Research Computing Executive Committee

Annual Report

Summary of SRCPAC Activities

2016 Achievements
Monthly Group Utilization of Yeti
Yeti Publications
Yeti Operations Committee
Support
Educational Activities
Frontiers of Computing Systems
Habanero!
RFP Committee

Partnership Between CUIT & SRCPAC

Future Growth & Capacity
Amazon Web Services
Electronic Lab Notebooks

Shared Research Computing

Established robust, reliable infrastructure (*machines and people*)

Supports Columbia community in research *and* education

.... to work
.... to grow
.... to *innovate*

Summary of SRCPAC Activities

2016 Achievements
Yeti Report
Research Computing Services Update
Education Impacts
Research Impacts
Next Round Yeti: Habanero

Kathryn Johnston, Chair of SRCPAC & Professor and Chair of Astronomy

2016 Achievements

Markers of Success

Shared Research Computing: Grown into an institutional service and pivotal resource!

New shared cluster – Habanero:
31 faculty groups, purchasing ~200 servers and over 4700 cores (twice the size of previous rounds).

Successful Recruitment Tool:
6 Inquiries this year.

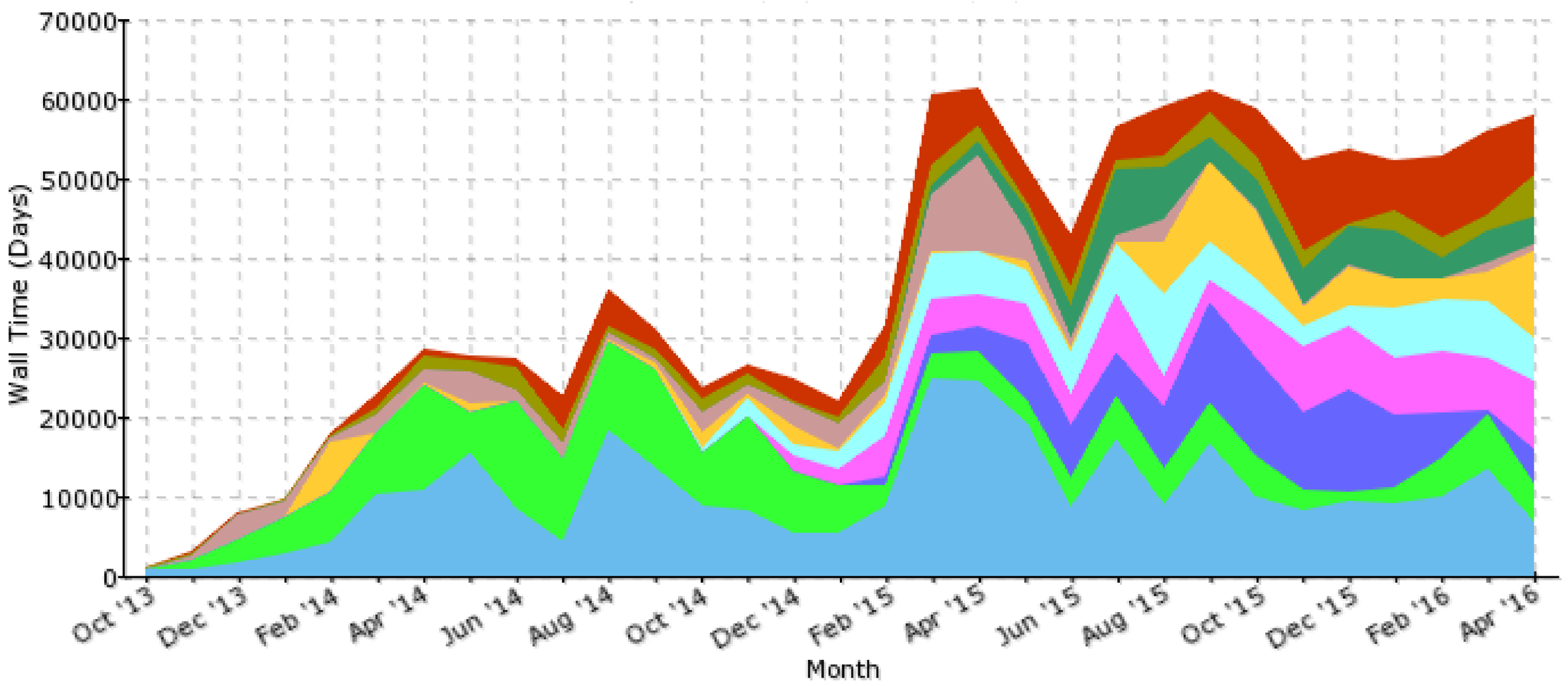
Committed Research Computing Services Team:
CUIT has two new added staff.

