



# Ever Forward: Lustre at Indiana University

Lustre User Group, Japan

November 2, 2017

Stephen Simms

Manager, High Performance File Systems

[ssimms@iu.edu](mailto:ssimms@iu.edu)

Indiana University



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY

Greetings from Indiana!





# And Indiana University



# Indiana University has 8 Campuses

In 2016 Total Student Count 114,160

IU has a central IT organization

- UITS
- 1200 Full / Part Time Staff
- Approximately \$160M Budget
- HPC resources are handled by Research Technologies division



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# Lustre at Indiana University



**RESEARCH  
TECHNOLOGIES**

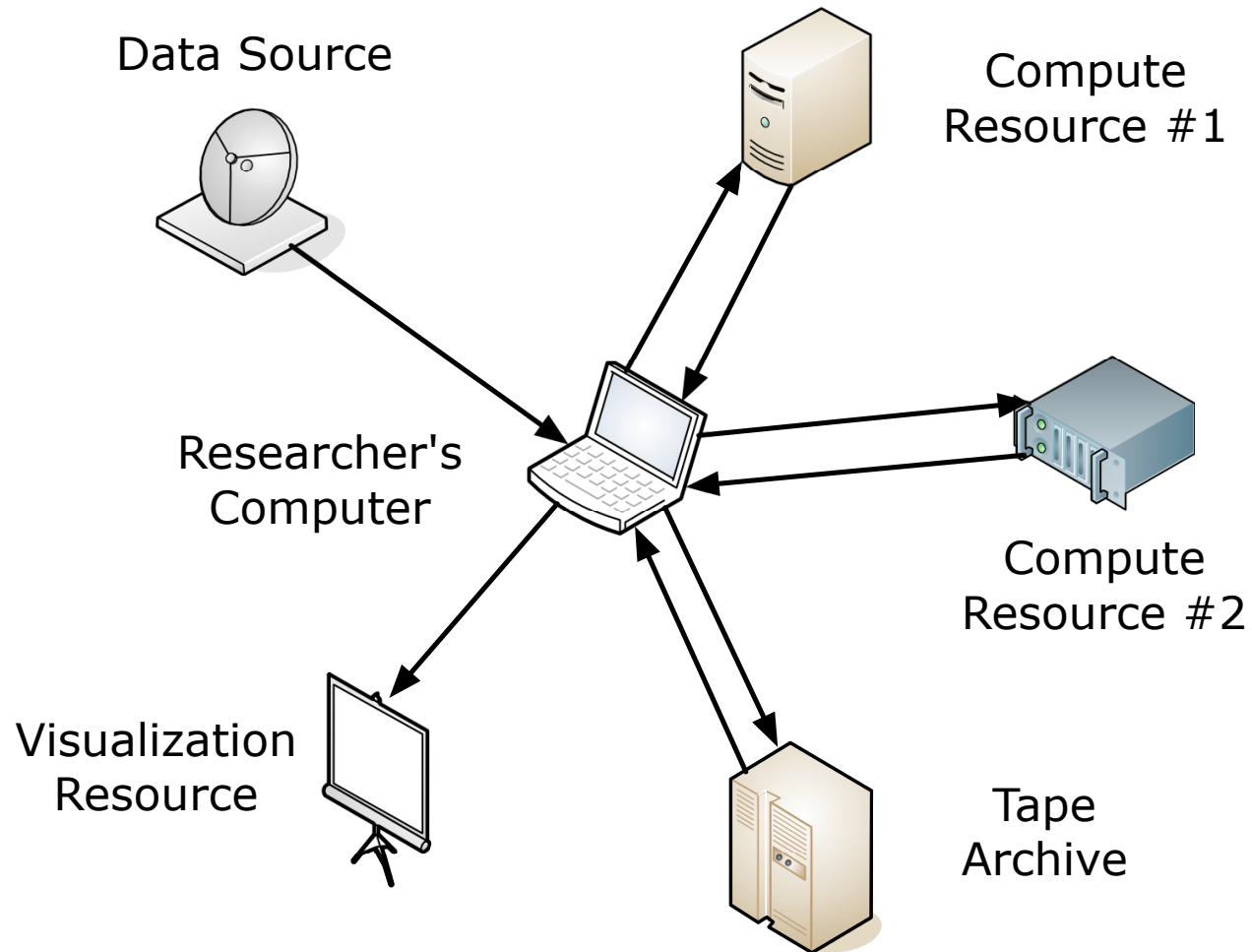
INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# Multi-Resource Workflow



