# WXES3182
# LATIHAN ILMIAH II
# SEM II, 2005/2006

**Project Title:**
**DATA GRID**

**Supervisor:**
Dr. TEH YING WAH

**Moderator:**
Pn. NORJIHAN ABDUL GHANI


| | |
|---|---|
| STUDENT NAME: | ROSELINE LAU MING TAI |
| MATRIC NO. | WEK 030183 |

## Abstract

This document gives a comprehensive briefly about data grid, overview of the cluster system and access portal. This cluster system will support both in submitting job and analysis will seamlessly tie together resources.

Besides, this document focuses on the relation between the access portal and cluster Rock version 4.0. The project is aim to develop an example of access portal for the cluster. In this project, it is expected to archieve objectives including able to submit the job to the cluster, access portal that contain job submission form.

This document also contains a snapshot of the interfaces of the access portal. It has divided into few modules which are login, news, job submission, job status, upload files, search job, post messages and logout. There is different between normal user and admin, which admin more modules included upload files, user control and viewing reports. The functional requirement for each module will be explained further in this report. The hardware and software requirements for the implementation of this project have been considered too.

Today, data grid, cluster and access portal becoming more popular and important aspect either for company and organization to capture the expected level of complexity, the vision that worked out to considerable detail, even though some of these details are likely to be adjusted in future.

## Acknowledgements

First of all, I would like to express the deepest appreciation to my supervisor, Dr. Teh Ying Wah for his comment and meticulous supervision through the lengthy process in the development of the first part of this project. Thanks for his confidence with me to handle this project and for his advice, guiding, understanding and constructive criticism.

Thanks to my moderator of semester II, Mdm. Norjihan Abdul Ghani and moderator of semester I, Mdm. Chew Esyin for their advices and comments. Thanks to my teammates that accompanying along this project.

Besides, thanks also to Mdm. Junnaini Ismun from Faculty of Engineering for explaining on Cray & Dell Cluster, Mr. Kevin Johnson, master student of our faculty and expert in high performance computer especially about cluster and Mr. Khairul Ikmal from Open Source company that willing spend his time for advice and guidance.
And not forgotten, I am grateful that full support and understanding from my family.

Lastly, I would like to thank everyone either directly or indirectly who provided me the facilities in laboratory and greatly appreciate all the lecturers in FSKTM for their willingness to share their knowledge and experiences.

# TABLE OF CONTENTS

## LIST OF FIGURE

# Chapter 1 – Introduction

## 1.1 Project Background

Today, the world is moving towards the information technology age and Malaysia is heading towards that direction too. The rapid development and advancement in information technology had imposed radical changes in human life. In most organization, computer are used widely to simplify tasks. Information is no longer recorded using the manual system or traditional methods but it stored in databases and even can submit the jobs let them work on it.

Data grid is a large scale of data movement beyond computation and sharing of data across security realms. Besides, it needs access multiple data sets from more than one data provider. It has a power grid analogy with multiple power generators complex transmission networks with switching. It can be known as best movement of data across computer networks with access control, quotas, guarantee of data access and complex usage costing.

In this point, data grid is sharing and coordinated use of diverse resources in distributed "virtual organizations". This sharing has to happen across multiple, mutually distrustful administrative domains. It is mean if they working together on the same grants, they need to share hardware resources, application, data sets and results.

To able user to be easier to submit the job, the access portal has been suggested. Because most people are not familiar with the command line under the linux platform as most clusters are in linux platform.

Access portal have been designed and implemented based on the web service paradigm. The most important term is flexible and extensible service framework. It is link highly available commercial products.

*"Flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources."*

From *The Anatomy of the Grid* by Foster, Kesselman and Tuecke

(The expert of grid)

**1.2 Project Motivation**

Based on research, the usage of the access portal for cluster in data grid begins being popular among the large organization in European countries. Lots more organization especially universities in certain countries corporate to work on it. In Malaysia, it is still a new method to develop.

Due to that, this project is to develop access portal for job submission of the cluster in the data grid. This project is expected to let user submit the job in the easiest way.

**1.3 Project Objective**

The following points are the main objectives that need to be achieved in this project:

- Design and develop a access portal of mining2 according to requirements of Cray and Dell Cluster for Faculty of Engineering .

- Users can access this portal from any location through a standard web browser, and no additional software or configuration needed at client side.
- Provide a reliable portal for users to do job submission to the cluster.
- Increase knowledge flow within the organization to share important knowledge and information without fear of loss. It will help the users to get the information they want.
- To create an integrated system can reduce manual work.

### 1.4 Project Scope

#### 1.4.1 Project target user

This project is developing focus on access portal on the job submission for cluster in data grid. Due to the time constraint and limitation of resources, the project developed on the simplest among the portals from world wide.

In overview, the access portal will involve managerial, executive and operation level of an organization. All the users can access the portal and submit the jobs they wish to.

#### 1.4.2 Functional Scope

Basically the scopes of modules login, news, upload files, job submission, view job status, view reports, user controls and log out in access portal are as follow:-

i. Be able to submit jobs
   - Users can submit whatever the jobs they wish to cluster to settle for them including from calculation to stimulation.

ii. Be able to view the status

- Users able to view the job status whether the job is complete, delay, in process or fail.

iii. Be able to share knowledge

- Let the users able to retrieve information such as history of the job submitted.

iv. Be able to upload new job script (only by administrator)

- More job script can be added and more job can be running on the cluster

v. Be able to view reports or summary of the activities (only by administrator)

- Can be view overall reports about the activities and user's job submission

vi. Be able keep track of user activation (only by administrator)

- Able to keep track of who and what user currently do in the access portal

vii. Be able to sign up and edit profile

- Able to add new user and edit or update the user or administrator's profile

## 1.5 Project Constraints

Generally every development faces its own set of constraints. There are several major constraints have been identified when this project has been carried out.

### 1.5.1 Limited time

As can be referred in the project schedule earlier, roughly 2 semester has been given to finish the project, starting from the feasibility study phase till maintenance phase. It is not full time development, as others subject that need to be completed. So, the time allocated is really limited and tight.

### 1.5.2 Lack of experience

Next major constraints is due to inexperience. Theoretical and practical knowledge gained is not enough. Since the project is considered very new in Malaysia, a lot of times have to be spent on researches and studies on this project.

### 1.5.3 Limited human resources

Since the project is still very new Malaysia, there are lack of advisors and people can asked for advice. Even conference has been held and inviting the Professor from Thailand.

