CERN openlab

ANNUAL REPORT

2018

01 INTRO

A word from the DG

CERN openlab began a new three-year phase in 2018. Together with its collaborators from research and industry, CERN openlab is working to tackle tomorrow's ICT challenges.

Fabiola Gianotti, Director-General of CERN.

CERN openlab is a unique public-private partnership. It was founded by CERN in 2001 to provide a way to collaborate with leading technology companies to tackle tomorrow's ICT challenges. Together, we are able to accelerate the development of cutting-edge computing technologies, to the benefit of research in particle physics and beyond.

This collaboration is also of great value to the companies. Their membership enables them to test their latest technologies in CERN's uniquely challenging environment, receiving valuable feedback from our experts.

In 2018, Oracle became the second company, after Intel, to have spent over 15 years collaborating with CERN through CERN openlab. The continuing appeal of this collaboration is not only shown by the desire of already-engaged companies to continue working in CERN openlab for phase after phase, but also by the range of new companies signing up. In 2018, Micron Technology, IBM, E4 Computer Engineering, and Google all joined CERN openlab.

CERN openlab runs in successive three-year phases, making it possible for both CERN and the companies to take stock at regular intervals. 2018 marked the start of CERN openlab's sixth such phase, with work beginning on a series of new projects. These aim to address the key ICT challenges faced by our research community, as identified through an in-depth consultation process with experiments and teams across the laboratory (http://openlab.cern/whitepaper). Building upon the success of pilot collaborations in CERN openlab's fifth phase, advanced discussions are also now underway with a number of new laboratories and universities about joining CERN openlab.

In addition, CERN openlab is exploring a range of emerging, disruptive technologies that offer the potential to change fundamentally ICT processes at CERN and beyond. For example, in 2018 CERN openlab began investigation of technologies related to quantum computing. While these technologies are still at an early stage of development, they do hold significant potential, and CERN openlab is ideally positioned to help drive innovation in this area forward.

02 CONTEXT

Background

Founded in 1954, the CERN laboratory sits astride the Franco-Swiss border near Geneva. It was one of Europe's first joint ventures and now has 23 member states.

The laboratory

At CERN, physicists and engineers are probing the fundamental structure of the universe. They use the world's largest and most complex scientific instruments to study the basic constituents of matter — the fundamental particles. The particles are made to collide at close to the speed of light. This process gives the physicists clues about how the particles interact, and provides insights into the fundamental laws of nature.

The instruments used at CERN are purpose-built particle accelerators and detectors. Accelerators boost beams of particles to high energies before the beams are made to collide with each other or with stationary targets. Detectors observe and record the results of these collisions.

CERN is home to the Large Hadron Collider (LHC), the world's largest and most powerful particle accelerator. It consists of a 27-kilometre ring of superconducting magnets, with a number of accelerating structures to boost the energy of the particles along the way.

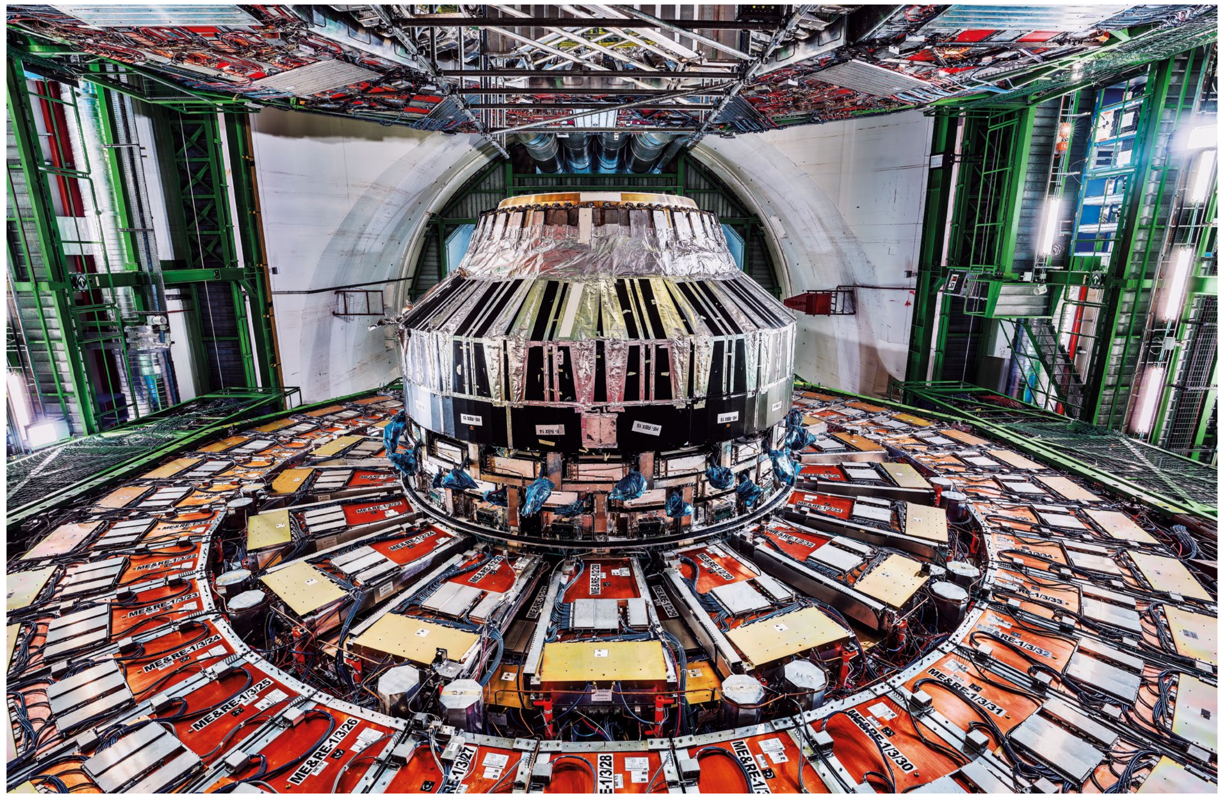
The LHC

The accelerator complex at CERN is a succession of machines that accelerate particles to increasingly high energy levels. Each machine boosts the energy of a beam of particles, before injecting the beam into the next machine in the sequence. In the LHC — the last element in this chain — particle beams are accelerated up to the record energy of 6.5 teraelectronvolts (TeV) per beam. Most of the other accelerators in the chain have their own experimental halls where beams are used for experiments at lower energies.

The proton source is a simple bottle of hydrogen gas. An electric field is used to strip hydrogen atoms of their electrons to yield protons. Linac 2, the first accelerator in the chain, accelerates the protons to the energy of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. The protons are finally transferred to the two beam pipes of the LHC. Inside the accelerator, two high-energy particle beams travel at close to the speed of light before they are made to collide. The beams travel in opposite directions in separate beam pipes: two tubes kept at an ultra-high vacuum. They are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets.

The beam in one pipe circulates clockwise while the beam in the other pipe circulates anticlockwise. It takes 4 minutes and 20 seconds to fill each LHC ring, and 20 minutes for the

The LHC is the world's largest and most powerful particle accelerator.



Seven experiments on the LHC use detectors to analyse the myriad particles produced by collisions.

protons to reach their maximum energy of 6.5 TeV. Beams circulate for many hours inside the LHC beam pipes under normal operating conditions. The two beams are brought into collision inside four detectors — ALICE, ATLAS, CMS and LHCb — where the total energy at the collision point is equal to 13 TeV. The particle beams have such a small diameter that the task of making them collide is akin to firing two needles 10 kilometres apart with such precision that they meet halfway.

Protons are not the only particles accelerated in the LHC. Lead ions for the LHC start from a source of vaporised lead and enter Linac 3 before being collected and accelerated in the Low Energy Ion Ring (LEIR). They then follow the same route to maximum energy as the protons. Colliding lead particles makes it possible to recreate conditions similar to those just after the big bang, known as "quark-gluon plasma".

The experiments

Seven experiments on the LHC use detectors to analyse the myriad particles produced by collisions. These experiments are run by collaborations of scientists from institutes all over the world. Each experiment is distinct, and characterised by its detectors. The biggest of these experiments, ATLAS and CMS, use general-purpose detectors to investigate the largest range of physics possible. Having two independently designed detectors is vital for cross-confirmation of any new discoveries made. ALICE and LHCb have detectors specialised for focusing on specific phenomena. These four detectors sit underground in huge caverns on the LHC ring.

The smallest experiments on the LHC are TOTEM, LHCf, and MoEDAL, which focus on "forward particles": protons or heavy ions that brush past each other rather than meeting head on when the beams collide. TOTEM uses detectors positioned on either side of the CMS interaction point, while LHCf is made up of two detectors which sit along the LHC beamline, at 140 metres either side of the ATLAS collision point. MoEDAL uses detectors deployed near LHCb to search for a hypothetical particle called the magnetic monopole.

It is important to note that while the main focus of research at CERN has moved in recent years towards the LHC, experiments at other accelerators and facilities both on-site and off also remain an important part of the laboratory's activities. In fact, CERN has a very diverse research programme, covering a wide range of physics topics, from the Standard Model to supersymmetry, and from dark matter to cosmic rays. Supporting all the experiments hosted at CERN is a very strong theory programme, which carries out world-leading research in theoretical particle physics.

The CERN data centre

The CERN data centre processes hundreds of petabytes of data every year. The data centre in Meyrin, Switzerland, is the heart of CERN's entire scientific, administrative, and computing infrastructure. All services — including email, scientific data management and videoconferencing — use equipment based in the data centre. The centre hosts around 10,000 dual-CPU servers with approximately 300,000 processor cores.

The servers undergo continual maintenance and upgrades to make sure that they will operate in the event of a serious incident. Critical servers are held in their own room, powered and cooled by dedicated equipment. Over 300 petabytes of data are stored on tape at the Meyrin site, with an additional 280 petabytes of capacity on disks.

The Wigner Research Centre for Physics

In June 2013, the Wigner data centre in Budapest was inaugurated. It serves as an extension to the main data centre in Meyrin. Today, the Wigner data centre hosts 3500 dual-CPU servers, with approximately 100,000 processor cores. It also offers around 100 PB of disk space.

The equipment in Wigner is managed and operated from CERN, in the same way as the equipment in the CERN data centre. Only activities requiring physical access to the equipment are performed by the Wigner data centre staff, such as installation of equipment into racks, repairs to the servers, etc.

The CERN and Wigner centres are connected via three independent and dedicated 100 Gb/s fibre-optic lines, with a bandwidth equivalent to the transfer of eight full DVDs per second. Network latency (the time taken between sending data and receiving on the other end) between the two sites, which are 1800 km apart, is about 25 milliseconds.

The Worldwide LHC Computing Grid (WLCG)

Physicists must sift through the many tens of petabytes of data produced annually by the LHC experiments to determine if the collisions have thrown up any interesting physics. CERN does not have the computing or financial resources to crunch all of the data produced by the LHC experiments on site, so in 2002 it turned to grid computing to share the burden with computer centres around the world. The Worldwide LHC Computing Grid (WLCG) — a distributed computing infrastructure arranged in tiers — gives a community of thousands of physicists near real-time access to LHC data.

With 170 computing centres in 42 countries, the WLCG is the most sophisticated data-taking and analysis system ever built for science. It runs more than 2 million jobs per day. The CERN data centre — working in unison with its extension in Budapest — forms "Tier-0" of the WLCG, the first point of contact between experimental data from the LHC and the grid.

With 170 computing centres in 42 countries, the WLCG is the most sophisticated data-taking and analysis system ever built for science.

03 HIGHLIGHTS

2018 at CERN

At the end of 2018, the second run of the LHC came to a close. Work is now under way to upgrade the collider and the experiments — as well as the various engineering and ICT systems that support these — in time for the start of the LHC's third run, beginning in 2021.

CERN also has a diverse and exciting experimental programme beyond the LHC. Below are a few highlights from 2018:

- The ALPHA experiment performed the most precise measurements of the spectrum of antihydrogen ever recorded.
- The new Alpha-G and GBAR experiments — also located in CERN's antimatter factory — prepared for new measurements on the gravitational properties of antimatter.
- The NA62 experiment saw hints of an ultra-rare charged kaon decay, which could be an indicator of physics beyond the Standard Model.
- The first particle tracks were seen in ProtoDUNE, a prototype for an international neutrino experiment.
- At the ISOLDE facility, researchers demonstrated — and explained — the changing shape of mercury isotopes.
- At AWAKE — an R&D project developing new particle-accelerator technology — researchers achieved the first ever acceleration of electrons in a proton-driven plasma wave. This could pave the way for smaller, more compact particle colliders in the future.

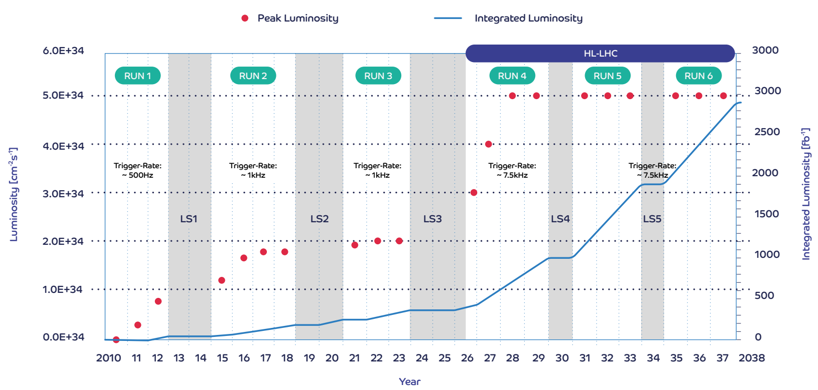
In December, operators of the CERN Control Centre turned off the Large Hadron Collider (LHC), ending the very successful second run of the world's most powerful particle accelerator. CERN's accelerator complex is now stopped for about two years to enable major upgrade and renovation works.

