

## Draft Minutes of SRCPAC Meeting of Thursday, September 5<sup>th</sup>

### 1:04 – 1:05 Call to Order - David Madigan

### 1:05 – 1:07 Shared Research Computing Facility Launch, October 29<sup>th</sup> – Victoria Hamilton

Announcement of Shared Research Computing Facility Launch on Tuesday, October 29<sup>th</sup> from 3 – 5:30 p.m. Dual goal is to (a) thank the funding agencies and internal administrative supporters, and (b) spread the word about this new service facility to the broader research community. Invitations for this event will be sent to SRCPAC members, and researchers should be encouraged to attend.

### 1:08—1:15 Update on Research Storage Pilot – Tom DiPrete

#### Details

- CUIT technical team working with functional lead Tom DiPrete to finalize service definition and articulate metrics of success of pilot
- Goal is a dollar per GB per year (competitive with Amazon)
- 60-70 eligible participants
- Pilot Phase will:
  - Limit accounts to a 2 terabyte per storage space maximum
  - Each PI will be allowed up to three accounts (1 personal and 2 group)
  - Accounts will start with 100 GB, and then incremental purchases up to limit available

#### Questions

1) Will storage spaces be accessible from new HPC cluster?

--> Not dynamically, but you can transfer files to HPC scratch storage.

2) Are there periodic backups of files as part of the research storage pilot?

--> Yes, periodic backups will be made of stored files for the research storage service pilot, and “snapshots” of stored files will be available over a limited time window (for example, 30 days). In addition, a periodic backup copy of stored files will be kept off-site. CLARIFICATION FROM DISCUSSION AT SRCPAC MEETING: In case of a hardware problem or failure, the storage hardware support contract will enable storage system recovery in 2-3 days. However, In the event of a disaster, such as physical damage to the data center by flooding, etc., recovery time for research data may be many weeks. The service level agreement for the research storage service pilot will contain the details of file backup and recovery policies.

3) Is this all same storage?

--> The storage is a physical unit storing both the HPC and the research storage clusters.

4) Available off campus?

--> Specifics need to be worked out, but yes, some form of off-campus services will be offered, possibly through Columbia’s VPN service

### 1:15 – 1:25 Update on new HPC – Victoria Hamilton

#### Loss of NSF MRI – Feedback from Jerry Ostriker

Leadership of the NSF Directorates are convinced that high performance computing is important. When it comes to a specific proposal, it's refereed by a specific group of people (from one discipline) without collective vision. As a result, they shoot down proposals *not* in their area. Collectiveness worked against us. Future efforts should perhaps be more targeted.

#### Updates re Yeti new 1600+ core HPC

- 10 departments across campus participated in the HPC purchase
- \$1.1M equipment (total includes HPC scratch storage and Research Storage Pilot)
- \$800k from NYS which purchased equipment infrastructure, networking etc. as well as nodes and storage
- Goal is a self-sustaining model for HPC service, but need immediate support to ensure good service from the get-go
- Contributions from A&S and SEAS (and CUIT) to underwrite CUIT staff support until the service is self-sustaining
- Aiming to have Yeti (will call the new HPC "Yeti") up and running for the Launch in October
- Rental service will be introduced in the Fall
- Free tier will be introduced after service for paying participants is operating well
- Rob Lane will host monthly workshops for users to help on-board
- Phase 1 Hotfoot will be retired as has exceeded recommended four year life. Phase 2 of Hotfoot was added later and therefore will be continued until it has finished its four year life.
- If Yeti proves successful, Goal is to implement a buying cycle so new machines are in place every two years.

#### Questions

1) Why not add nodes whenever someone has some funding?

→ We need to make sure the CUIT resources are primarily focused on serving the researchers, and a continuous stream of purchasing & installing would consume a lot of resources. Moreover, it would be considerably more difficult to decommission on the recommended four year cycle

2) What if there is significant demand for another machine earlier due to grant receipt?

→ The two year rule was recommended as best practice by other Universities, but SRCFAC will evaluate and respond to the needs of the research community. If Yeti proves such a success that there is big demand, the purchase cycle can be accelerated.

3) What's the capacity of the SRCF?

→ Different factors come into play to answer that question exactly, but a working estimate is that there are 100kW available in the data center for research computing, and Yeti is estimated to take 25 to 30kW. Therefore the SRCF could support two more systems the size of Yeti. Substantial demand for machines might also allow for some specialization in terms of machines. For example, folks who want very fast interconnects might elect to group together for one machine, while others might need a large shared memory.

Important to remember -- This is a pilot → Feedback is useful, so reach out to Victoria at any time ([Victoria.hamilton@columbia.edu](mailto:Victoria.hamilton@columbia.edu)).

