

Advancing the University's Mission Through Cyberinfrastructure

University of Missouri
CI Day
March 3, 2015

Jim Bottum

CIO, Clemson University
Presidential Fellow, Internet2



Personal

- **NSF Program Officer**, Office of Advanced Scientific Computing (1982-86)
- **Executive Director**, National Center for Supercomputing Applications, UIUC (1986-2001)
- **VP and Inaugural CIO**, Purdue University (2001-2006)
- **VP and CIO**, Clemson University (2006-Present)
- **Inaugural Presidential Fellow**, Internet2 (2012-Present)



Cyberinfrastructure

- **Infrastructure** - used since the 1920s to refer to the roads, power grids, telephone systems, bridges, rail lines, and similar public works required for an **industrial economy**.
- **Cyberinfrastructure** refers to distributed computer, information and communication technology - necessary for a **knowledge economy**.
- **Cyberinfrastructure** can now enable the acquisition, storage, management, integration, mining, and visualization of data in **virtually all disciplines**.



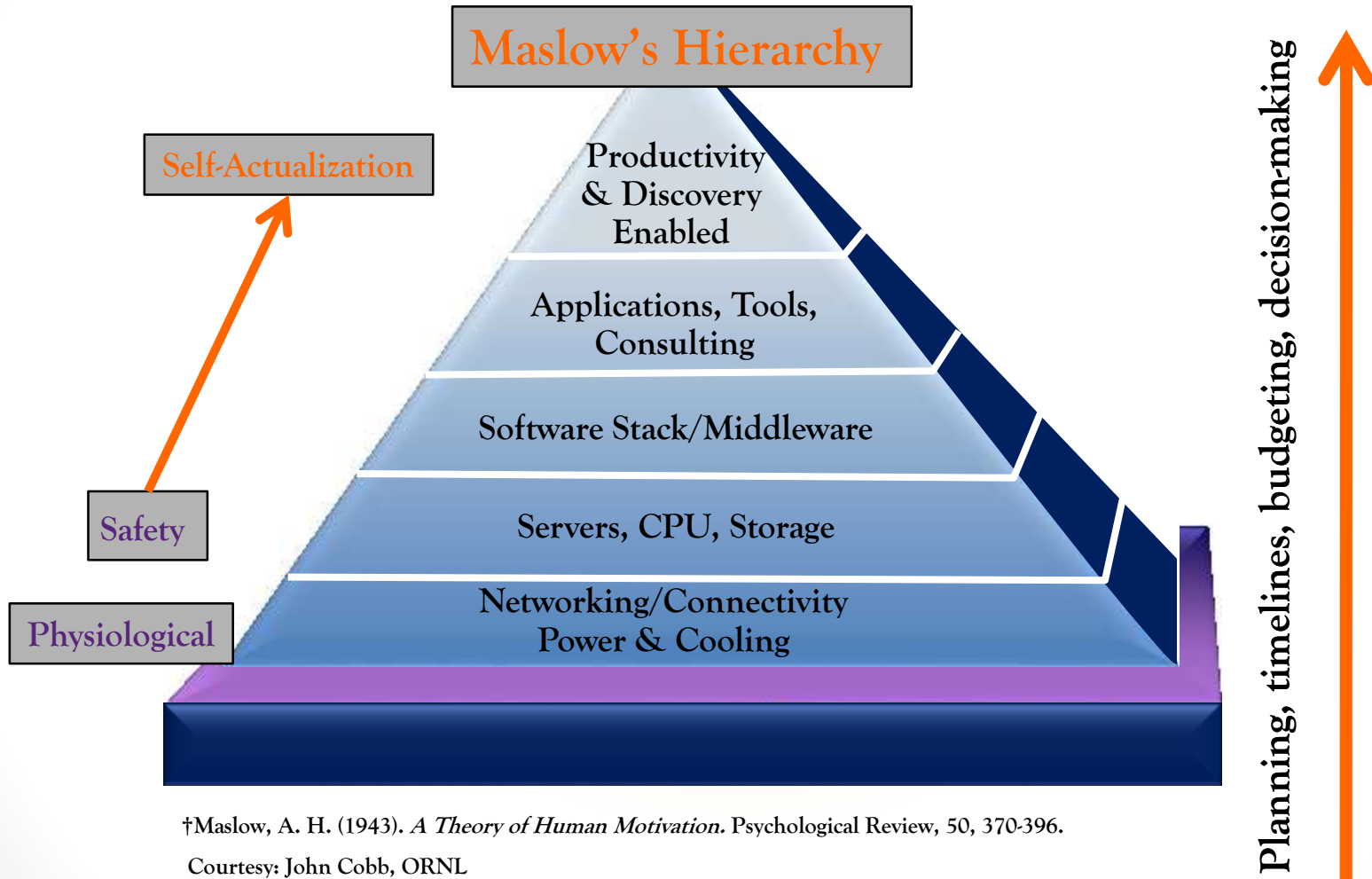
Is CI Important?

“Leadership in cyberinfrastructure may well become the major determinant in measuring pre-eminence in higher education among nations.”

**Arden Bement, Former Director,
National Science Foundation**



How do we think about it?



†Maslow, A. H. (1943). *A Theory of Human Motivation*. *Psychological Review*, 50, 370-396.

Courtesy: John Cobb, ORNL

CI was Supercomputing in 1986

Boutique business



512K
“Fat Mac”

\$14M Cray
Supercomputer
200 Mflops
4M words
memory

Physicist John
Kogut talks to the
Mac, the Mac talks
to the Cray

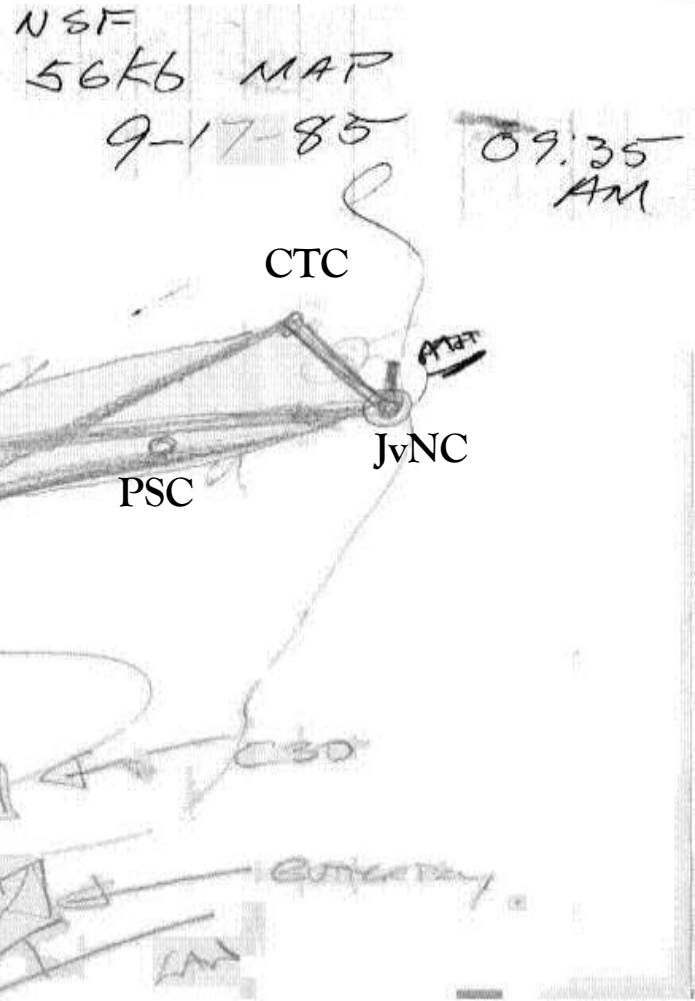
Technologies change but supporting researchers is largely the same

