

High-Performance Computing

Ioan Raicu

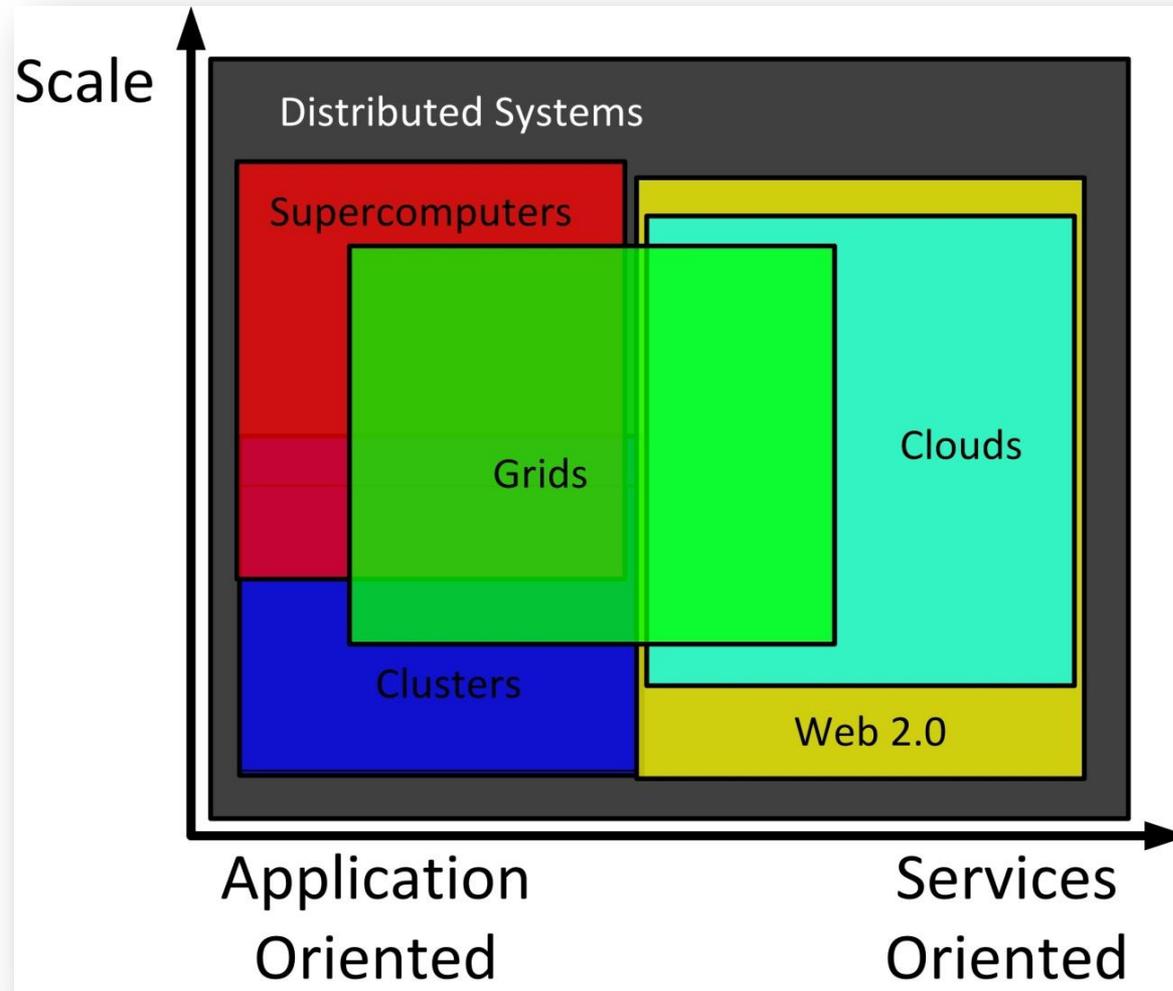
Computer Science Department
Illinois Institute of Technology

