A High-Performance Campus-Scale Cyberinfrastructure: The Technical, Political, and Economic

Presentation by Larry Smarr to the NSF Campus Bridging Workshop

October 11, 2010

Anaheim, CA

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

Dept. of Computer Science and Engineering

Jacobs School of Engineering, UCSD

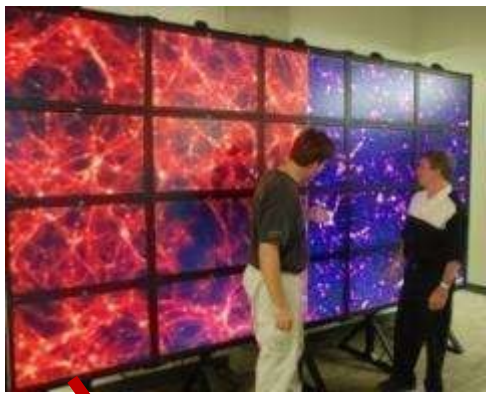
Follow me on Twitter: lsmarr

SDSC

it²

OptIPuter

Academic Research "OptIPlatform" Cyberinfrastructure: An End-to-End 10Gbps Lightpath Cloud



End User OptIPortal



Instruments



HD/4k Telepresence



HD/4k Video Cams



10G Lightpaths



HPC

TeraGrid

Campus Optical Switch



Data Repositories & Clusters



HD/4k Video Images

SDSC

it²

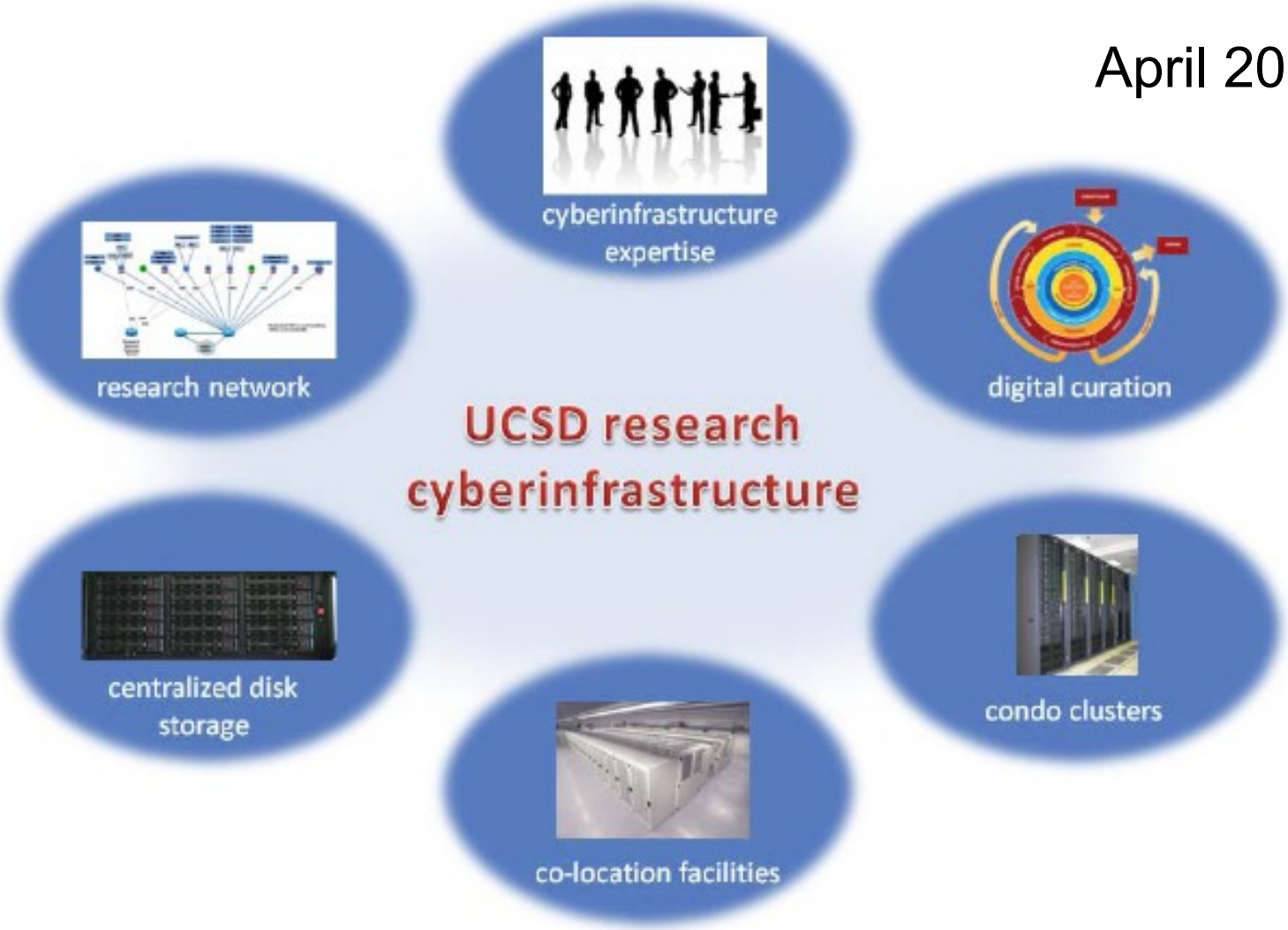
OptIPuter

"Blueprint for the Digital University"--Report of the UCSD Research Cyberinfrastructure Design Team

- **Focus on Data-Intensive Cyberinfrastructure**