Supercomputing Today



Centers & Networks - 1986

Original NSFNet designed at NCAR,
Boulder, CO, September 17, 1985



Map of United States

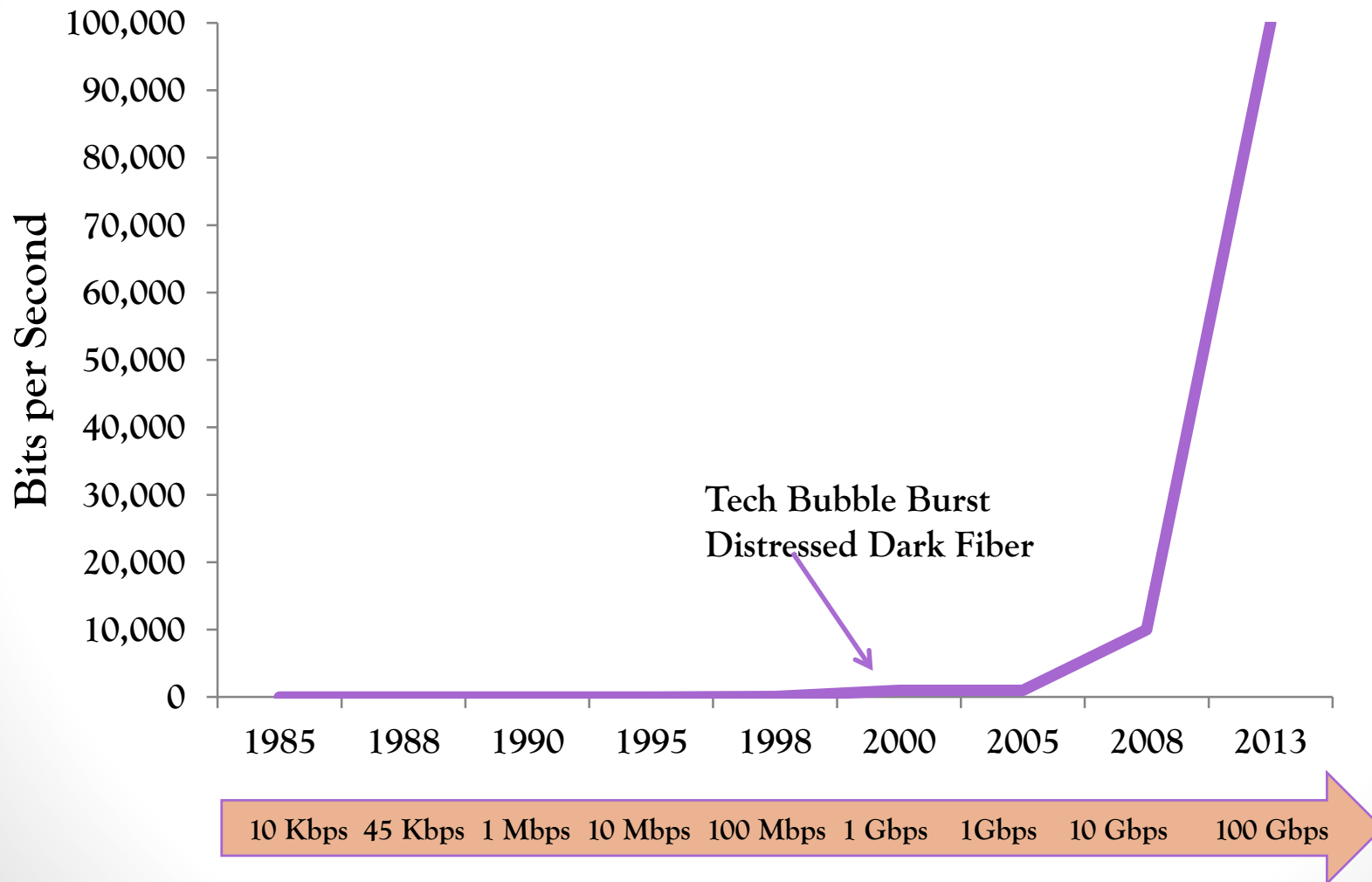
Leadership in Wide-Area Networking

100 Gb/s Internet2 Innovation Platform Pilot Site



R & E Networking Evolution

Exponential Network Growth



The Planetary Cloud Computer Is Connected to a Billion Cray-Speed Smartphones

1 Billion Smartphones Shipped In **InformationWeek**
2013

1/28/2014

University of Missouri
Alum and my boss and
mentor of 15 years – Dr.
Larry L. Smarr

The New York Times
The iPad in Your Hand: As Fast as a
Supercomputer of Yore

By JOHN MARKOFF MAY 9, 2011 3:45 PM 14 Comments

1988 Cray Y-MP



2010 iPad 2

ipad 2



The Changing Digital Age

2005



Photo Credit: Luca Bruno, AP

The Changing Digital Age

2013



Photo Credit: Michael Sohn, AP

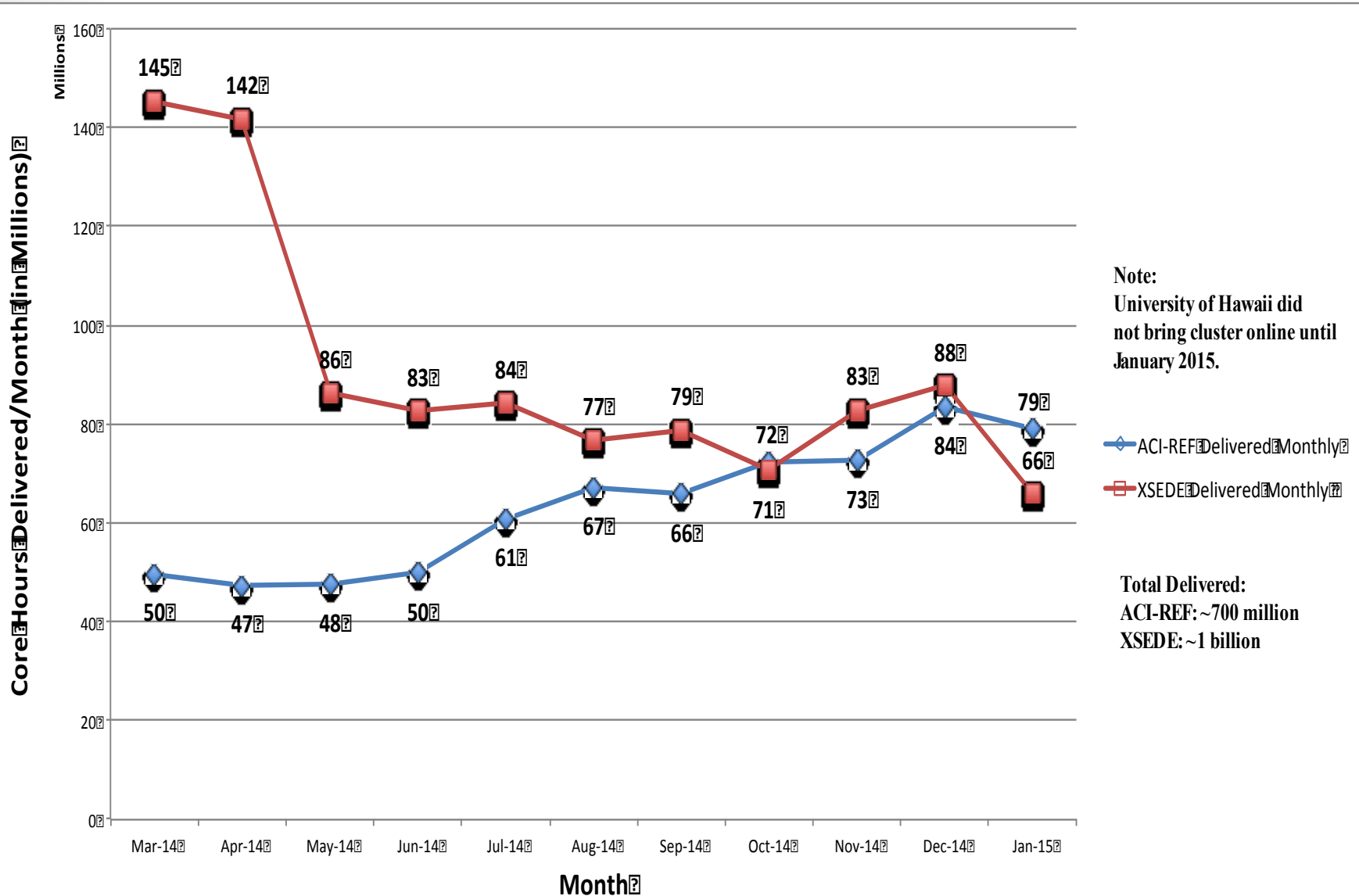


National programs primed the pump for cyberinfrastructure and we need the national programs...

...but the bulk of the action is at the campus level today.



