SCIENCENODE™

Home   **Archive**   **Contribute**   **Sponsor**

About   **Give Now**

Handling big data to understand antimatter at CERN's LHCb experiment

This year's International Supercomputing Conference (ISC'14) in Leipzig, Germany, is now just one month away. iSGTW speaks to Niko Neufeld ahead of his talk at the event, 'The Boson in the Haystack', which will take place during the session on 'Emerging Trends for Big Data in HPC' on Wednesday 25 June.

Posted on MAY 21 2014 6:49AM

Andrew Purcell
European editor

"In total there are about 50 million packets injected in our data-acquisition network every second!"

Neufeld co-designed the data-acquisition system of the LHCb detector, which is one of the four large experiments on the Large Hadron Collider (LHC) at CERN, near Geneva, Switzerland. The 'b' in LHCb stands for 'beauty', which is a sub-variety of a type of particles known as quarks. By studying the beauty quarks thrown out by particle collisions within the LHC, researchers at the LHCb experiment are able to investigate the slight differences that exist between matter and antimatter.

Niko, how did you first become involved with scientific computing?

I originally studied physics, since I was very interested in understanding how matter works at
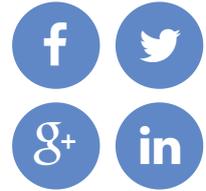
Tags

analysis

beauty

big data

CERN

CERN openlab

DAQ

data acquisition

International Supercomputing Conference

ISC'14   IT

Large Hadron Collider

LHC   LHCb

Niko Neufeld

Physics

quark

the most fundamental level. Once I came to CERN, my role involved a lot of computing and I gradually became more and more exposed to the technical aspects of the work done at the laboratory, particularly the data acquisition. Of course, I'd already done a lot of computing-related work as a student and was keen to pick this up again as I was always what you might today call a geek - I played and experimented a lot with computers when I was younger.

What exactly does your current role with the LHCb experiment entail?

I'm mainly responsible for the upgrade of LHCb's online computing system. That means all of the computing infrastructure that is related to transporting data from the detector, filtering it, temporarily storing it, and then sending it to mass storage. My job essentially ends once the data has left our experimental facility.

You currently get data coming off the LHCb particle detector at a rate of around 70 gigabytes per second. What are the challenges related to dealing with such a high volume of data?

The challenges were tougher a few years ago, but Moore's law always helps you to deal with these things over time. The problems have mainly been related to the traffic pattern. In a data-acquisition system, you have a rather uncommon network: it's very different to a typical campus network, where you normally have more randomized, many-to-

many communication. Obviously you still have hotspots with campus networks, and you have a server and multiple clients, too. However, a data-acquisition system is very different: at LHCb we have about 400 data sources geographically distributed over a large detector, and each data source has just a piece of an overall particle collision event. These pieces need to be brought together, in order to run a physics algorithm that will decide whether or not this data is interesting, and thus if it is worth keeping. The challenges come from the fact that we currently get new event-data every microsecond. We have 400 data packets running through our network to the same destination at exactly the same time, which puts tremendous stress on the network devices. In total there are about 50 million packets injected in our data-acquisition network every second!
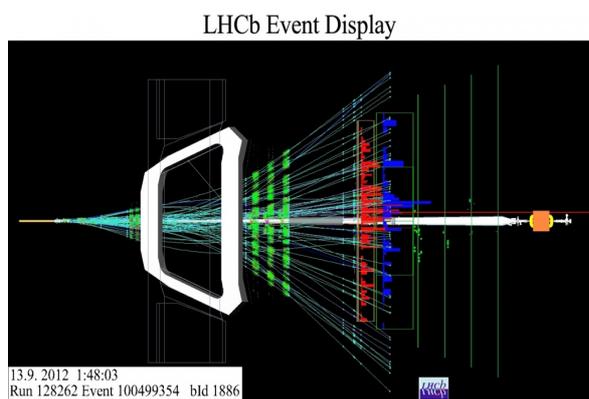
Event display of a proton-lead event in the LHCb detector, showing the reconstructed tracks of particles produced in the collision. The proton beam travels from left to right. Image courtesy CERN.

