Cancer Investigators Use Google Compute Engine to Accelerate Life-Saving Research

Institute for Systems Biology
Revolutionizing Science. Enhancing Life.

At a Glance

What they wanted to do
- Accelerate analysis of large cancer data sets
- Reduce the time to scale up to gain additional computing power
- Use open standards to allow researchers to employ the computing tools they prefer

What they did
- Chose Google Compute Engine to harness virtual computing power from Google's massive data center
- Worked with Google to identify complex relationships within data sets
- Used insights gained through Compute Engine to guide research

What they accomplished
- Greatly reduced the time needed to analyze a cancer data set
- Accessed powerful computational capabilities without adding hardware
- Anticipate using the time savings to accelerate the pace of medical discovery

Organization
The nonprofit Institute for Systems Biology is one of two dozen research centers taking part in a national project to map genetic changes in 20 types of cancer. Using Google Compute Engine, researchers are able to tap into the vast computational power of Google's data center to analyze complex cancer data sets in a fraction of the time, spurring hope that new treatments and cures can be found more quickly.

Challenge
Analyzing cancer data sets is a time-consuming task for researchers at the Institute for Systems Biology, even with the organization's significant computing infrastructure. Analyses may take several hours, days or even weeks to complete because of the scale and complexity of the data. The researchers knew that a powerful cloud-based computation tool could speed their work tremendously on The Cancer Genome Atlas project, a joint initiative of the National Cancer Institute and National Human Genome Research Institute.

"Shortening the computing time would allow our scientists to explore additional hypotheses and make our research more effective," explains ISB software architect Hector Rovira.

Rovira and Ilya Shmulevich, a professor at the institute and a principal investigator within the genome project, had already used Google App Engine and other Google services and were impressed by the ease of use, security and open standards. But they were worried that gaining the additional computing capacity could require considerable effort.

> "Google Compute Engine is just part of what we see as a whole new way for scientists around the world to work more effectively in the cloud, not only by accessing powerful computational resources, but by collaborating more easily on complex research."
> —Hector Rovira, software architect at the Institute for Systems Biology

Solution
Rovira and Shmulevich then heard about Google Compute Engine, which uses virtual machines harnessed by Google's data center to solve big computing problems. They started analyzing their data on the system in February 2012 with help from Google's Computational Discovery

About Google Compute Engine

Google Compute Engine allows developers to harness the scale and efficiency of the Google data center. It offers simple yet powerful clusters of virtual machines to solve big computing problems while supporting open standards and strong security and privacy controls.

For more information, visit
http://cloud.google.com/compute

"Until now, we haven't had a way to work with big data sets as effectively as with Google Compute Engine. Having a tool that lets researchers get answers fast will have a major impact on our work."
—Ilya Shmulevich, professor at the Institute for Systems Biology and a principal investigator in The Cancer Genome Atlas project

Department. The institute sends the Google team data sets containing publicly available clinical information and genomic measurements from the project's patient population – for example, information on DNA mutations in a cancer cell. Google then loads the data into Compute Engine, and the analysis helps guide the institute's research. The Google team also analyzed ISB's data using Exacycle, an experimental Google system that also offers researchers fast, large-scale data analysis.

Because Google Compute Engine can rapidly start up and shut down virtual machines, it has reduced computation times significantly. "We've been really impressed by the speed," Shmulevich says. "We can already see how its fast computation cycles will help us work faster and potentially explore more areas of research."

Results

Using Google Compute Engine, researchers have tapped into Google's powerful computational resources without buying additional hardware. The system has analyzed a cancer data set in two hours, compared with 15 hours on the institute's internal system. Compute Engine is built on open standards, so users can employ the computing tools they like best when doing their research.

"Until now, we haven't had a way to work with big data sets as effectively as with Google Compute Engine," Shmulevich says. "A tool that lets researchers analyze data and get answers quickly will have a major impact on our work."

Shmulevich and Rovira say Google Compute Engine may help provide valuable insight into the progression of cancer and improve the researchers' ability to predict how patients will respond to drugs. Even more importantly, they expect it to play a key role in transforming how scientific research is done by using it along with other Google services.

"Google Compute Engine is just part of what we see as a whole new way for scientists around the world to work more effectively in the cloud, not only by accessing powerful computational resources, but by collaborating more easily on complex research," Rovira says.