Grid Computing

Dr Simon See
High Performance Computing
Sun Microsystems Inc

The Grid

BLUEPRINT FOR A NEW COMPUTING INFRASTRUCTURE

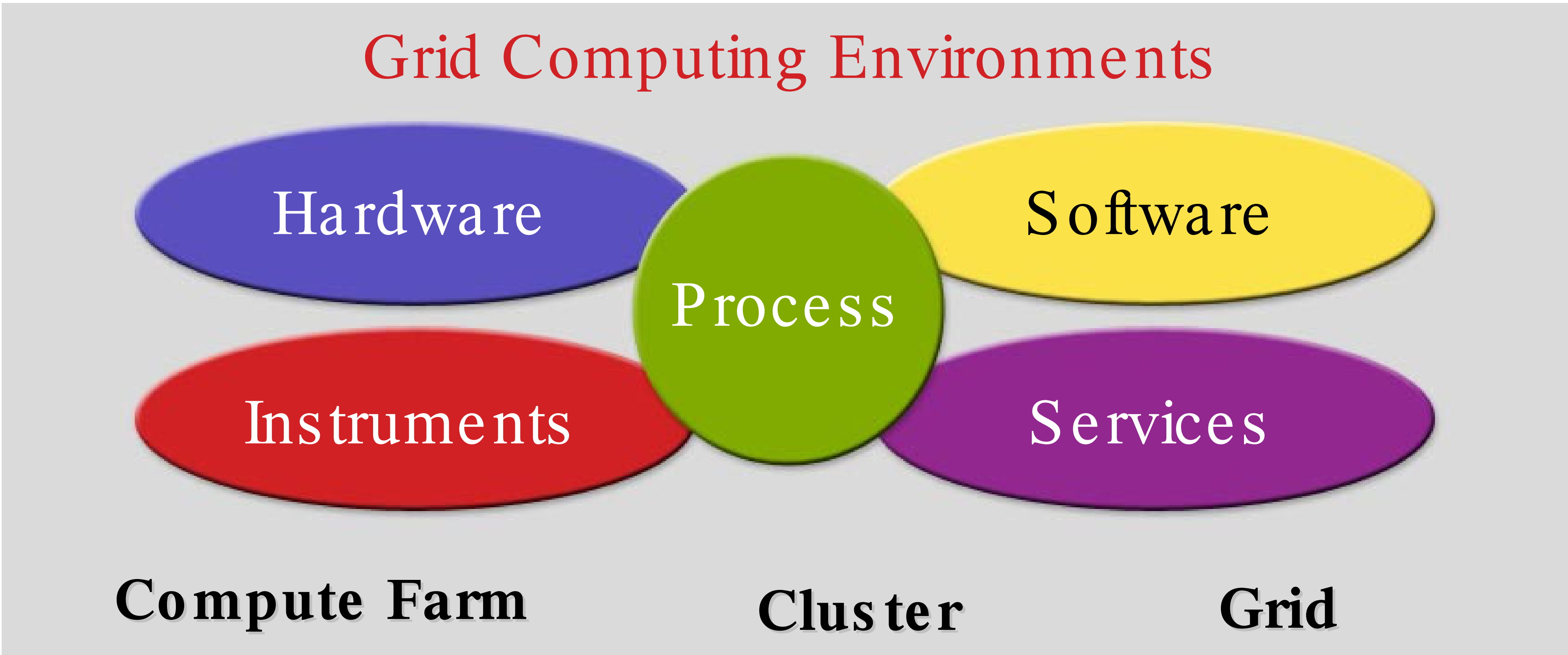
"The grid is an emerging *infrastructure* that will fundamentally change the way we think about – and use – computing. The grid will... create a universal source of computing power."

Grid Computing Environments:

ARCHITECTURE, IMPLEMENTATION AND MANAGEMENT

- Historical Perspective
- Industry Characteristics
- Requirements and Policies
- Architecture
- Implementation
- Usability and Management
- Extensibility and Web Services

Definitions



A Grid is a hardware/software/instruments/service infrastructure that provides **dependable, consistent, pervasive, inexpensive** * access to computational and data capabilities.

* Foster, Kesselman, 1999

Definitions

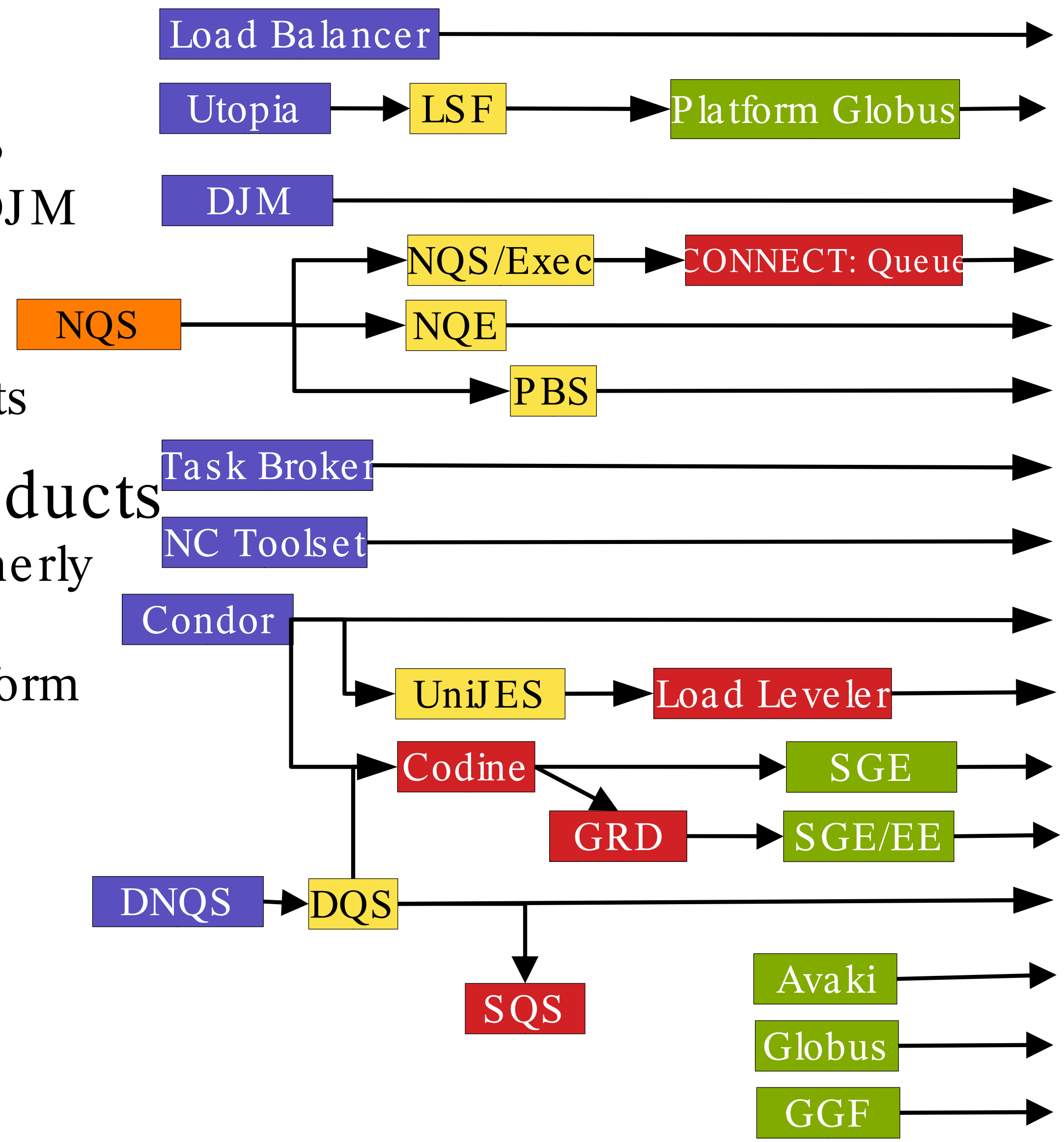
Grid Computing Environments

- Distributed Resource Management
  - SGE
  - LSF
  - NQS
  - DJM

- Parallel Computing Environments
  - PVM
  - MPL
  - MPI
  - FORTRAN 90
  - Linda

Grid Computing Genealogy

- Early Grid Technologies
  - Distributed Job Manager; DJM
  - Network Queuing System: NQS
  - University Research projects
- Mature Commercial Products
  - Sun Grid Engine (Sun, formerly Codine/GRD)
  - Load Sharing Facility (Platform Computing)
  - Load Leveler
- Industry Collaboration / Leadership
  - Globus
  - Global Grid Forum
  - Platform Globus



Industry Characteristics

Life Sciences

- Startup and cost efficient
- Custom research or limited use applications
- Multi-day application runs (BLAST)
- Exponential Combinations
- Limited administrative staff
- Complementary techniques

Electronic Design

- Time to Market
- Fastest platforms, largest Grids
- License Management
- Well established application suite
- Large legacy investment
- Platform Ownership issues

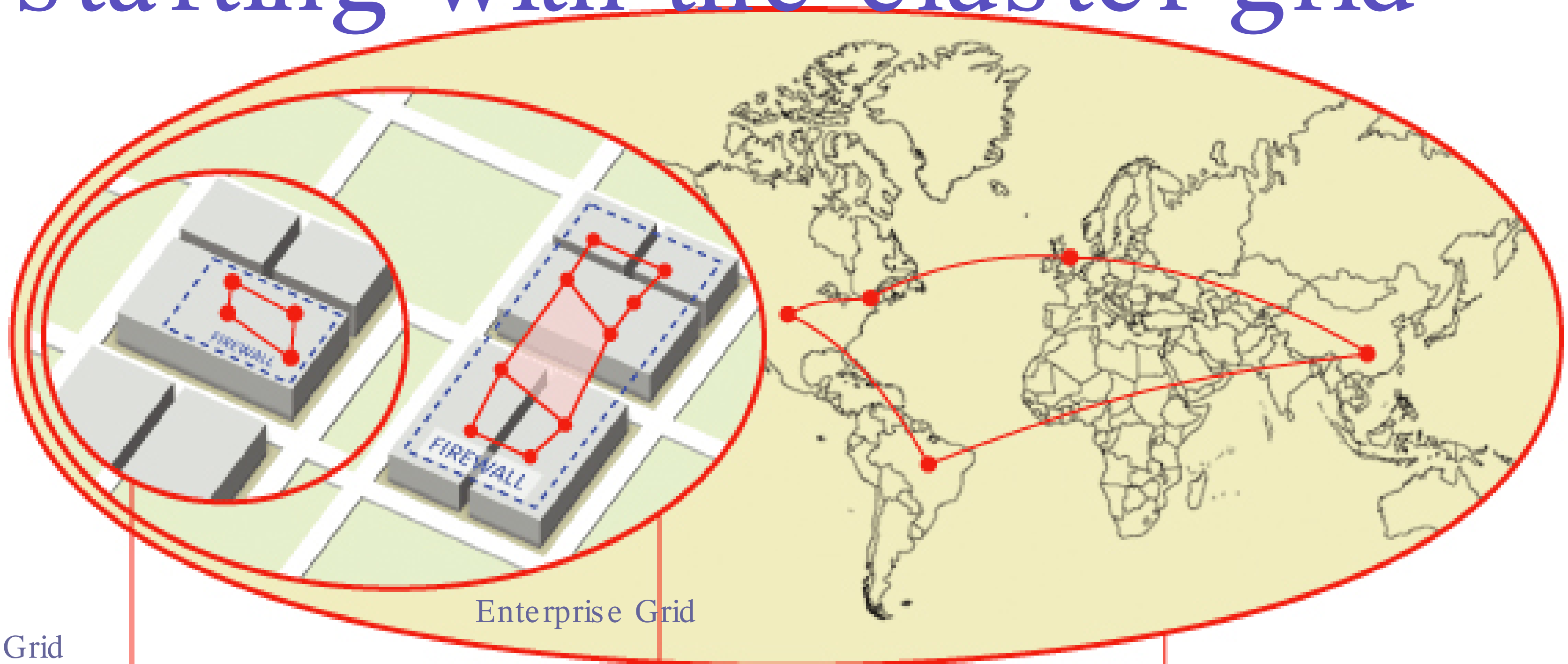
Financial Services

- Market simulations
- Time IS Money
- Proprietary applications
- Multiple Platforms
- Multiple scenario execution
- Need instant results and analysis tools

