Moving Towards Terabit/sec Scientific Dataset Transfers

High Energy Physicists Set New Record for Network Data Transfer

With 186 Gbps Sustained Data Rates High Energy Physicists Demonstrate Efficient Use of Long Range Networks to Support Leading Edge Science

SEATTLE, Washington – During the SuperComputing 2011 (SC11) conference in November, an international team of high energy physicists, computer scientists, and network engineers led by the California Institute of Technology (Caltech), University of Victoria, the University of Michigan, CERN, and Florida International University, together with other partners, has set a new record in data transfer rate over a 100 Gbps wide area network circuit. Using a 100 Gbps circuit set up by CANARIE and BCnet, the team was able to reach transfer rates of 98 Gbps between the University of Victoria Computing Centre located in Victoria, British Columbia, and the Washington State Convention Centre in Seattle, Washington. Together with the simultaneous data rate of 88 Gbps in the opposite direction, the team reached a sustained bidirectional data rate of 186 Gbps, demonstrating the capability of efficiently moving 2 Petabytes of data per day between two data centers.

Partners from the University of Florida, the University of California at San Diego, Vanderbilt University, Brazil (Rio de Janeiro State University, UERJ, and the São Paulo State University, UNESP) and Korea (Kyungpook National University, KISTI), were involved in the larger demonstration of massive data transfers to and from the Caltech booth at the SC11 conference, to locations within the US as well as in Brazil and Korea.

Caltech's exhibit at SC11 by the High Energy Physics (HEP) group and the Center for Advanced Computing Research (CACR) also demonstrated applications for globally distributed data analysis for the Large Hadron Collider (LHC) at CERN, along with its global network and grid monitoring system MonALISA (http://monalisa.caltech.edu) as well as its Fast Data Transfer application (http://monalisa.caltech.edu/FDT) developed in collaboration with the Polytechnica University (Bucharest). Caltech's worldwide collaboration system EVO (Enabling Virtual Organizations; http://evo.caltech.edu) developed with UPJS in Slovakia, and SeeVogh (http://www.seeogh.com) aimed at the private sector also were shown.

A focus of the exhibit was the HEP team's record-breaking demonstration of storage-to-storage data transfer over wide area networks, using a small set of data servers. The team achieved an aggregate disk-to-disk data rate of over 60 Gbps (7.5 GBytes/s), between clusters equipped with SSD disk arrays at the University of Victoria (10 servers) and Caltech booth at the convention center in Seattle (7 servers).

By setting new records for sustained data transfer among storage systems over continental and transoceanic distances, using simulated and real LHC datasets, the HEP team demonstrated its readiness to enter a new era in the use of state of the art cyber-infrastructure to enable physics discoveries at the high energy frontier, while

demonstrating some of the groundbreaking tools and systems they have developed to enable a global collaboration of thousands of scientists located at 350 universities and laboratories in more than 100 countries to make the next round of physics discoveries.

Advanced Networks, Servers and State of the Art Applications

The record-setting demonstrations were made possible through the use of a dedicated 100 Gbps wave between the Caltech booth and the University of Victoria Computing Centre, provided by the Canadian Research and Education Network CANARIE, and BCnet, the regional R&E network in British Columbia. Optical transport used for the demonstration was provided by Ciena's OME 6500 Optical-Packet Platforms.

The University of Victoria setup consisted of a Brocade MLXe-4 router with equipped with a 100GE line card connecting to the Ciena OME 6500 client interface. One 16 port 10GE line card on the MLXe-4 was used to connect directly to a cluster of 10 Dell servers equipped with raid arrays with six SSD disks each. Each of the servers was able to read from disk and deliver data to the network at 9.4 Gbps.