Increasing Education Role:
First undergraduate course uses cluster (genetics/bioinformatics); A&S and SEAS fund Education Tier.

Intellectual Hub:
Faculty-led initiatives enabled by and reliant upon SRCPAC/Yeti and CUIT/RCS.

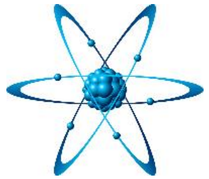
Monthly Group Utilization of Yeti

All Clusters from October 2013 to April 2016; 75% Utilization



84 Yeti Publications To Date

That We Know Of…



Physics & Astronomy
25

Social Sciences
18

Biomedical Sciences
22

Statistics, Computer Science & Engineering
19

Governance Model: Yeti Operating Committee

FY15-16

Charge: Allow Users to Guide Operating Decisions
FY16 Meetings: 2

9 Executive Committee Members

- 4 Representatives of Large Purchasers: Statistics, APAM, Lamont, CCLS
- 4 Representatives of Small Purchasers (1 Vacant Seat): Journalism, Biological Sciences, Chemistry
- Faculty Chair Representing Renters & Free Tier

Outcomes

- Longer jobs approved for Infiniband systems.
- Free tier users allowed to use more resources at once but for shorter periods of time.
- No changes requested at Spring meeting, indicating maturation of service and generally satisfied users.

Support: Research Computing Services

- 6 FTEs Supported by CUIT, A&S, SEAS, EVPR
  - In 2015, CUIT acquired two new salary lines for RCS
- Team Roles
  - Manager
  - Technical Lead
  - 2 x Systems Administration
  - 2 x User Support
- Assist Yeti and (soon) Habanero Operating Committees
- Provide Workshops for Novice Users
  - Topics: HPC, Linux, Scripting (New)
  - Each Class Held Once Per Semester
- Hold Q&A Sessions with Researchers (Upon Request)
- Positive Faculty Feedback from Prior SRCPAC Meetings

Educational Activities

Workshops, Training & Outreach

Successful experiment this year with a genomics/bioinformatics course utilizing the new Education Tier.

RCS and Libraries/Information Services collaborating on training workshops in HPC and Linux.

Ryan Abernathey (Earth & Environmental Sciences):

- Python Boot Camp
- Software Carpentry

A&S and SEAS each contributed funds to purchase Habanero nodes for the Educational Tier.

Looking Head: Frontiers of Computing Systems

New Working Group/Center in the Data Sciences Institute

Directed by Steve Nowick (Computer Science); Co-Directed by Chris Marianetti (APAM). For collaboration on high-performance system design and analysis, and applying massively-parallel computation to many application areas.

20+ faculty members from Engineering, Arts and Sciences, Earth Institute, Neuro-Technology Center, Medical School.

7 external collaborators from Sandia National Labs, NASA/JPL, IBM, Intel, Microsoft Research, NVIDIA.

An Exciting Intellectual Nexus Between SRCPAC and the Data Science Institute!

Announcing the Next Wave of Research Computing

The New Habanero Cluster!

