

An Exploration of Grid Computing to be Utilized in Teaching and Research at TU

Dr. Mohammed A.M. Ibrahim & Group of Grid Computing

Department of Information Technology, Faculty of Engineering and Information Technology, Taiz University
Email: sabri1966@yahoo.com

ABSTRACT

Taiz University (TU) has a hundreds of computing resources on different campuses for use in areas from offices work to general access student labs. However, these resources are not used to their full potential. Grid computing is a technology that is capable to unify these resources and utilize them in very significant way. The difficulties of funding a complete grid computing environment and also, the difficulties of grid tools makes teachers and researchers in TU unable to involve in teaching and research in grid computing or in distributed computing. These problems raised up our awareness to mitigate this problem by build a simple environment for Grid computing from resources are available in TU and the built environment we can use it for teaching and research. The objective of this paper is to build, implement and testing a grid computing environment (Globus Toolkit). To achieving this objective we built the hardware and software parts, and configured several basic grid services commands line and web portal. The test result for basic grid services have been indicated that our proposed grid computing model is promising and can use in teaching and research in TU. The paper takes a look at how grid computing is realizing this aim and have created unbelievable opportunities for students, teachers and researchers at TU in addition the result of this paper will make TU a pilot to the other universities in whole Yemen in field of Grid and distributing computing.

Keywords - **Grid Computing, Network Computing, Software and Hardware of Grid Computing, Grid Evaluation.**

Date of Submission: October 20, 2014

Date of Acceptance: November 20, 2014

1. INTRODUCTION

Grid computing is an approach where the end user can be offered any of the services that provide by a grid or a network of computer system located either in a Locally or in geographical area. in grid computing user can dynamically select and locate any recourses such as (processing power, disk storage, applications, etc). In another hand grid computing integrate hardware and software by network , for multiple different organizations to be available to any of the users of any of these multiple organizations. Grid computing has many goals such as : providing remote access to IT assets, and aggregating processing power more detail in [1]. huge number of research have studied the grid computing in different areas and also, many researchers have defined the grid computing, Foster and Kesselman [2] Define a computational Grid “as a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.” Grid computing is concerned with coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. TU saver a shortages in funding and also in skills for grid computing technology for mitigating these shortages we come up with this paper which is maiming to implement a grid computing environment (Globus Toolkit).We choose

to implement grid and its tools because Grids can more easily deal with the enormous increasing amounts of data and provide tools to access, securely share data between trusted sources, storage and to solve other problems. To achieving this objective we built the hardware and software parts, and configured several basic grid services of a grid, this grid not be targeted to specific domain, we followed a general simplified architecture that leads to general practical results and build a wide skills knowledge in grid computing technology, we separated out the group of the different components and functionalities of the grid into smaller groups or parts distributed over different machines.

2. TYPES OF GRID

Grid computing address various kinds of applications used in many areas, grid computing are categorized by the type which are summarized below [3]:

- **Computational:** A computational grid is focused on setting aside resources specifically for compute power. In this type of grid most of the machines are high-performance servers.
- **Scavenging:** A scavenging grid is most commonly used with large numbers of desktop machines. In [4,5] machines are scavenged for available CPU cycles and other resources. Owners of the desktop machines are usually

given control of when their resources are available to participate in the grid.

- **Data grid:** The main task of data grid is storing and providing access to data across multiple organizations. an example, that if multiple organizations using an application, each organization has unique data. A data grid would allow these organizations to share their data, manage the data, and manage security issues such as who has access to what data[6].

3. GRID ARCHITECTURE

Grid computing should has number of characteristics and features that are required by a grid to provide users with a computing environment and the detail of these characteristics can be find in [7, 8]. Foster, Kesselman, & Tuecke, proposed a new grid architecture exploration can be seen in [9] this architecture, identifies the basic components of a grid system, defines the purpose and functions of such components and indicates how each of these components interacts with one another, the new architectures shown in Figure 1.