The network setup at the Caltech booth used one Brocade MLXe-4 router with a 100GE and one 16 port 10GE line card. A Dell/Force10 Zetascale Z9000 data center switch was used to interconnect the 40GE server interfaces. An aggregated link of 12 10GE ports was used to interconnect the Z9000 and the MLXe-4. Force10 switch was also connected to ESnet ANI 100G testbed using the 40GE QSFP LR Optics from ColorChip.

The Caltech booth's cluster consisted of a set of servers equipped with 40GE Mellanox CX3 network interface cards. Three of the servers were using PCIe Generation 3 motherboards, enabling data rates of up to 35 Gbps per single network interface. Each of these servers was equipped with raid arrays of 16 SSD disks each, with a measured data rate to disk of 12.5 Gbps.

One of the key elements in this demonstration was Fast Data Transfer (FDT), an open source Java application developed by Caltech in collaboration with Polytehnica University in Bucharest. FDT runs on all major platforms and uses the NIO libraries to achieve stable disk reads and writes coordinated with smooth data flow using TCP across long-range networks. The FDT application streams a large set of files across an open TCP socket, so that a large data set composed of thousands of files, as is typical in high energy physics applications, can be sent or received at full speed, without the network transfer restarting between files. FDT can work on its own, or together with Caltech's MonALISA system, to dynamically monitor the capability of the storage systems as well as the network path in real-time, and send data out to the network at a moderated rate that achieves smooth data flow across long range networks.

Since it was first deployed at SC06, FDT has been shown to reach sustained throughputs among storage systems at 100% of network capacity where needed, in production use, including among systems on different continents. FDT also achieved a smooth bidirectional throughput of 191 Gbps (199.90 Gbps peak) using an optical system carrying an OTU-4 wavelength over 80 km provided by Ciena during SC08.

Lessons Learned: Towards A Compact Terabit/sec Facility

The SC11 demonstration also achieved its goal of clearing the way to Terabit/sec data transfers. The latest generation of servers based on the recently released PCIe v3 standard and equipped with line-rate 40GE interface cards from Mellanox as well as raid arrays with high-speed SSD disks allowed the team to reach a stable throughput of 12.5 Gbps from network to disk per 2U server. It is important to underline that pre-production systems with relatively few SSDs were used during this demo, and no in-depth tuning was performed due to limited amount of time in the preparation. We therefore expect these numbers will improve further, and approach the 40GE line rate within the next year. Such devices could be used as front-end caches for larger storage systems, or as a stand-alone storage system capable of delivering 100 Gbps of data with half a rack of equipment.

The LHC Program: CMS and ATLAS

The two largest physics collaborations at the LHC, CMS and ATLAS, each encompassing more than 3,000 physicists, students, engineers and technologists from 180 universities and laboratories, have embarked on a new round of exploration at the frontier of high energies. As the LHC experiments continue to take collision data, new ground will be broken in our understanding of the nature of matter and space-time and in the search for new particles. In order to fully exploit the potential for scientific discoveries during the next year, more than 100 petabytes (10^{17} bytes) of data have been processed, distributed, and analyzed using a global grid of 300 computing and storage facilities located at laboratories and universities around the world, and the data volume is expected to rise to the exabyte range (10^{18} bytes) as the LHC progresses in its collision rate and energy.

The key to discovery is the analysis phase, where individual physicists and small groups located at sites around the world repeatedly access, and sometimes extract and transport multi-terabyte data sets on demand from petabyte data stores, in order to optimally select the rare "signals" of new physics from potentially overwhelming "backgrounds" from already-understood particle interactions. The HEP team hopes that the demonstrations at SC11 will pave the way towards more effective distribution and use for discoveries of the masses of LHC data.