April 2009

No Data Bottlenecks--Design for Gigabit/s Data Flows



research.ucsd.edu/documents/rcidt/RCIDTReportFinal2009.pdf

Broad Campus Input to Build the Plan and Support for the Plan

- Campus Survey of CI Needs-April 2008
  - 45 Responses (Individuals, Groups, Centers, Depts)
  - #1 Need was Data Management
    - 80% Data Backup
    - 70% Store Large Quantities of Data
    - 64% Long Term Data Preservation
    - 50% Ability to Move and Share Data
- Vice Chancellor of Research Took the Lead
- Case Studies Developed from Leading Researchers
- Broad Research CI Design Team
  - Chaired by Mike Norman and Phil Papadopoulos
  - Faculty and Staff:
    - Engineering, Oceans, Physics, Bio, Chem, Medicine, Theatre
    - SDSC, Calit2, Libraries, Campus Computing and Telecom

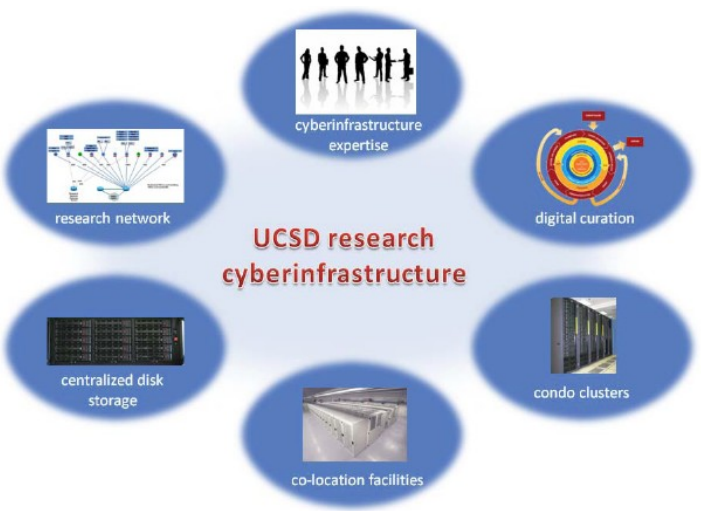
SDSC

it²

OptIPuter

Why Invest in Campus Research CI?