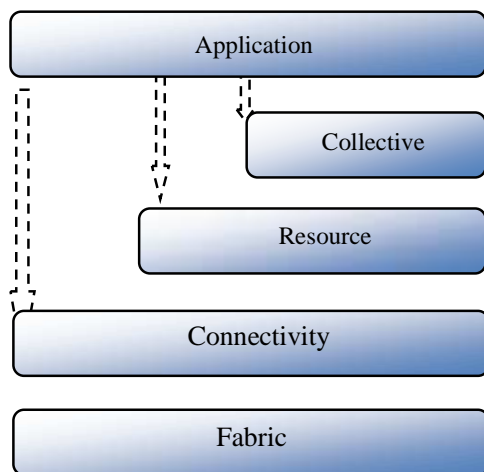


Figure 1. Grid Architecture

4. Grid Components

The major components that are necessary to form a grid computing and the detail of those components can be find in [10].

5. Implementation Requirements

To implement a grid computing model (Globus Toolkit) the hardware requirements and software requirements should be introduce as follows:

1. Hardware

The implementation test bad of grid computing model consisted of three machines each machine can be regarded as a source of computing power and data

storage capacity and all machines are identical in specification, which is listed below.

- Intel® Core ® 2 Due Processor (2CPUs) 2.40GHz, 2M cache
- Integrated Broadcom 802.11a/b/g/n Wi-Fi Adapter
- 2MB DDR2 Memory
- 250GB Serial ATA Hard Drive

These three machines will play the roles as servers as follow:

a). CA Server: Is the Digital certification authority server Used for issuing digital certificates (X.509 certificates) to grid users, resources and services.

b). MyProxy Server: is An online credential repository used to store X.509 proxy credentials, protected by a passphrase, for later retrieval over the network so you and other applications can access your credentials remotely. This eliminates the need for manually copying private key and certificate files between machines. MyProxy-Server can also be used for authentication to grid portals and credential renewal with job managers.

c). Portal Server: is a portal-server where the grid portal resides, the Grid portal is the access point to a web system; it provides an environment where the user can access the resources and services of the Grid, implement and monitor network applications and collaborate with other users.

2. Software Requirements

The basic software requirements that must be satisfied by any grid implementation are :

- **Ubuntu Desktop Edition 11.04**

We use Ubuntu as the main platform. Development tools and various other packages were installed as needed on each of the machine. **Globus Toolkit Version 4.2.1 (GT4)**

The open source Globus® Toolkit is a fundamental enabling technology for the "Grid," letting people share computing power, databases, and other tools securely online across corporate, institutional, and geographic boundaries without sacrificing local autonomy. The toolkit includes software services and libraries for resource monitoring, discovery, and management, plus security and file management[11,12].

- **Grid Portal**

we used the OGCE portal v2.5, it includes everything you need to get started building a Java-based Grid portal, including Tomcat 5.5 web server, GridSphere 2.1 portlet container, and OGCE Grid portlets.

- **Apache Tomcat 5.5**

Apache Tomcat is an open source web server and servlet container developed by the Apache Software Foundation (ASF). Tomcat implements

the Java Servlet and the JavaServer Pages (JSP) specifications from Sun Microsystems, and provides a "pure Java" HTTP web server environment for Java code to run in.

• **MyProxy v 4.2**

MyProxy is an open source software for managing X.509 Public Key Infrastructure (PKI) security credentials (certificates and private keys).

• **GridSphere 2.1 portlet container**

The GridSphere portal framework provides an open-source portlet based Web portal. GridSphere enables developers to quickly develop and package third-party portlet web applications that can be run and administered within the GridSphere portlet container.