Further information about the demonstration may be found at: http://supercomputing.caltech.edu

Acknowledgements

The demonstration and the developments leading up to it were made possible through the strong support of the partner network organizations mentioned, the U.S. Department of Energy Office of Science and the National Science Foundation, in cooperation with the funding agencies of the international partners, through the following grants: US LHCNet (DOE DE-FG02-08-ER41559), UltraLight (NSF PHY-0427110), DISUN (NSF PHY-0533280), CHEPREO/WHREN-LILA (NSF PHY-0312038, PHY-0802184 and OCI-0441095), the NSF-funded PLaNetS (NSF PHY-0622423) project a travel grant from NSF specifically for SC09 (PHY-0956884), FAPESP (São Paulo) Projeto 04/14414-2), as

well as the NSF FAST TCP project, and the US LHC Operations Program funded jointly by DOE and NSF.

We'd like to thank our industry partners involved in this year's demonstration for providing equipment and support: CIENA, Brocade, Mellanox, Dell and Force10 (now Dell/Force10), and Supermicro.

Quotes on the significance of the demonstrations:

"This year the HEP team was able to show the way towards effective use of the next generation of networks and data systems. By sharing our methods and tools with scientists in many fields, we hope that the research community will be well-positioned to further enable their discoveries, taking full advantage of 100 Gbps networks as they become available. In particular, we hope that these developments will afford physicists and young students throughout the opportunity to participate directly in the LHC's next round of discoveries as they emerge."

-- *Harvey Newman, Caltech professor of physics, head of the HEP team and co-lead of US LHCNet, and chair of the US LHC Users Organization*

"The 100Gb/s demo at SC11 is pushing the limits of network technology by showing that it is possible to transfer peta-scale particle physics data sample in a matter of hours to anywhere around the world."

-- *Randall Sobie, University of Victoria Professor and LHC physicist*

About Caltech: With an outstanding faculty that has been honored with 32 Nobel prizes and 66 National Medals of Science and Technology, and such off-campus facilities as the Jet Propulsion Laboratory, Palomar Observatory and the W. M. Keck Observatory, the California Institute of Technology is one of the world's major research centers and a premier institution of learning. The Institute conducts instruction in science and engineering for a student body of approximately 950 undergraduates and 1,400 graduate students who maintain a high level of scholarship and intellectual achievement. Caltech's 124-acre campus is situated in Pasadena, California, a city of 135,000 at the foot of the San Gabriel Mountains, approximately 30 miles inland from the Pacific Ocean and 10 miles northeast of the Los Angeles Civic Center. Caltech is an independent, privately supported university, and is not affiliated with either the University of California system or the California State Polytechnic universities. http://www.caltech.edu.

About University of Victoria:

About CACR: The mission of the Center for Advanced Computing Research (CACR) is to ensure that Caltech is at the forefront of computational science and engineering. CACR provides an environment that cultivates multidisciplinary collaborations and its researchers take an applications-driven approach and currently work with Caltech research groups in aeronautics, applied mathematics, astronomy, biology, engineering,

geophysics, materials science, and physics. Center staff have expertise in data-intensive scientific discovery, physics-based simulation, scientific software engineering, visualization techniques, novel computer architectures, and the design and operation of large-scale computing facilities. http://www.cacr.caltech.edu/.

About BCNET: BCNET is a not-for-profit, shared IT services organization that collaborates with its higher education members to explore and evaluate leading-edge information technology services to accelerate research, collaboration, learning and innovation. BCNET's high-capacity, advanced fibre optic network infrastructure provides a common and powerful platform to deliver services to support research and the strategic objectives of higher education institutions.

Owned, governed and funded primarily by its members and government, BCNET is guided by clear, member-defined principles that provide a framework for the types of projects or services it will undertake. Proposed services share the common goal of reducing members' costs, minimizing technology duplication and improving efficiencies, while meeting the collective mandate of its members.

Sixty-eight research and higher education institutions connect to BCNET, including federal and provincial research labs, federal cultural institutes, provincial health centres, universities and research institutions; another 72 colleges and schools connect to BCNET through the Provincial Learning Network.