Apart from the major constraints mentioned, several minor constraints have been identified as well. Books about access portal for cluster in data grid is very limited and it is hard to get more information from the books. So, all the references have to be obtained from the internet.

Lastly, greater hardware and software requirements are needed in order to implement this project.

## 1.6 Expected outcome

The expected and planned outcome of this project will consists of login and logout, job submission and view status, sign up and edit profile, view reports, user controls and upload new files only by administrator.

Each of these modules is independent but interrelated. The login module is implemented as security checking to allow only the authenticated user to access the system.

With the modules that stated previously in system, it could help user to submit the jobs without to know how complex job submission directly to cluster. Besides, it could allow users to view on the status of the job that have been submitted.

Administrator can edit user's profile but user can only edit own profile. New user can sign up in the access portal. The summary reports and user controls to keep track user's activities can only be view and control by the administrator. And lastly, files can be uploaded including job script and user manuals to add up more job submission within the access portal.

## 1.7 Project Schedule

| | Task Activities Name | Duration | Start | Finish |
|---|---|---|---|---|
| 1 | Prelimeary Study and Planning | 3 wks | Mon 7/18/05 | Fri 8/5/05 |
| 2 | Milestones | 0 wks | Sun 8/7/05 | Sun 8/7/05 |
| 3 | Literature Review | 2 wks | Mon 8/8/05 | Fri 8/19/05 |
| 4 | Milestones | 0 wks | Sun 8/21/05 | Sun 8/21/05 |
| 5 | Methodology | 2 wks | Mon 8/22/05 | Fri 9/2/05 |
| 6 | Milestones | 0 wks | Sun 9/4/05 | Sun 9/4/05 |
| 7 | System Analysis | 4 wks | Mon 9/5/05 | Thu 9/29/05 |
| 8 | Milestones | 0 wks | Sun 10/2/05 | Sun 10/2/05 |
| 9 | System Design | 4 wks | Mon 10/3/05 | Fri 10/28/05 |
| 10 | Milestones | 0 wks | Sun 10/30/05 | Sun 10/30/05 |
| 11 | Prototyping | 6 wks | Mon 10/31/05 | Fri 12/9/05 |
| 12 | Milestones | 0 wks | Sun 12/11/05 | Sun 12/11/05 |
| 13 | Software Development and Codin | 8 wks | Mon 12/12/05 | Thu 2/2/06 |
| 14 | Milestones | 0 wks | Sun 2/5/06 | Sun 2/5/06 |
| 15 | System Testing | 4 wks | Mon 2/6/06 | Fri 3/3/06 |
| 16 | Milestones | 0 wks | Sun 3/5/06 | Sun 3/5/06 |
| 17 | Implementation and Maintainance | 5 wks | Mon 3/6/06 | Wed 4/5/06 |
| 18 | Milestones | 0 wks | Sun 4/9/06 | Sun 4/9/06 |
| 19 | Documentation | 37 wks | Sun 8/7/05 | Fri 4/7/06 |

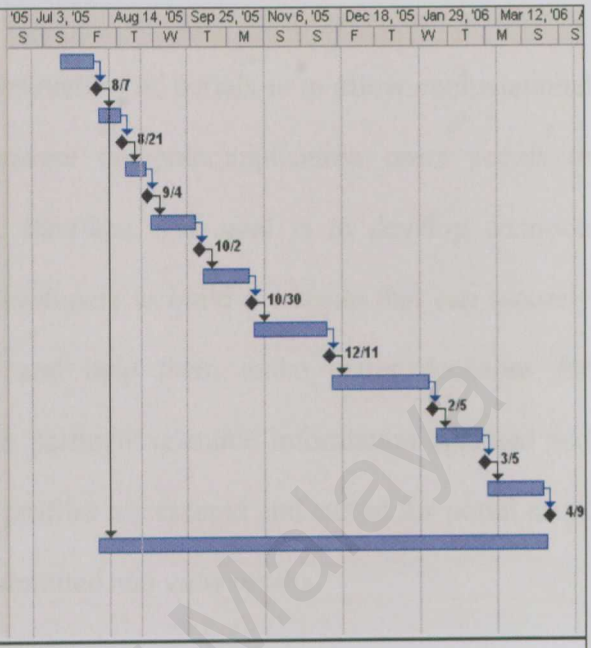



**Figure 1.1 Grantt Chart**

# Chapter 2 – Literature review

## 2.1 Introduction to access portal

One of the most exciting areas of the construction of portals is to allow computational scientists, researchers and high performance computer/application users access to resources via an easy to use web page interface. The goal is to develop common components that can be used by portal developers to build a website that can securely authenticate users to remote resources and help them make better decisions for scheduling jobs by allowing them to view pertinent resource information obtained and stored on a remote database. In addition, profiles are created and stored for portal users allowing them to track and monitor jobs submitted and view results.

## 2.2 Introduction to Rock Version 4.0 (cluster)

The National Partnership for Advanced Computational Infrastructure (NPACI) designed the NPACI Rocks toolkit in November 2000 to help simplify the building and management of clusters. Rocks known as a simple, self-contained, cluster-aware management system that is scalable and upgradeable. Then, it is slowly becoming the de facto cluster package. In October 2005, registered users of NPACI Rocks reported computational power that totals 149 trillion floating-point operations per second (TFLOPS).

As clusters expand, cluster deployment, maintenance, monitoring, and management can become complex and time-consuming processes. Before an application can tap the tremendous computing power available to an HPC cluster, administrators

must install and configure the cluster for monitoring and management. Manually deploying even a small HPC cluster is no small task. The job demands that each cluster node and its OS be identically configured and installed, including proper drivers, networks, parallel libraries, and management software.

### 2.3 Grid architecture

Architecture model and technology has been developed for the establishment and management of cross-organizational resource sharing. This new architecture, called grid architecture, identifies the basic components of a grid system. The grid architecture defines the purpose and functions of its components, while indicating how these components interact with one another. The main focus of the architecture is on interoperability among resource providers and users in order to establish the sharing relationships. This interoperability, in turn, necessitates common protocols at each layer of the architectural model, which leads to the definition of a grid protocol architecture as shown in figure 2.1.

This protocol architecture defines common mechanisms, interfaces, schema, and protocols at each layer, by which users and resources can negotiate, establish, manage, and share resources. Figure 1 shows the component layers of the grid architecture and the capabilities of each layer. Each layer shares the behavior of the underlying component layers.

The following describes the core features of each of these component layers, starting from the bottom of the stack and moving upward.

