Vision for the DOE Science Grid

Bill Johnston (LBNL), Ian Foster (ANL), and Ray Bair (PNNL)

The Vision for a DOE Science Grid

- Programs like SciDAC that use computing to address otherwise intractable science problems will require use of the computing, data, and instrument resources across multiple Labs and Universities

- Building the teams needed to solve these large scale scientific problems will require extensive use of distributed collaboration tools and facilities

- **It is the role of Grids – tools and middleware for widely distributed systems – to enable the degree of scalability in scientific computing necessary for DOE to accomplish its missions in science**
- **The DOE Science Grid's major objective is to provide the advanced distributed computing infrastructure that will be essential for DOE science in the future**

The vision for “Grids” is to revolutionize the use of computing in science by making the construction and use of large scale systems of diverse resources as easy as using today’s desktop environments.

Expected Outcomes

Grids will present a *uniform usage and management interface* to computing, collaboration, storage, and instrument systems, and provide the *capability of scalably connecting* these into large, on-demand systems. This should lead to:

- **Increased mobility and access to computing and data by computational scientists**
- **Routine collaboration among DOE Labs and their university partners through ready and secure access to collaboration tools, remote instruments, and petabyte size data sets**

- **Construction of systems of aggregated resources for solving large-scale problems that involve coupling multiple computational simulations and data archives**

- **New approaches to laboratory science through the coupling of large-scale computing and storage systems to instrument systems in order to provide real-time analysis of experiment data and feedback based experiment control**

- **Standardized architecture and tools that make it easier to incorporate new computer architectures, data systems, and instruments into a usable application environment**

Motivating Application Classes

- ***Scientific data analysis and computational modeling with a world-wide scope of participants*** **– e.g. High Energy Physics data analysis and climate modeling**
- ***Real-time data analysis for on-line instruments*****, especially those that are unique national resources – e.g. LBNL's and ANL's synchrotron light sources, PNNL's gigahertz NMR machines, etc.**

- ***Generation, management and use of very large, complex data archives*** **that are shared across an entire community – e.g. DOE's human genome data and NASA's EOS data**

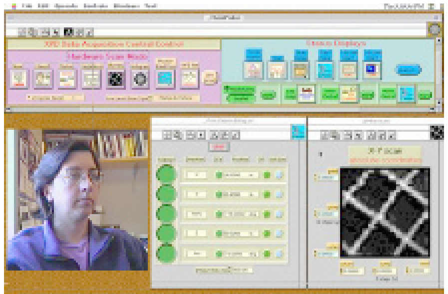
- ***Collaborative, interactive*** **analysis and visualization of massive datasets – e.g. DOE's Combustion Corridor project**

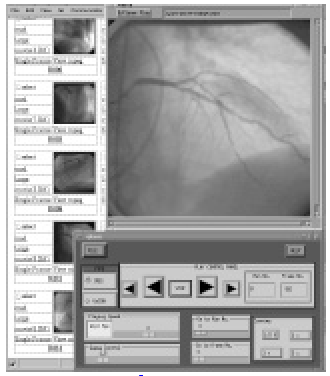
Addressing the requirements of these classes of applications in a general way, with common Grid infrastructure deployed across the DOE Labs and collaborating universities, will enable many different applications to routinely use widely distributed resources**.**

Current Environment

- Some good tools, with others being developed
- Very little persistent common infrastructure
- Result is that most distributed application environments deal with a restricted set of resources in an ad hoc way, and every project re-invents this scenario