RFP Committee

Four Committee Meetings in Spring 2016

Compute

7 Bids → 3 Finalist Presentations → 1 Winner: **HP**

Storage

9 Bids → 2 Finalist Presentations → 1 Winner: **DDN**

Physical Location of Hardware

Zuckerman Institute Data Center
(Basement of the new Jerome L. Greene Science Center)
MOU Not Yet Finalized

One Expansion Round Within 18 Months

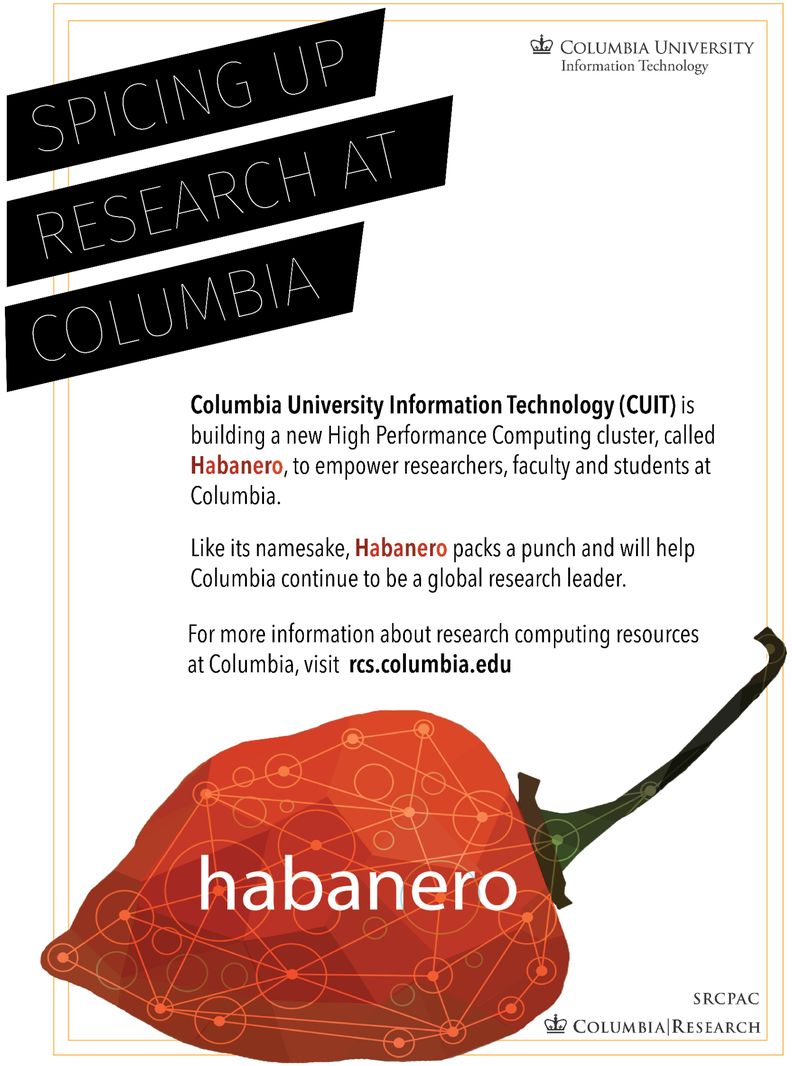
Use Estimated at 55kW This Year

Committee Members

David Kipping, Astronomy
Kyle Mandli, APAM
Chris Marianetti, APAM
Bob Mawhinney, Physics
Lorenzo Sironi, Astronomy
Brent Stockwell, Biological Sciences

New *Habanero Operating Committee*
Modelled After Highly Successful Yeti Operating Committee.