• **OGCE Portal v2.5**

OGCE portal v2.5 includes everything you need to get started building a Java-based Grid portal, including Tomcat 5.5 web server, GridSphere 2.1 portlet container, and OGCE Grid portlets

6. GRID COMPUTING MODEL DESIGN AND ARCHITECTURE

The main focus of this section is building the hardware and software parts, and configure several basic grid services of a grid in which we will implement and test them. Our proposed grid computing model will not be targeted to specific domain, we will follow a general simplified architecture that leads to general practical results, we are going to separate out the group of the different components and functionalities of the grid into smaller groups or parts distributed over different machines so we can examine and get a clear view of each one functionalities. Figure 2. shows our proposed architecture.

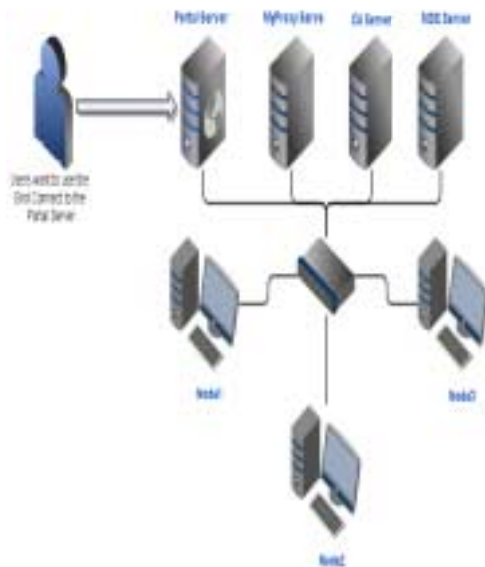


Figure 2. Proposed Grid Architecture Model

7. ACCESSING THE GRID

In order for the accessing the proposed architecture grid computing, user should create a set of keys for public key cryptography and request his certificate from the Certificate Authority and a copy of the public key of the CA.

1. Obtaining signed certificates from CA server:

Before users be able to request their certificates from a specific CA, they first should configure their hosts to trust that CA by coping CA's public key to setup GSI in their hosts, Figure 3. and the following procedure describe the steps to establish the GSI communication [13]:

1. Copy the Certificate Authority's public key to our grid host with which we setup GSI.
2. Creating private key and a certificate request.
3. after that we send the certificate request to CA by e-mail or another more secure way if you are running a production system and need to positively identify the sender.

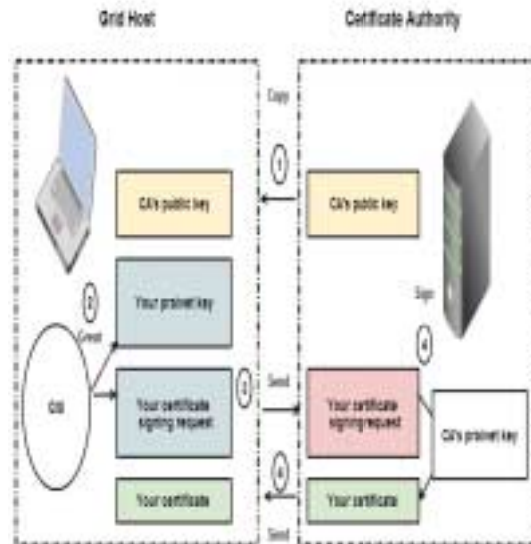


Figure 3. Obtaining Signed Certificates from CA Server

When that procedure has been completed the signed digital certificate should to be received, at that time user will have three important files on his own grid host and these files are:

- The CA's public key
- The grid host's private key
- The grid host's digital certificate

Mapping Global name of a Grid user to a local user name in each resource of the grid resources that the user is going to use.

2. Accessing the grid through the grid portal

A certain things should be done first before the user can access and use the grid resources and services through the grid portal from any machine that has access to the portal web server, before anything a kind of trust

relationship between portal web server and the underling grid architecture should exist. If a user wants to access the grid through the portal the portal should be able to access the grid on that user behalf, this can be achieved with the help of Myproxy Server which manages grid users credentials and delegation of user privileges to the portal server or any other machine that has access to the grid. The following subsections describe all the process in order.

a). Delegation of Credential to MyProxy server:

After the user has acquired his Grid credential he can use this credential to delegate his privileges to the grid portal. To delegate (store) a proxy credential in the MyProxy server, the user would run myproxy-init client program (contained in the MyProxy server package) on the machine (or logged in a secure manner - e.g. an encrypted Secure Shell session) where he can access his X.509 credential (using the passphrase that encrypts the private key associated with this credential) and delegate a proxy credential to the MyProxy server repository along with authentication information.