Problem Solving Environments

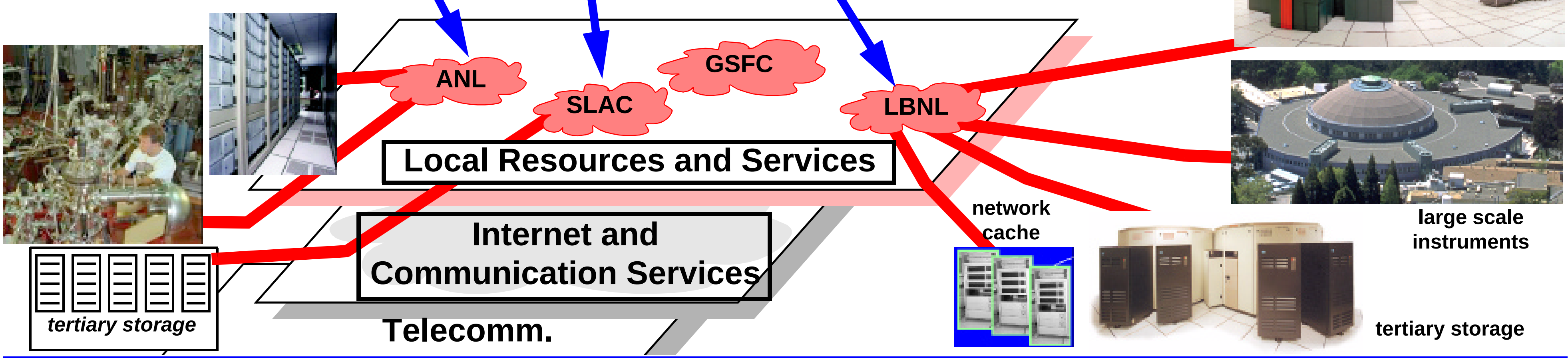




Distributed Environment

Currently: PSEs typically deal with a restricted set of resources in an ad hoc way: There is no common middleware that provides the uniformity needed to expand the scope of these applications.

Future: The DOE Science Grid will provide a shared, and when possible, uniform view of resources in order to facilitate access and construction of distributed systems.



Approach and Goals for the DOE Science Grid

- **Grids are built through collaborative efforts, and at the same time facilitate collaboration: The DOE Science Grid will be a collaboration among DOE Labs**

- **Deployment of existing technology (Globus, Condor, Grid portals, etc.) will provide for relatively rapid impact – the first version of the Science Grid will be built from computing and storage resources in the DOE computer science and applications communities**

Approach and Goals

- **ESNet will provide the development and support for Grid directory and security services - this will ensure persistent and usable infrastructure across the DOE Labs**

- **Grid support for building collaboration services will both facilitate construction of the Grid and more readily provide collaboration tools to users**

- **A Science Grid operational model will be developed that provides for easy user access and for local control of resources that are connected to the Science Grid**

- **Strong security will be provided from the start in order to address authentication, authorization, and infrastructure assurance in open science networks for both applications and Grid services**
- **As Grid services are debugged and validated they will be deployed on the NERSC systems to provide a uniform supercomputing environment – the production Science Grid represents a new service delivery model for DOE computing, data, and instrument resources**

- **Applications will be identified that both test Grids and benefit from Grids – a powerful application development and production environment will result**
- **Analysis and resolution of issues uncovered during the deployment of the Science Grid will lead to R&D on technologies to provide new capabilities**
- **Active cooperation with the other major players in the Grid community, e.g. the NSF PACIs and NASA's IPG, is facilitated through the Grids Forum – this will ensure that all of these groups can leverage each other's work (as is already happening)**

Experience with Grid-like Systems

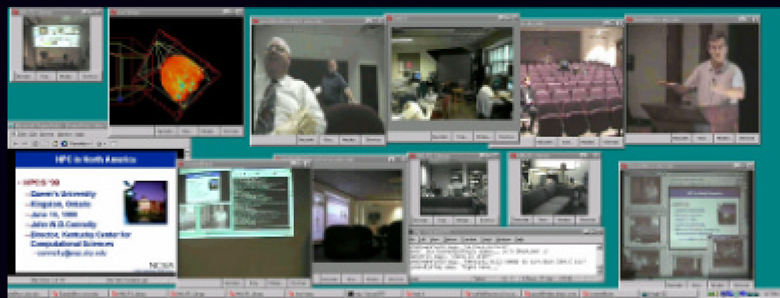
- **SF Express**
- **Access Grid**
- **Kaiser on-line cardio-angiography system**