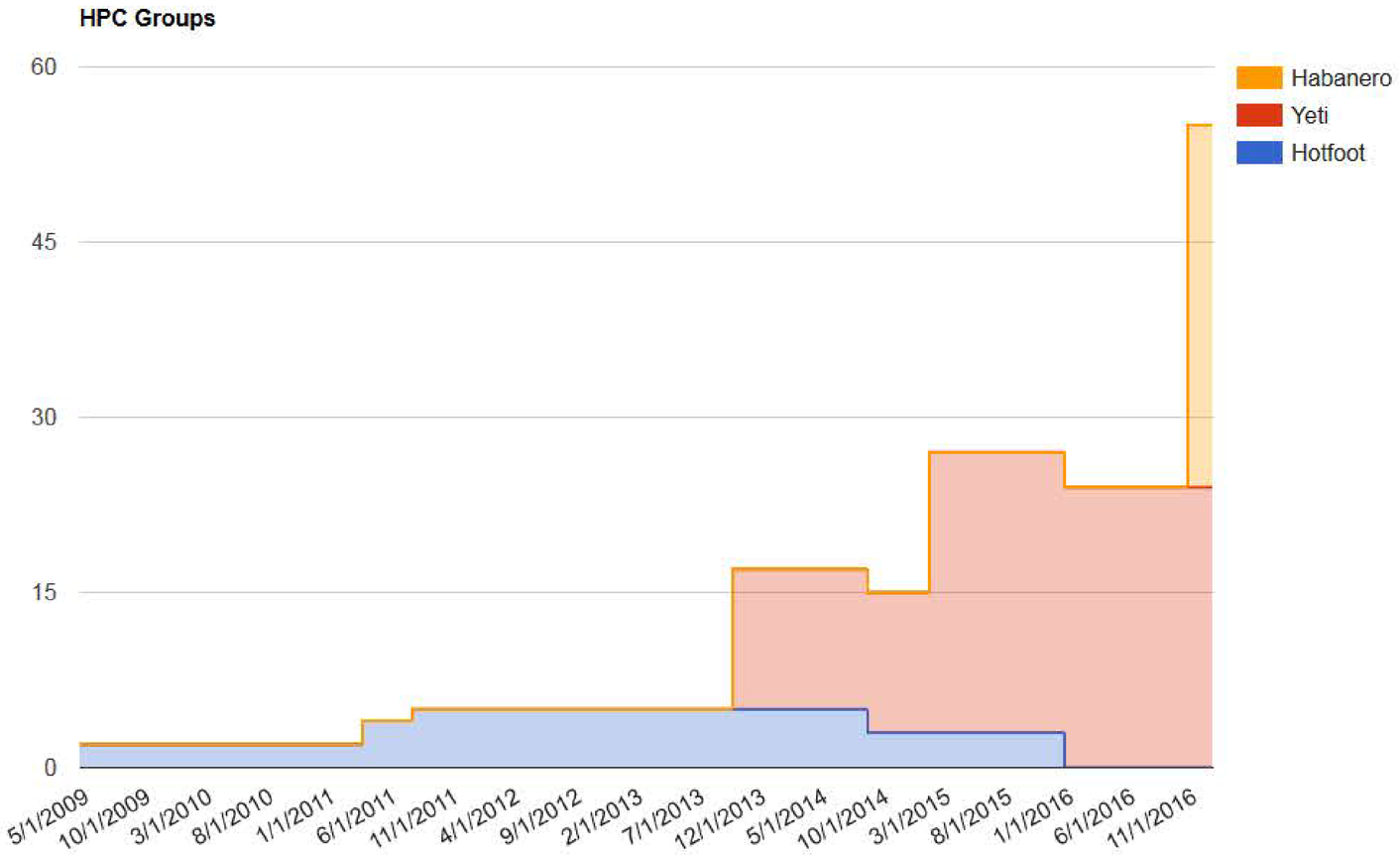
The Next Wave of Research Computing



The New Habanero Cluster!

196 Nodes
4704 Cores
304 Terabytes
31 Groups
$1.3 Million Spent
…and counting!

System Growth



Phase	Groups	Nodes	Cores
HOTFOOT LAUNCH	2	32	256
HOTFOOT EXPANSION #1	2	24	288
HOTFOOT EXPANSION #2	1	8	96
YETI LAUNCH	12	101	1616
YETI EXPANSION	12	66	1056
HABANERO LAUNCH	31	196	4704

31 Groups Engaged in Habanero

13 Arts and Sciences
Ryan Abernathey (DEES)
Andrei Beloborodov (Physics)
Harmen Bussemaker (Biological Sciences)
Juliana Capaldi (Social Science Computing Cluster)
David Kipping (Astronomy)
Szabi Marka (Physics)
Brian Metzger (Physics)
Andy Millis (Physics)
Ingrid Richter (Psychology)
Ozgur Sahin (Biological Sciences)
Department of Statistics
Brent Stockwell (Biological Sciences)
Tian Zheng (Statistics)

2 Medical Center
Larry Abbott (Neuroscience)
Christine Benanti (Psychiatry)

10 Engineering
Daniel Bienstock (APAM)
Ton Dieker (IEOR)
Julia Hirschberg (Computer Science)
Chris Marianetti (APAM)
Kathy Marte (ChemE)
Cev Noyan (APAM)
Steve Nowick (Computer Science)
Lorenzo Polvani (APAM)
Peter Schlosser (EEE)
Steve Sun (CEEM)