High Performance Computing

- Parallel Reservoir Simulations
- Geophysical Ray Tracing
- Custom in-house codes
- Well established application suite

Grid Evolutionary Strategy: starting with the cluster grid



Cluster Grid

Departmental Computing

- Simplest Grid deployment
- Maximum utilization of departmental resources
- Resources allocated based on priorities

Enterprise Grid

Enterprise Computing

- Resources shared within the enterprise
- Policies ensure computing on demand
- Gives multiple groups seamless access to enterprise resources

Global Grid

Internet Computing

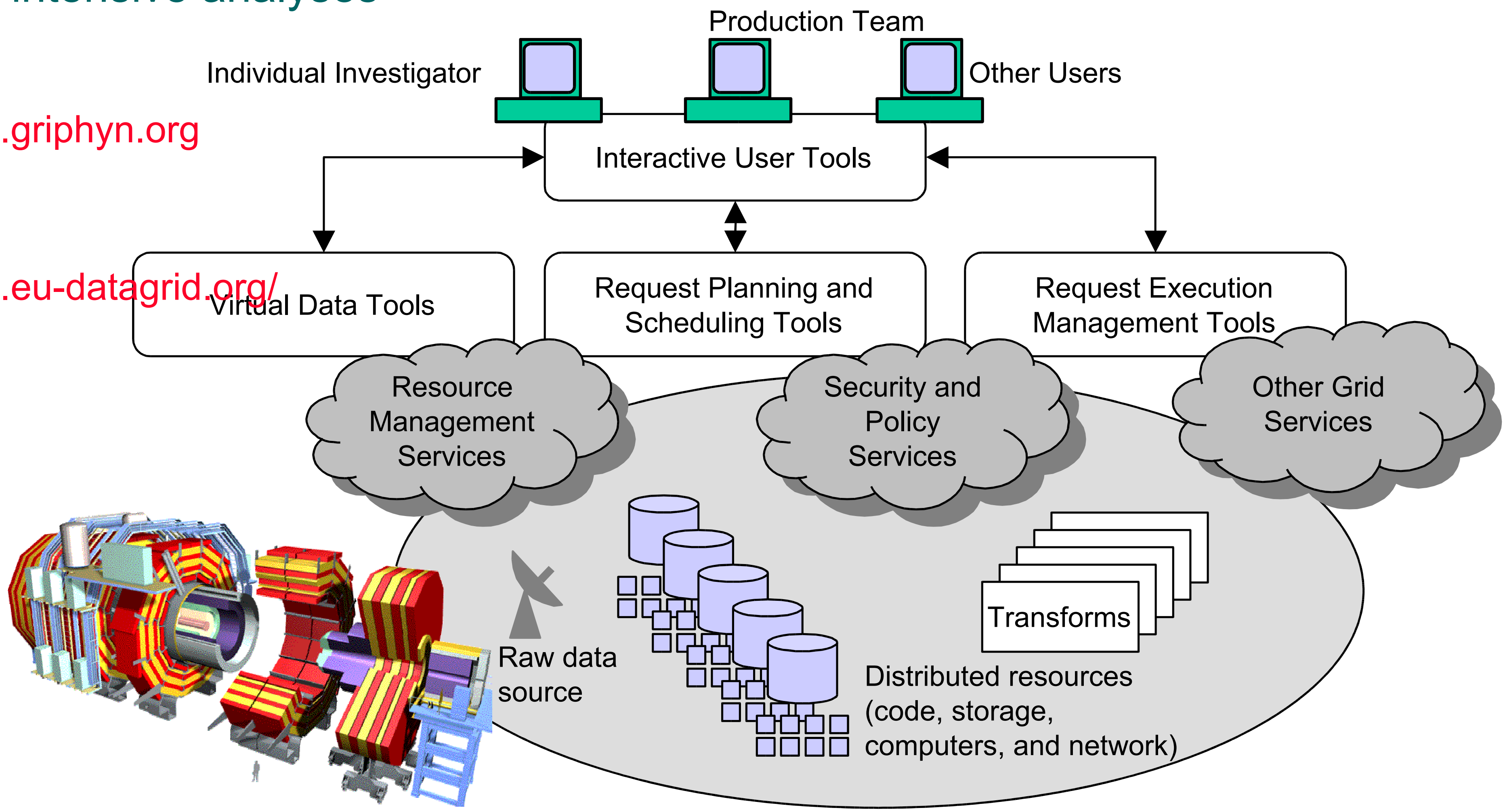
- Resources shared over the Internet
- Global view of distributed datasets

Example Grid Communities

- **GriPhyN and European DataGrid**
  - Enable (thousands of) physicists to pool data & computing resources for data-intensive analyses

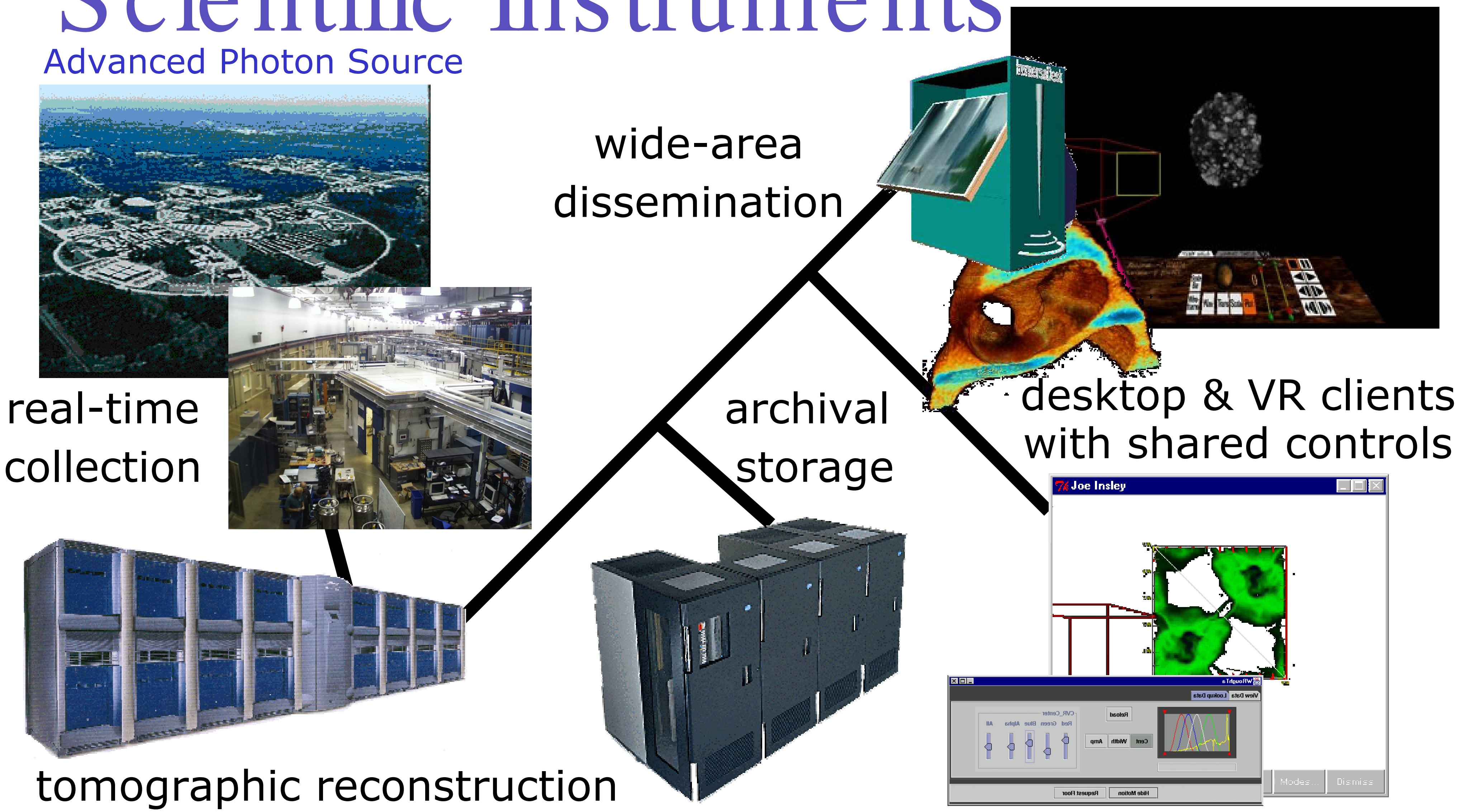
http://www.griphyn.org

http://www.eu-datagrid.org/



Online Access to Scientific Instruments

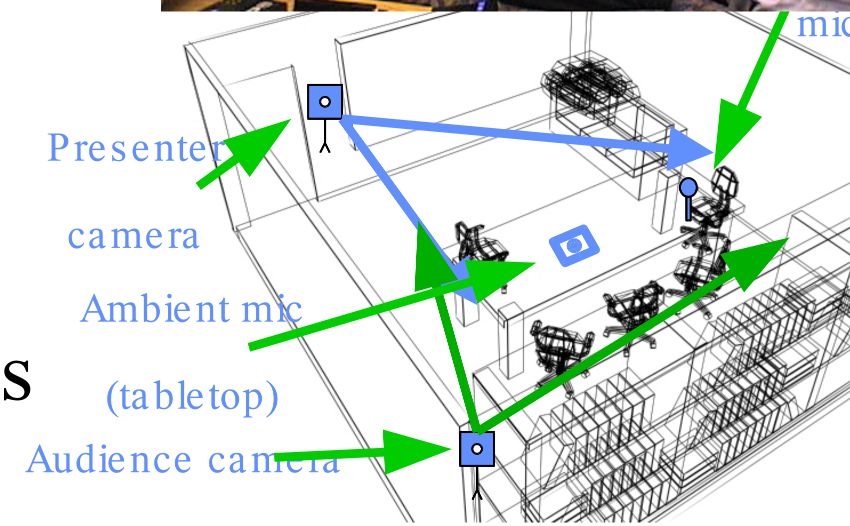
Advanced Photon Source



DOE X-ray grand challenge: ANL, USC/ISI, NIST, U.Chicago

Example Grid Communities

- Access Grid Collaboration
  - Enable collaborative work at dozens of sites worldwide, with strong sense of shared presence
  - Combination of commodity audio/video tech + Grid technologies for security, discovery, etc.
  - 40+ sites worldwide, number rising rapidly



http://www.accessgrid.org

The Grid Problem

- Flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resource

From "The Anatomy of the Grid: Enabling Scalable Virtual Organizations"

- Enable communities ("virtual organizations") to share geographically distributed resources as they pursue common goals -- *assuming the absence of…*
  - central location,
  - central control,
  - omniscience,
  - existing trust relationships.