XSEDE & ACI-REF Core Hours Delivered



ACI-REF Universities



Project:
NSF Award
ACI-1341935
\$5.3M, 2-year award

Purpose:
Build a community of
facilitators across the
partner campuses to
support research

Year 1 Progress

“Without Bob Freeman’s [Harvard ACI-REF] work through ACI-REF I do not think I would have been able to complete my bioinformatics project. I was not aware of ACI-REF at the time I started my HPC bioinformatics work.”

-Zack Lewis, Harvard University PhD Candidate, on Harvard ACI-REF Bob Freeman

“Putting someone so knowledgeable and so capable of explaining procedures in simple but powerful terms greatly enhances the effective power and utility of the millions invested in the hardware.”

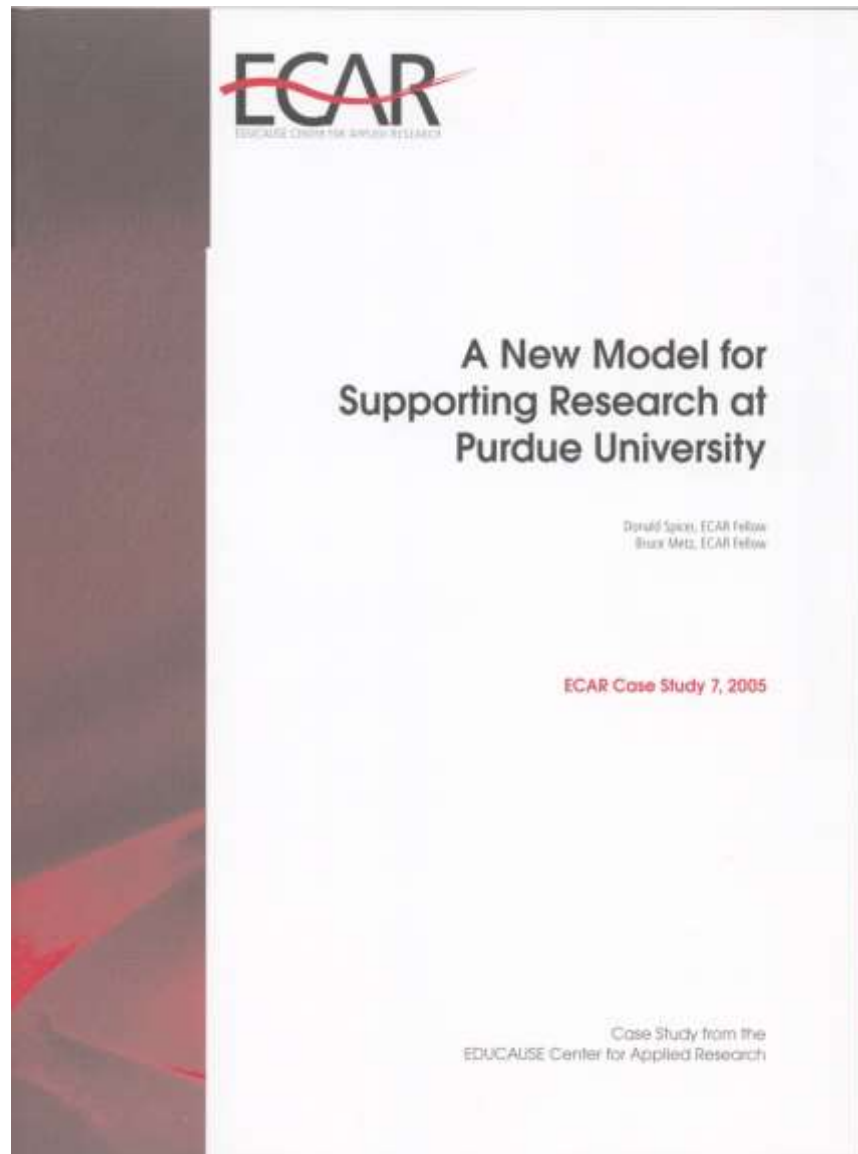
-Thomas Givnish, Professor of Botany commenting on UW-Madison infrastructure and ACI-REF Lauren Michael

It's About Campuses and People!

- Computational scientists, visualization specialists, and data scientists from various domains that provide outreach and support to users across the sciences and humanities.
- Quite simply... the men and women behind the magic!



CI Strategy: Purdue



All Universities Are Wrestling With These Issues...



Advancing Research Computing on Campuses: Best Practices Workshop

ACI-REF and NCSA Collaboration

March 17-19, 2015

Clemson, South Carolina



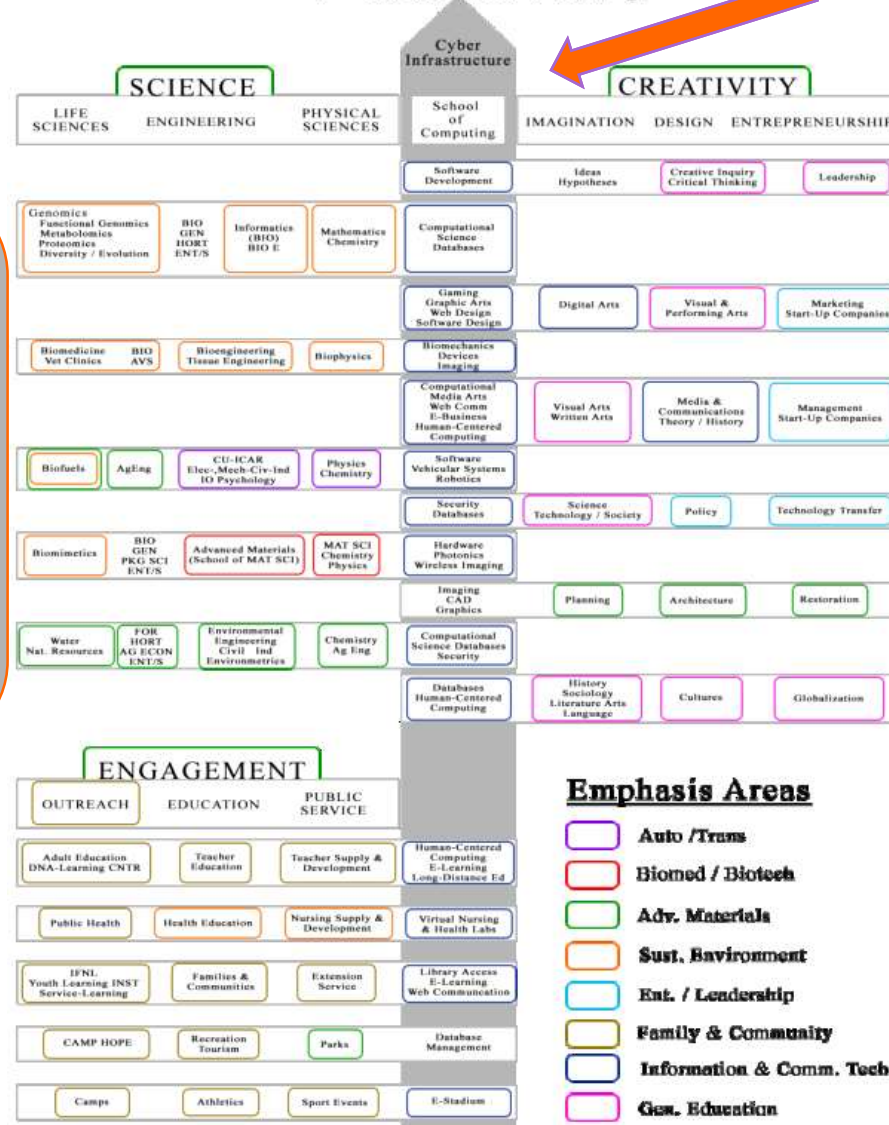
One Campus' Evolution – The Clemson Experience



Institutional Alignment

Cyberinfrastructure Backbone: Investment, NOT Cost

The Innovative University



Emphasis Areas

- Auto / Trans
- Biomed / Biotech
- Adv. Materials
- Sust. Environment
- Ent. / Leadership
- Family & Community
- Information & Comm. Tech
- Gen. Education

