Tech / #InTheCloud

SEP 12, 2016 @ 05:00 AM 59 VIEWS

CERN Tests Data Exploration Using Big Data, Analytics, And The Cloud

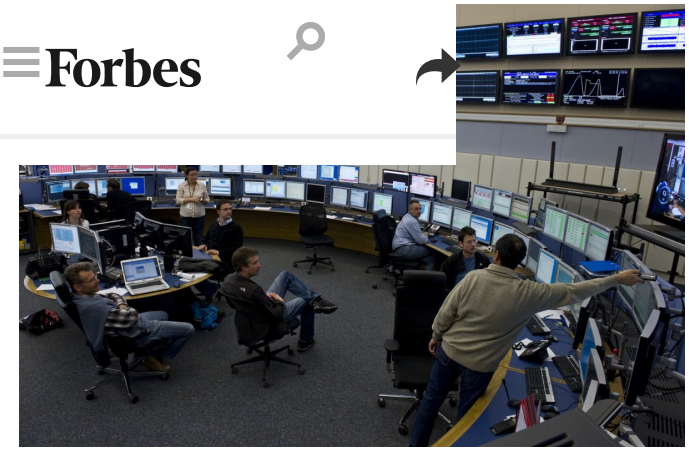
OracleVoice

Simplify IT, Drive Innovation FULL BIO

Sasha Banks-Louie, Oracle

For more than half a century, scientists at the CERN particle physics laboratory (more formally, the European Organization for Nuclear Research) have been probing the inner workings of the universe. Today, the organization, which was the birthplace of the World Wide Web, is assessing the potential of the cloud to make its research IT infrastructure more scalable and economical to operate.

Since 2008, scientists from around the globe have been using the largest and most powerful particle accelerator in the world, CERN's Large Hadron Collider (LHC), to create conditions similar to the Big Bang, and from their studies, gain a better understanding of how our universe works.



Scientists at the CERN particle physics laboratory have been using the world's most powerful particle accelerator to better understand how our universe works. (Photo courtesy CERN openlab.)

Just four years after the LHC was inaugurated, researchers discovered the Higgs boson, previously the last unverified piece of what is known as the Standard Model of particle physics. Nevertheless, the standard model explains only about 5% of how the universe actually works. Thus, as well as working to better define the characteristics of the Higgs boson, CERN and the worldwide particle physics research community are now hoping to make new discoveries that may shed some light on the rest.

In order to achieve this, researchers are analyzing huge numbers of subatomic particle collisions, which create many petabytes of raw data. Over the coming years, this analysis will require more computing power and storage capacity than CERN's data centers—and budget—can currently handle.

CERN's research is largely funded by 22 member countries, which together contribute roughly $1.1 billion annually. The proportion of this dedicated to IT has been largely flat for the past 10 years.

Oracle: Is The Data Scientist Era Coming To An End?

Oracle: Java (New Big Data Opportunities

Together, the LHC experiments produce more than 30 petabytes of data per year. CERN's Data Centre (including its remote extension in Budapest, Hungary) provides 250 petabytes of disk storage space and around 200,000 computing cores. Analysis of the physics data is made possible by a global network of more than 170 computer centers known as the Worldwide LHC Computing Grid. Each day, more than 2 million jobs run on this network.

"Over the next 10 to 15 years, we expect the total amount of data produced by the experiments on the LHC to increase significantly," says Alberto Di Meglio, head of CERN openlab, a public-private partnership that accelerates the development of cutting-edge solutions for the worldwide LHC community and wider scientific research network. Through this initiative, CERN collaborates with leading ICT companies and research institutes.

And that's just the data coming from the LHC. A successor to the LHC, known as the Future Circular Collider (FCC), is being studied. It would be around four times larger and produce even more data still.

In addition to managing its own on-site data centers, Di Meglio believes CERN can efficiently pursue a hybrid model:

orage services
e cloud is
portant
component of our research infrastructure," says Di Meglio. However, managing, storing and analyzing data in the cloud on the scale required by CERN can be very complex.

CERN openlab Uses Big Data Discovery

CERN is also working with Oracle Big Data Discovery to see how it can be used to more efficiently and intelligently analyze technical engineering information produced by approximately 50,000 sensors and other metering devices the center uses to capture operational data from its accelerator complex. The data is analyzed to ensure these accelerators are operating at their full potential—and if not, identify what resources are needed so that they do.

- **Related: Find more on Oracle Big Data Discovery**

Using the reliability and simulation tools that are built into the Oracle Big Data Discovery platform, the CERN openlab team is able to correlate fault conditions

ıption, power
l cryogenics.

These correlations are particularly important because we are conducting a variety of analyses for accelerator conditions and modes such as cooling down, warming up, and injecting beams of energy," says Johannes Gutleber, senior engineer at CERN in charge of accelerator reliability and availability assessment.

"Previously, we had a plethora of homegrown custom and third-party applications that partially stored custom files or dumped them into a relational database management system. While this was a great way to capture the data, getting it out to analyze it was a completely different story," says Gutleber.

