e-Infrastructures Reflection Group (http://e-irg.eu:8080/news-blog? p_p_id=49&p_p_lifecycle=1&p_p_state=normal&p_p_mode=view&_49_struts_action=%2Fmy_

News (…   /   ... / News bl…   /   CERN s…

Navigation

News (http://e-irg.eu:8080/news)

- **News blog (http://e-irg.eu:8080/news-blog)**
- Magazine (http://e-irg.eu:8080/magazine)
- Video news (http://e-irg.eu:8080/video-news)
- Subscribe / Unsubscribe (http://e-irg.eu:8080/subscribe-unsubscribe-mailing-list)

This news blog provides news about the e-IRG and related e-Infrastructure topics.

⦿ (http://e-irg.eu:8080/news-blog? p_p_id=33&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column- 2&p_p_col_count=1&_33_struts_action=%2Fblogs%2Fview)
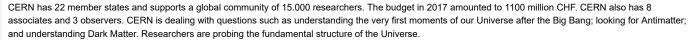
CERN seeking answers and collaboration with industry to future challenges in data processing and storage

26/07/18 09:26

The first keynote speaker at ISC'18 in Frankfurt, Germany, was Maria Girone, CTO at CERN, the European Organisation for Nuclear Research in Geneva, Switzerland. She talked about the challenges that are coming up for CERN in the next decade and the way these are being tackled already today with the newest technologies, including machine learning, artificial intelligence, advanced simulation, HPC, and quantum computing.

At CERN, accelerators have been built and are being operated. Data is processed wordwide on the Grid. The Large Hadron Collider (LHC) has huge computer challenges. Solutions for these challenges are trying to be found using scientific technologies and artificial intelligence and through interactive work with other scientists.

CERN is the European Laboratory for Particle Physics, founded in 1954 in Geneva, Switzerland, under the slogan "Science for Peace". The laboratory straddles the Franco-Swiss border near Geneva in a large complex consisting of offices and installations. CERN has 22 member states and supports a global community of 15.000 researchers. The budget in 2017 amounted to 1100 million CHF. CERN also has 8 associates and 3 observers. CERN is dealing with questions such as understanding the very first moments of our Universe after the Big Bang; looking for Antimatter; and understanding Dark Matter. Researchers are probing the fundamental structure of the Universe.

According to Maria Girone CERN's mission involves research, technology, education and collaboration. Researchers are trying to advance the frontiers of knowledge by revealing the secrets of the Big Bang and searching for what was the matter. They are developing new technologies for accelerators and detectors. CERN is committed to information technology in using the Web and the Grid. Other areas of interest are diagnosis and therapy in medicine. CERN is training the scientists and engineers of tomorrow and unites people from different countries and cultures.

Maria Girone went on to talk about CERN's accelerators. The LHC is the world's largest and most powerful particle accelerator. There are four interaction points, known as CMS, Alice, Atlas and LHCb. The LHC is built around 100m underground and has a circumference of 27 km. The particles are accelerated to close to the speed of light. Maria Girone showed that the LHC is a machine of records in different aspects.

Looking at the future, Maria Girone told the audience that the LHC is a schematic of an 80-100 km long tunnel. The High Luminosity (HL)-LCH is a ongoing project. FCC is a design study. The second LHC operation run 2 will last about 20 years. A new collider is coming.

Maria Girone elaborated about the experiments at CERN. The detectors are like gigantic digital cameras built in cathedral-size. Experiments are run by collaborations of scientists from institutes all over the world. Scientists deal with aspects of the LHC in their own laboratory. Two general-purpose detectors cross-confirm discoveries, such as the Higgs boson.

Atlas is 46m long with a diameter of 25m. It weighs 7000 tonnes. There are 100 million electronic channels and 3000 km of cables. CMS is 22m long and has a 15m diameter. It weighs 14.000 tonnes. It is the most powerful superconducting soneloid ever built. Alice studies the "Quark Gluon Plasma", the state of matter which existed moments after the Big Bang. LHCb studies the behaviour difference between the b quark and the anti-b quark.

At CERN, a large amount of data is being collected about the collision between proton and neutron. There are up to about 1 billion particle collisions. This can generate up to a petabyte of data per second. The data is being filtered in real time, selecting potentially interesting events, the so-called trigger. Data is being generated at a rate of 40 million times per second: PB/s; at a rate of 100.000 selections per second: TB/s; and at a rate of 1000 selections per second: GB/s.

