



An artistic view of one of the CERN Data Centre's tape libraries used for long-term storage. (Photo Jeff Frost)

Computing

Extensive upgrade and migration

To prepare for Run 2, the LHC experiment teams and the Worldwide LHC Computing Grid (WLCG) collaboration have upgraded the computing infrastructure and services. During Long Shutdown 1 (LS1), the IT department doubled the capacity available to the LHC experiments, with the addition of some 100 petabytes (PB) of disk storage and almost 60 000 new cores. The compute capacity of the CERN private cloud has nearly doubled during the last year, now providing over 150 000 computing cores. LS1 offered an ideal opportunity to migrate the archived data from legacy cartridges and formats to higher density ones. This involved migrating around 85 PB of data, and was carried out in two phases during 2014 and 2015. As an overall result of this two-year migration, no less than 30 000 tape cartridge slots were released to store more data.

New run, new records

With the start of Run 2, new data-taking records were achieved. 40 PB of data were written on tape at CERN in 2015. Out of the

30 PB from the LHC experiments, a record-breaking 7.3 PB were collected in October, and up to 0.5 PB of data were written to tape each day during the heavy-ion run. By way of comparison, CERN's tape-based archive system collected in the region of 70 PB of data in total during the first run of the LHC. The WLCG also set a new record in 2015 by running a total of 51.1 million jobs in October.

Data preservation

CERN manages the largest archive of scientific data in the high-energy physics (HEP) domain. By the end of 2015, CERN's tape-based archive system held 132 PB of data, corresponding to 410 million files spread across approximately 20 000 tape cartridges, distributed over seven tape libraries providing a total of 86 production tape drives. Most of these data, from past and present HEP experiments, are to be preserved indefinitely.

Bits in tape media are smaller than bacteria, so any damage to the tape can destroy significant amounts of data. Consequently,



The participants of the 39th CERN School of Computing in September 2015. (OPEN-PHO-CCC-2016-001-1)

to ensure long-term data preservation, it is important to protect the physical environment against contamination hazards. In 2015, CERN deployed an improved airflow-monitoring system based on custom-made environmental sensors. The sensors were prototyped and built in-house and are hosted in an empty tape drive tray inside the tape libraries. They are comparable to commercial systems in terms of precision and reaction time, but at about a fiftieth of the cost and requiring no maintenance. The hardware design, as well as all software and firmware components, are freely available via open-source licenses.

In June 2015, the first workshop and Collaboration Board meeting of the Data Preservation and Long-Term Analysis in High-Energy Physics (DPHEP) collaboration took place at CERN, aiming to address the general issues of data preservation in HEP. A significant milestone was achieved with the publication of a Status Report detailing the progress made since the publication of the DPHEP Blueprint in May 2012.

Open source for open science

The impact of CERN's broader open-source software efforts has been growing during 2015. Zenodo – a free open-data repository that benefited from European Commission co-funding for its development and that allows researchers to share articles, data and software – has grown considerably in 2015, especially as the repository of choice for publishing science software via a link with the popular software sharing site Github. More than 3000 different science packages have been released as a result.

The open-source Invenio software, initially developed at CERN, has been refactored in 2015 to enhance the reusability of individual modules, resulting in diverse software projects around the world building on these new features. The CERN spin-off company TIND has also seen strong interest and steady growth in the number of customers wishing to have Invenio installations built for them, including CALTECH and the United Nations.

The number of sites running instances of the Indico conferencing package surpassed 200 in 2015, reflecting CERN's commitment to making collaborating and conferencing more efficient. CERN also released in 2015 a dashboard for its videoconferencing infrastructure based on Vidyo, which resulted in universities, hospitals and companies installing CERN's open-source software.

Through a number of collaborations, CERN's disk-storage system for LHC computing, known as EOS, has also been made available to companies and user communities to use with their "big data" systems, or to use its data-distribution capabilities to build distributed data repositories.

Contributions to open-source projects

In addition to leading and developing its own open-source software, CERN has contributed to external projects in order to adapt them to the Organization's needs. One of the larger contributions is to ownCloud, which is used to provide the CERNBox service – a secure and scalable equivalent to Dropbox for CERN users. Other projects receiving contributions include OpenStack, which is key in enabling the deployment of cloud services in the Data Centre on the Meyrin site and remotely at the Wigner Data Centre in Budapest, and Ceph, the most popular network-block-storage backend for OpenStack.

