ANNUAL REPORT

2019

01

INTRO

A word from the DG

Since its foundation in 2001, CERN openlab has been working to help accelerate the development of cutting-edge computing technologies. Such technologies play a vital role in particle physics, as well as in many other research fields, helping scientists to continue pushing back the frontiers of knowledge.

CERN openlab is a public-private partnership, through which CERN collaborates with leading ICT companies and other research institutions on R&D projects related to scientific computing. Today, a record 25 projects are underway, addressing topics such as machine learning, data analytics, big-data storage, and much more.

With the High-Luminosity Large Hadron Collider (HL-LHC) set to launch later this decade, it is crucial to ramp up work to address the associated ICT challenges (openlab.cern/whitepaper). Overall computing requirements are expected to increase by a large factor, and long-term data storage will go from the order of petabytes to exabytes. Furthermore, ICT challenges go far beyond those related to the handling of physics data from the experiments: there are also important challenges to be tackled related to accelerator control systems and the efficient running of our Organization.

Members of CERN's research community are already exploring how to adapt many important processes to new computing architectures. The computing models used in scientific research are also becoming more flexible, with high-throughput computing, high-performance computing, and cloud computing all playing important roles. One of the primary drivers behind this is the move towards heterogeneous computing architectures.

As well as making sure the right hardware and software are in place, it is vital to build skills and knowledge around the new technologies. CERN openlab plays an important role in this, capitalising on its deep connections with industry to organise workshops and other training sessions for CERN's community.

In particular, the CERN openlab summer-student programme plays an important role in training the ICT specialists of the future. In 2019, 40 students from 19 countries were selected from over 1600 applicants to the programme. It is always a great pleasure to welcome them at CERN and see them bring new ideas and fresh perspectives.

Fabiola Gianotti, Director-General of CERN.

Speaking of new ideas, I am also pleased to see the investigations that CERN openlab has begun — in collaboration with leading industries — into state-of-the-art technologies like machine learning, neuromorphic computing, and quantum computing. To take the last of these as an example, CERN openlab is now participating in five different projects related to quantum computing, involving companies like Google and IBM.

It is crucial to explore the potential of emerging new modes of computing such as these, so as to make the right decisions about the computing models we will employ as our field and related ICT challenges evolve over this new decade. This work also helps us to be ready to adopt these technologies — by having both the right software and people with the right skills in place — when they do reach maturity. CERN openlab, with its strong network of experts from academia and industry, has an important role to play in this domain.

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.

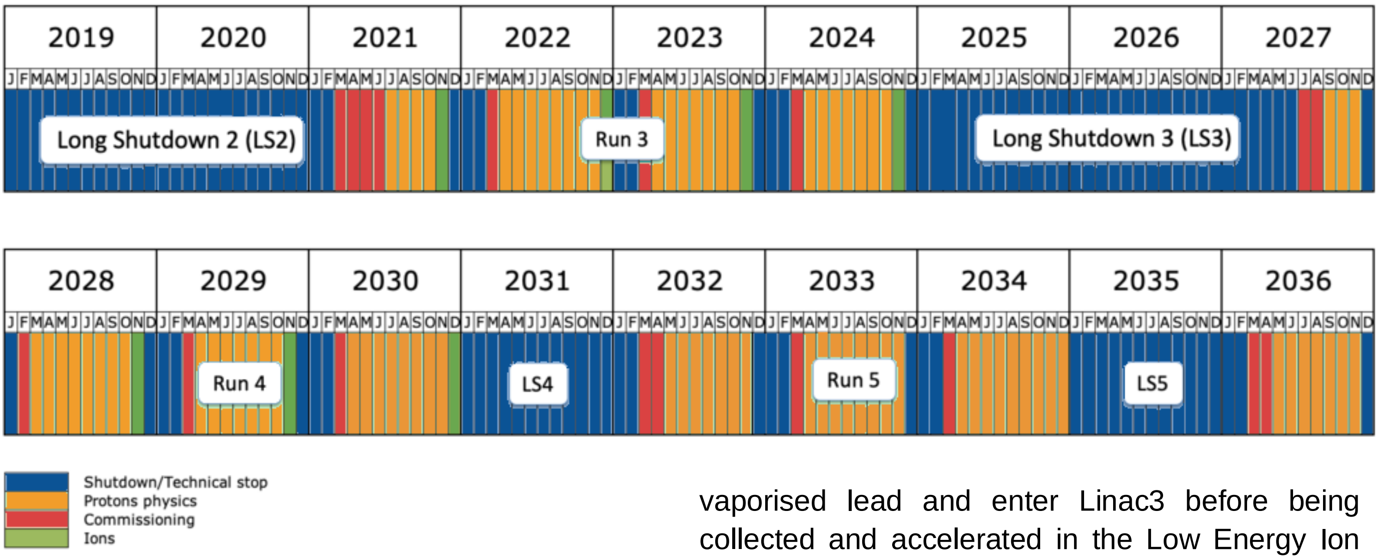
CERN is home to the Large Hadron Collider (LHC), the world's largest and most powerful particle accelerator. It consists of a 27-kilometre ring of superconducting magnets, with a number of accelerating structures to boost the energy of the particles along the way.

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 (when the LHC comes online again in 2021, this will be replaced by a new, more powerful accelerator called Linac4). 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

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



The planned upgrade schedule for the LHC.

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 protons to reach their maximum energy of 6.5 TeV. The 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 Linac3 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 High-Luminosity LHC upgrade

The LHC has been designed to follow a carefully set out programme of upgrades. The LHC typically produces particle collisions for a period of around three years (known as a ‘run’), followed by a period of about two years for upgrade and maintenance work (known as a ‘long shutdown’, which is commonly abbreviated to LS). We are currently in the second such shutdown period, known simply as LS2, with the LHC set to come online again for Run 3 in mid-2021.

Further upgrades over the coming years will lead to the launch of the High-Luminosity LHC (HL-LHC) in around 2027. The HL-LHC project aims to crank up the performance of the LHC, in order to increase the potential for discoveries starting in Run 4. The objective is to increase the luminosity by a factor of ten beyond the LHC’s design value. Luminosity is an important indicator of the performance of an accelerator: it is proportional to the number of collisions that occur in a given amount of time. The higher the luminosity, the more data the experiments can gather. This enables physicists to observe rare processes and study new particles with greater precision.

The HL-LHC is likely to require computing and storage capacity 2-3 times greater than today.

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 16,000 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. Around 340 petabytes of data are stored on tape at the Meyrin site, with roughly 230 petabytes of data stored on disks.

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 over 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 forms "Tier-0" of the WLCG, the first point of contact between experimental data from the LHC and the grid.

03

HIGHLIGHTS

2019 at CERN

Throughout 2019, work was carried out to upgrade the particle accelerators and experiments at CERN – as well as the engineering and ICT systems that support them – in time to start of the LHC's third run, beginning in mid-2021.

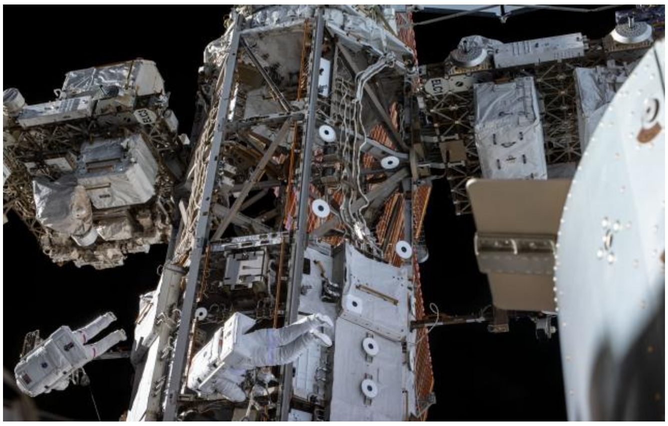
2019 was the first full year of the LHC's second 'long shutdown' period, known as 'LS2'. During this time, maintenance and upgrades are being carried out the entire accelerator chain for the LHC.

Linac4, a new linear accelerator to replace the retired Linac2 accelerator, was tested and connected to the accelerator chain. The Proton Synchrotron (PS), which celebrated its 60th anniversary in 2019, was also rejuvenated — as was the PS Booster. In addition, the Super Proton Synchrotron (SPS) was fitted with a new acceleration system and a new beam dump.

Throughout 2019, diggers also worked underground to excavate new structures for the HL-LHC. These new caverns, which will house equipment that is needed to achieve high luminosity, were joined to the existing LHC tunnel structures on 13 December.

However, it wasn't just the accelerators that were upgraded; important work was carried out at the experiments, too. For example, the heart of the ALICE detector was replaced, and both the ATLAS and CMS experiments are taking the opportunity of LS2 to install new muon detectors.

Of course, there were also theoretical breakthroughs from the experiments in 2019. ATLAS and CMS measured new properties of the Higgs boson; ALICE measured the elliptic flow of the bottomonium particle, shedding new light on the early universe; and LHCb saw a new flavour of matter-antimatter asymmetry, as well as observing new pentaquarks.



In 2019, a series of spacewalks was conducted to install a new cooling system for the Alpha Magnetic Spectrometer (AMS). The AMS detector is installed on the side of the International Space Station, and the collaboration is headquartered at CERN.

CERN also has a diverse and exciting experimental programme beyond the LHC. Below are a few highlights — both theoretical and technical — from 2019:

- The NA62 experiment spotted two potential rare particle decays, putting the standard model to the test.
- The NA64 experiment cast light on dark photons, a hypothetical dark-matter particle.
- The BASE experiment proposed a new way to probe dark matter using antimatter.
- The AEgIS experiment found a new way to make long-lived positronium for antimatter gravity experiments.
- The CLOUD experiment used natural cosmic rays to examine their direct effect on clouds.
- The ProtoDUNE experiment tested a new technology to detect neutrinos.

Another highlight of 2019 was the successful testing of a first-of-its-kind superconducting line to transport electricity. The plan is to use this technology for the HL-LHC, but it could also perhaps be used in future to transport electricity more efficiently to towns and cities.

Finally, 2019 was also an important year for CERN in terms of culture. As well as marking 30 years since the birth of the World Wide Web at the laboratory, 2019 also saw CERN hold its first major Open Days event in six years. Across a weekend in September, 75, 000 people came to visit the laboratory. The registration system for the visitors was developed through a CERN openlab project with Oracle, and CERN openlab was presented to visitors that came to the CERN IT Department.

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: Fons Rademakers, chief research officer; Miguel Marquina, educational programmes advisor; Kristina Gunne, chief finance and administration officer; Alberto Di Meglio, head of CERN openlab; Maria Girone, chief technology officer; Andrew Purcell, chief communications officer; Federico Carminati, chief innovation officer; Hans Baechle, junior communications 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 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 second 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 (openlab.cern/whitepaper), published at the end of 2017. Challenges related to data-centre technologies and infrastructures, computing performance and software, machine learning and data analytics, and quantum technologies 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.

Participants in the CERN openlab Technical Workshop 2020.

ORACLE

intel

SIEMENS

Micron

Google

be studys
Research and innovation

IBM

E4
COMPUTER ENGINEERING

CQC
Cambridge Quantum Computing

COMTRADE

open systems

INFN
Istituto Nazionale di Fisica Nucleare

Scimpulse foundation

KING'S College LONDON

ODTÜ METU

gluoNNet
knowledge exchange for smart decisions

TU/e Technische Universiteit Eindhoven University of Technology

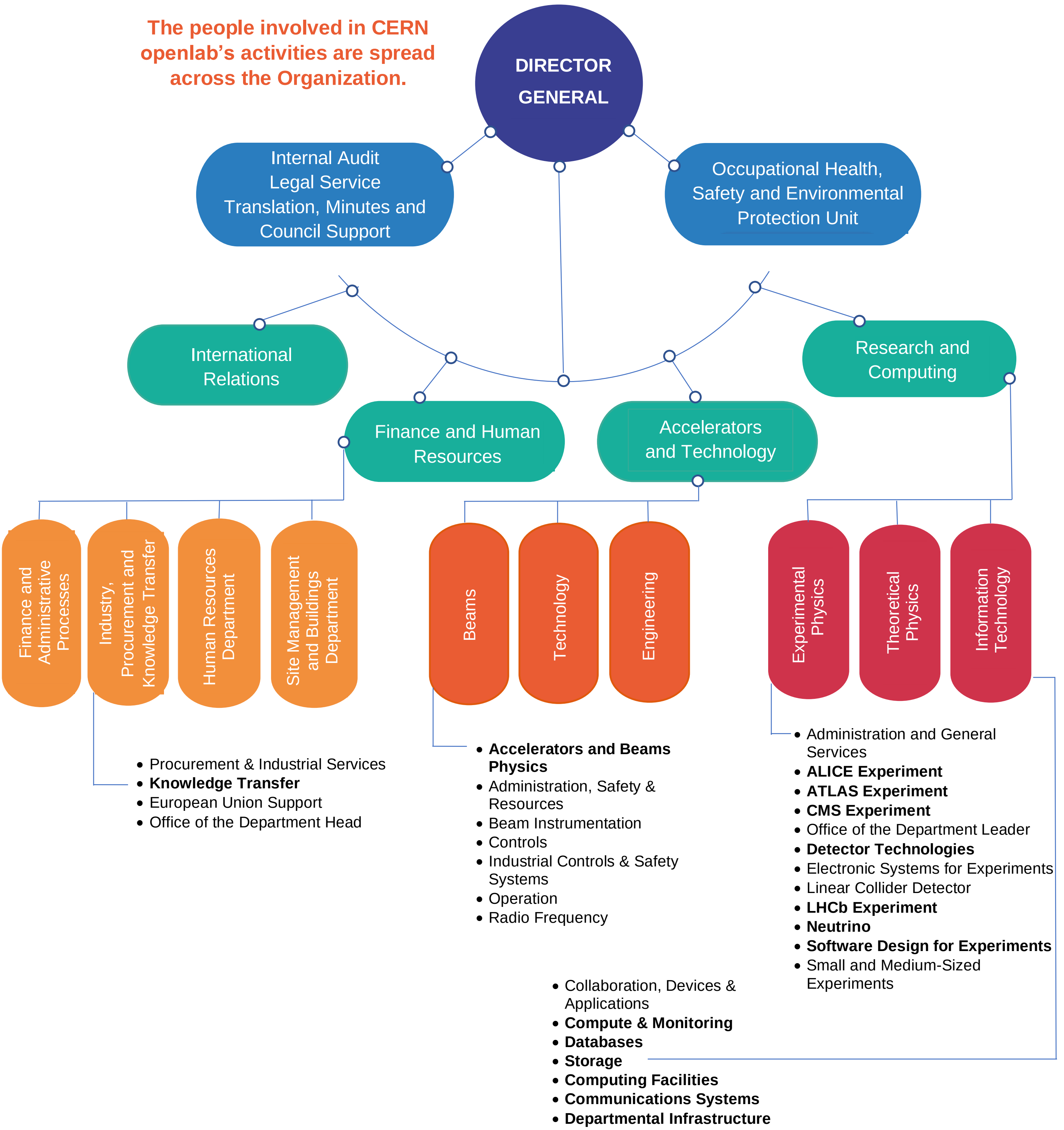
SAMARA UNIVERSITY

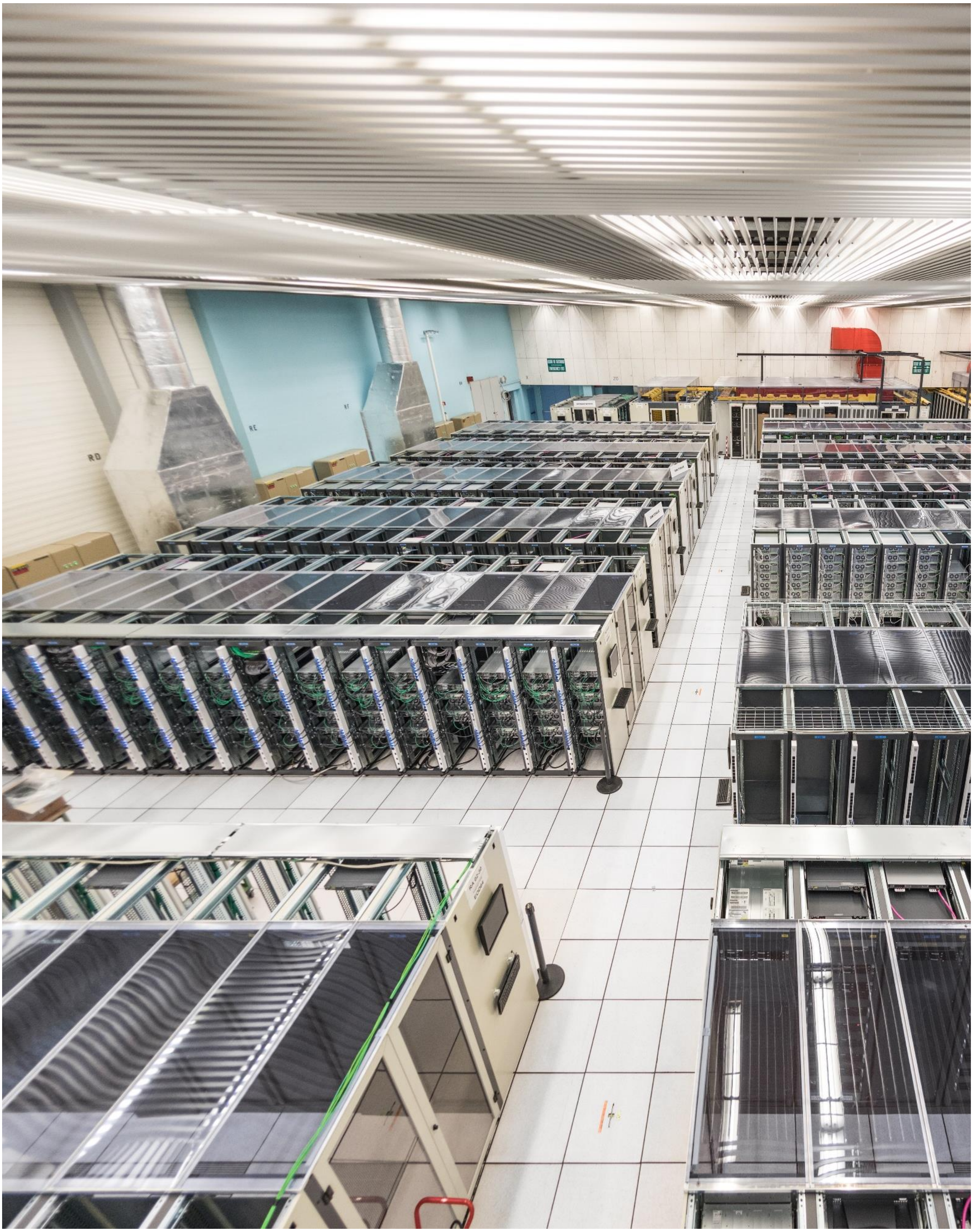
Fermilab

Newcastle University

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 25 technical projects that ran in 2019 can be found in this section. These projects are organised into five 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. It is also important to ensure that software fully exploits the features offered by modern hardware architectures, such as many-core GPU platforms, acceleration coprocessors, and hybrid combinations of CPUs and FPGAs. At the same time, it is paramount that physics performance is not compromised in drives to maximise 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:
Quantum technologies

The HL-LHC is likely to require computing and storage capacity 2-3 times greater than today. Even corrected for the evolution of technology, this will result in a shortage of computing resources. Keeping with its mandate, CERN openlab is exploring new and innovative solutions to help physicists to bridge this resource gap which may put in danger the achievement of the goals of the HL-LHC experimental programme.