- Fabric layer

- The fabric layer defines the interface to local resources, which may be shared. This includes computational resources, data storage, networks, catalogs, software modules, and other system resources.

- Connectivity layer
  - The connectivity layer defines the basic communication and authentication protocols required for grid-specific networking-service transactions.
  - Authentication protocol which used in grid computing ensures only the authorized virtual organization user log in to portal.
  - There is some of the characteristic of authentication solution in grid.
    i. Single sign on
       - User just need log on to the portal once. After they log on, they can accept any resource which shred in system. There is no further user interaction.

    ii. Delegation
       - Application should be access the resources that are required. The resources are under user's authorization. Besides that, application should be able to conditionally delegate a subset of its right to online application

    iii. Integration with the various local security solution
       - Other security solution Kerberos and UNIX security also can integrate in the security function. The resource provides can be add or these local security solution to authentication users

    iv. User- based trust relationship

- User can be resources from multiple provides to cooperate or interact with each other in configuring the security environment

- Resource layer
  - This layer uses the communication and security protocols (defined by the connectivity layer) to control secure negotiation, initiation, monitoring, accounting, and payment for the sharing of functions of individual resources.
  - The resource layer calls the fabric layer functions to access and control local resources. This layer only handles individual resources, ignoring global states and atomic actions across the resource collection pool, which are the responsibility of the collective layer.
  - There are two primary classes in the resource layer protocol. First, information protocol is used to obtain information about the structure and state of a resource. For example, information of protocol is the configuration, current load and usage policy of a resource. Second, management protocols are used to negotiate access to shared resource. The negotiation included these following fields such as resource requirements and operation to be perform. Advanced reservation and quality of service are included in resource management.
- Collective layer
  - While the resource layer manages an individual resource, the collective layer is responsible for all global resource management and interaction with collections of resources. This protocol layer implements a wide variety of

sharing behaviors using a small number of resource-layer and connectivity-layer protocols.

- Responsible in data exchange between fabric layer resources. Hence the communication requirements are include transport, routing and naming. The communication will transfer the resources to actual location to excute applications

- Application layer
  - The application layer enables the use of resources in a grid environment through various collaboration and resource access protocols.

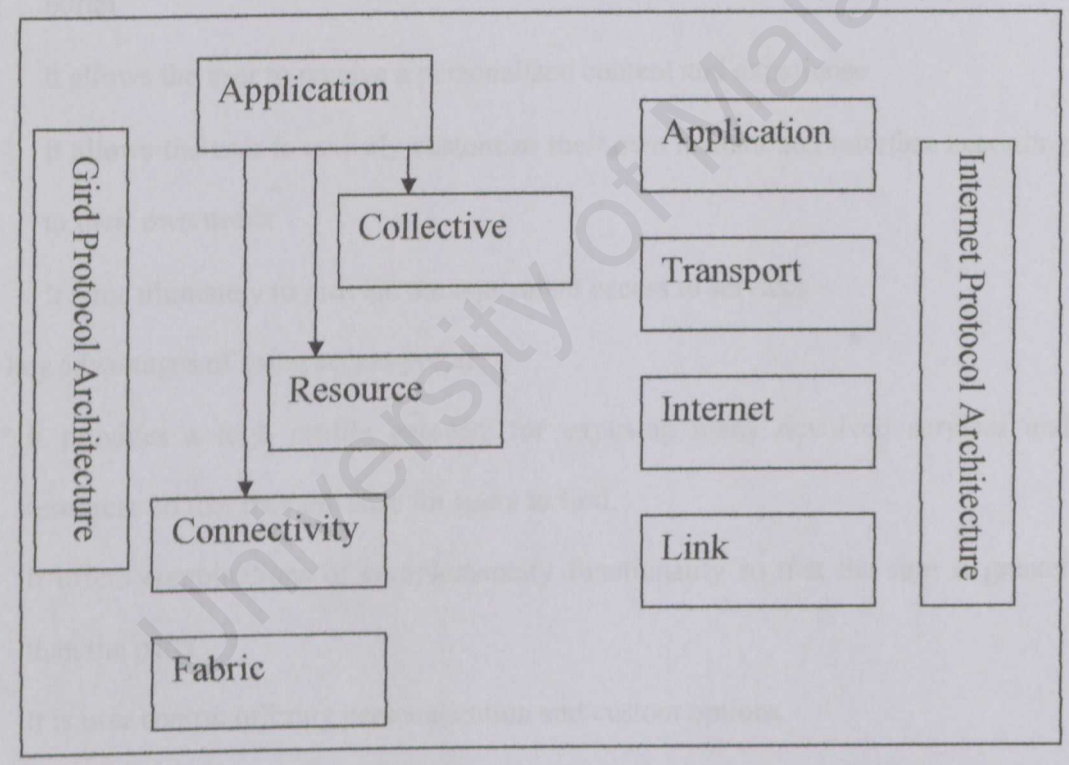



**Figure 2.1 Grid Architecture**

**2.4 WHY access portal?**

Commonly referred to as simply a portal, a Web site or service that offers a broad array of resources and services, such as job submission, forums, search engines, and on-line viewing status. It is understandable that some may feel that this initiative is perhaps technology driven; or adds little more than the current web intranet services; or is perhaps too oriented towards centralized management.

In much more simple terms this means:

- The portal supplies a secure gateway to access institutional systems and services.
- It brings together a collection of services and information that provided in the portal
- It allows the user to receive a personalized content and experience.
- It allows the user to actively customize their own content and interface according to their own needs
- It aims ultimately to provide the equivalent access to services

The key advantages of using access portal:

- It provides a high profile gateway for exposing many devolved services and resources so that they are easy for users to find.
- It offers combinations of complementary functionality so that the sum is greater than the parts.
- It is user centric offering personalization and custom options.
- It is built upon industry standards and helps promote shared institutional standards and quality.
- It offers new functionality that is not available elsewhere.

- It promotes greater service availability, moving towards time and user location independence.

