

Components of Technology Supporting Data Intensive Research

Ron Hutchins

Associate Vice Provost for Research and Technology
and CTO

Georgia Institute of Technology

24 January, 2012

NSF Dear Colleague Letter: Cyberinfrastructure Framework for 21st Century Science and Engineering (CF21)

- “Today, every discipline of science and engineering is being **revolutionized** by the widespread use of comprehensive Cyber-infrastructure (CI). **Computing power, data volumes, and network capacities are all on exponential growth paths**, collaborations are growing dramatically, and all forms of CI---and **multiple communities spanning multiple agencies and international domains**---often must be brought to bear to address a single complex grand challenge problem, such as climate change. All of these developments are part of a ***revolutionary new approach*** to scientific discovery in which advanced computational facilities (e.g., ***data systems, computing hardware, high speed networks***) and instruments (e.g., telescopes, sensor networks, sequencers) are coupled to the development of quantifiable models, algorithms, software and other tools and services to provide **unique insights into complex problems** in science and engineering.

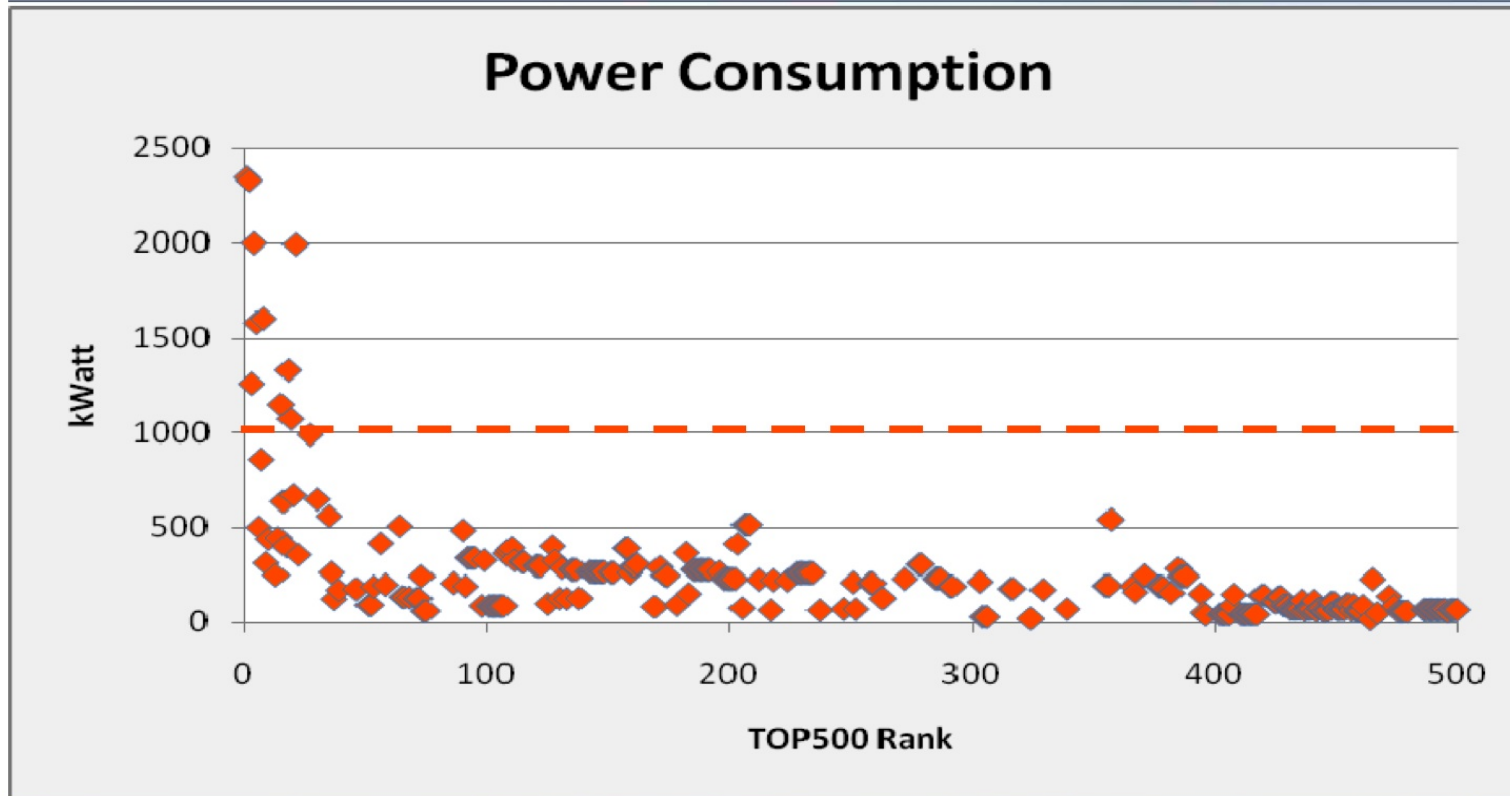
Challenges in Computing, Storage, & Networking