Following a successful workshop on quantum computing held at CERN in 2018, CERN openlab has started a number of projects in quantum computing that are at different stages of realisation.

R&D topic 5:
Applications in other disciplines

The fifth 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 high-energy physics brings together all these aspects.

DEVELOPING A TICKET RESERVATION SYSTEM FOR THE CERN OPEN DAYS 2019

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

Project coordinator:
Artur Wiecek

Technical team:
Antonio Nappi, Franck Pachot, Luis Rodriguez Fernandez, Nuno Guilherme, Matos De Barros, Thomas Løkkeborg, Viktor Kozlovszky

Collaborator liaisons:
Pauline Gillet-Mahrer, Christian Jacobsen, Eric Mitha, Cristobal Pedregal-Martin, David Ebert, Dmitrij Dolgušin

Project goal

The goal of this project was to deliver a web-based ticketing system, with zero downtime, for the CERN Open Days 2019. The system had to be capable of handling multiple ticket-release dates and high loads, as well as providing a user-friendly and responsive interface.

Background

With the LHC stopped for upgrade and maintenance during 2019, CERN took the opportunity to organise one of its major Open Days events in September. At these events, members of the public are able to visit many areas that are normally restricted. Our team's responsibility was to design, deliver, and maintain the ticketing system for the Open Days. The event took place over a full weekend, and was attended by 75,000 people.

Progress in 2019

The project was delivered in 75 days and was maintained for three months from mid-June. The system was made available in English and French, and it had a multi-layer architecture. Thanks to the cutting-edge technologies used, adherence to coding standards, and the fine-tuning possibilities offered by the various layers, we were able to successfully deliver a highly available and scalable system.

We used an Angular framework, with Angular material design on the front-end application layer and a REST API based on the Java Spring boot application as the back-end layer. These were run in containers, hosted in a Kubernetes environment.

Each of the components were individually scalable based on the load, as was the Oracle Autonomous Transaction Processing database. All the components were hosted on the Oracle cloud, using cloud-native services like a load balancer, a container engine for Kubernetes, and Oracle Analytics Cloud. The system was stress tested and was capable of handling 20,000 parallel users across a period of six minutes.

The reservation system played an important role in the organisation of this large event.

Next steps

The project came to a successful close on 16 September 2019. As part of the project, we tested and evaluated the capabilities of the Oracle Cloud, which will serve as the foundation for a planned upcoming project on disaster recovery.

The project members would also like to thank the following people at CERN for their assistance:

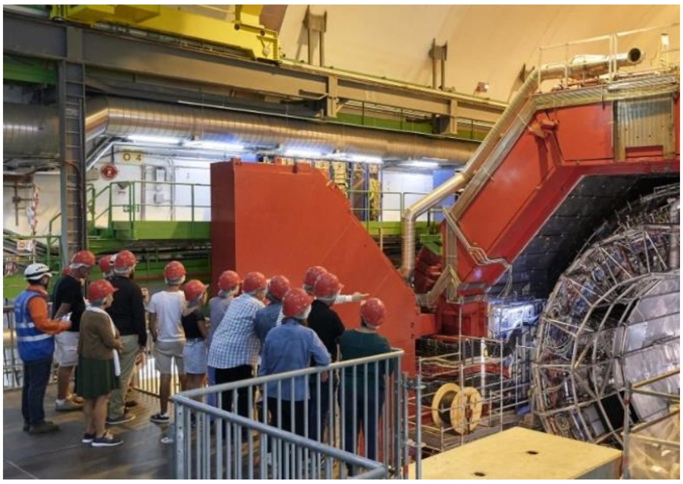
- *Sébastien Masson, Manuel Martin Marquez, Aimilios Tsouvelekakis and Aparicio Cotarelo for their consultation services.*
- *Francois Briard, the project owner.*
- *Tobias Betz, for his assistance with the mail templates and the countdown timer.*
- *Liviu Valsan, Sebastian Lopienski, and Jose Carlos Luna Duran for their assistance with security.*
- *Bas Wallet and Ewa Lopienska for their work on the user interface and experience.*
- *Ana Godinho, Odile Martin, Achintya Rao, and Corinne Pralavorio for their translation work.*

Publications

- V. Kozlovszky, Open Days reservation system's high level overview – 2019. Databases at CERN blog. 2019. *cern.ch/go/ff6k*
- V. Kozlovszky, Internationalization of the 2019 Open Days reservation system. Databases at CERN blog. 2019. *cern.ch/go/6qQB*

Presentations

- E. Grancher, V. Kozlovszky, 100% Oracle Cloud: Registering 90,000 People for CERN Open Days (16 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/6xZZ*
- E. Grancher, 11 Months with Oracle Autonomous Transaction Processing (18 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/8PQ6*



Visitors taking a look at the ALICE detector during the CERN Open Days 2019.

DYNAMICAL EXASCALE ENTRY PLATFORM – EXTREME SCALE TECHNOLOGIES (DEEP-EST)

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

Project coordinator:
Maria Girone (for DEEP-EST task 1.7)

Technical team:
Viktor Khristenko (for DEEP-EST task 1.7)

Project goal

The main focus of the project is to build a new kind of a system that makes it possible to run efficiently a wide range of applications — with differing requirements — on new types of high-performance computing (HPC) resources. From machine-learning applications to traditional HPC applications, the goal is to build an environment which is capable of accommodating workloads that pose completely different challenges for the system.

Background

DEEP-EST is a project funded by the European Commission that launched in 2017, following on from the successful DEEP and DEEP-ER projects. The project involves 27 partners in more than 10 countries and is coordinated from Jülich Supercomputing Centre at Forschungszentrum Jülich in Germany.

Overall, the goal is to create a modular supercomputer that best fits the requirements of diverse, increasingly complex, and newly emerging applications. The innovative modular supercomputer architecture creates a unique HPC system by coupling various compute modules according to the building-block principle: each module is tailored to the needs of a specific group of applications, with all modules together behaving as a single machine.

Specifically, the prototype consists of three compute modules: the cluster module (CM), the extreme scale booster (ESB), and the data-analytics module (DAM).

- Applications requiring high single-thread performance are targeted to run on the CM nodes, where Intel® Xeon® Scalable Processor (Skylake) provide general-purpose performance and energy efficiency.
- The architecture of ESB nodes was tailored for highly scalable HPC software stacks capable of extracting the enormous parallelism provided by Nvidia V100 GPUs.
- Large memory capacities using Intel® Optane™ DC persistent memory and different acceleration capabilities (provided by Intel Stratix 10 and Nvidia V100 GPU on each node) are key features of the DAM; they make it an ideal platform for data-intensive and machine-learning applications.

CERN, in particular the CMS Experiment, participates by providing one of the applications that are used to evaluate this new supercomputing architecture.

Progress in 2019

We ported the software used at CMS for the reconstruction of particle collision events for hadronic and electromagnetic calorimeters. We then optimised these workloads for running on Nvidia V100 GPUs, comparing performance against the CPU-based systems currently used in production.

Next steps

We will incorporate MPI offloading into the 'CMSSW' software framework used at the CMS experiment, in order to be able to run different parts of the reconstruction on different hardware. We will also explore the use of FPGAs for reconstruction at CMS. Finally, we will explore the possibility of capitalising on new and faster memory interconnection between CPUs and accelerators, using Intel® OneAPI for development and debugging.

Publications

The following deliverables were submitted to the European Commission:
Deliverable 1.2: Application Use Cases and Traces
Deliverable 1.3: Application Distribution Strategy
Deliverable 1.4: Initial Application Ports

Presentations

- V. Khristenko, CMS Ecal Reconstruction with GPUs (23 October). Presented at CMS ECAL DPG Meeting, Geneva, 2019. *cern.ch/go/FC6j*
- V. Khristenko, CMS Hcal Reconstruction with GPUs (8 November). Presented at CMS HCAL DPG Meeting, Geneva 2019. *cern.ch/go/P6Js*
- V. Khristenko, Exploiting Modular HPC in the context of DEEP-EST and ATTRACT projects (22 January). Presented at CERN openlab Technical Workshop, Geneva, 2020. *cern.ch/go/rjC7*

The DEEP-EST prototype.

ORACLE MANAGEMENT CLOUD

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

ORACLE

Project coordinator:
Eva Dafonte Perez, Eric Grancher

Technical team:
Aimilios Tsouvelekakis

Collaborator liaisons:
Simone Indelicato, Vincent Leocorbo, Cristobal Pedregal-Martin, David Ebert, Dmitrij Dolgušin

Project goal

We are testing Oracle Management Cloud (OMC) and providing feedback to Oracle, including proposals for the evolution of the platform. 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 uses and provides specialised monitoring solutions to teams across the laboratory that use database infrastructure. Since the beginning of 2018, we have an agreement in place with Oracle to test OMC, which offers a wide variety of monitoring solutions.

At CERN, as at other large organisations, it is very important to be able to monitor at all times what is happening with systems and applications running both locally and in the cloud. Thus, we conducted tests of OMC involving hosts, application servers, databases, and Java applications.

Progress in 2019

Improvements proposed to Oracle during the previous year were implemented within new releases of the platform in 2019. Initial investigation shows that the platform has been enhanced with features covering most of our needs.

Furthermore, we deployed the registration application for the CERN Open Days in Oracle

Cloud Infrastructure (see project ‘Developing a ticket reservation system for the CERN Open Days 2019’ for further details). The application made use of the Oracle Autonomous Transaction Processing (ATP) database to store visitor information. The behaviour of the ATP database was monitored using OMC, providing meaningful insights into the stresses put on the database and the database-hosting system during the period in which registration for the event was open.

Next steps

The next step for OMC is to use it as a monitoring platform for all the projects that are to be deployed in the Oracle Cloud Infrastructure.

Presentations

- A. Tsouvelekakis, Oracle Management Cloud: A unified monitoring platform (23 January). Presented at CERN openlab Technical Workshop, Geneva, 2019. *cern.ch/go/Z7j9*
- A. Tsouvelekakis, Enterprise Manager and Management Cloud CAB (April). Presented at Oracle Customer Advisory Board, Redwood Shores, 2019. *cern.ch/go/tZD8*
- A. Tsouvelekakis, CERN: Monitoring Infrastructure with Oracle Management Cloud (September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/mMd8*

At CERN, as at other large organisations, it is very important to be able to monitor at all times what is happening with systems and applications running both locally and in the cloud.

HETEROGENEOUS I/O FOR SCALE

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

Project coordinator:
Maria Girone, Viktor Khristenko

Collaborator liaisons:
Ulrich Bruening (University of Heidelberg), Mondrian Nuessle (Extoll GmbH)

Project goal

We are working to develop a proof of concept for an FPGA-based I/O intermediary. The potential impact would be to change the way data ingestion happens when using remote storage locations. In view of the enormous amounts of data to be employed in the future for data analytics, it is crucial to efficiently manage the flow in order to harness the computational power provided by high-performance computing (HPC) facilities.

Background

One of the common aspects of all data-intensive applications is the streaming of recorded data from remote storage locations. This often imposes constraints on the network and forces a compute node to introduce complex logic to perform aggressive caching in order to remove latency. Moreover, this substantially increases the memory footprint of the running application on a compute node. This project, abbreviated to 'HIOS', aims to provide a scalable solution for such data-intensive workloads by introducing heterogeneous I/O units directly on the compute clusters. This makes it possible to offload the aggressive caching functionality onto these heterogeneous units. By removing this complicated logic from compute nodes, the memory footprint decreases for data-intensive applications. Furthermore, the project will investigate the possibility to include additional logic, coding/decoding, serialisation, I/O specifics, directly onto such units.

An integral part of the project will be the ability to integrate developed units directly with current HPC facilities. One of the main outcomes of the project will be the reduced time required to extract insights from large quantities of acquired information, which, in turn, directly impacts society and scientific discoveries.

HIOS is one of the 170 breakthrough projects receiving funding through the ATTRACT initiative. ATTRACT, which is part of the European Union's Horizon 2020 programme, is financing breakthrough ideas in the fields of detection and imaging.

Progress in 2019

Development work began in September 2019. The first task was to provide the full specification and initial implementation that will become the proof of concept.

Next steps

In 2020, work will be carried out to complete the proof of concept and to submit the deliverables defined within the ATTRACT project.

Maria Girone and Viktor Khristenko, CERN members of the project team, pictured in the CERN data centre.

ORACLE WEBLOGIC ON KUBERNETES

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

Project coordinator:
Artur Wiecek

Technical team:
Antonio Nappi, Lukas Gedvilas, Luis Rodriguez Fernandez, Theodoros Rizopoulos, Aimilios Tsouvelekakis

Collaborator liaisons:
Monica Riccelli, Will Lyons, Maciej Gruzka, Cristobal Pedregal-Martin, David Ebert, Dmitrij Dolgušin

Project goal

This project is working to improve our deployment of the Oracle WebLogic infrastructure, in order to make it portable, repeatable, and faster. This will help us to be more efficient in our daily work.

Background

The Oracle WebLogic service has been active at CERN for many years, offering a very stable way to run applications core to the laboratory. However, we would like to reduce the amount of time we spend on maintenance tasks or creating new environments for our users. We therefore started to explore solutions that could help us to improve the way in which we deploy Oracle WebLogic. Kubernetes has now made our deployment much faster, reducing the time spent on operational tasks and enabling us to focus more on developers' needs.

Progress in 2019

Progress was made in a range of areas in 2019. We first reorganised the structure of our repositories, in order to better organise our work and to split Docker images into smaller ones — in line with trends towards microservices. We have defined and unified the way we were managing secrets in Kubernetes and in virtual machines. We also improved the way our users interact with the new infrastructure by providing them with a new way to deploy applications via REST API, using standard technologies like Rundeck.

In addition, we sped up and simplified the configuration of WebLogic using an open-source tool provided by Oracle. This has enabled us to build the WebLogic domain when a container was starting, instead of storing it in Docker images.

Next steps

The project itself is now relatively stable, with much testing already done. We are in the migration phase, with many development-and-testing environments already moved to Kubernetes. Our main goal is now to migrate the remaining such environments in 2020, as well as the production environment. We would also like to integrate new standard technologies, like Prometheus and Fluentd, into our systems for monitoring and logging.

Presentations

- A. Nappi, WebLogic on Kubernetes at CERN (May 16). Presented at WebLogic Server Summit, Rome, 2019.
- A. Nappi, One Tool to Rule Them All: How CERN Runs Application Servers on Kubernetes (16 September) Presented at Oracle Code One 2019, San Francisco, 2019. *cern.ch/go/DbG9*
- D. Ebert (Oracle), M. Martin, A. Nappi, Advancing research with Oracle Cloud (18 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/LH6Z*
- E. Screven, A. Nappi, Cloud Platform and Middleware Strategy and Roadmap (17 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/d8PC*
- M. Riccelli, A. Nappi, Kubernetes: The Glue Between Oracle Cloud and CERN Private Cloud (17 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/Bp8w*

Members of the CERN IT department's group for database services. Each of those pictured are working on at least one of the five current CERN openlab projects with Oracle.

