

Minutes of the HiPCAT Meeting
University of Texas Health Science Center at San Antonio
27 Jan 2006
Author: A. Kerstens

HiPCAT Attendees:

- Robert Badgett (UTHSCSA)
 - Anthony Chronopoulos (UTSA)
 - Borries Demeler (UTHSCSA)
 - Lori Gibler (TTU)
 - Nick Grishin (UTSW)
 - Steve Harris (UTHSCSA)
 - Luis Hernandez (UTEP)
 - Lennart Johnson (UH)
 - Andre Kerstens (UTEP)
 - Chuck Koelbel (Rice)
 - Josten Ma (UH)
 - Jeremy Mann (UTHSCSA)
 - Rosalinda Mendez (UH)
 - Lee Panetta (TAMU via Insors)
 - David Parker (BCM)
 - Jerry Perez (TTU)
 - Karl Schulz (TACC)
 - Alan Sill (TTU via Insors)
 - Phil Smith (TTU)
 - Warren Smith (TACC)
 - Michela Taufer (UTEP)
 - Kiran Thyagaraja (Rice)
 - Jerry York (UTHSCSA)
 - Zhiwei Wang (UTHSCSA)
 - Jae Yu (UTA)
-
- Meeting takes place at UTHSCSA and is opened by Phil Smith at 8:30
 - Introductions by everybody in the room and on the AccessGrid. AccessGrid: Lee Panetta at TAMU.
 - Presentation by Jerry York – VP and CIO UTHSCSA

The academic and admin building in which the meeting takes place is the newest building on campus. UTHSCSA has classroom and rooms with ‘standardized patients’ for medical student training. UTHSCSA is in the top 3 medical schools of the country and is growing tremendously. The aging center on campus is number 1 in the country. Four campuses are spread over the region (1 central and 3 regional). 9000 hours of video conferencing per year with regional campuses. UTHSCSA recently launched a MD/PhD program. UTHSCSA is the only tier-1 university in South Texas. They just hired new VP of Research.

Network connectivity is 10/100Mb to the desktop. They have started to use the LEARN network more and more. They are in the process of putting wireless in every room on campus (40% done). UTHSCSA sees 2 to 3 million email messages per month of which 60% is blocked as spam.

Question from Jerry Perez: will there be a HPC center soon? – They are currently discussing this possibility.

Question from Phil Smith: what is the bandwidth of the campus backbone? – 1 GB now, 10 GB soon.

- Presentation by Jim Williams – Director of LEARN (Lonestar Education And Research Network)

Jim shows a map of PoP nodes. San Antonio has a Wavenet PoP. If you want to connect to NLR for dedicated services contact Jim. Luis mentions that the El Paso NLR leased link will be ready in March (to the city) and will be connected to the UTEP campus in April. Texas Tech still has no dark fiber solution, but this is being worked on. Future network ring should go all around the state including the far West. Tasks for 2006: finish the initial deployment; what layers should be provided (1,2 or 3)?; how to best connect the unconnected institutions?

Question from Phil Smith: who is not connected to the LEARN network? – Many (smaller) institutions are not yet, but they are working to get everybody connected up.

Question from Jerry York: are there any non-educational inst. connected? – They are currently talking to NASA.

- Talk Steve Harris – Dept of Periodontics

Steve provides an introduction to the biological problem and how HiPCAT can help. He shows his lab website where gene collections can be found. Finite element methods are used for modeling.

Question from Jerry Perez: is there a load-balancing present between bone and muscle? – nobody knows, but possibly yes.

Question from Phil Smith: Phil wants to get more bioinformatics at TT: the question is how to bring computer scientists and researchers closer together? – Steve says to have more meetings like this one. Specific bioinformatics initiatives can help.

