An Introduction to Grid Computing

Bina Ramamurthy
bina@cse.buffalo.edu
http://www.cse.buffalo.edu/gridforce
Partially Supported by NSF DUE CCLI A&I Grant 0311473

Topics for Discussion

- Current Status of Information Technology
- Motivation to explore the Grid
- Grid services
- Grid high-level concepts
- Sample Application
- Grid-based Workflow

What is a grid?

- Grid is a ***sophisticated*** framework that enables sharing of a variety of resources among distributed applications.
- Open standard
- Large scale operations
- Automatic
- Intelligent
- Spontaneous
- Interoperable
- Service-oriented

What is a grid? (A formal definition)

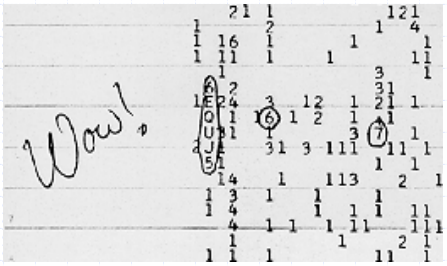
- Grid specifies a standard architecture, infrastructure, protocols and application program interface (API) for building an open enterprise system.
- It can provide
  - Middleware supporting network of systems to facilitate sharing, standardization and openness.
  - Infrastructure and application model dealing with sharing of compute cycles, data, storage and other resources.
  - A framework for high reliability, availability and security.
  - Interoperation of batch-oriented and service-based architectures.
  - Standard service level feature definitions and higher level concepts for inter and intra-business collaboration.

GridForce

- Our adaptation is evolving into a comprehensive framework we call GridForce (**Grid For Collaboration and Education**):
  - Course curriculum,
  - Laboratory exercises (labs),
  - Infrastructure to support labs,
  - Research projects,
  - Industrial outreach.

Beginnings of The Grid

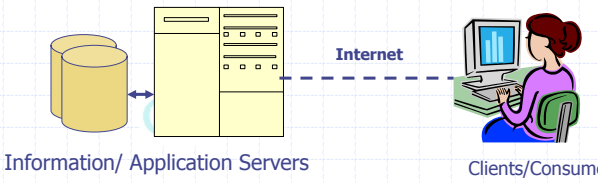
- Beginnings of the grid in Search for Extra Terrestrial Intelligence (seti@home project)
- http://planetary.org/html/UPDATES/seti/index.html
  - The Wow signal http://planetary.org/html/UPDATES/seti/SETI@home/wowsignal.html

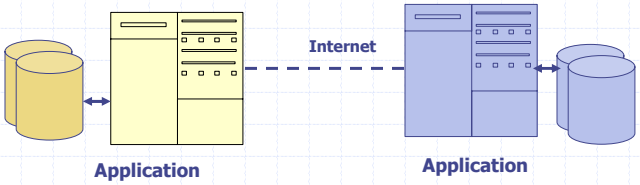


Current Grid Users

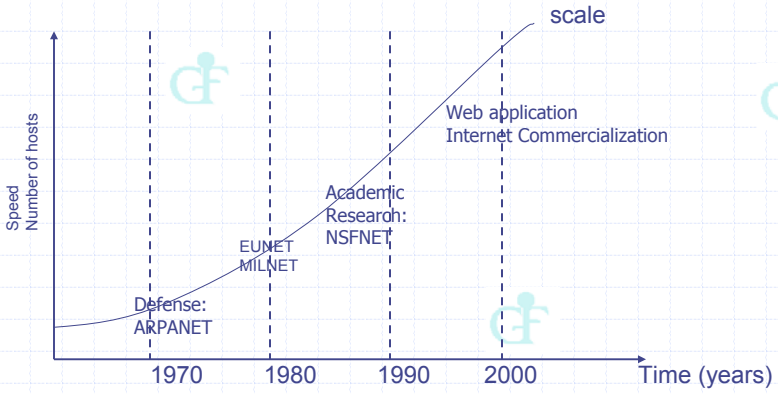
- A survey of 180 companies last summer by research firm Summit Strategies found that 4% of respondents had implemented a grid, and 12% were currently evaluating the technology.
- Gartner predicted in 2002 that grid-based distributed systems will return by 2004-2007.
- Oracle Server 10g: g stands for grid. (Oracle 9i: i was for Internet)
- Grid middleware from companies such as DataSynapse and Platform provides users the ability to manage workloads across the shared resources.
- IBM used grid-base infrastructure for 2004 US Open: Enterprise Networks Aug 2004.
- Burlington Coat Factory is investing its IT future on a grid-based, virtualized architecture: Enterprise Network June 2004.
- HR outsourcer Hewitt Associates put grid to work for crunching pension calculations.
- …

Current Status

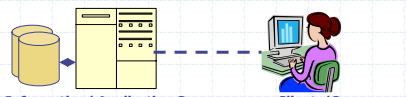




How did we get here?



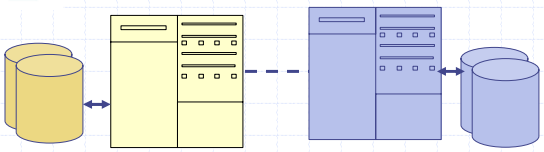
Where are we heading?



Business to Consumers (B2C)

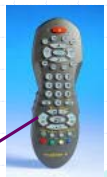
Web-enabling information
Web-enabling applications/forms
HTML

Business to Business (B2B)
Application to application



Web Services, XML
Standards for specifying operation in
SOAP (Simple Object Access Protocol)

Beyond Search Engines: Enabling Information Technology (IT) Applications



Simple Search (stateless)

Financial: Build Portfolio

Environment: Plan Forestation

Medicine: Find Cure

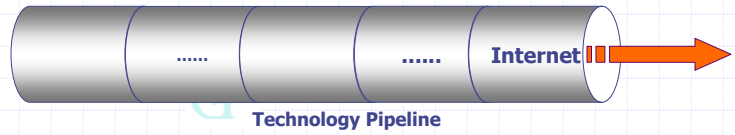
Travel: Plan a Trip

Complex multi-business applications

Web Services Standard

- A common operation on the Internet is search, the results of which is consumed by humans.
- We want to develop complex multi-business applications that are beyond the current search-type applications.
- Webservices (WS) is a standard that has been introduced by W3 consortium to address this important transition.
- Grid takes the web services to the next level: a grid service (GS) is a web service.
- GS = WS + state + standard features for security, reliability, integration, …
- Grid specifies a standard architecture, infrastructure, protocols and application program interface (API) for an open enterprise system.

Technology Pipeline

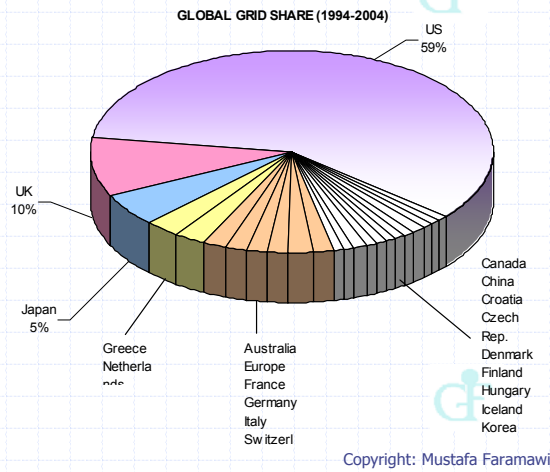


Technology Pipeline

Technology Pipeline

Grid Growth

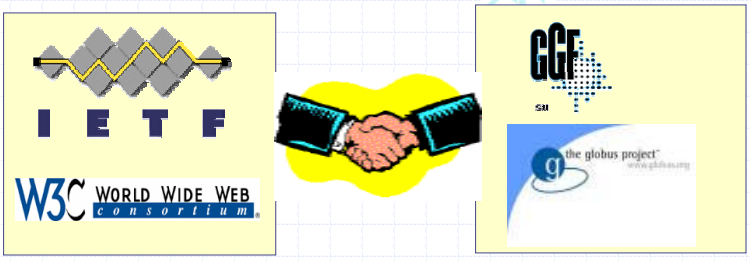
Global Grid Share



Grid Organizations

- Global Grid Forum (GGF):
  - www.globalgridforum.org
  - a community-initiated forum of thousands of individuals from industry and research leading the global standardization effort for grid computing.
- The Globus Alliance:
  - www.globus.org
  - conducts research and development to create fundamental technologies behind the "Grid," which lets people share computing power, databases, and other on-line tools securely across corporate, institutional, and geographic boundaries without sacrificing local autonomy.

Future Outlook



Internet and Web Standards

Grid Standards

- It is expected
  - either the Internet will evolve into the grid or
  - the grid concepts will be adapted into the Internet standard.
- Similar to current push in IT to "web enabling", future will have you "grid enable".
- Bottom line: it is worthwhile learning about the grid to strategize for the future of IT in your business.

What can the Grid do?

- Grid specifies a standard architecture, infrastructure, protocols and application program interface (API) for building an open enterprise system.
- It can provide
  - Middleware supporting network of systems to facilitate sharing, standardization and openness.
  - Infrastructure and application model dealing with sharing of compute cycles, data, storage and other resources.
  - A framework for high reliability, availability and security.
  - Interoperation of batch-oriented and service-based architectures.
  - Standard service level feature definitions and higher level concepts for inter and intra-business collaboration.

Types of Grid

- **Batch-oriented**

1. Compute-intensive jobs processing using sophisticated scheduling and resource discovery.
2. High performance applications
3. High Throughput applications
4. The Condor Project
5. Example: Condor
6. Our installation: CSECCR grid

- **Service-Oriented**

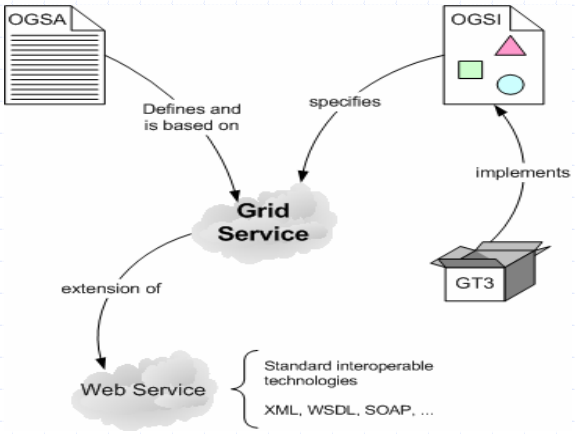
1. View all the resources and functions as services.
2. Build application models around services.
3. Anatomy of the grid
4. Physiology of the grid
5. It is this genre of grid that will move the grid technology towards business applications.
6. Example: Globus
7. Our installation: CSELinux Grid

Service-oriented Standards

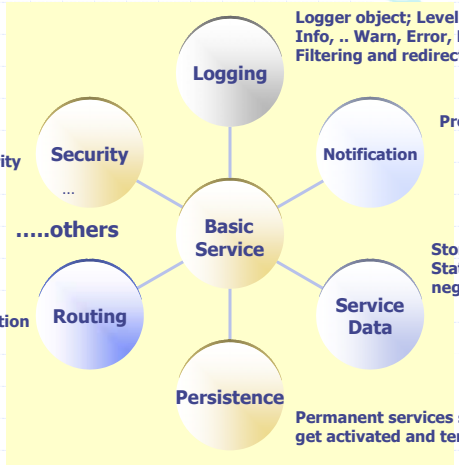
- Open Grid Services Architecture (OGSA)
- Open Grid Services Infrastructure (OGSI)
- Globus Toolkit (Gt3) is a reference implementation
- We will discuss next:
  - service-level concepts and
  - higher-application-level concepts.

OGSA, OGSI and WS

From tutorial: Satomayor's GT3 Tutorial
http://www.casa-sotomayor.net/gt3-tutorial-working/



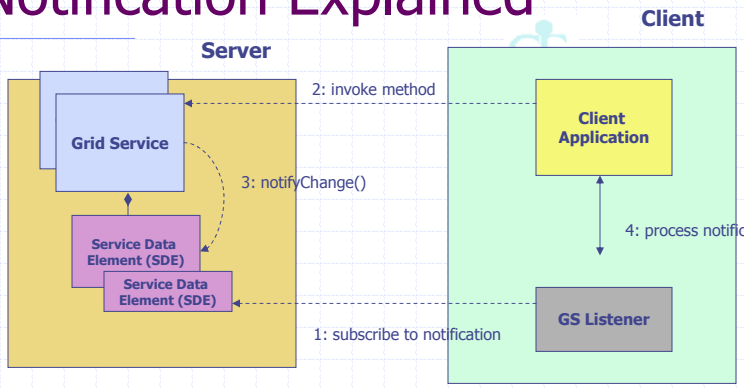
Standard Features of Grid Service



Sample Grid Service: Notification

- Foundational concepts: messaging, queues, source and sink for messages, subscription model, loose coupling, push and pull notification
- Grid related concepts: Service data element (SDE), OGSINotification API
- SDE is XML structure for holding service chaaracteristics/state.
- Implement a service that is a producer of notification.
- Notification can be triggered by change in SDE.
- Implement a client application that invokes a service that produces notification; an associated listener that consumes the notification.

Notification Explained



Example: Grid service (GS) can be a Math Service with notifyChange to SDE on invocation of add Subtract methods.
GWSDL file: extends="ogsi": GridServiceogsi:NotificationSource (declarative vs programmatic)
Listener has: NotificationSinkManager to which is added a listener to Math Service's GSH and SDE.
Listener has deliveryNotification() method to process notification.

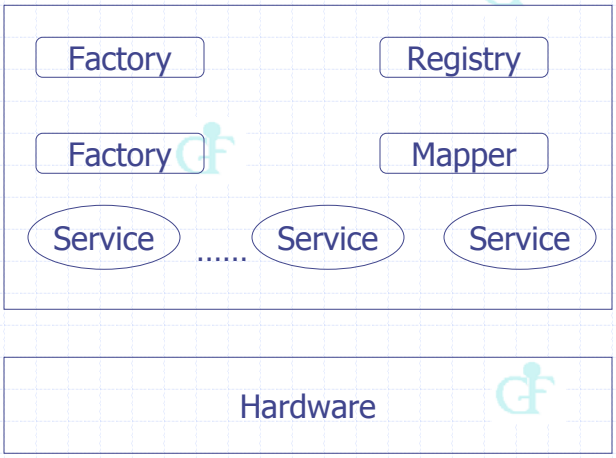
Higher Level Grid Concepts

- Virtualization of services and resources
- Federation of Data
- Provisioning
- Lifecycle Management
- Virtual Organization

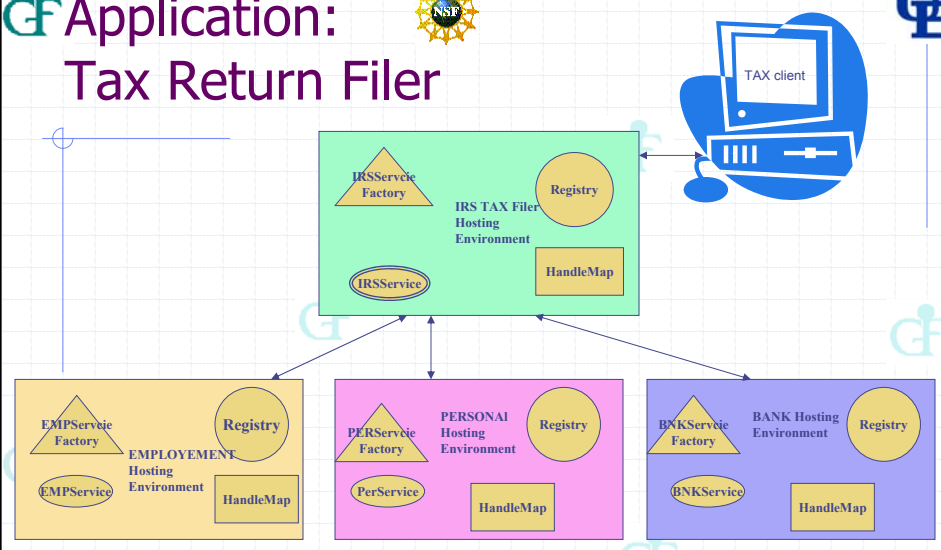
Virtualization

- Encapsulating service operations behind a common message-oriented service interface is called service virtualization.
- Isolates users from details of service implementation and location.
- Assumes support of a standard architecture.
- Webservices (WS) can do this, however grid life cycle management, fault handling and other features we have seen in the GT3 tutorial are not available with WS.
- OGSI specification addresses these issues using a core set of standard services.

Virtual Organization (VO)



Application: Tax Return Filer



Concepts illustrated: Virtual organization (VO) called IRS/Tax Filer that brings together virtualized capabilities of physical organizations of banking, personal profiles, and employment. Grid service handle (GSH) and Grid service reference (GSR), registry and handlemap, discovery of services, index services, application of notification, logging.

UB Infrastructure(1): CSELinux Grid

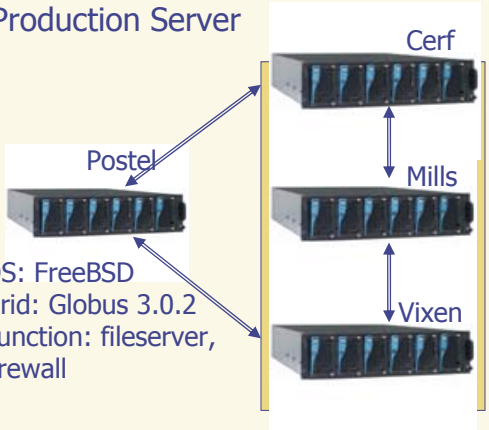
- Goal: To facilitate development of service-oriented applications for the grid.
- Two major components: Staging server and Production grid Server.
- Grid application are developed and tested on staging server and deployed on a production server.
- Production grid server:
  - Three compute nodes with Red Hat Linux and Globus 3.0.2 instance.
  - One utility gateway node with Free BSD and Globus 3.0.2.

CSELinux: Development Environment

Staging Server

OS: Solaris 8.0
Grid: Globus 3.0.2
Function:
Debug and test services

Production Server



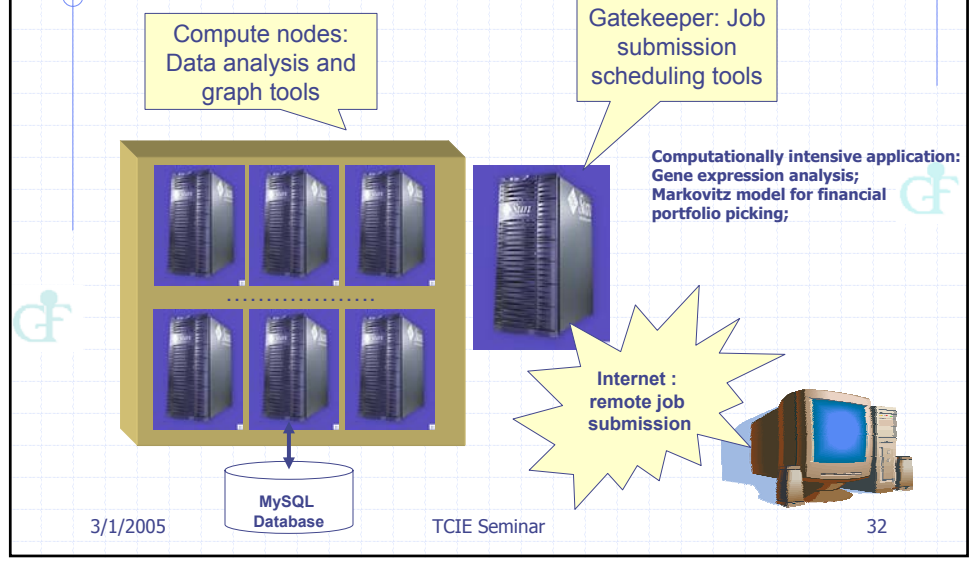
OS: FreeBSD
Grid: Globus 3.0.2
Function: fileserver, firewall

OS: Red Hat Linux 9.2
Grid: Globus3.0.2
Function: Deploy services

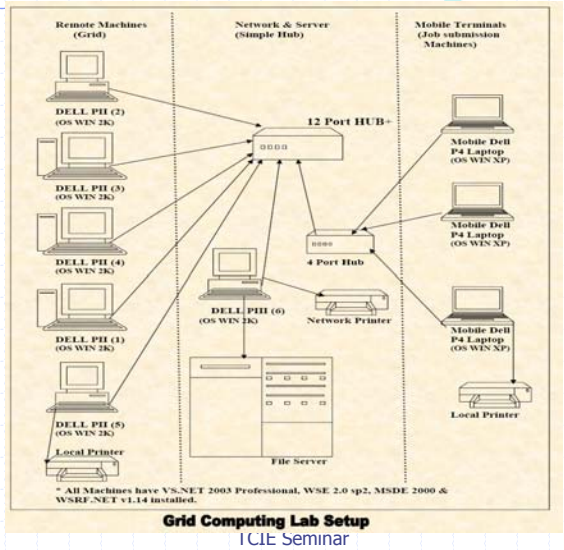
UB Infrastructure(2): CSECCR Grid

- Goal: To run jobs submitted in a distributed manner on a Condor-based computational cluster Condor.
- Composed of 50 Sun recycled used Sparc4 machines, which form computational nodes, headed by a front-end Sun server.
- The installation scripts are custom-written facilitating running of jobs in a distributed manner.
- Partially supported by Center for Computational Research (CCR).

CSECCR Grid

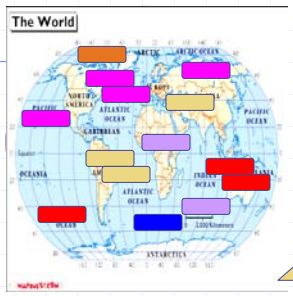


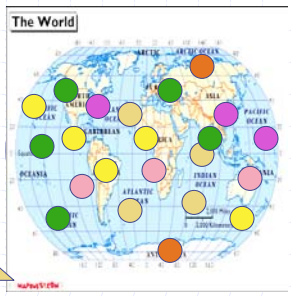
WSRF.NET Grid



Grid Computing Lab Setup

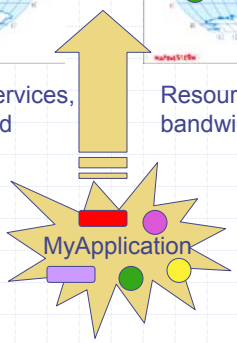
Services and Resources



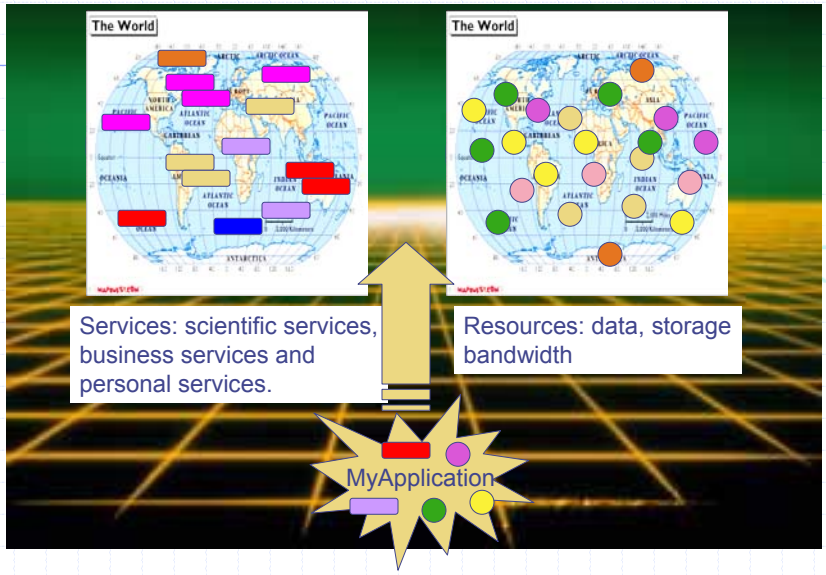


Services: scientific services, business services and personal services.

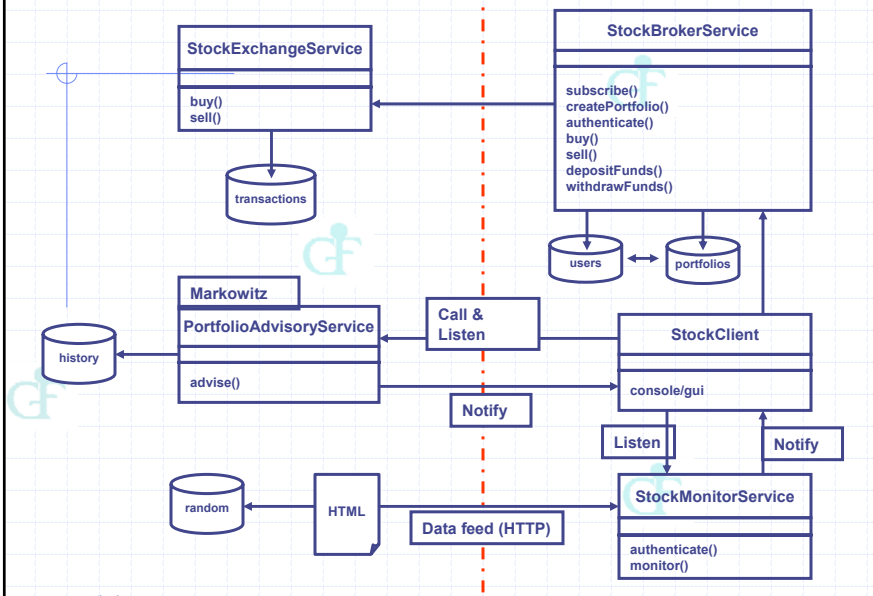
Resources: data, storage bandwidth



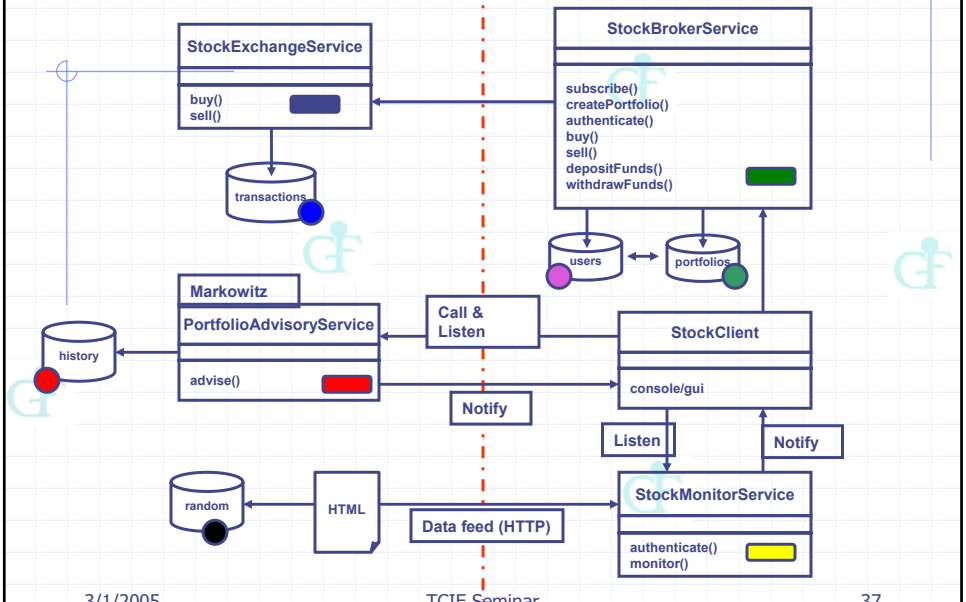
Grid Application using the services and resources

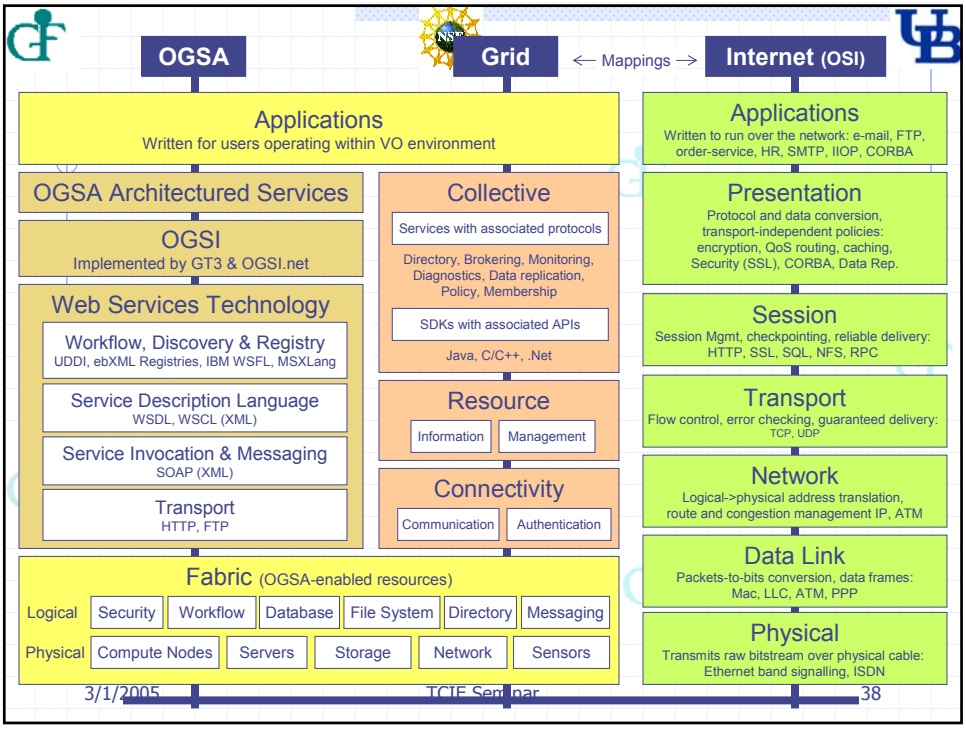


Grid-based Stock Portfolio Building Application: Design Diagram By Faramawi, a Student in CSE4/586 Course

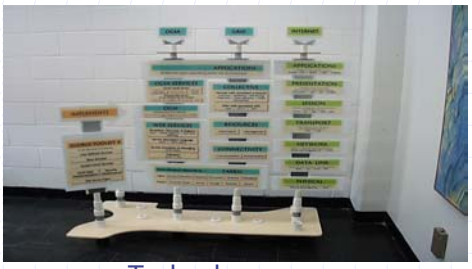


Grid-based Stock Portfolio Building Application: Design Diagram By Faramawi, a Student in CSE4/586 Course





The Grid Conundrum



Technology

Grid ← → Internet

State of the technology

Flux ← → Solid

System level operation ← → Protocol level operation

Where are the opportunities for CSE educators and students?

Getting to know the grid?

- Start with reading the literature on Condor and Globus grid.
- http://www.globus.org/research/papers/anatomy.pdf
- http://www.globus.org/research/papers/ogsa.pdf
- http://www.globus.org/research/papers.html
- Try out the grid tutorials and reference implementations.
- Explore newer businesses and business models.
  - Example: storage service, personal database service (personal identity management)
  - Work on a reference implementation of grid specification.
  - http://www.extreme.indiana.edu/swf-survey/