EOS PRODUCTISATION

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

COMTRADE

Project coordinator:
Luca Mascetti

Technical team:
Elvin Sindrilaru

Collaborator liaisons:
Gregor Molan, Branko Blagojević, Ivan Arizanović, Svetlana Milenković

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, the project aims to add native 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 the 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 2019

Comtrade's team continued to acquire further knowledge of EOS, profiting from their visit to CERN and from working side-by-side with members of the development and operations teams. This helped them to improve their work on EOS installation, documentation, and testing.

In particular, a dedicated document describing best practices for operating EOS in large-scale environments was produced, as well as a full-stack virtual environment hosted at Comtrade. This shows the potential of the system when used as a geographically distributed storage system.

Next steps

The project will focus on improving and updating the EOS technical documents, for future administrators and operators. The next main goal is to host dedicated hardware resources at CERN to support prototyping of an EOS-based appliance. This will enable Comtrade to create a first version of a full storage solution and to offer it to potential customers in the future.

In addition, the team will investigate the possibility of developing a native Windows client for EOS.

Presentations

- L. Mascetti, EOS Productisation (23 January). Presented at CERN openlab Technical workshop, 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*

Comtrade's Gregor Molan presents his company's proposals at CERN.

KUBERNETES AND GOOGLE CLOUD

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

Project coordinator:
Ricardo Manuel Brito da Rocha

Collaborator liaisons:
Karan Bhatia, Andrea Nardone, Mark Mims, Kevin Kissell

Project goal

The aim of this project is to demonstrate the scalability and performance of Kubernetes and the Google Cloud, validating this set-up for future computing models. As an example, we are using the famous Higgs analysis that led to the 2013 Nobel Prize in Physics, thus also showing that real analysis can be done using CERN Open Data.

Background

As we look to improve the computing models we use in high-energy physics (HEP), this project serves to demonstrate the potential of open and well established APIs, such as Kubernetes. They open up a wide range of new possibilities in terms of how we deploy our workloads.

Based on a challenging and famous use case, we're working to demonstrate that these new tools — together with the virtually unlimited capacity offered by public cloud providers — make it possible to rethink how analysis workloads can be scheduled and distributed. This could lead to further improvements in the efficiency of our systems at CERN.

The project also provides an excellent opportunity to show how, given enough resources, anyone can replicate important physics analysis work using the open data published by CERN and the LHC experiments.

Progress in 2019

The initial goal of the project has been fulfilled: we have demonstrated that Kubernetes and the Google Cloud is a viable and extremely performant set-up for running HEP analysis work. The code required, as well as the set-up, is fully documented and publicly available (see publications).

The outcome of the project was presented at a number of high-profile conferences, including a keynote presentation at KubeCon Europe 2019, an event attended by over 8000 people. A live demo of the whole set-up, using data from the CERN Open Data Portal, was shown on stage.

The set-up — as well as the data set used — has been prepared for publication as a Google Cloud official tutorial. This will enable anyone to trigger a similar execution using their own public cloud resources. This tutorial will be published in early 2020, once the text has been finalised.

Next steps

This project was initially self-contained, with a clear target for the presentation at KubeCon Europe 2019. However, the project has now grown beyond this initial, limited scope. Future steps should include:

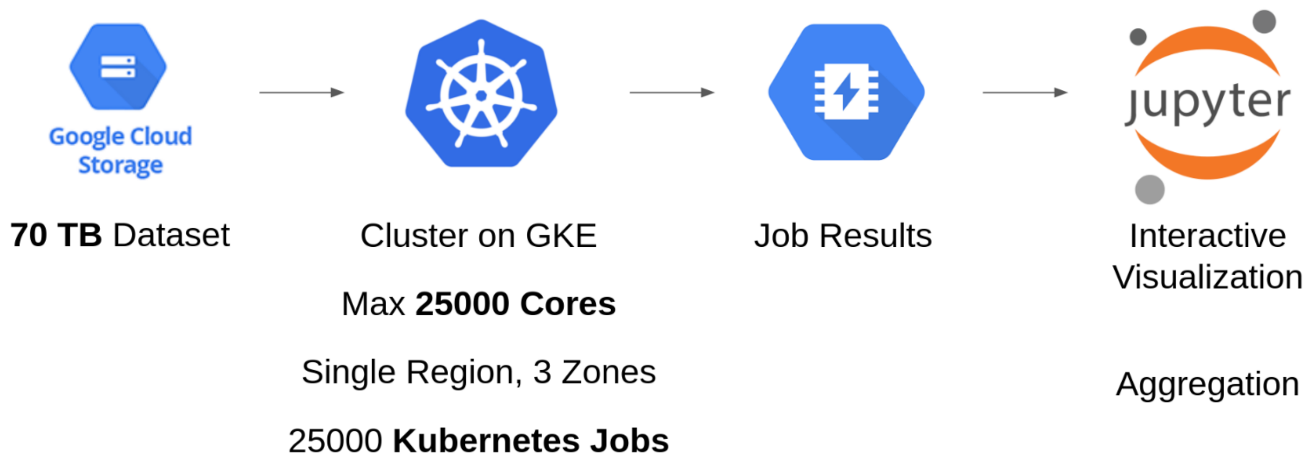
- Further investigating how using public cloud can improve physics analysis.
- Working to provide on-demand, bursting to public cloud capabilities for our on-premise resources.
- Seeking to understand how we can best define policies and accounting procedures for using public cloud resources in this manner.

Publications

- R. Rocha, L. Heinrich, Higgs-demo. Project published on GitHub. 2019. *cern.ch/go/T8QQ*

Presentations

- R. Rocha, L. Heinrich, Reperforming a Nobel Prize Discovery on Kubernetes (21 May). Presented at Kubecon Europe 2019, Barcelona, 2019. *cern.ch/go/PlC8*
- R. Rocha, L. Heinrich, Higgs Analysis on Kubernetes using GCP (19 September). Presented at Google Cloud Summit, Munich, 2019. *cern.ch/go/Dj8f*
- R. Rocha, L. Heinrich, Reperforming a Nobel Prize Discovery on Kubernetes (7 November). Presented at the 4th International Conference on Computing in High-Energy and Nuclear Physics (CHEP), Adelaide, 2019. *cern.ch/go/6Htg*
- R. Rocha, L. Heinrich, Deep Dive into the Kubecon Higgs Analysis Demo (5 July). Presented at CERN IT Technical Forum, Geneva, 2019. *cern.ch/go/6zls*



Higgs boson analysis set-up on the Google Cloud.

HIGH-PERFORMANCE DISTRIBUTED CACHING TECHNOLOGIES

R&D TOPIC 2: COMPUTING PERFORMANCE AND SOFTWARE

Project coordinator:
Giovanna Lehmann Miotto

Technical team:
Adam Abed Abud, Danilo Cicalese, Fabrice Le Goff, Remigius K Mommsen

Collaborator liaisons:
Claudio Bellini, Aleksandra Jereczek, Grzegorz Jereczek, Jan Lisowiec, Maciej Maciejewski, Adrian Pielech, Jakub Radtke, Jakub Schmiegel, Malgorzata Szychowska, Norbert Szulc, Andrea Luiselli

Project goal

We are 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 the 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 2019

We were able to test the first Intel® Optane™ persistent memory devices, enabling us to benchmark the behaviour of DAQDB on this new type of hardware. A testbed with four very powerful machines was set up, hosting Optane persistent memory and SSDs, and results are encouraging. However, more work is needed to reach the performance and scalability goals required for the next generation of High-Luminosity LHC experiments (in particular ATLAS and CMS), as well as by the DUNE experiment.

Next steps

The project formally came to a close in 2019, but several developments and tests will continue in 2020. This will enable us to continue exploring how the new storage technologies, and DAQDB, can be effectively used in data-acquisition systems.

Publications

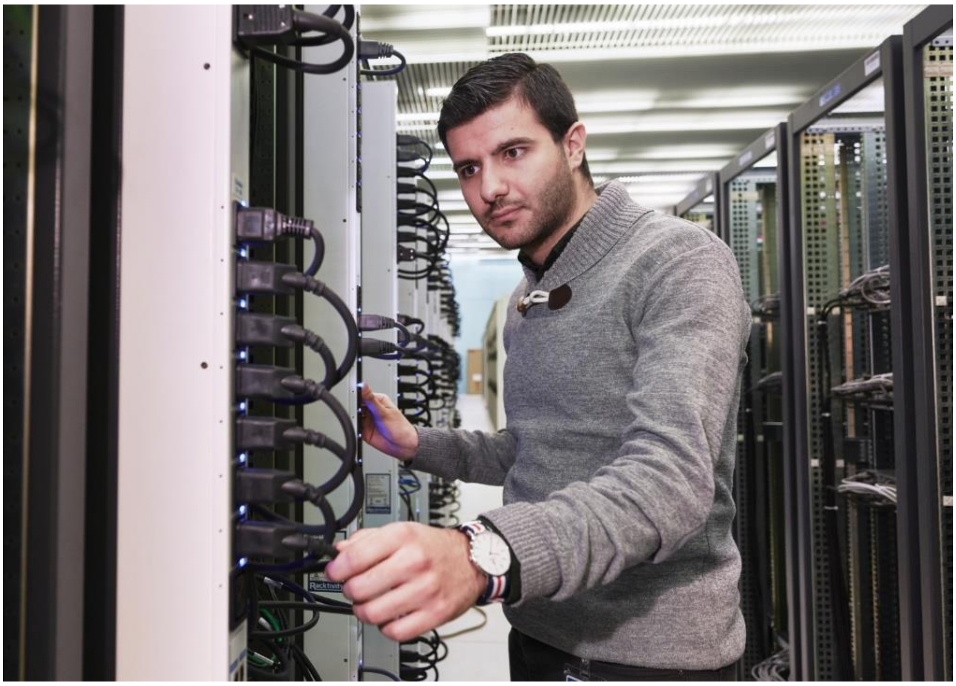
- D. Cicalese et al., The design of a distributed key-value store for petascale hot storage in data acquisition systems published in EPJ Web Conf. 214, 2019. *cern.ch/go/xf9H*

Presentations

- M. Maciejewski, Persistent Memory based Key-Value Store for Data Acquisition Systems (25 September). Presented at IXPUG 2019 Annual Conference, Geneva, 2019. *cern.ch/go/9cFB*

- G. Jereczek, Let's get our hands dirty: a comprehensive evaluation of DAQDB, key-value store for petascale hot storage (5 November). Presented at the 4th International Conference on Computing in High-Energy and Nuclear Physics (CHEP), Adelaide, 2019. *cern.ch/go/9cpL8*

- J. Radtke, A Key-Value Store for Data Acquisition Systems (April). SPDK, PMDK and VTune (tm) Summit 04'19, Santa Clara, 2019. *cern.ch/go/H6Rl*



The project formally came to a close in 2019, but several developments and tests will continue in 2020.

TESTBED FOR GPU-ACCELERATED APPLICATIONS

R&D TOPIC 2: COMPUTING PERFORMANCE AND SOFTWARE

Project coordinator: *Maria Girone*

Technical coordinators:
Andrea Bocci, Felice Pantaleo, Maurizio Pierini, Federico Carminati, Vincenzo Innocente, Marco Rovere, Jean-Roch Vlimant, Vladimir Gligorov, Daniel Campora, Riccardo De Maria, Adrian Oeftiger, Lotta Mether, Ian Fisk, Lorenzo Moneta, and Sofia Vallecorsa

Technical team:
Mary Touranakou, Thong Nguyen, Javier Duarte, Olmo Cerri, Jan Kieseler, Roel Aaij, Dorothea Vom Bruch, Blaise Raheem Delaney, Ifan Williams, Niko Neufeld, Viktor Khristenko, Florian Reiss, Guillermo Izquierdo Moreno, Luca Atzori, Miguel Fontes Medeiros

Collaborator liaisons:
Cosimo Gianfreda (E4), Daniele Gregori (E4), Agnese Reina (E4), Piero Altoé (NVIDIA), Andreas Hehn (NVIDIA), Tom Gibbs (NVIDIA)

Project goal

The goal of this project is to adapt computing models and software to exploit fully the potential of GPUs. The project, which began in late 2018, consists of ten individual use cases.

Background

Heterogeneous computing architectures will play an important role in helping CERN address the computing demands of the HL-LHC.

Progress in 2019

This section outlines the progress made in the six main use cases that were worked on in 2019, related to computing performance and software, as well as machine learning and data analytics.

1. Simulating sparse data sets for detector geometries
We are working to generate realistic detector geometries using deep generative models, such as adversarially trained networks or variational autoencoders. To this end, we are investigating custom loss functions able to deal with the specificity of LHC data. We plan to optimise the model inference on GPUs, with a view to delivering a production-ready version of the model to the LHC experiments.

Work got underway in late 2019. We began designing and training the model on two benchmark data sets (the Modified National Institute of Standards and Technology data set and the LHC Jet data set) before moving to the actual problem: generation of detector hits in a realistic set-up. Once we have converged on the full model design and the custom set-up (loss function, etc.), we plan to move to a realistic data set and scale up the problem.

2. 'Patatrack' software R&D incubator
The Patatrack initiative is focused on exploiting new hardware and software technologies for sustainable computing at the CMS experiment. During 2019, the Patatrack team demonstrated that it is possible to run some of its particle-collision reconstruction algorithms on NVIDIA GPUs. Doing so led to a computing performance increase of an order of magnitude. The algorithms were initially developed using NVIDIA's CUDA platform. They were then ported on an ad hoc basis to run on conventional CPUs, with identical results and close to native performance. When run on an NVIDIA Tesla T4 GPU, the algorithm achieves twice the performance compared to when it is run on a full dual socket Intel Xeon Skylake Gold node.

Performance portability will be explored during 2020. A comparison of the tested solutions — in terms of supported features, performance, ease of use, integration in wider frameworks, and prospects — will also be carried out.

3. Benchmarking and optimisation of TMVA deep learning
ROOT is an important data-analysis framework used at CERN. This use case is focused on optimising training and evaluating the performance of the ROOT's TMVA (Toolkit for Multivariate Data Analysis) on NVIDIA GPUs. During 2019, thanks to the contribution of Joanna Niermann (a student supported by CERN openlab), a new implementation of the convolution operators of the TMVA was performed. This made use of NVIDIA's cuDNN library and led to a significant boost in performance when training or evaluating deep-learning models.

We also performed a comparison study with Keras and TensorFlow. This showed better computational performances for the TMVA implementation, especially when using smaller models. Our new implementation was released in ROOT's new 6.20 version.

In addition, we developed new GPU implementations using cuDNN for recurrent neural networks. These also showed very good computational performance and will be integrated into the upcoming ROOT version 6.22.

4. Distributed training
There is a growing need within the high-energy physics community for an HPC-ready turnkey solution for distributed training and optimisation of neural networks. We aim to build software with a streamlined user interface to federate various available frameworks for distributed training. The goal is to bring the most performance to the end user, possibly through a 'training-as-a-service' system for HPC. In 2019, we prepared our software for neural-network training and optimisation. User-friendly and HPC-ready, it has been trialled at multiple institutions. There are, however, outstanding performance issues to be ironed out. We will also work in 2020 to build a community around our final product.

5. Integration of SixTrackLib and PyHEADTAIL
For optimal performance and hardware-utilisation, it is crucial to share the particle state in-place between the codes SixTrackLib (used for tracking single particles following collisions) and PyHEADTAIL (used for simulating macro-particle beam dynamics). This helps to avoid the memory and run-time costs of maintaining two copies on the GPU. The current implementation allows this by virtue of implicit context sharing, enabling seamless hand-off of control over the shared state between the two code-bases. After a first proof-of-concept implementation was created at an E4-NVIDIA hackathon in April 2019, the solution was refined and the API of the libraries was adopted to support this mode of operation.

Work carried out within this project made it possible for PyHEADTAIL to rely on SixTrackLib for high-performance tracking on GPUs, resulting in performance improvements of up to two-to-three orders of magnitude compared to state-of-the-art single-threaded CPU-based code. We also exposed SixTrackLib to new applications and use cases for particle tracking, which led to several improvements and bug fixes.

We are now working on further optimisation, as well as extending it to new applications. We will also work to expand the user community in 2020.

6. Allen: a high-level trigger on GPUs for LHCb
'Allen' is an initiative to develop a complete high-level trigger (the first step of the data-filtering process following particle collisions) on GPUs for the LHCb experiment. It has benefitted from support through CERN openlab, including consultation from engineers at NVIDIA.

The new system processes 40 Tb/s, using around 500 of the latest generation NVIDIA GPU cards. Allen matches — from a physics point of view — the reconstruction performance for charged particles achieved on traditional CPUs. It has also been shown that Allen will not be I/O or memory limited. Plus, not only can it be used to perform reconstruction, but it can also take decisions about whether to keep or reject events.