- Talk Robert Badgett – Dept. of medicine

The common diagnosis in primary care is based on 60% guessing. UTHSCSA has its own live search engine for Medline: <http://sumsearch.uthscsa.edu>. This tool is currently working using the Pubmed database. It will soon be working as a 'systematic textbook'. Medline currently has 15 million records in xml with a size of 52 GB. It is growing 6 GB per year and is updated weekly.

Question Jae Yu: how long does a search take currently? – around 30 secs.

Question Steve Harris: would it be possible to use natural language processing in searching? – Bob says that maybe they should, but it is very complex: searching in medicine couldn't reliably be done like Google page ranking, because medicine has more time-dependent data.

Question from Phil Smith: how many times do you observe mismatches in the system? – this is difficult to answer, because every patient is different and there are a lot of subtle questions.

- Talk Emre Brooks – CS dept. graduate student

The talk is titled: Parallelization of the 2-dimensional Spectrum Analysis and Genetic Algorithms for Modeling Solutes in Analytical Ultracentrifugation Experiments – Pathways to TIGRE. Borries explains the biological aspects of the experiment. There is coarse- and fine-grained parallelization involved. Lively discussion going on between members during the presentation.

Question Jerry Perez: are you going to put a Globus component in this setup? – Yes, the plan is to run this on TIGRE (incl. Globus components)

Question Phil Smith: why do you use genetic algorithms and not something else? – for the sake of flexibility: you don't know what your parameter space is going to look like during the experiment.

- Talk Brad Pollock – Director center for epidemiology and biostatistics, bioinformatics initiative

Recently funded infrastructure grant implications are: a new phd in biostatistics/bioinformatics and a positive educational and scientific impact. Three scientific liaisons are being recruited. The bioinformatics core facility is being discussed by Zhiwei Wang, the director of the facility.

Question Phil Smith: how much money was available to start this program? – The initial funds were \$2,000,000.

Question Nick Grishin: where will future funding be coming from? – NIH, NSF, State funding.

- Presentation Michael Papka and Thomas Uram – Argonne Nat. Lab

AccessGrid has seen 6000+ downloads and 8000+ certificates issued. It has multi-platform support (Windows, Mac OS X, Linux). Several deploying projects: Westgrid, HPC Europe, etc. There is a large community of third-party developers of AG tools. AccessGrid 3 design and implementation will be very much based on standards. AG3 has just gone into beta testing. AccessGrid homepage: www.accessgrid.org. AccessGrid portal: www.agcentral.org.

Question Warren Smith: when will the final AG 3 be coming out? – Thomas says very soon, probably less than a couple of months.

Question Rosalinda Mendez: how are third-party tools and their support handled? – Mike says that developers of their respective AG tool are supposed to support their tools themselves. Argon only handles the core AG software.

- Talk Jerry Perez – Main CDLT representative HiPCAT

Jerry shows the new Insors website on www.hipcat.net. Insors was originally started from AG version 1. Jim Miller, CEO of Insors, gives an overview of Insors. Insors can interoperate with the AccessGrid without many problems. They are working on interfacing with AG 3 right now. Insors is windows only currently, but a beta version for Mac is ready to come out soon.

- Discussion concerning conferencing for HiPCAT between the HiPCAT members, the Argon developers and Jim Miller.

Support for integration with Skype and other tools for VoIP plus integration with H323 and SIP: Insors – yes, AG – no (Insors is already taking that space from their point of view)

Question Jae Yu: what does a typical Insors installation cost? – 1 license for 1 camera is \$1,600. A typical meeting room setup is between \$30k and \$40k. A server is around \$50k for commercial users, but educational/research users get a 50% discount.

Question Rosalinda Mendez: there are rumors that the DOE will not support AG anymore soon. How true is this? – Mike says there are at least 18 months of funding from NIH available and also some DOE money. The community will keep it alive as well as long as there will be a need for AG.