During this second run (2015–2018), the LHC performed beyond expectations, achieving approximately 33 million billion proton-proton collisions at an energy of 13 teraelectronvolts (TeV) and large datasets for lead-lead collisions at an energy of 5.02 TeV. These collisions produced an enormous amount of data, with more than 300 petabytes now permanently archived on tape in CERN's data centre. By analysing this data, the LHC experiments have already produced a large amount of results, extending our knowledge of fundamental physics and of the universe.

Over the past few years, the LHC experiments have made tremendous progress in the understanding of the properties of the Higgs boson. A cornerstone of the Standard Model of particle physics — the theory that best describes the elementary particles and the forces that bind them together — the Higgs boson was discovered at CERN in 2012 and has been studied ever since. In particular, physicists are analysing the way it decays or transforms into other particles, to check the Standard Model's predictions. Over the last three years, the LHC experiments extended the measurements of rates of Higgs boson decays, including the most common, but hard-to-detect, decay into bottom quarks, and the rare production of a Higgs boson in association with top quarks. The ATLAS and CMS experiments also presented updated measurement of the Higgs boson mass with the best precision to date.

Besides the Higgs boson, the LHC experiments produced a wide range of results and hundreds of scientific publications, including the discovery of exotic new particles such as Ξcc^{++} and pentaquarks with the LHCb experiment, and the unveiling of so-far unobserved phenomena in proton–proton and proton-lead collisions at ALICE.

During the coming two-year period, known as "Long Shutdown 2" (LS2), the whole accelerator complex and the detectors will be reinforced and upgraded for the next LHC run, starting in 2021, and the High-Luminosity LHC (HL-LHC) project, which will start operation after 2025. Increasing the luminosity of the LHC means producing far more data, and will lead to significant ICT challenges. For example, it is expected that — were current software, hardware, and analysis techniques to be used — the computing capacity required would be around 50-100 times greater than during the LHC's last run. Equally, data storage needs are expected to be in the order of exabytes by this time.



The planned upgrade schedule for the LHC.

04
ABOUT

The concept

CERN openlab is a unique public-private partnership that accelerates the development of cutting-edge ICT solutions for the worldwide LHC community and wider scientific research. Through CERN openlab, CERN collaborates with leading ICT companies and research institutes.

Members of the CERN openlab management team. From left to right: Maria Girone, chief technology officer; Fons Rademakers, chief research officer; Kristina Gunne, finance and administration officer; Alberto Di Meglio, head of CERN openlab; Andrew Purcell, communications officer; Ioanna Katsina Dimoula, junior communications officer; Federico Carminati, chief innovation officer.

Within the CERN openlab framework, CERN provides access to its advanced ICT infrastructure and its engineering experience — in some cases even extended to collaborating institutes worldwide. Testing in CERN's demanding environment provides the collaborating companies with valuable feedback on their products, while enabling CERN to assess the merits of new technologies in their early stages of development for possible future use. This framework also offers a neutral ground for carrying out advanced R&D with more than one company.

Industry collaboration can be at the associate, contributor, or partner level. Each status represents a different level of investment, with projects lasting typically between one and three years. The collaborating companies engage in a combination of cash and in-kind contributions, with the cash being used to hire young ICT specialists dedicated to the projects. The associate status formalises a collaboration based on independent and autonomous projects that do not require a presence on the CERN site. The contributor status is a collaboration based on tactical projects, which includes a contribution to hire an early-career ICT specialist supervised by CERN staff to work on the common project, in addition to the hardware and software products needed by the projects. The partners commit to a longer-term, strategic programme of work and provide three kinds of resources: funding for early-career researchers, products and services, and engineering capacity. The partners receive the full range of benefits of membership in CERN openlab, including extensive support for communications activities and access to dedicated events.

CERN openlab was established in 2001, and has been organised into successive three-year phases. In the first phase (2003–2005), the focus was on the development of an advanced computing-cluster prototype called the "opencluster". The second phase (2006–2008) addressed a wider range of domains. The combined knowledge and dedication of the engineers from CERN and the collaborating companies produced exceptional results, leading to significant innovation in areas such as energy-efficient

Frédéric Hemmer presents an award to representatives of Oracle, marking 15 years of fruitful collaboration with the company through CERN openlab. From left to right: Maria Girone, CERN openlab CTO; Alberto Di Meglio, head of CERN openlab; Frédéric Hemmer, head of the CERN IT Department; Cris Pedregal, technology director, Oracle Development; David Ebert, a director for industry solutions (public sector, education and research, healthcare) at Oracle.

computing, grid interoperability, and network security. CERN openlab's third phase (2009-2011) capitalised and extended upon the successful work carried out in the second phase. New projects were added focusing on virtualisation of industrial-control systems and investigation of the then-emerging 64-bit computing architectures. The fourth phase (2012-2014) addressed new topics crucial to the CERN scientific programme, such as cloud computing, business analytics, next-generation hardware, and security for the ever-growing number of networked devices. The fifth phase (2015-2017) tackled ambitious challenges covering the most critical needs of ICT infrastructures in domains such as data acquisition, computing platforms, data storage architectures, compute provisioning and management, networks and communication, and data analytics. It also saw other research institutes join CERN openlab for the first time.

This annual report covers the first year of CERN openlab's sixth phase (2018-2020). The ICT challenges to be tackled in this phase were set out in a white paper (https://openlab.cern/whitepaper), published at the end of 2017. Challenges related to data-centre technologies and infrastructures, computing performance and software, and machine learning and data analytics are being addressed.

At CERN openlab's annual technical workshops, representatives of the collaborating companies and research institutes meet with the teams, who provide in-depth updates on technical status. Collaborating companies and research institutes also elect representatives for CERN openlab's annual Collaboration Board meeting, which is an opportunity to discuss the progress made by the project teams and to exchange views on the collaboration's plans.

The CERN openlab team consists of three complementary groups of people: young engineers hired by CERN and funded by our collaborators, technical experts from the companies involved in the projects, and CERN management and technical experts working partly or fully on the joint activities.

The names of the people working on each project can be found in the results section of this report. The members of CERN openlab's management team are shown in the image on the previous page.

Attendees at the 2018 CERN openlab Collaboration Board.

SIEMENS
intel
ORACLE
Micron
HUAWEI
Google

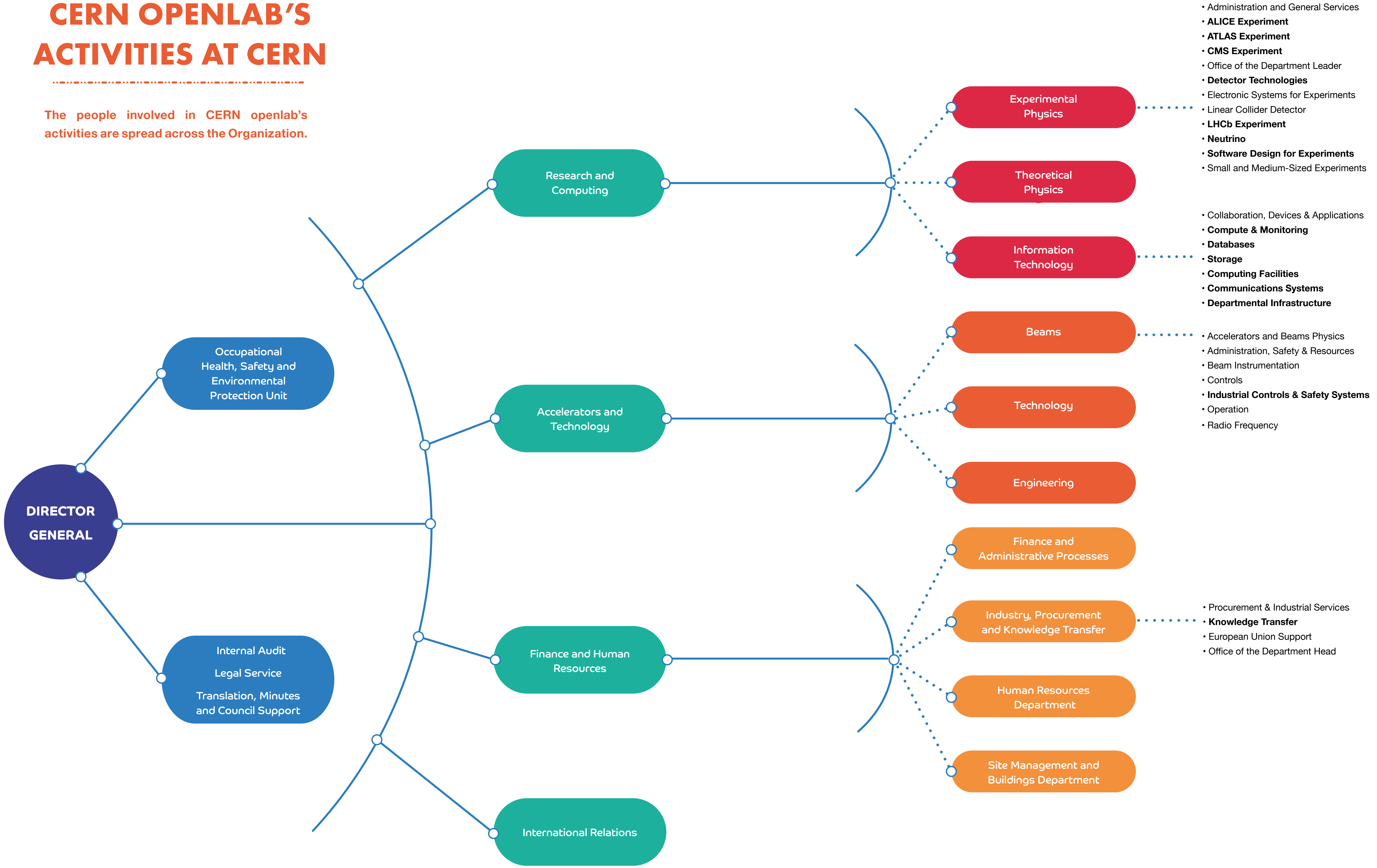
rackspace
IBM
Extreme
Connect Beyond the Network
E4
COMPUTER ENGINEERING

COMTRADE
open systems

INFN
Istituto Nazionale di Fisica Nucleare
IVI INNOVATION VALUE INSTITUTE
KING'S College LONDON
Newcastle University
EMBL-EBI
EUROPEAN SOCIETY OF PREVENTIVE MEDICINE
TU/e Technische Universiteit Eindhoven University of Technology
Fermilab
Scimpulse foundation

POSITIONING CERN OPENLAB'S ACTIVITIES AT CERN

The people involved in CERN openlab's activities are spread across the Organization.



05 RESULTS

Our projects

Information about each of the 17 technical projects that ran in 2018 can be found in this section. These projects are organised into four overarching R&D topics.

R&D Topic 1:
Data-centre technologies and infrastructures
Designing and operating distributed data infrastructures and computing centres poses challenges in areas such as networking, architecture, storage, databases, and cloud. These challenges are amplified and added to when operating at the extremely large scales required by major scientific endeavours.

CERN is evaluating different models for increasing computing and data-storage capacity, in order to accommodate the growing needs of the LHC experiments over the next decade. All models present different technological challenges. In addition to increasing the capacity of the systems used for traditional types of data processing and storage, explorations are being carried out into a number of alternative architectures and specialised capabilities. These will add heterogeneity and flexibility to the data centres, and should enable advances in resource optimisation.

R&D topic 2:
Computing performance and software
Modernising code plays a vital role in preparing for future upgrades to the LHC and the experiments. It is essential that software performance is continually increased by making use of modern coding techniques and tools, such as software optimising compilers, etc. It is also important to ensure that software fully exploits the features offered by modern hardware architecture, such as many-core GPU platforms, acceleration coprocessors, and innovative hybrid combinations of CPUs and FPGAs. At the same time, it is of paramount importance that physics performance is not compromised in drives to ensure maximum efficiency.

R&D topic 3:
Machine learning and data analytics
Members of CERN's research community expend significant efforts to understand how they can get the most value out of the data produced by the LHC experiments. They seek to maximise the potential for discovery and employ new techniques to help ensure that nothing is missed. At the same time, it is important to optimise resource usage (tape, disk, and CPU), both in the online and offline environments. Modern machine-learning technologies — in particular, deep-learning solutions applied to raw data — offer a promising research path to achieving these goals.

Deep-learning techniques offer the LHC experiments the potential to improve performance in each of the following areas: particle detection, identification of interesting collision events, modelling detector response in simulations, monitoring experimental apparatus during data taking, and managing computing resources.

R&D topic 4:
Applications in other disciplines
The fourth R&D topic is different to the others in this report, as it focuses on applications in other disciplines. By working with communities beyond high-energy physics, we are able to ensure maximum relevancy for CERN openlab's work, as well as learning and sharing both tools and best practices across scientific fields. Today, more and more research fields are driven by large quantities of data, and thus experience ICT challenges comparable to those at CERN.

CERN openlab's mission rests on three pillars: technological investigation, education, and dissemination. Collaborating with research communities and laboratories outside the high-energy physics community brings together all these aspects.

CLOUD STORAGE

R&D TOPIC 1: DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

rackspace

Project coordinators:
Daniel van der Ster, Tim Bell

Technical team:
Julien Collet

Rackspace liaisons:
Brian Stein, Philip Williams

Project goal
The goal is to characterise and improve the performance of the Ceph block storage for OpenStack, the software platform used to manage CERN's private computing cloud. Ceph is a free-software storage platform that provides interfaces for object-, block-, and file-level storage. Block storage is critical for the cloud, and plays a key role in supporting challenging use cases, such as those involving very large volumes of data, the need for high availability, and more.

Our work involves the development of tools to help us understand the critical bottlenecks in our Ceph system. This will enable us to precisely target future efforts to improve performance.