A diverse range of algorithms have been implemented efficiently on Allen. This demonstrates the potential for GPUs not only to be used as 'accelerators' in high-energy physics, but also as complete and standalone data-processing solutions.

Allen will now be used by the LHCb experiment as the new baseline solution for the next run of the LHC.

Next steps
As outlined above, work related to each of these use cases will continue in 2020. Work will also begin on a number of new use cases.

A full list of publications and presentations related to this project can be found on the CERN openlab website: cern.ch/go/7rkR.

DATA ANALYTICS IN THE CLOUD

R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS

ORACLE®

Project coordinator:
Eva Dafonte Perez, Eric Grancher

Technical team:
Luca Canali, Riccardo Castellotti

Collaborator liaisons:
Barry Gleeson, Vincent Leocorbo, Don Mowbray, Cristobal Pedregal-Martin, David Ebert, Dmitrij Dolgušin

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 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 the detectors of the LHC experiments. 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 2019

In 2019, we tested and deployed data-analytics and machine-learning workloads of interest for CERN on OCI. Testing began with the deployment of Apache Spark on Kubernetes, using OCI resources.

During this initial phase, we were able to successfully deploy two workloads for processing physics data at scale:

- *Reduction of big data from the CMS experiment:*

This use case consists of running data-reduction workloads for data from particle collisions. Its goal is to demonstrate the scalability of a data-reduction workflow based on processing ROOT files using Apache Spark.

- *Spark deep-learning trigger:*

This use case entails the deployment of a full data-preparation and machine-learning pipeline (with 4.5 TB of ROOT data) using Apache Spark and TensorFlow.

This activity has led to a number of improvements. In particular, we were able to improve the open-source connector between OCI and the Hadoop Distributed File System: we made it compatible with recent versions of Spark and we developed a mechanism to distribute workloads.

Next steps

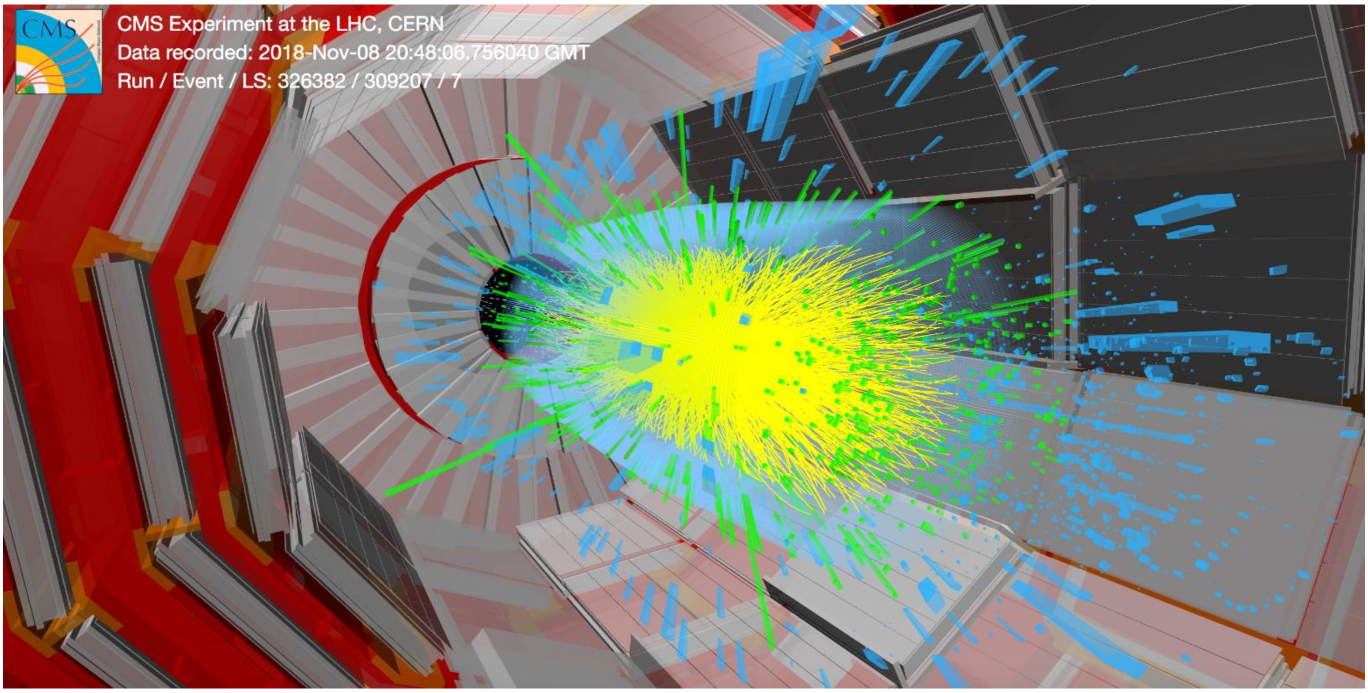
In 2020, the focus of the project will also include work to improve user interfaces and ease of adoption. We will develop a proof-of-concept integration of CERN's analytics platform (SWAN) with OCI resources.

Publications

- M. Bień, Big Data Analysis and Machine Learning at Scale with Oracle Cloud Infrastructure. Zenodo (2019). *cern.ch/go/lhH9*
- M. Migliorini, R. Castellotti, L. Canali, M. Zanetti, Machine Learning Pipelines with Modern Big Data Tools for High Energy Physics. arXiv e-prints,p. arXiv:1909.10389 [cs.DC], Sep 2019. *cern.ch/go/8CpQ*

Presentations

- L. Canali, "Big Data In HEP" - Physics Data Analysis, Machine learning and Data Reduction at Scale with Apache Spark (24 September). Presented at IXPUG 2019 Annual Conference, Geneva, 2019. *cern.ch/go/6pr6*
- L. Canali, Deep Learning Pipelines for High Energy Physics using Apache Spark with Distributed Keras on Analytics Zoo (16 October). Presented at Spark Summit Europe, Amsterdam, 2019. *cern.ch/go/xp77*



A particle-collision event recorded by the CMS detector.

DATA ANALYTICS FOR INDUSTRIAL CONTROLS AND MONITORING

R&D TOPIC 3:
MACHINE LEARNING AND DATA ANALYTICS

SIEMENS

Project coordinator:
Fernando Varela Rodriguez

Technical team:
Filippo Tilaro, Rafal Kulaga, Piotr Golonka, Peter Sollander, Fernando Varela, Marc Bengulescu, Filip Siroky

Collaborator liaisons:
Thomas Hahn, Juergen Kazmeier, Alexey Fishkin, Tatiana Mangels, Elisabeth Bakany, Ewald Sperrer

Project goal

This project is working to render the industrial control systems used for the LHC more efficient and more intelligent. The aim is to develop a data-analytics platform that capitalises on the latest advances in artificial intelligence (AI), cloud and edge-computing technologies. The ultimate goal is to make use of analytics solutions provided by Siemens to provide non-expert end users with a turnkey data-analytics service.

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 2019

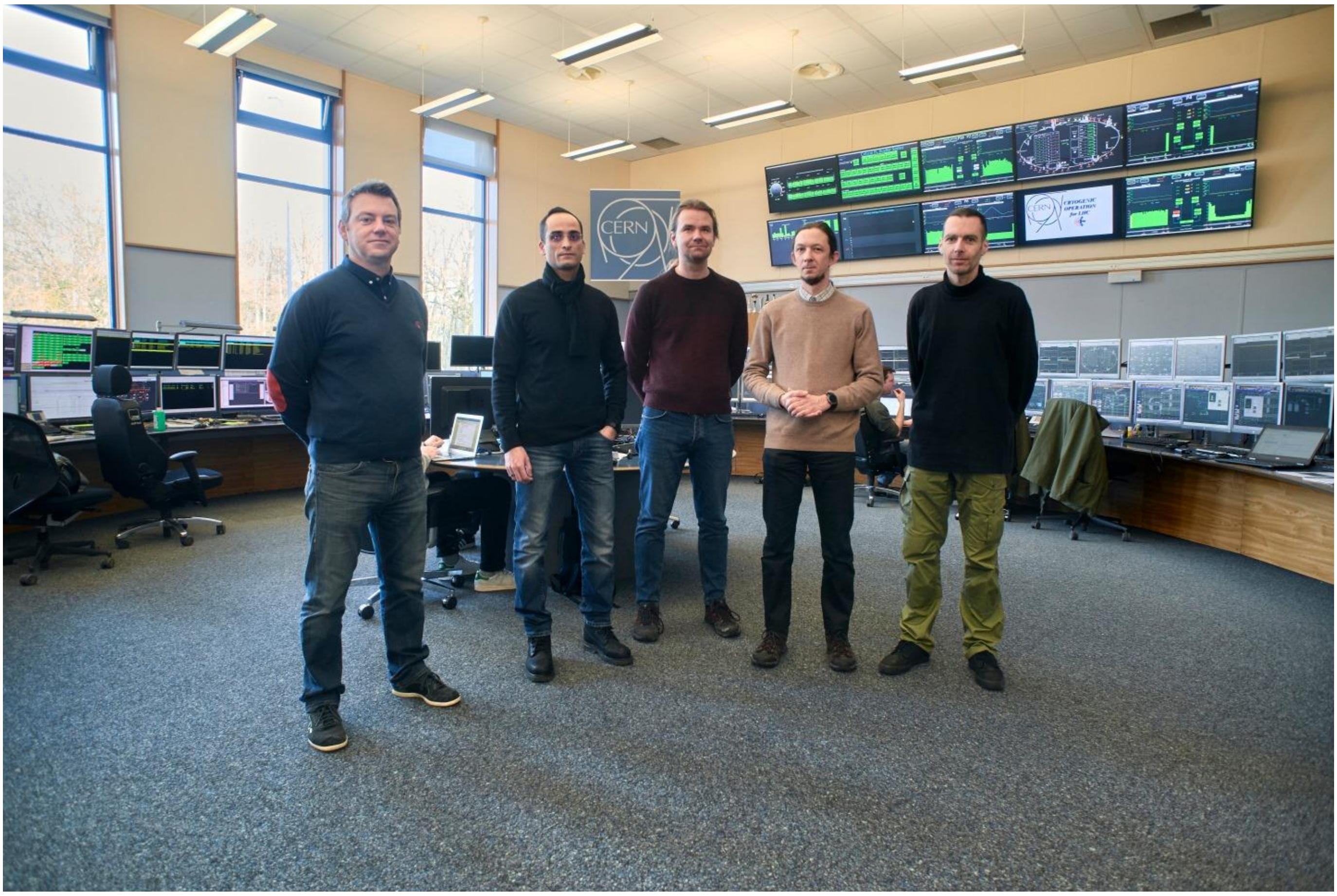
In the first half of 2019, we focused on the monitoring of various LHC control systems, using two distinct analytics solutions from Siemens: Smart IIoT, a framework used to monitor a multitude of control signals in a distributed manner, and ELVis, a web-based platform for handling multiple streams of time-series data from sensors.

Achieving tighter integration between ELVis and Smart IIoT was one of the main objectives for the first half of 2019. A single interface was developed to enable users to define complex event-processing rules, configure the cloud and edge infrastructure, and monitor the execution of the analyses.

In the second half of 2019, Filip Siroky, a new fellow funded by CERN openlab, joined the team. His work has focused on the following: optimising the ion-beam source for the Linac3 accelerator at CERN; deploying Siemens's Distributed Complex Event Processing (DCEP) technology to enable advanced data analytics and predictive maintenance for the oxygen-deficiency sensors in the LHC tunnel; and integrating an array of Siemens IoT infrared sensors for detecting room occupancy into the central room-booking system at CERN.

Next steps

One of the main objectives for 2020 is to integrate the DCEP technology with the control systems of other equipment groups at CERN: cryogenics, electricity, and cooling and ventilation. The other aim is to provide a service for collection of generic AI algorithms that could easily be employed by people who are not data scientists, helping them to perform advanced analytics on controls data.



The team focuses on the development of a data-analytics platform that combines the benefits of cloud and edge computing.

EXPLORING ACCELERATED MACHINE LEARNING FOR EXPERIMENT DATA ANALYTICS

**R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS**

Project coordinator:
Emilio Meschi, Paola Sala, Maria Girone

Technical team:
Thomas Owen James, Dejan Golubovic, Maurizio Pierini, Manuel Jesus Rodriguez, Anwesha Bhattacharya, Saul Alonso-Monsalve, Debdeep Paul, Niklas Böhm, Ema Puljak

Collaborator liaisons:
Mark Hur, Stuart Grime, Michael Glapa, Eugenio Culurciello, Andre Chang, Marko Vitez, Dustin Werran, Aliasger Zaidy, Abhishek Chaurasia, Patrick Estep, Jason Adlard, Steve Pawlowski

Project goal

The project has two threads, each investigating a unique use case for the Micron Deep Learning Accelerator (a modular FPGA-based architecture).

The first thread relates to the development of a real-time streaming machine inference engine prototype for the level-1 trigger of the CMS experiment.

The second thread focuses on prototyping a particle-identification system based on deep learning for the DUNE experiment. DUNE is a leading-edge, international experiment for neutrino science and proton-decay studies. It will be built in the US and is scheduled to begin operation in the mid-2020s.

Background

The level-1 trigger of the CMS experiment selects relevant particle-collision events for further study, while rejecting 99.75% of collisions. This decision must be made with a fixed latency of a few microseconds. Machine-learning inference in FPGAs may be used to improve the capabilities of this system.

The DUNE experiment will consist of large arrays of sensors exposed to high-intensity neutrino beams. The use of convolutional neural networks has been shown to substantially boost particle-identification performance for such detectors. For DUNE, an FPGA solution is advantageous for processing ~5 TB/s of data.

Progress in 2019

For the CMS experiment, we studied in detail two potential use cases for a machine-learning approach using FPGAs. Data from Run 2 of the LHC was used to train a neural network. The goal of this is to improve the analysis potential of muon tracks from the level-1 trigger, as part of a 40 MHz 'level-1 scouting' data path. In addition, a convolutional neural network was developed for classifying and measuring energy showers for the planned high-granularity calorimeter upgrade of the CMS experiment. These networks were tested on the Micron FPGA hardware and were optimised for latency and precision.

For the DUNE part of the project, we tested the Micron inference engine and characterised its performance on existing software. Specifically, we tested it for running a neural network that can identify neutrino interactions in the DUNE detectors, based on simulated data. This enabled us to gain expertise with the board and fully understand its potential. The results of this benchmarking were presented at the 24th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2019).

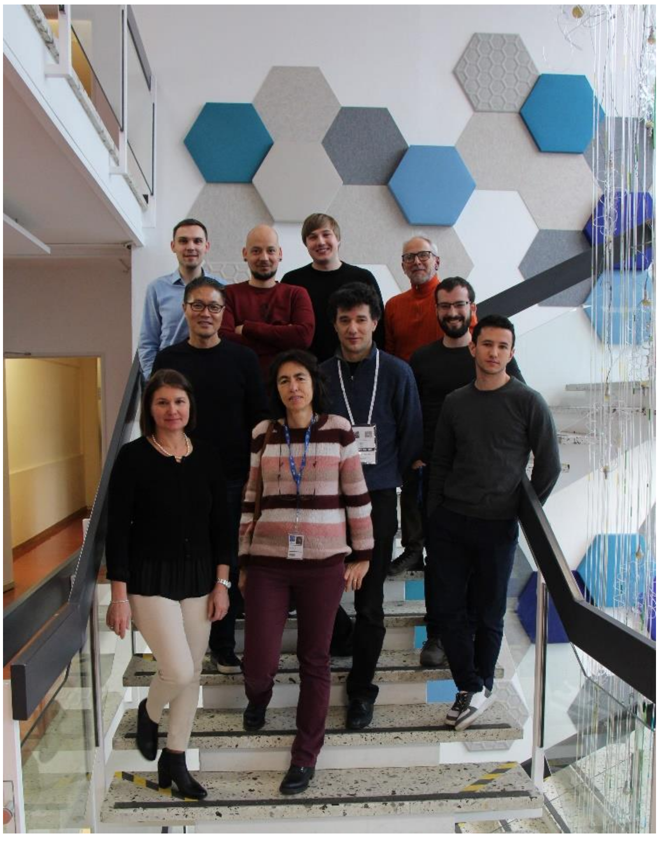
Next steps

The CMS team will focus on preparing a full scouting system for Run 3 of the LHC. This will comprise a system of around five Micron co-processors, receiving data on high-speed optical links.

The DUNE team plans to set up the inference engine as a demonstrator within the data-acquisition system of the ProtoDUNE experiment (a prototype of DUNE that has been built at CERN). This will work to find regions of interest (i.e. high activity) within the detector, decreasing the amount of data that needs to be sent to permanent storage.

Presentations

- M. J. R. Alonso, Fast inference using FPGAs food DUNE data reconstruction (7 November). Presented at 24th International Conference on Computing in High Energy and Nuclear Physics, Adelaide, 2019. *cern.ch/go/bl7n*



Members of the project team.