Elements of the Problem

- Resource sharing
  - Computers, storage, sensors, networks, …
  - Sharing always conditional: issues of trust, policy, negotiation, payment, …
- Coordinated problem solving
  - Beyond client-server: distributed data analysis, computation, collaboration, …
- Dynamic, multi-institutional virtual orgs
  - Community overlays on classic org structures
  - Large or small, static or dynamic

Data Grids for High Energy Physics

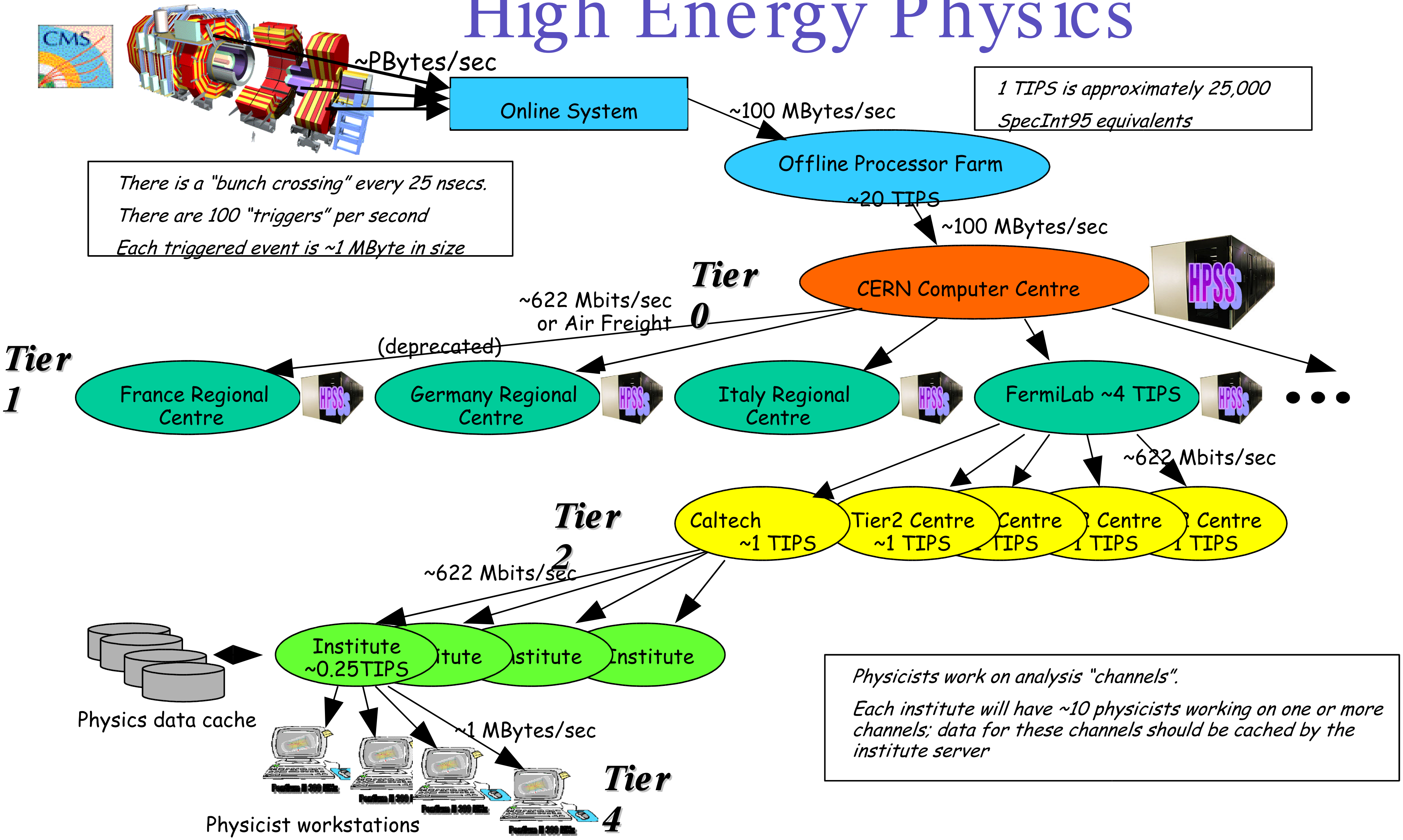
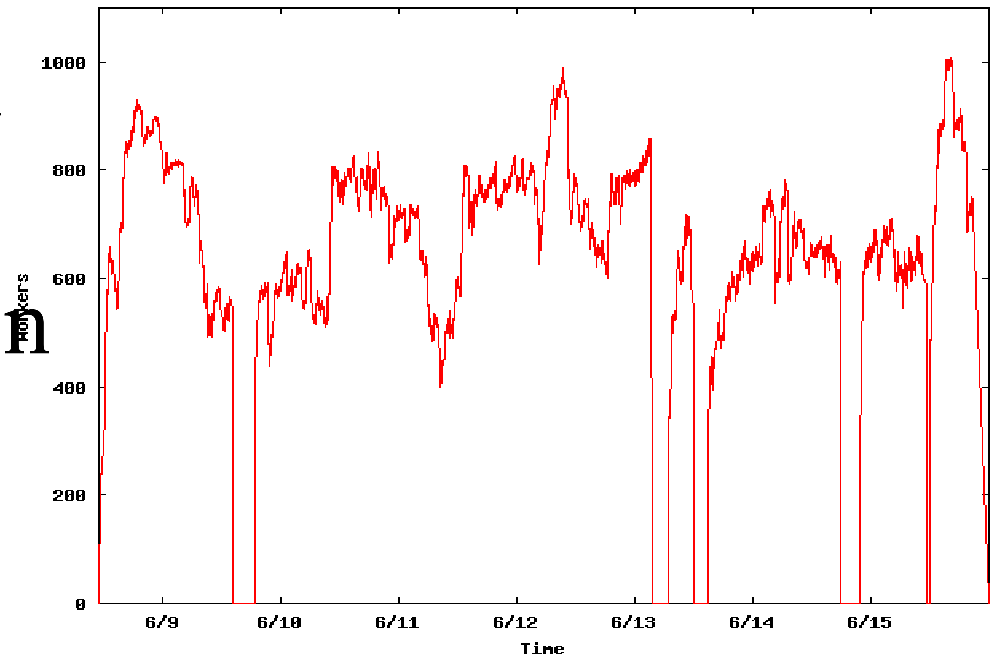


Image courtesy Harvey Newman, Caltech

Mathematicians Solve NUG30

- Looking for the solution to the NUG30 quadratic assignment problem
- An informal collaboration of mathematicians and computer scientists
- Condor-G delivered 3.46E8 CPU seconds in 7 days (peak 1009 processors) in U.S. and Italy (8 sites)

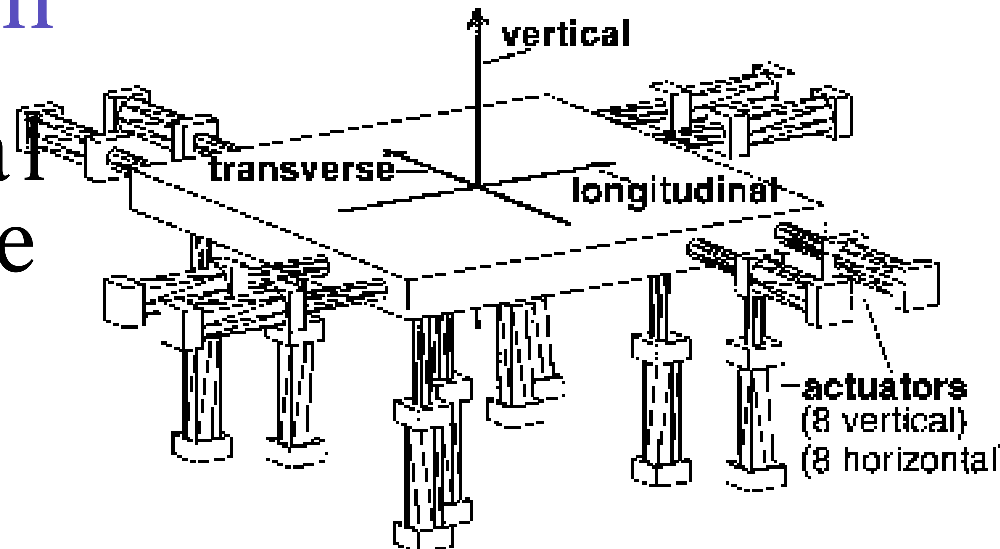


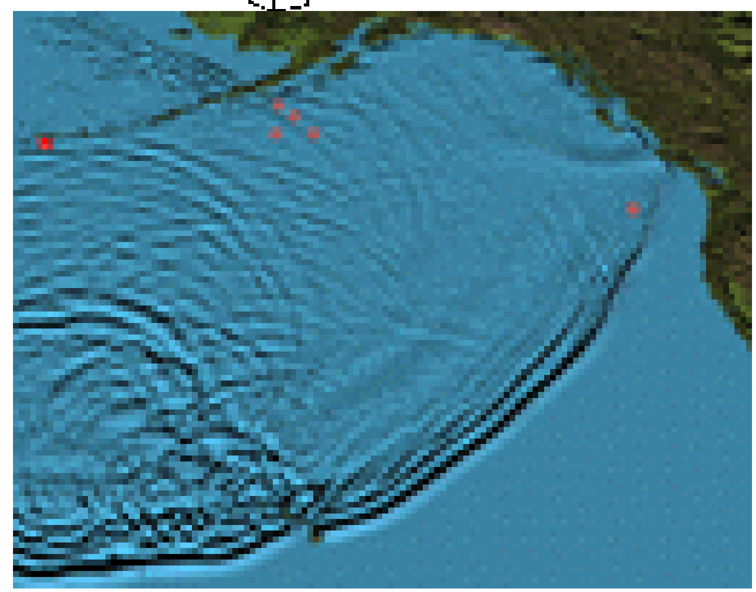
14,5,28,24,1,3,16,15,
10,9,21,2,4,29,25,22,
13,26,17,30,6,20,19,
8,18,7,27,12,11,23

MetaNEOS: Argonne, Iowa, Northwestern, Wisconsin

Network for Earthquake Engineering Simulation

- NEESgrid: US national infrastructure to couple earthquake engineers with experimental facilities, databases, computers, & each other
- On-demand access to experiments, data streams, computing, archives, collaboration





NEESgrid: Argonne, Michigan, NCSA, UIUC, USC

Home Computers Evaluate AIDS Drugs

- Community =
  - 1000s of home computer users
  - Philanthropic computing vendor (Entropia)
  - Research group (Scripps)
- Common goal= advance AIDS research

Broader Context

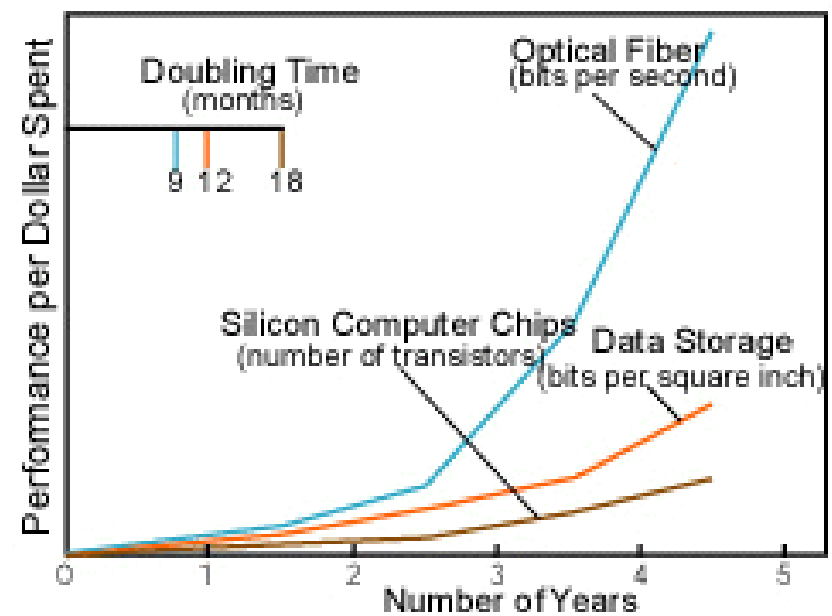
- "Grid Computing" has much in common with major industrial thrusts
  - Business-to-business, Peer-to-peer, Application Service Providers, Storage Service Providers, Distributed Computing, Internet Computing…
- Sharing issues not adequately addressed by existing technologies
  - Complicated requirements: "run program X at site Y subject to community policy P, providing access to data at Z according to policy Q"
  - High performance: unique demands of advanced & high-performance systems

Why Now?