## **1:25 – 1:30 Cloud HPC – Rob Lane**

### Stages

- With support from Statistics, CUIT started to Investigate options to extend the capacity of HPC with a tightly integrated system to allow “cloud-bursting”. Unfortunately, the need to designate where your data resides in order to operate on it makes the original cloud bursting idea too cumbersome for researchers.
- Now exploring the ability to offer HPC services in the cloud, as a separate complement to Yeti
- Might provide flexibility without constraints of purchasing equipment and needing local research staff or students to set up a system
- Not quite the same as running local systems; the motivation to share doesn’t make sense in the cloud as there are nodes on demand.
  - Similar scripts would be used to run a job in the cloud...but located up in a cloud versus being in the node
  - Financial model is not yet clear. Depends to some extent how much overhead is required from CUIT
  - Going Forward – Get Yeti up and running...then revisit the potential of Cloud for SRCPAC
  - Goal is to make sure to keep a toe in the water so can always offer best portfolio of options to research community.
  - Future cloud-based service may also serve as stopgap between purchasing cycles.

## **1:30—2:00 SRCPAC Sub-committees and Operating Governance of new HPC – David Madigan**

### Evolution of Current Sub-committees:

At inception, SRCPAC formed six sub-committees. Some have fulfilled their objective and should be disbanded: Internal Survey Sub-Committee and External Peer Review Sub-Committee. The effort to create a Manhattanville Sub-committee to explore potential for further economies-of-scale ran afoul of the fact that each building is developing its own plans, and there is no strategy to share common infrastructure. The decision was regrettably made to abandon this effort. The Cloud/External Resources Committee should be reinvigorated with new faculty when SRCPAC meets again in December. The Storage sub-committee is functioning well under Tom DiPrete’s direction. The Hotfoot operations committee likewise should continue as only roughly half of the machine will be decommissioned. It will dissolve naturally when Phase II Hotfoot is decommissioned.

### **Proposed Yeti Operating Committee**

Following a thorough discussion, SRCPAC determined that the Shared Research Computing community would be best served by creating a small, dedicated, knowledgeable and informed group to make decisions regarding Yeti operational issues. Therefore, SRCPAC recommends the following structure for the Yeti Operating Committee.

- Two representatives from the three research groups that invested in the largest number of nodes
- One representative from the two groups that invested in a mid-level of nodes
- One representative from the five groups who invested at the entry level

- One representative for the renters (appointed by SRCPAC Chair until sufficient renters to select their own)
- Yeti Chair will be appointed from among the representatives by the Chair of SRCPAC
- Staff: CUIT Research Computing Services (RCS)

The five Members of the Yeti committee will be expected to make decisions based on what makes the most sense for the community as a whole (that is, a committee member will not just reflect the interests of his or her own group.) Experience at other universities and of Hotfoot suggests that in most cases, informed and dedicated people can reach a consensus. In the event that a consensus is not reached, each member of the committee will have a single vote. RCS will staff the committee, but will have no vote. The Yeti Committee will report on their activities at each SRCPAC meeting, highlighting any particularly contentious issues. SRCPAC reserves the right to restructure the Yeti Operating Committee to reflect the evolution of the resource and changing needs. SRCPAC also maintains the right to over-rule decisions for the good of the community.

Appendix: Attendee List

**APPENDIX: SRCPAC MEETING ATTENDEE LIST**  
**September 5, 2013**

1. Tiffany Shaw, DEES/APAM
2. Javad Lavaei, EE
3. Pierre Gentine, EAEE
4. Amy Nurnberger, CU Libraries/IS
5. Rebecca Kennison, CU Libraries/IS
6. Mark Newton, CU Libraries/IS
7. Thomas Lippincott, CCLS
8. Raj Bose, Researching Computing Services (CUIT)
9. Jerry Ostriker, Astro
10. Hatim Diab, CCLS
11. Kathryn Johnston, Astro
12. Don Lemma, CUIT
13. Victoria Hamilton, Research Initiatives (EVPR)
14. David Madigan, EVP A&S
15. Andreas Hielscher, Biomedical Engineering
16. Naomi Henderson Naik, LDEO
17. Chris Marianetti, APAM
18. Thomas DiPrete, Sociology/CPRC
19. Wojciech Kopczuk, Economics
20. Robert Cartolano, CU Libraries/IS
21. Breck Witte, CU Libraries/IS
22. Marc Spiegelman, APAM/DEES
23. Abhishek Joshi, Research Computing Services (CUIT)
24. Halayn Hescoock, CUIT
25. Rob Lane, CUIT
26. Peter Tripp, Psychology
27. Antonio Melgarejo, Columbia Astrophysics Lab
28. Harmen Bussemaker, Biological Sciences
29. Alex Bergier, CUIT
30. Mahdad Parsi, LDEO
31. Marie B. Tracy, Research Initiatives (EVPR)