Big Compute: The Collision of where HPC is Meeting the Challenges of Big Data

Jason Stowe, Cycle Computing

At Cycle Computing we're seeing several large trends as it relates to Big Data and Analytics. We started talking about this concept of Big Compute back in Oct. 2012 with this blog post: BigData, meet BigCompute: 1 Million Hours, 78 TB of genomic data analysis, in 1 week [1]. In many ways, it's the collision of where HPC is meeting the challenges of Big Data. As our technical capabilities continue to expand in the ways we can collect and store data, the problem of how we access and use data is only growing. We're seeing several themes around this issue, and feel strongly that the ability to easily orchestrate and access Big Data on the Cloud provides clear solutions. The Cloud allows us the ability to access extremely large amounts of data — some of which is being collected to in near real-time — and allows us to tap into virtually unlimited computing power. Below are a few themes that highlight this.

Ask the Right Question

There is a shift underway where researchers, engineers, and analysts, can change the very way they think about problems. Previously, we have been limited by the computing resources we have — the clusters we have on premise. Today, we can change the very way we ask our questions. Ask the right questions — and use the Cloud to create the size of system needed to answer your questions.

Unprecedented Collaboration

Today, scientific breakthroughs now come from teams of people, instead of the lone scientist. Often the teams are in fact collaborating on different continents. It's amazing that technology has enabled this type of worldwide collaboration. But it's even more exciting to consider how HPC in the Cloud is taking this ability to a

completely new level.

Enabling Technologies for Streaming Analytics

A few technologies come to mind as innovative & allowing for these themes to come alive. Some of these include NoSql databases, RabbitMQ [2], and something we at Cycle Computing are calling Jupiter. Jupiter was designed to enable low overhead, streaming computations and analytics on hundreds to hundreds of thousands of cores. Highly resilient, and with extremely low overhead, Jupiter has already proven itself when we used it to conduct a record-breaking 156,000+ core Cloud computing run over all eight AWS [3] Regions, running Schrodinger Materials Sciences [4] tools — called The MegaRun [5]. Jupiter was critical in making this happen - and will be a key tool for Cloud computing runs of all size in the future.

Our Role

Cycle Computing is focused on creating technologies that provide greater access to computing power and capabilities through the Cloud. We're driven knowing that greater access will lead to a new era in scientific discovery & engineering invention. And while a lot of attention is paid to the really big and record-setting things we're doing at Cycle Computing, the reality is, most of our customers are doing everyday work, by using 40 to 4,000 cores. We like to demonstrate what is possible in the Cloud through our orchestration tools by continually breaking industry records — but we love enabling real-world HPC work on the Cloud on a daily basis.

Jason Stowe is Chief Executive Officer at Cycle Computing. He may be reached at editor@ScientificComputing.com.

Source URL (retrieved on *04/19/2015 - 1:41pm*):
http://www.scientificcomputing.com/blogs/2014/03/big-compute-collision-where-hpc-meeting-challenges-big-data

Links:
[1] http://blog.cyclecomputing.com/2012/10/bigdata-meet-bigcompute-1-million-hours-utility-supercomputing.html
[2] http://www.rabbitmq.com/
[3] http://aws.amazon.com/
[4] https://www.schrodinger.com/materials/
[5] http://www.cyclecomputing.com/discovery-invention/use-cases/