## 2.5 Case study

### 2.5.1 PANDA GRID Portal

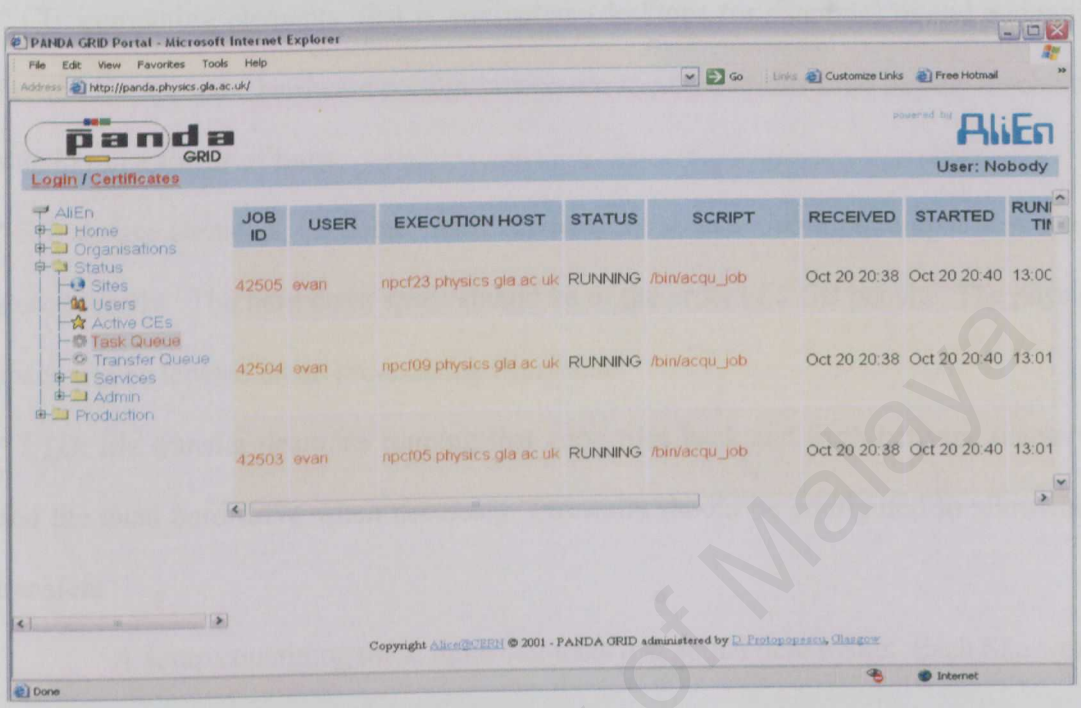



**Figure 2.2 Panda Grid Portal**

Considering the PANDA projected acquisition rate and the immediate need of large scale simulations of the PANDA detector, no single institution (not even GSI) will have the computing resources to process this amount of data in a timely manner. Therefore it is necessary to put together our resources in a coordinated way. The GRID is designed exactly for this purpose.

The PANDA GRID can be developed only with the contribution of the PANDA member institutions. Glasgow University provides the infrastructure as well as local resources. However, a GRID project implies pooling collaboration's resources in a coherent, structured manner.

Pooling resources means that each institution contributes their local resources and gets access to remote resources contributed by others.

Local resources mean:

* CE: computing elements, that is computers (desktops for example) or/and computing farms. No special hardware configurations are required at this time but the machine(s) should run a flavor of linux.

* SE: storage elements, i.e. some hard drive space so that job input/output files can be stored locally. The hard drive space should be of the order of 5 GB per site. The physical space can be located on an pre-existing filesystem.

* FTD: file transfer deamons running that copy files back and forth between the server and the local hard drive when necessary. Firewalls should be configured to allow these transfers.

A setup containing these three elements that called here a Site. Each Site would in principle correspond to a PANDA member institution.

A new Site to the AliEn-PANDA Grid can be added by a person who can perform administration tasks at the respective site location. The administration of a GRID Site can be done by one person and would require and estimated time of 5 hours/month. Heads of institutions should allocate the manpower to do this.

### 2.5.2 GENIUS Web Portal

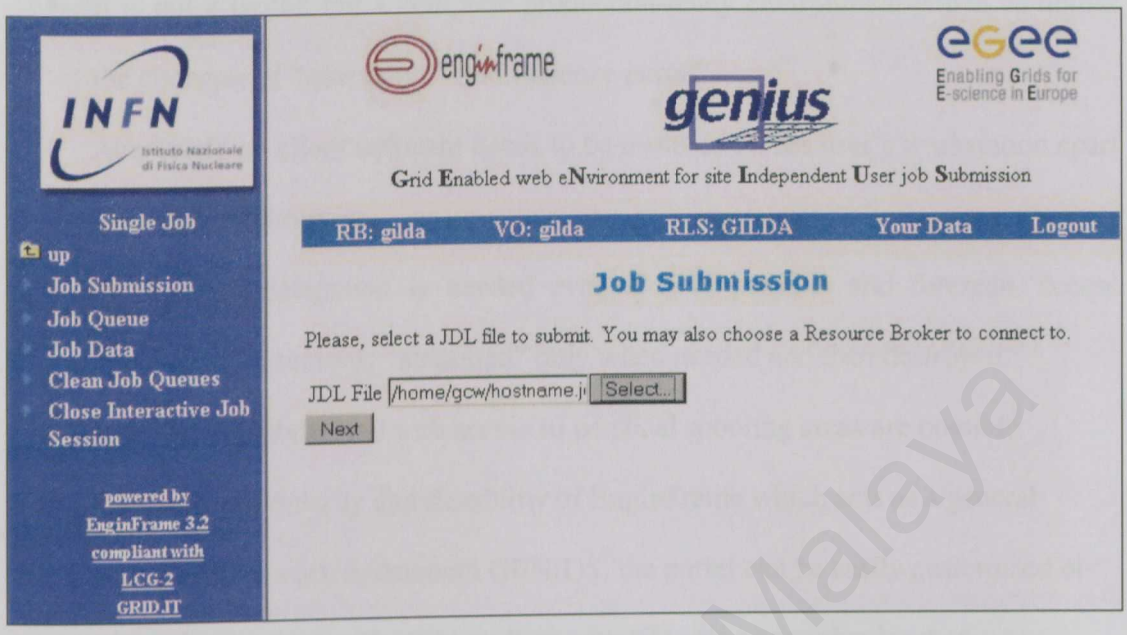



**Figure 2.3 GENIUS Web Portal**

GENIUS is the short name for Grid Enabled web eNvironment for site Independent User Job Submission. The EnginFrame have been used as a basis to build GENIUS. The user can seamlessly run jobs on the grid by using GENIUS job submission services.

The functions provided by the GENIUS web portal are:

- List the available computing elements (sites/queues) where a given job can run on.
- Submit a job to a Resource Broker (which can be selected by the user) or to a particular computing element.
- Monitor the current status of a job.
- Inspect on the flow the output of a job and save it both on the UI machine and on the local workstation.