- Moore's law improvements in computing produce highly functional endsystems
- The Internet and burgeoning wired and wireless provide universal connectivity
- Changing modes of working and problem solving emphasize teamwork, computation
- Network exponentials produce dramatic changes in geometry and geography

Network Exponentials

- Network vs. computer performance
  - Computer speed doubles every 18 months
  - Network speed doubles every 9 months
  - Difference = order of magnitude per 5 years
- 1986 to 2000
  - Computers: x 500
  - Networks: x 340,000
- 2001 to 2010
  - Computers: x 60
  - Networks: x 4000



Moore's Law vs. storage improvements vs. optical improvements. Graph from **Scientific American** (Jan-2001) by Cleo Vilett, source Vined Khoslan, Kleiner, Caufield and Perkins.

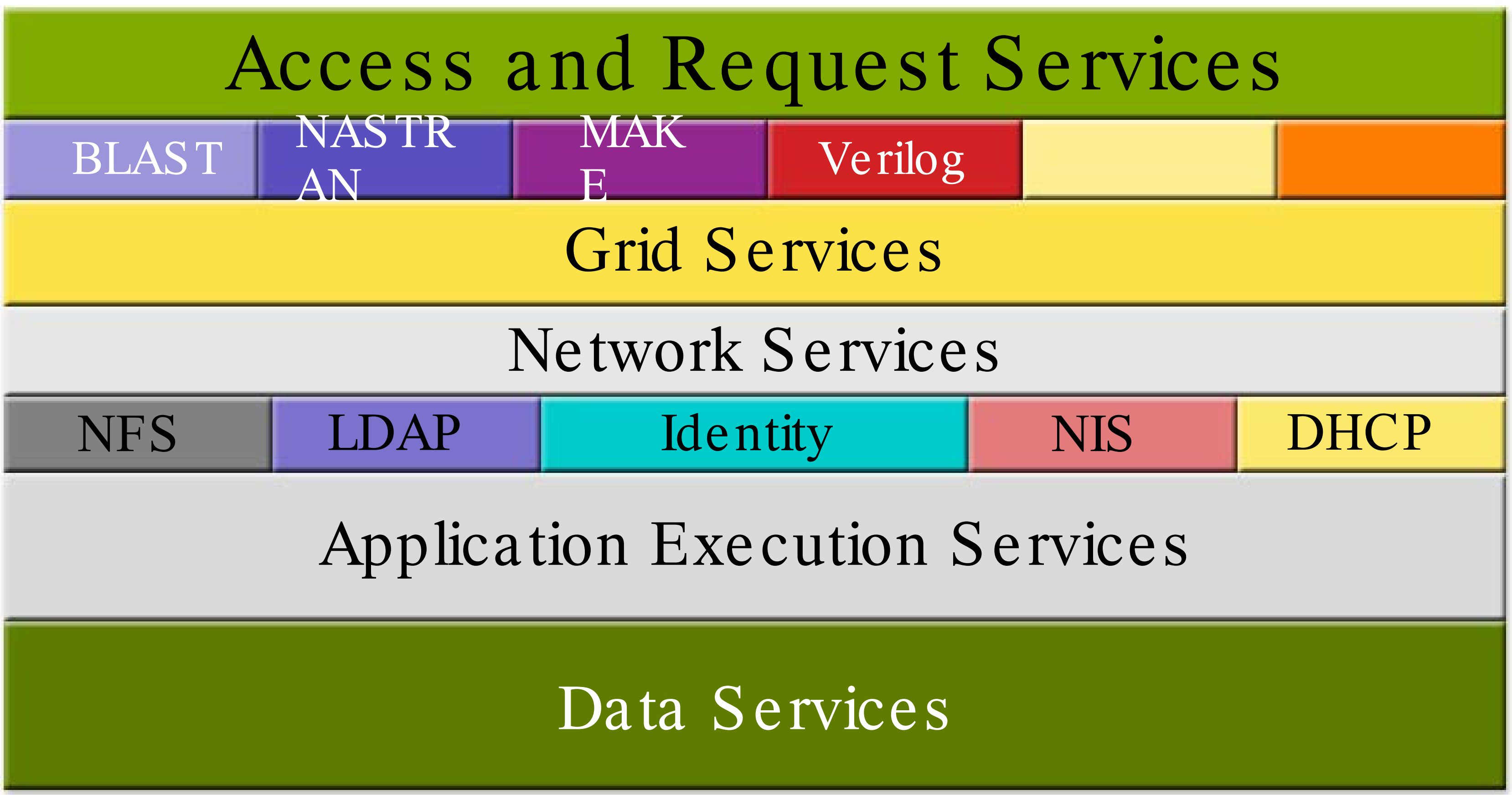
Business Drivers

- Reduced Costs
    - Efficient use of resources, TCO, ROI
    - Utilize latent cycles
- Shorten Time to Market
- Improve business decisions, reduce risk
- Faster, Better, Cheaper
- New Capabilities, improved quality, innovation

Application Requirements

- Basic Information
  - Primary stakeholders
  - Operating Systems and support mechanisms
  - Deployment model
  - I/O and data requirements
- Execution Profile
  - Dynamic memory allocation, footprint
  - Performance characteristics, reference benchmarks
  - Network I/O
  - Job duration characteristics
  - Utilization trends
- Application Characteristics
  - Dependencies
  - Priorities vs. other applications
  - Data Management post execution
  - Checkpointing capable

