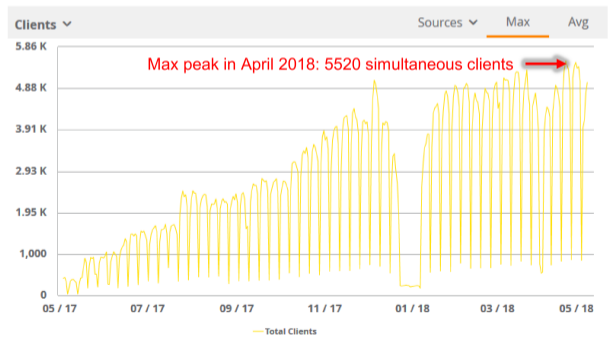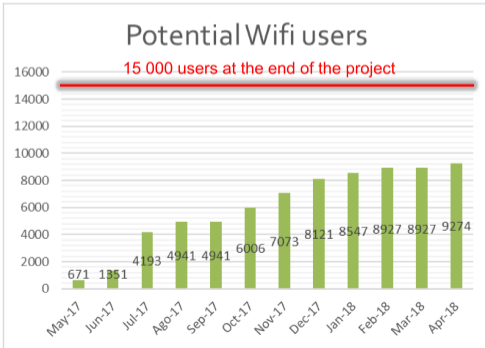
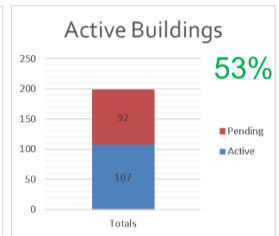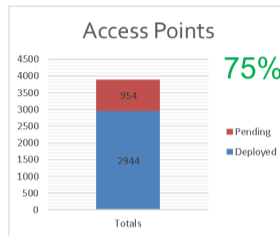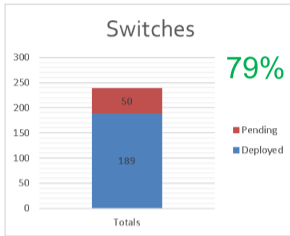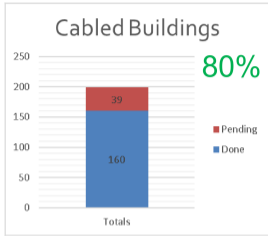- On LHCOPN, LHCONE and the external network

Obtained the last /22 IPv4 prefix (1024 addresses) from RIPE

Planning switch/router upgrade for the campus network

- growing demand for non-user devices-with requirements for services such as PoE and VLANs which aren't needed for user devices.

We expect more and more user devices (laptops/tablets/phones, but also desktops) to use the new Wi-Fi service.

Wi-Fi project

Cabled Buildings
80%

Pending: 39
Done: 160
Totals

Switches
79%

Pending: 50
Deployed: 189
Totals

Access Points
75%

Pending: 954
Deployed: 2944
Totals

Active Buildings
53%

Pending: 92
Active: 107
Totals

Potential Wifi users

15 000 users at the end of the project

Month	Users
May-17	671
Jun-17	1351
Jul-17	4193
Ago-17	4941
Sep-17	4941
Oct-17	6006
Nov-17	7073
Dec-17	8121
Jan-18	8547
Feb-18	8927
Mar-18	8927
Apr-18	9274

Clients ⌄   Sources ⌄   Max   Avg

Max peak in April 2018: 5520 simultaneous clients

5.86 K
4.88 K
3.91 K
2.93 K
1.95 K
1,000
0

05 / 17   07 / 17   09 / 17   11 / 17   01 / 18   03 / 18   05 / 18

Total Clients

Wi-Fi project

Wi-Fi project

Digital Repositories

CDS Videos released in production
- First step in the migration of CDS to a new technology stack based on Invenio 3

REANA (Reuse Analysis) v.0.1.0 released.
- Working on providing infrastructure to enable reusability of physics analysis using complex computational workflows

CERN Open Data
- New site based on Invenio 3 released
- + 1PB of CMS Run 2012 data

Future:
- CDS Books: new circulation, acquisition and ILL service for the CERN library - will also be available to other institutes through Invenio Free Open Source Software
- REANA: Finalise the integration with CWL

Techlab

cern.ch/techlab

CERN Techlab is an IT department project aiming at improving the efficiency of the computing architecture and making better utilisation of the processors available today.

New Nvidia GeForce / Titan EULA

- *No Datacenter Deployment. The SOFTWARE is not licensed for datacenter deployment, except that blockchain processing in a datacenter is permitted.*

May affect large scale deployments.

Workarounds being investigated:

- Look at other GPU vendors
- Explore "academic offers" for datacenter deployment of GPUs
- Use open source drivers

This Thursday

Techlab benchmarking web portal
    by Maxime Reis

MAIt

A strategic move towards CERN's philosophy of Open Software for Core Services

A recent change in Microsoft Licensing Terms with CERN led IT to rethink its principles of engagement:

- ▶ Deliver open source Core Services
- ▶ Avoid vendor lock-in
- ▶ Solutions inclusive to all CERN users

More details at 28th May ITUM: https://indico.cern.ch/event/709461/

Talk on Monday

Computer Security Update by Stefan Lueders

Privacy at CERN

Office of Data Privacy Protection

Align to best practices and recommendations

Operational Circular on the Protection of Personal Data (in review)

Privacy Notices of IT services (in review)

Talks from CERN this week

TRIDENT Tool for collecting and understanding performance hardware counters

- ▶ by Servesh Muralidharan

Computer Security Update

- ▶ by Stefan Lueders

CERN Site Report

- ▶ by Andrei Dumitru

WLCG Archival Storage group report

- ▶ by Vladimir Bahyl

Next generation of large-scale storage services at CERN

- ▶ by Jakub Moscicki

Talks from CERN this week

Batch on EOS Extra Resources moving towards production

- ▶ by Markus Schulz

Techlab benchmarking web portal

- ▶ by Maxime Reis

ExDeMon: a new scalable monitoring tool for the growing CERN infrastructure

- ▶ by Daniel Lanza Garcia

Evolution of technology and markets

- ▶ by Bernd Panzer-Steindel and Helge Meinhard

Proposal for a technology watch WG

- ▶ by Helge Meinhard

Talks from CERN this week

Monitoring Infrastructure for the CERN Data Centre

- ▶ by Asier Aguado Corman

Evolution of the Hadoop and Spark platform for HEP

- ▶ by Zbigniew Baranowski

BoF session: Tape storage

- ▶ by Vladimir Bahyl

Status update of the CERN private cloud

- ▶ by Spyridon Trigazis

Baremetal provisioning in the CERN cloud

- ▶ by Spyridon Trigazis

home.cern