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



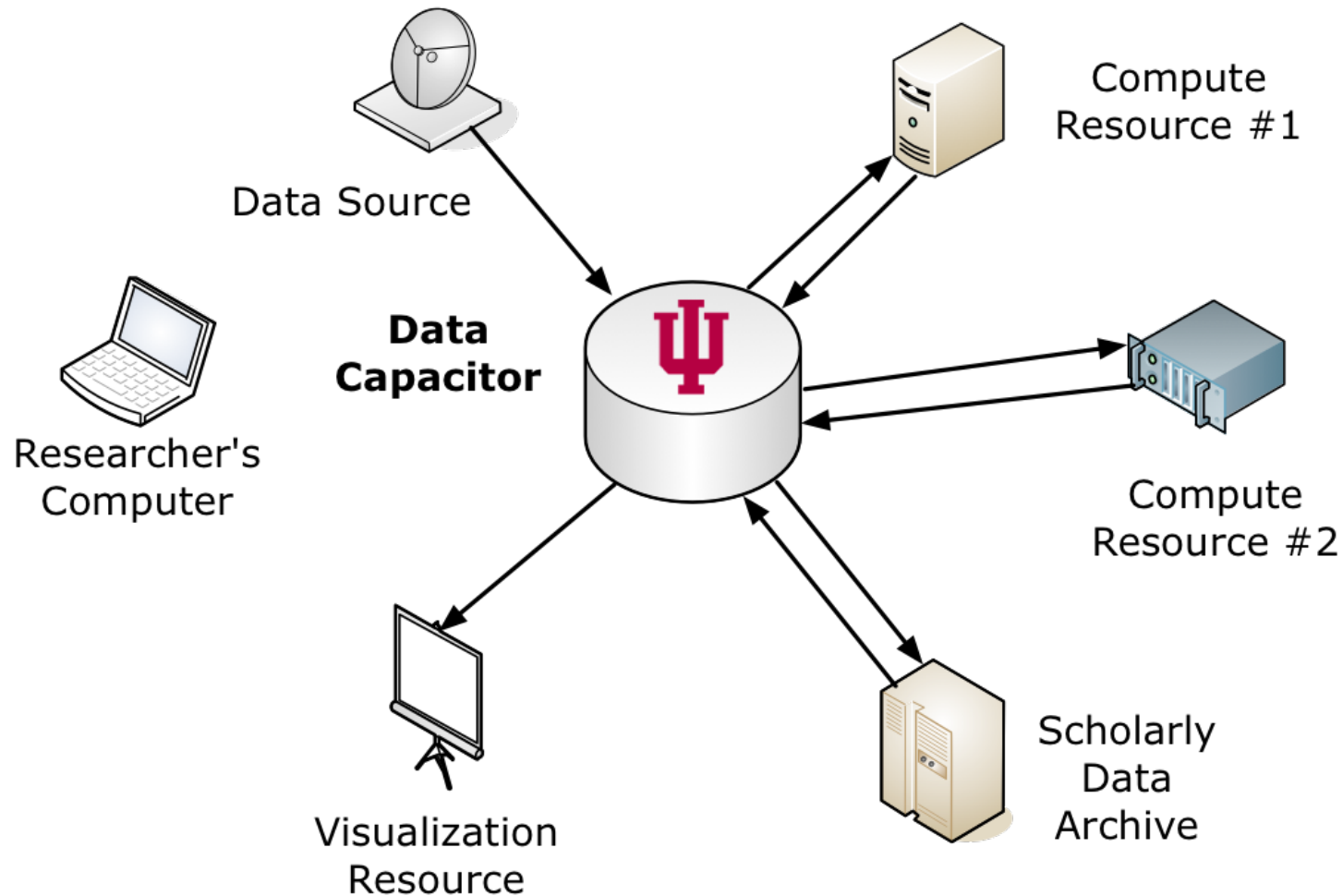
**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY





# The Idea in 2005



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# 2006 Data Capacitor

- First implemented in 2006
  - NSF MRI Grant
  - 535 TB – DataDirect Networks S2A9550
  - Site-wide file system with clients in Bloomington and Indianapolis spanning 80 km
  - Partnership between Dell and Data Direct Networks
  - 24 OSSs x 10 Gb/s



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# Lustre is scalable – 55 PB at LLNL





# Lustre is fast – 1 TB/s at ORNL



<http://www.flickr.com/photos/dvd5/163647219/>



# Lustre can support thousands of clients





# Expanding Lustre's Reach



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# 2006 Single 10Gb Client Tests