- Competitive Advantage
- Growing Campus Demand
- Leadership Opportunities
- Complementarity With National Programs
- Preservation of Digital Knowledge is Vital to the Scientific Method
- Institutional Obligations to Preserve Data:
  - OMB Circular A-110/CFR Part 215
  - Preserve Data for 3 Years After Grant
- Escalating Energy/Space Demands
- Integration with UC-Wide Initiatives



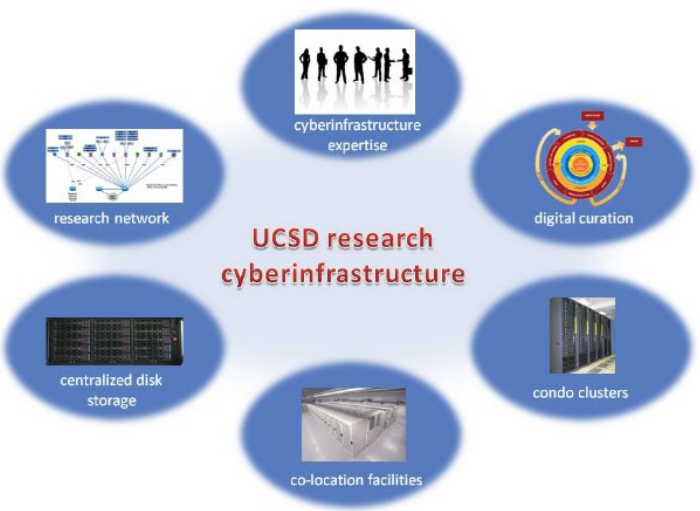
SDSC

it²

OptIPuter

Why Invest Now?

- Doing More With Less
- Exploit UCSD's Under-Developed Synergies
- SDSC Deployment of the Triton Resource
- The Longer We Wait