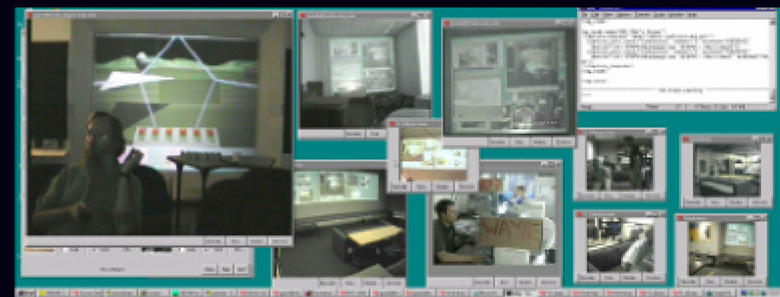
Access Grid Project Goals

- Enable Group-to-Group Interaction and Collaboration
  - Connecting People and Teams via the Grid
- Improve the User Experience: Go Beyond Teleconferencing
  - Provide a Sense of Presence
  - Support Natural Interaction Modalities
- Use Quality but Affordable Digital IP Based Audio/video
  - Leverage IP Open Source Tools
- Enable Complex Multisite Visual and Collaborative Experiences
  - Integrate With High-end Visualization Environments
  - ActiveMural, Powerwall, CAVE Family, Workbenches
- Build on Integrated Grid Services Architecture
  - Develop New Tools Specifically Support Group Collaboration

MCS ✦ FUTURESLAB ARGONNE ✦ CHICAGO

Access Grid Milestones





- AG Tutorials @ Argonne
- Multisite AG Testbed Operational June 14, 1999
  - ANL, ACCESS-DC, UNM, UK, BU, Utah, EVL, NCSA
- AG Test Cruises/Nanocruises Ongoing
- AG Toolkit v0..0 Functionality defined
  - Multiscreen NT workstations, Multipipe SGIs
  - ImmersaDesk "remote node only" functionality
- AG Web site Operational (AG software, configs downloadable)
- AG Debut at Alliance Chautauqua's and DOE CorridorOne Projects
- Initial AG Toolkit Version Released in FY99
- AG interfaces to "Grid Services Architecture" definition

MCS ✦ FUTURESLAB

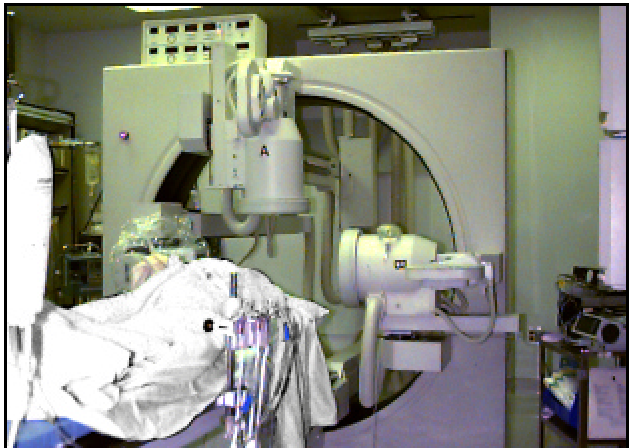
ARGONNE ✦ CHICAGO

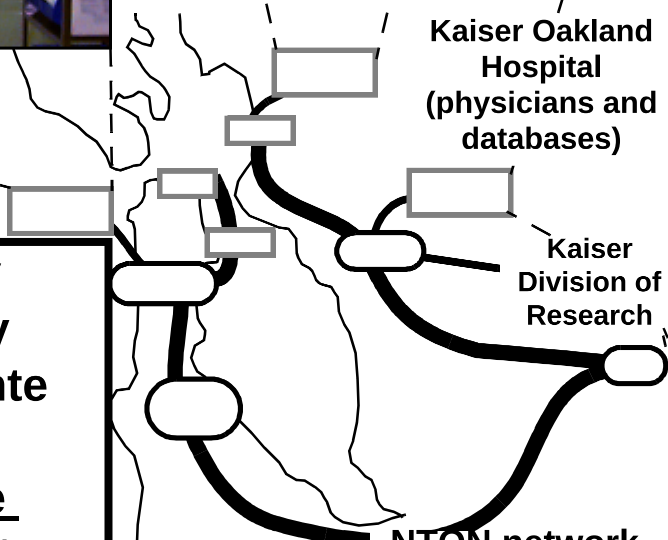
WALDO real-time digital library system and DPSS distributed cache [9] for data cataloguing and storage

Tertiary Storage

Compute servers for data analysis and transformation

Kaiser San Francisco Hospital Cardiac Catheterization Lab (X-ray video imaging system, ≈ 130 mbit/s, 50% duty cycle 8-10 hr/day)

