Grid Explorations

Marios D. Dikaiakos
University of Cyprus

http://www.cs.ucy.ac.cy/mdd

Talk Outline

- **From Parallel to Grid Computing**
- Grid Resource Characterization through Benchmarking
- Navigating and Searching the Grid Information Space
- Conclusions

Parallel computing

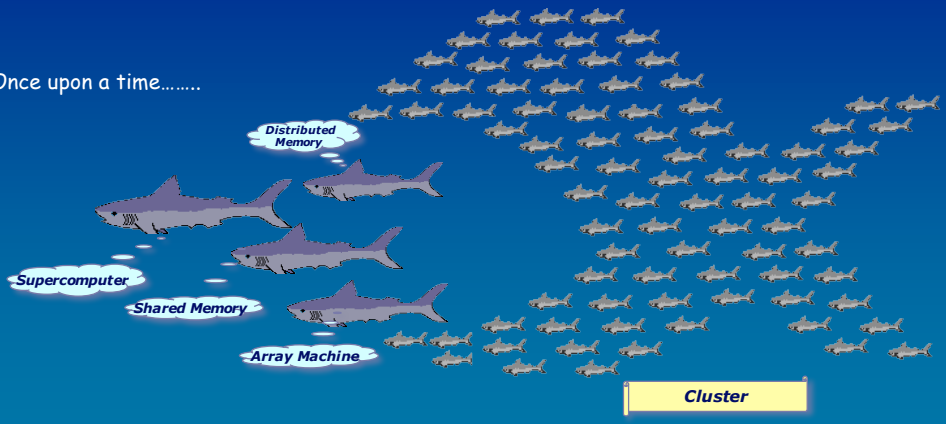
- Used to solve large problems, mainly from science/engineering.
- Different parallel machine models (SIMD, MIMD).
- A large variety of parallel hardware (Vector, Array Processors, Shared Memory and Distributed Memory Machines).
- Alternative computing models: data/task parallelism, SPMD.
- Lots of parallel programming approaches: implicit parallelism, shared-memory and/or message-passing semantics, high-level languages.
- Software tools: compilation, automatic parallelization, performance analysis & modeling, parallel debugging, application development.

Parallel computing: implicit assumptions

- "Tightly coupled."
- Homogeneous processing elements and interconnection.
- Communication through a "closed" interconnection network (bounded latency, high bandwidth).
- Parallel computer under a single administrative domain.
- User needs direct access to the parallel machine.

The Dead Supercomputer Society and the "killer micros"

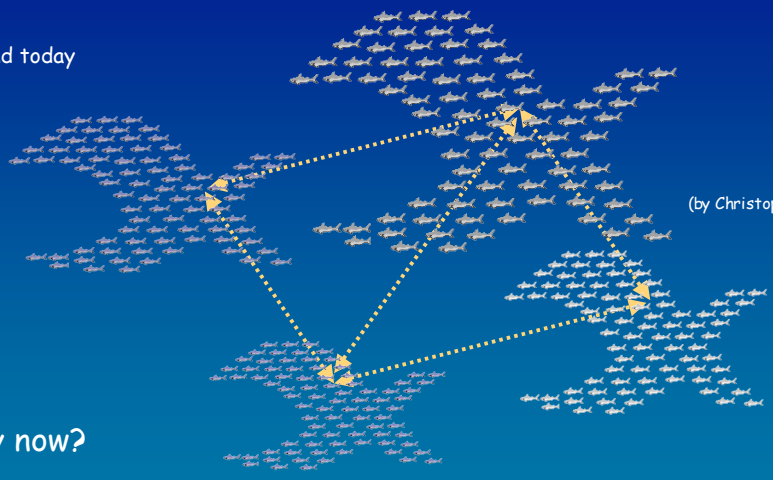
Once upon a time……..



(by Christophe Jacquet)

The GRID distributed computing idea

…and today



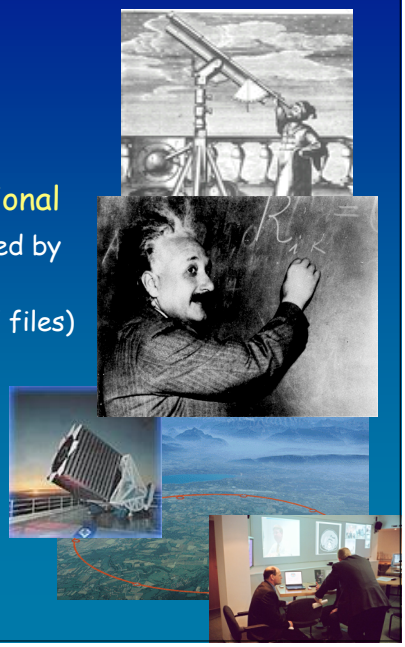
(by Christophe Jacquet)

Why now?

- Technological progress.
- Trends in Science.
- Internet Computing trends (e-Services, Overlays, P2P).

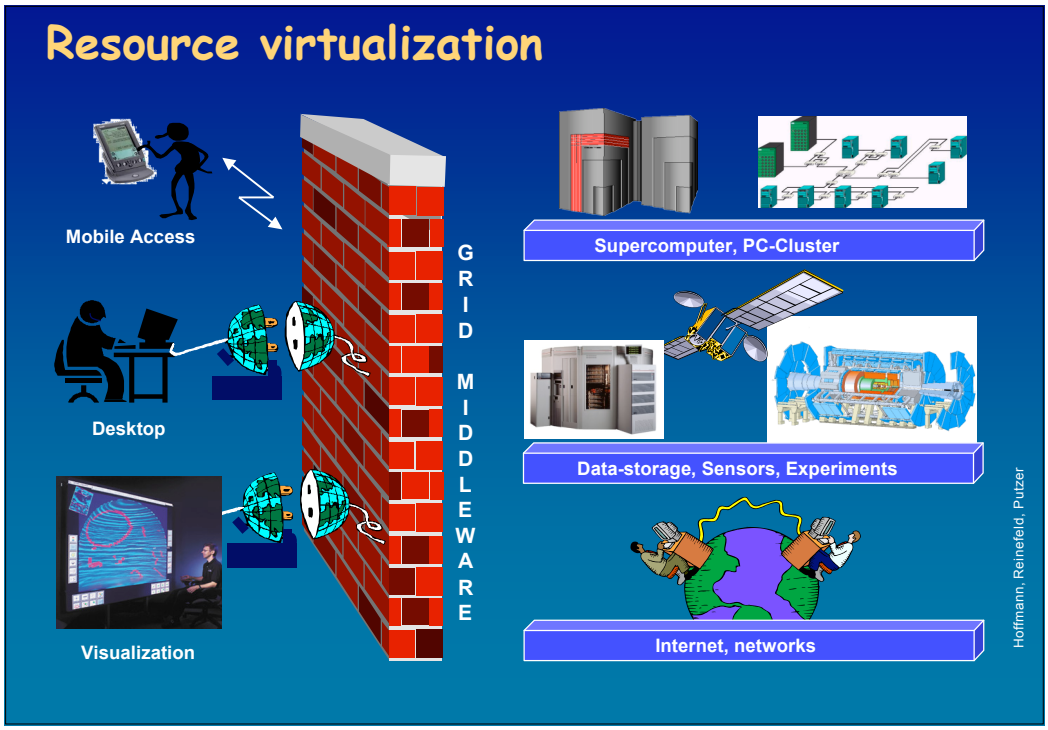
Why Now..? *From Science to e-Science*

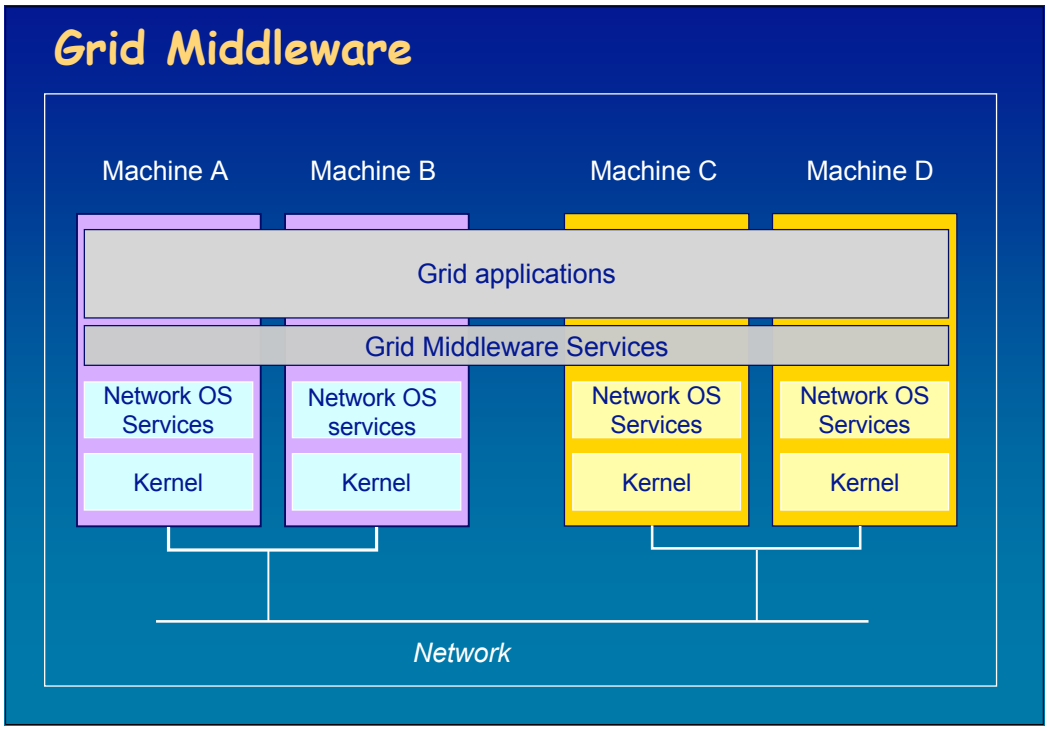
- In the past, science has been mainly empirical and theoretical.
- Recently, science is becoming computational (Data captured by instruments or data generated by simulators, processed by software, placed in databases / files. Scientists analyze database / files)
- Large Hadron Collider (LHC) at CERN:
  - 10 Petabytes/year of data.
  - ~100,000 of today's fastest PC processors.