Reliability Analysis and Cloud-Based Disaster Recovery

CERN can determine which combinations of investments in infrastructure and technical systems would result in the most beneficial outcomes for physics research.

"We are now using Oracle Big Data Discovery to set up an architecture for the reliability and availability analysis of the systems within the proposed FCC accelerator complex. Ultimately we are looking to build a dashboard where we can look at the entire machine complex and determine the extent to which introducing fault tolerances and upgrading accelerators would improve reliability without breaking the budget," says Gutleber.

The CERN openlab team is also testing Oracle Database Cloud to understand how

While we are just beginning to test disaster recovery in the cloud, we believe it will prove to be a powerful and cost-effective solution for data security and business continuity. We have been pleased with our initial investigations with the Oracle Database Backup Cloud Service and the Oracle Java Cloud Service solutions, and look forward to continuing our investigations with Oracle Database Cloud," says Eric Grancher, group leader of database services in the CERN IT Department and CERN coordinator for the collaboration with Oracle through CERN openlab.

- **Related: Oracle Database Cloud**

Based on testing carried out at the laboratory, Manuel Martin Marquez, a CERN data scientist, believes that the visual interface of Oracle Big Data Discovery has the potential to help the research scientists—most of whom aren't data scientists—extract and interpret data from the Hadoop platform. They could then easily create analytics applications and share their findings, says Marquez.

"Essentially, Oracle Big Data Discovery can make life easier for the users by doing the dirty work for them," says Marquez.

- **Related: Get more details about Oracle Java Cloud Service**

Tech

SEP 30, 2015 @ 06:00 AM 19,783 VIEWS

Is The Data Scientist Era Coming To An End?

OracleVoice
Simplify IT, Drive Innovation FULL BIO

Margaret Harrist, Oracle

Wringing value from big data is becoming easier, thanks to much-improved software toolkits that could potentially ease the pent-up demand for hard-core data scientists.

In our data-driven world, that progress couldn't come fast enough. A quick search for data scientist jobs on LinkedIn nets more than 30,000 positions available in the US alone. Those openings call for three distinct skillsets:

- The business-savvy data scientist, typically hired by lines of business.
- The programmer data scientist who's adept with statistical analysis toolkits, often hired to work in the IT department.

can build
models,
s and marketing
agencies.

- It's a rare person who has all three of these abilities, yet a number of job openings require all of the above. But as big data tools get easier to use, organizations will have less need for the highly technical (and now scarce) data scientist, says Jeff Pollock, Oracle vice president of product management.

Source: iStockphoto

In the early days of big data, algorithms had to run serially in R, a language designed for statistical analysis, and then the data had to be loaded into Hadoop for further work, which meant heavy lifting on the I/O side of things as well.

"Now we're seeing a lot more of the most popular R algorithms being ported to run in parallelized big data environments, so you no longer have to do this I/O work," Pollock says. "You're actually bringing the algorithms to the data instead of bringing the data to the algorithms."

That also means organizations can split up complex statistical analysis into many smaller projects and run them at the same time.

“The tools didn’t exist just a few years ago, so you had to build custom statistical models or write your own program, basically build things from the ground up,” he says. Which is why the demand for data scientists began to explode.

Oh, Say Can You See

Another big change is the advent of visualization tools (such as Oracle Big Data Discovery) that make it possible for non-technical business people to analyze mountains of unstructured data in Hadoop as well as structured data in a data warehouse. The user isn’t even aware of the statistical models and machine-learning libraries being used to parse the data and populate the graphics on the screen.

“Business users are used to reporting tools that are great when you know the questions you need to ask, and you need to see your daily reports or your weekly reports about those questions,” Pollock says. “But a discovery tool is basically what a data scientist would use when it’s not clear what questions to even ask.”

So if an inventory manager wants to find out which products sell better based on seasonal patterns or weather events, a

ı correlations
r thought to
e was
Walmart's discovery that strawberry Pop-Tarts are a top-selling item when people are preparing for a hurricane. Without data discovery and statistical analysis, who would have ever thought to query about that correlation?

Now, easy-to-use discovery tools make this kind of pattern analysis available to ordinary business users—no PhD in mathematics or statistics required.

Shift Happens

Will the demand for data scientists wane as big data technology and tools mature? Consider two recent analyst reports:

- Gartner finds that more than 75% of companies are investing or planning to invest in big data analysis in the next two years, and business unit heads rather than CIOs will initiate almost half of those projects.
- Ovum predicts that the big data software sector will grow nearly sixfold by 2019. "The experimental era of big data is coming to an end," says Tom Pringle, Ovum practice leader and co-author of the report. "Organizations are formalizing their use of big data technology to realize the business value they expect to find."

While there will always be a role for mathematicians and statisticians in this data-driven business environment, Pollock says, "what they'll be focused on is building out the heavy predictive modeling instead of having to be Hadoop superheroes."

Find more about big data on Oracle.com:

- Attend Big Data Sessions at Oracle OpenWorld

Comment on this story

Report Corrections Reprints & Permissions

Active on Twitter

6 Reasons To Buy An 'Xbox One S' Instead Of A 'PS4 Pro'