Application Stack

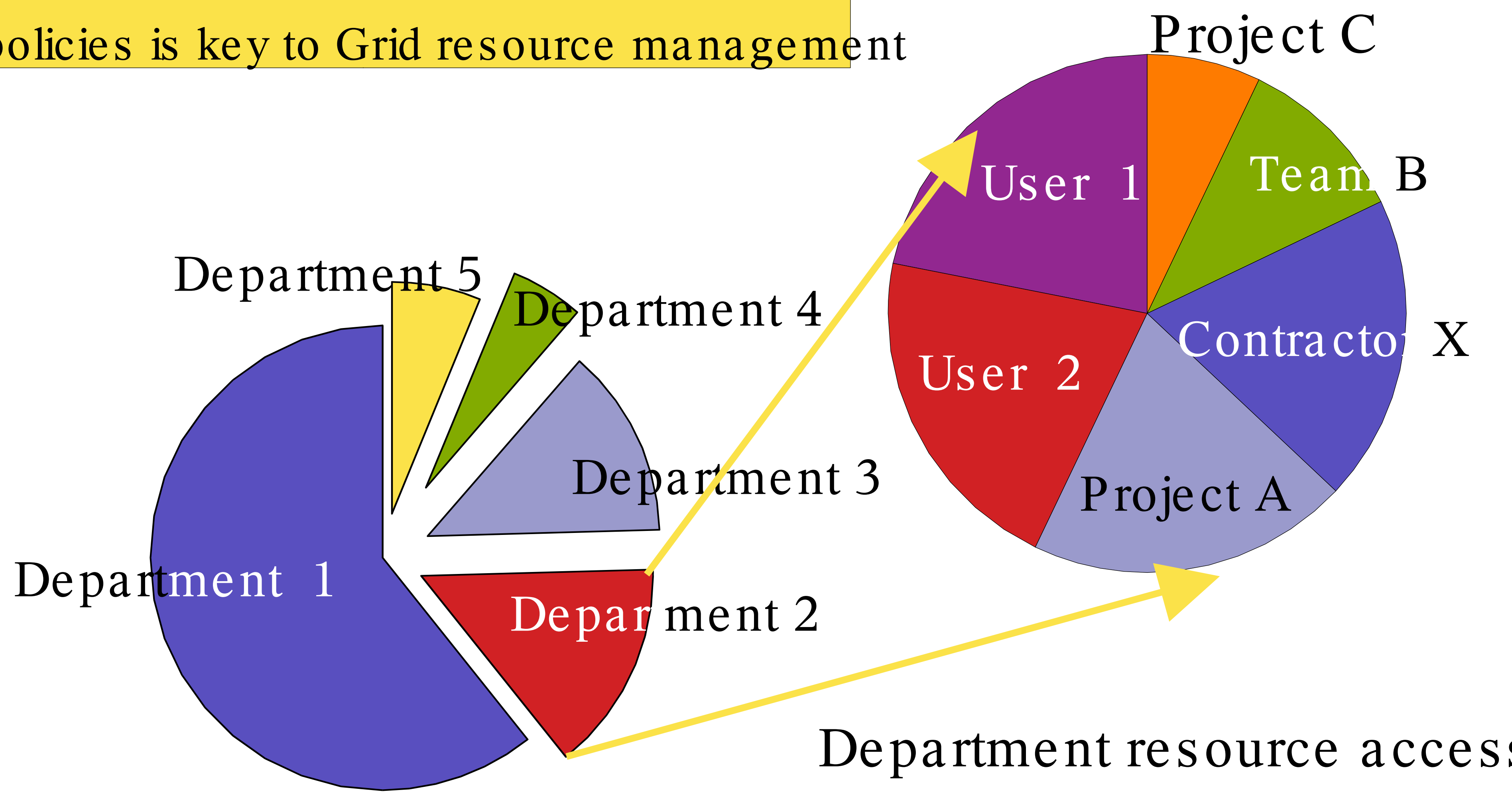


Architectural Requirements

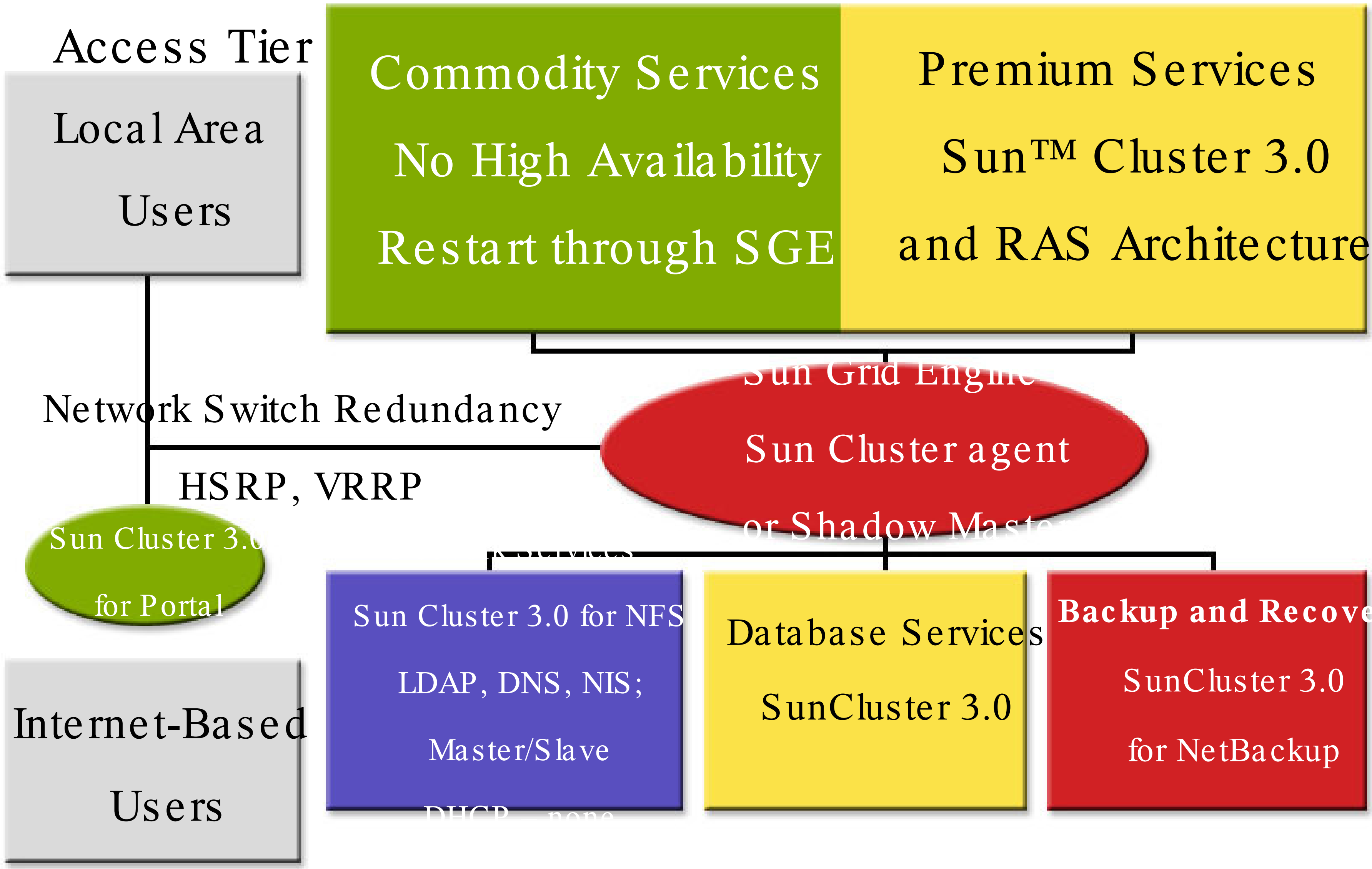
- Availability
  - Downtime impact of Grid environment
  - Impact of interruption of individual jobs
  - Acceptable maintenance windows
- Scalability
  - Anticipated growth over 1–3–5 years
  - Desired scaling strategy and response to peak loads
  - Strategy for technology refresh, evolution
- Manageability
  - Skill set/workload of administration personnel
  - Expected stability of applications
  - Code management, software distribution mechanisms
- Security
  - User authentication mechanisms
  - Internet access to grid
  - Data Security requirements

Grid Policy Management

Clear definition and documentation of priorities and policies is key to Grid resource management

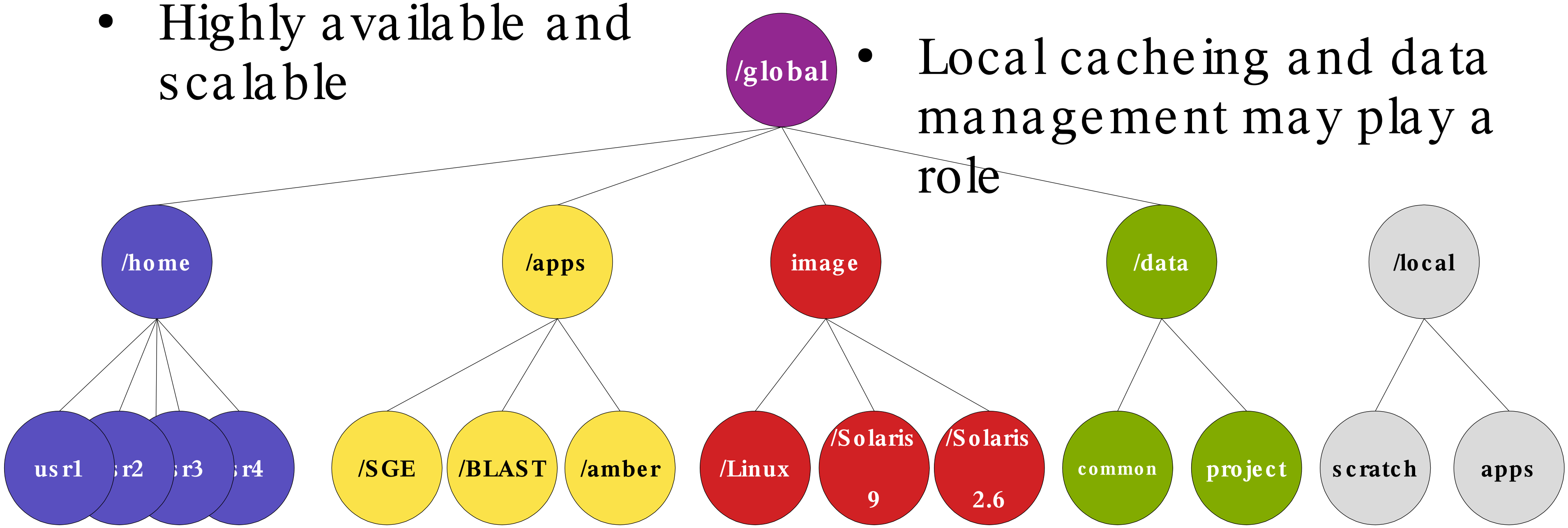


Availability Model



Data Model

- Network File System
  - Dominates Grid Environments
  - Naming Service
  - automounter
- Highly available and scalable
- Key to efficient deployment and management
- SAN Technologies usually present behind data service
- Local cacheing and data management may play a role



Scalability

- Network architecture constrains grid scalability
- NFS Architecture scales w/ support of SAN + Server Technologies
- Scalability Model
  - Network Architecture
  - Storage Architecture
  - Operations Model

Security

- Grid Security lays on top of existing Security architecture, good or bad
- Security Issues
  - Grid may contains resources owned by different distrustful organizations
  - Data needs to get in/out of the grid
  - Users need access to the grid
  - Insecure protocols such as NFS and automount
- Security Requirements
  - Single sign-on
  - Protection of Credentials
  - Interoperability w/ local security standards
  - Secure communication
  - Multiple implementations

Security Recommendations

- Understand Security Policy
  - Multiple Trust domains, policy domains
  - Access from employees, partners, customers
  - Sensitivity of data, information, resources, communication
- Secure the Environment
  - Security architecture techniques
    - Apply to Grid Service Infrastructure
    - Segregate the Grid from external networks with Firewalls
    - Harden operating system on Grid Nodes
  - Authentication / Authorization
    - Enforce access to the Grid
    - Authorization at the user/group level
  - Do not allow users direct connection to the grid nodes
  - Develop secure data transfer

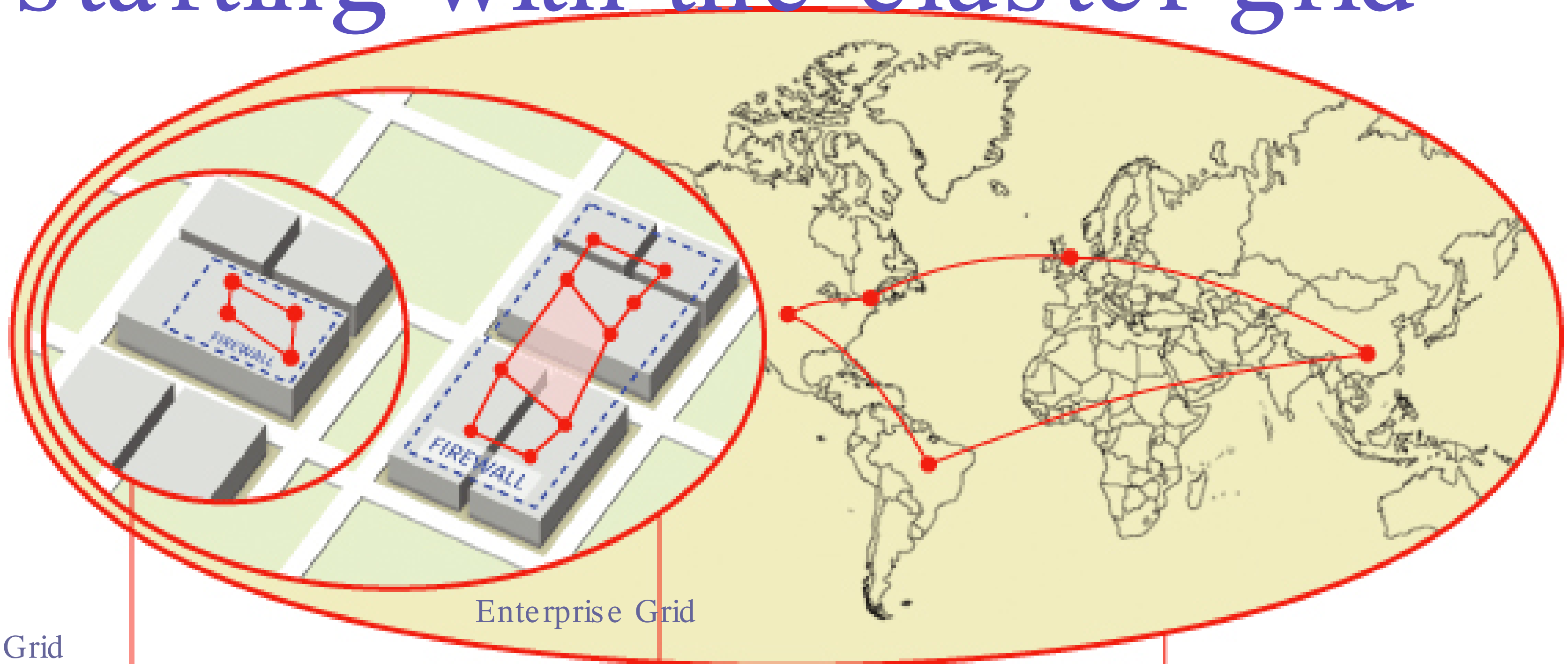
Additional Requirements

- Data Distribution
  - Location, volume, refresh, and security of data
- Usability
  - Skill set and client environment can impact usability of environment
  - Psychological factors important
- Operations Management
  - Limited staff to manage hundreds or thousands of CPUs
  - Resources added in large blocks
  - Change control is extremely critical operation

Grid Engine

- Resource management software that aggregates compute power and delivers it as a network service.
- Provides consistent and pervasive access to high-throughput computational capabilities, and the job accounting and statistics needed to monitor and improve resource utilization.