The Grid

- Middleware infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals and institutions (Foster, Kesselman, Tuecke).
- Enables communities ("Virtual Organizations") to share geographically distributed resources as they pursue common goals --
- Key assumptions: ***absence of…***
  - Homogeneity
  - Central location
  - Central control
  - Existing trust relationships


Resource virtualization
Mobile Access
Desktop
Visualization
GRID MIDDLEWARE
Supercomputer, PC-Cluster
Data-storage, Sensors, Experiments
Internet, networks
Hoffmann, Reinefeld, Putzer

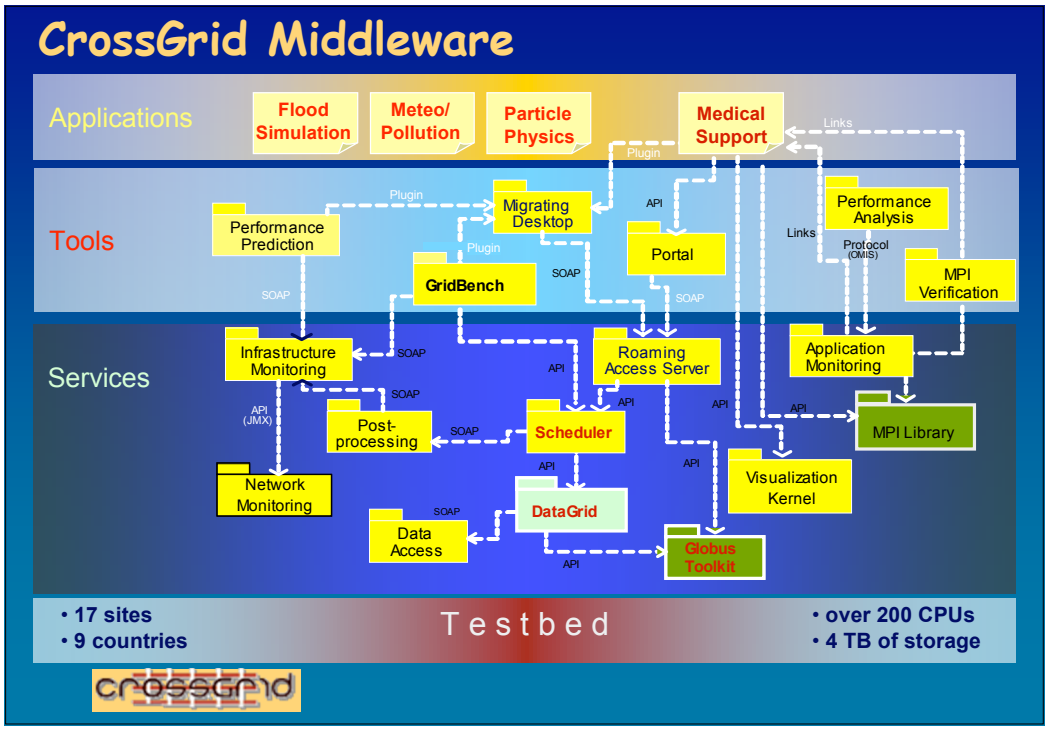

Grid Middleware
Machine A
Machine B
Machine C
Machine D
Grid applications
Grid Middleware Services
Network OS Services
Network OS services
Network OS Services
Network OS Services
Kernel
Kernel
Kernel
Kernel
Network

Key Components… and challenges

- Resource management: discovery, monitoring, control, access, data management..
- Job management: definition, submission, control, Web-based access..
- Administration: configuration & policy management, security, fault discovery, software packaging and distribution..
- Application development: programming tools, new programming paradigms, higher-level abstractions..
- Common protocols, standards, APIs, services.

Grid Middleware Systems

- GLOBUS Toolkit (ANL,ISI/USC)
  - GSI, GRAM, GridFTP, MDS, Grid Services.
- Condor (Wisc)
  - "Cycle-stealing," high-throughput computing, check-pointing and migration.
- UNICORE (R.Z. Juelich)
  - A "vertically integrated" Grid components supporting batch-job submission to distributed supercomputing sites.
- EDG, CrossGrid, LCG2, GridLab, gLite (EU-IST)
  - Toolkits using some components from several other projects, packaged and tested together.

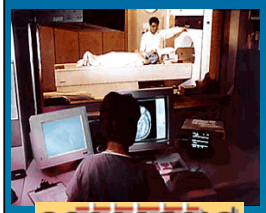


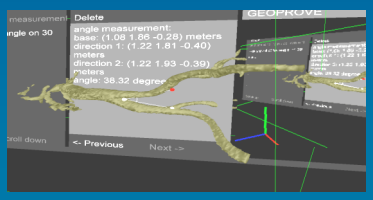
Talk Outline

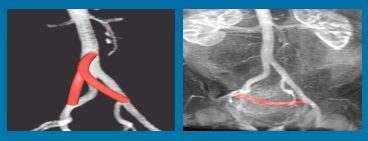
- From Parallel to Grid Computing
- **Grid Resource Characterization through Benchmarking**
- Navigating and Searching the Grid Information Space
- Conclusions

Vascular surgery planning: a Grid application

- Goal: support bypass surgical treatment of vascular diseases by simulating vascular reconstruction.
  - Examine alternative scenarios for bypass placement.
  - Improvements in blood flow determine which bypass is best.
  - 3D arterial model presented to the surgeon in an immersive environment.
  - Parallel MPI code computes blood flow properties.



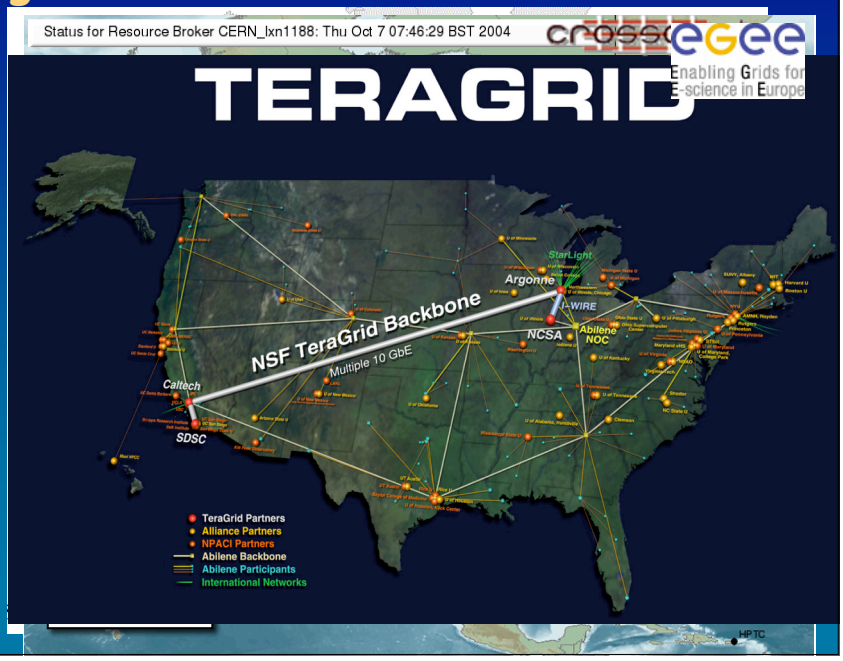


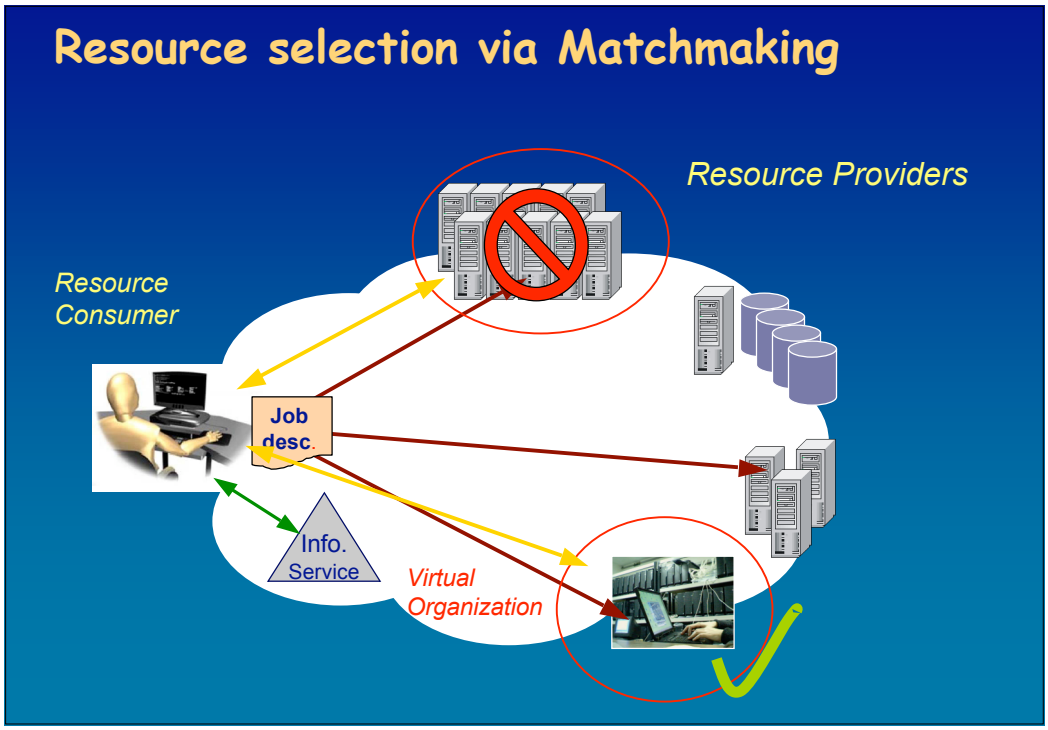


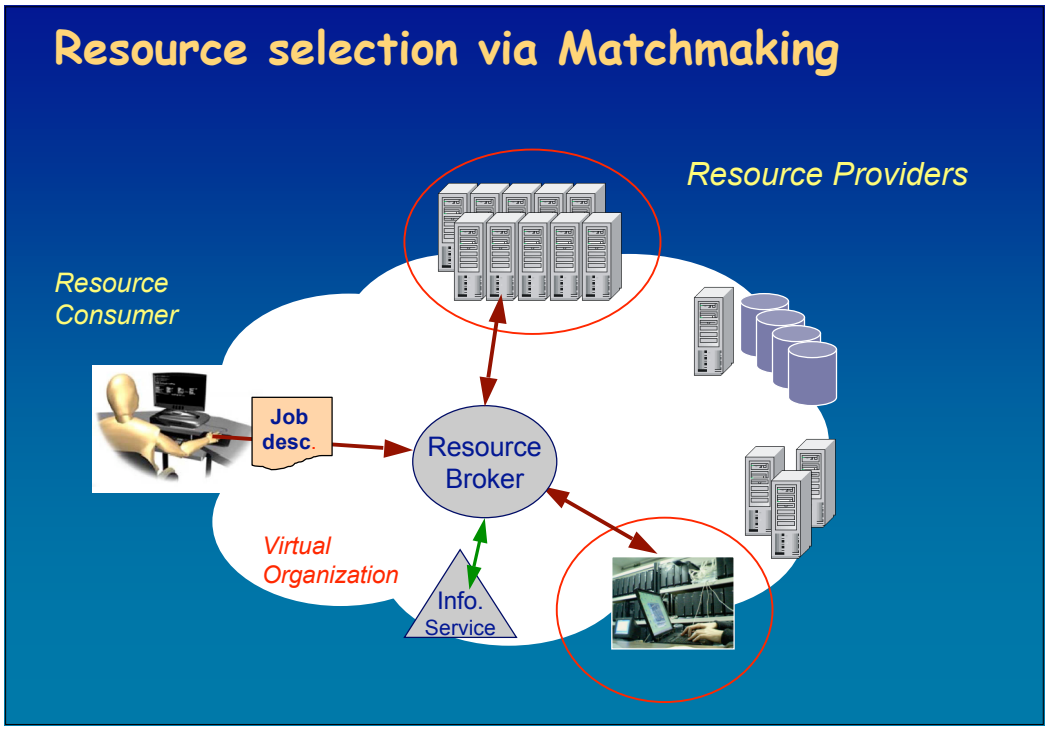
crossGrid

Source: P. Sloot, Univ. of Amsterdam & CrossGrid project

Choosing Grid resources




Resource selection via Matchmaking
Resource Providers
Resource Consumer
Job desc.
Info. Service
Virtual Organization


