European Network on High Performance and Embedded Architecture and Compilation

HiPEAC blog

« Back to press room (/press/blog/)

Jan. 17, 2018

HiPEAC blog: 'My family has traditionally practised law, and I have continued that tradition by studying the laws of the universe' - Maria Girone, CERN openlab

Experiments undertaken at the Large Hadron Collider at CERN, the European Organization for Nuclear Research, generate vertiginous amounts of data. At the HiPEAC conference in January 2018, CERN openlab Chief Technology Officer Maria Girone will give a keynote talk on the computing challenges behind this incredible tool for scientific research. In this interview, Maria sets out some of these challenges and gives us a glimpse into the future of high-energy physics research at CERN.

What first got you interested in physics? What is it about high-energy physics that fascinates you?

I was interested in being a scientist as early as primary school. I wanted to understand how the world came to be, how it works, and what it is made of. I explored the study of many sciences while I was growing up before choosing to concentrate on physics, and later high-energy physics. My family has traditionally practised law, and I have continued that tradition by studying the laws of the universe.

What are some of the specific demands which this field places upon computing?

High-energy physics (HEP) has long been a leader and a driver of computing for science. There are computing contributions to the LHC programme from all the populated continents in the world. HEP developed the largest distributed computing infrastructure for science called the

Worldwide LHC Computing Grid (WLCG). Tools, interfaces, and protocols all needed to be developed to make the global system function. HEP has some of the largest scientific datasets that place high demands on storage, networking, and data access. Processing this data requires hundreds of thousands of processor cores working in concert across many sites to process collected data and generate simulations.

What were some of the challenges involved in the computing coordination of the Compact Muon Solenoid (CMS) experiment at the LHC?

The CMS experiment is one of the two large general-purpose experiments on the Large Hadron Collider. Being the CMS computing coordinator (CC) is a challenging job. The CC is responsible for the global computing system and all of the sites and services. The CC schedules the execution of all the centrally organized workflows and makes sure the huge analysis community has the services and the processing infrastructure needed to maximize the physics output.

My tenure as CMS computing coordinator was during a transition period: I took over at the beginning of the LHC's first long shutdown (a planned two-year period for upgrade work and maintenance). This shutdown is one of the few opportunities we have for significant changes in the computing infrastructure. I supervised the deployment and commissioning of the CMS data federation; this fundamentally changed how data could be accessed by processing re-sources and users. In addition, there were improvements in the analysis submission infra-structure and workflow management systems to commission. I was also the CC for the beginning of the LHC's second run (Run 2); the LHC ran at higher luminosity — and hence with more collisions — than ever before.

Since 2015, I have been working as the CTO of CERN openlab, another challenging and exciting role. CERN openlab is a public-private partnership between CERN and leading ICT compa-nies. Its mission is to accelerate the development of cutting-edge ICT solutions for the worldwide LHC community and wider scientific research.

Can you give us an analogy to help us understand the scale of the data generated at CERN?

In 2017, the CERN data centre recorded over 70 petabytes (PB) of data to tape, with 40 PB coming from the LHC experiments. At the end of 2017, the total amount of data permanently archived on tape in the CERN data centre stood at 230 PB; this is the equivalent of around two millennia of Blu-ray-quality video.

What are some of the main challenges for capturing, storing and processing these data volumes?

There are a lot of challenges for data handling and processing in HEP. The first relates to the trigger systems: despite the huge amount of data we collect, many orders of magnitude more data are rejected by the triggers (typically, roughly only one event in a million is selected). One of our biggest challenges is data mining: how to avoid missing an expected signal by not selecting it with the trigger and how to find new physics in the events we collect.

Once data is collected, the challenges involve data management and data access. There are nearly 200 sites in the WLCG and the appropriate data needs to be accessible to make the computing centres useful. Petabytes of data are moved weekly, verified against data corruption in transfer, and made accessible to applications for processing and analysis.

What can we expect from the High Luminosity Large Hadron Collider (HL-LHC)? What new demands will the upgrade place upon the computing infrastructure?

The HL-LHC is a new computing frontier for HEP. The trigger rates for the LHC experiments increase by almost an order of magnitude and the number of interactions per crossing will be 150-200, which is up a factor of five from today. The reconstruction, data-analysis and data-storage challenges will require nearly 100 times more resources. Even with technology improvements expected before the start of the HL-LHC in 2026, we currently foresee about a factor of 10 resource gap in what can be achieved without funding increasing. The experiments, the WLCG, and CERN openlab are all working on research and development projects to utilize more efficient techniques, to migrate to specialized hardware for specific applications, and to adopt new facilities and architectures to improve our efficiency. CERN openlab recently published a white paper identifying 16 key ICT challenges faced as we approach the HL-LHC era.

In this news

(/network/institutions/7832/cern/)

About HiPEAC

Since 2004, the HiPEAC (High Performance and Embedded Architecture and Compilation) project has provided a hub for European researchers in computing systems; today, its network, the biggest of its kind in the world, numbers around 1,500 specialists. The project offers training, mobility support and dissemination and recruitment services, along with numerous networking facilities to its members. The latest incarnation of the project, HiPEAC 5, was launched on 1 December 2017 and is delivered by 13 partners, led by Ghent University. It is a Coordination and Support Action funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 779656.

Press contact: Madeleine Gray (mailto:communication@hipeac.net), Communication Officer

Share

Tweet

The HiPEAC project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 779656.