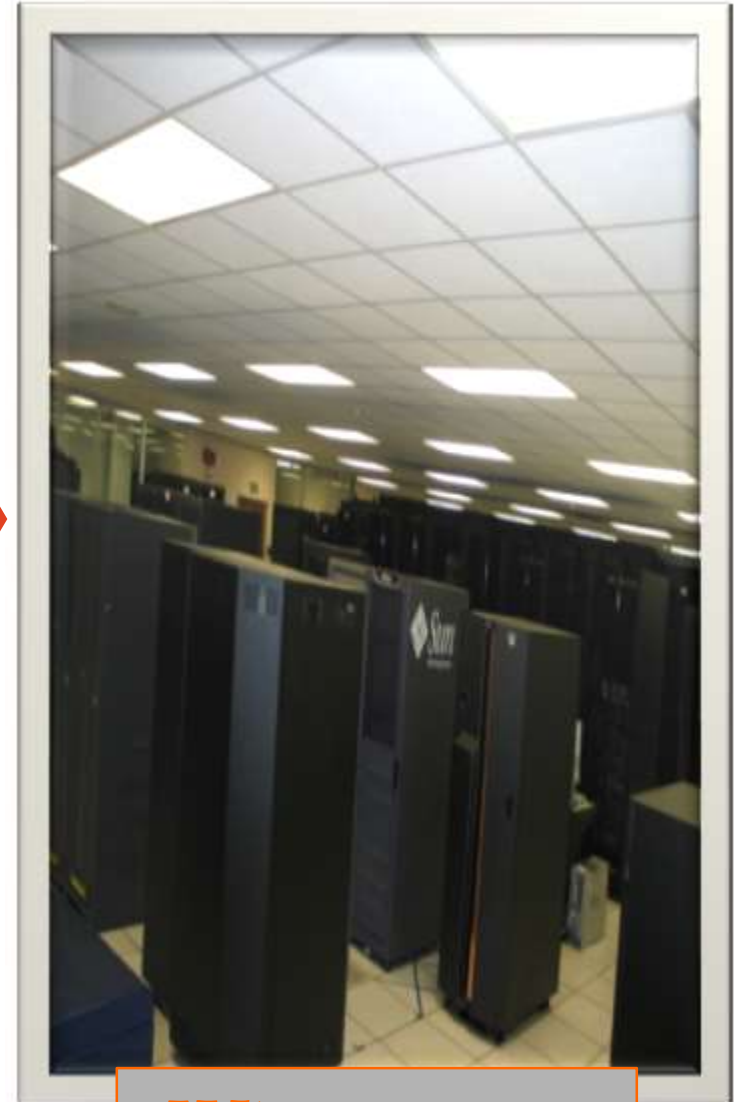
“Cyberinfrastructure is the primary backbone that ties together innovation in research, instruction, and service to elevate Clemson to the Top 20”

Doris Helms,
Former Clemson Univ.
Provost

Reality: Aligning Aspirations With Priorities



Fall 2006



Winter 2007

Don't overlook this resource...



Open Science Grid

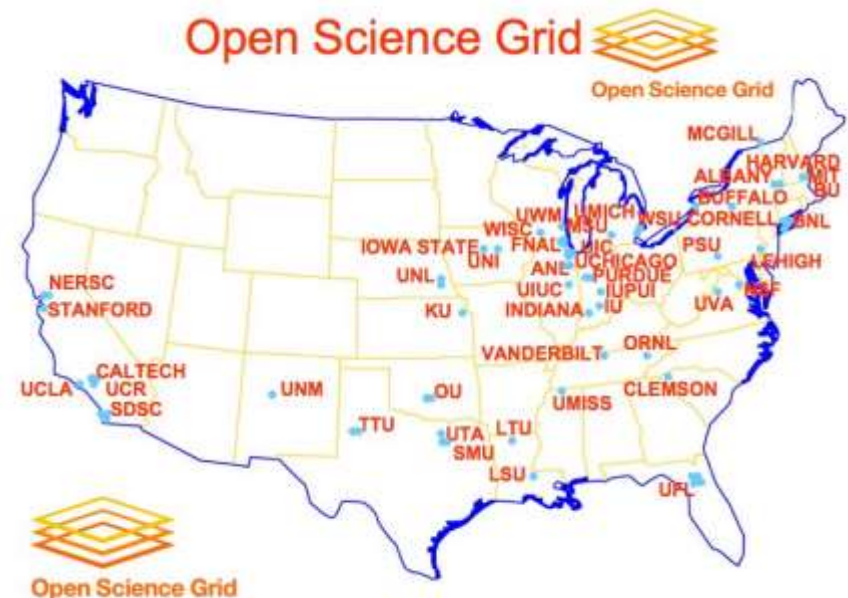
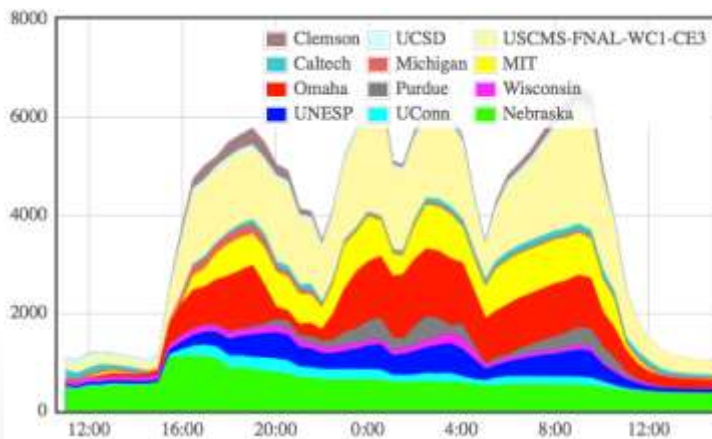
The Open Science Grid (OSG) supports science such as

- High Energy Physics: CMS and ATLAS
- Nanoscience: NANOHUB
- Structural Biology: SBGrid
- Community VO (multiple sciences): Engage