  - The Harder It Will Get
  - The More Opportunities Will Be Lost



SDSC it²

OptIPuter

Implementing the Campus Research CI Plan

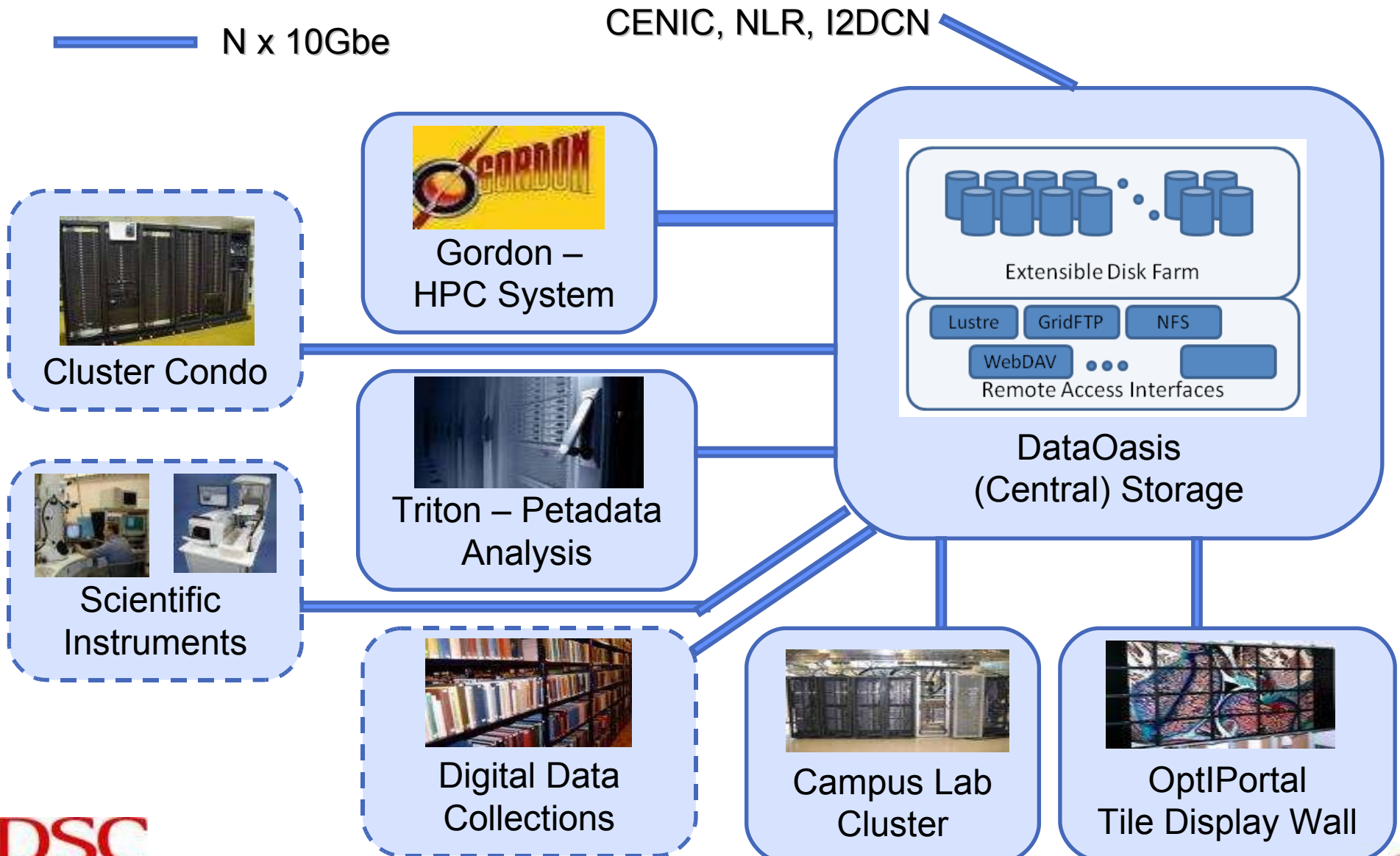
- Cyberinfrastructure Planning & Operations Committee
  - Chancellor Fox Appoints Fall 2009
  - Mission: Develop a Business Plan for the Self-Sustaining Operations of a Research Cyberinfrastructure
  - Report Delivered April 2010
- Business Plan Components
  - Direct Campus Investment
  - Energy Savings
  - PI Contributions
  - ICR
  - Separate Budgets for Startup and Sustaining
- Create an RCI Oversight Committee