About CANARIE: CANARIE Inc. is Canada's Advanced Research and Innovation Network. Established in 1993, CANARIE manages an ultra high-speed network that supports leading-edge research and big science across Canada and around the world. One million researchers, scientists and students at over 1,100 Canadian institutions, including universities, colleges, research institutes, hospitals, and government laboratories have access to the CANARIE Network. Together with 12 provincial and territorial advanced network partners, CANARIE enables researchers to share and analyze massive amounts of data, like climate models, satellite images, and DNA sequences that can lead to groundbreaking scientific discoveries. CANARIE is a non-profit corporation supported by membership fees, with the major investment in its programs and activities provided by the Government of Canada.

CANARIE keeps Canada at the forefront of digital research and innovation, fundamental to a vibrant digital economy. For additional information, please visit: www.canarie.ca.

About CERN: CERN, the European Organization for Nuclear Research, has its headquarters in Geneva. At present, its member states are Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, the Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, and the United Kingdom. Israel, Japan, the Russian Federation, the United States of America, Turkey, the European Commission, and UNESCO have observer status. For more information, see http://www.cern.ch.

About the University of Michigan: The University of Michigan, with its size, complexity, and academic strength, the breadth of its scholarly resources, and the quality of its faculty and students, is one of America's great public universities and one of the

world's premier research institutions. The university was founded in 1817 and has a total enrollment of 54,300 on all campuses. The main campus is in Ann Arbor, Michigan, and has 39,533 students (fall 2004). With over 600 degree programs and $739M in FY05 research funding, the university is one of the leaders in innovation and research. For more information, see http://www.umich.edu.

About Florida International University: Florida International University is one of Florida's State University System institutions and the premier public research university in the South Florida metropolitan area, the country's fourth largest. The university is located in Miami-Dade County with campuses in West and North Miami. The university with its large and diverse student body is the largest minority serving institution in the nation and is ranked first in awarding STEM degrees to underrepresented minorities. The university was founded in 1972 with only 5,667 students. In the four short decades since the university has grown phenomenally and now boasts a total enrollment of more than 39,000 students; 60 percent of which are of Hispanic descent. With an increasing emphasis on graduate research, the university has the Carnegie Foundation's highest ranking and in 2008, received over $100 million in external contracts and grants. http://www.fiu.edu

About the Politehnica University (Bucharest, Romania): Founded in 1818, Politehnica University of Bucharest (UPB: http://www.upb.ro) is the largest and the best Technical University in Romania. UPB is a full member of several international organizations such as CESAER, EUA and AUF, and has bilateral co-operation agreements with similar universities, mainly in Europe, the U.S., Singapore and Japan. UPB also participates in projects funded by NATO and the EU 6th and 7th Frameworks. 26,000 undergraduate, masters and Ph. D students are enrolled at UPB, including more than 1,500 in diverse areas of Computational Science and Engineering (CSE). The Romanian National Center for Information Technology (NCIT) is part of UPB and is run by the CSE Department. UPB has extensive experience in monitoring distributed resources, in projects such as MonALISA, a fully distributed monitoring system based on autonomous, self-describing agent-based subsystems which has been developed over the last seven years by Caltech and UPB. The UPB team also has been involved in the EU-NCIT-EU IST Excellency project, focused on Grid computing and Collaborative work, and the FP7 project SENSEI. UPB also benefits from its results in the FP7 P2P-Next project, such as unified P2P technologies for live streaming and progressive downloading.

About UERJ (Rio de Janeiro): Founded in 1950, the Rio de Janeiro State University (UERJ; http//www.uerj.br) ranks among the ten largest universities in Brazil, with more than 23,000 students. UERJ's five campuses are home to 22 libraries, 412 classrooms, 50 lecture halls and auditoriums, and 205 laboratories. UERJ is responsible for important public welfare and health projects through its centers of medical excellence, the Pedro Ernesto University Hospital (HUPE) and the Piquet Carneiro Day-care Policlinic Centre, and it is committed to the preservation of the environment. The UERJ High Energy Physics group includes 15 faculty, postdoctoral, and visiting Ph.D. physicists and 12 Ph.D. and master's students, working on experiments at Fermilab (D0) and CERN (CMS). The group has constructed a Tier2 center to enable it to take part in the Grid-based data

analysis planned for the LHC, and has originated the concept of a Brazilian "HEP Grid," working in cooperation with USP and several other universities in Rio and São Paulo.

