SURAgrid Update – March 2009

NEWS

- - SURAgrid Spring All Hands Meeting - -

The SURAgrid All Hands Planning Committee is considering all of the community input, the ongoing work of building and maintaining SURAgrid, needs of our science researchers, and the funding landscape. Confirmed on the agenda is a joint session with the Coalition for Academic Scientific Computation (CASC), an educational nonprofit 501(c)(3) organization with 57 member institutions representing many of the nation's most forward thinking universities and computing centers. The joint session will include presentations by the NIH on access management and their federated identity service, the NSF Office of Cyberinfrastructure, and time for a dialogue between SURAgrid and CASC.

The All Hands meeting will begin at 1pm on Wednesday April 29th and end at noon on Friday May 1st. For those of you making your travel arrangements, here are a few suggested lodging options:

- Morrison Clark Inn, located at 1015 L St NW (walking distance to SURA), is advertising a rate of $199/night. Please call them directly at 202-898-1200. This is a special in honor of the first 100 days of the new presidency. If you'd like a copy of the special announcement, please contact Linda Akli, akli@sura.org.
- The Henley Park Hotel located at 926 Massachusetts Ave NW (walking distance to SURA)
- The Four Points Sheraton located at 1201 K Street, NW, (walking distance to SURA)
- Internet2 meeting attendees can take advantage of the Internet2 room block rate of $224 at the Crystal Gateway Marriott. (close to Reagan National Airport and the metro). For more information see http://events.internet2.edu/2009/spring-mm.
- CASC meeting attendees may take advantage of the CASC room block rate of $209 per night at the Westin Arlington Gateway Hotel (1 block from NSF and the metro). More information is available at http://www.casc.org/meetings/09apr/index.html

- - SURAgrid Membership Continues to Grow - -

We are pleased to announce that North Carolina A&T State University, Greensboro, N.C., has been approved by SURAgrid Governance Committee as our newest SURAgrid Participating member. Dr. Yaohang Li is NCAT's SURAgrid point of contact and currently serves as a SURA IT Committee member. Dr. Li is a faculty member in the departments of Computer Science and Bioengineering. One of his research interests is protein structure modeling, and he has developed a protein loop structure model with National Science Foundation funding. NCAT is a research university with 11,000 students, a PhD program in computational science, and ranks third in the UNC System in terms of research funding - with over $45 million in FY2008.

We welcome North Carolina A&T State University to SURAgrid and look forward to working together!

- - SAS at TTU - -

Contributed by: Jerry Perez, Research Associate, High Performance Computing Center at TTU

A dialogue about SAS in the SURAGrid environment was started in the February SURAgrid monthly call, where Cheryl Doninger, Director of the grid computing solutions group at SAS was the featured presenter. This article contributed by Jerry Perez, Research Associate at TTU's, High Performance Computing Center, presents a case study of SAS in the grid environment at TTU, a SURAgrid contributing member.

Problem Description

TTU developed statistical resampling methods to determine whether announcements and other historical events affect stock prices. Resampling is a compute-intensive method where the data are sampled repeatedly (say 10,000 times) with or without replacement. In addition, each resampled data set required costly matrix inversions. Adding to this computational complexity, the resampling procedure itself was studied using 10,000 simulations for a total of 100,000,000 data sets to be processed. The problem grew even larger when 10 parameters were attached to each simulation which resulted in 1 billion data sets.

Solution
SAS/CONNECT provides an essential tool for distributing jobs across a network. TTU combined the distribution capabilities of SAS/CONNECT with high-powered SAS analytics to implement their financial application on a grid.

TTU's SAS grid is comprised of 200+ high powered Windows machines in the computer labs of the Rawls College of Business Administration at TTU; only 100 of these machines are available at any given time because of the number of SAS licenses purchased by TTU. Available licenses are managed though a keyserver application. Thus, the computing environment can be conceptually viewed as a virtual 100 node (2.66 GHz per node) super computer with 100 Gigs of combined RAM. These computers are used during the day by students to complete their daily assignments; SAS grid jobs have been run while students are using them without affecting performance. However, the prime opportunity to leverage these resources for grid computing is during off peak hours and nights when students have no need for these machines.

Benefits
The grid computing capabilities of SAS/CONNECT offer a fantastic advantage over the sneaker grid in that the jobs to process the one billion data sets are all sent at the same time and all data are sent back to the client machine for automatic summarizing using SAS analytics. TTU used 40 minutes of compute time on their SAS compute farm and would have used 25 hours on a single machine. It should be noted that the savings (40 minutes versus 25 hours) is actually quite a bit more when considering the "false starts" and minor errors that always accompany job execution: there were approximately four false starts, so the total savings is really 5x40 minutes = 3.3 hours, vs. 5x25 hours = 125 hours. This represents a 97% performance gain from leveraging the SAS grid technology.

As a result of TTU's initial success with their SAS grid, they are currently implementing their next SAS grid application which is a portfolio selection and analysis project.