Background
Ceph is an open-source storage system which has become the de-facto standard solution for OpenStack clouds around the world. Here at CERN, Ceph block storage is capacity-oriented, offering acceptable latency and IOPS for the bulk of our block storage use-cases. We're working to deliver lower-latency and higher-IOPS storage, thus bringing a range of benefits to our users, such as the ability to run databases or interactive processing applications. Gaining a better understanding of the critical bottlenecks in the performance of this system will help us to take informed decisions about hardware procurement. This is particularly important where there is a trade-off between performance and capacity.

Progress in 2018
First, we developed a benchmarking suite for the assessment of cluster performance, in order to highlight the variations in performance between the different layers of Ceph. In parallel, various contributions to the Ceph project were developed and merged, in collaboration with the upstream team.

A Ceph "top" tool, enabling detailed evaluation of a cluster's workload, was also implemented in 2018. This tool helps Ceph administrators to understand the behaviour of the cluster. We worked to integrate it as a Ceph core feature, resulting in a first basic implementation.

Finally, a preliminary performance evaluation of a new all-flash, hyper-converged cluster was conducted. The first results show high levels of IOPS, a promising result that highlights the potential for deploying new use cases on top of Ceph at CERN.

Next steps
We plan on finalising the implementation of the built-in Ceph top tool, and will continue evaluating real CERN use cases on upcoming all-flash architectures.

Presentations
- D. van der Ster, Mastering Ceph Operations: Upmap and the Mgr Balancer (13 November). Presented at Ceph Day Berlin, Berlin, 2018. *cern.ch/go/6hzm*
- J. Collet, Ceph @ CERN (28 November). Presented at JTech Ceph day, Paris, 2018.

The project team is working to characterise and improve the performance of the Ceph block storage for OpenStack, the software platform used to manage CERN's private computing cloud.

EOS PRODUCTISATION

R&D TOPIC 1: DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

COMTRADE

Project coordinator:
Luca Mascetti

Technical team:
Elvin Sindrilaru

Comtrade liaisons:
Gregor Molan, Ivan Arizanovic, Branko Blagojevic

Project goal

This project is focused on the evolution of CERN's EOS large-scale storage system. The goal is to simplify the usage, installation, and maintenance of the system. In addition, we will add support for new client platforms, expand documentation, and implement new features/integration with other software packages.

Background

Within the CERN IT department, a dedicated group is responsible for the operation and development of storage infrastructure. This infrastructure is used to store the physics data generated by the experiments at CERN, as well as the files of all members of personnel.

EOS is a disk-based, low-latency storage service developed at CERN. It is tailored to handle large data rates from the experiments, while also running concurrent complex production workloads. This high-performance system now provides more than 300 petabytes of raw disks.

EOS is also the key storage component behind CERNBox, CERN's cloud-synchronisation service. This makes it possible to sync and share files on all major mobile and desktop platforms (Linux, Windows, macOS, Android, iOS), with the aim of providing offline availability to any data stored in the EOS infrastructure.

Progress in 2018

We are now in the third phase of this project. The team at Comtrade has been working to acquire further knowledge of EOS, with this activity carefully organised into nine separate work packages.

Three Comtrade engineers also visited CERN and spent two weeks working side-by-side with members of the development and operations teams (helping to handle hardware failures, reconfigurations, software upgrades, and user support). We were then able to work together to create a set of technical documents describing the main aspects of EOS, for use by future administrators and operators.

In addition, we set up a proof-of-concept system using container technology. This shows the potential of the system to be used as a geographically distributed storage system and will serve as a demonstrator to potential future customers.

Next steps

We will continue our evolving work on EOS installation, documentation, and testing. We will prepare a dedicated document outlining "best practices" for operating EOS in large-scale environments.

An additional goal is to provide future customers with a virtual full-stack environment hosted at Comtrade. This would consist of an EOS instance enabled with the latest-generation namespace, a sync-and-share endpoint (using CERNBox), and an interactive data-analysis service (based on SWAN, the JupyterHub notebook used at CERN).

The EOS system now provides more than 300 petabytes of raw disks.

Presentations

- L. Mascetti, Comtrade EOS productization (23 January). Presented at CERN openlab technical workshop, Geneva, 2019. *cern.ch/go/W6SQ*
- G. Molan, EOS Documentation and Tesla Data Box (4 February). Presented at CERN EOS workshop, Geneva, 2019. *cern.ch/go/9QbM*

EXTREME FLOW OPTIMIZER

R&D TOPIC 1: DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

Extreme
Connect Beyond the Network

Project coordinator:
Tony Cass

Technical team:
Adam Krajewski, Stefan Stancu, Edoardo Martelli, David Gutierrez Rueda

Extreme Networks liaisons:
Manjunath Gowda, Murtaza Bawahir, Mythil Raman, Christoph Kaelin, Giacomo Bernardi, Salvador Ferrer

Project goal

The Extreme Flow Optimizer (EFO) project aims to enhance Extreme Networks' EFO application. Our plan is to use it, in conjunction with the Extreme Workflow Composer (EWC) software, to build an automated traffic-steering system for our network. EFO improves visibility of network traffic and manages volumetric DDoS threats by applying dynamic flow steering. EWC is an event-driven IT automation platform, well suited for network automation purposes.

Background

As technology evolves, there are more and more use cases where traditional, static network configurations prove too rigid. This project uses the EFO and EWC software to provide increased programmability and flexibility in our network.

The project has led to upgrades to CERN's intrusion detection system (IDS), which relies on EFO and EWC to achieve scalability and to implement advanced features, such as the offloading of the inspection of bulk data transfers.

Progress in 2018

Throughout 2018, important contributions were made to the general EFO product development, and the IDS prototype at CERN was brought to maturity and put into production.

In the first part of the year, the CERN fellow funded by Extreme Networks continued to be fully integrated in the EFO development team. There, he made a major contribution to the redesign of the software architecture, making use of Docker containers. The fellow also provided expert consultancy regarding EFO to a major customer of Extreme Networks in Switzerland.

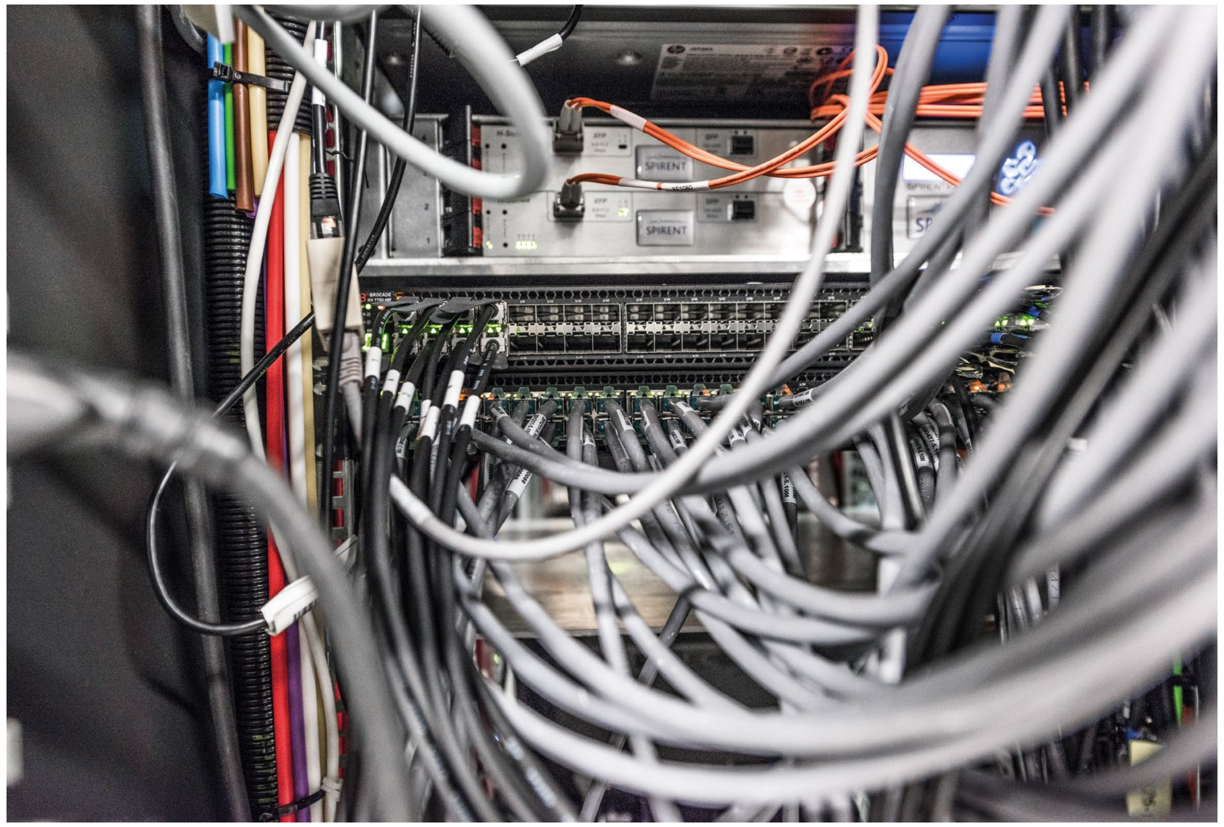
Later in the year, effort was directed into finalising the traffic orchestrator for the upgraded IDS system at CERN. The IDS receives a copy of the traffic that crosses CERN's network boundary and load-balances it across a pool of servers, each running the open-source Bro Network Security Monitor system. The traffic orchestrator uses Extreme Networks' SLX 9540 hardware platform, a high-end data-centre switch with advanced hardware capabilities for traffic orchestration. In addition, the EWC software provides increased automation capabilities through modular and configurable workflows, abstracting the network device configuration.

Next steps

The EFO project came to a close in October 2018, following three years of successful work. Throughout the project's lifetime, numerous contributions and enhancements were made to the EFO software, and the project ultimately resulted in an upgraded, scalable IDS for CERN. This IDS relies on a traffic orchestrator that makes use of both hardware and software from Extreme Networks.

Presentations

- A. L. Krajewski, Extreme Networks project Highlights (11 January). Presented at CERN openlab technical workshop, Geneva, 2018. *cern.ch/go/B8tw*
- A. L. Krajewski, Bro optimisations and network topologies (27 June). Presented at WLCG Security Operations Center WG workshop, Geneva, 2018. *cern.ch/go/w7Gh*



EFO improves visibility of network traffic and manages volumetric DDoS threats by applying dynamic flow steering.

INFRASTRUCTURE MONITORING AND AUTOMATION OF RESOURCE DEPLOYMENT

R&D TOPIC 1: DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

ORACLE

Project coordinators:
Eva Dafonte Perez, Eric Grancher

Technical team:
Viktor Kozlovszky, Luis Rodríguez Fernández, Artur Wiecek, Scott Hurley

Oracle liaisons:
Vincent Leocorbo, Cristobal Pedregal-Martin

Project goal

By learning from industrial standards and best practices, we're working to further improve and expand CERN's infrastructure monitoring and deployment processes for our Java platform.

Background

CERN's IT department is home to a dedicated group responsible for database services, referred to as IT-DB. This group maintains server infrastructures required by departments across the laboratory. In order to successfully maintain these infrastructures, it is important to be constantly aware of our systems' status. The monitoring infrastructure consists of several components; these collect essential information about the machines and the applications running on them.

The IT-DB group also provides tools and custom Java libraries that play an important role in the processes of teams across the laboratory. It is therefore vital to deliver stable, quality applications that build on industrial standards and best practices.

Progress in 2018

In 2018, we began evaluation of the commercial Java monitoring tools and features provided by Oracle. We set up a test environment and we managed — through an SSL secured channel — to establish a connection between the commercial monitoring clients and the test application servers. The outcome of our experimental work has been published on our group's blog (see list of publications).

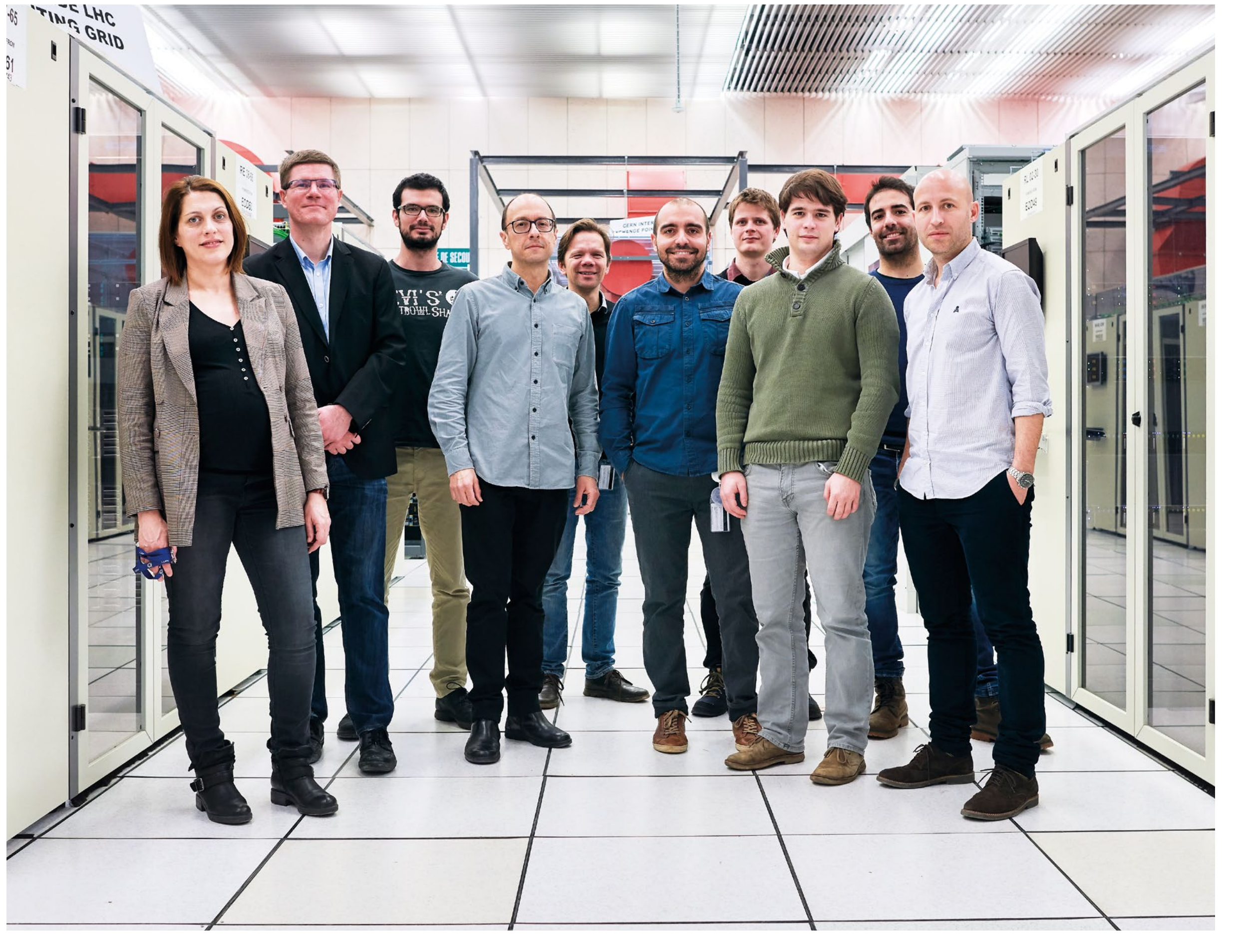
At the end of the year, we also started work to update and improve our Java tools by incorporating the latest automated-testing and deployment practices from industry.