977 MB/s between ORNL and IU (95.3%)

Using 10Gb TeraGrid connection

Dell 2950 client

2x dual 3.0 GHz Xeon

4GB RAM

1 Myricom Myri10G card

Ethernet mode

Lustre 1.4.7.1

16 (of 24) Data Capacitor OSSs



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY

# 2007 Bandwidth Challenge Idea

Use the Data Capacitor across 2000 miles

Build a second small one in Reno, Nevada

Show science that RT supports every day

- We support a diverse community

Saturate a single 10Gb connection

Use a production network

- Connected - I2, NLR, TeraGrid

- Acquisition and Visualization (IU)
  - Live Chemistry Instrument Data
- Rare Archival Material (RIT)
  - Humanities
- Acquisition, Analysis, and Visualization
  - Trace Data (TUD)
    - Computer Science
  - Simulation Data (IU)
    - Life Science
    - High Energy Physics



**RESEARCH  
TECHNOLOGIES**

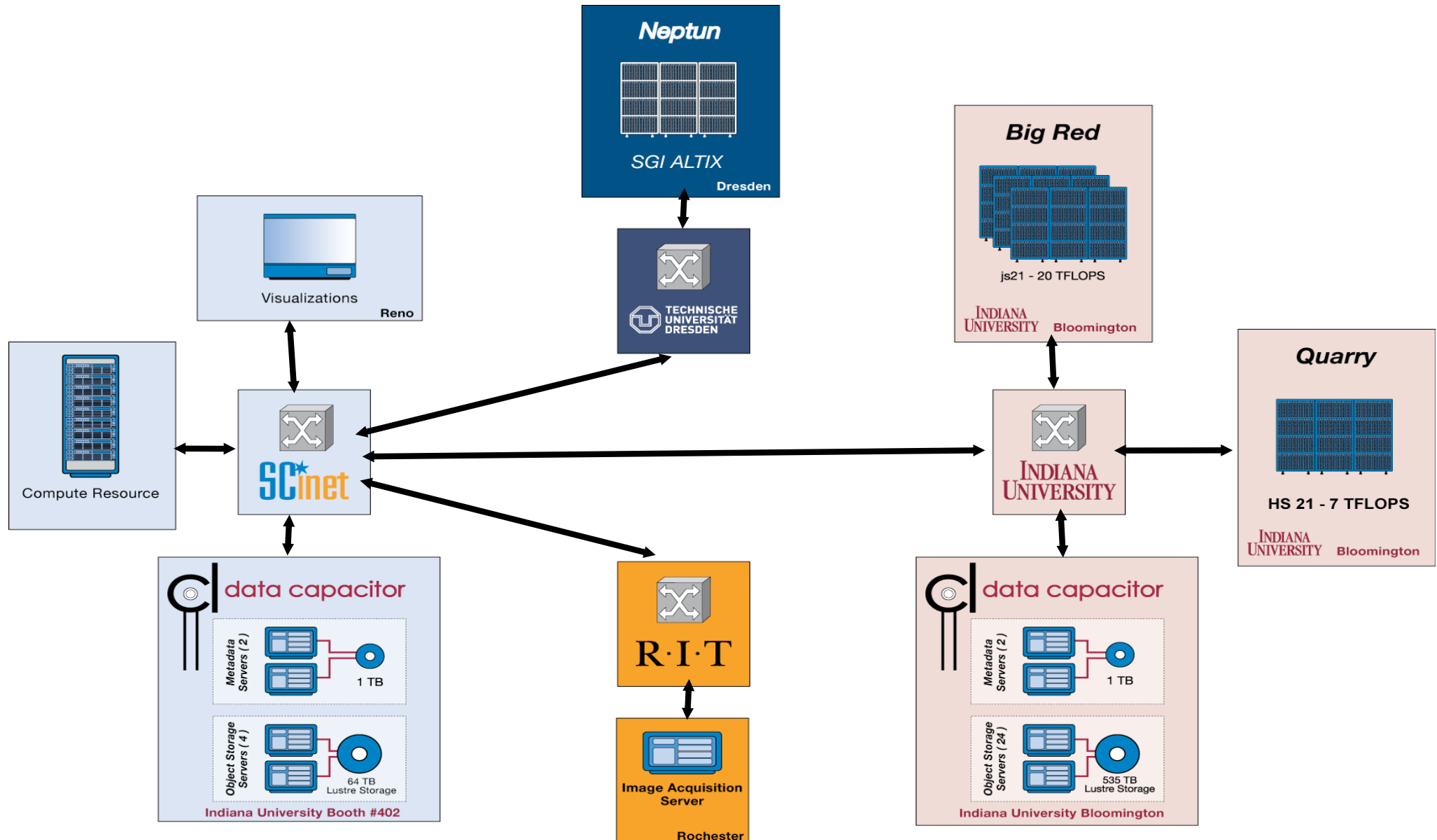
INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