CS 595

Hot Topics in Distributed Systems: Data-Intensive Computing

September 1st, 2010

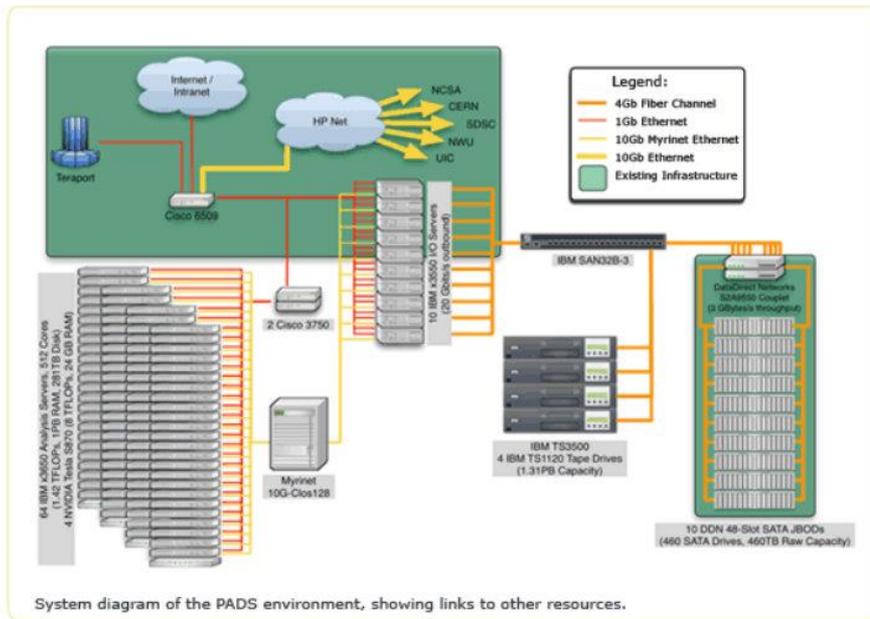
Distributed Systems



Cluster Computing: PADS

PADS

Computer clusters using commodity processors, network interconnects, and operating systems.



PADS is a petabyte (10^{15} -byte)-scale online storage server capable of sustained multi-gigabyte/s I/O performance, tightly integrated with a 9 teraflop/s computing resource and multi-gigabit/s local and wide area networks. Its hardware and associated software enables the reliable storage of, access to, and analysis of massive datasets by both local users and the national scientific community.

The PADS design results from a study of the storage and analysis requirements of participating groups in astrophysics and astronomy, computer science, economics, evolutionary and organismal biology, geosciences, high-energy physics, linguistics, materials science, neuroscience, psychology, and sociology. For these groups, PADS represents a significant opportunity to look at their data in new ways, enabling new scientific insights. The infrastructure also encourages new collaborations across disciplines. PADS is also a vehicle for computer science research into active data store systems, and provides rich data on which to investigate new techniques. Results will be made available as open source software.