Resource selection via Matchmaking
Resource Providers
Resource Consumer
Job desc.
Resource Broker
Virtual Organization
Info. Service

Ask a Grid Information Service!

dn: GlueSubClusterUniqueID=cgce.ifca.org.es,

dn: GlueCEUniqueID=cgce.ifca.org.es:2119/jobmanager-pbs-short,
Mds-Vo-name=ifcapro,mds-vo-name=local,o=grid

```
GlueChunkKey: GlueClusterUniqueID=cgce.ifca.org.es
GlueHostApplicationSoftwareRunTimeEnvironment: CG2_0_4
GlueCEUniqueID: cgce.ifca.org.es:2119/jobmanager-pbs-short
GlueCEInfoGatekeeperPort: 2119
GlueCEInfoHostName: cgce.ifca.org.es
GlueCEInfoLRMSType: pbs
GlueCEInfoLRMSVersion: OpenPBS_2.4
GlueCEInfoTotalCPUs: 20
GlueCEStateEstimatedResponseTime: 0
GlueCEStateFreeCPUs: 20
GlueCEStateRunningJobs: 0
GlueHostMainMemoryVirtualSize: 1144
```

GlueForeignKey: GlueClusterUniqueID=cgce.ifca.org.es

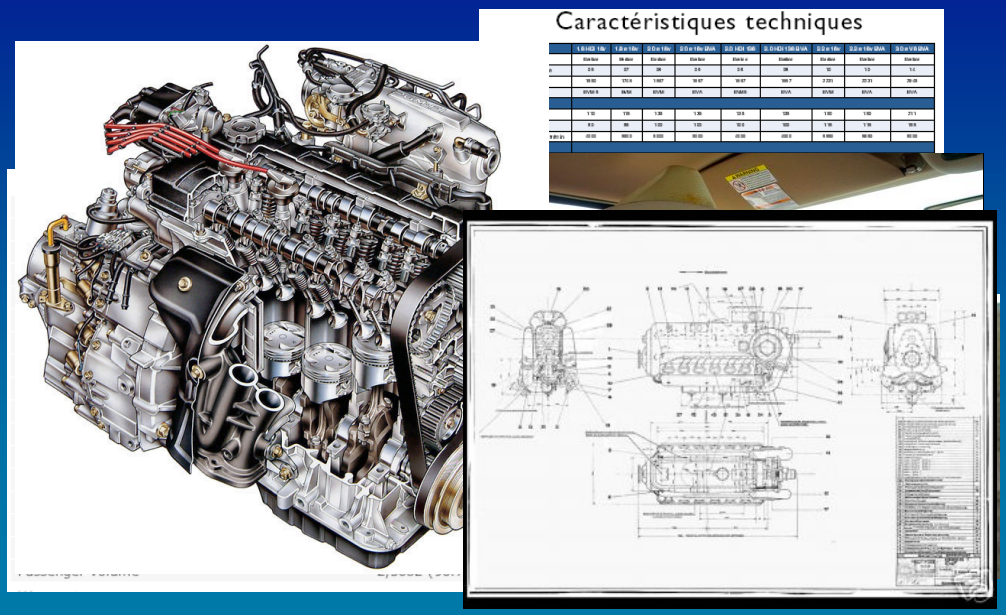
Static Info

Dynamic Info

Motivation and Focus

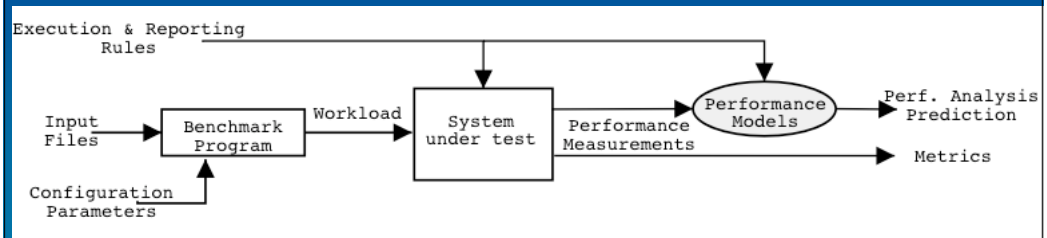
- How can we characterize the performance of Grid resources?
  - Support more advanced criteria for matchmaking: performance, functionality, reliability, robustness, cost…
  - Drive the design and configuration of Grid infrastructures.
  - Open marketplaces based on performance negotiation.
  - Developing models for performance prediction.
- "We have no real idea how the Grid and Grid applications could be characterized from the point of view of performance" *(APART Working Group on Automatic Performance Analysis, Rackeve Workshop, 11/2003)*
- Do this in an end-to-end fashion.

Choosing an automobile



Test-driving computers…

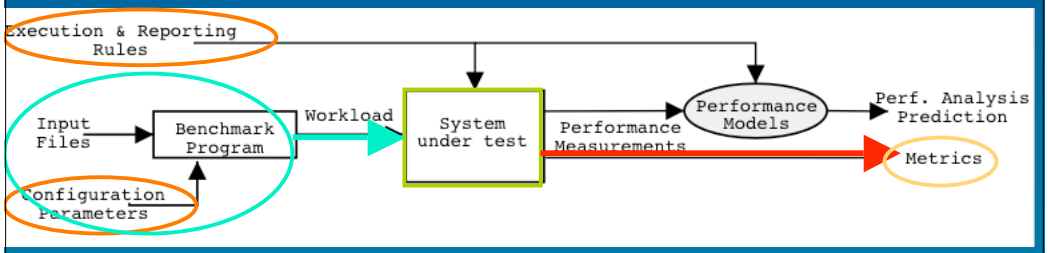
- **Benchmarks:** standardized programs designed or chosen to investigate performance properties of computer systems:
  - System characterization
  - Fair comparison
  - System design assessment
  - Supporting quantitative research



- Conditions of execution & measurement: well-defined and publicly available.
- Benchmarks are required to be: portable, fair, relevant, easy to measure, easy to explain.

Key challenges for benchmarking Grids

- Measuring a fuzzy target
- Trusting our measurements
- Describing performance (metrics)
- Managing cost
- In search of relevance



GridBench

- The GridBench Framework:
  - A software tool for characterizing the performance of Grids and Grid resources quantitatively, using benchmarks.
  - Supports the administration of Grid benchmarks, the archival, publication, browsing and analysis of metrics.
- GridBench Suite of Benchmarks:
  - A layered suite of benchmarks deployed on a Grid testbed.
  - Geared towards high-performance and high-throughput computing needs.

Elements of the GridBench Framework

- An "ontology" of performance metrics.
- GBDL, a platform-independent XML language:
  - Specification of configuration.
  - Representation of results.
- A translator from GBDL to different Job Description Languages:
  - RSL (Globus)
  - JDL (Condor/DataGrid/CrossGrid/EGEE).
- A "user-friendly" administration tool and GUI.
- A database for archiving metrics.
- A performance analysis tool and GUI.

GridBench software

- Implemented in JAVA + MySQL + Linux (ports to IRIX, AIX)
- Open Source Benchmarks in C/C++/Fortran + MPI
- Available for downloading: http://grid.ucy.ac.cy/GridBench
- Part of CrossGrid Middleware distribution
- Adopted and in use by Grid Ireland
- Working towards demonstrating it on EGEE

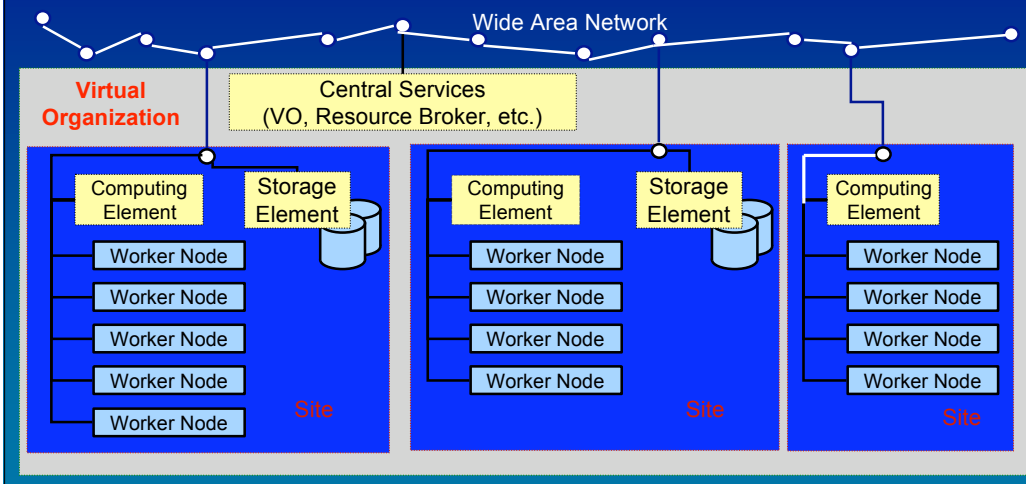
Talk Outline

- From Parallel to Grid Computing
- **Grid Resource Characterization through Benchmarking**
  - **Metrics and Benchmarks.**
  - Filtering polluted measurements.
  - GridBench GUI and use.
- Navigating Grids
- Grid Search Engines
- Conclusions & Future Work

An ontology of performance metrics

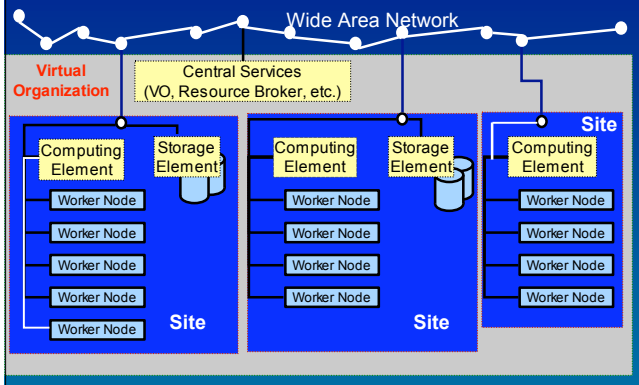
- Performance capacity of Grid infrastructure can be described by:
  - The performance of a hierarchical collection of measurable entities (CPUs, memory performance, computers, clusters, collections of clusters..)
- Thus:
  - Small sets of metrics not adequate for Grids.
  - Definition, organization, storage, and interpretation requires advanced, open data models, amenable to post-processing (statistical, data mining, AI).
- ...what about interpretation of metrics?