- Discussion Lennart Johnsson concerning the fiber network in Houston and the relationship to the LEARN network. They want to take it down into the Galveston area as well since there are a number of institutions and campuses there.
- Discussion LEARN by Luis Hernandez (UTEP) and all other members:

The current network link to UT Austin is an OC3 (155 Mb/s). UTEP and the city of El Paso are utilizing in the order of 90 Mb/s of this bandwidth. Current traffic over Internet2: 25 Mb/s sustained; most of this traffic is coming from Mexico (UTEP is the Internet2 gateway for Mexico). Jae is explaining that UT Arlington is pushing about 150 Mb/s and expect to go to 2Gb/s in the future.

Examples of projects that could benefit of LEARN:

Digital Library

Astronomy (telescopes)

Physics

Biochemists

HPC computer centers

Currently it is well known that scientists/researchers won't use the infrastructure if they know upfront that it can't handle the experiments. Therefore it is necessary to have the bandwidth in place.

Computation comes in 3 components: cpus, storage and network. UTEP has quite a bit of processor power and storage, but is currently lacking in bandwidth. In the near future it will be possible to route traffic meant for the public internet over the LEARN network.

- TIGRE project discussion by Warren and Alan (on Insors)

TIGRE was started officially on 12/01/05. Applications that could help TIGRE are welcome. Contact Warren or Alan if you have one.

- Institutional updates
 - UH/TLCC (Lennart Johnson, Josten Ma, Rosalinda Mendez) – they mostly talked about the disaster recovery story of last year.
 - Rice (Chuck Koelbel) – new Cray XD1 cluster installed, Kiran has joined Rice IT.
 - Baylor (David Parker) – Jay Boisseau visit, new 50 node cluster.
 - UT Austin/TACC (Warren Smith) – new IBM P5 system, upgrading Dell cluster, new scientific comp. classes taught.
 - UTSW (Nick Grishin) – new imaging center (collaboration Dallas and SW), completing project with TACC after a year.
 - TTU (Phil Smith) – setup community cluster project, upgrading Dell cluster to 348 cpu's, introduction Lori Gibler (liaison between researchers and IT)
 - UTA (Jae Yu) – UTA selected as tier-2 computer center (center will be extended by 500 compute nodes per year), finished new chemistry building incl. computer room.
 - UTEP (Andre Kerstens) – cluster A/C story, Pat Teller chair SC08 in Austin.
 - UTHSCSA (Borries Demeler) – TIGRE project collaboration, Phil promises to get a HiPCAT support letter to Borries, workshop organization.
- Discussion scientific computing classes presented by Karl Schultz of TACC (Jay couldn't make it).

TACC technology areas: HPC, Visualization & Data Analysis (VDA), Data Information Systems (DIS), Distr. and Grid Computing (DGC). Courses are hosted by the UT Austin CS department.

- 1) Introduction to scientific computing
- 2) Parallel computing for science and engineering
- 3) Visualization & data analysis for science and engineering
- 4) Distr. and grid computing for science and engineering

These classes will include lab exercises as well (answer to question Michela Taufer).

EPIC (www.eotepic.org) will fund these classes.

Chuck mentions the Connections project at Rice, which is part of the Creative Commons.

- Open discussion:
 - Letter of support from HiPCAT for NSF proposals that use TIGRE. Support is not in terms of money (for the investigators) but in human cycles (i.e., experience, ideas). Blessing of the project means that the investigators should provide something for the HiPCAT group. Need for a mutual exchange between HiPCAT and investigators.
 - Lennart Johnson and Chuck Koelbel share the opinion that Borries should get a letter from TIGRE. Phil will distribute draft that Borries writes to TIGRE group.
 - Next face-to-face meeting will be somewhere in September. Phil will email for final date.
 - Problems with interconnectivity between AG, Insors and phone connections have to be figured out.
 - Jerry Perez mentions that every institution should have a dedicated CDLT contact for the video conference meetings. Andre volunteers for UTEP.
- Meeting adjourned at 15:45 CST.