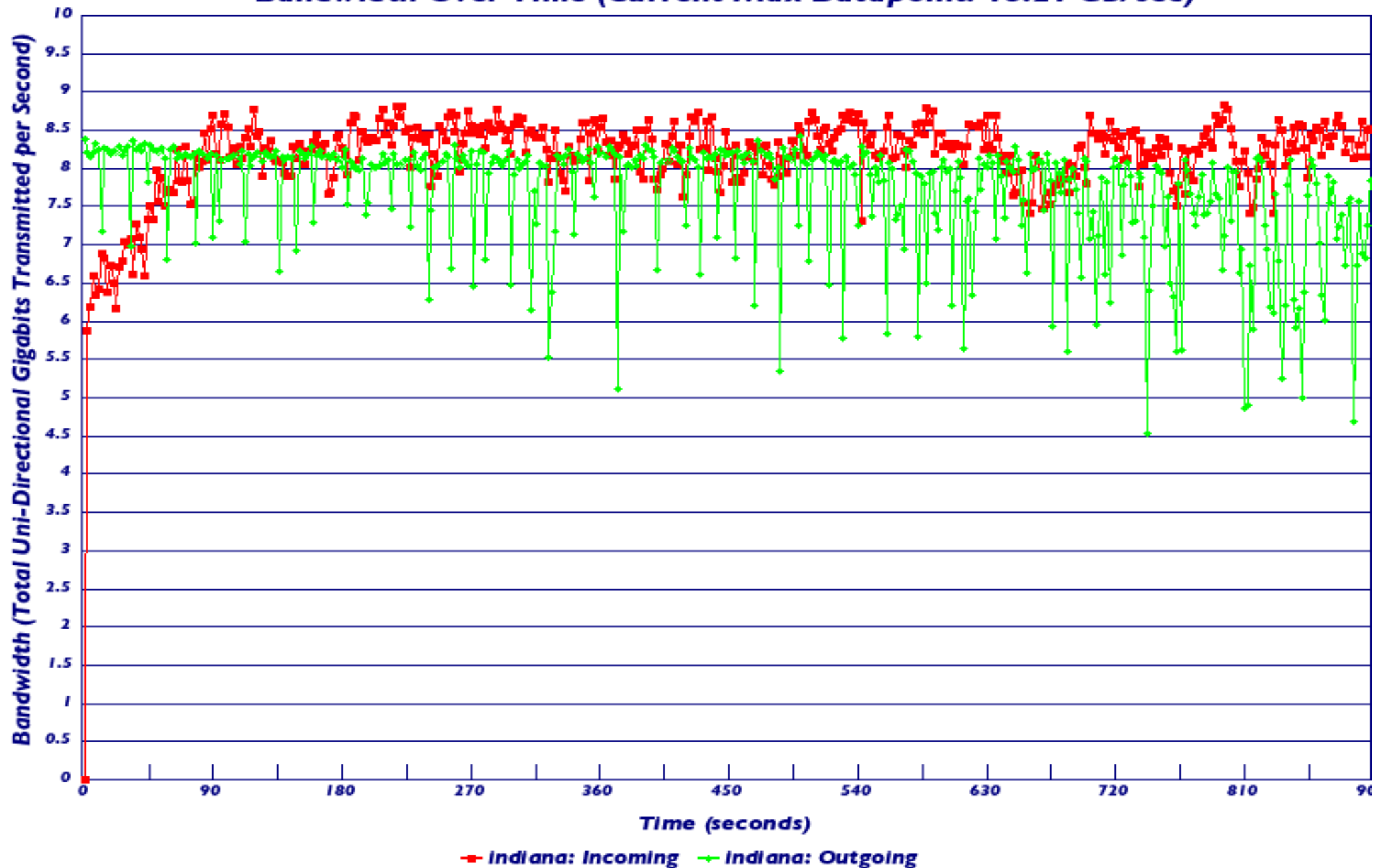
# 2007 Bandwidth Challenge Configuration