U of Missouri

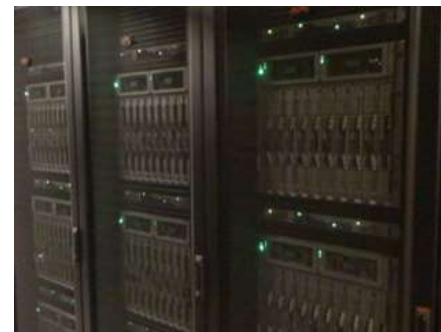
Dong Xu – 45,143 hrs

Plant Biology

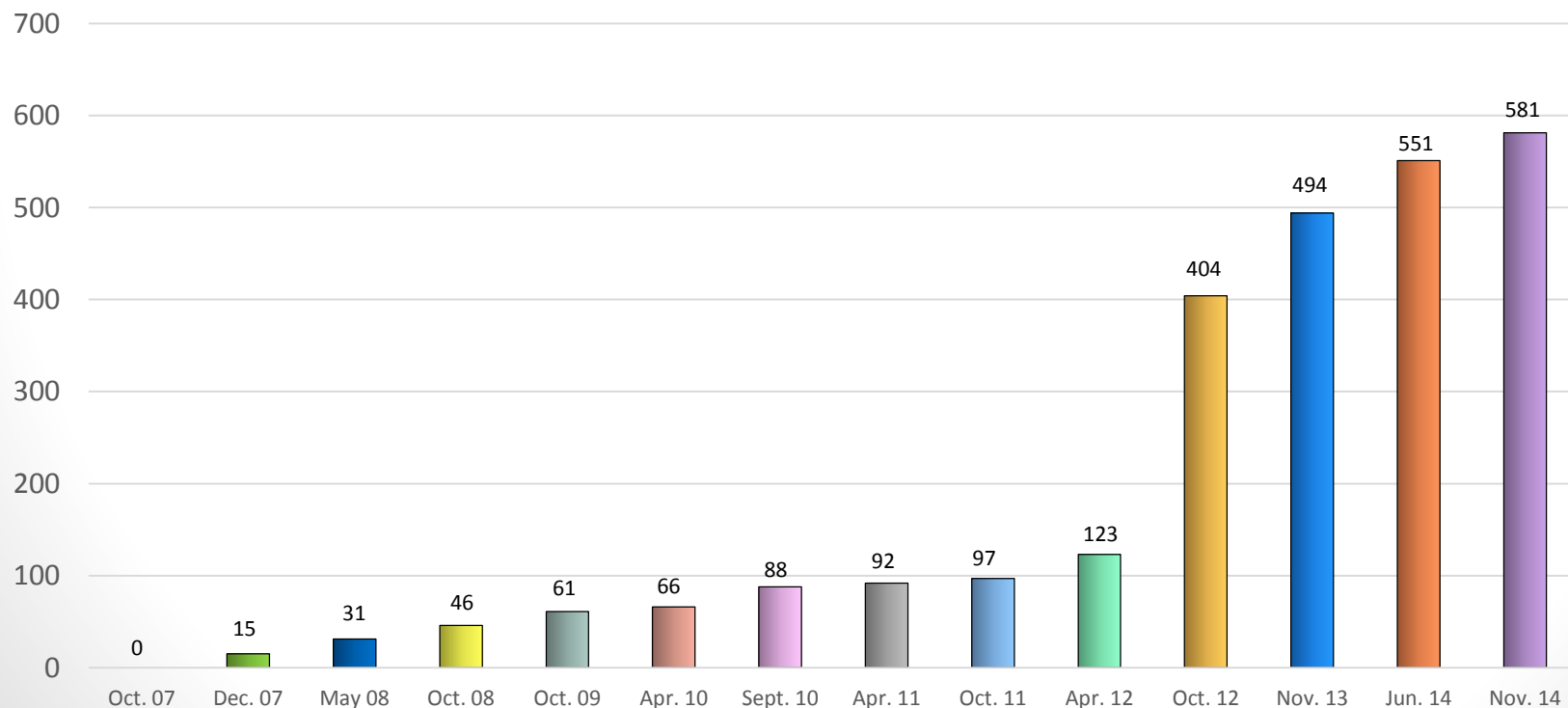


“Palmetto” HPC Infrastructure

- Community HPC Clusters
- Highly leveraged instrument for research
- November 2014: **Top 5** among public academic institutions (without a national center)
- November 2014: **#89 worldwide**



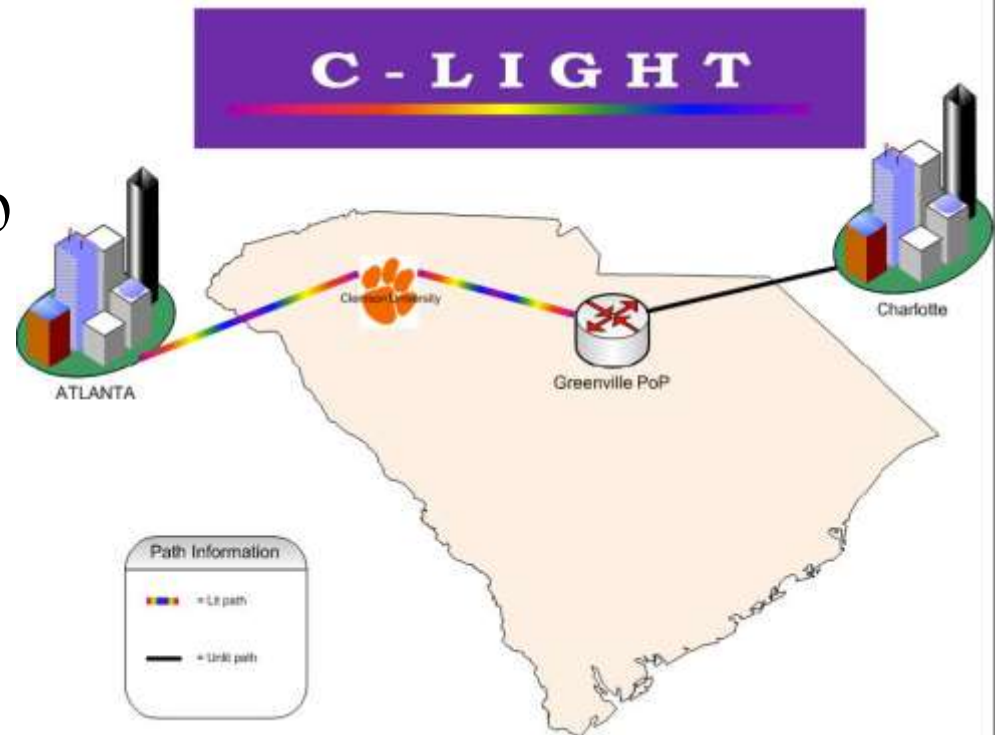
Computing Capacity - TeraFLOPS



We Want HPC But We Don't Have A Network...

Clemson in 2007: 100Mb/S
8 Months to Upgrade to 200 Mb/S

Clemson in 2012: 100Gbs
Internet2 Innovation
Platform Pilot Site
'Condo of Condos' Model



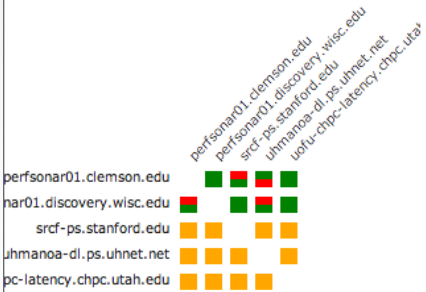
C-Light: Privately gifted fiber
Atlanta, GA → Charlotte, NC
LLC established May 2007

First South Carolina regional optical network, CLight, integrated with national research networks

100 Gb/S Network Outputs

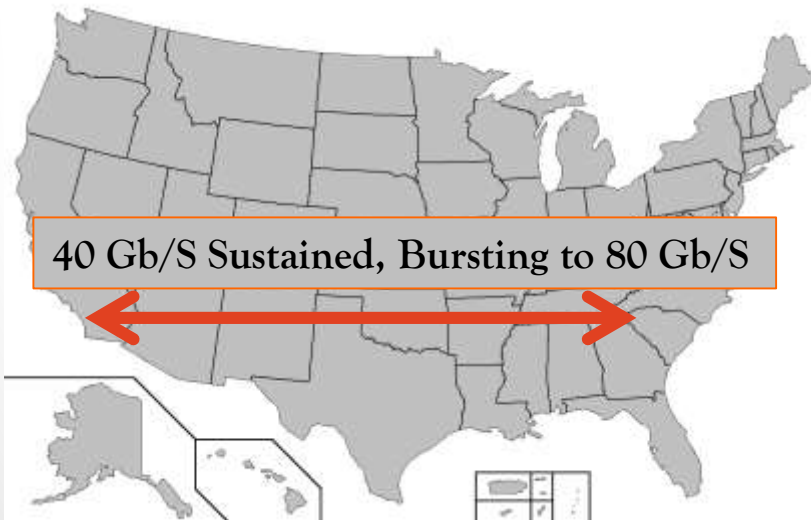
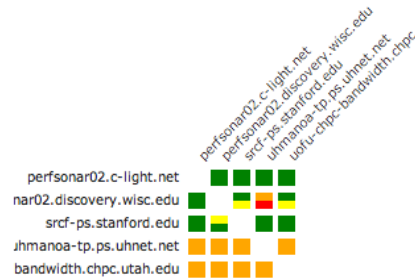
OWAMP

■ Loss is 0
 ■ Loss is greater than 0
 ■ Unable to retrieve data
 ■ Check has not yet run



BWCTL

■ Throughput >= 100Mbps
 ■ Throughput >= 10Mbps
 ■ Throughput < 10Mbps
 ■ Unable to retrieve data
 ■ Check has not yet run