6 Other
Peter deMenocal (Lamont)
Kathy McKeown (Data Science)
Mahdad Parsi (Lamont)
Donna Schillington (Lamont)
Renata Wentzcovitch
Raj Bose (Zuckerman)

SRCPAC & CUIT Partnership

Future Growth & Capacity
Amazon Web Services
Electronic Lab Notebooks

Gaspare LoDuca, Chief Information Officer & Vice President, Information Technology

Future Growth & Capacity

HPC and Beyond

Main Data Center – Private Cloud Resources and Yeti HPC cluster; Potential for future clusters

Zuckerman Data Center – Habanero cluster

Amazon Web Services – Public Cloud Resources

Amazon Web Services

Cloud Computing Agreement

CUIT has signed an enterprise agreement with Amazon Web Services, thereby developing a process to link existing AWS accounts to the enterprise agreement. In addition, provisioning new accounts can be done automatically. Both services are available now via the ServiceNow service catalog.

The Research Computing Services (RCS) team will provide advisory, system, and software support services for anyone wanting to get set up on AWS compute services.

Electronic Lab Notebooks

CUIT & Libraries Effort

Co-funded by CUIT and Libraries/Information Services, and working together with EVPR, Columbia has obtained an enterprise license with LabArchives, a provider for electronic lab notebooks. The enterprise license includes the professional edition and the classroom edition. LabArchives has signed a BAA, and we are in final phases of CUMC Security certification.

This service will be at **no charge** to researchers and instructors. Benefits include secure, backed-up, collaboration space, audit trail to protect intellectual property, and support for agency data management requirements.

With formal go-live is planned for July 1, 2016, several labs have elected as early adopters:

- Brent Stockwell Lab (Biological Sciences)
- Jeff Kysar Lab (Mechanical Engineering)
- Oliver Hobert Lab (Biological Sciences)

Shared Research Computing

Established robust, reliable infrastructure (*machines and people*)

Supports Columbia community in research *and* education

.... to work
.... to grow
.... to *innovate*

Questions?

Kathryn Johnston
Chair of SRCPAC
Professor & Chair of Astronomy
kvj@astro.columbia.edu

Research Computing Services (CUIT)
rcs@columbia.edu

2016 RCEC Membership Roster

G. Michael Purdy (*Chair*)

Executive Vice President for Research

Primary

David Madigan
Executive Vice President
for Arts and Sciences

Mary Boyce
Dean, Engineering

Ann Thornton
University Librarian

Gaspare LoDuca
Vice President of Information
Technology/Chief Information Officer

Justin Pearlman
Chief of Staff
Office of the Provost

Kathryn Johnston
Chair of SRCPAC

Alternate

Amber Miller
Dean of Science
Faculty of Arts and Sciences

Shih-Fu Chang
Senior Executive Vice Dean
Engineering

Rob Cartolano
Associate Vice President for Digital
Programs & Technology Services

Victoria Hamilton (*Staff*)

Executive Director of Research Initiatives

Reporting Structure

SRCPAC Committees & Working Groups

Active or *Upcoming*

- *Habanero Operating Committee*
- Yeti Operating Committee
- Yeti Executive Committee
- Education Subcommittee

Research Computing Executive Committee

Shared Research Computing Policy Advisory Committee

Completed or *Suspended*

- Columbia Survey Working Group
- External Peer Survey Working Group
- Manhattanville Liaison Working Group
- RFP Committee
- Research Storage Working Group
- *Cloud Subcommittee*
- *Hotfoot HPC Operations Committee*
- *Intercampus Subcommittee*

Why Are We Doing This?

Meeting HPC Needs

Local Clusters
National Computing Centers
Cloud Computing

Why ***Share*** a Local Cluster?

Researchers Gain

- Time
- Local Expertise
- Access to a Larger Machine
- Flexibility

Columbia Gains

- Energy and Space
- Shared Staff and Hardware Costs
- Recruitment Advantage
- Happy Faculty!