Hipcat Face-to-face Meeting 22-Sep-06

University of Houston

Attendees

Guy Almes – A&M
Kim Andrews – Rice
Judy Cox – UH
Borries Demeler – UTHSCSA
Eric Enquist – UH
Nick Grishin – UTSW
Chris Hempel – TACC
Tsung-I Mark Huang – UH
Lennart Johnsson – UH
Akbar Kara – UH
Andre Kerstens – UTEP
Chuck Koelbel – Rice
Josten Ma – UH
Jeremy Mann – UTHSCSA
Rosalinda Mendez – UH
Lee Panetta – A&M
Paul Roberts – UH
Tom Seville – UH
Phil Smith – TTU
Warren Smith – TACC
David Steffen – BMC
Michela Taufer – UTEP
Pat Teller – UTEP
Kirin Thyagaraja – Rice
Spiros Vellas – A&M

Notes

- Lennart Johnsson gives an overview of TLCC:
    - Different centers and research programs
    - Josten provides a bit more detail on SWTC and TIMES
    - Their 250 node Itanium cluster is still the main platform to serve the UH scientific community
    - They've build a distributed campus storage network
    - Visualization lab: Eric Enquist provides some more detail; the current SGI machines will soon be replaced by a commodity Linux cluster
    - Connection to the Houston fiber ring has been established
    - Short overview of high-tech classrooms with video conferencing

    - Overview of the web site services that TLCC offers
    - Proposal activities ($ awarded; about 20% of UH funding); the departments that are involved. Rosalinda provides a bit more detail on who they are trying to attract and about their outreach activities (e.g. take kids to a ranch and use technology to observe nature)
- Dennis Fowdy – Vice-president of IT
    - Makes a point that collaboration is important on several levels
    - Welcomes everybody and leaves the meeting
- Lennart introduces Paul Roberts:
    - Paul is on the LEARN and RENoH (Research Education Network of Houston) engineering teams
    - Pat asks where most of their funding comes from; Lennart answers that they don't have a real breakdown and that it comes from many different sources. Operational funding mostly comes from campus IT.
    - Many collaborations across institutions in Houston which drives the infrastructure development
    - Paul gives an overview of RENoH circuits and who uses them. Dark fiber, point-to-point and ring network and shared Ethernet service. Lennart explains that half of the network is under control of the university and half under control of the faculty. Network brings huge cost savings compared to the previous solution (leased lines from carrier).
- Break at 9:45
- Lei Huang (postdoc of Barbara Chapman) presents the OpenMP project. Barbara Chapman was his advisor during his PhD studies.
    - Most future architectures are or will be based on multi-core processors
    - OpenMP: compiler directives for shared memory programming
    - OpenMP is being extended to clusters. Intel brought out first product: Cluster OpenMP (using DSM). Barbara's group utilizes Global Arrays
    - OpenUH compiler available for download: http://www2.cs.uh.edu/~openuh/
- Akbar Kara introduces the LEARN network
    - LEARN stands for 'Lonestar Education and Research Network'
    - They have 33 members (mostly in higher education)
    - Operating expenses are paid for by the members/users
    - The network backbone consists of mostly dark fiber
    - San Antonio to El Paso layer 1 and 2 link is leased from and operated by NLR (National Lambda Rail) and has a capacity of 10 Gb. The link will be fully active in November 2006.
    - LEARN is capable of 72 x 10 Gb network connections
    - Pat asks whether it would make sense to connect to other statewide networks; Akbar answers that that makes sense only if a business case can be build for it. Paul mentions Quilt: an organization that tries to pull disparate networks together by holding workshops, etc.
    - Phil asks whether people are actually asking for capacity on LEARN. Akbar answers that that is the case. Phil asks if the commercial telcos kick back because of loss of income. Akbar thinks that they do (he's not involved in such discussions though) and that there are clauses in their contracts that

delay-sensitive traffic (e.g. voice/video) are not allowed to be offered by LEARN.
    - Paul says that for a basis fee of $2000 you can be connected to LEARN and have a fast link to any other campus/research center on the network.
    - LEARN applications: transport, NLR, Internet2/NewNet aggregration, 10 Gb Teragrid, in-state collaborations (hipcat, grids, texas digital library, etc)
    - Phil asks if there is dark fiber available between Dallas and Lubbock. Akbar answers that that is the case, but because there is not much competition in that region it is expensive.
- At 11:30 am the EC is taking half an hour in another meeting room to meet in private. In this meeting Phil Smith is reelected to be Director of Hipcat for the next two years. In the meantime there is a presentation by Chris Benhaddou on networking projects at UH
- At 12 the EC has ended their private meeting and returns to the public meeting.
- Presentation by Daewon Byun on air quality forecasting in east Texas. This is a collaborative project between A&M and UH. Data is available at http://imaqs.uh.edu (login needed for AFQ details – user: uhaqf, passwd: imaqs_aqf)
- Presentation by Guy Almes on the short-term future plans of the LEARN community and the creation of a research council
    - Lee Panetta asks what the ideal size for the new research counsil would be. Guy answers that they are currently thinking about this, but thinks around 16 is probably a good number.
    - Borries asks what disciplines will be represented in the research council. Guy answers that this will be highly dependent on the users themselves, but it will probably consist of people that can see the practical value of the network and not only know how to operate it and know what it costs. Pat adds that it should probably consist of both types of people: practical use and management/big thinkers.
- Presentation by Larry Pinsky on CERN/LHC and NASA projects at TLC2/UH
    - The ALICE project is trying to recreate and study the early big bang times
    - A spin-off project from ALICE is looking at a new methodology used for cancer therapy
    - Lee says that finding 1000 processors in Texas should be trivial; most agree. The TIGRE project could possibly deliver on a request for many processors on a short notice and for short periods of time. Larry says that there are other non-trivial issues like getting raw data, privacy (hippa) and convincing doctors that this can make a difference in getting the number of secondary cancer cases down.
    - ALICE software is very robust and runs on a multitude of platforms because the system has been engineered to run on grids.
- Warren updates the group on TIGRE
    - There are currently 12 people active in the TIGRE project
    - The TIGRE portal is web-service enabled. This is the only grid portal around that supports such services according to Phil.
    - First experiment: porting Borries's Ultrascan application to TIGRE was very successful and satisfied all milestones.