About UNESP (São Paulo): Created in 1976 with the administrative union of several isolated institutes of higher education, the São Paulo State University, UNESP, has 39 institutes in 23 different cities in the State of São Paulo. The university has 33,500 undergraduate students in 168 different courses and almost 13,000 graduate students. UNESP has just inaugurated the Center for Scientific Computing with the goal of empowering research groups with high performance computing, storage and networking resources. The new state-of-the-art data center houses the GridUNESP main cluster and the SPRACE cluster.

About SPRACE: The São Paulo Regional Analysis Center (SPRACE) is a Worldwide LHC Computing Grid Tier-2 center operating in association with the Open Science Grid. The SPRACE researchers are members of the Fermilab´s DZero experiment since 2004 and of the CERN's CMS Collaboration. SPRACE has been leveraging competences in different research areas by sharing the technical expertise generated by High Energy Physics experiments, including high speed networks, high performance computing, and grid computing architectures. SPRACE has inspired the GridUNESP project which has deployed the first Campus Grid in Latin America.

About GridUNESP: UNESP has deployed a distributed computational system which is the largest Campus Grid initiatives in Latin America, with computing resources located on seven different campuses. The GridUNESP computational infrastructure includes almost 400 servers, over 200 terabytes of storage space and an advanced networking infrastructure for inter-cluster connection. GridUNESP has established a formal partnership with the Open Science Grid Consortium and is employing the OSG middleware stack to integrate its computational resources and share them with other research and education institutions worldwide.

About Kyungpook National University (Daegu): Kyungpook National University is one of leading universities in Korea, especially in physics and information science. The university has 13 colleges and 9 graduate schools with 24,000 students. It houses the Center for High Energy Physics (CHEP) in which most Korean high-energy physicists participate. CHEP (chep.knu.ac.kr) was approved as one of the designated Excellent Research Centers supported by the Korean Ministry of Science.

About KISTI: KISTI is a specialized institute providing STI (Science and Technology Information) services based on national supercomputing center and advanced research networks (KREONET, GLORIAD-KR and KRLight) to promote global competitiveness in science and technology by actively challenging the rapidly changing world paradigm. KRLight supports end-to-end dedicated lightpath provisioning for high end applications such as HEP as a GLIF Open Lightpath Exchange of Korea. KISTI aims to be the world leader in science and technology. Please visit http://www.kisti.re.kr/english/index.jsp for more information.

About the National Science Foundation: The National Science Foundation (NSF) is an independent federal agency created by Congress in 1950 "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense…" With an annual budget of about $6.06 billion, it is the funding source for approximately 20 percent of all federally supported basic research conducted by America's colleges and universities. In many fields such as mathematics, computer science and the social sciences, NSF is the major source of federal backing.

About the DOE Office of Science: DOE's Office of Science is the single largest supporter of basic research in the physical sciences in the nation and ensures U.S. world leadership across a broad range of scientific disciplines. The Office of Science also manages 10 world-class national laboratories with unmatched capabilities for solving complex interdisciplinary problems, and it builds and operates some of the nation's most advanced R&D user facilities, located at national laboratories and universities. These facilities are used by more than 21,000 researchers from universities, other government agencies, and private industry each year.

Contacts:

Harvey B. Newman, Caltech Professor of Physics, newman@hep.caltech.edu
(626) 395-6656

Deborah Williams-Hedges, Caltech Acting Director of Media Relations, debwms@caltech.edu (626) 395-3227