The CERN data centre processes hundreds of petabytes of data every year. CERN's data centre in Meyrin is the heart of the laboratory's computing infrastructure. The Meyrin Centre is located in Switzerland and has 300.000 processor cores; 180 PB on disk; and 230 PB on tape. There is a data centre in Hungary too. The two centres are connected by three Gb/s fibre-optic links.

Physicists must sift through the 30-50 PBs produced annually by the LHC experiments. This is done in real time. They run complex algorithms to achieve structured data and they are running simulations too.

The Tier-0 centres are located at CERN and in Hungary where the data recording, reconstruction and distribution happen. Tier-1 is offering permanent storage, re-processing, and analysis. Tier-2 is providing simulation, and end-user analysis.

The Worldwide LHC Computing Grid (WLCG) gives thousands of physicits across the globe near real-time access. The Worldwide LHC computing Grid integrates computer centres worldwide. There are 170 computing centres in 42 countries, also including Antarctica. CERN handles about 1 million CPU cores, 1 EB of storage, 2 million jobs/day, 10-100 GB/s links, 340 GB/s transatlantic traffic and 3 PB are moved per day.

Data organisation, management and access in WLCG is huge business, as Maria Girone told the audience. Making hundreds of petabytes of data accessible globally to scientists is one of the biggest challenges of WLCG. CERN is now consolidating the storage infrastructure by sharing the operations of the storage infrastructure to save on the costs.

The LHC has been designed to follow a carefully set out programme of upgrades. In run 3, Alice and LHCb will be upgraded. In run 4, Atlas and CMS will be upgraded. The other planned upgrades - run 5 and 6 - will greatly increase the scientific reach. The High Luminosity-LHC will take place in run 4 to 6.

The rate of new physics is 1 event in 10^{12}. Selecting a new physics event is like chosing 1 grain of sand in 20 volley ball courts. More collisions help physicists to observe rare processes and study them with greater precision.

The HL-LHC will come online around 2026 enabling more collisions and more complex data.

CMS had an event from 2017 with 78 reconstructed vertices. ATLAS will have a simulation for HL-LHC with 200 vertices. LHCb and ALICE will move offline processing closer to the online data collection chain. The goal is to perform processing and data analysis in near real time. Solutions are under investigation. There will be new HLT farms for Run 3. It will be a flexible and efficient system with ambitious PUE ration. The ALICE and LHCb experiments will increase their data acceptance rates for Run 3.

By Run 4, the detectors will become more granular and more radiation hard, Maria Girone explained. Reconstructing more particles with more granular detectors will be computationally more expensive. The ATLAS and CMS experiments will be significantly upgraded for the HL-LHC. Using current techniques, the required computing capacity increases by 50-100 times. This is an impressive amount of resources comparing to the budget.

For data storage, Maria Girone explained, it is vital to explore new technologies and methodologies since CERN is reaching exascale of data.

Closing the resource gap in the next decade requires close collaboration with industry to enable new technology, evolution and improvements, software innovation, new architectures, techniques and methods, as well as improvements in hardware. CERN openlab is a unique collaboration, a science-industry partnership, fostering research and innovation. It is involved in the training of young scientists and the dissemination of new technologies. There are three main areas of research and development involving the scale out of capacity with public Clouds, HPC, and new architectures. Maria Girone told the audience that data centre performance will be increased with hardware accelerators and optimized software.

New techniques will be used including Machine Learning, Deep Learning, and Advance Data Analytics.

As for the data centre technologies and infrastructures, CERN is faced with a resource gap of an enormous magnitude so it is of great importance to fully exploit the available resources. CERN is one of the early adopters and largest contributors to OpenStack. Maria Girone explained that 90% of the resources are provided through a private Cloud. This allows for flexible and dynamic deployment. CERN is moving to containrers for even more flexibility.

Experiments have demonstrated that it is possible to elastically and dynamically expand production resources to commercial Clouds. Large-scale tests are happening with commercial Clouds. There is a joint procurement of R&D Cloud services for scientific research involving Helix Nebula Science Cloud. The Joint Pre-Commercial Procurement involves the following procurers: CERN, CNRS, DESY, EMBL-EBI, ESRF, IFAE, INFN, KIT, STFC, and SURFSara.