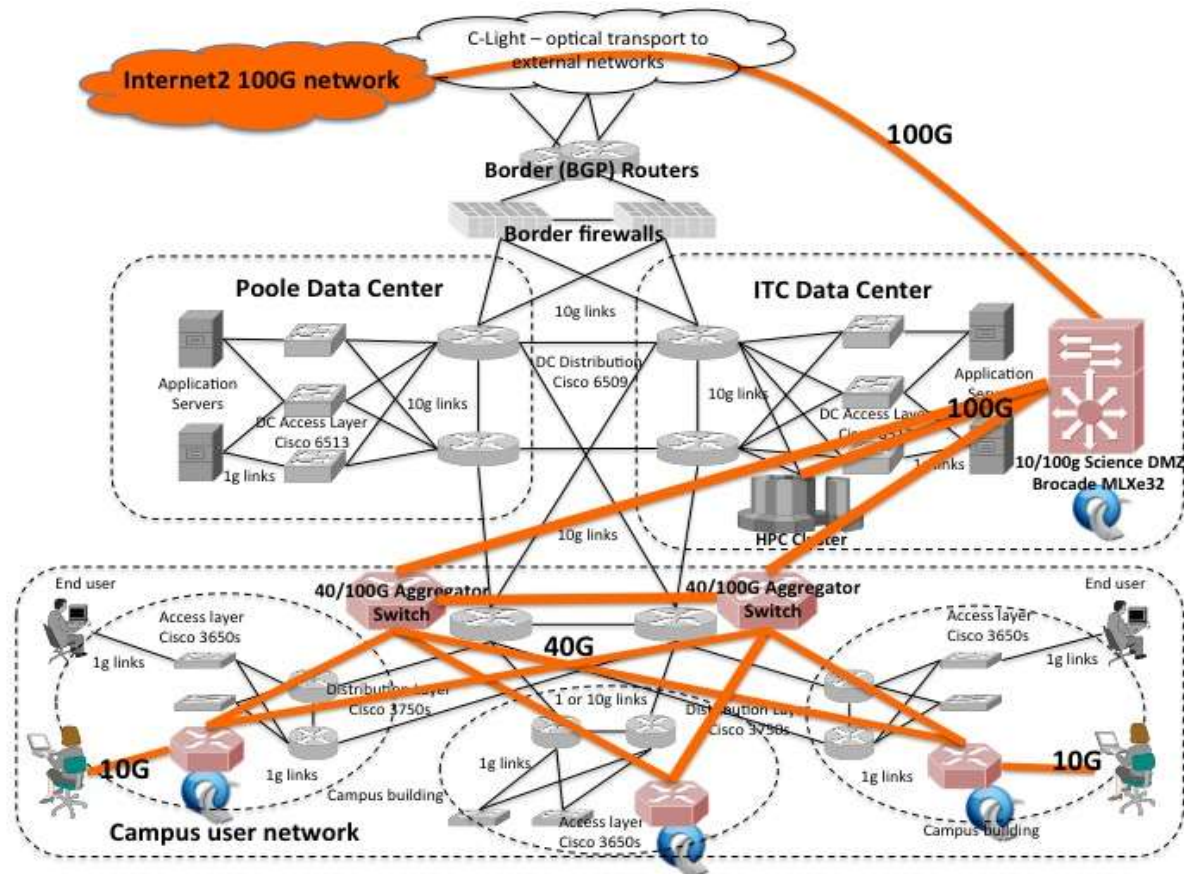


Partnership with USC and its Shoah Foundation (Holocaust Interview Archives)

- Integrated file systems
- Data sharing (Clemson Faculty/Libraries have data access)
- USC has DR at Clemson over 100 Gbps connection

Next Generation Networking Comes to Campus

National Science Foundation CC-NIE Award (\$1M)*



Allows Clemson to expand SDN and bring 10 and 40 Gbps to some labs and classrooms
Expertise developed in C-Light expansion, expertise base contributed to award

CloudLab

- One of two \$10M NSF awards for research in cloud computing
- CloudLab provides a “meta-cloud” for building clouds
- Build your own cloud on our hardware resources
- Agnostic to specific cloud software...
 - Run existing cloud software stacks (like OpenStack, Hadoop, etc.)
 - ... or new ones built from the ground up
- Control and visibility all the way to the bare metal
- “Sliceable” for multiple, isolated experiments at once

With CloudLab, it will be as easy to get a cloud tomorrow as it is to get a VM today.

(28)

It's Not Coming, It's Here...



INTRODUCING
BIG DATA
EXPO

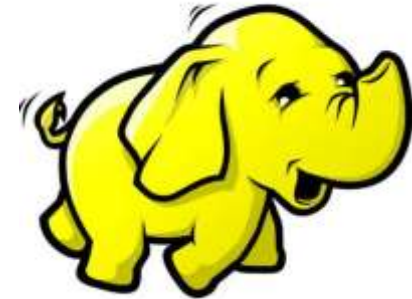
JUNE 11-14, 2012
BIGDATAEXPO.NET
@BIGDATAEXPO

“Any data that we don’t yet understand well enough to computerize can be called big data.”

George Strawn
EDUCAUSE Review

Hadoop Cluster at Clemson

- 19 nodes
 - 16 data nodes (Dell PowerEdge C220X)
 - 256 cores
 - 4TB memory
 - 192 TB storage
 - 3 management nodes (Dell PowerEdge C6220)
 - 24 cores each
 - 192 GB memory
 - 1,500 GB storage
 - 10G Ethernet (each)
 - Running: YARN's ResourceManager & NameNode Service
- Cluster configured with Hortonworks HDP 2.2 software stack
- Provides a production Hadoop environment with an easily accessible web interface.



Hadoop Training Classes

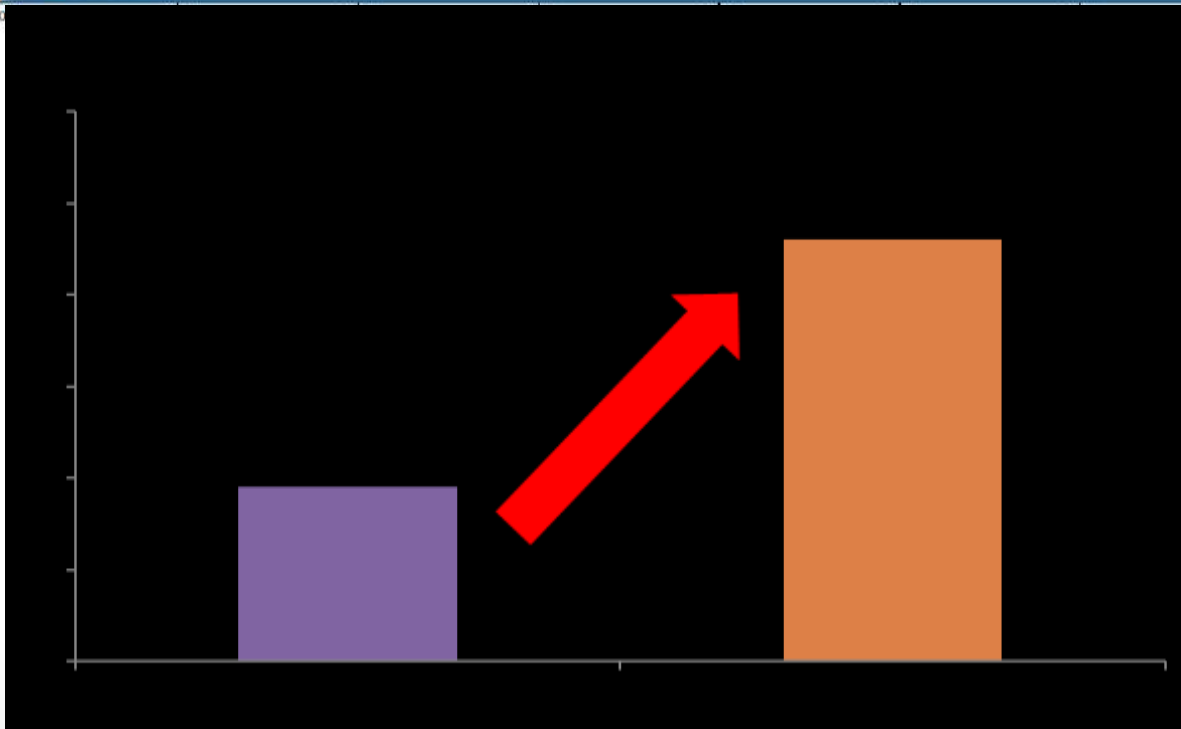
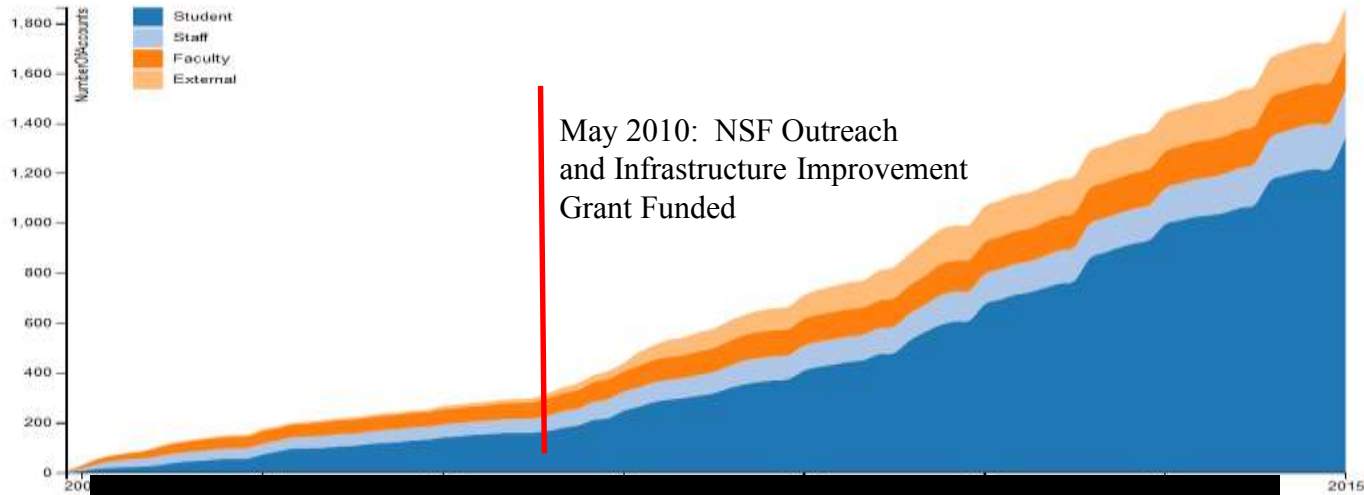
Department