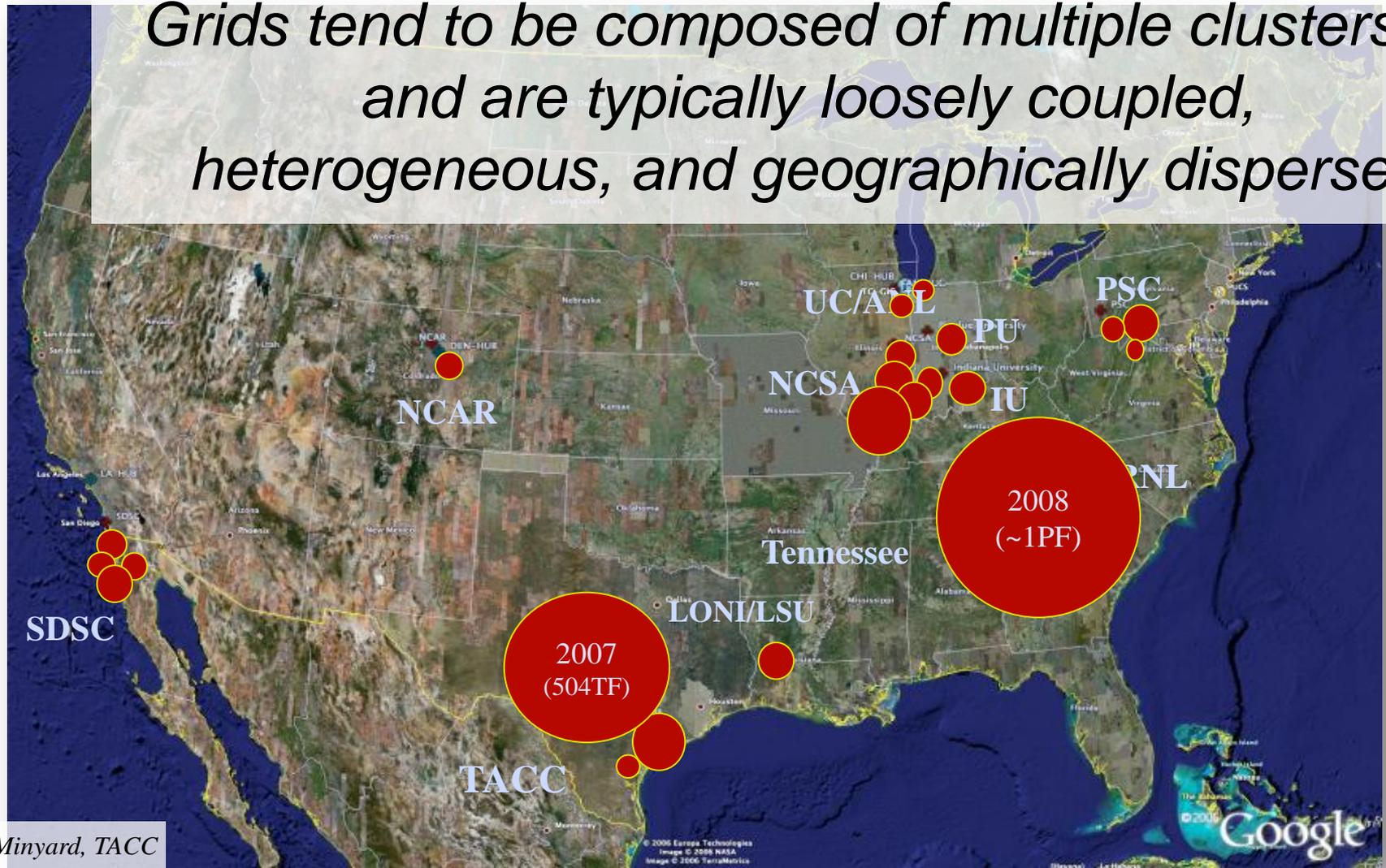
The PADS project is supported in part by the National Science Foundation under grant OCI-0821678 and by The University of Chicago.

[PADSstatus](#)

[myPADS](#)

Grid Computing: TeraGrid

Grids tend to be composed of multiple clusters, and are typically loosely coupled, heterogeneous, and geographically dispersed