EVALUATION OF POWER CPU ARCHITECTURE FOR DEEP LEARNING

R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS

IBM

Project coordinator:
Maria Girone and Federico Carminati

Technical team
Sofia Vallecorsa

Collaborator liaisons:
Eric Aquaronne, Lionel Clavien

Project goal

We are investigating the performance of distributed learning and low-latency inference of generative adversarial networks (GANs) for simulating detector response to 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 greatly reducing the need for detailed Monte Carlo (MC) simulations in generating particle showers. Detailed MC is computationally expensive, so this could be a way to improve the overall performance of simulations in high-energy physics.

Using the large datasets 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 2019

In accordance with the concept of data-parallel distributed learning, we trained a GAN model 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 dataset for training the model.

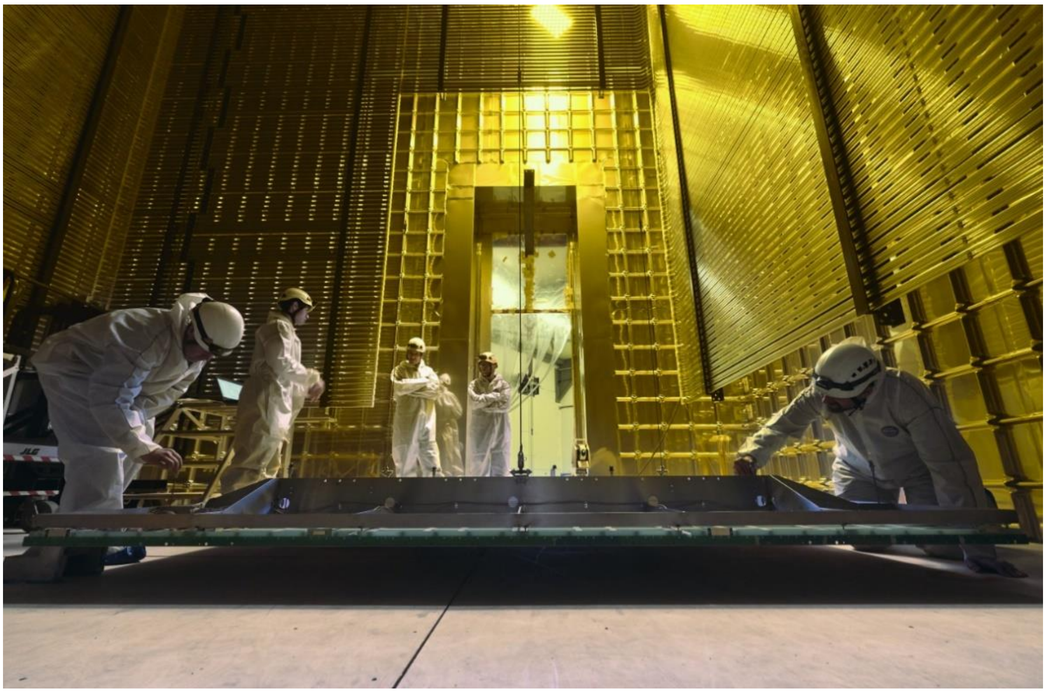
The model we benchmarked is called '3DGAN'. It uses three-dimensional convolutions to simulate the energy patterns deposited by particles travelling through high-granularity calorimeters (part of the experiments' detectors). More details about this can be found on the page about the fast-simulation project). In order to distribute the training workload across multiple nodes, 3DGAN uses an MPI-based tool called Horovod. Running on the test cluster, we achieved excellent scaling performance and improved the training time by an order of magnitude.

As planned, work also began in 2019 to 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. Initial work focused on developing a model — based on a combination of convolutional and graph networks — to reduce the noise in the raw data produced by the detector. Preliminary results on MC-simulated data are very promising.

Next steps

We will work to further optimise our noise-reduction model for the DUNE data, testing its performance on real data collected from a prototype experiment built at CERN called ProtoDUNE. Furthermore, we will investigate the feasibility of running the model in law-latency environments for real-time applications, using FPGAs.

Our plan is to then extend this approach to perform several other steps in the data-processing chain. In the longer term, our ultimate goal is to develop a tool capable of processing the raw data from DUNE, thus making it possible to replace the entire offline reconstruction approach.



Scientists working on the ProtoDUNE experiment.

FAST DETECTOR SIMULATION

R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS

Project coordinator:
Federico Carminati

Technical team:
Sofia Vallecorsa, Gul Rukh Khattak

Collaborator liaisons:
Claudio Bellini, Marie-Christine Sawley, Andrea Luiselli, Saletore Vikram, Hans Pabst, Sun Choi, Fabio Baruffa from Intel.

Valeriu Codreanu, Maxwell Cai, Damian Podareanu from SURFsara, B.V., which is also collaborating in the project

Project goal

We are using artificial intelligence (AI) techniques to simulate the response of the HEP detectors to particle collision events. Specifically, we are developing deep neural networks and, in particular, generative adversarial networks (GANs) to do this. Such tools will play a significant role in helping the research community cope with the vastly increased computing demands of the High Luminosity LHC (HL-LHC).

Once properly trained and optimised, generative models are able to simulate a variety of particles, energies, and detectors in just a fraction of the time required by classical simulation, which is based on detailed Monte Carlo methods. Our objective is to tune and integrate these new tools in the experiments' existing simulation frameworks.

Background

Simulating the response of detectors to particle collisions — under a variety of conditions — is an important step on the path to new physics discoveries. However, this work is very computationally expensive. Over half of the computing workload of the Worldwide LHC Computing Grid (WLCG) is the result this single activity.

We are exploring an alternative approach, referred to as 'fast simulation', which trades some level of accuracy for speed. Fast-simulation strategies have been developed in the past, using different techniques (e.g. look-up tables or parametrised approaches). However, the latest developments in machine learning (particularly in relation to deep neural networks) make it possible to develop fast-simulation tools that are both more flexible and more accurate than those developed in the past.

Progress in 2019

Building on our work from 2018, we focused on optimising a more complex model that can simulate the effects of several particle types to within 5-10 % over a large-energy range and for realistic kinematic conditions. The model is remarkably accurate: GANs can reproduce Monte Carlo predictions to within just a few percent.

Training time is, however, still a bottleneck for the meta-optimisation of the model. This includes not only the optimisation of the network weights, but also of the architecture and convergence parameters. Much of our work in 2019 concentrated on addressing this issue.

We followed up on the work, started in 2018, to develop distributed versions of our training code, both on GPUs and CPUs.

We tested their performance and scalability in different environments, such as high-performance computing (HPC) clusters and clouds. The results are encouraging: we observed almost linear speed-up as the number of processors increased, with very limited degradation in the results.

We also began work to implement a genetic algorithm for optimisation. This simultaneously performs training and hyper-parameter optimisation of our network, making it easier to generalise our GAN to different detector geometries.

Next steps

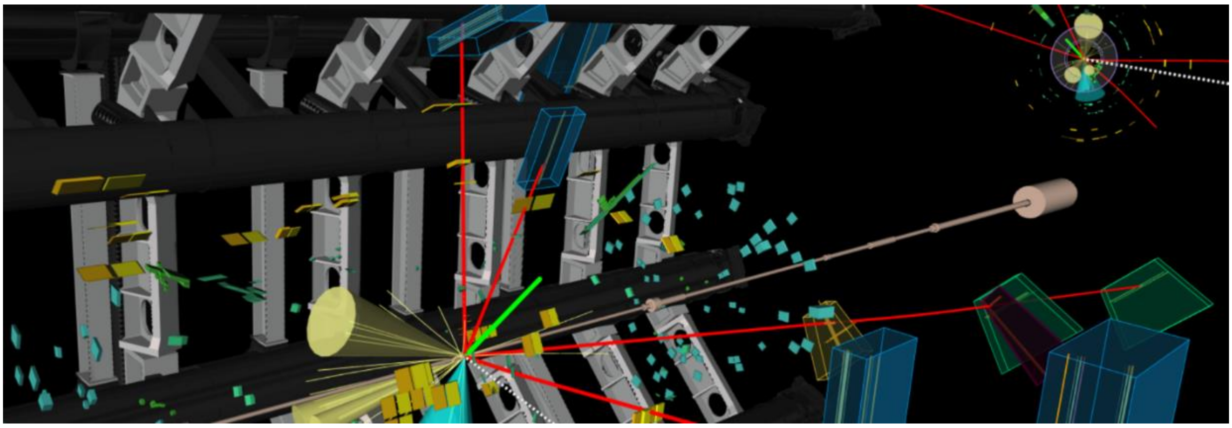
We will continue to investigate HPC training and will work on the optimisation of physics accuracy in the distributed training mode. We will also complete the development of the genetic-algorithm approach for hyper-parameter optimisation. On top of this, we will extend the tool to other types of detectors.

More broadly though, we now believe our model is mature enough to start planning its test integration with the classical approaches currently used by the LHC experiments. In addition, we will also extend the tool to cover other detectors not currently simulated.

Publications

- D. Brayford, S. Vallecorsa, A. Atanasov, F. Baruffa, W. Riviera, Deploying AI Frameworks on Secure HPC Systems with Containers. Presented at 2019 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, 2019, pp. 1-6. *cern.ch/go/fg6k*

- G. R. Khattak, S. Vallecorsa, F. Carminati, G. M. Khan, Particle Detector Simulation using Generative Adversarial Networks with Domain Related Constraints. Presented at 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, 2019, pp. 28-33. *cern.ch/go/rl9D*



Simulating the response of detectors to particle collisions is an important step on the path to new physics discoveries.

NEXTGENERATION ARCHIVER FOR WINCC OA

R&D TOPIC 3:
MACHINE LEARNING
AND DATA ANALYTICS

SIEMENS

Project coordinator:
Fernando Varela Rodriguez

Technical team:
Filippo Tilaro, Jakub Guzik, Anthony Hennessey, Rafal Kulaga, Piotr Golonka, Peter Sollander, Fernando Varela, Marc Bengulescu, Filip Siroky

Collaborator 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 2019

Two important milestones for the NextGeneration Archiver (NGA) project were achieved in 2019: preparation of a release for all ETM customers with WinCC OA 3.17 and start of deployment at the ALICE experiment.

Significant progress has been made with all areas of the NGA project, including providing support for redundancy, for complex queries, and for handling signal metadata. In order to improve the performance and scalability of queries, and to make sure that they do not negatively affect core components of the system, direct query functionality was also developed and tested.

In order to ensure reliability of the NGA in large systems with high throughput, several tests were performed at CERN. Existing test automation tools have been significantly extended in order to allow for better synchronisation of testing efforts at CERN and ETM.

Initial results from InfluxDB performance tests performed at CERN show that the technology will most likely not be able to replace the current Oracle technology used for systems with very large numbers of signals (in the range of hundreds of thousands). However, it could successfully act as a shorter-term storage, improving the performance of certain queries and enabling users to easily create web dashboards using Grafana.

Next steps

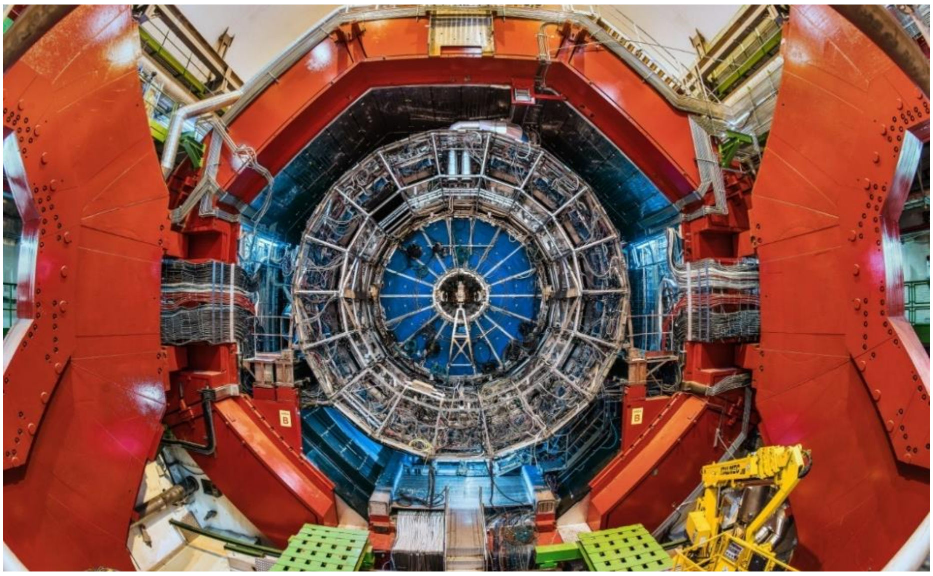
In 2020, work on the project will continue on many fronts. Increasing test coverage, especially for 'corner cases' and failure scenarios, remains one of the main priorities. Work on missing features will continue for all components of the NGA. Further tests of InfluxDB and Apache Kudu will help to determine their performance in large systems. The team will also provide support for ALICE as the experiment prepares to restart after the current long shutdown.

Publications

- P. Golonka, F. Varela-Rodriguez, Consolidation and Redesign of CERN Industrial Controls Frameworks, Proceedings for 17th Biennial International Conference on Accelerator and Large Experimental Physics Control Systems, New York, 2019. *cern.ch/go/8RRL*

Presentations

- F. M. Tilaro, R. Kulaga, Siemens Data Analytics and SCADA evolution status report (23 January). Presented at CERN openlab Technical Workshop, Geneva, 2019. *cern.ch/go/kt7K*



The ALICE-detector – one of the four main LHC experiments.

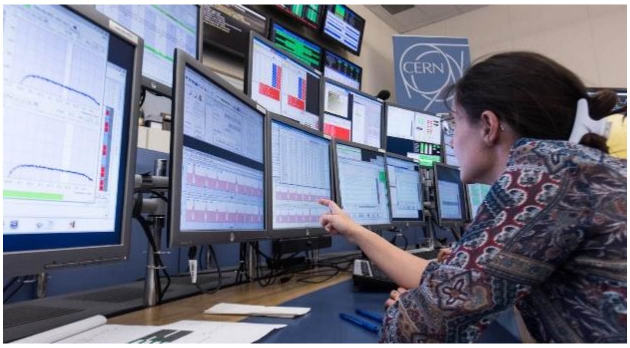
ORACLE CLOUD TECHNOLOGIES FOR DATA ANALYTICS ON INDUSTRIAL CONTROL SYSTEMS

R&D TOPIC 3
MACHINE LEARNING AND DATA ANALYTICS

Project coordinator:
Manuel Martin Marquez

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

Collaborator liaisons:
Çetin Özbütün, Reiner Zimmermann, Michael Connaughton, Cristobal Pedregal-Martin, Engin Senel, Cemil Alper, Giuseppe Calabrese, David Ebert, Dmitrij Dolgušin



An operator in the CERN Control Centre.

Project goal

CERN's control systems acquire over 250 TB of data per day from over 2 million signals from across the LHC and its experiments. Managing these extremely complex "Industrial Internet of Things" (IIoT) systems raises significant challenges related to data management, retrieval and analytics.

The project team is working with Oracle Autonomous Database and analytics technologies. The goal is to assess their capabilities in terms of integrating heterogeneous control IIoT data sets and improving performance and efficiency for the most challenging analytics requirements, while reducing operational costs.

Background

Keeping the LHC and the rest of the accelerator complex at CERN running efficiently requires state-of-the-art control systems. A complex IIOT system is in place to persist this data, making it possible for engineers to gain insights about temperatures, magnetic-field strengths, and more. This plays a vital role in ensuring the highest levels of operational efficiency.

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

Progress in 2019

During 2019, the team focused on three main aspects: (i) scaling data volumes, (ii) improving the efficiency of the potential solutions in terms of automatisations and reducing operational costs, and (iii) increasing data retrieval/analytics complexity using real-life scenarios.

We began migrating one of the largest and most complex control data sets to the object-storage system of Oracle Cloud Infrastructure (OCI). Due to the large volume of this data set (about 1 PB), different solutions — based on standard networks, the GÉANT network, and Oracle's appliance-based data-transfer solution — were tested. At the same time, the team worked together with development and management teams at Oracle to define the best data-model strategy to reduce associated costs and improve efficiency. To achieve this, a hybrid model was put in place. This model emphasises the benefits of object storage in the following manner: transparent external tables, based on parquet files, are used for data that is infrequently accessed, whereas normal database tables are used for data that requires close to real-time responses. To assess this, once a representative amount of data was available, real-life data loads were captured and simulated on OCI's Autonomous Database.

Next steps

In 2020, we will focus on finalising the migration of the historical data to OCI Object Storage, to make it available for Oracle Autonomous Database instances. This will require us to face challenges related to the task flow for Oracle's appliance-based data-transfer solution and the network configuration for GÉANT. In addition, we will work on automating data ingestion. In parallel, we will constantly increase real analytics load and asses solutions for interactive data exploration based on Oracle Autonomous Database technologies.