GENIUS offers the following advantages when it is compared with other web portals available "on the market". The advantages include:

- It is not a toolkit but a complete production-ready environment which combines the concepts of "user portal" and "science portal".
- Absolutely no client software needs to be installed on the user's workstation apart from the web browser.
- No security delegation is needed even if it is possible and foreseen, access passwords are securely "streamed" only when needed and then destroyed.
- Interactive analysis and web access to personal spooling areas are possible.
- Due to the modularity and flexibility of EnginFrame which acts as a general purpose framework underneath GENIUS, the portal can be easily customized or adapted to interact with other grids or new VO using any other kind of middleware.

Disadvantages of GENIUS portal:

- User is stuck to the particular host and the functionality which might be provided by GENIUS is restricted by the browser capabilities.
- User must have an account on that host and store there (at least temporarily) all files retrieved from the grid after job execution. Since size of output files can be large, their remote browsing or editing could be difficult.

### 2.5.3 PACI HotPage Grid Computing Portal

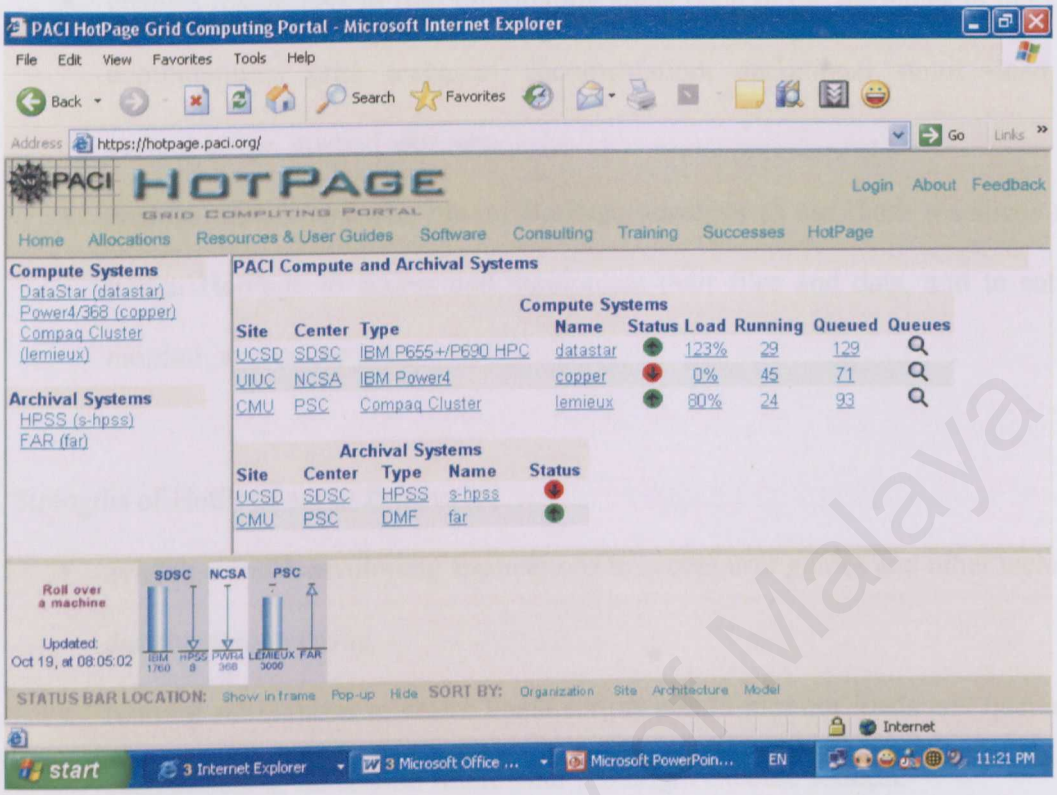



**Figure 2.4 PACI HotPage Grid Computing Portal**

In this PACI computational grid, it put together many resources in a coordinated way, including TeraGrid , SDSC, NCSA and PSC resources. There are four compute systems and two archival systems available for those resources in PACI. They are:

| Compute systems | Archival Systems |
|---|---|
| DataStar (datastar) | HPSS (s-hpss) |
| Power4/368 (copper) | FAR (far) |
| Compaq Cluster (lemieux) | |

Table 2.1 Compute systems and archival systems

HotPage is designed to make the computational science process easier. The functions provided by the HotPage are as below:

- enables researchers to find information about each of the resources in the PACI computational grid: technical documentation, operational status, load and current usage, queued jobs, etc.
- enables researchers who obtain HotPage accounts to use these resources from within HotPage to access and manipulate their files and data, and to submit, monitor, and delete jobs.

Strengths of HotPage are as below:

- available while developing applications to access user guides and other technical documentation easily.
- running simulations to create batch scripts and to monitor loads and queues on each system to determine where your job might execute soonest.
- easy, point-and-click interface to upload and edit files, submit jobs to any system, and download data or archive results.

Weaknesses of HotPage are as below:

- unable to manage allocations and track usage, store files in a single logical file space.
- metascheduler not available to simplify the process of submitting jobs to any and all available compute platforms.

## 2.6 Computing technology

### 2.6.1 Distributed computing

Distributed computing is parallel computing using multiple independent computers communicating over a network to accomplish a common objective or task. The type of hardware, programming languages, operating systems and other resources may vary drastically. It is similar to computer clustering with the main difference being a wide geographic dispersion of the resources.

There are many different types of distributed computing systems and many challenges to overcome in successfully designing one. The main goal of a distributed computing system is to connect users and resources in a transparent, open, and scalable way. Ideally this arrangement is drastically more fault tolerant and more powerful than many combinations of stand-alone computer systems.

Openness is the property of distributed systems such that each subsystem is continually open to interaction with other systems (see references). Web Services protocols are standards which enable distributed systems to be extended and scaled. In general, an open system that scales has an advantage over a perfectly closed and self-contained system.

Consequently, open distributed systems are required to meet the following challenges:

i. Monotonicity
   - Once something is published in an open distributed system, it cannot be taken back.

ii. Pluralism

- Different subsystems of an open distributed system include heterogeneous, overlapping and possibly conflicting information. There is no central arbiter of truth in open distributed systems.

iii. Unbounded non determinism

- Asynchronously, different subsystems can come up and go down and communication links can come in and go out between subsystems of an open distributed system. Therefore the time that it will take to complete an operation cannot be bounded in advance (see unbounded non determinism).

### 2.6.2 Parallel Computing

Parallel computing is the simultaneous execution of the same task (split up and specially adapted) on multiple processors in order to obtain results faster. The idea is based on the fact that the process of solving a problem usually can be divided into smaller tasks, which may be carried out simultaneously with some coordination.