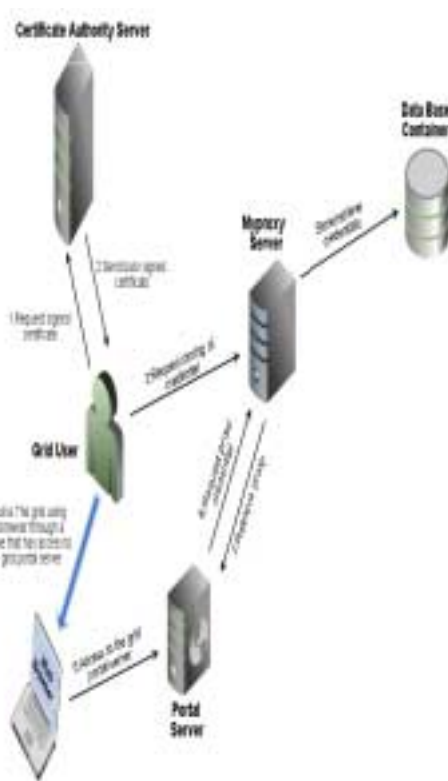


Figure 4. Delegation of Privileges to MyProxy server.

b). Retrieval of Credential from MyProxy server:

To obtain a proxy credential from the MyProxy server, the user must provide a username and the passphrase then user delegated a proxy credential in the MyProxy server. Figure 5. shows the steps for retrieval of proxy credential from the MyProxy server.

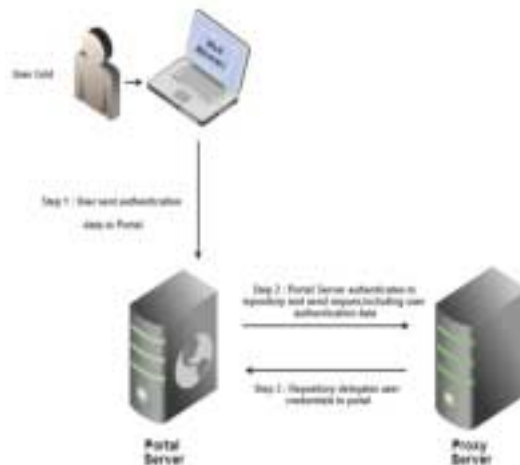


Figure 5. Retrieval Proxy Credential from MyProxy server.

2. Data Management

GridFTP: GridFTP is a high-performance, secure, reliable data transfer protocol optimized for high-bandwidth wide-area networks. The GridFTP protocol is based on FTP [14]. One of the major features of GridFTP is that it enables third-party transfer. Third-party transfer is suitable for an environment where there is a large file in remote storage and the client wants to copy it to another remote server our work in this section is similar to that in [15] , as illustrated in Figure 6.

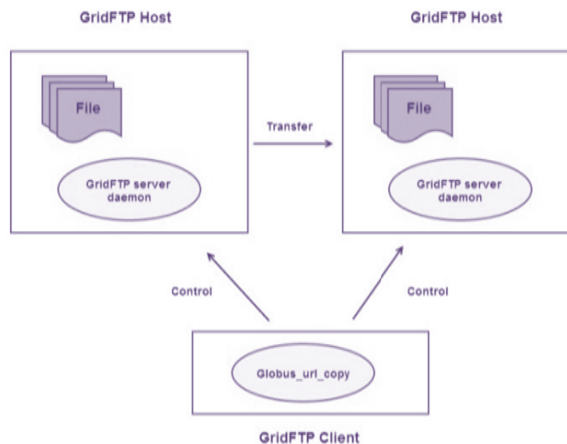


Figure 6. GridFTP third-party transfer .