Presentations

- E. Grancher, M. Martin, S. Masson, Research Analytics at Scale: CERN's Experience with Oracle Cloud Solutions (16 January). Presented at Oracle OpenWorld 2019, London, 2019. *cern.ch/go/S6qf*
- A. Mendelsohn (Oracle), E. Grancher, M. Martin, Oracle Autonomous Database Keynote (16 January). Oracle OpenWorld 2019, London, 2019.
- M. Martin, J. Abel (Oracle), Enterprise Challenges and Outcomes (17 January). Presented at Oracle OpenWorld 2019, London, 2019.
- S. Masson, M. Martin, Managing one of the largest IoT systems in the world with Oracle Autonomous Technologies (18 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/SBc9*
- D. Ebert (Oracle), M. Martin, A. Nappi, Advancing research with Oracle Cloud (17 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/9ZCg*
- M. Martin, R. Zimmermann (Oracle), J. Otto (IDS GmbH), Oracle Autonomous Data Warehouse: Customer Panel (17 September). Presented at Oracle OpenWorld 2019, San Francisco, 2019. *cern.ch/go/nm9B*
- S. Masson, M. Martin, Oracle Autonomous Data Warehouse and CERN Accelerator Control Systems (25 November). Presented at Modern Cloud Day, Paris, 2019.
- M. Martin, M. Connaughton (Oracle), Big Data Analytics and the Large Hadron Collider (26 November). Presented at Oracle Digital Days 2019, Dublin, 2019.
- M. Martin, Big Data, AI and Machine Learning at CERN (27 November). Presented at Trinity College Dublin and ADAPT Center, Dublin, 2019.
- M. Martin, M. Connaughton (Oracle), Big Data Analytics and the Large Hadron Collider (27 November). Presented at the National Analytics Summit 2019, Dublin, 2019. *cern.ch/go/CF9p*

QUANTUM GRAPH NEURAL NETWORKS

e

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Federico Carminati (CERN openlab)

Team members:
Sofia Vallecorsa (CERN openlab), Fabio Fracas (CERN openlab), Cenk Tüysüz (METU), Daniel Dobos (gluoNNet), Jean-Roch Vlimant (CALTECH), Bilge Demirkoz (METU), Karolos Potamianos (DESY/University of Oxford), Kristiane Novotny (gluoNNet)

Project goal

The goal of this project is to explore the feasibility of using quantum algorithms to help track the particles produced by collisions in the LHC more efficiently. This is particularly important as the rate of collisions is set to increase dramatically in the coming years.

Background

The Large Hadron Collider (LHC) at CERN is producing collisions at unprecedented collider energy. The hundreds of particles created during the collisions are recorded by large detectors composed of several sub-detectors. At the centre of these detectors there is usually a tracker detector, precisely recording the signal of the passage of charged particles through thin layers of active material. The trajectories of particles are bent by a magnetic field, to allow the measurement of the particle momentum. There is an expected ten-fold increase in the number of tracks produced per bunch crossing after the high-luminosity upgrade of LHC. Classical algorithms to perform the reconstruction of the trajectory of charged particles are making use of Kalman filter formalism and even though quite accurate, scale worse than quadratically with the number of tracks. Several ways are explored to mitigate the increase in the computing needs, such as new detector layout, deep learning and code parallelisation. Quantum computing has been shown to provide speed-ups for certain problems and different R&D initiatives are exploring how quantum tracking algorithms could leverage such capabilities. We are developing a quantum-based track-finding algorithm aimed at reducing the combinatorial background during the initial seeding stage for the Kalman filters. We are using the publicly available dataset designed for the recent Kaggle 'TrackML' challenge for this work.

Progress in 2019

We have established a consortium of parties interested in addressing this challenge. Members of the following organisations are now working together on this project: the Middle East Technical University (METU) in Ankara, Turkey; the Deutsches Elektronen-Synchrotron (DESY) in Hamburg, Germany; the California Institute of Technology (Caltech) in Pasadena, US; and gluoNNet, a humanitarian-focused big data analysis not-for-profit association based in Geneva, Switzerland. Quantum graph neural networks (QGNNs) can be implemented to represent quantum processes that have a graph structure. In the summer of 2019, we began work to develop a prototype QGNN algorithm for the tracking the particles produced by collision events.

Next Steps

Initial results are very encouraging. We have implemented a simplified GNN architecture and we were able to train it to convergence achieving results that are very similar to the ones obtained by the classical counterpart. Work is still ongoing in order to better understand the performance and optimise the training process.

Presentations

- F. Carminati, Particle Track Reconstruction with Quantum Algorithms (7 November). Presented at Conference on Computing in High Energy & Nuclear Physics (CHEP), Adelaide, 2019. *cern.ch/go/7Ddm*
- D. Dobos, HEP Graph Machine Learning for Industrial & Humanitarian Applications (26 November). Presented at Conference on HEPTECH AIME19 AI & ML, Budapest, 2019. *cern.ch/go/6qqx*

The project is exploring the feasibility using quantum algorithms to help track the particles produced by collisions in the LHC more efficiently.

QUANTUM MACHINE LEARNING FOR SUPERSYMMETRY SEARCHES

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Koji Terashi (University of Tokyo)

Team members:
Michiru Kaneda (University of Tokyo),
Tomoe Kishimoto (University of Tokyo),
Masahiko Saito (University of Tokyo)
Ryu Sawada (University of Tokyo),
Junici Tanaka (University of Tokyo),
Federico Carminati (CERN openlab),
Sofia Vallecorsa (CERN openlab),
Fabio Fracas (CERN openlab)

Project goal

The goal of this project is to develop quantum machine-learning algorithms for analysis of LHC collision data. The particular example chosen is the classification of supersymmetry signals from Standard Model background.

Background

The analysis of LHC data for the detection of effects beyond the standard model requires increasing levels of precision. Various machine-learning techniques are now part of the standard analysis toolbox for high-energy physics. Deep-learning algorithms are increasingly demonstrating their usefulness in various areas of analysis, thanks to their ability to explore a much larger dimensional space.

This seems to be an almost ideal area of application for quantum computing, which is offering a potentially enormous parameter space and a correspondingly large level of computational parallelism. Moreover, the quasi-optimal Gibbs sampling features of quantum computers may enhance the training of such deep-learning networks.

Progress in 2019

During this year, the Tokyo group has studied the performance of different quantum variational models for the classification of supersymmetry signals from the standard model background, where the signal is: $h \rightarrow \chi^{\pm}\chi^{\mp} \rightarrow WW (\rightarrow lvlv) + \chi^{0}_{1}\chi^{0}_{1}$, and the background comes from: $WW (\rightarrow lvlv)$. In particular, the efforts have focused on two approaches: quantum circuit learning (QCL) and quantum variational classifiers (QVC).

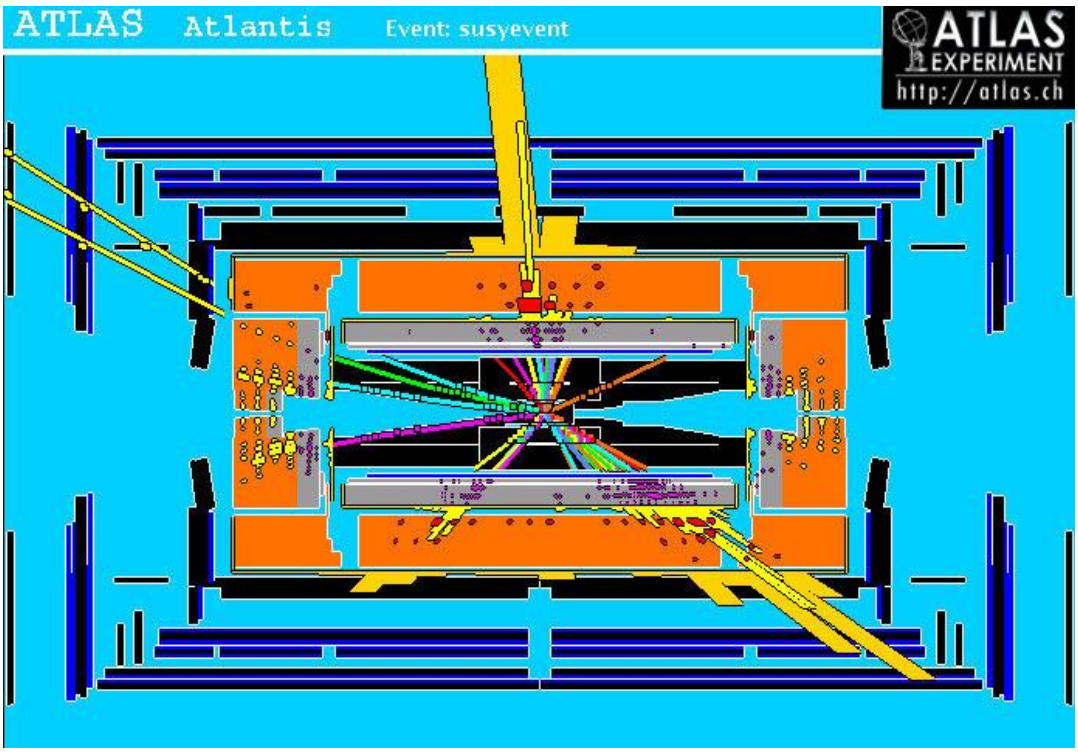
The supersymmetry dataset in the UCI machine-learning repository has been used for this study. A quantum circuit learning with a set of seven variables and a depth of three has been implemented. The results of 5000 iterations with COBYLA (constrained optimisation by linear approximation) minimisation have been compared with classical deep neural networks and and boosted decision trees. Promising results have been achieved. An initial implementation of a QVC, with a depth of two and a set of three input variables, is also being studied.

Next Steps

The current results are encouraging. The next steps of this work will be to test the results on real IBM Quantum Experience hardware. Initially, three variables will be used.

Presentations

- F. Carminati, Particle Track Reconstruction with Quantum Algorithms (7 November). Presented at Conference on Computing in High Energy & Nuclear Physics (CHEP), Adelaide, 2019. *cern.ch/go/7Ddm*
- D. Dobos, HEP Graph Machine Learning for Industrial & Humanitarian Applications (26 November). Presented at Conference on HEPTECH AIME19 AI & ML, Budapest, 2019. *cern.ch/go/6qqx*



The goal of this project is to develop quantum machine-learning algorithms for analysis of LHC collision data. The particular example chosen is the classification of supersymmetry signals from Standard Model background.

QUANTUM OPTIMISATION FOR GRID COMPUTING

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Google

Project coordinator:
Federico Carminati (CERN openlab)

Team members:
Mircea-Marian Popa (Politehnica University of Bucharest), Sofia Vallecorsa (CERN openlab), Fabio Fracas (CERN openlab), Costin Grigoras (CERN, ALICE), Latchezar Betev (CERN, ALICE), Mihai Carabas (Politehnica University of Bucharest), Popescu George Pantelimon (Politehnica University of Bucharest), Jaques Demongeot (Institut Polytechnique de Grenoble)

Project goal

The goal of this project is to develop quantum algorithms to help optimise how data is distributed for storage in the Worldwide LHC Computing Grid (WLCG), which consists of 170 computing centres, spread across 42 countries. Initial work focuses on the specific case of the ALICE experiment. We are trying to determine the optimal storage, movement, and access patterns for the data produced by this experiment in quasi-real-time. This would improve resource allocation and usage, thus leading to increased efficiency in the broader data-handling workflow.

Background

The WLCG has been essential to the success of the LHC's scientific programme. It is used to store and analyse the data produced by the LHC experiments. Optimal usage of the grid's resources is a major challenge: with the foreseen increase in the data produced by the LHC experiments, workflow optimisation — particularly for the data-placement strategy — becomes extremely important.

Simulating this complex and highly non-linear environment is very difficult; the complexity of the task goes beyond the capability of the computing hardware available today. Quantum computing could offer the possibility to address this. Our project — a collaboration with Google, the Polytechnic Institute of Grenoble and the Polytechnic University of Bucharest — will develop quantum algorithms to optimise the storage distribution.

Progress in 2019

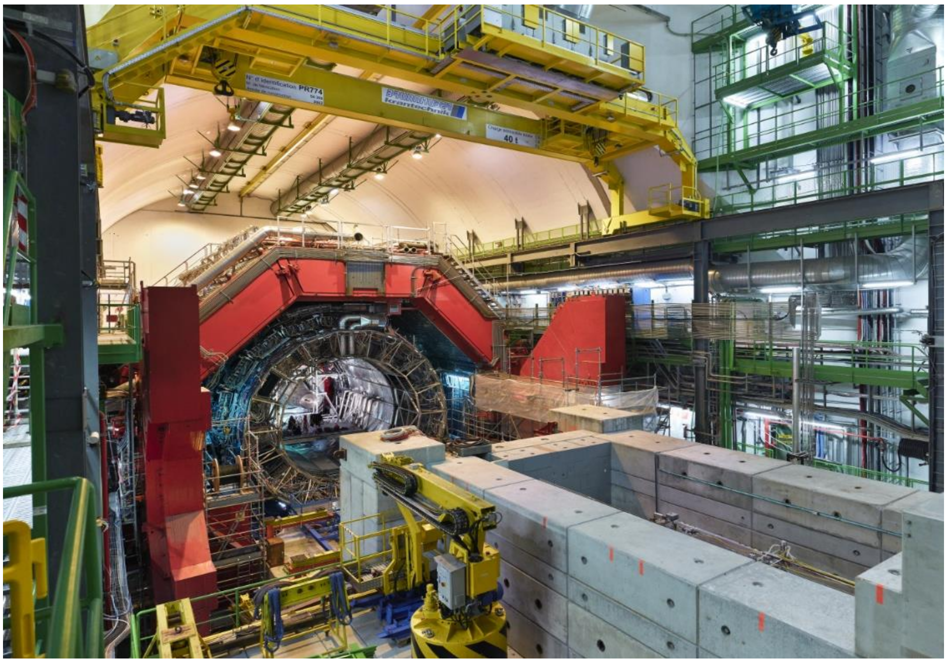
In May, this project was awarded one-year funding under the European Union's ATTRACT initiative. ATTRACT provides initial funding to 170 disruptive projects, each aiming to develop sensing and imaging technologies that will enable breakthrough innovations. This project, which has the full title of 'Quantum optimisation of Worldwide LHC Computing Grid data placement' is one of 19 projects funded in which CERN is involved. One of the major challenges faced by this project is the difficulty of defining a suitable description of the data set extracted from monALISA, the monitoring and scheduling tool used by the ALICE experiment for grid operations. We have now defined the problem in terms of reinforcement learning, one of the three paradigms of machine learning (alongside supervised and unsupervised learning). We have also started implementing the key components of the reinforcement-learning framework (in terms of environment and agent networks) that is to be used.

Next Steps

Currently the deep neural network describing the environment and the agent behaviours are being implemented as classical networks. Our first goal is to prove that this strategy can reproduce the expected monAlisa behaviour. At a later stage we will implement these components as quantum circuits using CIRQ and work on the optimisation of the training process. Quantum computing and grid/cloud computing are game-changing technologies with the potential to have a very large impact on the future. Our results will provide an excellent initial prototype: the work could be then extended and integrated with other existing initiatives at the scale of the whole Worldwide LHC Computing Grid. Eventually, benefits in terms efficient network usage, reduced computing time, optimised storage and therefore costs would be significant.

Presentations

- F. Carminati, Quantum Optimization of Worldwide LHC Computing Grid data placement (7 November). Presented at Conference on Computing in High Energy & Nuclear Physics (CHEP), Adelaide, 2019. *cern.ch/go/nP9M*



This ATTRACT-funded project is working to develop quantum algorithms that could be used to help optimise how data is distributed for storage in the WLCG.

QUANTUM SUPPORT VECTOR MACHINES FOR HIGGS BOSON CLASSIFICATION

R&D TOPIC 4:
QUANTUM TECHNOLOGIES

Project coordinator:
Sau-Lan Wu (Wisconsin University),
Ivano Tavernelli (IBM Zurich),
Federico Carminati (CERN openlab)

Team members:
Chen Zhou (University of Wisconsin),
Shaujun San (University of Wisconsin),
Wen Guan (University of Wisconsin),
Panagiotis Barkoutsos (IBM Zurich),
Jennifer Glick (IBM Zurich)

Project goal
This project is investigating the use of quantum support vector machines (QSVMs) for the classification of particle collision events that produce a certain type of decay for the Higgs boson. Specifically, such machines are being used to identify instances where a Higgs boson fluctuates for a very short time into a top quark and a top anti-quark, before decaying into two photons. Understanding this process — known by physicists as ttH production — is challenging, as it is rare: only 1% of Higgs bosons are produced in association with two top quarks and, in addition, the Higgs boson and the top quarks decay into other particles in many complex ways, or modes.

Background
QSVMs are among the most promising machine-learning algorithms for quantum computers. Initial quantum implementations have already shown performances comparable to their classical counterparts. QSVMs are considered suitable algorithms for early adoption on noisy, near-term quantum-computing devices. Several initiatives are studying and optimising input data representation and training strategies.

We are testing IBM's QSVM algorithm within the ATLAS experiment. Today, identifying ttH-production events relies on classical support vector machines, as well as another machine-learning technique known as 'boosted decision trees'. Classically, these methods are used to improve event selection and background rejection by analysing 47 high level characteristic features.