There are many different kinds of parallel computers (or "parallel processors"). They are distinguished by the kind of interconnection between processors and between processors and memories. It also classifies parallel (and serial) computers according to whether all processors execute the same instructions at the same time (single instruction/multiple data -- SIMD) or each processor executes different instructions (multiple instruction/multiple data -- MIMD). Parallel processor machines are also divided into symmetric and asymmetric multiprocessors, depending on whether all the

processors are capable of running all the operating system code and, say, accessing I/O devices or if some processors are more or less privileged.

There are also a variety of architectures that have been developed to accomplish parallel processing. One such example would be a "Ring Architecture" where processors are linked by a ring structure, allowing all processors to read and write data simultaneously, therefore accomplishing a true parallel processing environment. Using such architecture with a large grain set of processors, has demonstrated processing speed improvements of 10 times or more over classical computing.

While a system of n parallel processors is less efficient than one n-times-faster processor, the parallel system is often cheaper to build. Parallel computation is an excellent solution for tasks which require very large amounts of computation, have time constraints on completion and especially for those which can be divided into n execution threads. In fact, in recent years, most high performance computing systems, also known as supercomputers, have a parallel architecture.

## 2.7 Operating System

### 2.7.1 Microsoft Windows XP Professional

Microsoft Windows XP Professional is the new version of windows that bring user PC to life. Windows XP packs the punch user for top performance and enjoyment. User can get all the power and dependability of Microsoft's most advanced operating system technology. And to ensure that version of Windows XP is always the very latest, automatic updates are available at command from the Microsoft Web site. Windows XP delivers greater dependability and increased efficiency.

Furthermore, Windows XP brings a brand-new visual style to our computer that known as friendly, fresh, and streamlined. User will notice cleaner lines, richer colors, smarter organization, and easier ways to get where he/she want go and do what he/she wants to do. The simplified Start menu automatically puts user most frequently-used programs in plain view, never more than one click away. Besides, this operating system will find out everything user wants to know about Windows XP and computer. The all-new support center in Windows XP is user one-stop shop for:

- Clear how-to instructions.
- Engaging start-to-finish articles.
- Troubleshooting advice.
- Free product updates.

**2.7.2 Linux**

Linux is a powerful, non-proprietary, standards-based operating system that is currently the fastest growing computer operating system on the planet. Linux offers speed, performance, stability, and reliability.

Much of the success of Linux is due not only to its outstanding performance, but also by the fact that it is a non-commercial, non-proprietary, open source product. Much like the Internet, Linux brings with it a freedom and openness that has not been seen in the computer industry in many years. No longer are large, domineering companies like Microsoft dictating what users can and cannot do with their computers. No longer are computer users being treated as pawns in Microsoft's unethical (and illegal) practices of predatory capitalism. No longer are computer users the slaves of greed, aggregate wealth,

or the whims of a single, domineering software company whose only concern is the "bottom line". With Linux, there is no "bottom line". Instead, success is measured by performance and design excellence.

## 2.8 Programming language

### 2.8.1 ASP. NET

ASP.NET is a set of web development technologies marketed by Microsoft. Programmers can use it to build dynamic web sites, web applications and XML web services. It is part of Microsoft's .NET platform and is the successor to Microsoft's Active Server Pages (ASP) technology.

ASP.NET attempts to simplify developers' transition from Windows application development to web development by offering the ability to build pages composed of *controls* similar to a Windows user interface. A web control, such as a button or label, functions in very much the same way as its Windows counterpart: code can assign its properties and respond to its events. Controls know how to render themselves: whereas Windows controls draw themselves to the screen, web controls produce segments of HTML and JavaScript which form part of the resulting page sent to the end-user's browser. It encourages the programmer to develop applications using an event-driven GUI paradigm, rather than in conventional web-scripting environments like ASP and PHP. The framework attempts to combine existing technologies such as JavaScript with internal components like "Viewstate" to bring persistent (inter-request) state to the inherently stateless web environment.

ASP.NET uses the .NET Framework as an infrastructure. The .NET Framework offers a managed runtime environment (like Java), providing a virtual machine with JIT and a class library.

### 2.8.2 Java

Java is an object-oriented programming language developed by James Gosling and colleagues at Sun Microsystems in the early 1990s. The language, which was designed to be platform independent, is a derivative of C++ with a simpler syntax, a more robust runtime environment and simplified memory management. Java is not related to JavaScript, though they have similar names and share a C-like syntax.

Java is a general purpose programming language with a number of features that make the language well suited for use on the World Wide Web. Small Java applications are called Java applets and can be downloaded from a Web server and run on your computer by a Java-compatible Web browser, such as Netscape Navigator or Microsoft Internet Explorer.

### 2.8.3 C# programming language

C# is an object-oriented programming language developed by Microsoft as part of their .NET initiative. C# has a procedural, object oriented syntax based on C++ that includes aspects of several other programming languages (most notably Delphi, Visual Basic, and Java) with a particular emphasis on simplification (fewer symbolic requirements than C++, fewer decorative requirements than Java).

C# is, in some senses, the programming language which most directly reflects the underlying Common Language Runtime (CLR) on which all .NET programs run, and it depends strongly on this framework because it was designed specifically to take advantage of the features that the CLR provides. Most of C#'s intrinsic types all correspond to value-types implemented by the .NET Framework. A common misbelief is that they are garbage-collected, though they are not; they are true value-types and are stack allocated.

# Chapter 3 – Methodology

This chapter includes the description about the techniques and procedures that are used to gather the system requirement and will specify the justifications for the chosen methodology for the project.

## 3.1 Methodology

A methodology is collection of methods, procedures, techniques, tools and paradigms for solving a class of problem. It represents a particular approach or philosophy for building system.