# Bandwidth Challenge Results

**Bandwidth Over Time (Current Max Datapoint: 18.21 Gb/sec)**



# 2008 - Data Capacitor WAN

## Goals:

- Provide centralized short-term storage
- Enable Workflows
- Harness distributed resources
- Push beyond a demo

360TB of DataDirect Networks S2A9550

4 OSS x 10 Gbit

Rudimentary UID Mapping for Lustre 1.6



RESEARCH  
TECHNOLOGIES

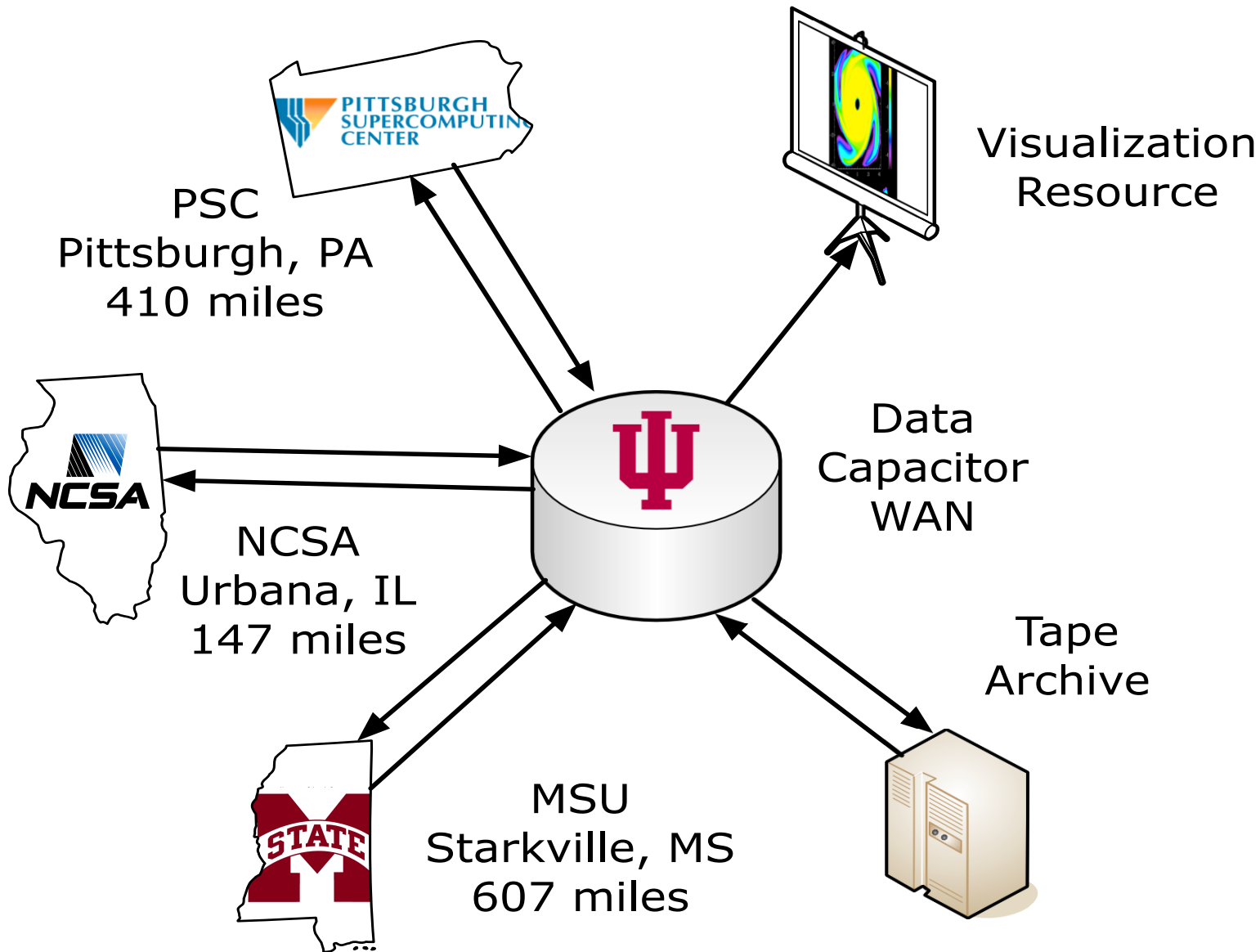
INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY

# Gas Giant Planet Research



# From 10Gb to 100 Gb Ethernet



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services

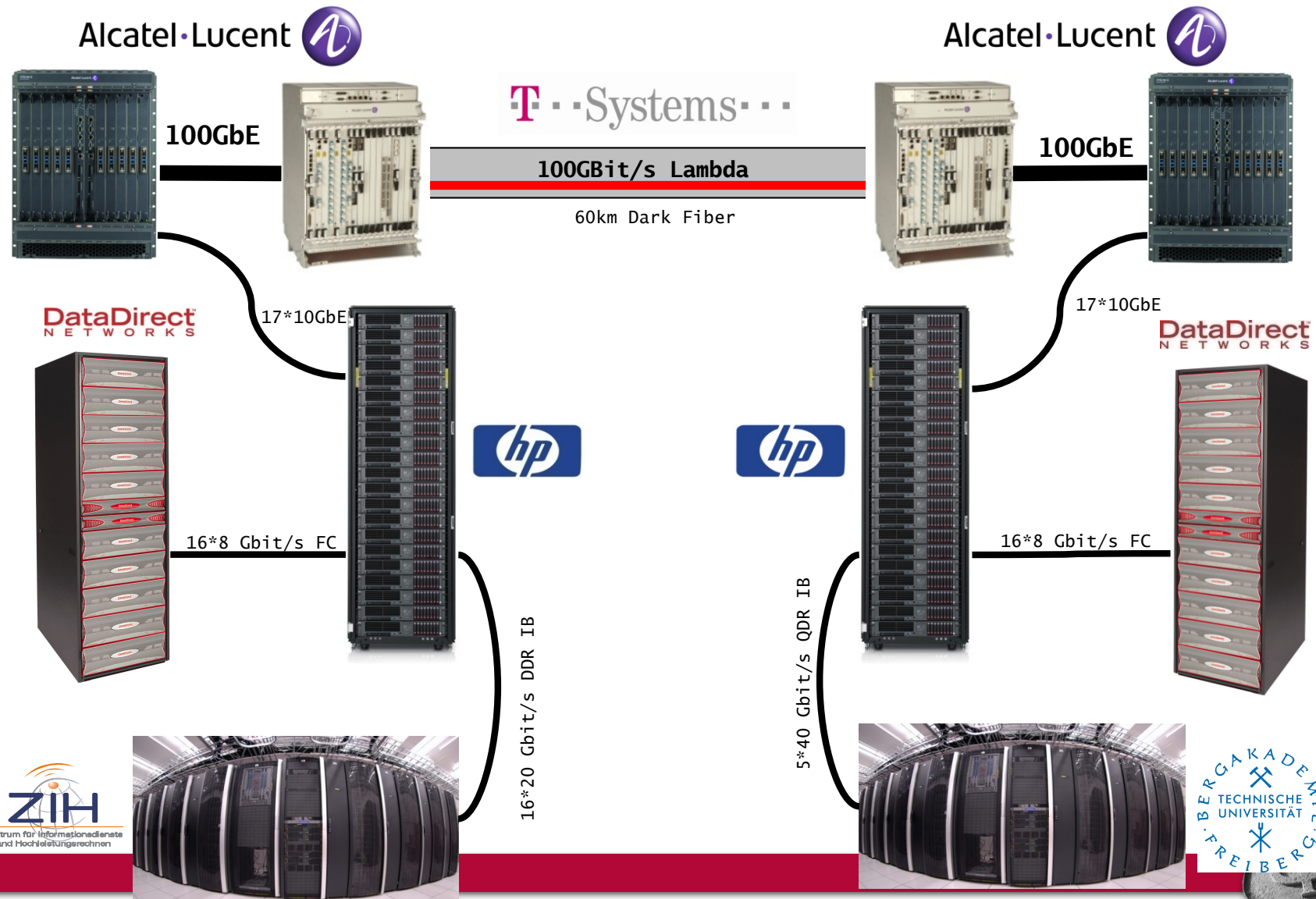


**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# 2010 Partnership with Dresden Germany



RESEARCH  
TECHNOLOGIES

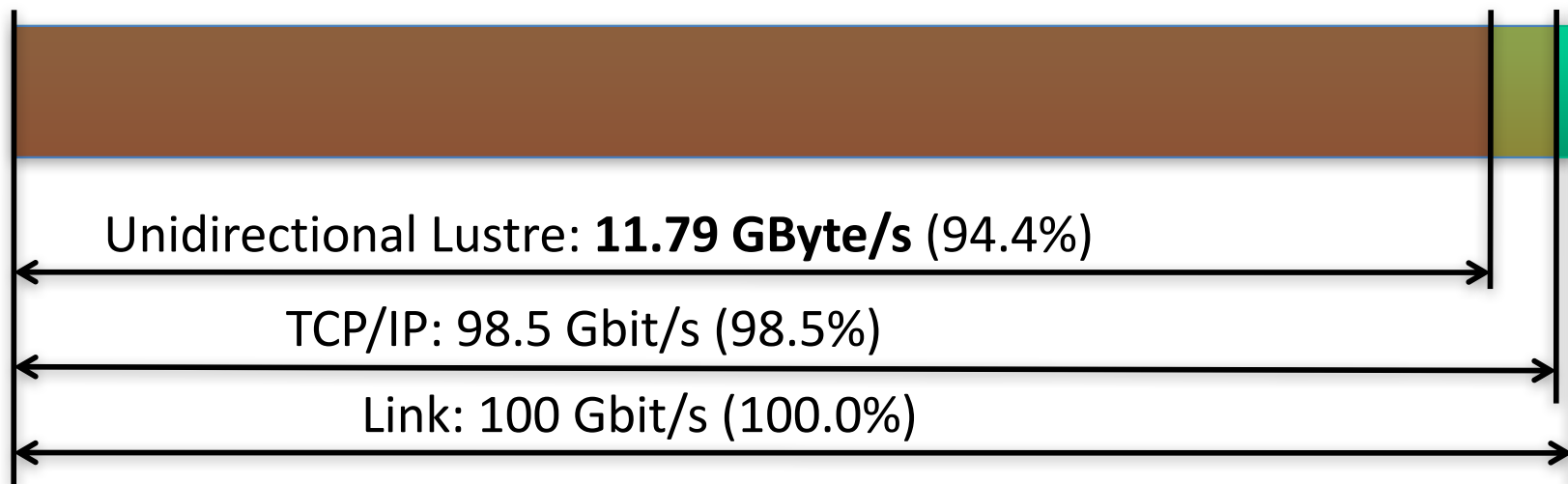
INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY

# Performance from Dresden to Freiberg (60km)



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY

# SC11 Demonstration

- Internet2 and ESnet, 50.5 ms RTT



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services

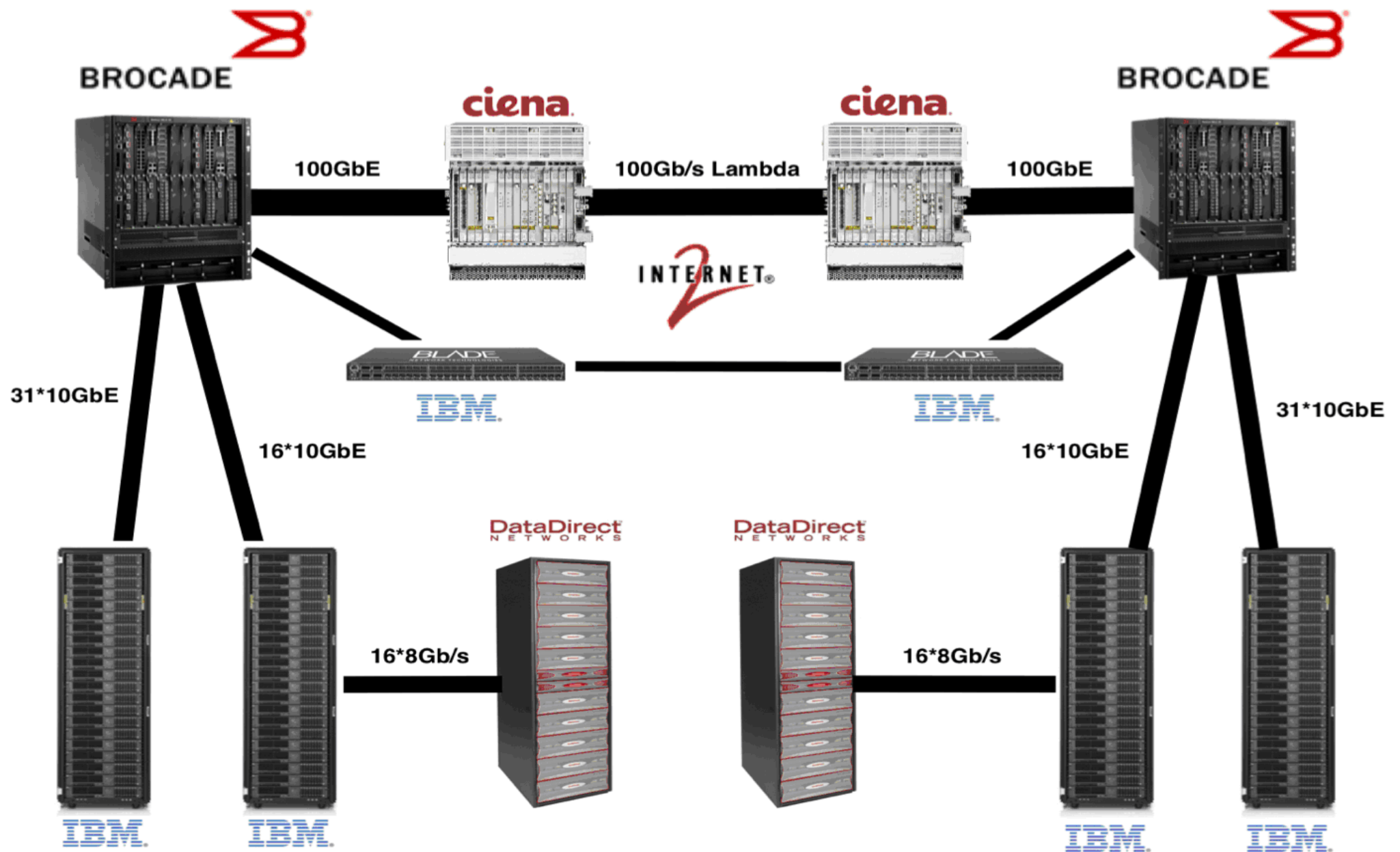


**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# SC11 Equipment – Seattle to Indianapolis





# SC 11 Performance – Limited by Time

	Measurement	Efficiency	Test Time
Link	50.5 ms		4 hours
TCP iperf	96 Gb/s	96%	6 hours
LNET	9.4 GB/s	77%	2 hours
IOR	6.5 GB/s	52%	2 hours
Applications	6.2 GB/s	50%	1 hour



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# 2012-2017 UID Mapping and Shared Key Code

Software Project Sponsored by OpenSFS  
Executed by IU with support from Intel

nodemap feature

UID Mapping – Spans Administrative Domains

Updates rudimentary MDS only UID mapping of DC-WAN

shared key feature

Allows secure connection between Lustre clients and servers

Both features are currently available in Lustre 2.10



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# In Use Today and Tomorrow



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# 2012 - Data Capacitor II

- 5.3 PB formatted
  - DataDirect Networks SFA12K Hardware
- 16 Object Storage Servers (96 GB RAM)
- 2 Metadata Servers (192 GB RAM)
- High Throughput
  - ~20 GB/s via Ethernet
  - >40 GB/s via InfiniBand
- 10Gb Ethernet systems: Karst, Carbonate Clusters
- FDR InfiniBand systems: Big Red II and II+ Cray Machines
- Currently running Lustre 2.5