Progress in 2019

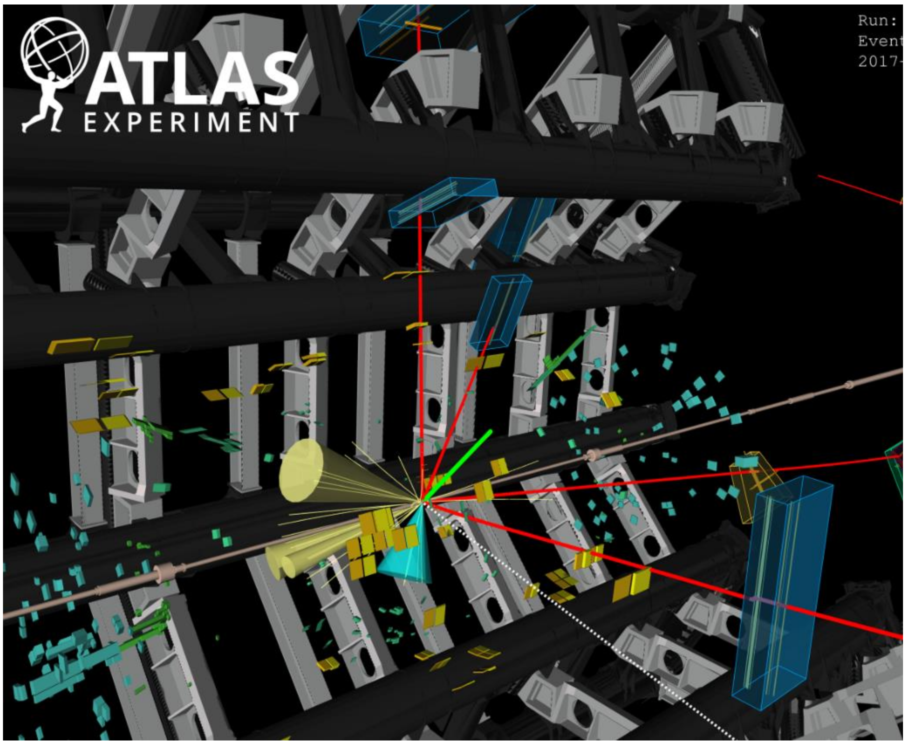
We are working to compare the QSVM to the classical approach in terms of classification accuracy. We are also working to ascertain the level of resources needed for training the model (time to convergence and training dataset size) and studying how different types of noise affect the final performance. In order to do this, we are making use of IBM's quantum simulator, with support from their expert team. Preliminary results, obtained using the quantum simulator, show that the QSVM can achieve performance that is comparable to its classical counterpart terms of accuracy, while being much faster. We are now simulating noise in different ways, in order to understand performance on real hardware.

Next Steps

Testing the algorithm on real hardware is one of the primary challenges. At the same time, we continue to work on the optimisation of the QSVM accuracy and we are studying the robustness of the algorithm against noise.

Presentations

- W. Guan, Application on LHC High Energy Physic data analysis with IBM Quantum Computing (March). Presented at 19th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT), Saas-Fee, 2019. *cern.ch/go/6DnG*



We are testing IBM's QSVM algorithm within the ATLAS experiment.

POSSIBLE PROJECTS IN QUANTUM COMPUTING

R&D TOPIC 4: QUANTUM TECHNOLOGIES

Project coordinator:
Federico Carminati (CERN openlab)

Team members:
Sofia Vallecorsa (CERN openlab), Fabio Fracas (CERN openlab), Su Yeon Chang (EPFL)

A call for collaboration
We present here a number of interesting projects in quantum computing that we have discussed with different partners, but for which there was no formal collaboration through CERN openlab in 2019. The objective is to attract interest from the user community and create a critical mass that could allow us to establish collaborations with users and vendors.

OpenQKD: Testbed for quantum key distribution
OpenQKD is a European Horizon 2020 project to test quantum key distribution (QKD) technology and to prepare a pan-European QKD deployment to protect European citizens and economies against the potential security threat posed by a quantum computer. The three-year project started in September 2019. With a budget of €18 million, 38 European partners are developing fibre-based and free-space QKD and deploying over 30 use cases at 16 sites across Europe. Among these, several use cases are planned in Geneva. One of them is the so-called Quantum Vault, which aims to protect digital assets against failures and attacks. As a proof of concept, the Quantum Vault is being realised in six Genevan data centres and telecom nodes. CERN openlab and the Poznan Supercomputing and Networking Centre in Poland are supporting the Quantum Vault by hosting one node at the CERN Data Centre. It is planned to actively involve CERN openlab by running a proper use case taking advantage of the Quantum Vault infrastructure.

Quantum generative adversarial networks
Generative adversarial networks (GANs) are among the most interesting models in classical machine learning. GANs are an example of generative models, i.e. models that learn a hidden distribution from the training dataset, and can sample new synthetic data. At CERN openlab, we have been investigating their use as an alternative to Monte Carlo simulation, obtaining remarkable results. Much faster than standard Monte Carlo algorithms, GANs can generate realistic synthetic data, while retaining a high level of accuracy (see our fast simulation project). Quantum GANs could have more representational power than classical GANs, making them better able to learn more complex distributions from smaller training datasets. We are now training a quantum GAN to generate images of a few pixels and we are investigating two possible approaches: a hybrid schema with a quantum generator learning the target PDF, using either a classical network or a vibrational quantum circuit as a discriminator (variational quantum generator), as well as a full quantum adversarial implementation (quGAN).

Track seeding optimisation
The Kalman filter is widely used in high-energy physics for track fitting of particle trajectories. It runs after an initial pattern-recognition step where detector 'hits' are clustered into subsets belonging to the same particle. Currently, several pattern-recognition approaches exist. Quantum computing can be used to reduce the initial combinatorial search using the 'quantum approximate optimisation algorithm'

developed by researchers at the Massachusetts Institute of Technology (MIT) in Cambridge, US. We are now studying the application of what is known as the 'variational-quantum-eigensolver algorithm' and implementing it on Intel's quantum simulator.

Quantum homomorphic encryption

The latest advances in machine learning and data analytics offer great potential for gaining new insights from medical data. However, data privacy is of paramount concern. Anonymisation via the removal of personal information is not an option, since medical records carry information that may allow the identification of the owner much more easily and securely than the name or the birth date. One possibility being studied is to encrypt sensitive information in such a way that makes data analytics possible without decryption. This is called homomorphic encryption. It is important to find an encryption strategy that is secure, while also ensuring it is possible to apply a large family of analytic algorithms to the data. While such encryption algorithms do exist, they require high-quality random numbers and they tend to be very demanding in terms of computing resources. Thus, this is a promising field of investigation for the utilisation of quantum computing.

The aims of the project are multiple: to transfer anonymized medical records protected by Quantum Keys, to develop a quantum homomorphic encryption (QHE) algorithm to apply on it and to analyse the data with QHE-friendly analysis tools (techniques based on machine learning or deep learning). The main project consists of four different parts, each realised in collaboration with different partners, both European and Korean: ID Quantique, Innosuisse (the Swiss Innovation Agency), Korea Institute of Science and Technology Information (KISTI), and the Seoul National University Bundang Hospital (SNUBH).

Quantum random number generator

We have recently established a collaboration with Cambridge Quantum Computing to test the performance of a new quantum random number generators and study its integration within simulation software used in high-energy physics.

RandomPower: evaluating the impact of a low-cost, robust 'true random power generator'

Researchers at the University of Insubria in Italy have invented a true random power generator (TRNG). This is based on the local analysis of the time series of endogenous self-amplified pulses in a specific silicon device. The principle has been validated with lab equipment and a low cost, small-form-factor board has been developed and commissioned with the support of an ATTRACT project. The board can deliver a stream of unpredictable bits at frequencies currently up to 1Mbps for a single generator, with the possibility to be scaled up. Randomness has been qualified through a test suite from the US National Institute of Standards and Technology, as well as beyond this. Together with the CERN openlab, the University of Insubria intends to evaluate the impact of the TRNG availability in a set of use cases, as follows:

- Modifying the Linux OS, replacing the embedded random number generation with the random power stream, in order to facilitate its adoption.
- Comparing the outcome of the application of generative adversarial networks using training sets guided by pseudo random number generators (PRNG) or the random power TRNG.
- Identifying classes of Monte Carlo simulations in high-energy physics where the use of PRNG could be particularly critical.

Moreover, the availability of a low-cost platform for high-quality random numbers may open up new possibilities in the use of homomorphic encryption, relevant for privacy-preserving data analysis; this will be thoroughly evaluated.

Next Steps

In 2020, we will continue to assess the merits of each of these lines of investigation. We will also continue our discussions with a range of companies and research institutes to identify areas for mutually beneficial collaboration. Where appropriate, we will work to formalise the investigations into full standalone projects.

In 2020, we will continue our discussions with a range of companies and research institutes to identify areas for mutually beneficial collaboration.

BIODYNAMO

R&D TOPIC 5: APPLICATIONS IN OTHER DISCIPLINES

Project coordinator:
Roman Bauer, Fons Rademakers

Technical team:
Lukas Breitwieser, Ahmad Hesam

Collaborator liaisons:
Uri Nevo, Marco Durante, Vasilis Vavourakis

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 and other institutions, supported by Intel. Together, we are working to develop a unique platform. This project is cofinanced by the CERN budget for knowledge transfer to medical applications.

ImmunoBrain Checkpoint, GSI Darmstadt and the University of Cyprus are also collaborating in BioDynaMo. Find out more about the project on the collaboration page*: biodynamo.org*.

Progress in 2019

We further integrated core CERN technologies into BioDynaMo. ROOT Notebooks provide life scientists with a web-based interface for (i) creating, running and visualising simulations; (ii) performing powerful analysis; and (iii) working interactively with a graphical user interface. Furthermore, users are now able to explore a large model parameter space on distributed computing infrastructures (e.g. cloud computers and computer clusters).

Based on feedback from our users, we improved BioDynaMo's API. Now, life-scientists can translate an idea into a simulation faster than before. We were also able to connect BioDynaMo with another simulator, to benefit from the strengths of agent-based and continuum-based modelling for cancer research.

In addition, we succeeded in integrating systems biology markup language (SBML) into BioDynaMo. This makes it possible for life scientists to reuse the large database of existing SBML models to simulate chemical reaction networks within each BioDynaMo simulation object.

Finally, BioDynaMo was featured in CERN's official teachers and students programme in 2019. Teachers and students from Dutch high schools learned through hands-on sessions how ICT technologies developed at CERN help in tackling challenges in the biomedical fields.

Next steps

BioDynaMo is currently able to simulate millions of cells on one server. To improve the performance yet further, we will focus on two main aspects. First, we will continue development on the distributed runtime, to combine the computational resources of many servers. Second, we will improve hardware acceleration to fully utilise (multiple) GPUs in a system. This will not only reduce runtime on high-end systems, but will also benefit users that work on a standard desktop or laptop.

Finally, thanks to new projects funded by the UK Engineering and Physical Sciences Research Council and the UK Medical Research Council in 2019, BioDynaMo has been extended for applications related to cryopreservation and the representation of realistic neural networks. Work in these areas will continue in 2020.

A full list of presentations and publications related to this project can be found on the CERN openlab website: cern.ch/go/RB9t.

The CERN members of the BioDynaMo technical team – Ahmad Hesam und Lukas Breitwieser – pictured in CERN's Data Centre.

FUTURE TECHNOLOGIES FOR MEDICAL LINACS (SmartLINAC)

**R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES**

SAMARA UNIVERSITY

Project coordinator:
Alberto Di Meglio

Technical team:
Yann Donon

Collaborator liaisons:
Dmitriy Kirsh, Alexander Kupriyanov, Rustam Paringer, Igor Rystsarev

Project goal

The 'SmartLINAC' project aims to create a platform for medical and scientific linear accelerators that will enable anomaly detection and maintenance planning. The goal is to drastically reduce related costs and unexpected breakdowns. The platform we develop will use artificial intelligence to adapt itself to different linear accelerators (linacs) operated in all kinds of environments.

Background

During a joint workshop held at CERN in 2017, involving the International Cancer Expert Corps and the UK Science and Technology Facilities Council, the need for simple-to-maintain-and-operate medical linacs was emphasised strongly. Maintenance can be one of the main sources of expenditure related to linacs; it is essential to reduce this cost in order to support the proliferation of such devices.

Following contacts with Samara National Research University in Russia in 2018, it was decided to create the SmartLINAC project. The university has a long history in the field of aerospace, which requires similar attention to fine detail and has led to the building up of expertise in big-data processing.

Progress in 2019

Following work to define the project's scope in 2018, as well as an initial feasibility study, the main project got underway in 2019. For the first stages of development within the project, data has been used from the Linac4 accelerator at CERN. In particular, we have used data from the 2 MHz radio-frequency source that is used to create the plasma; this presents periods of 'jitters' that influence the beam's quality.

By nature, these data sets are extremely noisy and volatile, leading to difficulties in interpretation and labelling. Therefore, the first research objective was to establish an appropriate data-labelling technique that would make it possible to identify 'jittering' periods. This has led to the creation of an anomaly-detection system that recognises early symptoms in order to make preventive maintenance possible. Several approaches based on statistics and neural-network technologies were used to solve the problem. These approaches are now being combined in order to offer a system that can be adapted to different sources.

The data has been shown to be extremely difficult for neural networks to categorise. Rather than using neural networks to detect anomalies themselves, we have therefore made use of them to define appropriate parameters for a statistical treatment of the data source. This will, in turn, lead to detection of anomalies.

Next steps

A first solution is already trained to function in the radio-frequency source environment of Linac4. Therefore, the first objective of 2020 is to start its on-site implementation and to set up continuous field tests. The next challenge will then be to consolidate our parameter-selection model and to test the technique on multiple data sources.

Publications:

- Y. Donon, A. Kupriyanov, D. Kirsh, A. Di Meglio, R. Paringer, P. Serafimovich, S. Syomic, Anomaly detection and breakdown prediction in RF power source output: a review of approaches, CEUR Workshop proceedings, 27th Symposium on Nuclear Electronics and Computing, Montenegro, 2019. *cern.ch/go/8g9c*

Presentations:

- Y. Donon, Smart Anomaly Detection and Maintenance Planning Platform for Linear Accelerators (3 October). Presented at the 27th International Symposium Nuclear Electronics and Computing (NEC'2019), Montenegro, 2019.

CERN's Linac4 linear accelerator is the first part of the LHC injection chain.

CERN LIVING LAB

R&D TOPIC 5: APPLICATIONS IN OTHER DISCIPLINES

Project coordinator:
Alberto Di Meglio

Technical team:
Taghi Aliyev, Jose Cabrero, Anna Ferrari

Collaborator liaisons:
David Manset (be-studys),
Marco Manca (SCImPULSE)

Project goal

The project goal is to develop a big-data analytics platform for large-scale studies of data under special constraints, such as information that is privacy-sensitive, or that has a varying level of quality, associated provenance information, or signal-to-noise ratio. Ethical considerations are also considered when necessary. This will serve as a proof-of-concept for federating and analysing heterogeneous data from diverse sources, in particular for medical and biological research, using ideas and expertise coming from CERN and the broader high-energy physics community.

Background

CERN is a living laboratory, with around 15,000 people coming to work at its main campuses every day. For operational purposes, CERN collects data related to health, safety, the environment, and other aspects of daily life at the lab. Creating a platform to collate and enable intelligent management and use of this data — while respecting privacy and other ethical and legal obligations — offers the potential to improve life at the lab. At the same time, such a platform provides an ideal testbed for exploring new data-analytics technologies, algorithms and tools, including ML/DL methods, encryption schemes, or block-chain-based ledgers. It also provides a natural bridge to collaborate with other scientific research domains, such as medical research and biology.

This project is being carried out in the context of CERN's strategy for knowledge transfer to medical applications, led by CERN's Knowledge Transfer group.