Next steps

In terms of monitoring, we will work to evaluate the behaviour tracking (Java "flight recording") feature of Oracle, comparing it with our existing monitoring solutions, and will use our Kubernetes cluster to evaluate the commercial monitoring features from Oracle. In terms of automation, we will update the remaining Java tools to ensure they can make use of the new automated testing and deployment approaches.

Publications

- S. Hurley, Configuring Technologies to Work with Java Mission Control. Databases at CERN blog. 2018. *cern.ch/go/Rzs8*
- S. Hurley, Java Mission Control Evaluation. Zenodo blog. 2018. *cern.ch/go/6kzW*
- V. Kozlovszky, JMX connection with SSL. Databases at CERN blog. 2018. *cern.ch/go/L9MG*



Members of the CERN IT department's database services group are collaborating with Oracle on five separate projects through CERN openlab.

OPENSTACK CLOUDS

R&D TOPIC 1: DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

HUAWEI

Project coordinator:
Tim Bell

Technical team:
Surya Seetharaman, Theodoros Tsioutsias

Huawei liaison:
Stephan Keuneke

Project goal

CERN's cloud infrastructure is among the world's largest private OpenStack installations. CERN and Huawei therefore have a common interest in driving and enhancing the development of OpenStack for large-scale applications. Work on this project is being carried out through the community processes of OpenStack, with the results being made fully open source.

Background

It is important to ensure that the management system used for allocating CERN's computing resources is as efficient as possible and enables scientists to gain access to resources on a simple, self-service basis. By running common software based on OpenStack — both in on-premises and commercial clouds — we can meet the end users' needs for LHC Run 3 and beyond in an efficient and seamless manner. The enhancements we make will be submitted upstream and included in future releases of OpenStack, meaning that they will be available for use both in CERN's private, on-premises cloud and in products from Huawei based on OpenStack.

Work on this project is being carried out through the community processes of OpenStack, with the results being made fully open source.

Progress in 2018

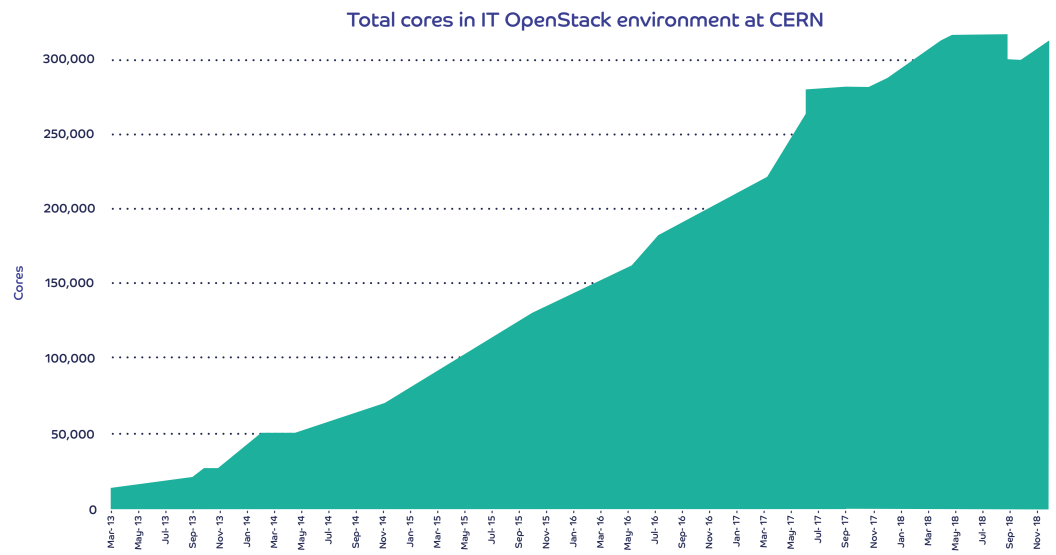
There are two main areas of focus for this project, namely OpenStack Cells v2 and pre-emptible instances.

Cells provides a way to build large clouds out of pools of resources; an initial implementation of this has been used at CERN since the in-house cloud went into production in 2013. The OpenStack community has developed a new version, Cells v2, which is now the default for all OpenStack clouds. It was put into production at CERN in 2018. We have worked closely with the upstream developers to design the solution, validate it at scale, and enhance the code — based on the experience we have gained.

The use of pre-emptible instances in OpenStack makes it possible for users to create virtual machines that can be rapidly stopped if higher priority work arrives, thus enabling CERN's computing resources to be managed more efficiently. The design has been approved by the OpenStack community and prototyped at CERN.

Next steps

The enhancements for Cells v2 are now included in the upstream OpenStack package, with the pre-emptible instances soon due to go into use on the production cloud. This will complete the deliverables for the project in 2019.



Total cores in the OpenStack environment at CERN.

Presentations

- B. Moreira, Moving from CellsV1 to CellsV2 at CERN (21 May). Presented at OpenStack Summit, Vancouver, 2018. *cern.ch/go/nK6l*
- B. Moreira, Containers on Baremetal and preemptible VMs at CERN and SKA (24 May). Presented at OpenStack Summit, Vancouver, 2018. *cern.ch/go/qhJ6*
- B. Moreira, Optimisations OpenStack Nova for Scientific Workloads (10 July). Presented at 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP), Sofia, 2018. *cern.ch/go/qhJ6*
- B. Moreira, S. Seetharaman, Scaling Nova with Cells V2 (13 November). Presented at OpenStack Summit, Berlin, 2018.
- T. Tsioutsias, Pre-emptible instances and bare metal containers (15 November). Presented at OpenStack Summit, Berlin, 2018. *cern.ch/go/qhJ6*

ORACLE MANAGEMENT CLOUD

R&D TOPIC 1: DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

Project coordinators:
Eva Dafonte Perez, Eric Grancher

Technical team:
Aimilios Tsouvelekakis, Artur Wiecek

Oracle liaisons:
Jeff Barber, Simone Indelicato, Vincent Leocorbo, Cristobal Pedregal-Martin

Project goal

We are testing Oracle Management Cloud and providing feedback to Oracle. We are assessing the merits and suitability of this technology for applications related to databases at CERN, comparing it with our current on-premises infrastructure.

Background

The group responsible for database services within CERN's IT department provides specialised monitoring solutions to teams across the laboratory which use database infrastructure. These solutions are used for a range of tools, from servers to applications and databases. Monitoring performance in this manner provides invaluable insights, and is key in helping those responsible for providing services at the laboratory to gain a full picture of what is going on with the infrastructure at all times. To accomplish this monitoring functionality, we use two different monitoring stacks: ElasticSearch for log management and InfluxDB for metrics management. This project is evaluating a unified monitoring solution provided by Oracle: Oracle Management Cloud.

Progress in 2018

Last year's evaluation took place in two distinct phases. The first was performed in February and March; this mainly focused on deploying the required components in our infrastructure, using only a subset of our datasets. The second evaluation phase — for which Oracle granted CERN a significant amount of cloud credits — lasted from June to December. During this time, we evaluated three components of the platform: log analytics, infrastructure monitoring, and application performance monitoring.

We used datasets from CERN's Oracle REST Data Services (ORDS) and CERN's Engineering Data Management Service (EDMS) — combining development, test, and production environments — to evaluate each of the three aforementioned components. From this, we were able to generate important graphs for logs and metrics, which — based on experience with our current, on-premises infrastructure — could be a significant boon when it comes to dealing with issues that arise. Based on our evaluation, we were able to provide in-depth feedback and propose possible enhancements that could be useful for other environments like ours.

Next steps

Our primary focus will be on continuing to work with Oracle on the evolution of the platform, based on the detailed feedback provided from our rigorous testing.

Presentations

- A. Tsouvelekakis, Oracle Management Cloud: A unified monitoring platform (23 January). Presented at CERN openlab Technical Workshop 2019, Geneva, 2019.

The group responsible for database services within CERN's IT department provides specialised monitoring solutions to teams across the laboratory which use database infrastructure.

ORACLE WEBLOGIC ON KUBERNETES

R&D TOPIC
1: DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

Project coordinators:
Eva Dafonte Perez, Eric Grancher

Technical team:
Antonio Nappi, Luis Rodríguez Fernández, Artur Wiecek

Oracle liaisons:
Vincent Leocorbo, Cristobal Pedregal-Martin, Monica Riccelli, David Cabelus, Will Lyon

Project goal
CERN is in the process of moving its Oracle WebLogic infrastructure to containers and Kubernetes, starting with the development environment. The goal is to achieve a robust, zero-downtime service. Taking advantage of the portability of Kubernetes, we want to evaluate Oracle cloud as a solution for disaster recovery.

Background
For over 20 years, CERN has run a production service to host critical Java applications. Many of these applications are central to the administration of the laboratory, while others are important for engineering or IT. We're working on solutions to help keep these applications running in case of major problems with the CERN data centre.

At CERN's database-applications service, there is ongoing work to migrate from virtual machines to Kubernetes. We're capitalising on this opportunity to evaluate how our services can run on public clouds — in particular, on Oracle cloud. This new architecture will increase the team's productivity, freeing up time to focus more directly on developers' needs.

Progress in 2018
In 2018, we consolidated the work of the previous year. We worked on two versions of Oracle WebLogic, thus ensuring backward compatibility with legacy applications and giving our users the opportunity to test the newer version. We also integrated a new open-source tool, called Oracle WebLogic Deploy Tooling, into our infrastructure. This is used to easily configure WebLogic domains starting from simple configuration files. Integration of this tool has enabled us to move the configuration of the WebLogic infrastructure outside the Docker images and to increase the speed at which images are generated. In addition, we developed tools to automate the deployment workflow of new applications on Kubernetes.

Another area of work in 2018 was the evaluation of Oracle WebLogic Operator. This is a new open-source tool that provides a WebLogic environment running on Kubernetes. We worked very closely with the Oracle team responsible for this tool, with a number of our feedback and suggestions having a direct impact on new releases.

Next steps
In 2019, we will mainly focus on ensuring that our production environment runs on Kubernetes. In addition, we will start to evaluate a disaster recovery plan to run on the Oracle cloud. We will also look into new options for our monitoring infrastructure; in particular, we will evaluate a tool called Prometheus.

Presentations

- L. Rodriguez Fernandez, A. Nappi, Weblogic on Kubernetes (11 January). Presented at CERN Openlab Technical Workshop, Geneva, 2018. *cern.ch/go/6Z8R*
- B. Cotarelo, Oracle Weblogic on Kubernetes (July). Presented at 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP), Sofia, 2018. *cern.ch/go/6MVQ*
- M. Riccelli, D. Cabelus, A. Nappi, Running a Modern Java EE Server in Containers Inside Kubernetes (23 October). Presented at Oracle Openworld, San Francisco, 2018. *cern.ch/go/b6nl*

At CERN's database-applications service, there is ongoing work to migrate from virtual machines to Kubernetes.

HIGH-PERFORMANCE CLOUD CACHING TECHNOLOGIES

R&D TOPIC 2: COMPUTING PERFORMANCE AND SOFTWARE

Project coordinator:
Giovanna Lehmann Miotto

Technical team:
Danilo Cicalese, Fabrice Le Goff, Jeremy Love, Remigius K Mommsen

Intel liaisons:
Grzegorz Jereczek, Maciej Maciejewski, Jakub Radtke, Jakub Schmiegel, Malgorzata Szychowska, Aleksandra Jereczek, Adrian Pielech, Claudio Bellini

Project goal

We're exploring the suitability of a new infrastructure for key-value storage in the data-acquisition systems of particle-physics experiments. DAQDB (Data Acquisition Database) is a scalable and distributed key-value store that provides low-latency queries. It exploits Intel Optane DC Persistent Memory, a cutting-edge non-volatile memory technology that could make it possible to decouple real-time data acquisition from asynchronous event selection.