## 3.2 The phases Software Development Life Cycle

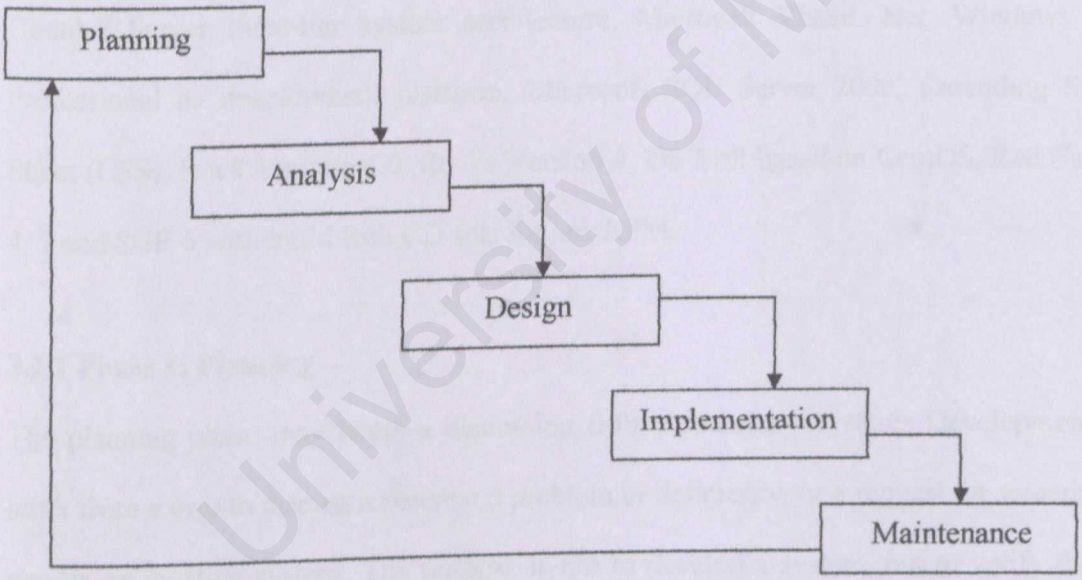



**Figure 3.1: A traditional Waterfall SDLC**

The SDLC methodology with prototyping is used to develop this system. It is a creation of the system and stimulation of all aspect of the system. SDLC methodology

support different degrees of prototyping that offer the end- user's abilities to review all aspects of the user interfaces and the structures of the documentation and reports before the system is generated. First comes to planning phase, then the analysis phase and follow by design phase. We use this methodology because it is very powerful and stated that what should we think about and what is being built. Besides, it is planning how it built should and force discipline process (analysis, planning and design) to avoid the pressures of developing the system before it is known what is built or any action to be taken. Besides, it will lower the risk to implement easy resource intensive and maintenance after we design the system. So, we can precisely define our system users' requirement and can be done much easier to build the cluster and access portal. This system needs special development tools and language such as Microsoft Internet Information Server (IIS), Client / Server three-tier system architecture, Microsoft Visual. Net, Windows XP Professional as development platform, Microsoft SQL Server 2000, Cascading Style Sheet (CSS), Rock Version 4.0, Rocks Version 4, OS Roll based on CentOS, Red Hat 3. 4. 3 and SGE 6 with build Roll CD sets for any RPM.

### 3.2.1 Phase 1: Planning

The planning phase may begin a discussion from a members Systems Development, a letter from a user to discuss a perceived problem or deficiency, or a request for something new in an existing system. The purpose is not to develop a system, but to verify that a problem or deficiency really exists, or to pass judgment on the new requirement. This phase is typically very short, usually not more than a day or two for a project. The end result, or deliverable, from the planning is either a willingness to proceed further, or the decision to 'call it quits'.

There are three factors, typically called constraints, which result in a 'go' or 'no-go' decision:

- **Technical**. The project cannot be completed with the technology currently in existence.
- **Time**. The project can be completed, but not in time to satisfy the user's requirements. This is a frequent reason for the abandonment of the project after the planning phase.
- **Budgetary**. The project can be completed, and completed on time to satisfy the user's requirements, but the cost is prohibitive.

When an agreement has been made to continue with the project, the second phase of the SDLC is implemented – The Analysis Phase.

### 3.2.2 Phase 2: Analysis

In the Analysis Phase, (sometimes called the Data Gathering Phase) we study the problem, deficiency or new requirement in detail. We have gathered enough information in Phase 1 of the SDLC to begin programming, the SDLC dictates that Phase 2 should be completed before any actual writing of the program begins. At the end of phase 2, the requirements statement have been in development: this provides details about what the system should do. It can easily form the basis of a contract between the user and system builder. The requirements statement will list all of the major details of the program.

### 3.2.3 Phase 3: Design

With a good design, find myself going back to modify pieces of code that already written will be reduced dramatically. The end result is a system that will behave in the way it was intended, and will generally have with a shorter overall program development time.

Design in the SDLC encompasses many different elements. Here is a list of the different components that are 'designed' in this phase:

- Input
- Output
- Processing
- Files

By the end of the design phase, a formal Requirements Statement and a rough sketch of what the user interface will be produced.

### 3.2.4 Phase 4: Implementation

In the implementation phase, the project reaches fruition. After the design phase of the SDLC is complete, the system is implemented. Any hardware that has been purchased will be delivered and installed software. User that will be using the program will also be trained during this phase of the SDLC.

During the Implementation phase, both hardware and software is tested. Although most problems found and fix, the user will uncover. This leads on to the sixth and final stage.

### 3.2.5 Phase 5: Maintenance

Phase 5 of the SDLC is the maintenance Phase. In this phase someone (usually the client, but sometimes a third party such as an auditor) studies the implemented system to ensure that it actually fulfills the requirements statement. Most important, the system should have solved the problem or deficiency, or satisfied the desire that was identified in phase 1 of the SDLC .

More than a few programs and systems have been fully developed that, for one reason or another, simply never met the original requirements. The Maintenance portion of this phase deals with any changes that need to be made to the system.

Changes are sometimes the result of the system not completely fulfilling its original requirements, but they could also be the result of user satisfaction.

# Chapter 4 – System Analysis

In this chapter, I focus on system analysis which including the requirement elicitation, requirement analysis, development tools considerations and justification. Meanwhile, the hardware and software requirement are also considered in order to come out with suitable and compact act of development tools and environment.

## 4.1 Introduction to system analysis – Rock Version 4.0 and Access Portal

System analysis is a problem- solving technique that decomposes a system into its component pieces for the purpose of studying how well those component part work and interact to accomplish their purpose.

The purposes are:

- Identify user's need
- Evaluate the system concept for feasibility study
- Allocate functions to hardware, software, people, database and others system elements

## 4.2 Techniques used to define requirements

Requirement elicitation is a critical part of the process. Effective and appropriate techniques must be used to define and elicit users' requirements. Several techniques need to be applied in order to get a complete requirement. The following are the techniques:-

### 4.2.1 Internet research

Internet is the main resource in searching materials especially when doing the literature review, some system solution in the web has been visited and analyzed. It is the fastest, easiest and most effective method in getting the latest information. The existing system also only can be found by using internet.