SDSC

it²

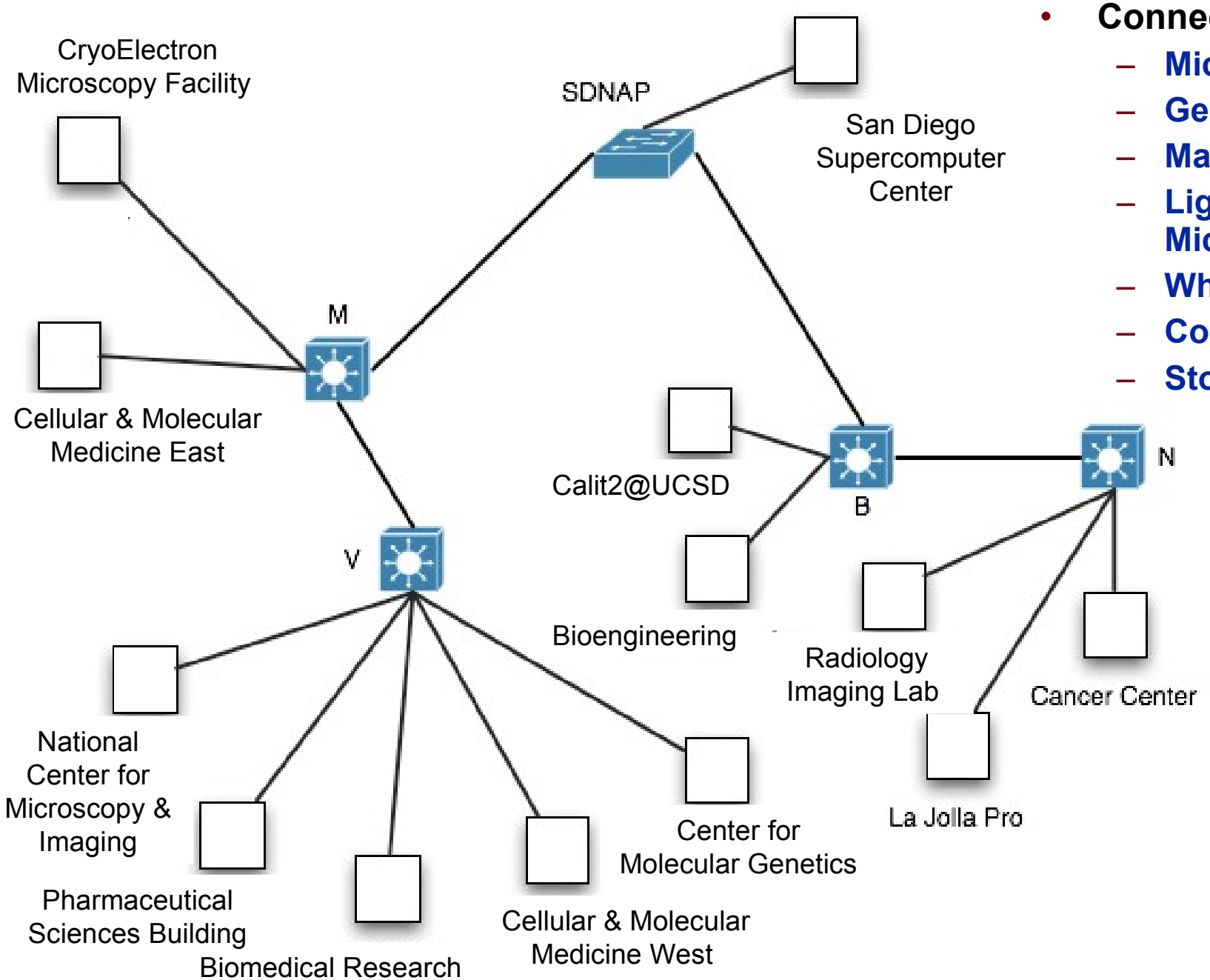
OptIPuter

UCSD Campus Investment in Fiber and Networks Enables High Performance Campus Bridging CI



Source: Philip Papadopoulos, SDSC, UCSD

UCSD Planned Optical Networked Biomedical Researchers and Instruments



- **Connects at 10 Gbps :**
  - **Microarrays**
  - **Genome Sequencers**
  - **Mass Spectrometry**
  - **Light and Electron Microscopes**
  - **Whole Body Imagers**
  - **Computing**
  - **Storage**

OptIPuter

Moving to a Shared Enterprise Data Storage and Analysis Resource:Triton Resource @ SDSC

Triton Resource

Large Memory PSDAF
- 256/512 GB/sys
- 9TB Total
- 128 GB/sec
- ~ 9 TF





Shared Resource Cluster
- 24 GB/Node
- 6TB Total
- 256 GB/sec
- ~ 20 TF



Large Scale Storage
- 2 PB
- 40 – 80 GB/sec
- 3000 – 6000 disks
- Phase 0: 1/3 TB, 8GB/s