Background

Upgrades to the LHC mean that the data rates coming from the detectors will dramatically increase. Data will need to be buffered while waiting for systems to select interesting collision events for analysis. However, the current buffers at the readout nodes can only store a few seconds of data due to capacity constraints and high cost of DRAM. It is therefore important to explore new, cost-effective solutions — capable of handling large amounts of data — that capitalise on emerging technologies.

Progress in 2018

During 2018, we worked to assess the potential of this new approach. We collaborated closely with Intel on this, coordinating with them on key design choices for the project. Dedicated servers are needed to make use of the newest hardware, such as persistent memory and NVMe SSDs. We set up the new hardware at CERN and integrated it into the existing data-acquisition software platforms for the ATLAS experiment. We then tested DAQDB thoroughly, providing feedback to the developers.

In addition, we explored a range of alternative solutions for modifying the current ATLAS data-acquisition dataflow and integrating the key-value store. We successfully integrated the data-acquisition software and were also able to separate the readout from the storage nodes, thus making them independent.

Next steps

In the first part of 2019, we will evaluate the system's performance in a small-scale test. We will then focus our attention on assessing and improving performance in a more realistic scenario, with a large-scale experiment. We will test Intel® Optane™ DC memory technology and Intel® Optane™ DC SSD. In parallel, we will begin work to integrate the system into other experiments, such as CMS and ProtoDUNE.

Presentations

- G. Jereczek, The design of a distributed key-value store for petascale hot storage in data acquisition systems (12 July). Presented at 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP), Sofia, 2018. *cern.ch/go/6hcX*

- M. Maciejewski, A key-value store for Data Acquisition Systems (12 September). Presented at ATLAS TDAQ week, Cracow, 2018.

- G. Jereczek, M. Maciejewski, Data Acquisition Database (12 November). Presented at The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC18), Dallas, 2018.

- M. Maciejewski, J. Radtke, The Design of Key-Value Store for Data Acquisition Systems (5 December). Presented at NMVe Developer Days, San Diego, 2018.

Upgrades to the LHC mean that the data rates coming from the detectors will dramatically increase.

INTEL BIG-DATA ANALYTICS

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

Project coordinators:
Luca Canali, Maria Girone, Eric Grancher

Technical team:
Evangelos Motesnitsalis, Viktor Kozlovszky, Viktor Khristenko, Matteo Migliorini, Vasileios Dimakopoulos, Matteo Cremonesi, Oliver Gutsche, Bian Bianny, Klaus-Dieter Oertel

Intel liaisons:
Claudio Bellini, Mike Riess

Project goal
At CERN, researchers are always exploring the latest scalable solutions needed to tackle a range of data challenges at the laboratory, related to both physics analysis and machine analytics. This project aims to help optimise analytics solutions for data integration, data ingestion, data transformation, performance, scalability, benchmarking, resource management, data visualisation, and hardware utilisation.

Background
The LHC experiments continue to produce large amounts of physics data, which offers numerous possibilities for new discoveries. Big-data technologies, such as Apache Spark, hold great potential for helping us to optimise our existing physics data-processing procedures, as well as our solutions for industrial control and online processing. Through this project, we are working to design and optimise solutions based on open-source big-data technologies. This work is being carried out in collaboration with Intel, the CMS experiment, the CERN IT department, the Fermi National Accelerator Laboratory (Fermilab) in the United States, and DIANA/HEP (a collaborative endeavour to develop state-of-the-art software tools for high-energy physics experiments).

Progress in 2018
In 2018, the project mostly focused on use cases related to the processing of physics data at scale. In particular, we built on two key data-engineering challenges that were tackled in the previous year: the development of a mature Hadoop-XRootD Connector library, which makes it possible to read files from the CERN's EOS storage system, and the Spark-ROOT library, which makes it possible to read ROOT files in Spark DataFrames (ROOT is an object-oriented program and library developed at CERN that provides tools for big data processing, statistical analysis, visualisation, and storage). We were able to produce, scale up, and optimise physics data-processing workloads on Apache Spark and test them with over one petabyte of open data from the CMS experiment.

In addition, we worked to address challenges related to the application of machine-learning solutions on physics data, using Intel BigDL (a distributed deep-learning library for Apache Spark) alongside a combination of Keras (an open-source neural network library) and TensorFlow (an open-source machine-learning framework). This led to promising results. The compatibility of the developed workloads with popular open-source analytics and machine-learning frameworks makes them very appealing, with various analysis groups from the CMS experiment choosing to carry out further development of these solutions.

Next steps
We will repeat the workload tests on top of virtualised/containerised cloud-native infrastructure, complete with Kubernetes. This will include running at CERN and performing tests on public clouds.

Furthermore, we also have plans for extending the techniques developed in the project to tackle more workloads. For example, we will work to address more complex physics data-processing challenges, such as use cases related to machine learning for online data processing (streaming).

Presentations

- E. Motesnitsalis, Hadoop and Spark Services at CERN (19 April). Presented at DataWorks Summit, Berlin, 2018. *cern.ch/go/LF8r*
- E. Motesnitsalis, From Collision to Discovery: Physics Analysis with Apache Spark (April). CERN Spring Campus, Riga, 2018.
- E. Motesnitsalis, Big Data Technologies at CERN (April). CERN Spring Campus, Riga, 2018.
- V. Khristenko, Physics Analysis with Apache Spark in the CERN Hadoop Service and DEEP-EST Environment (25 May). Presented at IT Technical Forum, Geneva, 2018.
- E. Motesnitsalis, From Collision to Discovery: Physics Analysis with Apache Spark (7 August). Presented at IT Lectures CERN openlab Summer Student Programme, Geneva, 2018. *cern.ch/go/W9HF*
- E. Motesnitsalis, Big Data at CERN (20 September). Presented at Second International PhD School on Open Science Cloud, Perugia, 2018. *cern.ch/go/JL8Q*
- M. Cremonesi et al., Using Big Data Technologies for HEP Analysis (July). Presented 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP), Sofia, 2018. *cern.ch/go/D8wv*

The project team is working to design and optimise solutions based on open-source big-data technologies.

ORACLE CLOUD TECHNOLOGIES FOR DATA ANALYTICS ON INDUSTRIAL CONTROL SYSTEMS

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

ORACLE

Project coordinators:
Eva Dafonte Perez, Eric Grancher

Technical team:
Manuel Martin Marquez, Sébastien Masson, Franck Pachot

Oracle liaisons:
Cemil Alper, Dimitry Dolgushin, David Ebert, Vincent Leocorbo, Pauline Maher, Cristobal Pedregal-Martin, Reiner Zimmermann

Project goal

This project is working to assess the capabilities of Oracle Autonomous Data Warehouse Cloud (ADWC) and Oracle Autonomous Analytics Cloud (AAC). These technologies are being tested for use in handling the masses of data that come from the control and monitoring systems in place for CERN's accelerator complex. Specifically, our goal is to try to use these technologies to integrate different existing datasets, to improve the performance and efficiency for the most important and challenging data retrieval/analysis, and to unlock new possibilities for data exploration.

Background

The LHC is one of the largest, most complex machines ever built. Keeping it — and the rest of the accelerator complex at CERN — running efficiently requires state-of-the-art control systems. More than 2.5 terabytes of monitoring data is generated per day, coming in from over a million signals spread across the accelerators and detectors. A complex "Industrial Internet of Things" (IIoT) system is in place to persist this data, making it possible for scientists and engineers to gain insights about temperatures, magnetic field strengths, beam intensities, and much more. This plays a vital role in ensuring the highest levels of operational efficiency.

The current system to persist, access, and analyse the controls and monitoring data is based on Oracle Database. Today, significant effort is dedicated to improving performance and coping with increasing demand — in terms of both data volume and analysis of bigger datasets.

Progress in 2018

We organised our work in 2018 into three phases. In the initial phase, we carried out a high-level feasibility study of ADWC and AAC, making sure the technology could manage the extreme demands of our IIoT systems and our complex analytics queries. In this phase, we also explored the flexibility of provisioning, as well as the ability of the technology to automate updates, backups, and patches.

The second phase was dedicated to the evaluation of various procedures for migrating the data from our current on-premises architectures to Oracle's cloud services. In particular, we considered the complexity of the data format, partitioning, indexing, etc. This work made it possible for us to evaluate the initial workload and data-analysis performance on a representative subset of the data, helping us to gain insights into the advanced optimisation features of AAC. We were also able to use Oracle Hybrid Columnar Compression to reduce storage requirements to about a tenth of what they previously were, as well as reducing the requirement for full scans. Thus, the performance for data retrieval and analytics tasks was significantly improved. On top of this, the system offered transparent and automated access to Oracle's "Exadata SmartScan" and "Exadata Storage Indexes" features. This reduced — or, in some cases, removed entirely — the dependency on indexes.

In the last phase, we also worked with AAC to offer seamless data analytics based on collaborative and interactive dashboards. Our most recent work focuses on elasticity and scalability. In particular, we are working to increase the data volume used to one terabyte and increase the complexity of the workloads and analysis.

CERN's accelerator complex creates more than 2.5 terabytes of monitoring data per day, coming in from over a million signals spread across the accelerators and detectors.

Next steps

This will lead to a comparison between the Autonomous Database's capacities and other database platforms, including the current on-premises setup.

Presentations

- E. Grancher, M. Martin Marquez, S. Masson, Boosting Complex IoT Analysis with Oracle Autonomous Data Warehouse Cloud (23 October). Presented at Oracle Openworld 2018, San Francisco, 2018. *cern.ch/go/RBZ6*
- Eric Grancher, M. Martin Marquez, S. Masson, Managing one of the largest IoT Systems in the world (December). Presented at Oracle Global Leaders Meeting – EMEA, Sevilla, 2018.
- M. Martin Marquez, Boosting Complex IoT Analysis with Oracle Autonomous Data Warehouse Cloud (June). Presented at Oracle Global Leaders Meeting – EMEA, Budapest, 2018.

DATA ANALYTICS FOR INDUSTRIAL CONTROLS AND MONITORING

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

SIEMENS

Project coordinator:
Fernando Varela

Technical team:
Filippo Tilaro, Jakub Guzik, Rafal Kulaga, Piotr Golonka, Peter Sollander, Fernando Varela, Enrique Blanco

Siemens liaisons:
Thomas Hahn, Juergen Kazmeier, Alexey Fishkin, Tatiana Mangels, Mikhail Kalinkin, Elisabeth Bakany, Ewald Sperrer

Project goal

Our aim is to make control systems used for the LHC more efficient and "smarter". We are working to develop a data-analytics platform that capitalises on the latest cloud- and edge-computing technologies. Specifically, this platform will make use of two new analytics solutions being developed by Siemens, internally referred to as "Smart Industrial Internet of Things" (Smart IIoT) and "ELVis".

Background

The HL-LHC project aims to increase the integrated luminosity — and hence the rate of particle collisions — by a factor of ten beyond the LHC's design value. Monitoring and control systems will therefore become increasingly complex, with unprecedented data throughputs. Consequently, it is vital to further improve the performance of these systems, and to make use of data-analytics algorithms to detect anomalies and to anticipate future behaviour. Achieving this involves a number of related lines of work. This particular project focuses on the development of a data-analytics platform that combines the benefits of cloud and edge computing.

Progress in 2018

We focused on the monitoring of various LHC control systems, using two distinct analytics solutions from Siemens. The first is Smart IIoT, a framework used to monitor a multitude of control signals in a distributed manner. The second is ELVis, a web-based platform for processing, visualisation and archiving multiple streams of time-series data from sensors.

In 2018, these were both deployed and integrated into CERN's industrial installations, combining both cloud and edge computing into a single, scalable platform. Different versions and releases were assessed to compute (online) thousands of analytical rules against heavy streams of measurements.

Several machine-learning algorithms were developed to detect anomalies in different control systems. For example, a probabilistic model-based approach has now been adopted for leak detection in cooling and ventilation systems. In 2018, we also began work to help optimise the ion beam source for one of the linear accelerators at CERN (Linac 3).

Despite this work still being at an early stage, the Siemens analytical solutions deployed at CERN have already enhanced our control-systems monitoring and have reduced operational cost by extending the operational life of some devices.

Next steps

A tighter integration between ELVis and Smart IIoT will be one of the main objectives for 2019. A single user-interface will make it easier to define complex event-processing rules, configure the cloud and edge infrastructure, and monitor the execution of the analyses. A new version of the domain-specific language will also be implemented to integrate control events and time-series data processing.

Presentations

- F. Tilaro, F. Varela, Model Learning Algorithms for Anomaly Detection in CERN Control Systems (25 January). Presented at BE-CO Technical Meeting, Geneva, 2018. *cern.ch/go/7SGK*
- F. Tilaro, F. Varela, Industrial IoT in CERN Control Systems (21 February). Presented at Siemens IoT Conference, Nuremberg, 2018.
- F. Tilaro, F. Varela, Optimising CERN control systems through Anomaly Detection & Machine Learning (29 August). Presented at AI workshop for Future Production Systems, Lund, 2018.
- F. Tilaro, F. Varela, Online Data Processing for CERN industrial systems (12 November). Presented at Siemens Analytics Workshop, Munich, 2018.

The team is working to develop a data-analytics platform that capitalises on the latest cloud- and edge-computing technologies.

DATA ANALYTICS IN THE CLOUD

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

ORACLE

Project coordinators:
Eva Dafonte Perez, Eric Grancher

Technical team:
Luca Canali, Riccardo Castellotti

Oracle liaisons:
Barry Gleeson, Vincent Leocorbo, Don Mowbray, Cristobal Pedregal-Martin

Project goal

This project is testing and prototyping solutions that combine data engineering with machine-learning and deep-learning tools. These solutions are being run using cloud resources — in particular resources and tools from Oracle Cloud Infrastructure (OCI) — and address a number of use cases of interest to CERN's community. Notably, this activity will make it possible to compare the performance, maturity, and stability of solutions deployed on CERN's infrastructure with the deployment on the OCI.