- **Reliable File Transfer Service (RFT):** is a web service that provides interfaces for controlling and monitoring third party file transfers using GridFTP servers. The client controlling the transfer which hosted inside of a Grid service so it can be managed using the soft state model and queried using the Service Data interfaces available to all Grid services figure 7. shows the RST and GridFTP16].

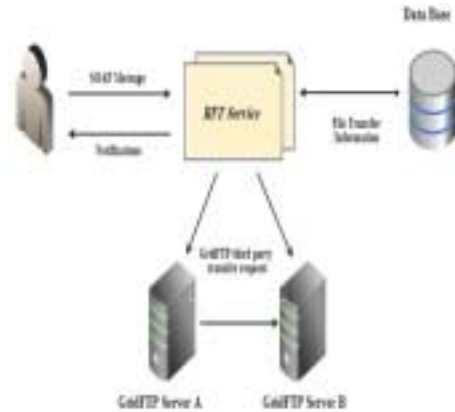


Figure 7. RST and GridFTP

8. INSTALLATION AND CONFIGURATION STEPS OF GRID COMPUTING MODEL

step 1:

Install and configure the software required for build our grid computing in all machines and some of these software are as follows:

- Globus Toolkit installer, from Globus Toolkit, Java J2SE 1.5.0+ SDK from Sun, IBM, HP, or BEA (do not use GCJ). IODBC (compile requirement for RLS) ,Tomcat (required by WebMDS, optional for other services), gLite Java VOMS parsing

•Setting up the Resource or Machine

This Settings applies to all node machines that is considered as a resource donor, in our infrastructure these machines are:

- node1.grid.local
- node2.grid.local
- node3.grid.local

step 2: Install Portal Server

Step 3: Setup Certificate Authority Server and this step include the following:

- 1) *1. Create users: user account, A generic globus account.*
- 2) *Run the setup script.*
- 3) *Configure the Subject Name:*
 - CA Name components:
 - Identifies the particular certificate as the CA certificate.
 - Identifies the CA from other CAs created by SimpleCA.
4. Configure the CA's email
5. Configure the expiration date
6. Enter a passphrase
7. Confirm generated certificate
8. Complete setup of GSI

Step 4:

1. Configure MyProxy version 4.0 and Myproxy-server can be configured automatically by the myproxy-server-setup command
2. Manage User Credentials
 - A. Initialize Credentials
 - B. Retrieve Credentials

Step5:

1. Distribute Service Credentials
 - A. Initialize Credentials
 - B. Retrieve Credentials

9. RESULT OVERVIEW.

- 1). After finished the Installation and Configuration steps to the proposed model for grid computing, figure 8. shown the result or the work has been done:

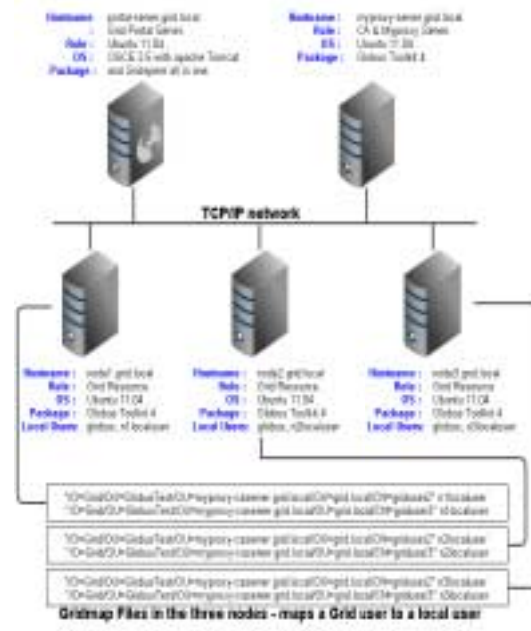


Figure 8. System after Installation and Configuration

2). Testing Grid Nodes

a). Testing Java WS Core

after setup, install and configure our model of Grid we started testing if it is work properly we have found a promised and encouragement results that indicate our grid model can be used as an grid computing environment for research and for teaching in TU , figure 9. shown promised result In each grid's nodes that is configured to run the Globus container, globus user can run all commands which used for grid computing such as:

globus-start-container this command use only if security requirements has been set up , but if security requirements has not yet been set up user can run the following command: **globus-start-container -nosec**