UCSD Research Labs





Campus Research Network



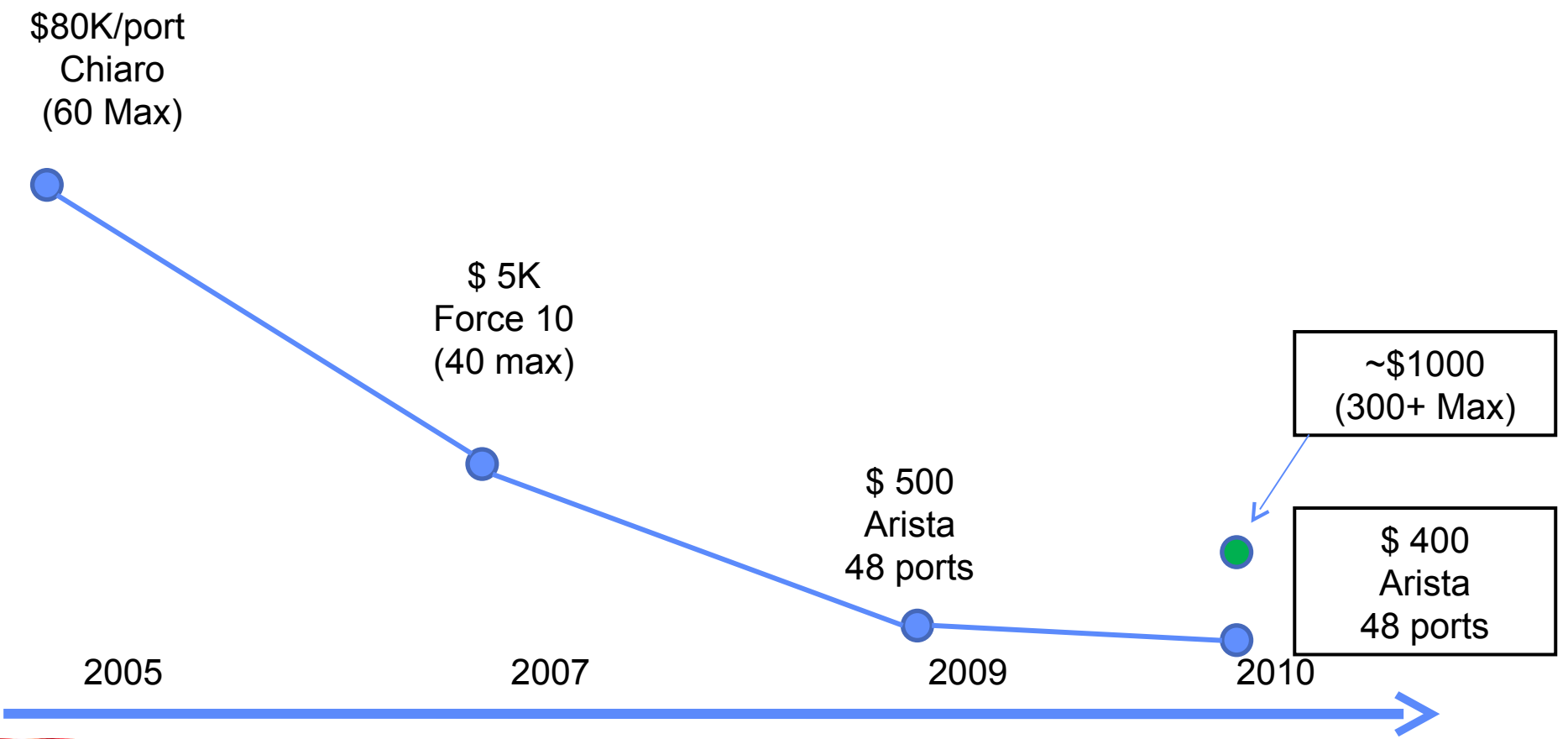
SDSC

SDSC

OptIPuter

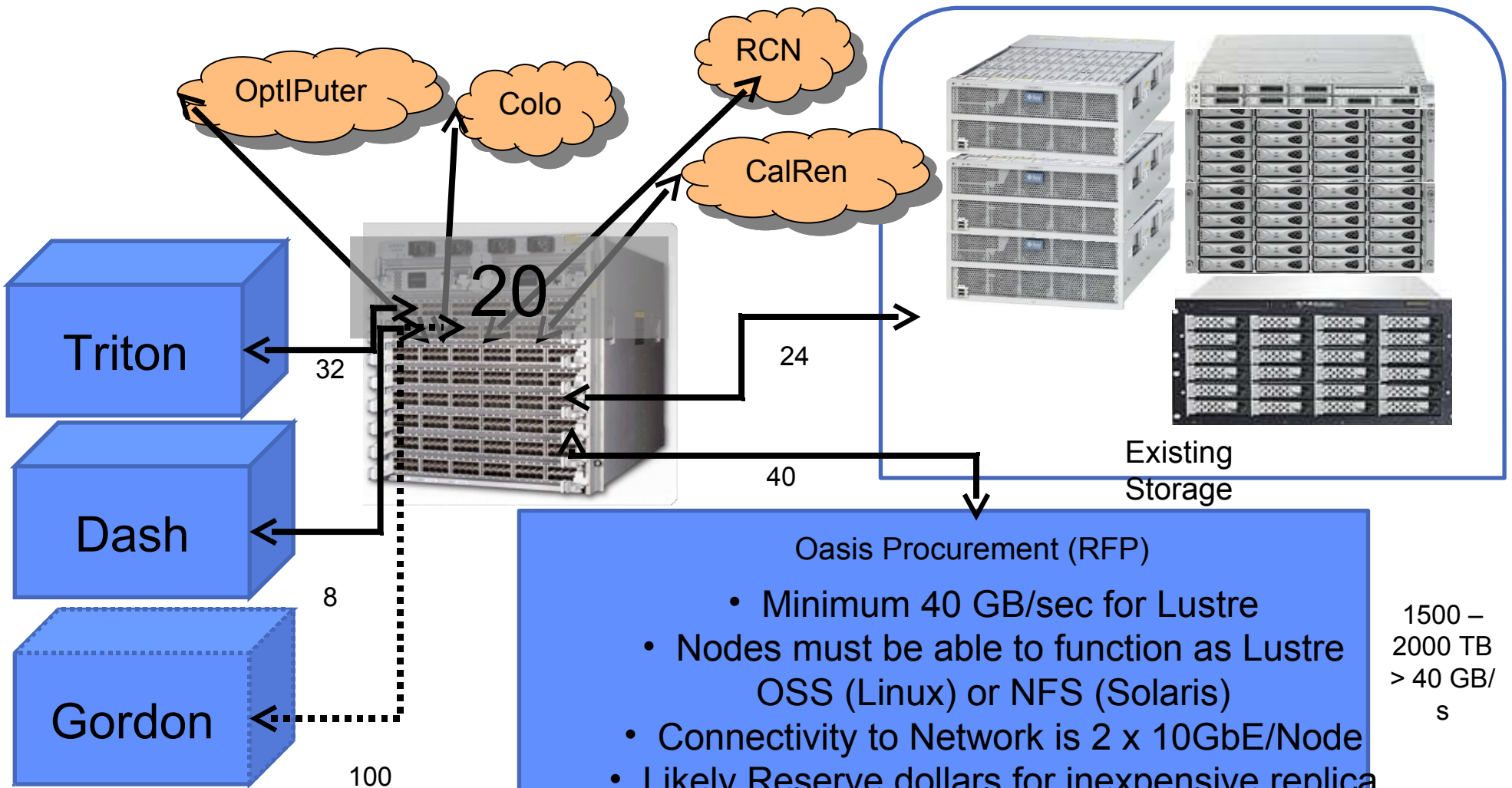
Rapid Evolution of 10GbE Port Prices Makes Campus-Scale 10Gbps CI Affordable

- Port Pricing is Falling
- Density is Rising – Dramatically
- Cost of 10GbE Approaching Cluster HPC Interconnects



Source: Philip Papadopoulos, SDSC, UCSD

10G Switched Data Analysis Resource: Data Oasis (RFP Underway)



SDSC

it²

OptIPuter

High Performance Computing (HPC) vs. High Performance Data (HPD)

Attribute	HPC	HPD
Key HW metric	Peak FLOPS	Peak IOPS
Architectural features	Many small-memory multicore nodes	Fewer large-memory vSMP nodes
Typical application	Numerical simulation	Database query Data mining
Concurrency	High concurrency	Low concurrency or serial
Data structures	Data easily partitioned e.g. grid	Data not easily partitioned e.g. graph
Typical disk I/O patterns	Large block sequential	Small block random
Typical usage mode	Batch process	Interactive

Source: SDSC

GRAND CHALLENGES IN DATA-INTENSIVE SCIENCES

OCTOBER 26-28, 2010

SAN DIEGO SUPERCOMPUTER CENTER , UC SAN DIEGO

Confirmed conference *topics* and speakers :

- *Needs and Opportunities in Observational Astronomy* - Alex Szalay, JHU
- *Transient Sky Surveys* – Peter Nugent, LBNL
- *Large Data-Intensive Graph Problems* – John Gilbert, UCSB
- *Algorithms for Massive Data Sets* – Michael Mahoney, Stanford U.
- *Needs and Opportunities in Seismic Modeling and Earthquake Preparedness* - Tom Jordan, USC
- *Needs and Opportunities in Fluid Dynamics Modeling and Flow Field Data Analysis* – Parviz Moin, Stanford U.
- *Needs and Emerging Opportunities in Neuroscience* – Mark Ellisman, UCSD
- *Data-Driven Science in the Globally Networked World* – Larry Smarr, UCSD

SDSC

it²

OptIPuter