Background

Big-data tools — particularly related to data engineering and machine learning — are evolving rapidly. As these tools reach maturity and are adopted more broadly, new opportunities are arising for extracting value out of large data sets.

Recent years have seen growing interest from the physics community in machine learning and deep learning. One important activity in this area has been the development of pipelines for real-time classification of particle-collision events recorded by detectors of the LHC detectors. Filtering events using so-called "trigger" systems is set to become increasingly complex as upgrades to the LHC increase the rate of particle collisions.

Progress in 2018

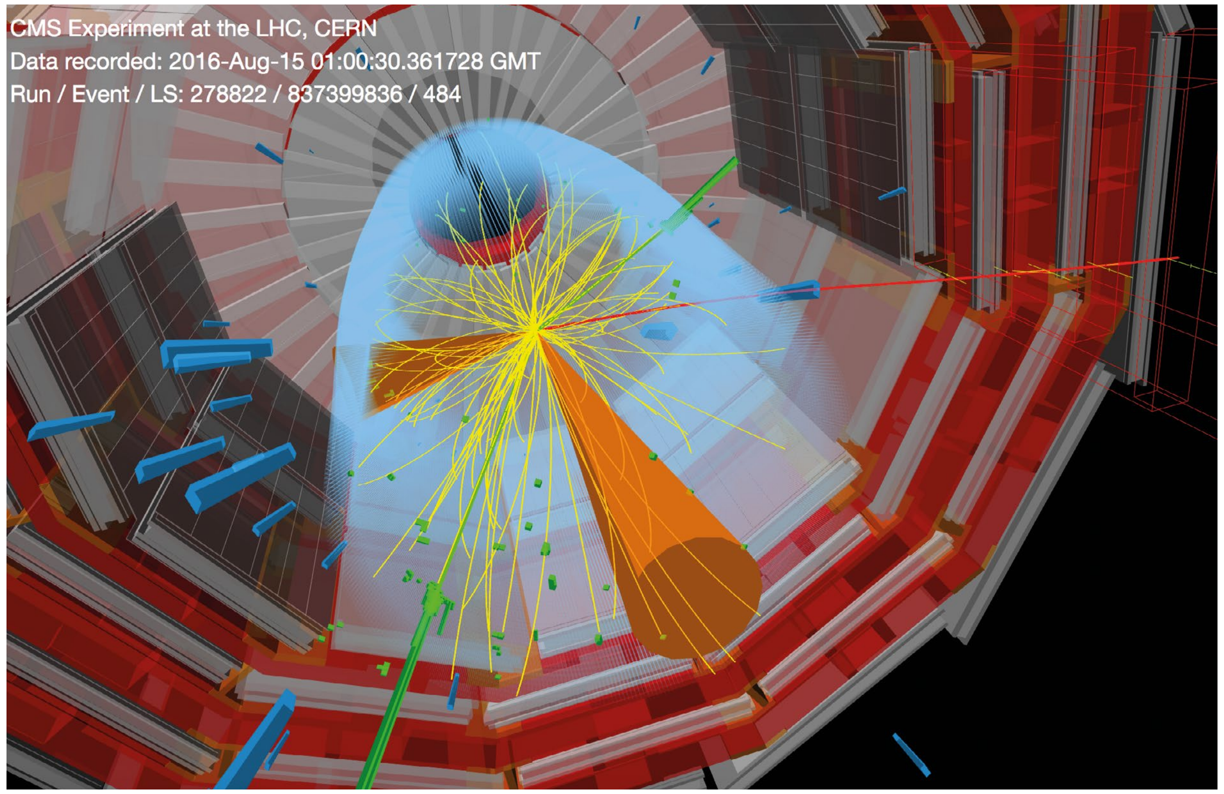
The project launched at the end of 2018. We began by developing and porting data pipelines — related to data-analysis and machine-learning use cases of interest — to the cloud (Kubernetes). The workloads had originally been developed to run on CERN's Hadoop and Spark infrastructure.

Next steps

In 2019, the project will investigate two main use cases. Firstly, the replication of data-reduction systems employed at the CMS experiment, with a view to exploiting the scalability of OCI to improve upon current performance. Secondly, the deployment of specific training of the models on OCI using GPUs to test the performance limits of such models on cloud-native solutions. The training models in question are those detailed in the paper listed under publications.

Publications

- T. Nguyen et al., Topology classification with deep learning to improve real-time event selection at the LHC, 2018. *cern.ch/go/8trZ*



Visualisation of a particle-collision event at the CMS experiment.

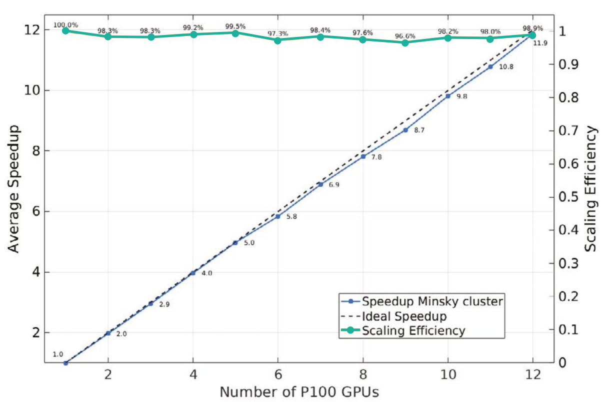
EVALUATION OF POWER CPU ARCHITECTURE FOR DEEP LEARNING

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

Project coordinators:
Maria Girone, Federico Carminati

Technical team:
Ahmad Hesam, Daniel Hugo Cámpora Pérez, Niko Neufeld

IBM liaisons:
Eric Aquaronne, Lionel Clavien



Training time scaling performance for 3D GAN.

Project goal

We are investigating the performance of distributed learning and low-latency inference of generative adversarial networks (GANs) for simulating particle collision events. The performance of a deep neural network is being evaluated on a cluster consisting of IBM Power CPUs (with GPUs) installed at CERN.

Background

GANs offer potential as a possible way of eliminating the need for classical Monte Carlo (MC) simulations in generating particle showers. Classical MC is computationally expensive, so this could be a way to improve the overall performance of simulations in high-energy physics.

Using the large data sets obtained from MC-simulated physics events, the GAN is able to learn to generate events that mimic these simulated events. Once an acceptable accuracy range is achieved, the trained GAN can replace the classical MC simulation code, with an inference invocation of the GAN.

Progress in 2018

In accordance with the concept of data-parallel distributed learning, we trained the GAN on a total of twelve GPUs, distributed over the three nodes that comprise the test Power cluster. Each GPU ingests a unique part of the physics data set for training the model. The neural network was implemented with a combination of software frameworks, optimised for Power architectures: Keras, TensorFlow, and Horovod. We used an MPI to distribute the workloads over the GPUs. As a result, we achieved great scaling performance, and we were able to improve the training time by an order of magnitude. With the trained model, we achieved a speedup of four orders of magnitude, compared to using classical MC simulation.

At the LHCb experiment, a convolutional neural network was also tested as a way of identifying particles from a certain type of electromagnetic radiation trace observed in particular sub-detectors. We fed a dataset with more than 5 million MC-generated particles through a deep neural network consisting of 4 million parameters. We used the same cluster to accelerate our exploration of this approach, achieving promising results.

Next steps

At the LHCb experiment, work will continue to improve the particle-identification performance of our new approach by incorporating new parameters into the model. We are seeking to identify the topology of the neural network that will best suit our problem; collaborating closely with IBM is key to achieving this.

We will also prototype a deep-learning approach for the offline reconstruction of events at DUNE, a new neutrino experiment that will be built in the United States. We believe that IBM's Power architecture could be well suited to handling the large amounts of raw data that will be generated by this experiment.

At the LHCb experiment, work will continue in 2019 to improve the particle-identification performance of our new approach by incorporating new parameters into the model.

Presentations

- A. Hesam, Evaluating IBM POWER Architecture for Deep Learning in High-Energy Physics (23 January). Presented at CERN openlab Technical Workshop, Geneva, 2018. cern.ch/go/7BsK
- D. H. Cámpora Pérez, ML based RICH reconstruction (8 May). Presented at Computing Challenges meeting, Geneva, 2018. cern.ch/go/xwr7
- D. H. Cámpora Pérez, Millions of circles per second. RICH at LHCb at CERN (7 June). Presented as a seminar in the University of Seville, Seville, 2018.

FAST SIMULATION

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

Project coordinator:
Federico Carminati

Technical team:
Sofia Vallecorsa, Gulrukh Khattak, Andrea Luiselli, Hans Pabst

Intel liaisons:
Claudio Bellini, Marie-Christine Sawley, Saletore Vikram

Project goal

We are developing fast-simulation tools based on machine learning — rather than primarily using classical Monte Carlo — to simulate particle transport in the detectors of the LHC experiments. Such tools could play a significant role in helping the research community to cope with the LHC's increasing computing demands.

The tools we are developing should be able to simulate a large variety of particles, energies, and detectors — all in a fraction of the time needed for classical simulation of particle transport. Our final objective is to integrate our tool in the existing code. This work is being carried out in collaboration with SURFsara and Cineca, as well as with Intel.

Background

Over half of the WLCG's computing workload is the result of a single activity, namely detector simulation. A single code, called Geant4, is used for this. Optimising this code has the potential to significantly reduce computing requirements, thus unlocking resources for other tasks.

The project is developing fast-simulation tools based on machine learning to simulate particle transport in the detectors of the LHC experiments.

Fast-simulation techniques have been developed in the past. However, the latest developments in machine learning (particularly in relation to deep neural networks) make it possible to develop fast-simulation techniques that are both more flexible and more accurate than existing ones.

Progress in 2018

Training time has turned out to be a major bottleneck for the meta-optimisation of our generative adversarial network. This includes not only the network weight, but also its architecture and the convergence parameters. Much of our work in 2018 concentrated on addressing this. We implemented distributed versions of our training code both on GPUs and CPUs, and we tested their performance and scalability in different environments (HPC clusters and cloud). The results are very encouraging: we observed almost linear speedup as the number of processors increased, with very limited or no degradation in results.

The other main area of work in 2018 related to the extension of the fast simulation tool to incorporate a larger set of kinematic conditions. We successfully extended the parameters related to incoming particles, integrating the angle of impact in the conditioning parameters. The tool is now mature enough to start planning its test integration with a classical Monte-Carlo code, such as Geant4.

Next steps

We plan to continue improving the accuracy of the simulation, with particular attention to the tails of particle showers and single-cell energy distribution. We will also continue to investigate HPC training and explore various framework for hyper-parameter training. Finally, we will extend the simulation tool to different detectors, and collaborate on its integration into the existing simulation frameworks.

Publications

- F. Carminati et al., A Deep Learning tool for fast detector simulation. Poster presented at the 18th International Supercomputing Conference 2018, Frankfurt, 2018. http://*cern.ch/go/zcT8*
- G. Khattak, Training Generative Adversarial Models over Distributed Computing System (2018), revised selected papers. *cern.ch/go/8Ssz*

Presentations

- F. Carminati, Quantum Computing for High Energy Physics Applications (21 February). Presented at the PhD Course on Quantum Computing at University of Pavia, Pavia, 2019. *cern.ch/go/Z8Tx*
- F. Carminati, S. Vallecorsa, G. Khattak, 3D convolutional GAN for fast simulation (5 March). Presented at IXPUG Spring Conference, Bologna, 2018. *cern.ch/go/9TqS*
- F. Carminati, G. Khattak, S. Vallecorsa, Three-dimensional energy parametrized adversarial networks for electromagnetic shower simulation (7 October). Presented at 2018 IEEE International Conference on Image Processing, Athens, 2018. *cern.ch/go/lVr8*
- F. Carminati, V. Codreanu, G. Khattak, H. Pabst, D. Podareanu, V. Saletore, S. Vallecorsa, Fast Simulation with Generative Adversarial Networks (12 November). Presented at The International Conference for High Performance Computing, Networking, Storage, and Analysis, Dallas, 2018. *cern.ch/go/Z6Wg*
- F. Carminati, G. Khattak, D. Moise, S. Vallecorsa, Data-parallel Training of Generative Adversarial Networks on HPC Systems for HEP Simulations (18 December). Presented at 25th IEEE International Conference on High Performance Computing, Data, and Analytics, HiPC, Bengaluru, 2018.

NEXT GENERATION ARCHIVER FOR WINCC OA

R&D TOPIC 3: MACHINE LEARNING AND DATA ANALYTICS

SIEMENS

Project coordinator:
Fernando Varela

Technical team:
Filippo Tilaro, Jakub Guzik, Rafal Kulaga, Piotr Golonka, Peter Sollander

Siemens liaisons:
Thomas Hahn, Juergen Kazmeier, Alexey Fishkin, Tatiana Mangels, Mikhail Kalinkin, Elisabeth Bakany, Ewald Sperrer

Project goal

Our aim is to make control systems used for the LHC more efficient and "smarter". We are working to enhance the functionality of WinCC OA (a SCADA tool used widely at CERN) and to apply data-analytics techniques to the recorded monitoring data, in order to detect anomalies and systematic issues that may impact upon system operation and maintenance.

Background

The HL-LHC programme aims to increase the integrated luminosity — and hence the rate of particle collisions — by a factor of ten beyond the LHC's design value. Monitoring and control systems will therefore become increasingly complex, with unprecedented data throughputs. Consequently, it is vital to further improve the performance of these systems, and to make use of data-analytics algorithms to detect anomalies and anticipate future behaviour. Achieving this involves a number of related lines of work. This project focuses on the development of a modular and future-proof archiving system (NextGen Archiver) that supports different SQL and NOSQL technologies to enable data analytics. It is important that this can be scaled up to meet our requirements beyond 2020.