- A question comes up whether the HiPCAT group should write new proposals to acquire new hardware for the TIGRE grid. Most agree this is probably a good idea. Where should funding be coming from? Federal or state?
- Where should the new hardware be based? Central or distributed? Phil says it should be distributed because of the grid concept; this is also a question of available floor space and cooling/power/manpower availability.
- Nick Grishin proposes that users should pay for TIGRE services in the future. Phil says that such a cost-model wouldn't work very well for most institutions.
- Borries says that users should write TIGRE money in their proposals if they are using the system. This will be discussed in the last agenda point.

- HiPCAT institutional updates:
  - UTEP:
    - Donated cluster from TSRI, La Jolla. Currently being made operational
    - D@H live; Phil asks whether boinc could be used for TIGRE. Alan Sill at TTU is looking at this. Web URL of docking project: http://docking.utep.edu. Ask Andre for an invitation code to help test the project
    - Topaz, a gridftp client, has been released; Andre mentions that this might be usable for TIGRE too. URL: http://gcl.utep.edu/projects/topaz
  - TAMU:
    - SURA-IBM collaboration
    - Spiros says that they plan to buy a 640 processor cluster from IBM (power5+ processors). Power/cooling infrastructure has to be upgraded because of weight/power/cooling requirements. They have purchased data direct drives and use the GPFS file system.
  - Rice:
    - Upgraded their Itanium cluster
    - Upgraded their Cray XD1 cluster
    - Common authentication system implemented for all HPC systems
    - Received 3 major parts of a DOE grant with an approx. value of $10 million
  - TACC
    - Upgrade of Lonestar cluster is being finalized: 1300 nodes (2600 processors, 5200 cores), 8 GB per node. 40 TB Lustre file system. Phil asks whether their LSF license has become more expensive because of the upgrade. The answer is yes, but only marginally.
    - Cluster2006 attendance in Barcelona; they organized a Multicore programming workshop. Cluster2007 will be in Austin next year
    - New computing facility has been finished
  - UTHSCSA
    - Renovation of Borries' lab
    - Purchase of a new 20 node cluster (32 bit)
    - Emry Brooks has made Ultrascan grid-capable

        - A paper has been accepted at SC06
        - Borries acquired course approved to teach Ultrascan
        - David Steffen asks what software are Borries's students using for modeling. Borries answers: Shake-and-Bake, mpi-blast, sam, etc
        - Genomics center of excellence is now in existence.
    - BCM
        - David Steffen mentions that many clusters on campus are still isolated; he tries to find out about these and pull them together.
        - A discussion about Matlab on clusters is going on for 5 minutes. Pat asks if anybody has ever used the parallel version of Matlab. Answer is that at TACC they have done some scalability studies, but nothing scientific. There are some groups lined up at TACC that like to use it though so more investigations will have to be done.
        - R (open source statistics package) is heavily used in Baylor.
        - Phil asks whether Baylor uses the bioinformatics package of Matlab. David hasn't found a function for it yet. Underlying data is a big problem.
    - UTSW
        - They organized a meeting among 3 Dallas universities (similar to HiPCAT but on a bigger scale)
        - They have some faculty positions available. Ask Nick for more info.
    - TTU
        - Upgrade their main cluster to the Lustre filesystem; 30 TB of total disk space including backup.
        - Pat asks why they chose Lustre. Phil answers that it's free and the test period was very successful.
    - UH
        - Smaller projects going on. Most have been mentioned during the meeting.
- Borries brings up the question on how to ask for TIGRE funding in proposals. What is the best way of doing this? Phil mentions that the EC should definitely speak about this and come up with some guidelines. Asks Borries to write him an email about this. Pat mentions to get a letter/commitment from the state. David Steffen thinks that it won't be so hard to make people pay for Tigre resources if the cost savings can be made clear.
- Pat asks the group as the SC'08 general chair to help her find good threads that can run through the SC'08 conference. Biology and medicine would be interesting subjects. There is also a big interest in energy (Horst Simon of NERSC is leading this effort). Anybody who has an idea should drop an email to Pat.
- 4 pm: Phil adjourns this highly successful meeting. We all just loved the food!!