A simple reference model



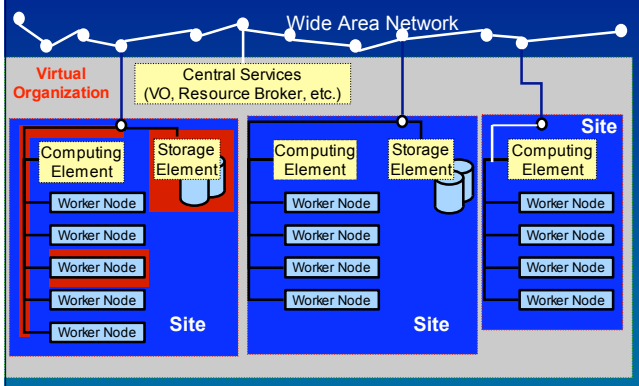
- ✓ Inspired by the DataGrid/CrossGrid/LCG architecture (Globus 2-based)
- ✓ Represented by the GLUE Schema.
- ✓ Necessary for defining benchmarking-targets and interpreting metrics.

Capturing metrics



- Performance measurements at the different levels of the Grid architecture.

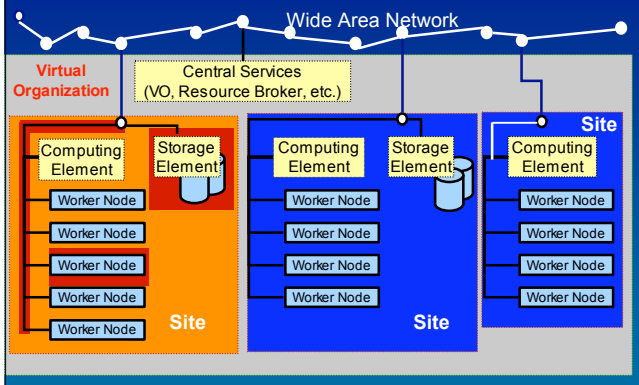
Capturing metrics



Individual Resources
(cluster nodes, mass storage)

- Performance measurements at the different levels of the Grid architecture.

Capturing metrics

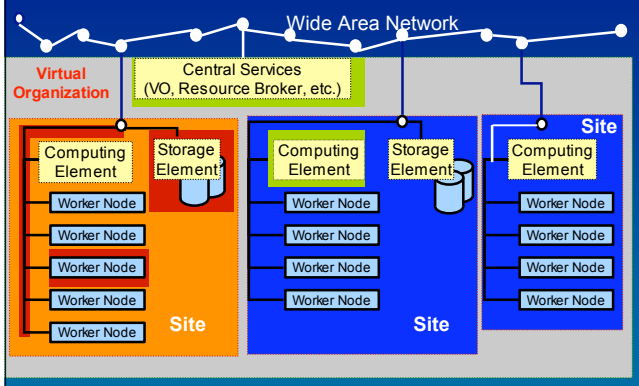


Individual Resources
(cluster nodes, mass storage)

Sites
(clusters, SMPs)

- Performance measurements at the different levels of the Grid architecture.

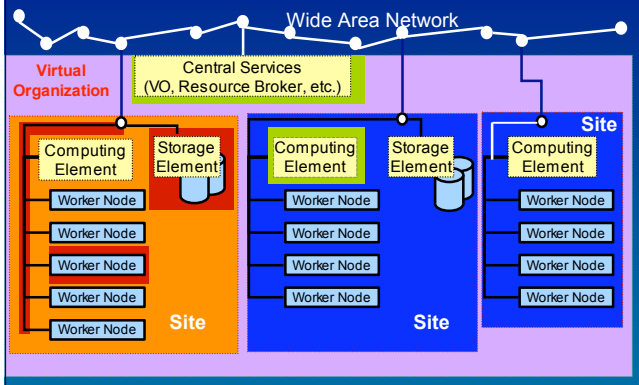
Capturing metrics



Individual Resources
(cluster nodes, mass storage)
Sites
(clusters, SMPs)
Middleware
(middleware layer providing access to shared resources)

- Performance measurements at the different levels of the Grid architecture.

Capturing metrics

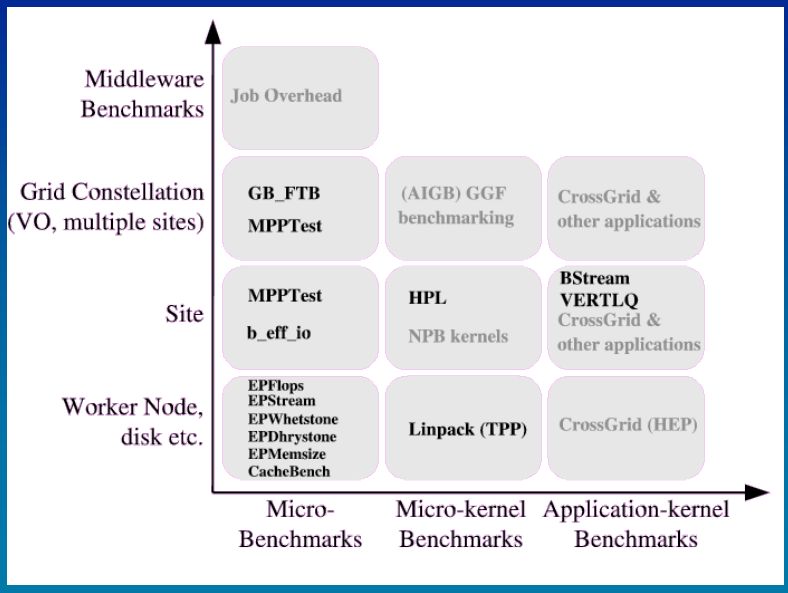


Individual Resources
(cluster nodes, mass storage)
Sites
(clusters, SMPs)
Middleware
(middleware layer providing access to shared resources)

Virtual Organization
(multiple sites, VO)

- Performance measurements at the different levels of the Grid architecture.

The GridBench suite of benchmarks



Outline

- From Parallel to Grid Computing
- **Grid Resource Characterization through Benchmarking**
  - Metrics and Benchmarks.
  - **Filtering polluted measurements.**
  - GridBench GUI and use.
- Navigating and Searching the Grid Information Space
- Conclusions & Future Work

Identifying polluted measurements

- Often benchmark runs in co-location with other jobs:
  - "Fellow passengers:" co-allocated by the Resource Broker.
  - "Free-riders:" unauthorized users, unknown to the VO.
  - "Runaways:" O/S processes, zombies, etc.
- These jobs "pollute" measurements and can affect seriously the characterization accuracy.
- Their effect on metric accuracy can be:
  - Identified through monitoring data.
  - Reduced through access control (for fellow passengers) and remote healing (for runaways).

Filtering polluted measurements

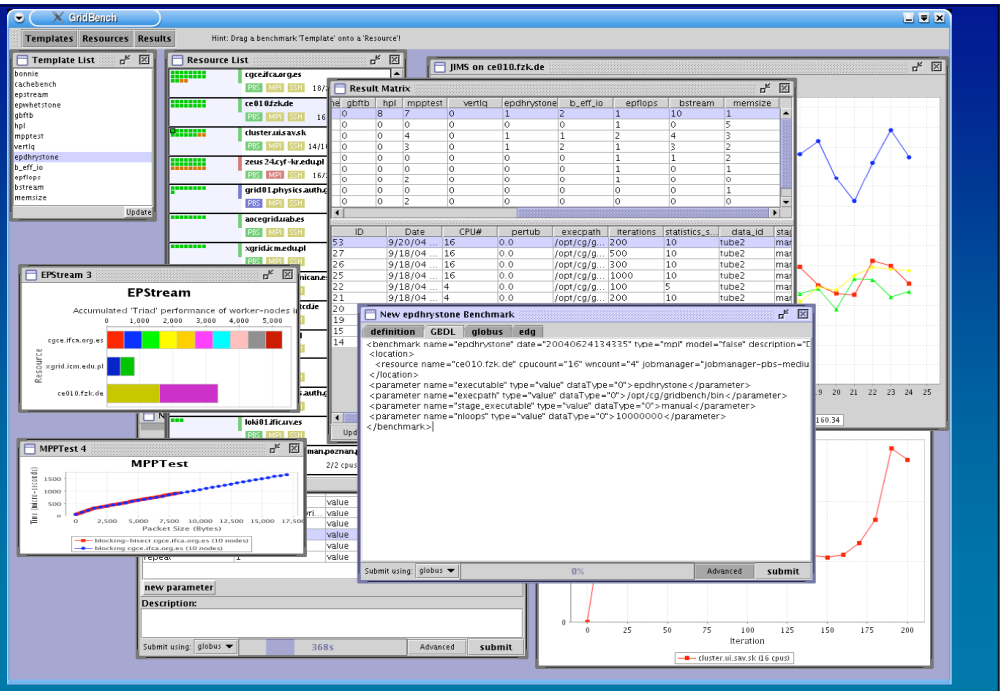
- GridBench retrieves monitoring information from Grid monitoring services, for a time-window encapsulating a benchmarking experiment.
- This is specified in the GBDL document describing a particular experiment:

```
<component name="data-transfer" ID="xfer01">...</component>
<monitor type="RGMA" source="ccwp71.in2p3.fr:3306"
  query="select * from NetworkTCPThroughput
         where NMIdSource='adc0003.cern.ch'
           and NMIdDestination='ccwp7.in2p3.fr'
  <parameter name="begin">comp-begin="xfer01"</parameter>
  <parameter name="end">comp-end="xfer01"</parameter>
</monitor>
```

- Retrieval from monitoring services is conducted via monitoring-client plug-ins called by GridBench.

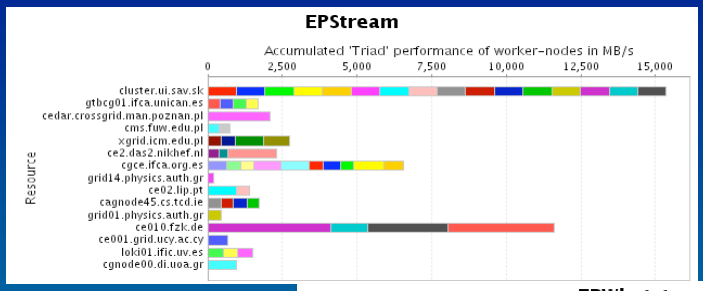
Outline

- From Parallel to Grid Computing
- **Grid Resource Characterization through Benchmarking**
  - Metrics and Benchmarks.
  - Filtering polluted measurements.
  - **GridBench GUI and use.**
- Navigating and Searching the Grid Information Space
- Conclusions & Future Work

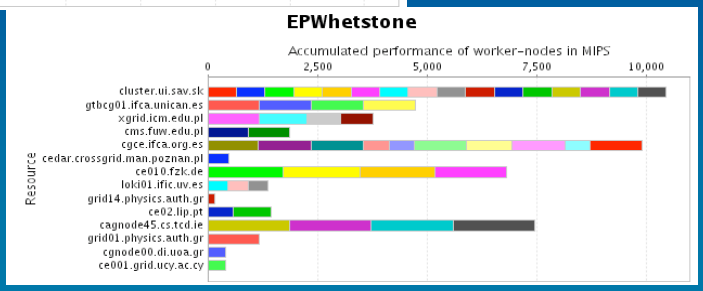


http://grid.ucy.ac.cy/GridBench

Sample results

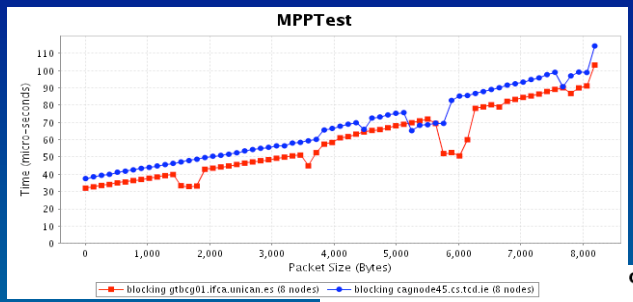


Memory Bandwidth

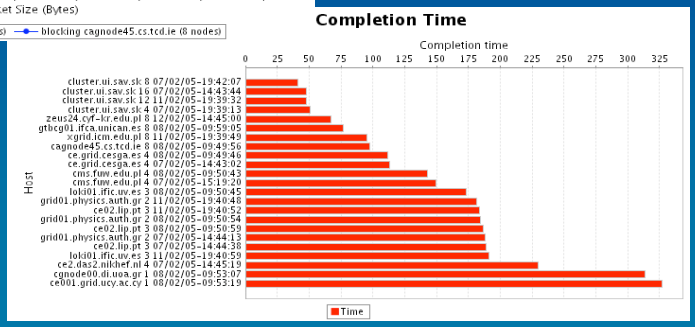


CPU Performance

Sample results



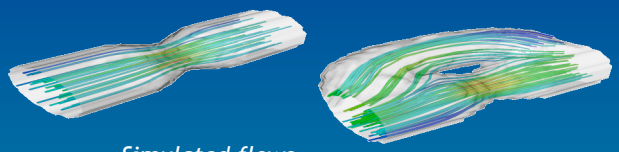
MPI Performance



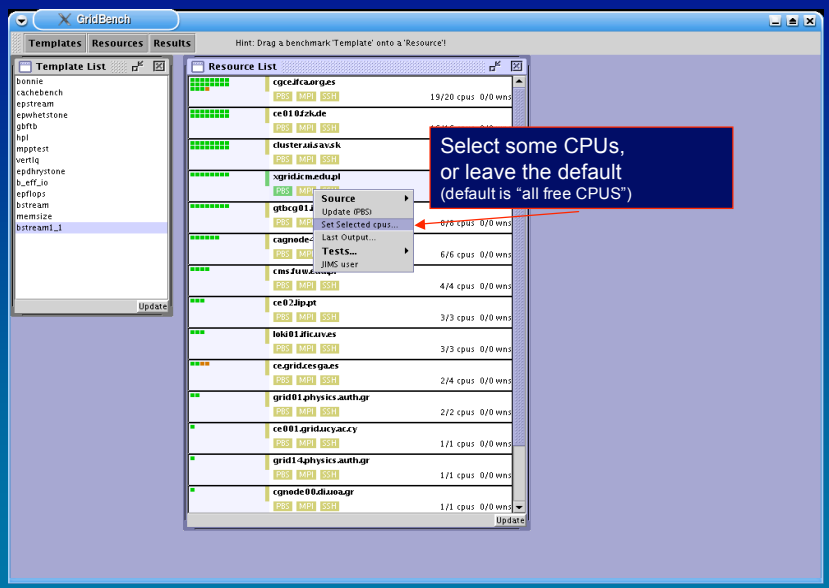
Application-kernel Performance

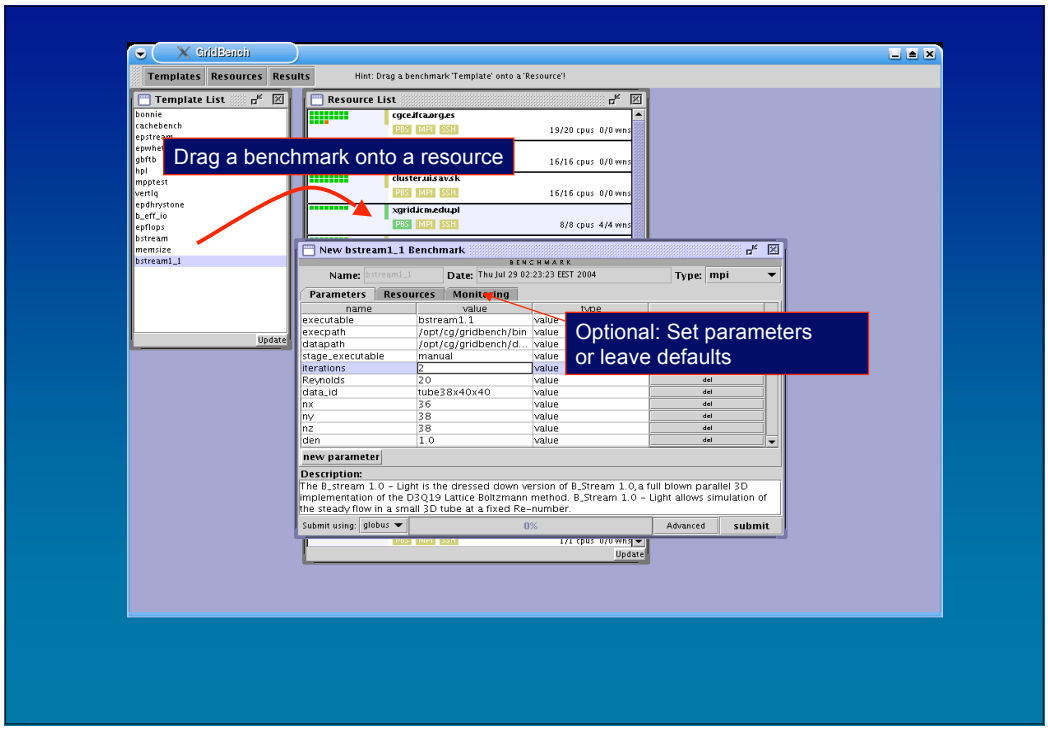
Resource Selection with GridBench

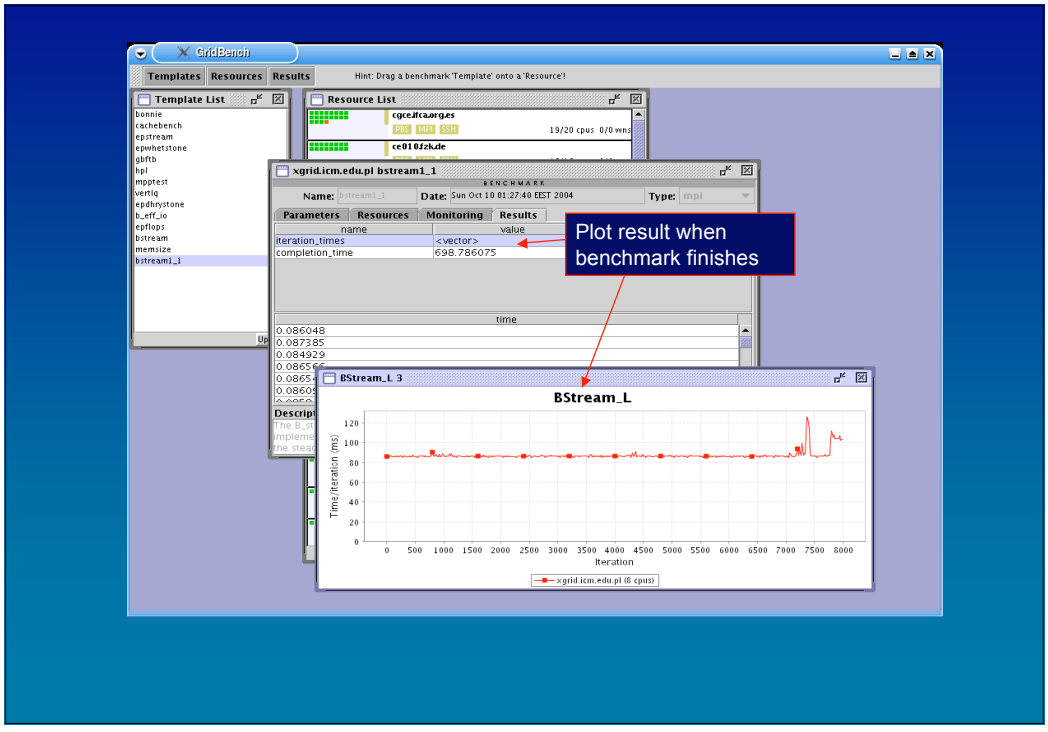
- Where should I run my fluid-flow code for Surgical Planning?

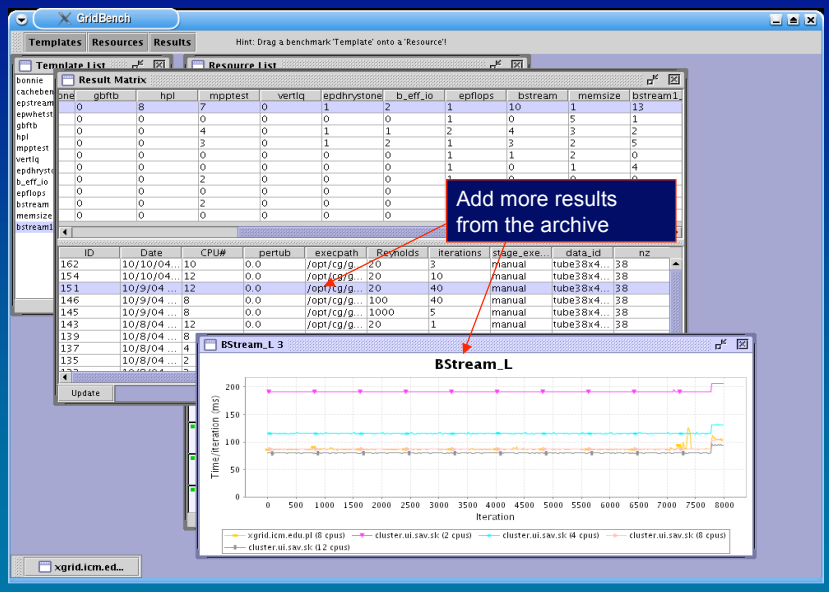


Simulated flows




GridBench
Templates
Resources
Results
Hint: Drag a benchmark 'Template' onto a 'Resource'!
Template List
Resource List
Drag a benchmark onto a resource
New bstream1_1 Benchmark
Name:
Date:
Type:
mpi
Parameters
Resources
Monitoring
Optional: Set parameters or leave defaults
Description:
Submit using:
globus
Advanced
submit


GridBench
Templates
Resources
Results
Hint: Drag a benchmark 'Template' onto a 'Resource'!
Template List
Resource List
xgrid.icm.edu.pl bstream1_1
Name:
Date:
Type:
Parameters
Resources
Monitoring
Results
Plot result when benchmark finishes
iteration_times
<vector>
completion_time
698.786075
time
BStream_L 3
BStream_L
Time/iteration (ms)
Iteration



Remarks

- Virtualization and resource heterogeneity turn Grid Benchmarking into a:
  - Challenging and expensive process.
  - Necessary undertaking for performance-based decisions.
- Isolated metrics are of little use. We need instead ontologies of metrics.
- Virtualization and the lack of central control, put the accuracy of benchmarking measurements to question:
  - Need a combination of metrics and monitoring information to filter-out invalid measurements.

Outline

- From Parallel to Grid Computing
- Grid Resource Characterization through Benchmarking
- **Navigating and Searching the Grid Information Space**
- Conclusions & Future Work

A Scenario for the Grid's future

- The Grid as a Wide-Scale Distributed System:
  - Millions of resources of different kinds.
  - Services and Policies in place.
  - Relationships (permanent and transient) between organizations, software, data, services, applications…
  - Different middleware platforms.
  - Common (?) protocols, standards and API's.
- The hope is that Grid will grow larger and will reach an acceptance as wide as the Web.

The Grid information problem

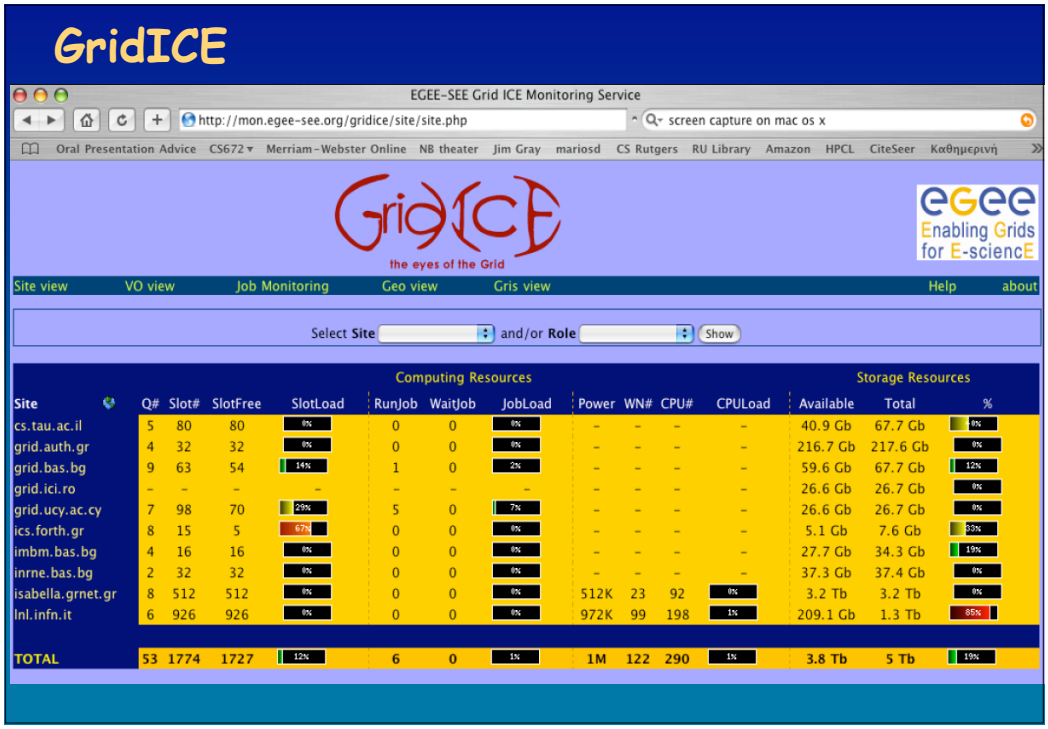
- How are individuals and organizations going to harness the capabilities of a fully deployed Grid:
  - Massive and ever-expanding base of resources.
  - Huge corpus of available programs, services, and data.
- Users need tools to discover and identify information about resources that are:
  - *Interesting* (discovery)
  - *Relevant* (classification)
  - *Accessible* and *available* under known *policies of use, cost* (inquiry)

Globus MDS

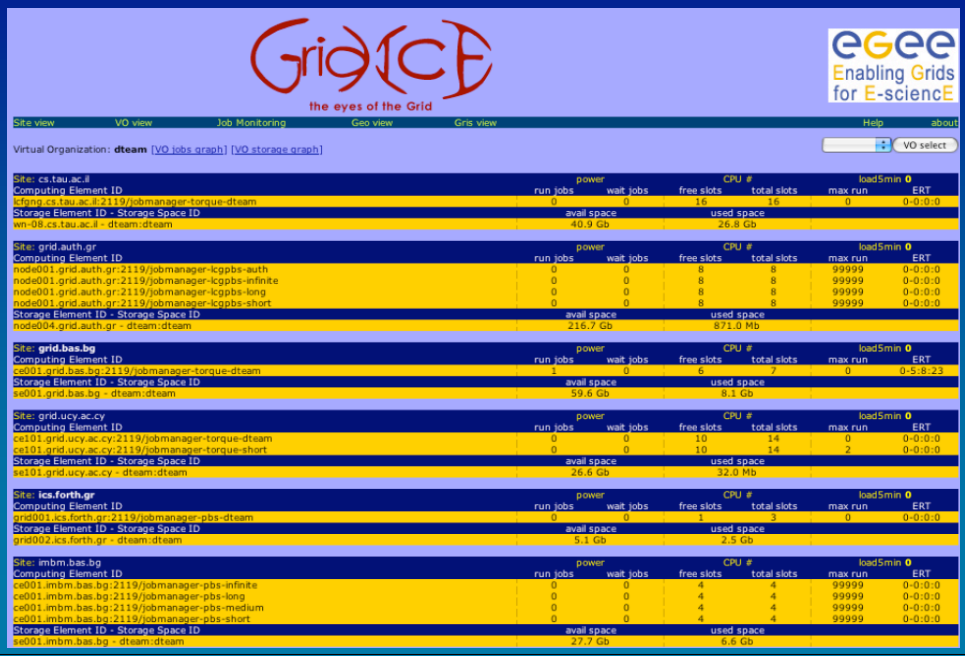
```
dn: GlueSubClusterUniqueID=cgce.ifca.org.es,
       GlueClusterUniqueID=cgce.ifca.org.es,
       Mds-Vo-name=ifcapro,mds-vo-name=local,o=grid
objectClass: GlueClusterTop
objectClass: GlueSubCluster
objectClass: GlueSchemaVersion
objectClass: GlueInformationService
objectClass: GlueKey
GlueSchemaVersionMajor: 1
GlueSchemaVersionMinor: 1
GlueChunkKey: GlueClusterUniqueID=cgce.ifca.org.es
GlueHostApplicationSoftwareRunTimeEnvironment: CG2_0_4
GlueHostApplicationSoftwareRunTimeEnvironment: CROSSGRID
GlueHostApplicationSoftwareRunTimeEnvironment: LCG-2
GlueHostApplicationSoftwareRunTimeEnvironment: MPICH
GlueHostApplicationSoftwareRunTimeEnvironment: MPICH
GlueHostApplicationSoftwareRunTimeEnvironment: MPICH-G2
GlueHostArchitectureSMPSize: 2
GlueHostBenchmarkSF00: 328
GlueHostBenchmarkSI00: 409
GlueHostMainMemoryRAMSize: 627
GlueHostMainMemoryVirtualSize: 1144
GlueHostNetworkAdapterInboundIP: FALSE
GlueHostNetworkAdapterOutboundIP: TRUE
GlueHostOperatingSystemName: Redhat
GlueHostOperatingSystemRelease: 2.4.20-30.7.legacysmp
GlueHostOperatingSystemVersion: 1 SMP Fri Feb 20 10:12:55 2004
GlueHostProcessorClockSpeed: 1261
GlueHostProcessorModel: Intel(R) Pentium(R) III family 1266MHz
GlueHostProcessorVendor: GenuineIntel
GlueSubClusterName: cgce.ifca.org.es
GlueSubClusterUniqueID: cgce.ifca.org.es
```

```
dn: GlueCEUniqueID=cgce.ifca.org.es:2119/jobmanager-pbs-short,
                Mds-Vo-name=ifcapro,mds-vo-name=local,o=grid
    objectClass: GlueCETop
    objectClass: GlueCE
    objectClass: GlueSchemaVersion
    objectClass: GlueCEAccessControlBase
    objectClass: GlueCEInfo
    objectClass: GlueCEPolicy
    objectClass: GlueCEState
    objectClass: GlueInformationService
    objectClass: GlueKey
    GlueSchemaVersionMajor: 1
    GlueSchemaVersionMinor: 1
    GlueCEName: short
    GlueCEUniqueID: cgce.ifca.org.es:2119/jobmanager-pbs-short
    GlueCEInfoGatekeeperPort: 2119
    GlueCEInfoHostName: cgce.ifca.org.es
    GlueCEInfoLRMSType: pbs
    GlueCEInfoLRMSVersion: OpenPBS_2.4
    GlueCEInfoTotalCPUs: 20
    GlueCEStateEstimatedResponseTime: 0
    GlueCEStateFreeCPUs: 20
    GlueCEStateRunningJobs: 0
    GlueCEStateStatus: Production
    GlueCEStateTotalJobs: 0
    GlueCEStateWaitingJobs: 0
    GlueCEStateWorstResponseTime: 0
    GlueCEPolicyMaxCPUTime: 900
    GlueCEPolicyMaxRunningJobs: 2
    GlueCEPolicyMaxTotalJobs: 4
    GlueCEPolicyMaxWallClockTime: 7200
    GlueCEPolicyPriority: 1
    GlueCEAccessControlBaseRule: VO:cg
    GlueForeignKey: GlueClusterUniqueID=cgce.ifca.org.es
```