Figure 15. Login page of the grid portal

At the time we login, all basic deployed portals will be displayed in taps as in shown in figure 16. deployed portlets are the following:



Figure 16. Taps of Deployed Portal.

3. Using Grid Portal to Get User proxy Credentials from Myproxy server:

Figure 17 shows the File Management Portal. Files can be uploaded or downloaded through the browser. *Myproxy-portlet* provides us with the credentials to use the grid due to after logging to the grid portal the user is not able to use the grid services until getting a proxy credentials of an already registered grid user from myproxy server, using a grid user name and password via the following address: myproxy-server.grid.local address.



Figure 17. File Management Portal

4. Grid Security Infrastructure GSI: is responsible for providing APIs and tools for authentication, authorization and certificate management to our proposed grid model figure 18. shows the GSI proxy credentials loaded into an account and will be valid for two hours.



Figure 18. GSI Proxy Credentials Loaded from Myproxy Server

5. File Transfer and Job Submission Through the Grid Portal:

Figure 19 shows the Job Submission Portal, Information of fields available and resources of sending and receiving the job must be entered as a GRAM resource in a configuration file in order to show up in the Host list as once can see in the figure 19,20.



Figure 19. Job Submission Portlet



Figure 20. File Transfer Portlet not Connected

10. CONCLUSION & FUTURE WORKS

TU's campuses have more one thousand computing resources in computer labs and offices most of them connecting to networks and almost their time are idle time. Grid computing is mechanism to integrate or unify all of these resources and harness their computing power effectively, all these features of grid motivated us to brought TU into the grid computing field. Globus Toolkit provide the necessary features such as the ability of utilize idle time on any machines around campuses and this will give TU powerful high performance computing resources without any additional cost, in addition the results we have gained them from the Adoption and testing the proposed grid computing model in two ways commands line and web portal interfaces which are mitigate the complexity of the shortage of using grid computing fields in TU. After the successes of this work which was as starting point to get the grid computing technology work in place and the skills and knowledge in grid computing that we have gained, we about to start scale up our grid model to integrate all computing resources in labs and offices to built large-scale grid computing and make it available to students, faculty, and researchers. we have built our grid model and installed Globus Toolkit Version 4.2.1 on three machines including security components, Apon GridPortal, Ubuntu Desktop Edition 11.04, OGCE portal v2.5, Apache Tomcat 5.5, MyProxy v 4.2, GridSphere 2.1, portlet container and others grid software need.

Many tasks involved in effective use of the grid were discussed with example executions in section 6. Among these were credential management by requesting credentials with gridcert-request, creating proxy credentials with grid proxyinit, and using the MyProxy Server. result gained from commands line and portal interfaces were shown, Java WS Core, GridFTP, RFT, WS GRAM, MyProxy Server, Portal Server, Grid Portal to Get User proxy Credentials, GSI, Transfer and Job Submission Through the Grid Portal using command-line and portal interfaces both mechanisms were works correctly. Results with the final deployment of GridSphere and GridPort, proxy management, file management, and the resource monitoring services work without a problem.

After a successful testing period for the proposed grid computing and gained the promising and encouragement results which motivated us to start integrating the computing resources at TU into what we plan to called it the campus grid and we believe our future large-scale campus grid computing will help to further unite the campus' computing power and give students, faculty and researchers a stronger system with which to work.