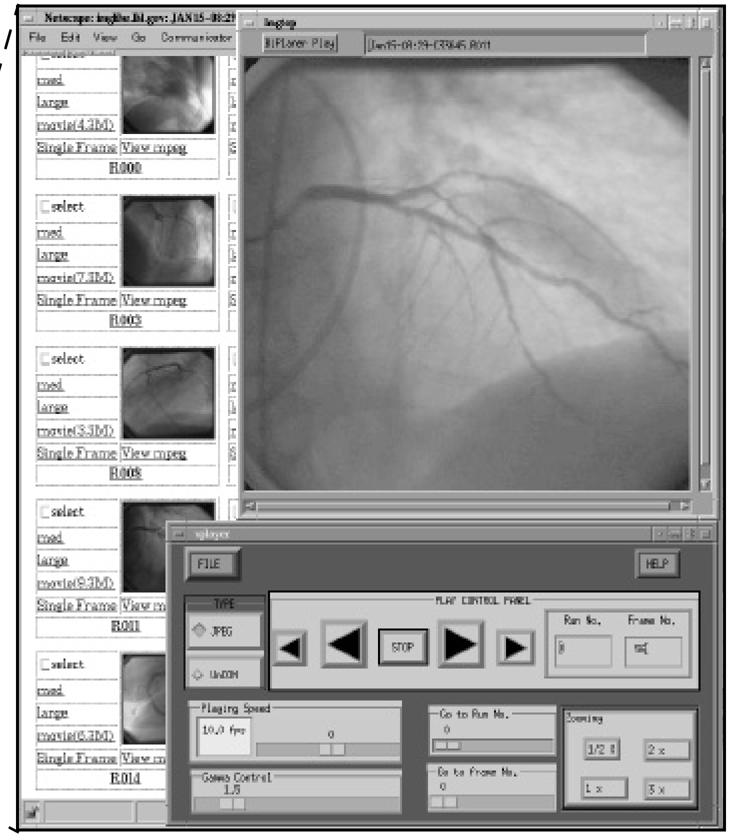




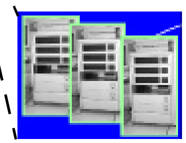
Kaiser Oakland Hospital (physicians and databases)

Kaiser Division of Research

NTON network testbed



The PSE: Automatically generated user interfaces providing indexed access to the large data objects (the X-ray video) and to various derived data.



Lawrence Berkeley National Laboratory and Kaiser Permanente Health Care

On-line Health Care Imaging Experiment

DOE Science Grid

Technology Vision

Grid middleware, services, and resource managers will provide a uniform and location independent view of distributed computing, storage, communication, instrument, and collaboration resources.

The DOE Science Grid will provide persistent infrastructure for resource naming, discovery, access, security, and operational support.

DOE Labs will connect resources to the Science Grid to facilitate collaboration and to build large-scale systems.