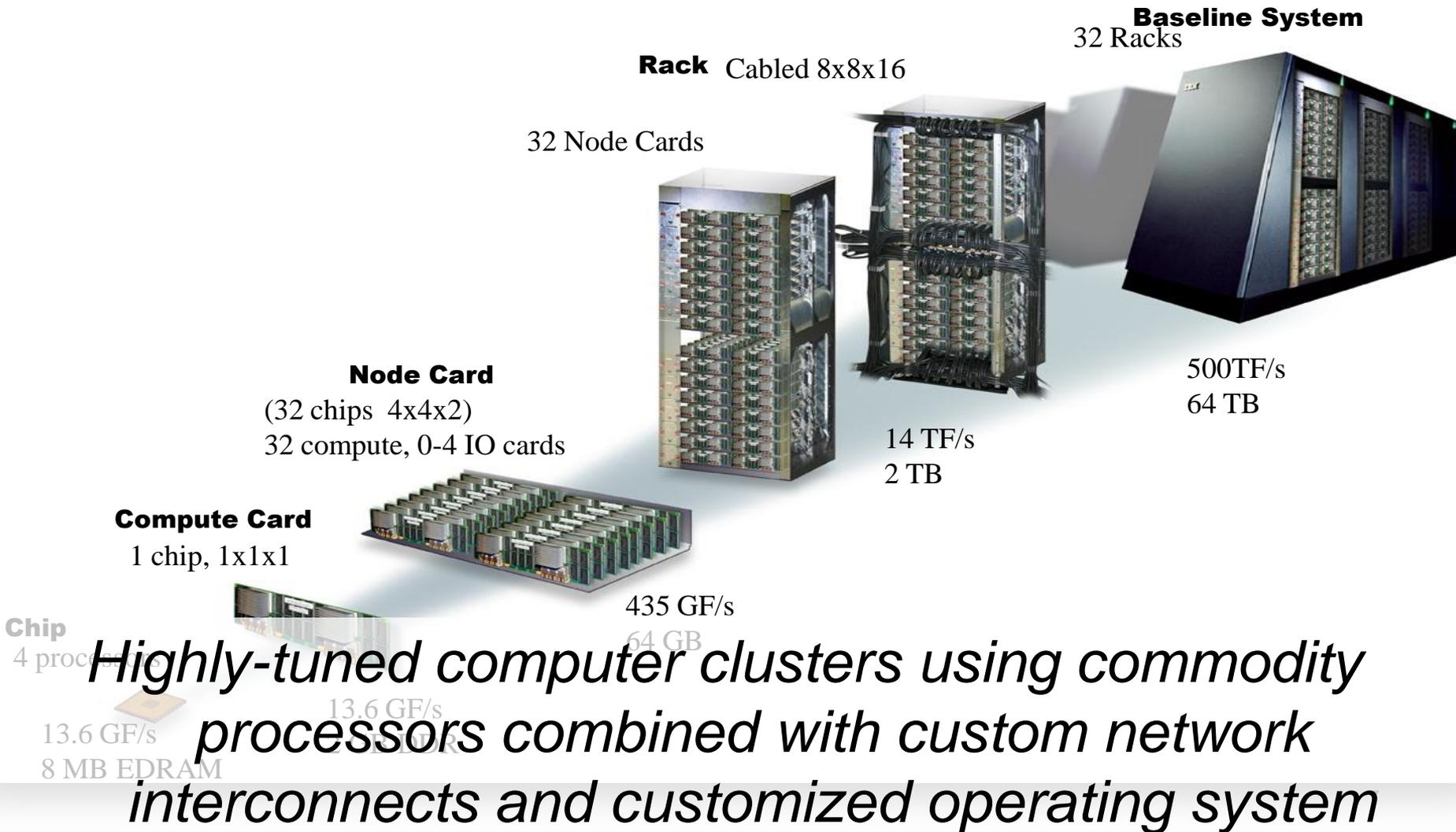
What is the TeraGrid?

- An instrument (cyberinfrastructure) that delivers high-end IT resources - storage, computation, visualization, and data/service hosting - almost all of which are UNIX-based under the covers; some hidden by Web interfaces
 - 20 Petabytes of storage (disk and tape)
 - over 100 scientific data collections
 - 750 TFLOPS (161K-cores) in parallel computing systems and growing
 - Support for Science Gateways
- The largest individual cyberinfrastructure facility funded by the NSF, which supports the national science and engineering research community
- Something you can use without financial cost - allocated via peer review (and without double jeopardy)

Major Grids

- TeraGrid (TG)
- Open Science Grid (OSG)
- Enabling Grids for E-scienceE (EGEE)
- LHC Computing Grid from CERN
- Grid Middleware
 - Globus Toolkit
 - Unicore

Supercomputing: IBM Blue Gene/P



Top 10 Supercomputers from Top500

- Cray XT4 & XT5
 - Jaguar #1
 - Kraken #3
- IBM BladeCenter Hybrid
 - Roadrunner #2
- IBM BlueGene/L & BlueGene/P
 - Jugene #4
 - Intrepid #8
 - BG/L #7
- NUDT (GPU based)
 - Tianhe-1 #5
- SGI Altix ICE
 - Plaiedas #6
- Sun Constellation
 - Ranger #9
 - Red Sky #10

HPC: High-Performance Computing

- Synonymous with supercomputing
- Tightly-coupled applications
- Implemented using Message Passing Interface (MPI)
- Large of amounts of computing for short periods of time
- Usually requires low latency interconnects
- Measured in FLOPS

Questions