Progress in 2018

Significant progress was made on all components of the NextGen Archiver in 2018. The front end, the InfluxDB, and Oracle back ends now support almost all basic functionality planned for the initial release in 2019.

Reaching this level of functionality has made it possible to test the NextGen Archiver and deploy it in pilot systems at CERN. At the ProtoDUNE experiment, the InfluxDB back end is used to enable online access to archived data from Grafana, open platform for analytics and monitoring. At the ALICE experiment, members of the collaboration are developing a custom back end to stream control systems data to a new online-offline physics data readout system.

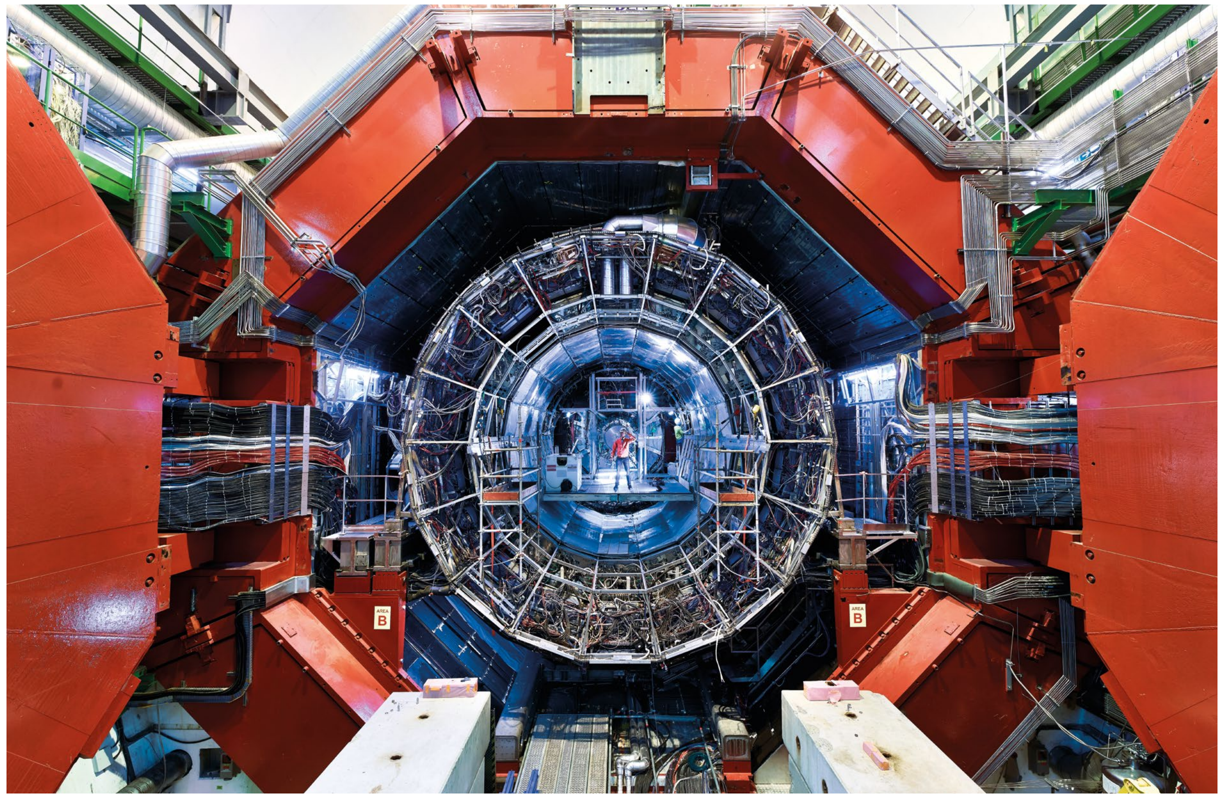
Several large-scale tests of the NextGen Archiver were performed at CERN in 2018 to stress-test the archiver and assess its scalability. These complemented the functional tests written at ETM, the Siemens-owned company behind the tool. The results of these tests have already influenced the query architecture of the archiver and helped us to improve its performance.

In 2018, tools for future benchmarking of Apache Kudu's read- and write-performance were developed, with the help of a CERN openlab summer student.

Next steps

Work on all components of the NextGen Archiver — as well as testing efforts — will continue in 2019. This year also brings important deadlines and deliverables for the project, including the release of version 1.0 and the start of tests in the ALICE infrastructure in May (to be put in production in the final quarter of the year).

The project team would also like to thank Enrique Blanco, Benjamin Bradu, and Jean-Charles Tournier for the expert consultancy they have provided.



At the ALICE experiment, members of the collaboration are developing a custom back end to stream control systems data to a new online-offline physics data readout system.

BIODYNAMO

R&D TOPIC 4: APPLICATIONS IN OTHER DISCIPLINES

Project coordinator:
Fons Rademakers

Technical team:
Lukas Breitwieser, Ahmad Hesam, Klaus-Dieter Oertel

Collaborator liaisons:
Roman Bauer (Newcastle University), Claudio Bellini (Intel)

Project goal

We are aiming to create a platform through which life scientists can easily create, run, and visualise three-dimensional biological simulations. Built on top of the latest computing technologies, the BioDynaMo platform will enable users to perform simulations of previously unachievable scale and complexity, making it possible to tackle challenging scientific research questions.

Background

Within the life-sciences community, computer simulation is being used more and more to model increasingly complex biological systems. Although many specialised software tools exist, establishing a high-performance, general-purpose platform would be a major step forward. CERN is therefore contributing its deep knowledge in large-scale computing to this collaboration with Newcastle University in the UK, supported by Intel. Together, we are working to develop a unique platform.

The team is aiming to create a platform through which life scientists can easily create, run, and visualise three-dimensional biological simulations.

Progress in 2018

In 2018, we added a neuroscience module that enables scientists to simulate the growth of dendrites — tree-like extensions of nerve cells — in 3D space. This made it possible for Jean de Montigny, a PhD student at Newcastle University, to use BioDynaMo to investigate retinal development. His research focuses on understanding the organisation of retinal ganglion cells, a type of neuron located near the inner surface of the retina. These cells are known for exhibiting semi-regular patterns of spatial arrangement, called retinal mosaics. Three main theories for retinal mosaic development exist. BioDynaMo is now being used to investigate these theories and the mechanisms that could influence the shape of these cells' dendritic trees.

Work to further increase performance also continued in 2018. Preliminary testing shows a performance increase of an order of magnitude, comparing single-thread performance of BioDynaMo against the benchmark software. BioDynaMo is fully parallelised and also optimised for non-uniform memory architectures. Thus, it scales well with the number of physical cores within one compute node, even for servers with multiple sockets. In addition, we also explored the possibilities of using heterogeneous computing resources (GPU and FPGA) to further reduce simulation time.

Next steps

Although the current simulation engine exploits the parallelism of modern hardware within a single compute node, the complexity of the biological model that can be simulated is limited. We will therefore develop a scale-out architecture to distribute computations across many compute nodes in a cloud infrastructure.

Publications

- A. Hesam, Faster than the Speed of Life: Accelerating Developmental Biology Simulations with GPUs and FPGAs (Master's thesis), Delft University of Technology, Netherlands, 2018. *cern.ch/go/f9v6*

Presentations

- A. Hesam, Biodynamo project status and plans (11 January). Presented at CERN openlab Technical Workshop, Geneva, 2018. *cern.ch/go/F8Cl*
- L. Breitwieser, BioDynaMo (1 February). Presented at University Hospital of Geneva Café de l'Innovation, Geneva, 2018.
- L. Breitwieser, The Anticipated Challenges of Running Biomedical Simulations in the Cloud (12 February). Presented at Early Career Researchers in Medical Applications @ CERN, Geneva, 2018. cern.ch/go/spc8
- N. Nguyen, Distributed BioDynaMo (16 August). Presented at CERN openlab summer students' lightning talks, Geneva, 2018.
- A. Hesam, Faster than the Speed of Life: Accelerating Developmental Biology Simulations with GPUs and FPGAs (31 August). Master's thesis defense, Delft, 2018. cern.ch/go/f9v6
- L. Breitwieser, The BioDynaMo Project: towards a platform for large-scale biological simulation (17 September). Presented at DIANA meeting, Geneva, 2018. *cern.ch/go/kJv7*

SMART PLATFORMS FOR SCIENCE

R&D TOPIC 4: APPLICATIONS IN OTHER DISCIPLINES

Project coordinator:
Alberto Di Meglio

Technical team:
Taghi Aliyev

King's College London liaison:
Mario Falchi

Project goal
We are developing a platform that will support a complete data-analysis life cycle, from data discovery through to access, processing, and end-user data analysis. The platform will be easy to use and will offer narrative interfaces.

As part of the development, we are working together with a number of teams at CERN on data integration and pipeline preservation. In particular, we are working closely with the teams behind REANA, a system for reusable analyses of research data, and Zenodo, open-access repository operated by CERN.

Background
In many research communities today, reproducibility, communication, and data pipelines are implemented in suboptimal ways. Through this project, we are working to create a powerful system to capture and facilitate the habits of researchers. Our platform will allow for negotiation and sharing of common values among scientists within a given field and will help us to understand the reasoning behind why certain choices are made. Rather than providing a simple toolkit for researchers, we are creating a rich system through which researchers can challenge the value chains within their own respective fields and potentially enhance their approach to performing research.

Progress in 2018
Throughout 2018, we gathered and worked on a range of initial use cases for the platform. Contacts were established with companies like IBM and non-profit organisations like GEnIAI (a local initiative working to enhance the lives of citizens in Geneva). As part of our collaboration with the two named organisations, we are now deploying solutions and ideas developed through the project to help tackle everyday challenges related to information retrieval and the answering of questions.

As part of the work with the GEnIAI community, we are working on the implementation of chat bots that could be used by members of the public in the Canton of Geneva, and are part of the Responsive City Camp Geneva initiative. This initiative has been endorsed by the Canton of Geneva, as well as by many organisations in the region. Initial ideas and results are to be presented at the Applied Machine Learning Days conference on 28 January 2019 in Lausanne, Switzerland.

Next steps
In the coming year, we will mainly work to assess the effectiveness of the prototypes and implemented models. Based on the obtained results, we will then work to improve the platform further, before deploying the first product to the wider research community.

The goal of this project is to develop a platform that will support a complete data-analysis life cycle.

Publications
- A. Manafli, T. Aliyev: Natural Language Processing for Science. Information Retrieval and Question Answering. Summer Student Report, 2018. *http://cern.ch/go/Z9l9*

Presentations
- T. Aliyev, Smart Data Analytics Platform for Science (1 November). Presented at i2b2 tranSMART Academic Users Group Meeting, Geneva, 2018.
- T. Aliyev, AI in Science and Healthcare: Known Unknowns and potential in Azerbaijan (December). Presented at Bakutel Azerbaijan Tech Talks Session, Baku, 2018.

06 KNOWLEDGE

Education, training, and outreach

CERN openlab is designed to create and share knowledge through a wide range of activities and programmes.

CERN openlab is a knowledge factory. We work to disseminate this knowledge through both outreach activities and educational programmes. As well as promoting our technical work among a variety of stakeholders, we are working to train the next generation of ICT specialists. Thus, CERN openlab provides a means for its collaboration members to share a joint vision of the future of scientific computing. This vision is communicated to a wide audience, including partner clients, policy makers, members of the press, and the general public. Together, we can shape the future of scientific computing for the benefit of both research and wider society.

Visits and workshops

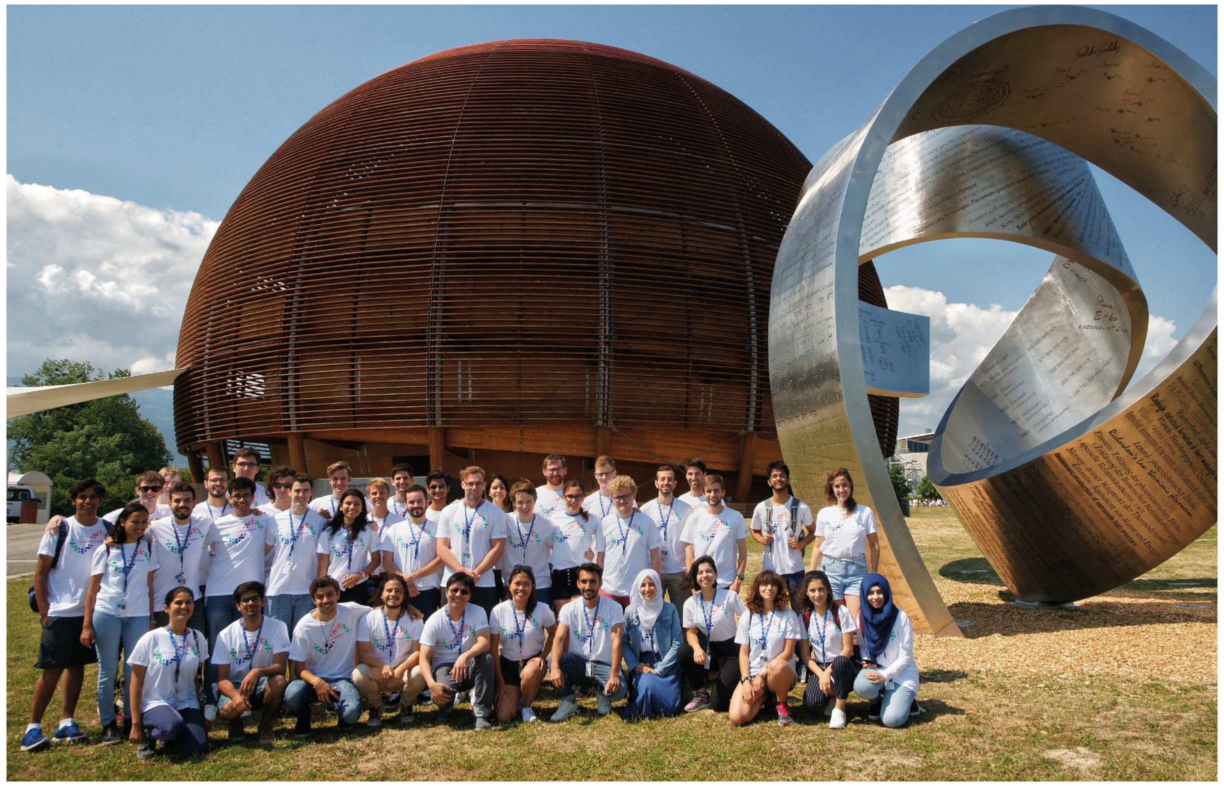
Top delegations from governments and industry frequently tour CERN: 120 protocol visits were organised in 2018. The CERN openlab concept and projects are systematically presented to the guests visiting the CERN IT department. CERN openlab partners have the opportunity to organise customer and press visits, too. Visiting groups are briefed about CERN openlab in a dedicated VIP meeting room, known as the CERN openlab "openspace". Over 150 press articles were published about our work over the course of the year.