CERN has contributed over 90 improvements to the latest OpenStack release and to significant new features in the Ceph open-source project. The initial objective with Ceph was to build a block-storage service for the CERN OpenStack cloud, but it has expanded to include R&D towards Ceph-based solutions to solve future LHC data-storage challenges. In 2015, the development team performed a number of scale tests in close collaboration with the key designers of Ceph, exploring the scalability limits of this system, which could potentially benefit other CERN partners.

Open data

In November 2014, CERN launched its Open Data Portal that allows the LHC experiments to share their data with a dual focus: firstly for the scientific community, including researchers outside of the CERN experimental teams as well as citizen scientists, and secondly for the purposes of training and education through specially curated resources. All data is in the public domain under the Creative Commons “zero” license, a first in CERN’s long history, and can be cited in scientific discourse using a unique digital object identifier.

In 2015, the CERN Open Data Portal has been ramped up with further data and code releases. This last year has seen increased use of these resources, and, in the spirit of open science, unexpected applications of the data, such as training in big-data analytics and data mining. CERN’s Open Data team has transferred experience and inspired many teams around the world by participating in conferences and online forums.

Science in the cloud

CERN is working towards building a European Open Science Cloud using the experience gained through various major projects together with the Helix Nebula initiative – a public-private partnership. The work of Helix Nebula instigated CERN to lead a Horizon 2020 project to create a procurement network of public research organisations, named PICSE (Procurement Innovation for Cloud Services in Europe), interested in making use of commercial cloud services to support their research programmes. In 2015, PICSE investigated the feasibility of a cross-border PCP (pre-commercial procurement) for shared procurement across public organisations. The European Commission decided to contribute funding to the HNSciCloud (Helix Nebula – The Science Cloud) PCP project led by CERN, which started in January 2016.

Education and sharing

Since the early seventies, the CERN School of Computing has been promoting advanced learning and knowledge exchange in scientific computing among young scientists and engineers involved in particle physics or other sciences. It is made up of three separate schools, each of them having a particular flavour and focus. In 2015, the Main School – which lasts two weeks and acts as a summer university, providing a series of lectures and hands-on exercises – took place at Kavala in Greece. Of the 76 participating students from 47 institutes, 68 passed the exam and received six ECTS credits.

The Meyrin and Wigner Data Centres together host around 15 000 servers, which are renewed every four to five years as they become obsolete for the purposes of CERN’s research. However, they remain suitable for less demanding applications. In March 2015, 224 servers and 30 switches were donated to the COMSATS Institute of Information Technology in Islamabad, Pakistan, to be used by scientists working on the ALICE experiment at the LHC. A few months later, in August 2015,

384 servers, 24 switches and 26 racks were donated by CERN to institutes in Mexico, which are using them for a variety of scientific and educational projects in the fields of physics, mathematics, energy and environmental sciences. To date, servers and switches have been donated to nine countries, namely Bulgaria, Egypt, Ghana, Morocco, the Philippines, Senegal, Serbia and now Mexico and Pakistan.

New phase for CERN openlab

In January 2015, CERN openlab entered its fifth three-year phase. Through this unique public-private partnership, CERN collaborates with leading ICT companies and research institutes to accelerate the development of cutting-edge solutions for the LHC community worldwide. Huawei, Intel, Oracle, and Siemens are all partner companies, Brocade, Cisco, IDT, Rackspace, and Seagate are contributors, and Comtrade and Yandex are associate members. For the first time, other public research organisations – the European Bioinformatics Institute (UK), the GSI Helmholtz Centre for Heavy-Ion Research (Germany) and Newcastle University (UK) – also joined as members. The topics selected for this new phase include next-generation data-acquisition systems, optimised hardware and software-based computing platforms for simulation and analysis, scalable and interoperable data storage and management, cloud-computing operations and procurement, and data-analytics platforms and applications.

CERN openlab held two new events in 2015: a first-of-its-kind open day in June 2015, and the CERN openlab “Innovation and Entrepreneurship” event in October, organised in collaboration with CERN’s Knowledge Transfer group and IdeaSquare. In addition, the CERN openlab Summer Student Programme continued to go from strength to strength, with forty students representing 27 nationalities coming to CERN.