Grid Architecture

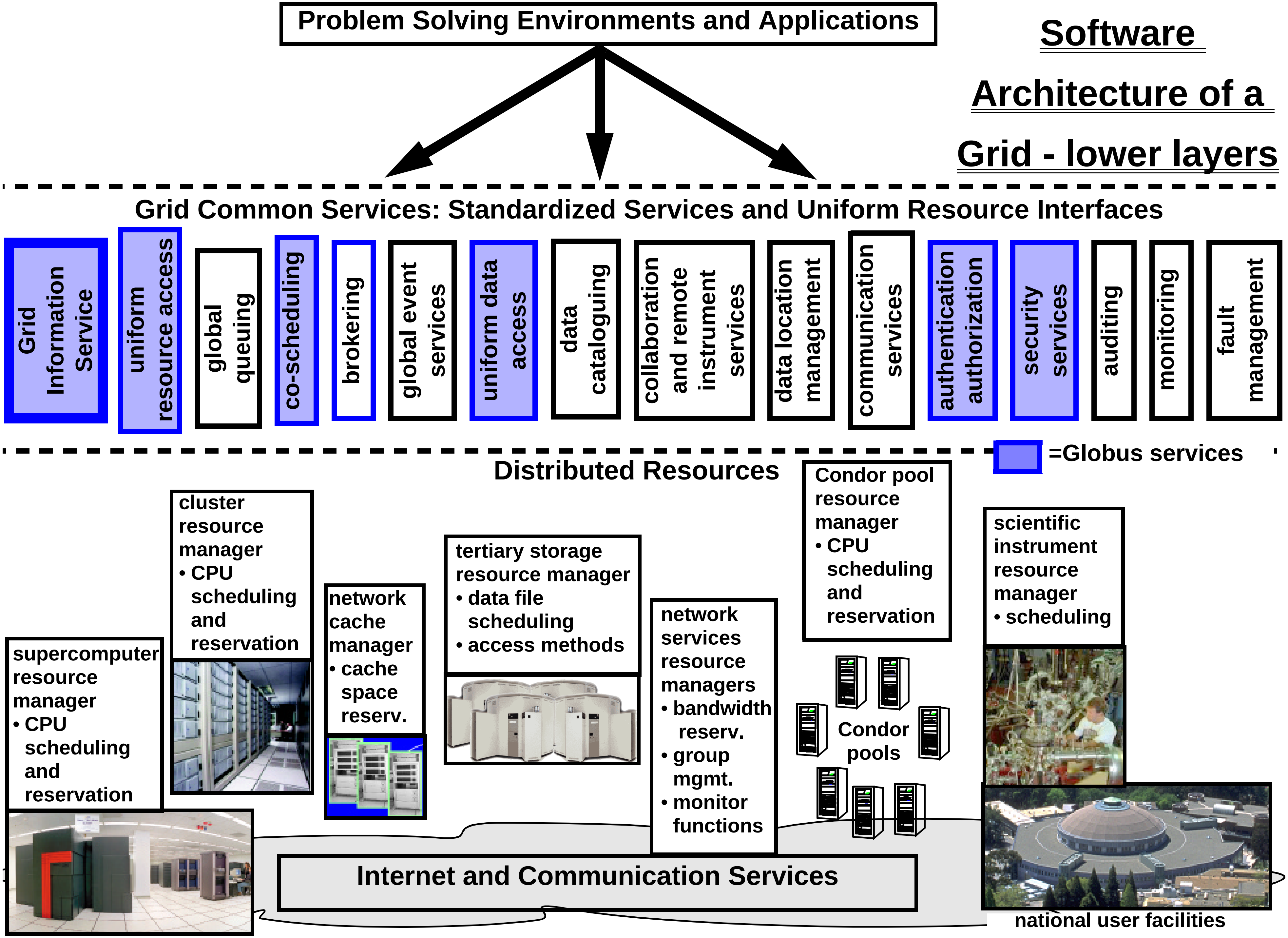
- Problem Solving Environments are the user interface to Grids, and are supported by Grid toolkits for
  - job submission, control, and tracking services
  - workflow management for specific classes of applications (e.g. physics data analysis frameworks or aircraft design parameter study managers)
  - policy based access control, etc.

- **Application development tools and services support various styles of programming in the Grid environment, as well as the development of Grid services themselves. E.g.:**
  - **uniform data access methods developed in the DataGrid project [14] will form the foundation for global storage management services such as MCAT/SRB [22] and the Storage Access Coordination System (STACS), HRM (HPSS Resource Manager) [15]**
  - **Globus I/O enabled MPI library provides coordinated, MPI communication between processes on separate systems**

- **The “Grid Common Services” locate, schedule, and provide uniform views of the underlying resources. E.g.: resource access, naming and location, and co-scheduling for computing, networking, and instrument systems**

- **Resource managers provide the basic functionality and access for the actual resources**
  - **some already exist - e.g. batch schedulers - however do not always support the required functionality (e.g. for advance reservation)**