Over 40 workshops, lectures, visits and other events related to CERN openlab's work were held throughout the year. Further information about these can be found on our website, including many recordings. On our website, you can also find a complete collection of press coverage, member case studies, press releases, videos, technical documents, presentations, and reports.

CERN openlab's work was presented at a number of high-profile industry events. One particular highlight was participating in the keynote presentation of Oracle CEO, Mark Hurd, at Oracle OpenWorld 2018; another was CERN openlab's CTO being invited to give the opening keynote at the prestigious ISC High Performance event in Germany.

New website

In 2018, CERN openlab launched a new website. Its structure has been adapted to reflect the primary areas of investigation in our new phase. There was also a major effort to streamline content and to ensure the site serves as an accessible gateway for people with no prior knowledge of our work. The site makes use of our new design package produced in 2017 and closely integrates with our popular social media channels. In 2018, we reached the milestone



Participants in CERN openlab's 2018 summer-student programme.

As in previous years, the team at Open Systems organised a packed and exciting tour in Zurich for the CERN openlab summer students.

of over 10 000 followers across Facebook and Twitter alone (including policy makers, journalists, and leading figures from both industry and research).

Summer-student fun

In 2018, 41 students — representing 22 different nationalities — took part in our annual summer-student programme. They each spent nine weeks at CERN, working on advanced computing projects with applications in high-energy physics and beyond. As part of the programme, the students attended a series of lectures given by IT experts on advanced CERN-related topics. Topics covered in 2018's programme included machine learning, computing security, grids and clouds, data acquisition, and code modernisation.

Within the CERN openlab summer-student programme, the students also visit CERN facilities and experiments, as well as other organisations. In 2018, the students went on a two-day trip to Zurich, where they visited IBM, ETH Zurich, and Open Systems.

Webfest innovation

Another highlight of the summer was the CERN Summer Student Webfest. The event is a hackathon, through which bright and creative minds meet over a weekend to build cool science projects using open web technologies. This year's Webfest, which was supported by CERN openlab, featured nine fascinating and original projects, related to topics such as data visualisation, robotics, education, and networking.

The event is open to all at CERN. However, this year's winning team consisted of five students participating in the CERN openlab summer-student programme. Their project, "CERN-Connect", is a web application that connects events and people across the CERN site using geolocalisation. The members of the team described the hackathon as providing an excellent opportunity to learn and develop new skills along with your friends.

Projects presented

The CERN openlab summer students had lots of fun — and learned much — during the trips, the Webfest, and the lecture series. However, the main focus of their time at CERN was undoubtedly their projects. These covered a diverse range of topics, including high-performance computing, big data, visualisation, machine learning, and much more.

The projects enabled the students to gain hands-on experience with some of the latest ICT solutions, working under the supervision of leading experts in the field.

On 14 and 16 August, the students presented their work in two dedicated public "lightning talk" sessions. In 5-minute presentations, each student explained the technical challenges they have faced and described the results of what they have been working on for the nine weeks they have spent at CERN.

The best presentations from each of the two sessions were selected by a panel of judges.

The winners of the first session were as follows:

- 1st: Nathan Jean C. Lacroix, Convolutional Neural Networks for Shelter Recognition in Satellite Images of Refugee Camps
- 2nd: Paul Samuel Maria Teuber, Efficient unpacking of required software from CernVM-FS
- 3rd: Aman Hussain, Deep Representation Learning for Trigger Monitoring

The winners of the second session were as follows:

- 1st: Xhesika Ruci, Wi-Fi Network Automation with Stackstorm
- 2nd: Rubab Zahra Sarfraz, Ceph Deployment with Rook
- 3rd: Sinclert Pablo Perez Castano, REANA: To reusability and beyond!

Nathan Lacroix, a student from ETH Zurich in Switzerland, was selected as this year's overall winner.

Maria Girone, CERN openlab CTO, giving the opening keynote presentation at the prestigious ISC High Performance conference in Frankfurt, Germany.

07
FUTURE

Next steps

More companies and research institutes are joining us to collaborate on tackling the challenges set out in our 2017 white paper. At the same time, we are beginning new explorations into a range of disruptive technologies.

2018 marked the start of CERN openlab's sixth three-year phase. Work was carried out on 17 different projects, each aiming to address the ICT challenges outlined in our 2017 white paper. Many of the projects were continuations or expansions of successful projects launched during our fifth three-year phase, while others were new.

A number of new companies also joined us, namely IBM, E4, Micron, and Google. IBM joined at the start of the year, meaning work is already well underway to investigate the potential of their Power architecture (see summary on pages 44 and 45 of this report). By contrast, the other three companies all joined in Q4 of 2018, meaning that work on joint projects begins in earnest in 2019. A summary of our planned work with these companies can be found below.

E4 Computer Engineering
E4 Computer Engineering has joined CERN openlab as a contributor, with a plan to collaborate on a one-year project that will establish a testbed for GPU-accelerated applications. In particular, the project will work to enable the direct programming of GPUs, as a way to accelerate calculations that underpin a number of key processes in the journey from particle collisions to new physics discoveries. At the same time, work will also be carried out to investigate the use of GPUs to train networks for machine-learning applications; such applications could, for instance, play an important role in helping experiments to cope with the increasing data rates that will be brought about by upgrades to the LHC. Already, ten specific use cases have been identified for investigation of GPU-accelerated applications. Early work will focus on support for planned upgrades to the LHCb experiment.

Micron Technology
Micron has joined CERN openlab as a partner, with plans to work together to explore a number of interesting new technologies over the next three years. Our first joint project focuses on testing Micron's advanced memory solutions as a way to potentially further machine-learning capabilities at the CMS experiment and for the ProtoDUNE detectors, prototypes for a major new international neutrino experiment to be built in the United States. Memory plays a vital role in processing, helping researchers gain valuable insights from data generated by the experiments.

Specifically, at CMS, we will work together to prototype a special kind of real-time inference engine for use in the experiment's data-filtering ("trigger") systems. For ProtoDUNE, the goal is to develop a new deep-learning-based approach — run on specially optimised hardware

Federico Carminati, CERN openlab's CIO, leading a panel discussion at the first CERN openlab workshop on quantum computing in high-energy physics.

Eckhard Elsen, Director for Research and Computing at CERN, speaking at the first CERN openlab workshop on quantum computing in high-energy physics.

provided by Micron — to speed up particle identification. For each of these two use cases, a detailed series of interim-steps and project milestones has been agreed for the first two years of the project's lifetime.

Google

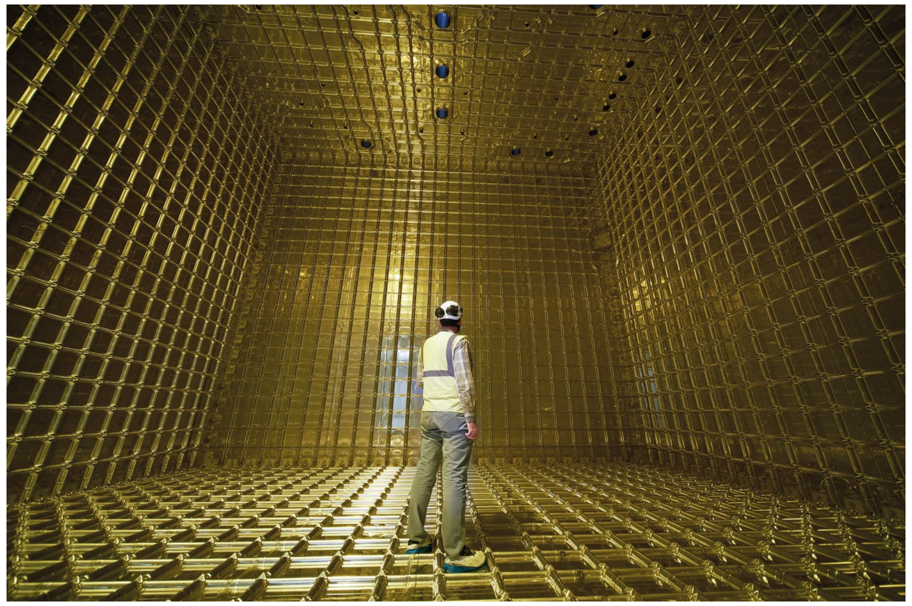
Google was the final company to join CERN openlab in 2018. Google signed an agreement with CERN, setting out the framework for our joint collaboration, in November. We are currently exploring possibilities for joint R&D projects related to cloud computing and machine learning, as well as in an important new area of investigation for CERN openlab: quantum computing.

Quantum computing

Over recent decades, the high-energy physics (HEP) community has acted as a driving force behind developments in a range of ICT areas, with the required computing power for planned projects often having outstripped what was foreseen to likely be available at the time of projects' conception. Today, significant developments are being made in the field of quantum computing, with both established computing vendors and start-up companies carrying out important activity in this field. It is therefore important to explore these technologies as one possible option for addressing future challenges.

The LHC's ambitious upgrade programme will result in significant ICT challenges over the next decade and beyond. It is therefore vital that we — the members of the HEP research community — continue to explore innovative new technologies, so as to ensure that we can continue to maximise the discovery potential of the world-leading research infrastructures at our disposal. Technologies related to quantum computing hold the promise of substantially speeding up computationally expensive tasks. We aim to assess the potential benefits of this technology and understand which activities within the HEP community are most well suited to their application.

Given both the potential and the challenges surrounding quantum computing, it is important to explore what these new technologies could bring to our field. To this end, CERN openlab organised a first-of-its kind workshop on quantum computing in high-energy physics in October 2018. Over 400 people followed the workshop, which provided an overview of the current state of quantum-computing technologies. The event also served as a forum to kick-off discussion of which activities within the HEP community may be amenable to the application of quantum-computing technologies.



Inside ProtoDUNE, a prototype for a major new international neutrino experiment that will be built in the United States.

The workshop brought members of the HEP community together with leading companies working on quantum computing technologies. Intel, IBM, Google, D-Wave, Microsoft, Rigetti and Strangeworks presented their latest work in this area at the event.

Discussions are currently being held with a number of companies working in this area, with a view to formalising joint R&D projects soon. These will help us to assess the future potential of quantum-computing technologies, their likely impact on computing models in HEP, and the feasibility of applying today's quantum-computing technologies to existing problems in HEP.

At the time of publishing, discussions with both Intel and Google were at a relatively advanced stage. Meanwhile, a specific project has been agreed with IBM. The agreement sees CERN join IBM's "Q Network'", with work beginning in 2019 to investigate how quantum machine-learning techniques could potentially be applied to classify particle collisions in the LHC experiments.

Development and disruption

Of course, quantum computing is just one of many avenues being explored as a way to address future ICT challenges, both at CERN and beyond. In 2019, we will expand our investigations into a number of emerging technologies that have the potential to disrupt key computing models used by the HEP community.

Nevertheless, the challenges set out in our 2017 white paper remain the core of our work, underpinning the majority of our growing number of joint projects with industry. Members of CERN openlab's management team are continuing to work very closely with representatives of experiments and departments across CERN to ensure we continue to address the latest, evolving ICT challenges faced by the laboratory's research community.

Through our ongoing dialogue with industry and research communities beyond HEP, we are able to ensure the challenges being addressed are of maximum relevance and can be of benefit to wider society. This is evidenced by the world-leading companies joining CERN openlab. We are looking forward to working with them — and the world-leading companies that are already established collaborators in CERN openlab — on a record number of joint R&D projects in 2019. We believe that these projects can make a real difference for the LHC research community, as well as for wider scientific computing.

ALBERTO DI MEGLIO
Head of CERN openlab
Alberto.Di.Meglio@cern.ch

MARIA GIRONE
Chief technology officer
Maria.Girone@cern.ch

FONS RADEMAKERS
Chief research officer
Fons.Rademakers@cern.ch

FEDERICO CARMINATI
Chief innovation officer
Federico.Carminati@cern.ch

ANDREW PURCELL
Communications officer
Andrew.Purcell@cern.ch

KRISTINA GUNNE
Administrative / finance officer
Kristina.Gunne@cern.ch

CERN openlab 2018 Annual Report

Editor: Andrew Purcell
Additional subediting: Hans Baechle

Design and production: CERN Design and Visual Identity Service

All images provided by the CERN Audio-Visual Production Service, except:
Christian Beutler (page 5); Jean-Claude Rifflard (page 31); Thomas Mc Cauley / CMS collaboration (page 43); Emma Ward (page 47); CERN openlab communications office (page 55, 56); Roberta Maggi (page 57)

An electronic version of this report is available on the CERN openlab website:
www.cern.ch/openlab/resources/annual-reports

ISBN 978-92-9083-534-9
Published by CERN

CERN openlab