Progress 2019

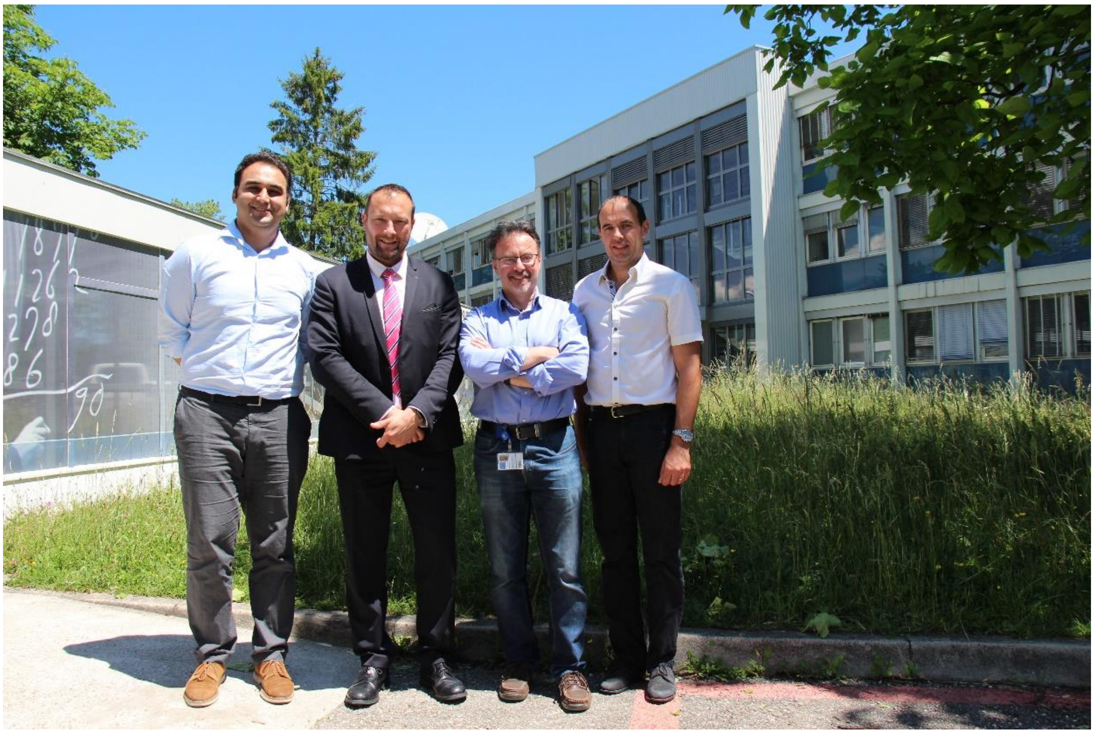
The CERN Living Lab project formally started in June 2019. A kick-off meeting was held with all project partners to discuss in detail shared interests and objectives. A second meeting took place in December 2019, focusing on the architecture and requirements for the big-data analytics platform. The platform architecture was defined and agreed. Also, four specific sub-projects were defined to address the data life-cycle in the presence of sensitive information, the data ingestion from foreign sources, the possibility of dynamically detecting and addressing the level of privacy protection required by different data transfer requests, and the use of homomorphic encryption as a possible privacy-preserving approach for cloud-based data analysis.

Next steps

In 2020, a proof-of-concept platform will be established, and a number of selected use cases will be deployed. Specifically, investigations will include classification and detection of the symptoms of Parkinson’s disease from wearable devices, optimisation of homomorphic encryption techniques for deep learning, and medical image analysis.

Presentations

- T. Aliyev, Meaningful Control of AI and Machine Ethics (7 June). Presented at Big Data in Medicine: Challenges and Opportunities, CERN, Geneva, 2019. *cern.ch/go/J7CF*
- A. Di Meglio, The CERN Living Lab Initiative (20 June). Presented at CERN Information Technology for the Hospitals, HUG, Geneva, 2019. *cern.ch/go/Fld8*
- T. Aliyev, Interpretability and Accountability as Necessary Pieces for Machine Ethics (2 July). Presented at Implementing Machine Ethics Workshop, UCD, Dublin, 2019. *cern.ch/go/7c6d*



Project members from be-studys and CERN meet at the Living Lab kick-off meeting in June, 2019.

HUMANITARIAN AI APPLICATIONS FOR SATELLITE IMAGERY

R&D TOPIC 5: APPLICATIONS IN OTHER DISCIPLINES

Project coordinator:
Sofia Vallecorsa, Federico Carminati

Technical team:
Taghi Aliyev, Yoann Boget

Collaborator liaisons:
Lars Bromley

Project goal

This project is making use of expertise in artificial intelligence (AI) technologies at CERN to support a UN agency. Specifically, we are working on AI approaches to help improve object recognition in the satellite imagery created to support humanitarian interventions. Such satellite imagery plays a vital role in helping humanitarian organisations plan and coordinate responses to natural disasters, population migrations, and conflicts.

Background

Since 2002, CERN has hosted UNOSAT, the Operational Satellite Applications Programme of UNITAR (The United Nations Institute for Training and Research) on the laboratory's premises. UNOSAT acquires and processes satellite data to produce and deliver information, analysis, and observations to be used by the UN or national entities for emergency response, to assess the impact of a disaster or a conflict, or to plan sustainable development in the face of climate change.

At the heart of this project lies the idea of developing machine-learning techniques that can help speed up analysis of satellite imagery. For example, predicting and understanding the movement of displaced persons by identifying refugee shelters can be a long, labour-intensive task. This project is working to develop machine-learning techniques that could greatly reduce the amount of time needed to complete such tasks.

Progress in 2019

Refugee camps often consist of more than 10,000 shelters and may need to be re-analysed several times in order to understand their evolution. Manual analysis typically leads to very high-quality output, but is very time-consuming. We have therefore worked with region-based convolutional neural networks to improve detection of new shelters in refugee camps, taking into account prior knowledge regarding the positions of existing shelters. The results were promising and the data pipeline created by our summer students has now been adapted and put to use by the UNOSAT experts. The retrained model yielded an average precision/recall score of roughly 80% and reduced the time needed for the task by a factor of 200 in some areas.

More recently, we also addressed the challenge of simulating synthetic high-resolution satellite images. High-resolution satellite imagery is often licensed in such a way that makes it difficult to share it across UN partners and academic organizations. This reduces the amount of image data available to train deep-learning models, thus hampering research in this area. We have developed a generative adversarial network (GAN) that is capable of generating realistic satellite images of refugee camps images. Our tool was initially based on a progressive GAN approach developed by NVIDIA. We have now developed this further, such that it can combine multiple simulated images into a cohesive larger image of roughly 5 million pixels. Several other lines of investigation — all related to AI technologies — are also being pursued within the scope of this project.

Next steps

Next year, we will pursue the initial work carried out on the GAN model in 2019 in a number of different directions. We will carry out a detailed performance study and will implement a distributed approach for parallel network training, as well optimising the use of computing resources. This should help us to reduce training time for the model and increase the maximum image size.

Publications:

- N. Lacroix, T. Aliyev, L. Bromley: Automated Shelter Recognition in Refugee Camps. Published on ZENODO, 2019. *cern.ch/go/v6rn*

Presentations:

- Y. Boget, ProGAN on Satellite images (15 August). Presented at CERN openlab summer student lightning talk session, Geneva, 2019. *cern.ch/go/P6NV*
- Y. Boget, S. Vallecorsa, Deep Learning for Satellite Imagery (24 September). Presented at IXPUG Annual Conference, Geneva, 2019. *cern.ch/go/m9n6*

This project is working to develop machine-learning techniques to aid understanding and predicting of the movement of displaced persons.

SMART PLATFORMS FOR SCIENCE

R&D TOPIC 5:
APPLICATIONS
IN OTHER DISCIPLINES

Project coordinator:
Alberto Di Meglio

Technical team:
Taghi Aliyev

Collaborator liaisons:
Marco Manca (SCImPULSE),
Mario Falchi (King's College London)

Project goal

The project goal is to design a platform that can analyse data collected from user interactions and can process this information in order to provide recommendations and other insights, thus helping improve the performance and relevance of user searches or learning objectives.

Background

Data-analysis systems often collect and process very different types of data. This includes not only the information explicitly entered by users ("I'm looking for…"), but also metadata about how the user interacts with the system and how their behaviour changes over time based on the results they get. Using techniques such as natural language processing (NLP) and smart chatbots, it is possible to achieve improved interaction between humans and machines, potentially providing personalised insights based on both general population trends and individual requests. Such a system would then be able to recommend further searches, actions, or links that may have not occurred to the user.

Such an approach could, for example, be used to design better self-help systems, automated first-level medical services, more contextual and objective-aware search results, or educational platforms that are able to suggest learning paths that address specific student needs.

This project is being carried out in the context of CERN's strategy for knowledge transfer to medical applications, led by CERN's Knowledge Transfer group.

Progress in 2019

The concept of the Smart Platforms project emerged in 2019 as a spin-off of the application of NLP techniques to genomic analysis in the GeneROOT project.

In 2019, a few initial discussions about possible applications were started in collaboration with educational institutes and public administrations, with the goal of developing the concept of smart chatbots that are able to improve human-machine interaction. As the project moved into the proof-of-concept phase, it became clear that the need to understand issues related to data-privacy and information sharing are still a critical roadblock for systems like this. The project was therefore merged into the CERN Living Lab, through which such concerns can be better addressed.

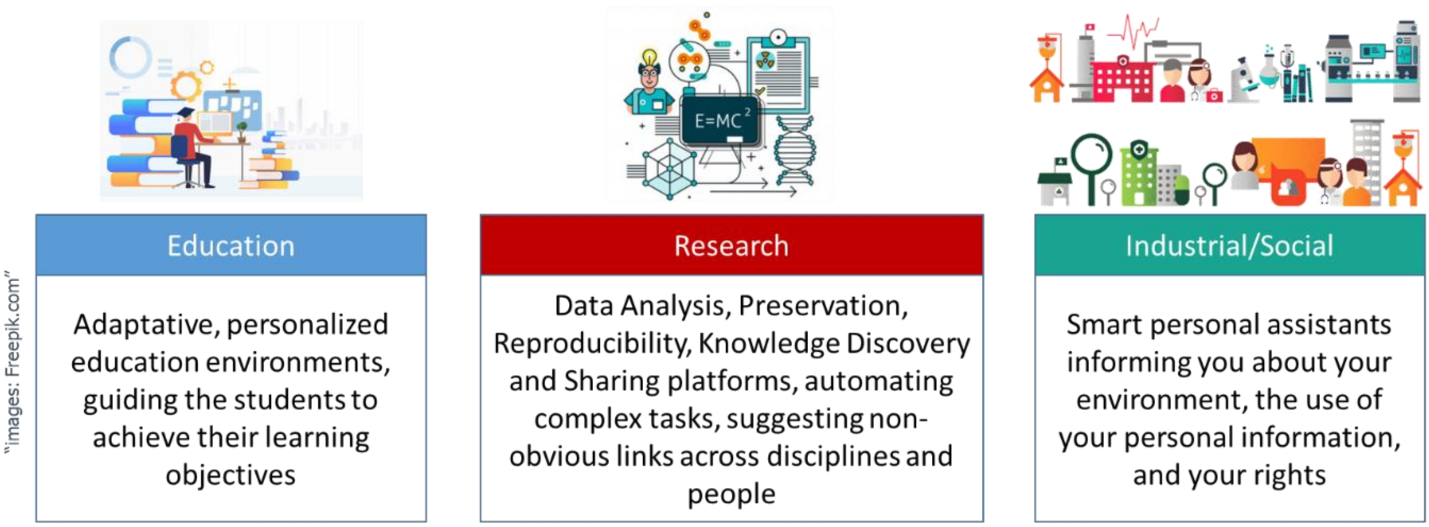
Next steps

The project has been merged into the CERN Living Lab as part of a general initiative to understand the implications of processing personal data and the related ethical constraints.

Presentations:

- A. Di Meglio, Introduction to Multi-disciplinary Platforms for Science, (24 January). Presented at CERN openlab Technical Workshop, CERN, Geneva, 2019. *cern.ch/go/XNt9*

.



A smart platform, of the kind outlined by this project, has a wide variety of potential uses.

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: 146 protocol visits were organised in 2019. 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". 45 press articles were published about our work over the course of the year.

Thirty workshops, lectures, visits and other events related to CERN openlab's work were held throughout the year. Here are three highlights:

- In June, CERN openlab co-organised a first-of-its-kind workshop on big data in medicine. The event marked the conclusion of a two-year pilot investigation into how CERN-developed technologies and techniques related to computing and big data could potentially be used to address challenges faced in biomedicine. The main goal of the workshop was to establish the terms for broader collaboration with the medical and healthcare research communities in future.

- In August, CERN openlab organised for Vivek Nallur of University College Dublin to give a lecture on AI and ethics. Nallur gave an overview of how AI technologies are being used in wider

Participants in the 2019 CERN openlab summer-student programme.

Vivek Nallur, assistant professor at the School of Computer Science at University College Dublin in Ireland, was invited to give a talk at CERN entitled 'Intelligence and Ethics in Machines – Utopia or Dystopia?'.

society today and highlighted many of the limitations of current systems. In particular, he discussed challenges related to the verification and validation of decisions made, the problems surrounding implicit bias, and the difficulties of actually encoding ethical principles.

- In September, the annual meeting of the Intel eXtreme Performance Users Group (IXPUG) was organised by CERN openlab and took place at CERN's Globe of Science and Innovation. The event provided an open forum, through which industry experts and researchers could share best practices and techniques for maximising software efficiency.

Further information about our past and future events 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 was also presented to visitors at the 2019 CERN Open Days.

CERN openlab was also presented to visitors to the IT department at the 2019 CERN Open Days. These took place on 14-15 September and saw over 75,000 people tour the laboratory.

A scholastic summer

In 2019, 40 students — representing 19 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 2019'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 2019, 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. The 2019 Webfest, which was again supported by CERN openlab, featured 14 fascinating and original projects, related to topics such as data visualisation, web games, education, and networking.

The event is open to all at CERN. However, this year's winning team consisted of four students participating in the CERN openlab summer-student programme. Their project, "CERNAccess", is a web application that translates sign language to text.

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 13 and 15 August, the students presented their work in two dedicated public "lightning talk" sessions. In five-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 of the two sessions were selected by a panel of judges. The winners were as follows:

- 1st: Hamza Javed, Fast inference on FPGAs (field-programmable gate arrays) for trigger systems in high-energy physics.
- 2nd: Raghav Kansal, Deep graph neural networks for fast HGCAL (high-granularity calorimeter) simulation.
- 3rd: Bartlomiej Borzyszkowski, Neuromorphic computing in high-energy physics.

The judges also picked out presentations by Priyanka Mathur and Riccardo Maganza for special mention.

The winners of the lightning-talk sessions receiving their prizes. From left to right: Alberto Di Meglio, head of CERN openlab; Hamza Javed, 1st price; Raghav Kansal, 2nd price; Bartlomiej Borzyszkowski, 3rd price; Priyanka Mathur, special mention; Riccardo Maganza, special mention; Frédéric Hemmer, head of the CERN IT department.

07
FUTURE

Next steps

We are working with leaders from academia and industry to tackle tomorrow's ICT challenges today.

More and more of science is becoming ever more data driven. This is increasing the importance — and value — of working together with industry, across research fields, to drive innovation in ICT.

Advances in computing technology are fundamentally changing the way we do science. Back in 2007, Jim Gray of Microsoft Research outlined what he called the fourth paradigm of scientific research. The idea is that science has gone through three main paradigms — *experimental science,* a thousand years ago*; theoretical science,* over the last few hundred years; *computational science*, over the last few decades — and is now entering a fourth: *data-intensive science*.

Leaving aside Popper, Kuhn, and the philosophy of what does — or doesn't — constitute *science*, the idea clearly has some practical merit. It is undoubtedly true that data volumes — both in science and other areas — are exploding, and this is having an effect on the way we do research. Gray argued that scientists risk being overwhelmed by a flood of observational data and that we need a new generation of scientific computing tools for management, sharing, visualisation, and analysis. Without these, researchers will struggle to cope, let alone be able to extract maximum research value from this flood of data.

It can be argued that a positive feedback loop is in place, with advances in computing technologies leading to a deluge of data, which in turn necessitates further innovation in computing, and so on and so forth.

Thanks to the cloud and improvements in networking technologies, we now have pervasive computing. Coupled with the data deluge, this has the potential to unlock exciting new possibilities for science. However, it's important to have the right technology in place, and at the right time. Standard classical computing architectures alone are no longer enough; through technologies like artificial intelligence and neuromorphic computing, we're trying to mimic the way the human brain makes sense of large volumes of data.

CERN openlab, with its deep connections to many of the ICT companies at the forefront of these developments, is ideally placed to help ensure the CERN research community derives maximum benefit from these innovations. Equally, CERN openlab can play an important role in helping to share the knowledge experts at CERN have developed in these areas with those from other research communities.

Much of CERN openlab's work over the past few decades has focused on evolution rather than revolution: we have often worked to help speed up existing approaches to scientific computing. Today, however, there is an opportunity to explore uncharted territory, with topics like quantum computing and artificial intelligence already being investigated through joint R&D projects with industry leaders.

Of course, no one is relying on such technologies to provide easy solutions to the computing challenges posed by the LHC's ambitious upgrade programme. Nevertheless, it is important to explore these technologies today in order to build expertise and to help us judge when (or if) technologies become ripe for adoption by our community.

We are working hard to strike the right balance between support for classical computing approaches and exploration of these new technologies. This will be at the forefront of our minds as we conduct investigations that will help us to lay the groundwork for the new three-year phase of CERN openlab that will begin in 2021. We look forward to discussions with innovators — from academia and industry — over the coming months as we finalise our vision for how CERN openlab can best support the LHC research community in the early years of this new decade.

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 2019 Annual Report

Editor: Andrew Purcell

Additional subediting: Hans Baechle and Anastasiia Lazuka

All images provided by the CERN Audio-Visual Production Service except:
Veronika McQuade (page 2); Christian Beutler (page 5); European Space Agency (page 11); CERN openlab communications office (pages 14, 31, 43, 67, 73, 74, 75); DEEP Projects (page 23); ATLAS Collaboration (pages 47, 55, 59); Thomas Mc Cauley / CMS collaboration (page 39); CNES (page 69).

ISBN 978-92-9083-583-7
Published by CERN

CERN openlab