REFERENCES

- [1] R. Al-Khannak, B. Bitzer. Load Balancing for Distributed and Integrated Power Systems using Grid Computing". *ICCEP 07, Capri, Italy, 23 November 2009, from IEEE database.*
- [2] S. Tarun and N. Sharma, "Grid Computing: A Collaborative Approach in Distributed Environment for Achieving Parallel Performance and Better Resource Utilization": *International Journal on Computer Science and Engineering (IJCSSE)*, January 2011 Vol. 3, Issue.01, 0975-3397
- [3] B. Jacob, L. Ferreira, N. Bieberstein, C. Gilzean. Enabling Applications for Grid Computing with Globus, Springer, CSA2011&WCC2011 Proceeding.
- [4] J. Xiao, D. Lin Survey of Security in Grid Services *The Fourth International Conference on Electronic Business ICEB, Beijing, May, 1st 200.*
- [5] P. Dabas A. Arya Grid Computing: An Introduction, *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 3, ISSN: 2277 128X, March 2013.
- [6] N. Thenmozhi¹, M. Madheswaran² Content Based Data Transfer Mechanism for Efficient Bulk Data Transfer in Grid Computing Environment" *International Journal of Grid Computing & Applications (IJGCA)* Vol.2, No.4, December 2011
- [7] X. Shi · H. Jin · S. Wu · W. Zhu · L. Qi Adapting grid computing environments dependable with virtual machines: design, implementation, and evaluations, *The Journal of Supercomputing An International Journal of High-Performance Computer Design, Analysis, and Use* ISSN 0920-8542 J Supercomput DOI 10.1007/s11227-011-0664-7.
- [8] M.A. Baker, R. Buyya, and D. Laforenza, The Grid: International Efforts in Global Computing, *SSGRR 2000 The Computer & eBusiness Conference, l'Aquila, Italy July 31. 2000 - August 6. 2000.*
- [9] C. Germain, V. Néri,... *Building an experimental platform for global computing* Grid2000 December 2000, IEEE Press.
- [10] Chervenak, I. Foster, C. Kesselman Salisbury, "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Data Sets", *J. Network and Computer Applications* (23). 187-200. (2001),
- [11] *Globus Toolkit 4.2.1 Download:*
<https://www.globus.org/toolkit/downloads/4.2.1/>.
- [12] *Globus Toolkit 4.2.1 Release Manuals:* <https://www.globus.org/toolkit/docs/4.2/4.2.1/>

- [13] S. Tarkoma Efficient Content-based Routing, Mobility-aware Topologies and Temporal Subspace Matching “, Report A-2, ISBN 952-10-3055-0, 2006.
- [14] H. Subramoni, P. Lai, R. Ketti, High Performance Data transfer in grid Environment using GridFTP over InfiniBand, *Proceedings of the 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, (May 2010).
- [15] Introduction to Grid Computing:
<http://www.redbooks.ibm.com/redbooks/pdfs/sg246778.pdf>
- [16] J. .Jacob1, R. E. Blessing, J. R. I. Ballasting. Impact of arrival rate of job / resource and Dual queues in the Matchmaking process in Grid Environment" *International Journal of Innovation and Applied Studies* ISSN 2028-9324 Vol. 2 No. 4 Apr. 2013, pp. 484-487 ©nnovative Space of Scientific Research Journals <http://www.issr-journals.org/ijias>
- [17] S. K. Patel; A. Seetha, A. K. Sharma, Grid Computing: Status of Technology In Current Perspective, *International Journal of software and hardware research in engineering*, ISSN NO. 234-4890, Volum 2 issue 6 June 2014

Authors Biography



Mohammed A.M. Ibrahim born in Taiz city, Yemen, received his BSc degree in computer science from Baghdad University, Iraq , in 1991, also he received his MSc in computer science & engineering from Shanghai University China, 1996, and his PhD degree in computer science and engineering from Shanghai Jiaotong University, China ,2003, his research interests include grid computing, cloud computing ,distributing computing ,network computing, network security.