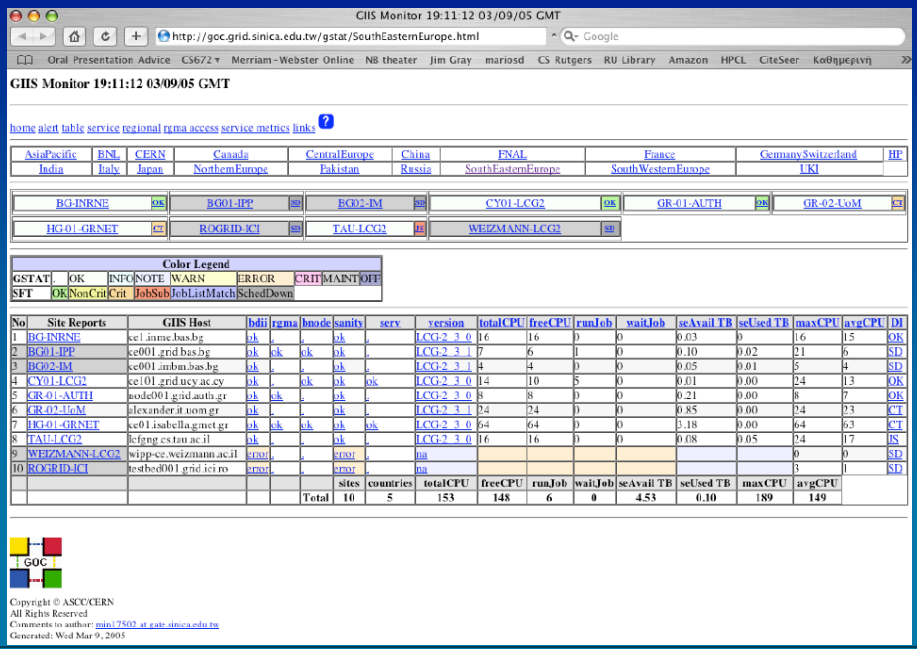
GridICE



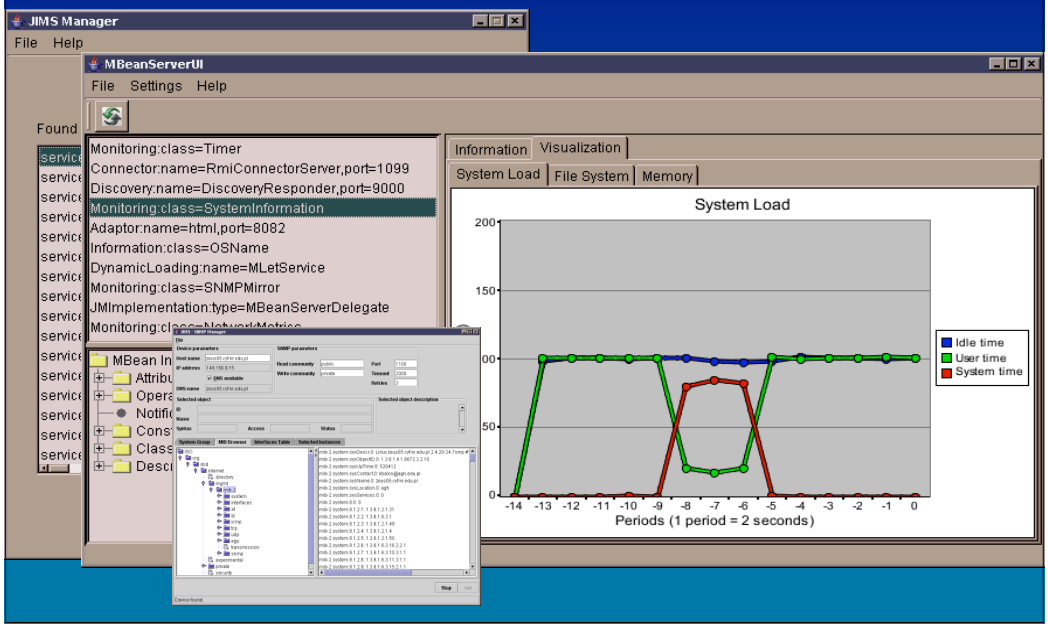
GridICE



MapCenter

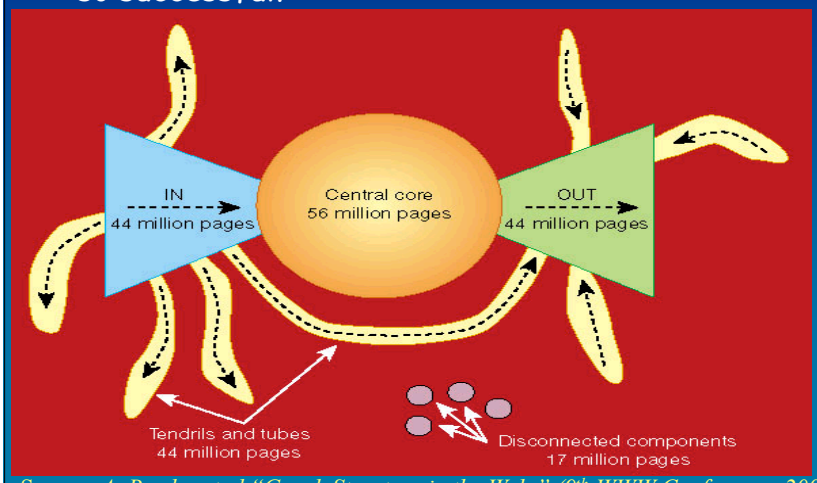


JIMS



Simple, yet powerful..

- Lessons learned from the Web or… why the Web has been so successful?



Source: *A. Broder et al "Graph Structure in the Web," (9th WWW Conference, 2000)*

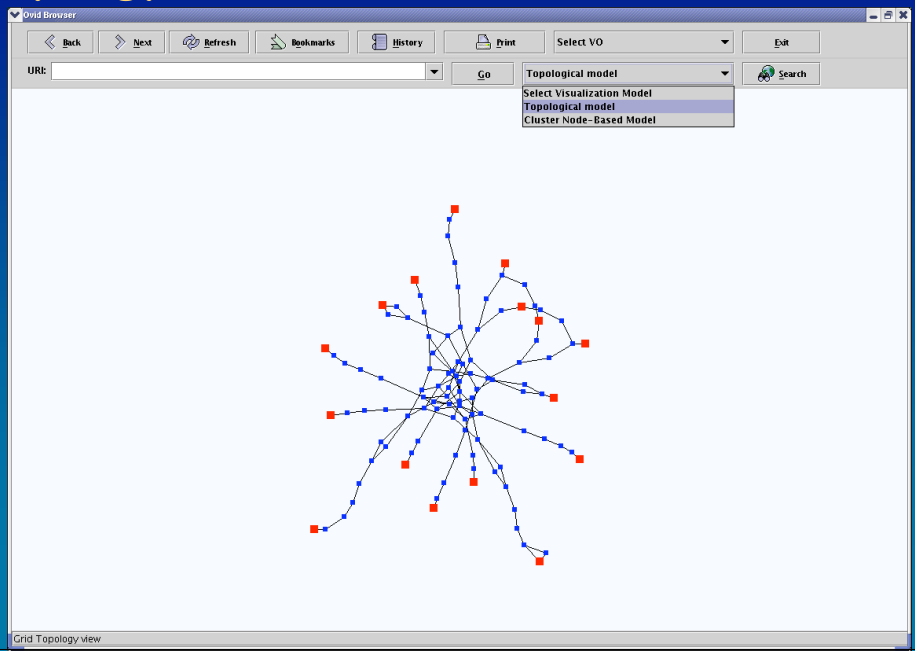
Navigating and Searching the Grid

- Key differences from the Web
  - Which are the information containers?
  - Hyperlinks?..
  - Information Retrieval makes no good..
- Navigation: the Ovid browser
  - Supporting the seamless navigation of users in the Grid information space.
- Towards a search engine for the Grid:
  - What is the information space we would like to search in?

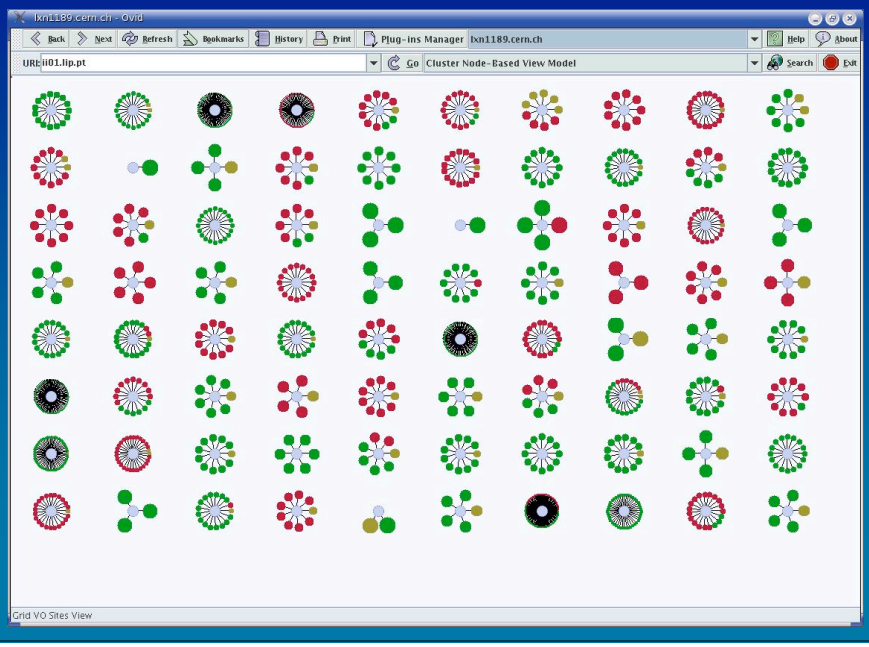
Ovid: A (universal) browser for Grids

- Navigational primitives designed to cope with network disorientation and information overloading;
- A small set of core graphical views, i.e. visual abstractions of Grid information;
- Support for embedding and implementing hyperlinks connecting related entities represented within different information views;
- A plug-in mechanism, for the seamless integration with Ovid of third-party monitoring clients;
- A modular software design (model-view-controller architecture), for the easy integration of different visualization algorithms.

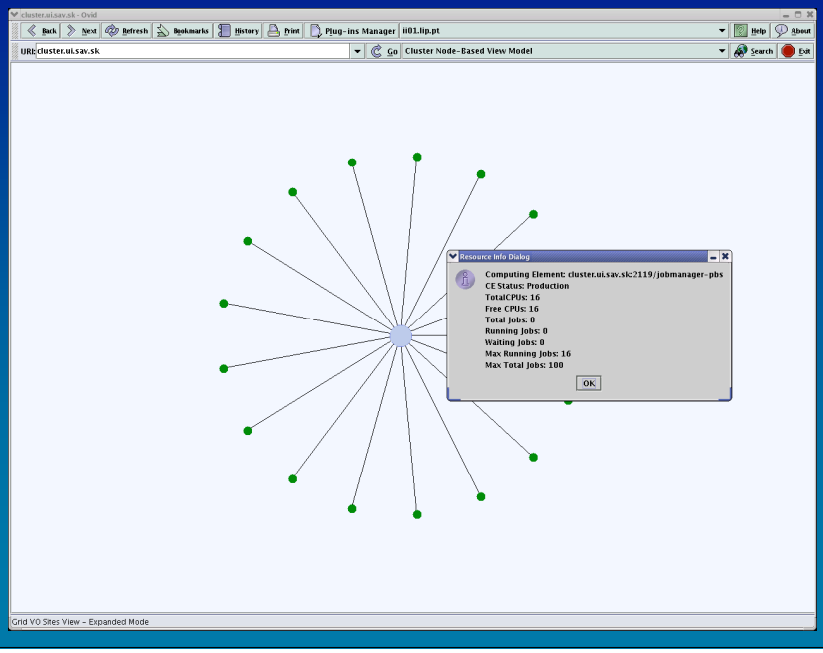
Topology of the CrossGrid VO



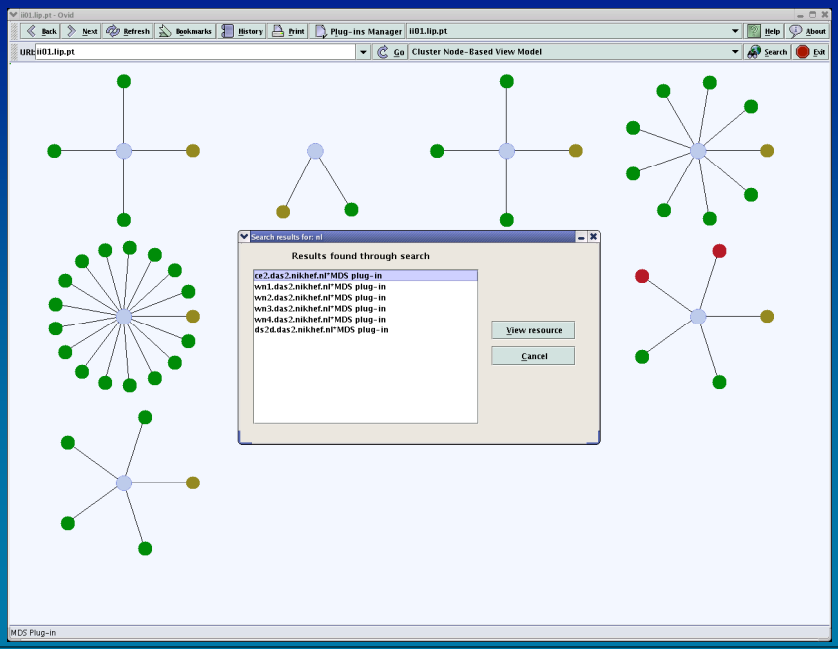
Ovid: EGEE VO



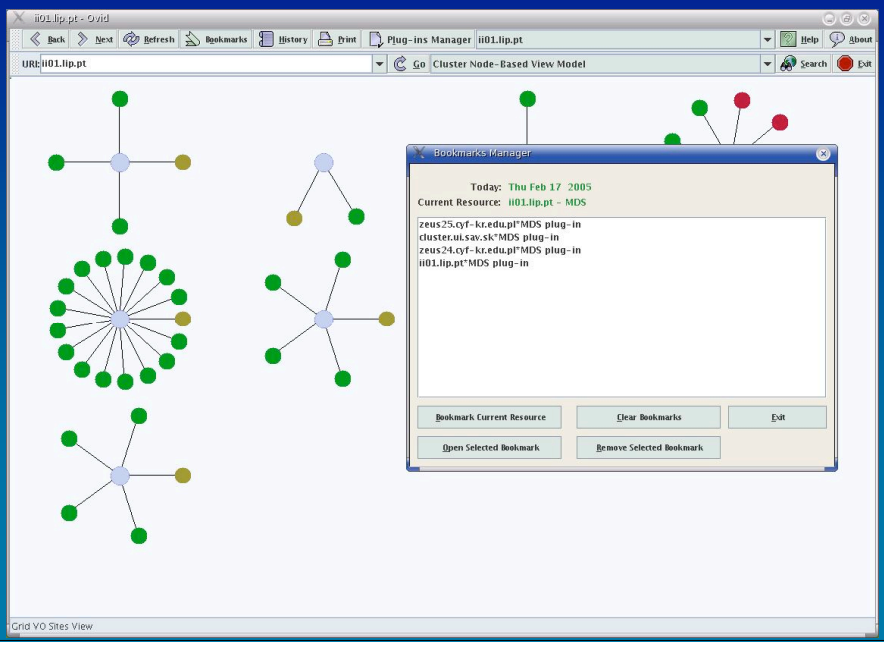
Ovid: Querying a Grid node



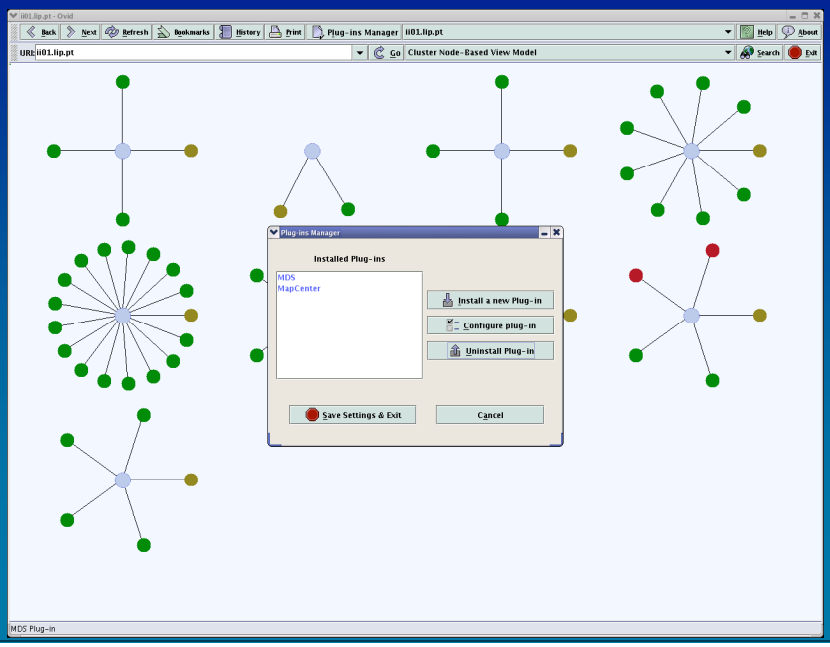
Ovid's Search interface



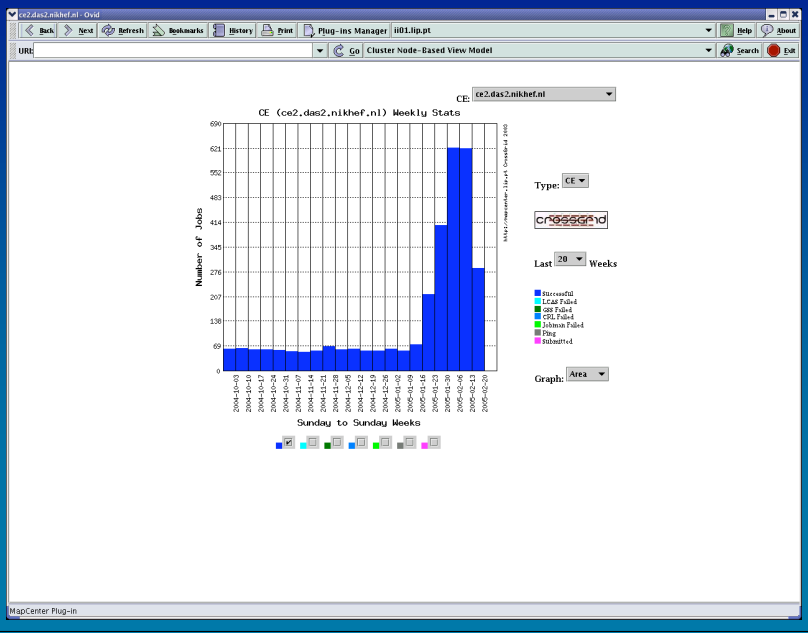
Ovid: Bookmarks manager



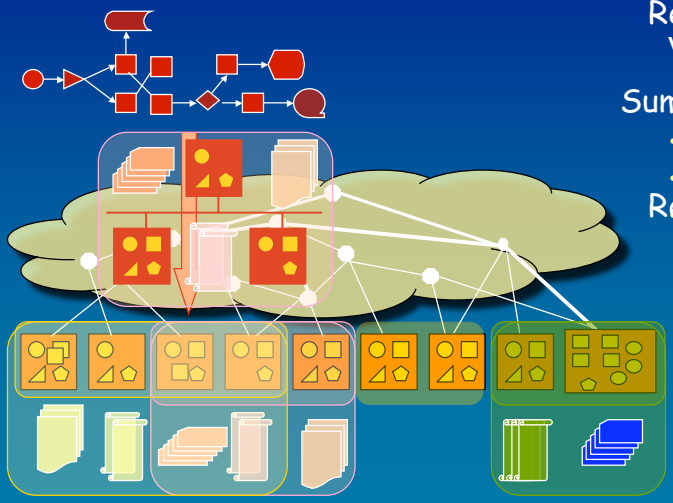
Ovid: Plug-in mechanism



MapCenter's plug-in



What information we want to search for?



Resource Specifications:

Virtual Organizations:

Software:

Summary & Statistics

- Policies
- Descriptions.
- Logs
- Associations

Resource Status:

Data-sets:

- Resources
- Availability.
- Monitoring data.
- Data
- Metadata
- Replicas

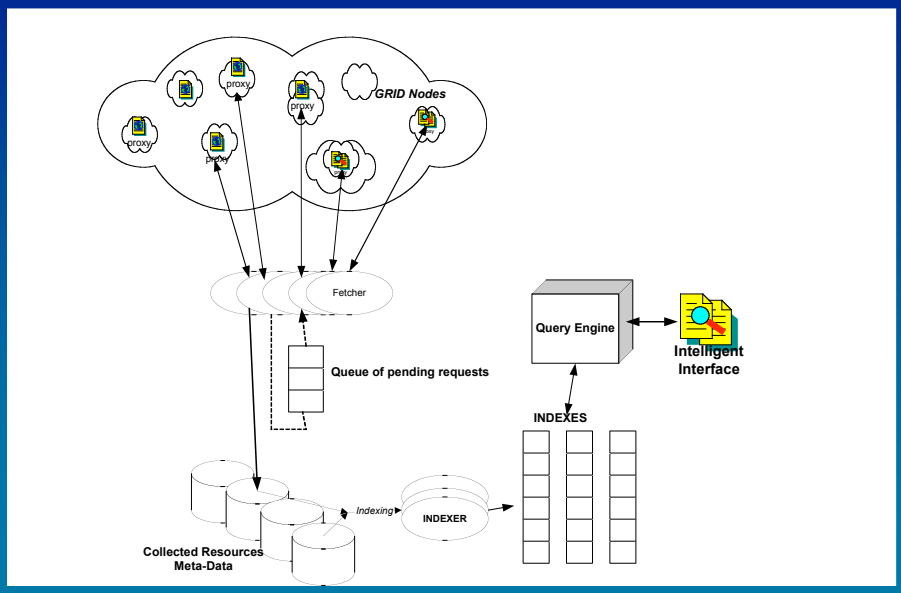
Services:

- Interface
- Metadata

Towards a Grid Search Engine (GRISEN)

- Based on the notion of "grid entity," which represents various (permanent or transient) resources on the Grid: computational, storage, and network; services, software and datasets; workflows and VO's; "best practices"; policies for use, pricing, QoS etc.
- Grid entities:
  - Capture characteristics of Grid-architecture components.
  - Have a common naming scheme.
  - Can be described by metadata using a common hierarchical data model (RDF or XML).
  - Have their metadata published in "proxies."

A Reference Architecture for GRISEN



Conclusions and Future Work

- Our motivation stems from the need to provide effective systems and tools for the users of future massive Grids that will enable:

⇨ The performance exploration of Grid resources and the selection of appropriate resources for dispatching Grid jobs.

⇨ End-user navigation in the Grid information space.

⇨ The searching for interesting/relevant information for various Grid-related resources.

Conclusions and Future Work

- Automatic configuration and management of Grid benchmarks is important to:
  - Keep performance-metrics ontology "fresh."
  - Support the automatic auditing of resource providers by VOs: performance capacity, policy compliance, reliability of information services, etc.
  - Derive indirect, collective metrics expressing "quality features" of Grid infrastructures: level of heterogeneity, infrastructure health, reliability, robustness.
  - Automate the process of metrics-filtering.
  - Use benchmarking as a mechanism for driving automatic remote healing.

Conclusions and Future Work

- Navigation and Searching in the Grid, is essentially a process of capturing, managing, and querying…
- … a very large, fragmented, incomplete, heterogeneous, and often inconsistent, implicit metadata space.
- Looking at issues of:
  - Metadata representation (RDF) and navigation.
  - Metadata integration.
  - The scalability of metadata management (RDF/Jena).


Enabling Grids for E-science in Europe
Wed, 9 Mar 2005 20:29:29 +0200
Questions ?
Wed, 9 Mar 2005 20:29:04 +0200
South Eastern Europe Grid Infrastructure