- Bioengineering
- CCIT HPC
- CCIT Software Development
- Chemical Engineering
- Chemistry
- Civil Engineering
- Customer Rel & Learning Tech
Cyberinfrastructure Tech Integ
- Economics
- Elec. & Computer Engr.
- Environmental Engr & Earth Sci
- Experiential Education
- General Engineering
- Genetics & Biochemistry
- Industrial Engineering
- Infrastructure Services & Ops

Department

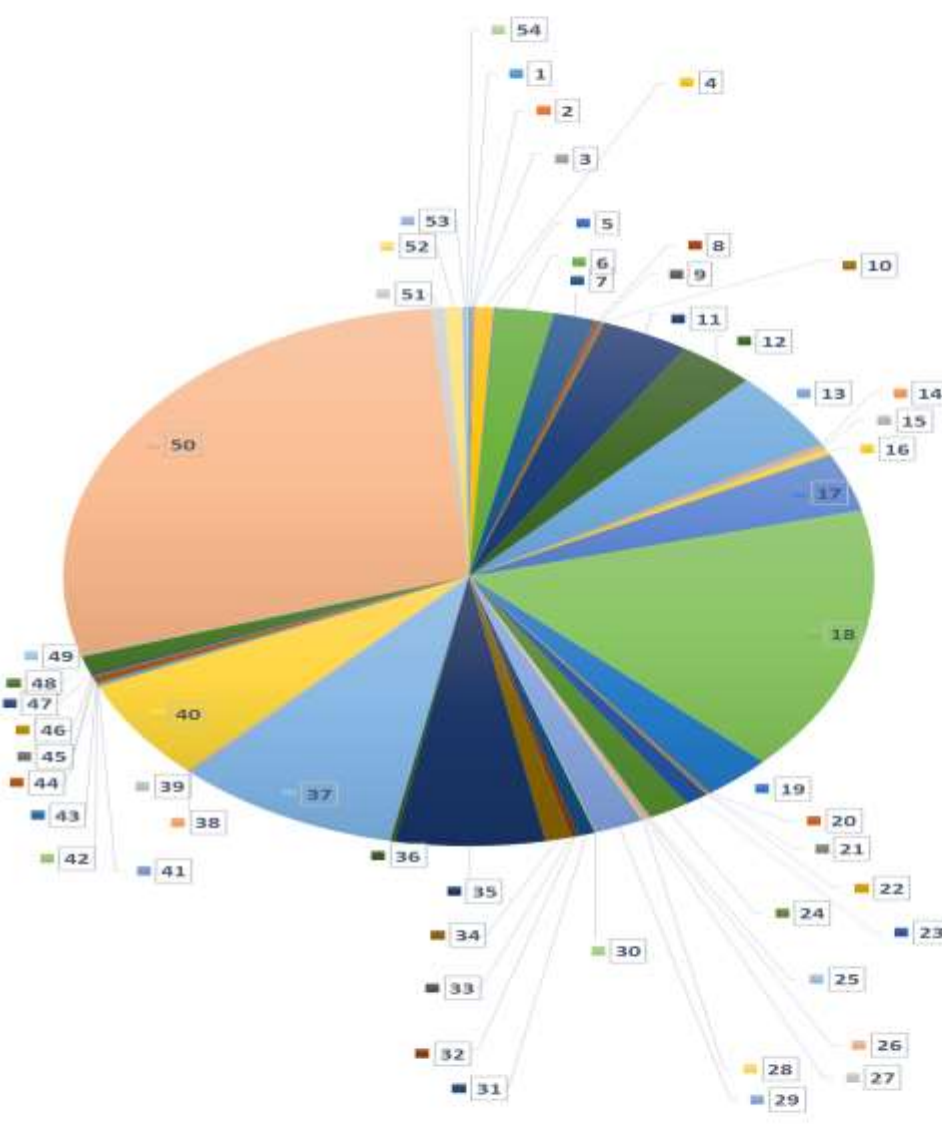
- International Programs
- Law Enforcement & Safety
- Management
- Mathematical Sciences
- Mechanical Engineering
- Medicaid IT Services
- Network Services & Telecomm
- Physics And Astronomy
- Public Health Sciences
- Research Safety
- School of Ag for Env Science
- School of Computing
- Univ Facilities Support Svcs
- VP Finance & Operations

Community Growth



Palmetto Cluster Areas of Study

Community Growth



Faculty Recruitment – Competitive Advantage

“As a computational materials scientist, **I was seeking a university** that was investing not only in developing supercomputing facilities but also in **providing the users an efficient support system**. In my experience, **the latter was usually missing in most universities**.

However, when **I asked to test Palmetto** before deciding to join Clemson, I quickly realized that Palmetto had both, and both were very topnotch.

That definitely put Clemson on top of my list as future destinations. In the last 8 months at Clemson, has further reinforced the impression, and the synergistic interactions between computational materials scientists, computer scientists and CITI that I have observed is simply exemplary.

Most importantly, I have been able to focus on science, rather than worrying about overheating of my cluster, that too without the hassle of waiting forever on queues - **for once, I am having my cake and eating it too!**”

-Sapna Sarupria – Chemical Engineering
Former Faculty at Princeton University

HPC: Research Impacts

Funding

- **\$96.8 M** – awards to Clemson HPC users since 2009
- **\$15.5 M** – awards to Clemson HPC users in FY14

Degrees Conferred

- At least **36** PhD degrees produced from research groups that make use of the Palmetto Cluster

Publications

- At least **115** publications made possible because of Palmetto Cluster (since FY10)*
 - Majority (~90%+) of these are refereed journals

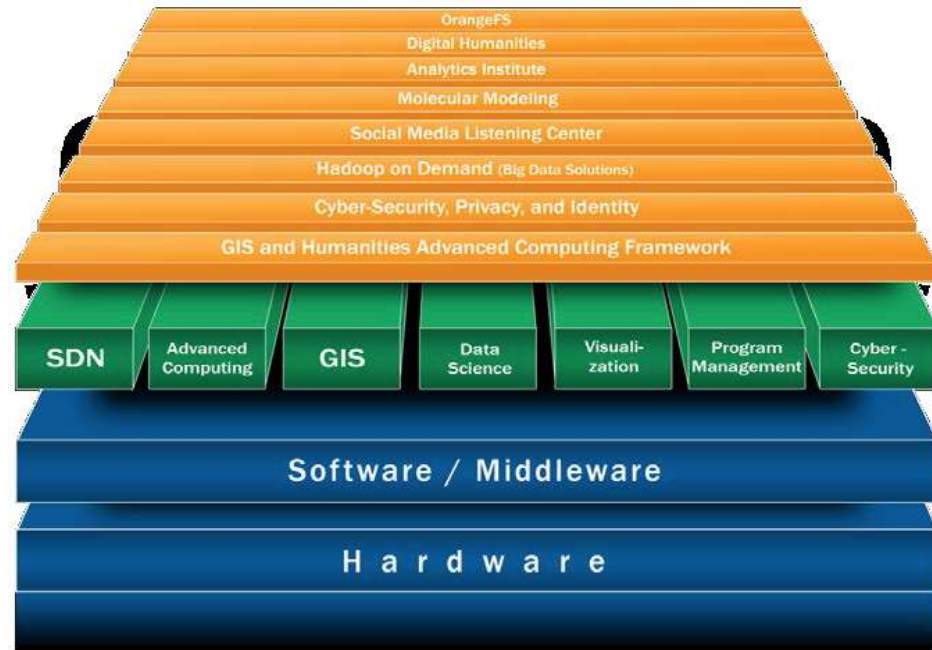
Center of Excellence in Next Generation Computing & Creativity