Grid Evolutionary Strategy: starting with the cluster grid



Cluster Grid

Departmental Computing

- Simplest Grid deployment
- Maximum utilization of departmental resources
- Resources allocated based on priorities

Enterprise Grid

Enterprise Computing

- Resources shared within the enterprise
- Policies ensure computing on demand
- Gives multiple groups seamless access to enterprise resources

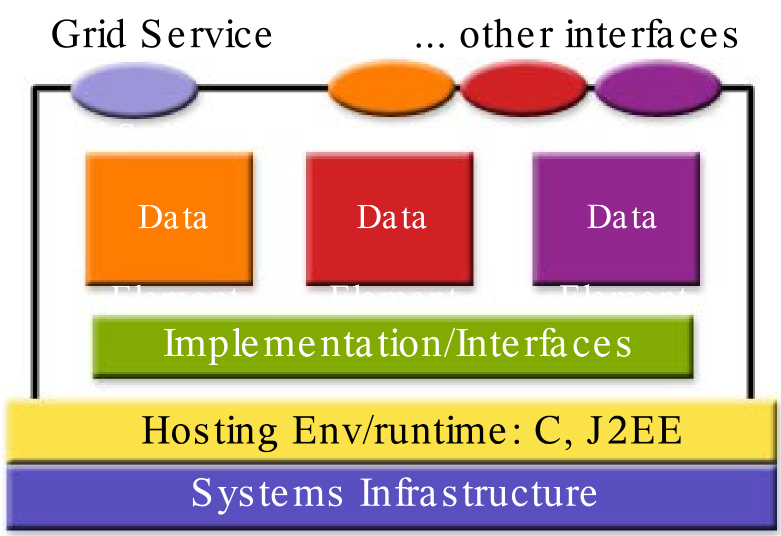
Global Grid

Internet Computing

- Resources shared over the Internet
- Global view of distributed datasets

Open Grid Services Architecture (OGSA)

- Evolution and Integration of Grid technologies and Web Services
- OGSA defines a "Grid Service"
- In terms of WSDL interfaces, defines mechanisms:
  - creating distributed systems
  - lifetime management
  - change management
  - credential management
  - notification



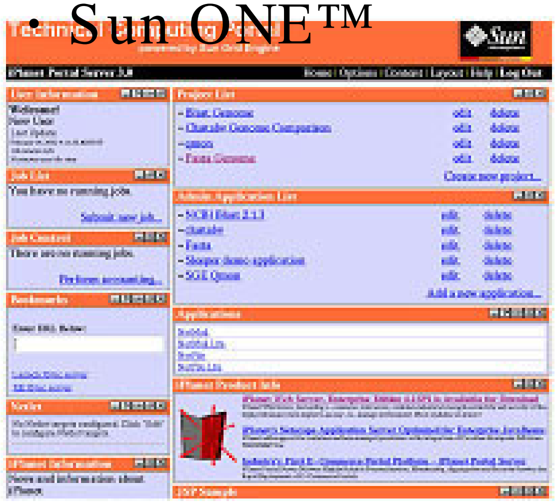
- Notification
- Authorization
- Service Creation
- Service Registry
- Manageability
- Concurrency

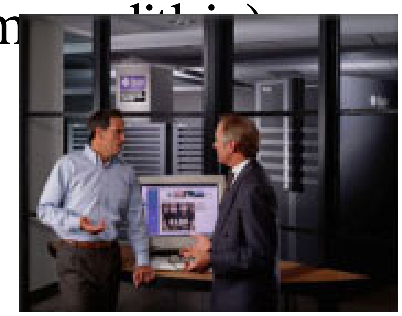
Sun Collaborative Computing

Sun Technologies

Constituencies

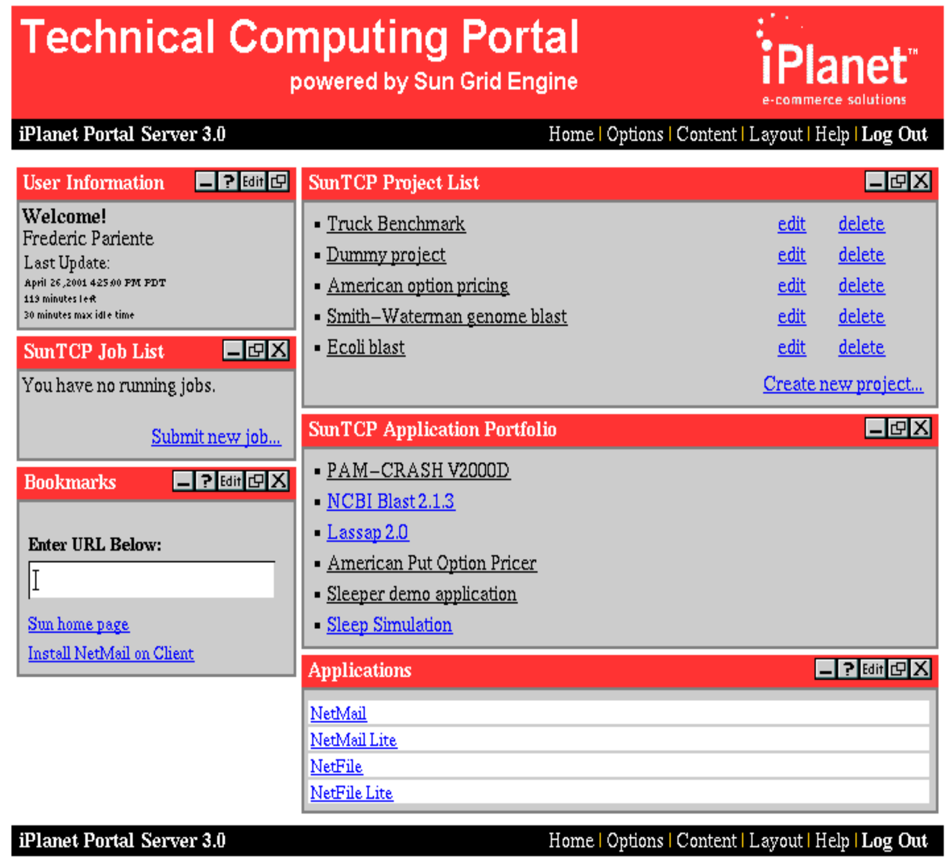
Client	Portal	Shared Pool of Resources
• Java™ • Jini™	• Sun ONE Studio™ Suite • Sun ONE™	• Virtualization of network, compute, storage resources—N1 initiative • Resource optimization • Thin-node and fat SMP clusters • Single OS (monolithic) scalability • RAS
• Java™ • Jini™		
• Scientist • Engineer	• Scientist developer • Engineer developer • Service Providers • ISVs	• Data Center managers • IT stakeholders





Usability

- Command line interaction to launch jobs
- GUI's often live in Unix® world
- Additional steps/products integrating with Windows OE
  - NFS Integration
  - Telnet windows
  - X Server
  - User ID Mgt.
- Technical Computing Portal
  - Centralize Job Mgt.
  - Standardize access and authentication
  - Extend standard interface to users, customers, partners



Sun Enabling Technologies

- Sun™ Grid Engine Software
  - High-throughput computational capabilities and resource utilization for cluster grids
  - Enterprise Edition adds policy management for shared-ownership campus grids
- Sun ONE™ Software
  - Portal Server: Internet services deployment platform
  - Services to quickly, securely deploy technically demanding portals

Supports Solaris, Linux, and Other Operating Systems

What is a Portal ?

- A Web Server plus
  - Security
  - Authentication (Unix, LDAP, secured,...)
  - Personalization
  - Application Access (business & technical)
  - Content Aggregation

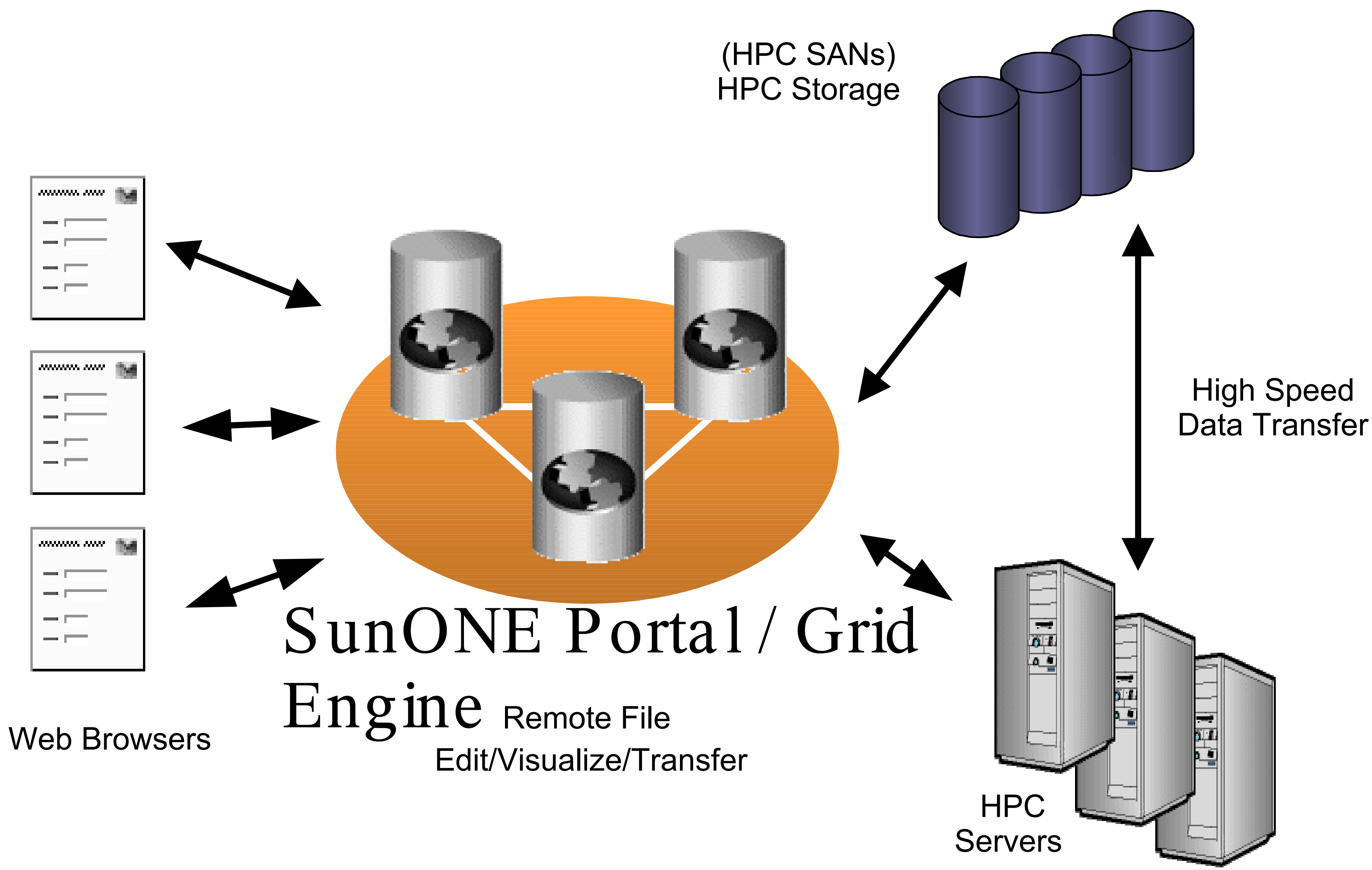
SunONE Portal Server©

What is an *LS-DYNA* Portal?