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# 2016 - Data Capacitor WAN 2

Indiana University's first attempt using ZFS and Lustre

Backed by 1.1 PB DataDirect Networks SFA-10K

Running Lustre 2.8

4 OSS nodes

40 Gb Ethernet

Project Focus / No Scratch Space

Uses Zpools and Lustre pools to create containers



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# 2016 - Data Capacitor RAM

- A solid state Data Capacitor (all non-system drives are solid state)
- 35 TB of SSD on Lustre 2.8.0 backed by LDISKFS on HP Servers
- 6 OSS/12 OSTs
  - Each OST is a RAID-0 array of 4 x 800 GB Intel DC 3500 SSDs
- 2 active MDS supporting DNE2
  - Each MDT is RAID-0 array of 4 800 GB Intel DC 3510 SSDs
  - We are currently using DNE1 because of Lustre 2.5.1 Cray clients
- 40 Gbps FDR-10 IB to HPC systems



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# 2017-2018 - Slate / Slate Condo

2 New File Systems

Persistent Storage for Researchers

Slate - 4 PB

Slate Condo - 8 PB

Will Run ZFS

Will Run Lustre 2.10

Working with DataDirect Networks



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# Acknowledgments

Thanks to

- DDN for a gracious invitation and ongoing partnership
- Intel for their continued support
- Michael Kluge and Technische Universitat Dresden
- Robert Henschel, Director of Science Community Tools at IU
- The High Performance File System Team at IU



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

# Thank You for Your Time!

## Questions?



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY