ISSN 2229-5518

Grid computing: An Emerging technology

G.S. Rohit

Abstract-Internet is the networking infrastructure which helps in connecting many users through interconnected networks through which users can communicate to each other. The World Wide Web is build on top of the internet to share information. The grid is again a service that is built on top of internet but is able to share computational power, databases, disk storage and software applications. This paper mainly focuses on significance Grid computing, its architecture, the grid middleware Globus toolkit and facts of wireless grid computing.

Index terms-Distributed computing, Globus, Grid layers, GridFTP, GSI, GT4, Security, Wireless.

1 INTRODUCTION

Grid computing, also known as the distributed computing or parallel computing is a cluster of well defined computers joined together by a network to achieve a common goal. It provides a virtual platform for the sharing of various computing resources. It has applications in different fields of life such as government, business, research, science etc. Grid computing is not just sharing of data between two different computers but between large scale of computers acting like individual agents coming to work together and solving large problems and other tasks. Here the bigger problems are broken down into smaller ones which are solved by these small computers without the use of a super computer. The advantage of such computing is that even if one of the computers goes down the job can be still carried out by some other computer unlike in the case of supercomputers. These computers can be connected to each other through different network such as private, public or the internet through an Ethernet cable. Through this network the devices like hard drives, cd drives, RAM and printers can be shared. Even though grid computing appears quite similar to distributed computing, the requirements for grid computing are much more complex. The distributed computing refers to managing thousands of computers which individually are more limited in their memory and processing power, however grid computing concentrates on efficient utilization of a pool of heterogeneous systems with optimal workload management .Grid computing helps in easy sharing of distributed heterogeneous hardware and software resources through dependable access. Grid provides computing has the capability for Security, data transfer, Job Submission and resource discovery thus making a perfect intranet. In places where Internet facilities are not available grid computing can act as a perfect alternative. For grid computing to happen, grid middleware is required.

- *G.S.Rohit is currently pursuing his bachelor's degree in engineering in Information Science at Dr Ambedkar Institute of Technology, Bangalore, India PH ,9008491494. E-mail-rohit.gs28@gmail.com*

Grid middleware is a specific software product which enables the sharing of heterogeneous resources. Example of popular grid middleware are Globus Toolkit, gLite, Legion and UNICORE. Globus is the most widely used grid middleware. It is an open source toolkit developed and provided by the Globus Alliance. The Globus Project provides software tools that make it easier to build computational grids and grid-based applications. The Globus grid forum developed standard interfaces, behaviours, core semantics for grid applications based on web services. Grid computing cannot be limited to fixed computing systems. With the increasing number of Smartphone users in the world, mobile devices should be utilized for this purpose. The paper discusses the given topics in detail.

2 HISTORY

The grid computing emerged in the mid 1990's in scientific computing, although it was around for decades. It was originally created to remove the geographical constraints and make proper use of unutilized resources.

In the early 1970's the idea of harnessing unused CPU cycles was born. A pair of programs called Creeper and Reaper ran on the Arpanet. In 1973, the Xerox Palo Research Centre (PARC) installed its first Ethernet network. Scientists John F. Shock and Jon A. Hupp created a worm as they called it which moved to different machines using idle resources.

Richarf Crandall a distinguished scientist at Apple, started putting idle, networked NeXT computers to work .He installed a software that allowed the machines when not in use to perform computations and combine the efforts with other machines in the network. The idea of grid was brought together by Ian Foster, Carl Kesselman and Steve Tuecke who are regarded as the fathers of grid. They created the Globus toolkit which incorporated not just the computation management but also the storage management, security management, security provision, data movement, monitoring and a toolkit for developing additional services based on the same infrastructure which includes trigger services, information aggregation, notification mechanisms and agreement negotiation.

3 ARCHITECTURE AND DESIGN

As the globus toolkit remains the de facto standard for building grid solutions we will go through the globus grid architecture. It consist of the following layers

1) Fabric layer: Interfaces to local control, including physical and logical resources such as files. The grid fabric layer helps in providing resources to which shared access is mediated by the grid protocols such as computational resources, storage systems, catalogs, network resources and sensors. Richer fabric functionality helps in sophisticated sharing operations at the same time. The Globus toolkit was designed to use an existing fabric component which includes vendor supplied protocols and interfaces. Whenever the fabric level behaviour is not provided, the Globus toolkit includes the missing functionality

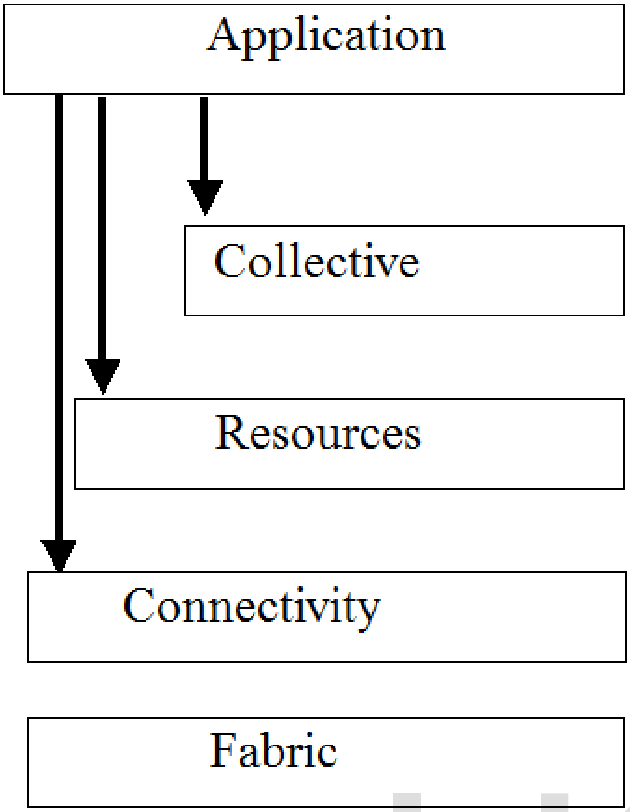

Fig 1: The layered Grid Architecture

IJSER

2) Connectivity layer: Helps in core communication and authentication protocol that supports grid specific transactions. Communication protocols enable in exchange of data between fabric layer resources. Authentification protocols that are build on communication services to provide cryptographically secure mechanisms for verifying the identity of resources. Authentication solutions for virtual organizations should have the following characteristics, they are

Single sign on: Every user must log in only once which will allow him to access multiple grid resources.

Delegation: Every user must be able to endow a program with the ability to run on that user's behalf so the program is able to access the resources on which users have been authorized.

User based trust relationships: In order for a user utilize resources from multiple providers then the security system must be such that the resource providers should interact with each other configuring security environment.

Integration with local security: Each site or resource must employ local security including Kerberos and UNIX security.

3) Resource layer: Allows the sharing of a single resource and it builds on connectivity layer communication and authentication protocols to define protocols for secure negotiation, monitoring, initiation and control of sharing operations on individual resources.

Two primary classes of resource layer protocols can be discussed

Information protocols: They are used to obtain information about the structure and state of resources

Management protocols: They are used to negotiate access to various shared resources, specifying resource requirements and operations performed.

4) Collective layer: Allows resources to be viewed as collections and sharing of resources. The Collective layer contains protocols and services not associated with any one specific resource but instead capturing interactions across the collection of resources.

5) Application layer: This layer uses appropriate components of each layer to support the application. Each of these layers may contain protocols, APIs and software development kits (SDK) to support the development of grid applications

The Integrated Grid computing infrastructure provides grid portals. These portals are web based user interfaces which interconnect the users and resources. The portals address various grid middleware issues and it supports applications in the area of Scientific and Engineering, Collaborative Computing and various application specific web enabled applications.

The Integrated Grid computing infrastructure provides grid portals. These portals are web based user interfaces which interconnect the users and resources. The portals address various grid middleware issues and it supports applications in the area of Scientific and Engineering, Collaborative Computing and various application specific web enabled applications.

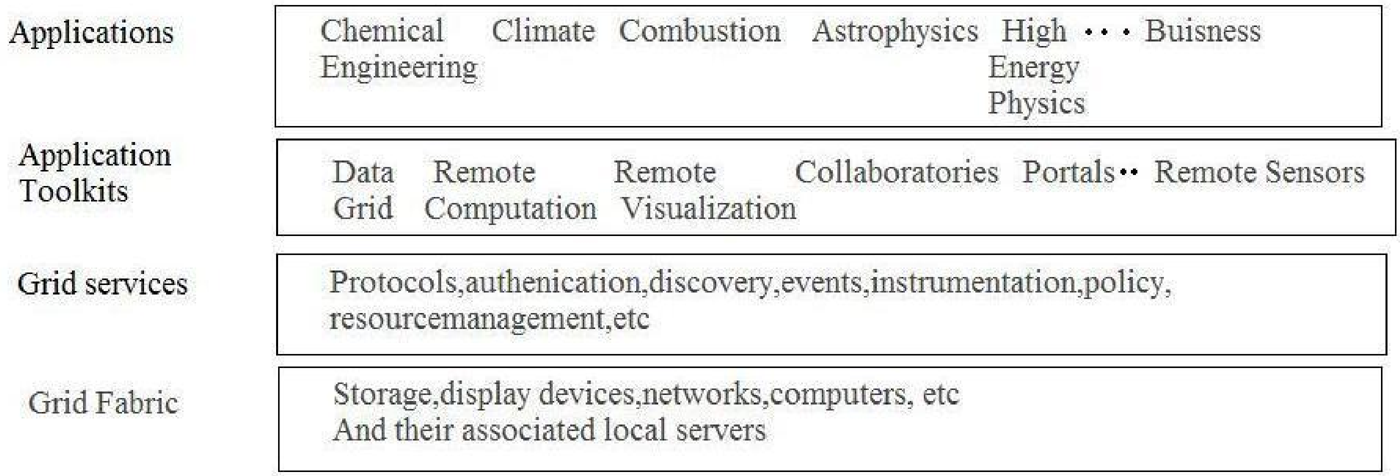

Fig 2 : The Integrated Grid Architecture

The important Components and services of each Globus grid layer is as follows.

Grid fabric (Layer one): The fabric of the grid comprises all the underlying systems, computers, operating systems, routers and other storage systems.
Grid services (Layer two): Grid services integrate the various components of grid fabric. Examples of services provided by Globus are :
GRAM: The Globus resource allocation manager is library service that provides capabilities to do remote submission job start up, locate, monitor, submit or cancel jobs. It unites Grid machines. It has also been designed to minimize the privileges required and to minimize the risks of service malfunction.

GridFTP: It is a high performance, secure, reliable data transfer protocol optimized for high bandwidth wide area networks. The GridFTP protocol is an extension of the standard File transfer protocol (FTP) for high speed and secure data transfer.

GIS-The Grid information service, GIS, was formerly known as Metacomputing Directory Service, MDS which provides information service.

GSI-The Globus toolkit incorporates multiple security components that helps to establish the identity of users or services, protect communication determine what actions are to be performed. The Grid security infrastructure is a library for providing generic security services for applications that run in the grid.
Application toolkits (Layer three) :The application toolkits use the grid services to provide higher level capabilities which are generally targeted to a specific application. The programming interface is constantly changing and is complex in nature. Thus to make the grid aware which consist of various heterogeneous and dynamically changing machines, it must run efficiently with various grid applications that are scheduled and proper resources allocated properly.
Specific applications (Layer four): A variety of applications can be developed that build on the services provided by the above three layers. Such specific applications include Chemical engineering, climate, High energy physics, compound chemistry, etc.
GT4 architecture overview
GT4 is a set of software components for building distributed systems that implement web services mechanisms. Web services provide a standard means of interoperating between different variety of platforms and frameworks. It addresses various service components such as Execution management, Security Services, Data Management, Information Services, Common Runtime Support (Python Web Core C, Java & extensible I/O).The Global Grid forum developed standard interfaces, behaviors, core semantics for grid applications based on web service. Before GT4, the globus The globus grid forum were divided into two parts: 1)Open Grid Services Architecture(OGSA) and 2)Open Grid Service Infrastructure(OGSI).OGSA defined standard mechanisms which were used for crating, naming, discovering grid service instances and addresses architectural issues relating to interoperable grid services. OGSI was based upon grid service specifications which specified a way in which client must interact to grid service. Globus toolkit 3(GT3) which is a predecessor of GT4 is a implementation of (OSGI).THE Globus Toolkit version 1 (GT1) and version 2 (GT2) are the predecessors of GT3.The technologies that are used to build the Globus toolkit components include GRAM, MDS, GridFTP. All these components use Grid Security Infrastructure (GSI) for security.
A web service is a software system designed to support interoperable machine to machine interaction over a network. It has an interface described in a machine process able format. Other systems interact with web service in a manner prescribed using message encoding (SOAP messages),typically conveyed by using HTTP with an XML serialization with other web related standards. Web services standardize the messages that entities in a distributed system must exchange in order to perform various operations. At the lowest level, this standardization concerns the protocol used to transport messages (typically HTTP), message encoding (SOAP), and interface description (WSDL) A client interacts with web service by sending it a SOAP message. The client may receive response messages in reply. At higher levels, other specifications define conventions for securing message exchanges (e.g., WS-Security), for management (e.g., WSDM), and for higher-level functions such as discovery and choreography.

IJSER

GT4 Architecture comprises of both a set of service implementations (i.e. server code) and associated client libraries. GT4 provides both Web Services (WS) components (on the left) and non-WS components (on the right). The white boxes in the client domain denote custom application and third party tools that access GT4 services or GT4-enabled services. All GT4 WS components use WS-Interoperability-compliant transport and security mechanisms, and can thus interoperate with each other and with other WS components.

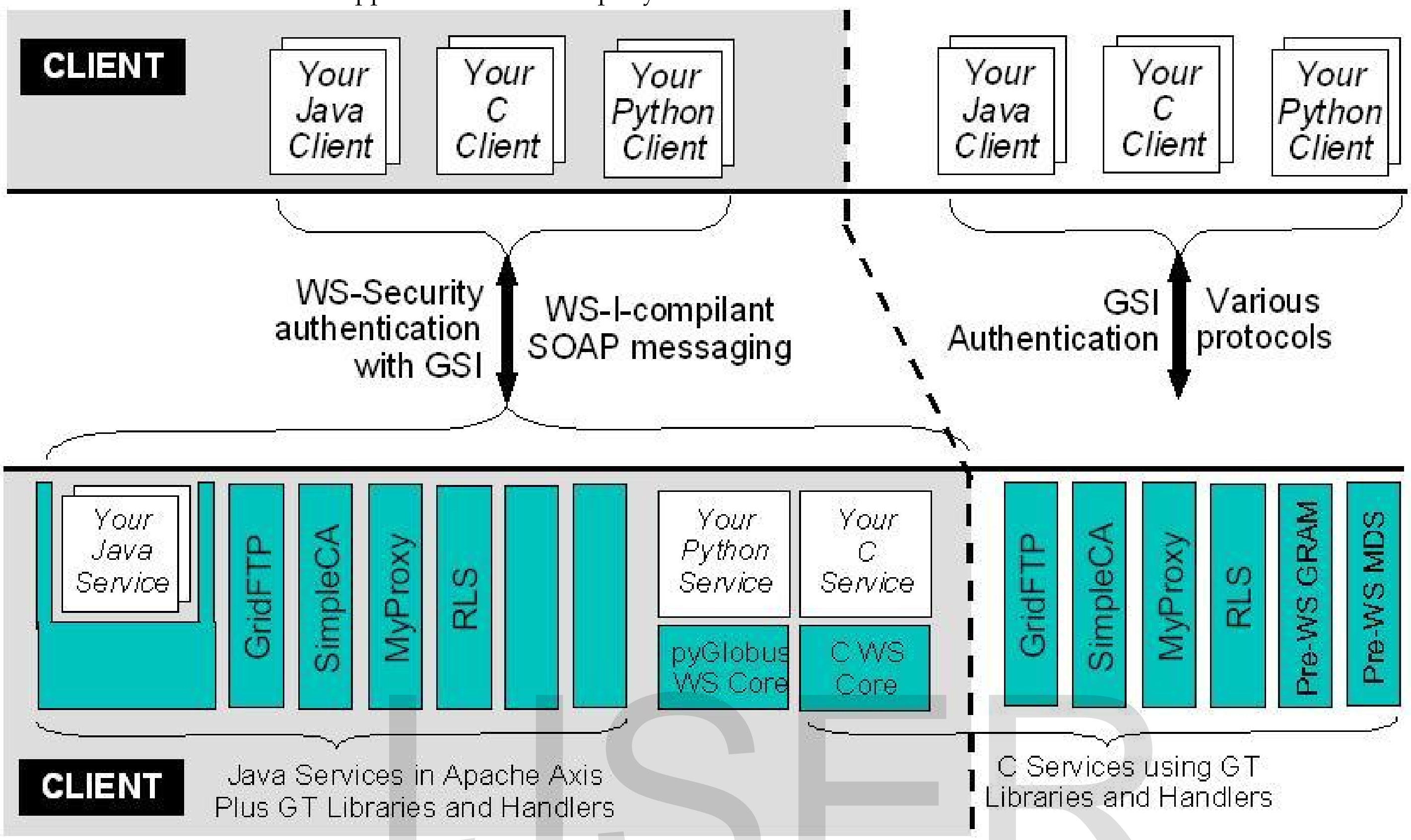

Fig. 3. Schematic view of GT4.0 components

A brief summary of GT4 elements are discussed below.

Service-oriented-architecture: GT4 software toolkit is designed so as to support applications in which set of services interact via standard protocols. The software includes both complete services and libraries implementing useful protocols. Developers can use these services and libraries, plus other related software, to build both simple and complex systems relatively quickly.

Infrastructure Service: GT4 includes built in services for accessing, managing, monitoring and controlling access to such infrastructure elements as computational and data resources.

Web Services: The GT4 software makes use of Industry-standard Web services protocols and mechanisms for service description, discovery, access, authentication, authorization.

Gt4 Containers: The GT4 software includes components that can be used to construct GT4 containers for hosing Web Services written in Java, C and Python.

The Globus toolkit is currently at version 5 is GSI-OpenSSH is a modified version of OpenSSH that adds support for X.509 proxy certificate authentication and delegation which provides a single sign for remote login and file transfer service. GSI-OpenSSH can be used to login to remote systems and transfer files between systems without entering password. It uses a valid proxy credential for authentification. GSI-OpenSSH forwards proxy credentials to the remote system on login, so commands which require proxy credentials (including GSI-OpenSSH commands) can be used on the remote system without the need to manually create a new proxy credential on that system.

Types of grid computing

Computational grid-The computational grid is defined as large number of computers grouped together with large pool of resources to carry out intensive computing.

Data Grid- Data grid helps is a grid computing system which provides access to data stored in different databases for data manipulation, handling, storage, support, publication and discovery.