References

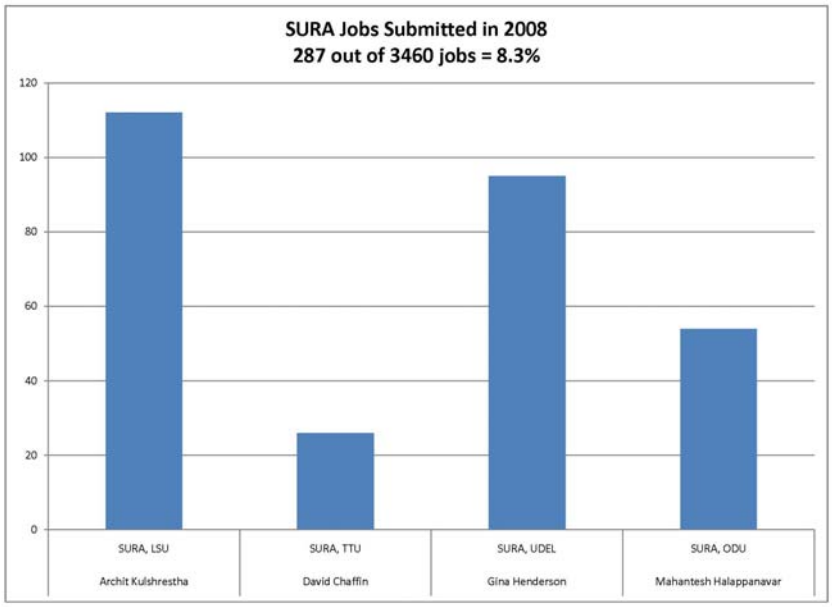
- Hein, S.E. and Westfall, P.H. (2004). Improving Tests of Abnormal Returns by Bootstrapping the Multivariate Regression Model with Event Parameters, Journal of Financial Econometrics 2, 451-471
- Bremer R., Perez J., Smith P. and Westfall, P.H. (2004), Grid Computing at Texas Tech University Using SAS. manuscript

- - First Steps towards SURAgrid Usage Data - -

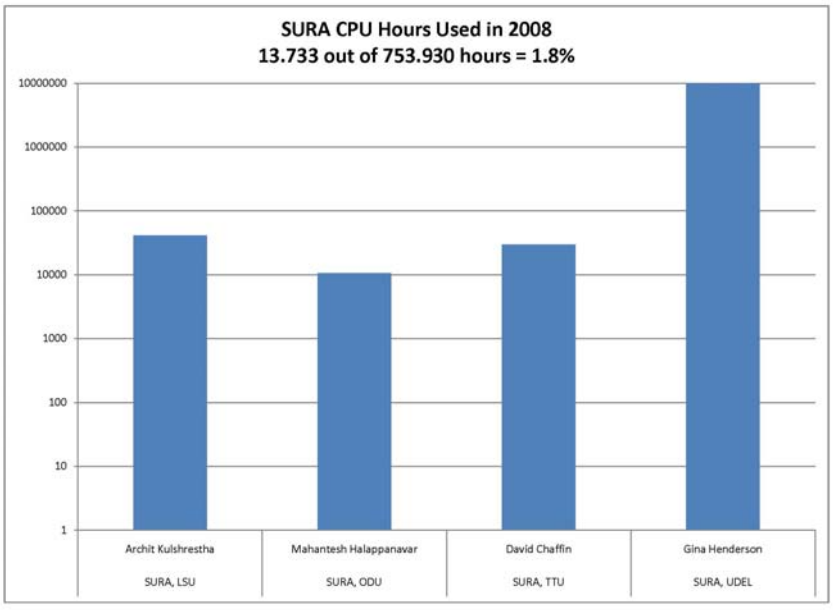
Contributed by: Nicole Geiger, Software Systems Engineer, IS&T - Advanced Campus Services, GSU

As SURAgrid is moving to gather accounting data across the various sites, some universities are already implementing this process locally. At Georgia State University, we are using IBM's LoadLeveler on our URSA system. This scheduler allows various types of data to be gathered. Below are two graphs. The first graph depicts the number of CPU hours used for SURAgrid users on URSA for the calendar year 2008 by individual user. It also shows the percentage of SURAgrid usage out of the total number of CPU hours for that time frame.

The second graph shows the number of SURAgrid jobs submitted by SURAgrid users for the same time period, as well as the percentage. These are just samples of the types of data that can be gathered with LoadLeveler and some of the other schedulers. What's next: Accounting data for SURAgrid across sites?



SURAgrid Users on URSA for Calendar Year 2008



- - UAB High-Performance Computing Update - -

Contributed by John-Paul Robinson, Lead Infrastructure Services, UAB IT

Cheaha, a UAB resource connected to SURAgrid, has reached a significant milestone that incorporates several incremental updates. With the acquisition of additional Dell hardware in August of last year and the incorporation of GridWay meta-scheduling technologies, Cheaha now offers an expanded compute pool of close to 3TFlops with the ability to seamlessly leverage additional compute resources on SURAgrid. For more information on this upgrade, please feel free to visit the UABgrid documentation page: http://docs.uabgrid.uab.edu/wiki/Cheaha

- - SURA Grows Hardware Partnership with IBM - -

From: SURA 2008 Year in Review

SURA set a significant milestone in our partnership with IBM with the addition of 74 IBM p575+ HPC nodes to augment existing systems purchased by four SURA member institutions: Louisiana State University, Georgia State University, Texas A&M University the University of Miami. SURA was able to broker an aggregated purchase these refurbished IBM nodes at a very substantial savings – well below current educational discount prices. Through this program, our member schools saved in excess of $3.7 million on this purchase. The systems are being added to existing systems at these members' sites and will add approximately 10 teraflops (TF) of capacity (nearly double the current capacity) to existing systems, which will also be added to the SURAgrid resource pool.

UPCOMING Events

- Mar 2 – 6, Open Science Grid Annual Meeting, Livingston, LA, http://indico.fnal.gov/conferenceDisplay.py?confId=2012
- Apr 27 – 29, Internet2 Spring Meeting, Arlington, VA, http://events.internet2.edu/2009/spring-mm
- Apr 28 – 30, CASC Spring Meeting, Arlington, VA, http://www.casc.org/meetings/09apr/index.html

All issues of SURAgrid update can be found at http://sura.org/programs/sura_grid_communications.html. Please submit comments and contributions for inclusion to Linda Akli at akli@sura.org.