Upgrade work carried out on the LHC in 2018 will see the rate of data coming off the detector increase 50 fold. What does this

mean for the data-acquisition challenges you face at the LHCb experiment?

In addition to the various custom protocols we've designed, we currently use expensive, telecommunications-class hardware to help solve our networking problems. The issue here is that using such hardware just won't scale financially for these massively increased data rates. Consequently, our plan is to use upcoming data-center technologies - whereby you have low-latency high-bandwidth interconnects - for data-acquisition. The telecommunications-class hardware we currently use is really overkill, since we're only using it for one of its features, namely its buffering capacity.

There are also going to be major challenges around the data-filtering process, too...

I believe that LHCb - along with some of the other LHC experiments - is planning to move the first stage of its data-filtering process, known as the 'level-1 trigger' from a hardware- to a software-based system. Are you involved in this work?

Yes, that's right. Traditionally, we've used hardware-based systems for the first stage of data filtering, since the data rates coming off the detector are simply too high for us to be able to use software running on commodity processors. However, the hardware-based system has significant limitations, in particular that you can only filter the data using a limited set of physics algorithms. High-bandwidth network interconnects

now mean that the radical solution of LHCb switching to a completely software-based data-filtering system is now possible for the first time. The software system will be more finely grained, meaning that we will increase our yield of useful data by at least a factor of two, and possibly up to a factor of five. It will also permit more flexibility for the experimentalists.

What can people working on - or hoping to work on - other 'big science' projects learn from coming to your talk at ISC'14 next month?

I want to make it clear to people that collecting a very large amount of data, and pushing it through a computing infrastructure for filtering and processing, can be done surprisingly cost-efficiently. Obviously, there's no silver bullet and you have to always look at the specifics of the problems you are facing. But if one looks at the technologies at hand in an unprejudiced manner, there are usually very cost-effective solutions to be found. This can be much more cost-effective than simply following the well-trodden path, and buying whatever expensive equipment is in fashion at the time.

Here at CERN, we're able to achieve things with very little custom-built hardware. We're not shy about talking to industry, or about integrating industrial components - albeit ones which may have been adapted slightly to meet CERN's specific needs - into our systems. CERN openlab is very important in this: it is vital that we are able to

communicate with the leading IT companies that really define the directions in which major technologies evolve. Talking to these companies and getting a feeling of what is going happen over the coming years is really important for CERN's long-term planning. CERN openlab is an excellent facilitator and is really appreciated by those of us working on the LHC experiments.

Finally, what's the appeal of ISC'14 for you personally? What are your main reasons for wanting to attend this conference?

The ISC events are always a good opportunity to find out about upcoming trends in the community and make new contacts. There are plenty of opportunities to talk to both people from IT companies and from other research domains. At this year's event, I plan to focus mostly on accelerators, but I'm also keen to find out more about developments in new, upcoming server technologies. In my view, even if speakers make their talks available online, you can't beat actually being there: there's great value in attending in person.

Join the conversation

Contribute   ☰

Do you have story ideas or
something to contribute? **Let
us know!**

OUR UNDERWRITERS

Thank to you our underwriters, who have
supported us since the transition from
International Science Grid This Week (iSGTW)
into Science Node in 2015. We are incredibly
grateful.

View all underwriters

CATEGORIES

**Advanced
computing**

Research networks

Big data

Tech trends

Community building

CONTACT

Science Node
Email:
editors@sciencenode.o
Website:
sciencenode.org

[IU GIVE NOW]

Copyright © 2022 Science Node ™ | **Privacy Notice** | **Sitemap**

Disclaimer: While Science Node ™ does its best to provide complete and up-to-date information, it does not
warrant that the information is error-free and disclaims all liability with respect to results from the use of the
information.