Collaboration grid-Through collaborative methods different problems can be solved with the advent of internet. For instance, person with a particular problem can discuss the aspects of project without disclosing their proprietary technologies.

Network grid-A Network grid helps in providing high performance communication service between two communication points.

Utility grid: This is regarded as the ultimate form of the Grid, in which not just the data and computation cycles are shared but the software or just about any other resource is also shared. The main services provided by utility grids are software and special equipments. For instance, the applications can be run on one machine and all the users can send their data to be processed to that machine and receive the required result back.

IJSER © 2013
http://www.ijser.org

4 Security concerns

As grid involves sharing of data, security is one of the key requirements for ensuring secure connection which involves Data Confidentiality (protection against spying), Authentification (proof of identity) and Data integrity(protection of message modified in transit).The globus designers recognized the grid operation had to be secure. In globus grid every user is issued with a trustworthy identification called a grid certificate so whenever a user tries to run a specific grid command, his grid certificate is sent along with the command so the particular target computer system will be able to determine as to who has issued the command. In security terms, proving one's identify is called authentication. Apart from authentication it has to check whether the user has the required permission. In Globus toolkit version 4 each user and resource is assumed to have X.509 public key credentials. Certificate authority (CA) is a program that issues certificates to each process, this certificate can be easily obtained by running the request command. The distinct WS and pre-WS authentication and authorization capabilities which are build on the standard X.509 end entity certificates and proxy certificates which are used to identify persistent entities such as users and servers and to support various temporary delegations of privileges to other entities. Ws security comprise of 1)Message level security mechanisms, which implement the WS-security standard and the WS- secure conversation specification to provide message protection for SOAP messages 2)Transport level security mechanism, which provides transport level security and 3)An authorization framework that allows for various authorization schemes. For non-WS components GT4 provides similar authentication, delegation mechanism but with lesser authorization options.