HPC technologies are significant resources. They are being tested by the experiments and optimized for highly parallel applications.

ATLAS reached more than 200k traditionalx86 HPC cores for the simulation of workflows. All experiments are exploring the use of heterogeneous HPC architectures. CERN will partner with EU-PRACE for optimizing the use of HPC resources. There are demonstrations with large-scale, dedicated HPC resources too. CERN is a partner of DEEP-EST, a blueprint project for heterogeneous HPC systems. The DEEP-EST project is a EU-funded co-design project, that is testing modular supercomputing architecture, offering a software environment and testing applications.

HEP at CERN has a vast investment in software with a significant effort to make efficient multi-threaded and vectorized CPU code. Accelerated computing devices such as GPUs and FPGAs offer a different model with the complexity of heterogeneous architectures. CERN is simultaneously exploring lower performance but lower power alternatives like ARM. Software optimization can also gain factors in performance.

The landscape is shifting at all levels. The environment has become much more complex with more opportunities. CERN has to make best use of the storage, the long-time memory, and high-performance accelerators like FPGAs, according to Maria Girone.

Higher data rates require more selective triggering and faster reconstruction. The LHCb is investigating FPGAs and GPUs to allow reconstruction of 5GB/s of events in real time. CMS is porting heavy "offline" tasks to real time processing for HL-LHC. It will integrate GPUs in the HLT farm to give high-qualtity reconstruction in 100 milliseconds latency as opposed to tens of seconds.

CERN will be exploiting co-processors for software-based filtering and real time reconstruction.

Quantum computing is also on the horizon, Maria Girone told the audience. CERN openlab is engaging in quantum computing with industry. This can substantially speed up training of deep learning and combinatorial searches. It is well suited for fitting, minimization and optimization and can directly describe basic interactions as well as lattice QCD calculations.

Machine Learning will be used for monitoring, automation and anomaly detection. Experiment and acclerator operations have similar challenges to industrial applications. Detectors and accelerators infrastructure health needs to be monitored. The quality of produced data needs to be validated. The resource usage needs to be optimized. CERN is working with industry to accomplish this.

With the current software and computer capacity an event like HL-LHC takes 10s of seconds, Maria Girone explained. CERN will examine the detector hit information and use 3D image recognistion techniques to identify objects. Physics objects will be recognized from learned patterns. This might dramatically increase the speed for reconstruction. Image recognition will thus be explored.

Simulation is one ot the most resource-intensive computing applications, Maria Girone explained. The main R&D areas consist in adapting the existing code to new computing architectures and in replacing complex algorithms with deep learning approaches such as fast simulation.

Researchers at CERN will be looking at adversarial networks to improve speed without giving up accuracy. CERN is collaborating with other communities who share similar technologies, such as the Square Kilometre Array (SKA). The SKA observatory's two telescopes will enable astronomers to study the sky in unprecedented detail. The first phase will be operational in the mid 2020s; the observatory will function for 50 years.

Maria Girone said that there will be a joint exascale data-storage and processing challenge between HL-LHC and SKA. She also talked about accelerating innovation and knowledge transfer to medical applications. CERN-MEDICIS is involved in the production of innovative isotopes for medical research. There is accelerator design for future hadron therapy facilities, medical imaging, dosimetry, and computing and simulation for health applications.

CERN has been pushing the boundaries of knowledge and technologye for more than 60 years. She concluded by citing Tom Hanks: "Magic is not happening at CERN, magic is being explained at CERN."

By Leslie Versweyveld | 67 Views,

Previous (http://e-irg.eu:8080/news-blog/-/blogs/431675?_33_redirect=http%3A%2F%2Fe-irg.eu%3A8080%2Fnews-blog%3Fp_p_id%3D33%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-2%26p_p_col_count%3D1%26_33_struts_action%3D%252Fblogs%252Fview)
Next (http://e-irg.eu:8080/news-blog/-/blogs/431904?_33_redirect=http%3A%2F%2Fe-irg.eu%3A8080%2Fnews-blog%3Fp_p_id%3D33%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-2%26p_p_col_count%3D1%26_33_struts_action%3D%252Fblogs%252Fview)