- Extreme-scale data/hosting facilities are necessary for web and cloud utility computing and storage – with dense power capability (30kw/rack). HPC doesn't go under a faculty member's desk any more.
- The Internet is no longer the “one” converged network to solve all problems – security, jitter, latency, privacy have all motivated a change to dedicated private research networks. The need goes far beyond standard ISP services
- Private fiber networks are proliferating in industry also: Insurance, Banking, Health Care, Wall Street: in 2009 more traffic was passed between tier 2 peerings than over the tier 1 Internet providers. Witness Google/YouTube
- Campus policies lean away from commercial off-site storage due to IP/export control/ etc. (some resolution exists, maybe)
- Multi-tenant facilities are necessary for separation and privacy.

Power! Sometimes a Lot!



Absolute Power Levels



What do Researchers Want? Philosophically...

- Control! If you can't give them what they want and the ability to control it, they will get it themselves – they won't bet their tenure on you easily or without experience.
- Custom technology solutions specifically tailored for their research needs.
- Invisible, easy to use/configure/manage stuff
- Free – unless you work with them on their proposal, they will not include funding for technology support in the general case.

Specifics – What Researchers Ask For

- Dedicated dark fiber for experimentation with new protocols (physical and above) and for bypassing campus firewalls, etc. for their dedicated private infrastructure.
- HPC with high performance storage/scratch – which implies dense power needs (above)
- High speed networking (L2 and L3) where needed – dedicated circuits/path but transparent to them (and free)
- Routing Tables, Flow traces, other network operational information
- DNS access for DNS research
- Long term network access for monitoring traffic (bot detection, etc) – start-up company magnet.
- Storage – different storage classes, multiple access methods.

... and...

- RNOC – a research support team that understands research and can quickly respond to requests
- Library as curator of research data for long term access as mandated by NSF.
- “D2D” as broader impact on research awards
- “GENI access / GENI Rack in SOX” – a larger scale infrastructure to work with to provide scale to the research infrastructure.
- Move the compute to the data for extreme data sets – this one is fairly new to me.

Resources in the Southeast and at Georgia Tech

- Southern Crossroads (FLR/SLR)
 - 3x10G Internet2, 10G NLR, 10G google, 10G ESNNet, Atlanta IX, peerings throughout the region and up the East Coast.
- <http://realestate.gatech.edu/hpc/index.php>
- SOX/Peachnet Cloud storage project – 1 PB, two locations, multiple access methods, replicated, etc.
- K-12 connections, Georgia Aquarium, other

Summary

- Researchers require special focus and care
- Standard help desk will not work – research support requires a understanding of technology and some learning into the research discipline to find the right fit.
- Mass customization – mass personalization is necessary in supporting research
- Infrastructure is expensive!
- Science demands collaboration which requires sometimes specialized infrastructure.
- If you disappoint them, they will not come back.