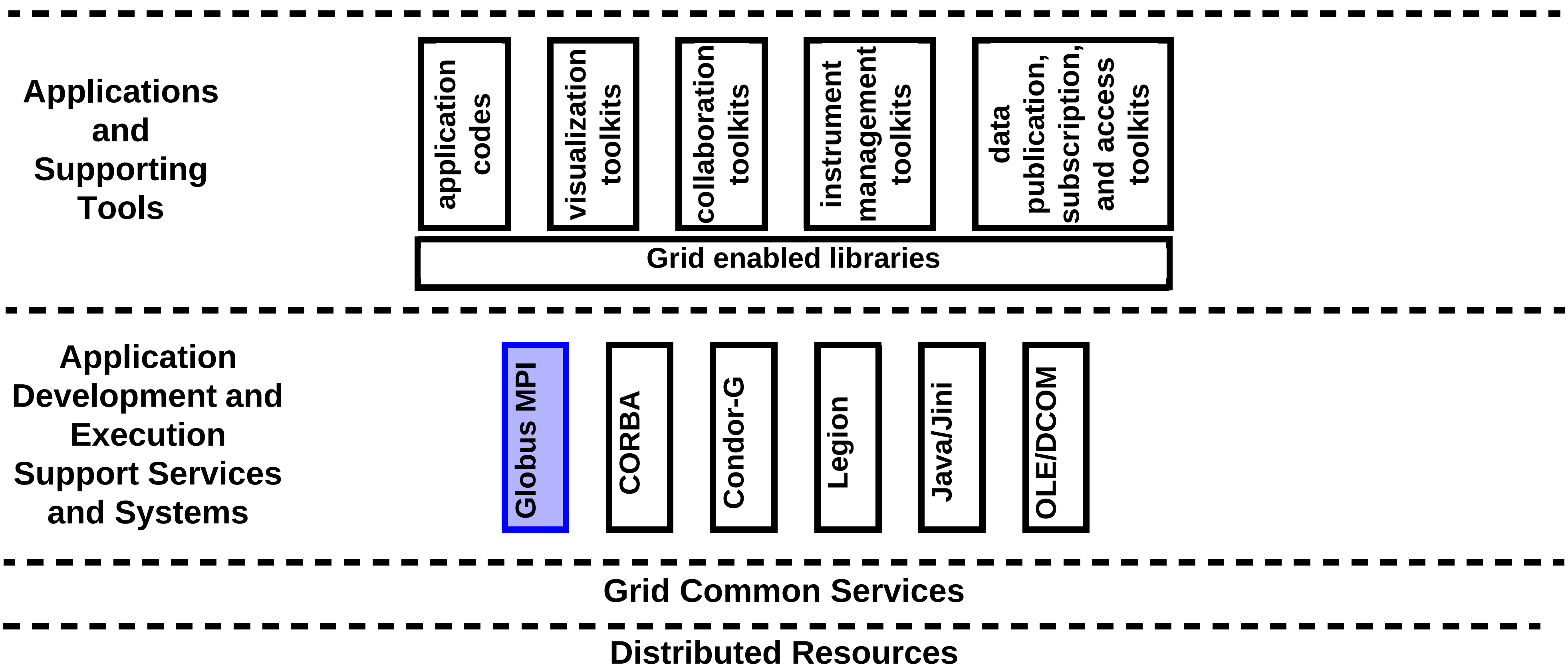
Software Architecture of a Grid - lower layers



Software Architecture of a Grid - upper layers

Problem Solving Environments

- Tools to implement the human interfaces
- Mechanisms to express, organize, and manage the workflow of a problem solution
- Access control
- E.g. SciRun [24], Ecce [25], "portals", WebFlow [26], ...



Roadmap for DOE Science Grid ("DSG")

Near term (1 -2 yr)

- Deploy existing tools (Globus, data access, and portals) across a collection of computer science R&D resources at DOE Labs to provide a prototype Science Grid infrastructure

  Outcomes:

  - enhanced collaboration among DOE Labs and their partners
  - demonstrate computing and storage resource sharing and load leveling
  - support for HENP large-scale data analysis
  - initial application use of the Science Grid

- **Provide uniform, strong, and versatile security services built on PKI/X.509 and Globus Security Infrastructure.**

  Outcomes:

  - facilitation of inter-site collaboration in the face of increased hacker sophistication and increased DOE security requirements

- **Define and support multiple levels of participation in the DSG in order to make the “cost of entry” commensurate with the type of participation**

  Outcomes:

  - **core sites will manage their own resources, provide system support for Grid services, and provide direct assistance to their application development community**
  - **sponsor sites will provide the services of a core site, and manage some of the Grid services of sponsored sites**
  - **sponsored sites can incorporate resources into the Grid, but will not have to manage all aspects of their Grid services**
  - **client site users or un-associated users only have to obtain an identity certificate and install the client-side software on their workstation to use the Science Grid – this will be made very easy**

- **Build a Grid services unit in ESNet to provide the foundation Grid naming and security services.**

  Outcomes:

  - persistent infrastructure for the Science Grid

  - operation of the root directory server for the Science Grid

  - operation of directory servers as a service for sponsored sites

  - operation of the Science Grid root Certification Authority that signs the certificates of other Grid (Lab) CAs

  - operation of a community CA for client sites or un-associated users

- **Provide Grid based services for collaboration tools and environments, and remote instrument interfaces.**

  Outcomes:

  - easier deployment of collaboration tools and remote instrument control systems that build on Grid directory services and security

- **Integrate the Science Grid data access tools with NERSC HPSS testbed**

  Outcomes:

  - Grid user access to HPSS

  - first step in integrating the production HPSS into the Grid

Roadmap (near term)

- **Collaborate with science application communities to get significant applications operating in the prototype Science Grid**
  Outcomes:
  - encourage application developers to use the Grid
  - first step in providing supported user services for the Grid

- **Collaborate with the ASCI DISCOM program**
  Outcomes:
  - promote a uniform DOE computing environment
  - cooperate on the development of Grid technologies

- **Participate actively in all working groups of the Grid Forum**

  Outcomes:

  - ensure that DOE requirements are addressed and incorporated into GF working documents and draft standards

Near/Medium term (2-3 years)

- **Science Grid services deployed across all Office of Science Labs**

  Outcomes:

  - an operational Science Grid actively supporting multi-institutional applications

- **Incorporate NERSC resources into the Science Grid**

  Outcomes:

  - homogenization the NERSC machine room floor and provide uniform "single sign-on" access to all NERSC resources

- **Grid services for cross-system, cross-site co-scheduling of heterogeneous resources: computing systems, storage systems, instruments, network Quality-of-Service, etc.**

  Outcomes:

  - routine construction of large scale and multi-disciplinary problem solving systems

- **Integrate Web portal technology with the Science Grid to provide Web based user access to Grid services, job tracking, etc.**

  Outcomes:

  - easier access to Grid services by the scientific community

Medium/Long term (3-5 years)

- **Global workflow management: global event driven, adaptive job step sequencers and managers, knowledge based application fault management, etc.**

  Outcomes:

  - **ability to coordinate and manage very complex tasks that span many sites and many different resources**

DOE SCience Grid Vision

The DOE Science Grid will provide uniformity, location independence, and capabilities for building complex, on-demand, large scale distributed systems from scientific computing, data storage, scientific instruments, and collaboration resources that are spread across the DOE Labs and their partners.

This will lead to revolutionary new capabilities for solving large-scale scientific problems.

References and Acronyms

[1] Globus is a middleware system that provides a suite of services designed to support high performance, distributed applications. Globus provides:

- Resource Management: Components that provide standardized interfaces to various local resource management systems (GRAM) manage allocation of collections of resources (DUROC). All Globus resource management tools are tied together by a uniform resource specification language (RSL).
- Remote Access: Components that enable remote access to files (GASS and RIO) and executables (GEM).
- Security: Support for single sign-on, authentication, and authorization within the Globus system (GSI) and (experimentally) authorization (GAA).
- Fault Detection: Basic support for building fault detection and recovery into Globus applications.
- Information Infrastructure: Global access to information about the state and configuration of system components of an application (MDS).
- Grid programming services: Support writing parallel-distributed programs (MPICH-G), monitoring (HBM), etc.

www.globus.org provides full information about the Globus system.

[2] *The Grid: Blueprint for a New Computing Infrastructure*, edited by Ian Foster and Carl Kesselman. Morgan Kaufmann, Pub. August 1998. ISBN 1-55860-475-8. http://www.mkp.com/books_catalog/1-55860-475-8.asp

[3] "Grids as Production Computing Environments: The Engineering Aspects of NASA's Information Power Grid," William E. Johnston, Dennis Gannon, and Bill Nitzberg. Eighth IEEE International Symposium on High Performance Distributed Computing, Aug. 3-6, 1999, Redondo Beach, California. (Available at http://www.nas.nasa.gov/~wej/IPG)

[4] "Vision and Strategy for a DOE Science Grid" - http://www.itg.lbl.gov/~wej/Grids

[5] See www.nas.nasa.gov/IPG for project information and pointers.

[6] See http://www-itg.lbl.gov/NGI/ for project information and pointers.

[7] The Particle Physics Data Grid has two long-term objectives. Firstly: the delivery of an infrastructure for very widely distributed analysis of particle physics data at multi-petabyte scales by hundreds to thousands of physicists. Secondly: the acceleration of the development of network and middleware infrastructure aimed broadly at data-intensive collaborative science. http://www.cacr.caltech.edu/ppdg/

[8] Tierney, B. Lee, J., Crowley, B., Holding, M., Hylton, J., Drake, F., "A Network-Aware Distributed Storage Cache for Data Intensive Environments", Proceeding of IEEE High Performance Distributed Computing conference (HPDC-8), August 1999.