Partnership with IT and Faculty

- Working collaborations between faculty, staff and students, the Center seeks to have a transformational impact on research and education.
- Academic interests driving IT priorities

Focus Areas

- Security, Identity, and Privacy
- Advanced Computing
- Next-Generation Networking
- Data Science
- Digital Humanities
- Creativity (cross-discipline)



Funding

- Private Sector/Industry Partnerships
- Federal Grants
- Clemson University

Center of Excellence in Next Generation Computing & Creativity

- Partnered with Faculty in Electrical & Computer Engineering, Arts, Architecture & Humanities, and Athletics with focus areas in:
 - Advanced Networking and Software Defined Networking
 - Information Security and Assurance Research
 - Digital Literacy & Creativity
- Dell, BigSwitch, Adobe, & Clemson Partnership on Next-Generation Tools & Technologies



Door to Humanities: Adobe Partnership

Major gift (\$11.3M) to the University which enables:

- Creative Cloud access for all Clemson faculty, staff, and students
- Access to Digital Publishing Suite
- Training Opportunities
- Student Competitions
- Adobe Digital Studio in Cooper Library
- Access to Adobe Marketing Cloud & Anywhere in Progress



Social Media Listening Center(s)



- Reps from other universities and industry continue to visit the SMLC. Most recently, University of Florida; Corning.
- Corporate projects with KEMET, Rawle-Murdy, Hampton Inn
- Enough projects to keep three interns (each working 10-15 hours a week) busy each semester

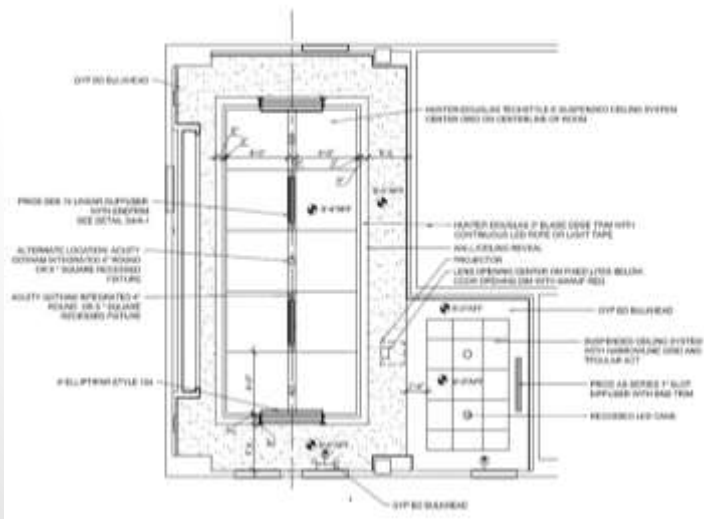
- External clients looking to interns as potential new hires after graduation
- Communication Studies faculty were on broadcast TV regarding significant breaking news; formal partnership born from the collaboration

Partnership with Dell and Salesforce.com



Security Operations Center (SOC)

- Student centered living laboratory; practical experience in security related event detection, response, and management.
- “Showcase” center located on campus.
- Idea is to bring technology, process and procedures from industry into the center.
- Create a talent pool for industry to draw from and to enhance the internship experience.
- Scheduled to open Spring 2015



Advanced Training Facilities

New generation of multipurpose advanced computing student laboratories.

The Digital Resources Laboratory

<http://citi.clemson.edu/drl>



- 15-panel visualization wall
- 16-node computational cluster
- 40 Gbps to the room
- SDN
- Video teleconferencing
- Centrally located, open to all faculty



Stay In The Creative Phase!



Massachusetts Green High Performance Computing Center (MGHPCC)

Hydroelectric Power

- MIT, Harvard, UMASS, NEU, BU
- 5MW day one connected load
- Airside economizers (green)
- ca. 640+ racks in “20 rack pods”
- 10% special computing spaces
- Open Feb 2013, first science was May (ATLAS)

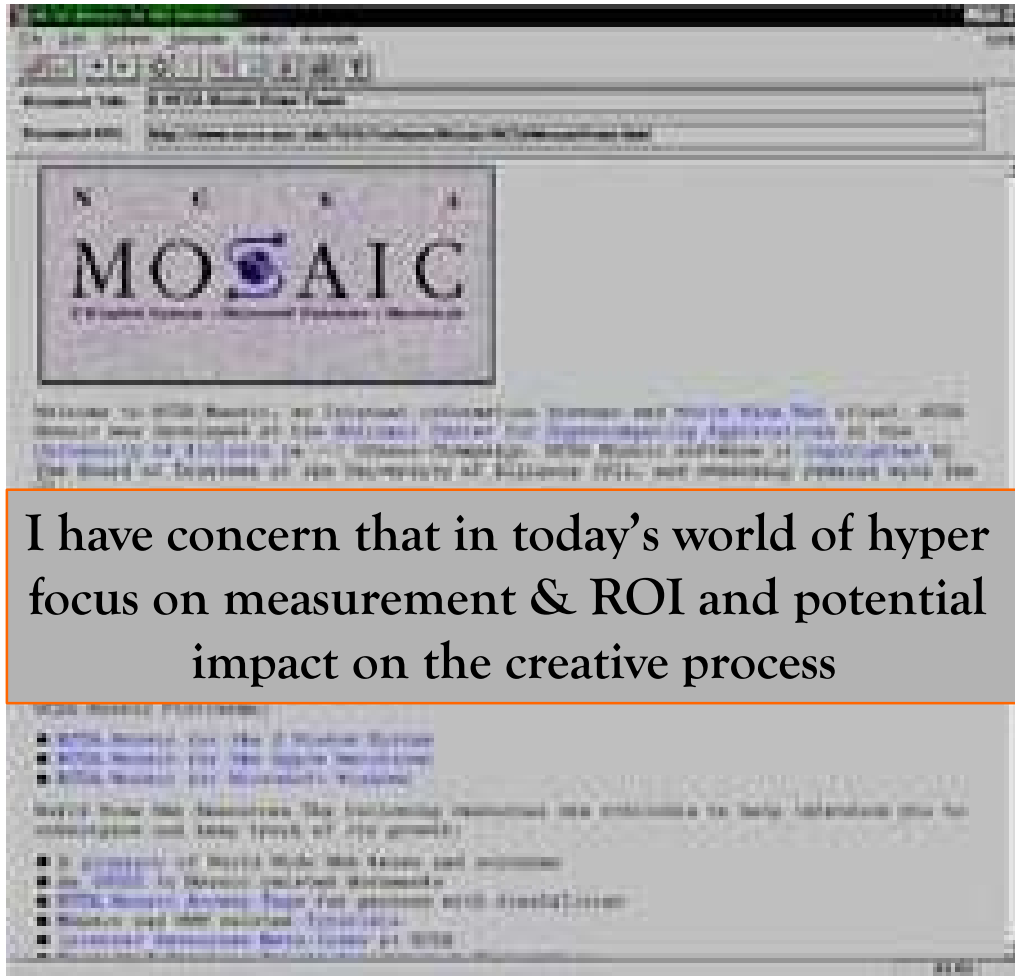


Clemson MRI Proposal: Acquisition of a Regional Advanced Computing System

- NSF Solicitation # 15-504
- Project budget: \$5,697,489
- Proposes multi-campus high performance computing equipment acquisition.
- Would provide dedicated HPC access for faculty & other researchers at the partner campuses for 3-year award period.
- Partner campuses committed operating funds for at least 2 additional years after award period (5 years total).
- Partners:
 - Clemson University
 - College of Charleston
 - Georgia State University
 - University of Georgia



And They Called It Mission Creep...



I have concern that in today's world of hyper focus on measurement & ROI and potential impact on the creative process

Steepest growth during recession of 1990

Thanks for having me!

Discussion