MAGIC Meeting Minutes
July 2, 2014

Attendees

Chip Elliott	GENI/BBN
Kelly Gaither	UTA/TACC
Dan Katz	NSF
Ji Lee	NCO
Grant Miller	NCO
Seung-Jong (Jay) Park	LSU
Lavanya Ramakrishnan	LBNL
Don Riley	U. Md

Action Items

Proceedings

This MAGIC Meeting was chaired by Dan Katz of the NSF. Seung-Jong (Jay) Park of LSU provided a description of the LSU environment to support big data analysis that is being developed under an NSF award.

Bridging, Transferring and Analyzing Big Data over 10Gbps Campus-Wide Software Defined Networks (BIC-LSU): Seung-Jong (Jay) Park

LSU has an NSF CC-NIE award to improve campus-level network services to support a wide range of scientific application big data movements. Improvements are being made to:
- Storage
- Networks
- Computing systems for big data

LSU big data applications include genome sequencing, chemistry experiments and simulation, data mining and visualization, hazard simulation and modeling and astrophysics.

Previously LSU storage systems (HDD speed was less than 1 Gbps. Network speeds were less than 1 Gbps and MPI was used for computation on HPC clusters. Under the upgrade the new system will provide:
- SSD around 20 Gbps
- 20 Gbps networking
- Hadoop and Giraph supercomputing

Under the new big data science infrastructure: Application servers will interface with edge OpenFlow switches that will access over 2 X 10Gbps networks, an aggregation OpenFlow switch that interfaces, using 10 Gbps networks, with core OpenFlow switches. The core OF switches are linked, using 40 Gbps networking to a 100Gbps switch that accesses the HPC clusters using Hadoop and also accesses Internet 2 networking.

Big data is stored using local SDSD-based high speed storage which is provided in cooperation with Samsung. Big data is transferred at up to 20 Gbps to Hadoop virtual clusters, to LSU internal collaborators and to external collaborators over Internet2. The big data is processed using Hadoop and Giraph software tools.

FOR OFFICIAL GOVERNMENT USE ONLY
c/o National Coordination Office for Networking and Information Technology Research and Development
Suite II-405 · 4201 Wilson Boulevard · Arlington, Virginia 22230
Phone: (703) 292-4873 · Fax: (703) 292-9097 · Email: nco@nitrd.gov · Web site: www.nitrd.gov

Storage: Samsung developed 8 SSD based servers providing disk to disk transfer up to 20 Gbps.
Transfer: OpenFlow control plane uses a core switch (Pluribus OF switch).   Edge switches are Dell or HP OpenFlow
Compute: Computation is carried out on a virtual Hadoop cluster to run Hadoop and Giraph jobs

LSU applications using these capabilities include processing 1 TB of metagenome data, requiring 10 TB of memory and 100 TB of storage for de novo genome assembly. They use map reduce architecture, Pregel framework, Apache Giraph (an open source implementation of Pregel), and Parallel Giraph based Assembler (PGA). Demonstration of this architecture and tools demonstrates improvements in efficiency and capability of big data applications. For the complete briefing, please see the MAGIC Website at:
http://www.nitrd.gov/nitrdgroups/index.php?title=MAGIC_Meetings_2014 under the July 2, 2014 meeting

Upcoming Meetings:
July 22-25, XSEDE, San Diego
July, NCAR: Data management for environmental and climate research. Speeding up data movement.
November, SC14, New Orleans: WISPE meeting focused on software

Next MAGIC Meetings:
August 6, 2014, 2:00-4:00 EDT, NSF
September 3, 2014, 2:00-4:00 EDT, NSF