[9] "Real-Time Generation and Cataloguing of Large Data-Objects in Widely Distributed Environments," W. Johnston, Jin G., C. Larsen, J. Lee, G. Hoo, M. Thompson, and B. Tierney (LBNL) and J. Terdiman (Kaiser Permanente Division of Research). Invited paper, International Journal of Digital Libraries - Special

Issue on “Digital Libraries in Medicine”. May, 1998. http://www-itg.lbl.gov/WALDO/

[10] MAGIC: “The MAGIC Gigabit Network.” See: http://www.magic.net

[11] TerraVision-2: VRML based data fusion and browsing - www.ai.sri.com/TerraVision

[12] “A Monitoring Sensor Management System for Grid Environments,” Brian Tierney, Brian Crowley, Dan Gunter, Mason Holding, Jason Lee, Mary Thompson. To appear, HPDC-9, July, 2000. Available at http://www-didc.lbl.gov/JAMM/

[13] A collaborative effort to enable desktop access to remote resources including, supercomputers, network of workstations, smart instruments, data resources, and more - computingportals.org

[14] “The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets.” A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, S. Tuecke, (to be published in the Journal of Network and Computer Applications).

[15] “Storage Access Coordination Using CORBA,” A. Sim, H. Nordberg, L.M. Bernardo, A. Shoshani and D. Rotem. Proceedings of the International Symposium on Distributed Objects and Applications. See http://gizmo.lbl.gov/sm/

[16] The Clipper Project: Computational Grids providing middleware that supports applications requiring configurable, distributed, high-performance computing and data resources. See http://www-itg.lbl.gov/~johnston/Clipper

[17] The Grid Forum (www.gridforum.org) is an informal consortium of institutions and individuals working on wide area computing and computational Grids. Current working groups include Security (authentication, authorization), Scheduling and Resource Management, Grid Information Services, Application and Tool Requirements, Advanced Programming Models, Grid User Services and Operations, Account Management, Remote Data Access, Grid Performance

[18] "New Capabilities in the HENP Grand Challenge Storage Access System and its Application at RHIC" http://rncus1.lbl.gov/GC/docs/chep292lp1.doc
"STACS is ... responsible for determining, for each query request, which events and files need to be accessed, to determine the order of files to be cached dynamically so as to maximize their sharing by queries, to request the caching of files from HPSS in tape optimized order, and to determine dynamically which files to keep in the disk cache to maximize file usage."

[19] "DeepView: A Collaborative Framework for Distributed Microscopy." IEEE Conf. on High Performance Computing and Networking, Nov. 1998. See http://vision.lbl.gov/ (projects -> collaborative computing)

[20] **Akenti**: "**Certificate-based Access Control for Widely Distributed Resources**," Mary Thompson, William Johnston, Srilekha Mudumbai, Gary Hoo, Keith Jackson, Usenix Security Symposium '99. Mar. 16, 1999. (See http://www-itg.lbl.gov/Akenti)

[21] GAA: "**Generic Authorization and Access control API**" (GAA API). IETF Draft. http://ghost.isi.edu/info/gss_api.html)

[22] Storage Resource Broker (SRB) provides uniform access mechanism to diverse and distributed data sources. http://www.sdsc.edu/MDAS/

[23] Condor is a High Throughput Computing environment that can manage very large collections of distributively owned workstations. http://www.cs.wisc.edu/condor/

[24] SCIRun is a scientific programming environment that allows the interactive construction, debugging and steering of large-scale scientific computations. http://www.cs.utah.edu/~sci/software/

[25] Ecce - www.emsl.pnl.gov

[26] WebFlow - A prototype visual graph based dataflow environment, WebFlow, uses the mesh of Java Web Servers as a control and coordination middleware, WebVM. See http://iwt.npac.syr.edu/projects/webflow/index.htm

[27] "QoS as Middleware: Bandwidth Reservation System Design." Gary Hoo and William Johnston, Lawrence Berkeley National Laboratory, Ian Foster and Alain Roy, Argonne National Laboratory and University of Chicago. To appear, Eighth IEEE International Symposium on High Performance Distributed Computing, Aug. 3-6, 1999, Redondo Beach, California. (See http://www-itg.lbl.gov/Clipper/QoS)