5 Wireless grid computing: Grid computing in mobile phones

As Grid computing tackles the problem of low internet penetration in various developing countries, the alternative to expensive prices of computing hardware can be dealt with the use of Smartphone. In 21st century the boom of technology has resulted in millions of Smartphone users. They include mobile devices of the form IPod, I Pad, Android devices, Blackberry mobile, Palm mobile etc. These devices are not only equipped with powerful processors but also large memory and storage-1 GB RAM and 2GHz processor is a common trend. With The sales of these Smartphone growing day by day, these Smart phones are being considered as the next wave of computing. One of the attempts to allow grid in mobile phones was done by enabling access to WSRF(web service resource framework) from mobile devices. Globus toolkit was chosen as the platform to develop the framework in conjunction with its implementation of WSRF.WSRF is an evolution of OGSI(open grid service infrastructure)which was a standard implemented by Globus. The software platform used was Java ME and in order to demonstrate the functionality of the framework and provide a user-friendly manner which will utilize its functionalities, a sample application was created and built upon framework libraries which were previously created. The Emulation platform was Nokia S60,as after research it was decided to restrict the case study implementation to focus on particular phone platform for deployment. This framework allowed users to browse and view files on the Globus server, submit jobs and transfer files between multiple servers. However its limitations were by the restrictions that were imposed by Java ME which works on limited environment were performance in not always ideal. Also one of the important limitations was utilizing the framework on non Nokia devices.

The mobile device capabilities are increasing by each passing day, however the proper integration has not been achieved yet. These are generally due to the following challenges faced

1) The Smart phones are typically restricted to reduce CPU, memory, secondary storage, and bandwidth capabilities as compared to desktop computers.

2) It also has increased heterogeneity, unpredictable long periods of complete disconnectivity, unreliable, low-bandwidth and dynamic network as the devices can enter or leave in unpredictable manner. As compared to typical desktop computer the mobile devices have 1/3th processing power,1/8th memory,1/10th network bandwidth,1/5th storage capacity. While mobile services continue to improve the inequality between resource constraints of mobile and fixed devices will remain.

The other major challenges are:

Hardware challenges: The battery technology has matured slowly over the past decade but has failed to keep up with increased power demands contemporary high-end mobile devices.

Network challenges: To overcome the network challenges applications must be written keeping the disconnections in mind.

Middleware challenges: Middleware are too heavy with respect to computational capacities of mobile devices and are not network centric.

Security challenges: Security is one of the most important issue as mobile devices are susceptible to wide range of attacks.

However due to lesser cost of installation per CPU cycle and dedicated system will eventually fall short of increasing demands, a mobile grid will evolve to speak with time with respect to computational power due to inclusion of faster devices and exclusion of inferior devices. Also with the advent of HTML5, mobile grid has an opportunity to grow as HTML5 along with JavaScript can provide a hardware and software independent means for developing grid applications.

6 Conclusion

- Grid is the next generation internet, it is not an alternative to internet but it is rather an additional set of protocols and services that build on Internet protocol and services.
- Grid is a source of free cycles
- The complete integration of fixed computing systems and mobile devices can happen in the future

References

[1] Ian Foster, Carl Kesselman, Steve tuecke : The Anatomy of grid
[2] Rajeev Wankar : “Grid computing with Globus: An Overview and Research challenges” International journal of computer science and applications Vol 5, No 3, pp 56 - 69
[3] Jan Christian, Mangs Shilhong, Huang Junwei Cao: Enabling access to WSRF fom Mobile Devices
[4] Ketan B. Parmar, Nalinibhai N Jani,Pranav S. Shrivatsav, Mitesh H. Patel: Mobile Grid Computing :Facts or Fantasy? International journal of multidisciplinary science and engineering Vol ,No 1,January 2013
[5] A Globus Toolkit Premier (Describing Globus Toolkit Version 4)
[6] GRIPSI-2007: An overview of grid computing http://www.cdac.in/html/events/beta-test/archives/gripsi-2007/gripsi-2007-gridcomp-overview.html
[7] AvarsysIncorporated 2005 http://www.avarsys.com/grid_computing_history.html
[8] Grid computing web blog 2005 http://thegridweblog.blogspot.in/2005/10/types-of-grids.html
[9] Wikipedia http://en.wikipedia.org/wiki/Grid_computing
[10] GSI-5.2.1Open-SSH http://globus.org/toolkit/docs/5.2/5.2.1/gsiopenssh/
[11] Grid Café http://www.gridpp.ac.uk/demos/gridsite-admin.cgi?cmd=print&file=flashcafe3.html

IJSER