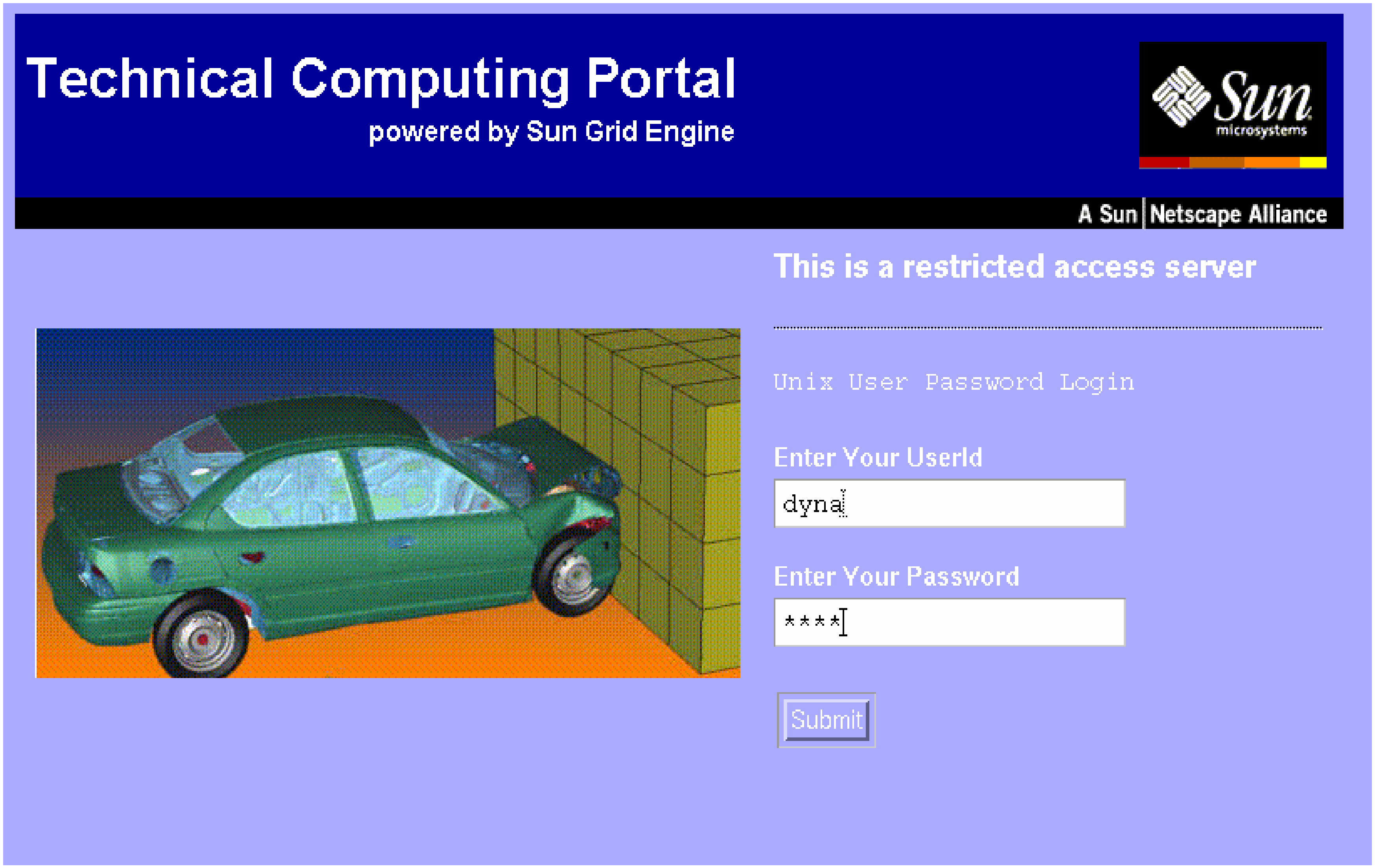
- LS-DYNA access through a Browser
  - Make system complexities *transparent*
  - Remote Access / Job Submission
  - Remote Job Management
  - HTML forms for job submission
  - Upload / Download files remotely
  - Separate Interfaces for Suppliers / Partners

Secure anytime, anywhere access

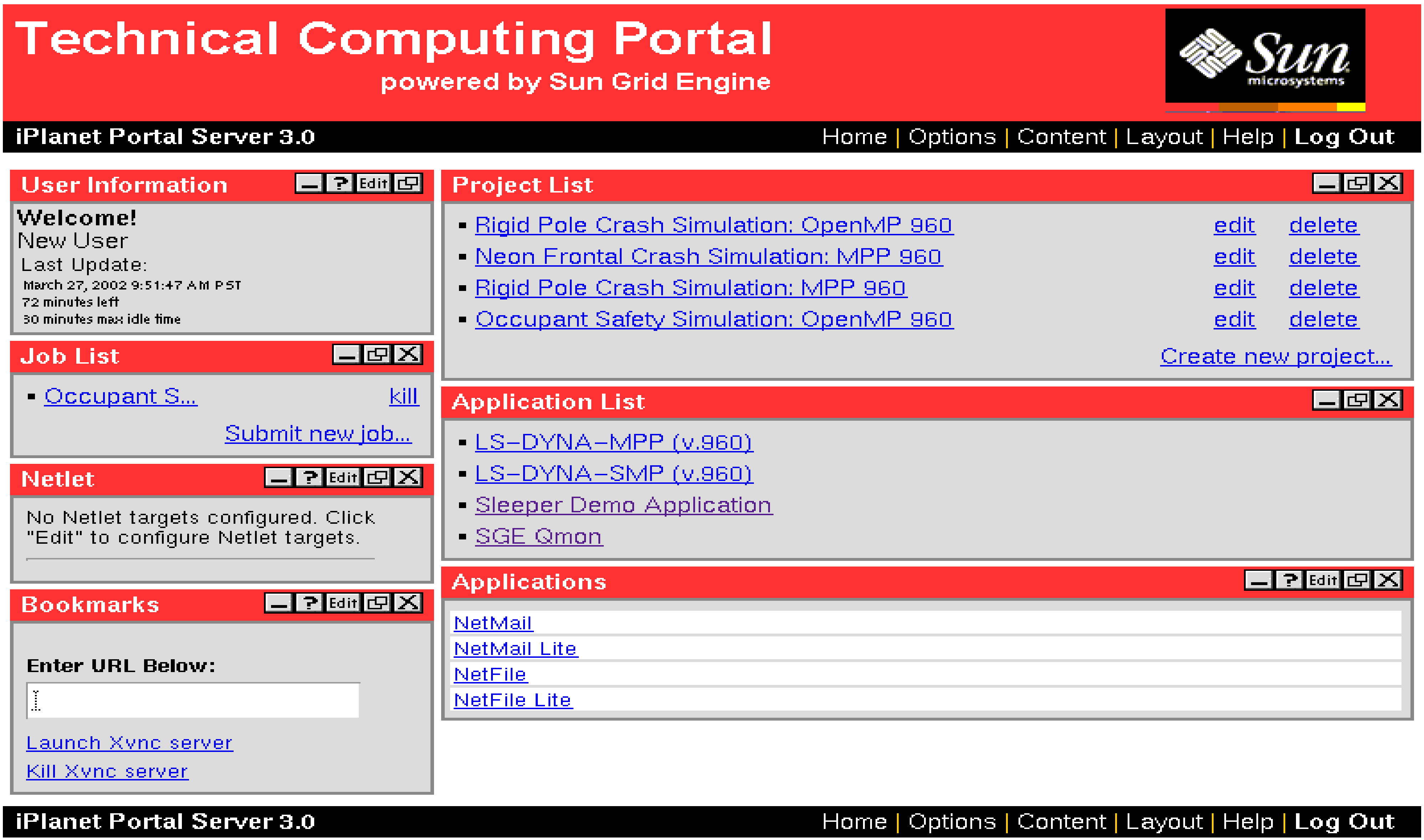
LS-DYNA Portal Architecture



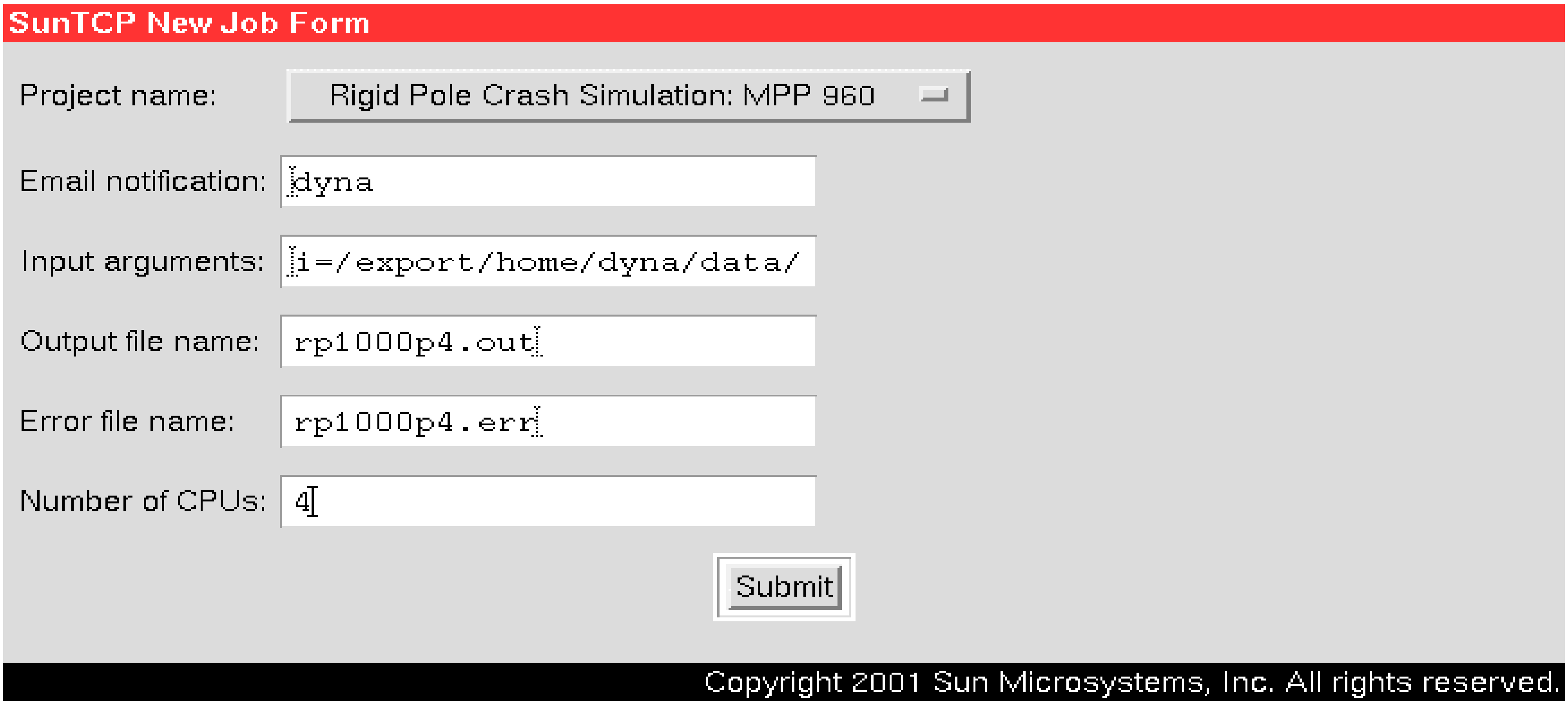
Login Screen



User Access View



Job Submission



SunTCP New Job Form

Project name: Rigid Pole Crash Simulation: MPP 960

Email notification: dyna

Input arguments: i=/export/home/dyna/data/

Output file name: rp1000p4.out

Error file name: rp1000p4.err

Number of CPUs: 4

Submit

Job Input Form

Netscape:

LS–DYNA–MPP Input Selection

Input Data :	Rigid Pole
Run to ncycle :	1000
Memory setting :	1000000
Pfile name :	

Continue

Help

File Viewer

SunTCP Project Information

Project name: Rigid Pole Crash Simulation: OpenMP 960

Application: SMPDYNAFORM

Exports:

Files:			
	.suntcp-project	head	tail
	.suntcp-su	head	tail
	rp200p2.out	head	tail
	rp200p2.err	head	tail
	messag	head	tail
	d3hsp	head	tail
	d3plot	head	tail
	d3thdt	head	tail
	status.out	head	tail
	d3dump01	head	tail
	d3plot01	head	tail
	d3thdt01	head	tail

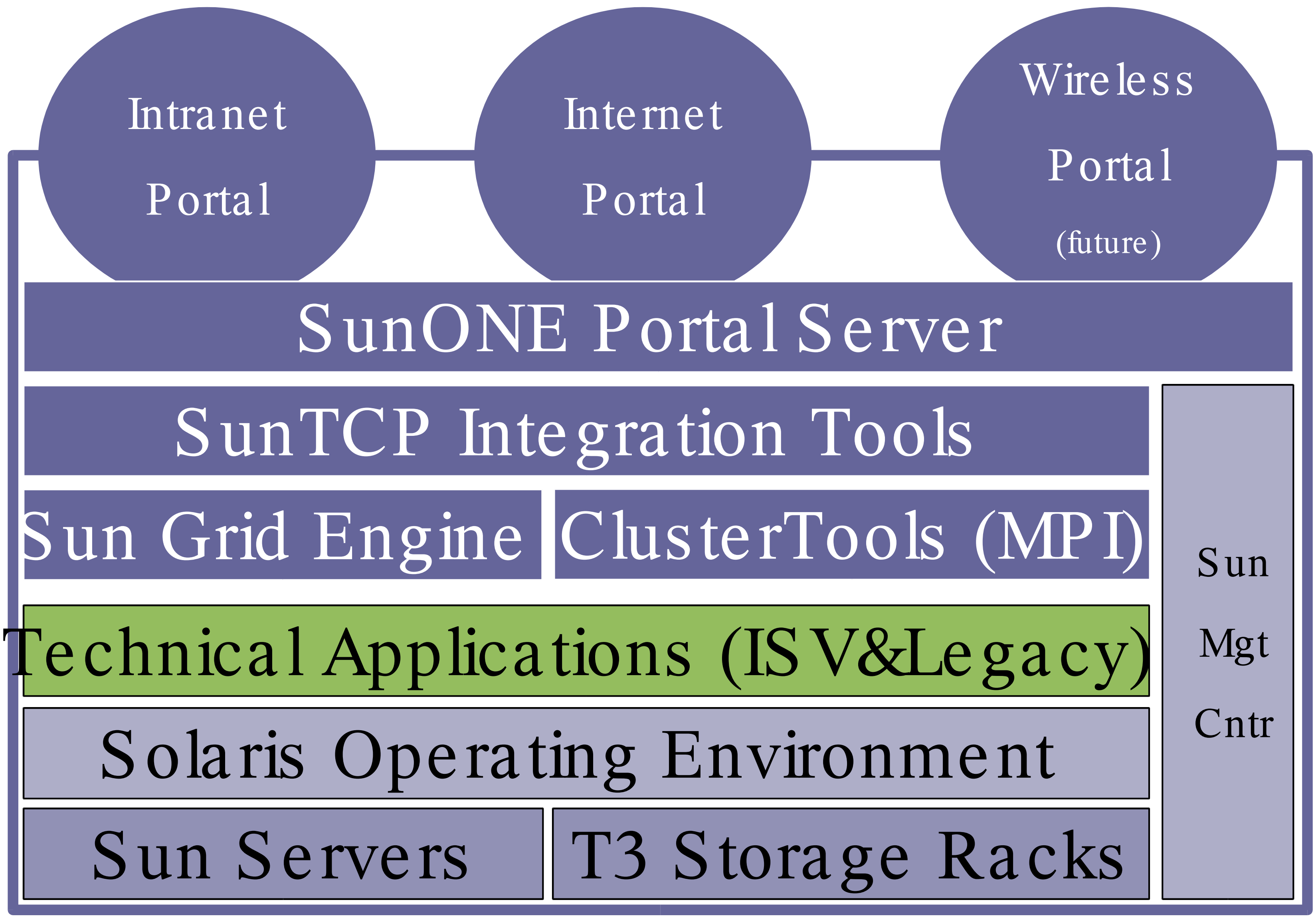
Continue

Output File

SunTCP File View

```
estimated clock time to complete  =      12399 sec (        3 hrs 26 mins)
      1 t 0.0000E+00 dt 1.16E-06 flush i/o buffers
      1 t 0.0000E+00 dt 1.16E-06 write d3plot file

******** termination cycle reached ********

    200 t 2.3176E-04 dt 1.16E-06 write d3dump01 file
    200 t 2.3176E-04 dt 1.16E-06 write d3plot file

N o r m a l   t e r m i n a t i o n

Storage required for explicit solution:      1067632

T i m i n g   i n f o r m a t i o n
                       CPU(seconds)   %CPU  Clock(seconds) %Clock
 ----------------------------------------------------------------
 Initialization ....... 1.7300E+00    4.26    9.0089E-01    4.42
 Element processing ... 3.6780E+01   90.52    1.8315E+01   89.91
 Binary databases ..... 1.7000E-01    0.42    8.8151E-02    0.43
 ASCII database ....... 0.0000E+00    0.00    1.4046E-03    0.01
 Contact algorithm .... 1.9500E+00    4.80    1.0620E+00    5.21
   Interface ID       1 1.9200E+00    4.73    1.0542E+00    5.18
 Contact entities ..... 0.0000E+00    0.00    0.0000E+00    0.00
 Rigid bodies ......... 0.0000E+00    0.00    3.2400E-03    0.02
 Implicit Nonlinear ... 0.0000E+00    0.00    0.0000E+00    0.00
 Implicit Lin. Alg. ... 0.0000E+00    0.00    0.0000E+00    0.00
 ----------------------------------------------------------------
 T o t a l s            4.0630E+01  100.00    2.0370E+01  100.00

Problem time       =    2.3176E-04
Problem cycle      =       200
Total CPU time     =        41 seconds (    0 hours  0 minutes 41 seconds)
CPU time per zone cycle  =  35635 nanoseconds
Clock time per zone cycle=  17835 nanoseconds

Number of CPU's      2
Start time    03/27/2002 10:18:48
End time      03/27/2002 10:19:10
Elapsed time        22 seconds (    0 hours  0 minutes 22 seconds)
```

TCP Hardware and Software Stack



With TCP, You Can...

Enjoy secure *anytime, anywhere access* – submit jobs remotely

Make HPTC resources available to other users without duplicating the investment

Upload files with the click of a button

Improve resource management and harness idle computing cycles

Manage job submission via forms

Dynamically check on the status of jobs

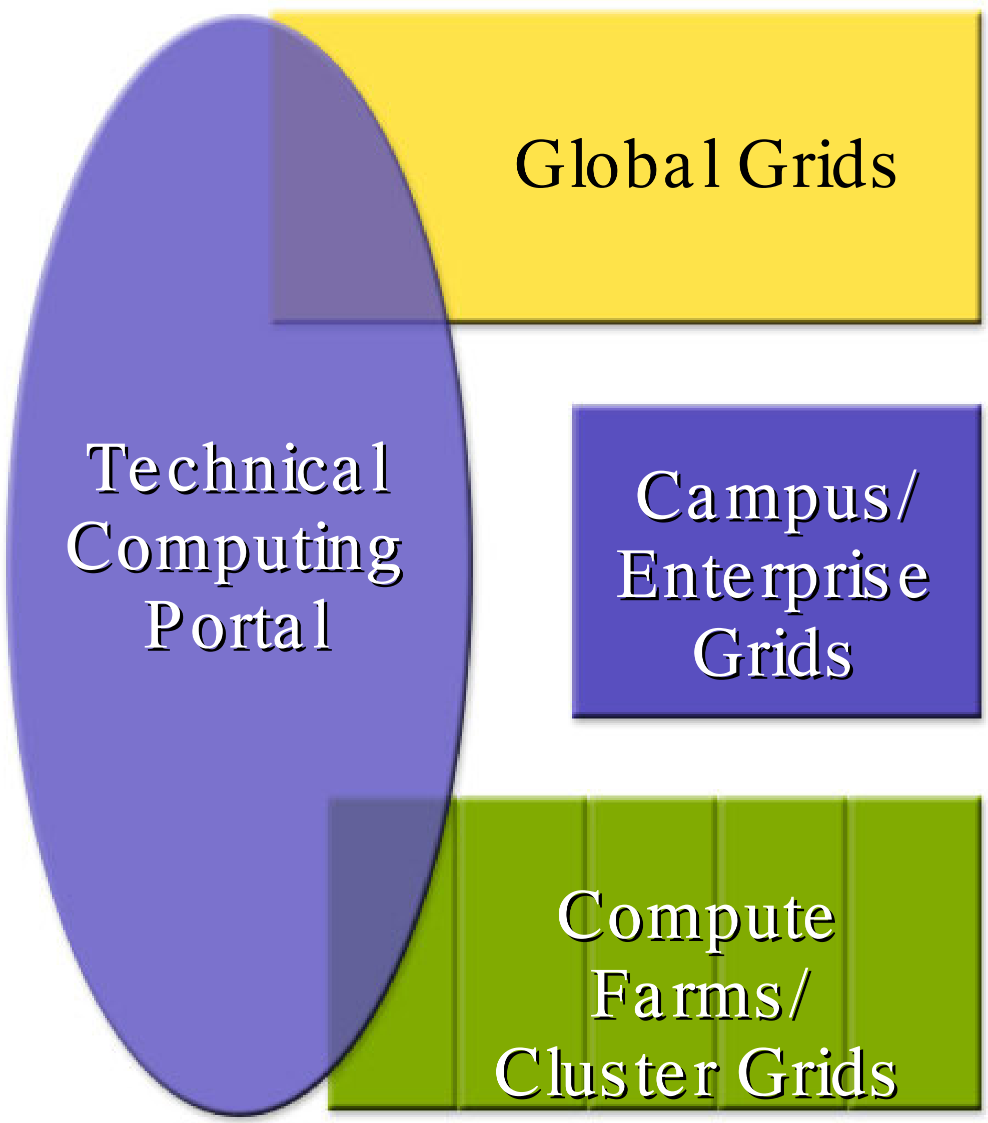
Receive email notification when jobs are complete

Download/visualize output files remotely

Easily share results with collaborators

Quickly add legacy apps for sharing and remote use

Sun Powers the Grid



- Caprion Pharmaceuticals
- Cognigen Corp.
- EPCC
- Ford Motor Company
- Ohio Supercomputer Center
- Motorola
- RWTH Aachen
- White Rose initiative

Ford Motor Company

- Sun HPTC solution
  - 500 dual-CPU workstations
  - Sun Grid Engine software
- Business Impact
  - Run MCAD applications on 500 CPUs by day
  - Run MCAE applications 7 x 24 on same systems configured as a grid
    - 500 CPUs daytime; 1000 CPUs nights and weekends
  - Greatly increased throughput
  - More analysis drives higher quality research and lower cost
  - Faster design decisions reduce time to market

Caprion Pharmaceuticals:

LEADING PROTEIN ANALYSIS COMPUTE FARM

- Sun HPTC solution
  - Based on SPARC/Solaris
  - Range of Sun Fire servers (280R to 6800)
  - Forte™ (Sun Studio) Development Tools
  - Sun Professional Services
- Bioinformatics firm, leader in proteomics
  - Discovers and develops pharmaceutical products
- Criteria
  - Build compute farm that can address the biggest issues in life sciences today
  - Process hundreds of terabytes of data per year
  - Create architecture that can grow exponentially
  - Better collaboration, faster time-to-market

Aachen University

- Leading European technical university
  - Replace outdated, vector-based supercomputer
  - Four teraflops goal
  - Collaboration focus on CFD and VR Tools
- Criteria
  - ISV relationships
  - Scalability, availability
  - Price/performance
  - Vendor with collaboration-oriented approach
- Sun HPTC solution
  - Four Sun Fire 15K systems, 16 Sun Fire 6800 systems
  - Aggregates to 1.2 teraflops
  - Sun™ HPC ClusterTools software
  - Sun™ Grid Engine for statewide access

HPC Virtual Labs (HPCVL)

- Consortium of Four Canadian Universities
  - Recently deployed massive Sun HPC environment
  - Expected to be among top 100 HPC sites in the world
- Criteria
  - Innovative partnership program
  - Robustness, reliability
  - Price/performance
  - Future product roadmap
- Sun HPTC Solution
  - 10 Sun Fire™ 6800s and 2 Sun Fire 15Ks
  - Sun Enterprise™ 220R, Sun StorEdge™ T3 arrays
  - Solaris™ 8 Operating Environment
  - Sun HPC ClusterTools™, Forte™, Sun Grid Engine, Sun Resource Manager™, Sun Management Center, Solstice DiskSuite™ software

Conclusions

Grid Computing has the potential to extend computing into a utility-type environment. If this is to happen, Grid computing must be built upon an infrastructure designed to provide carrier-grade Quality of Service