### 4.2.2 Trial on the existing system

There are some existing access portal for certain data grid offered free trial on job submission. So, I can try on how it works and have ideal on it. From the experience of using it, I can see and judge the system from the user's perspective.

### 4.2.3 Brainstorming and discussion

After getting all the information needed, then it is the time for brainstorming that includes supervisor and members that involved in this projects. It is very importance as it sense of formulating better or enhance system requirements, system architecture and design, implementation techniques and solve the ambiguities that faced in this project.

## 4.3 Operating System Platform

### 4.3.1 Rocks 4.0 cluster – Red Hat Linux

Linux (also known as GNU/Linux) is a computer operating system. It is one of the most prominent examples of free software and of open-source development; unlike proprietary operating systems such as Windows, all of its underlying source code is available to the public for anyone to freely use, modify, improve, and redistribute.

As used here, the term "Linux" describes a complete Unix-like operating system based on a combination of the Linux kernel and components from the GNU Project and elsewhere.

A Linux distribution bundles many applications with the core system, and can provide simplified installation and upgrades. Desktop environments such as GNOME and KDE are sometimes associated with Linux and referred to as integral components of the system, though they are independent packages that function on several Unix-like systems. Initially, Linux was primarily developed and used by individual enthusiasts. Since then, Linux has gained the support of major corporations such as IBM, Sun Microsystems, Hewlett-Packard, and Novell for use in servers and is gaining popularity in the desktop market[1]. Proponents and analysts attribute this success to its vendor independence (the opposite of vendor lock-in), low cost, security, and reliability.

Red Hat Linux was one of the most popular Linux distributions, assembled by Red Hat. It is one of the "middle-aged" Linux distributions; 1.0 was released in November 3, 1994. It is not as old as Slackware, but certainly older than many other distributions. It was the first Linux distribution to use RPM as its packaging format, and over time has served as the starting point for several other distributions, such as the desktop-oriented Mandriva Linux (originally Red Hat Linux with KDE), Yellow Dog Linux (which started from Red Hat Linux with PowerPC support), and ASPLinux (Red Hat Linux with better non-Latin character support).

#### 4.3.2 Access Portal - windows

Microsoft Windows is a series of popular proprietary operating environments and operating systems created by Microsoft for use on personal computers and servers.

Microsoft first introduced an operating environment named *Windows* in November, 1985, as an add-on to MS-DOS. This was in response to Apple Computer's computer system, the Apple Macintosh, which used a graphical user interface (GUI). Microsoft Windows eventually came to dominate the world personal computer market with market analysts like IDC estimating that Windows has around 90% of the client operating system market. All recent versions of Windows are fully-fledged operating systems.

The widespread use of Microsoft's operating system has benefited from not being tied to the success of one hardware manufacturer and from Microsoft's willingness to license the operating system to manufacturers. This is in contrast with Apple Computer, which does not license Mac OS X to other manufacturers, and Sun, which did not license Solaris before it was made free and open-source. However, the wide spectrum of possible hardware permutations with Microsoft Windows is also a major source of computer problems because of hardware-software incompatibilities for consumers.

### 4.3.3 Linux Vs Windows

Apart from debates over security, application availability, and open source software vs proprietary software arguments, many computing professionals debate the relative technical merits of Linux vs Windows. Directy comparisons are difficult for a number of reasons. Primarily, due to the closed source nature of Windows and open source nature of Linux, it is much easier to gain a greater understanding of the inner workings of Linux as compared to Windows. This means that, when comparing the way these two operating systems perform various tasks, some amount of reverse engineering, benchmarking, and guesswork must be used to determine how and why Windows performs certain tasks.

- kernel space vs user space

Microsoft has often been criticized for including many components, such as Internet Explorer and Windows Media Player, into the Windows kernel. Microsoft and Windows proponents claim that these libraries are a necessary part of the Operating System, and should therefore be included in the Kernel.

Linux proponents claim that the kernel should include only what absolutely must run in kernel space, such as code that must communicate directly with the hardware or file system, and that all other code should be run in user space.

- Memory management

Linux proponents claim that Linux's memory manager is better at avoiding swapping and allocating memory to programs.

Windows proponents claim that Windows use of the swap file offers better performance because it better anticipates what memory will not be used for some time, and can better load into memory what the user is most likely to soon need. But in Windows approach, claiming that use of the swap file should be avoided when possible, because of the bottleneck encountered during reading and writing to the hard disk. The Windows memory manager is also criticized for memory leaks.

- Stability

Linux proponents claim that Windows has severe problems with stability, often citing the once common Blue Screen Of Death. Both Windows and Linux proponents agree that a common source of instability is due to bugs in various device drivers. Linux users often claim that Windows itself lacks stability, and claim that reboots required for driver installation and security patches, as well as problems with Windows

memory management and instability in the Windows API cause the overall system to lack the Stability of Linux.

- Device Drivers

Since Device drivers run in kernel space, unstable drivers can cause system instability in Windows.

Linux runs many drivers in user space, which Linux proponents claim allows Linux systems to remain stable when a driver crashes.

### 4.4 Database Management System

#### 4.4.1 MySQL – Linux

MySQL is a multithreaded, multi-user, SQL Database Management System (DBMS) with an estimated six million installations. MySQL AB makes MySQL available as free software under the GNU General Public License (GPL), but they also dual-license it under traditional proprietary licensing arrangements for cases where the intended use is incompatible with the GPL.

- Architecture

MySQL Cluster has a few important concepts behind its design, which give both benefits and disadvantages.

- Replication

NDB uses synchronous replication through a two-phase commit mechanism in order to guarantee that data are written to multiple nodes upon committing the data. Normally there are two copies of the data in the cluster, however it can be configured to allow one to four copies at any single time.

- Horizontal Data Partitioning

Data within NDB tables are automatically partitioned across all of the data nodes in the system. This is done based on a hashing algorithm based on the PRIMARY KEY on the table. This is totally transparent to the end application.

- In-memory Storage

NDB currently stores all data and indexes in memory. NDB only writes the data to disk asynchronously. The reason it can do this safely is due to the first point above, replication.

- Shared Nothing

NDB has no single point of failure by design. Assuming a normal setup, any single node, system, or piece of hardware can fail without the entire cluster failing.

### 4.4.2 Microsoft SQL Server – windows

Microsoft SQL Server is a relational database management system (RDBMS) produced by Microsoft. It supports Microsoft's version of Structured Query Language (SQL), the most common database language. It is commonly used by businesses for small- to medium-sized databases, and in the past five years, even some larger enterprise databases.

MS SQL Server uses a variant of SQL called T-SQL, or Transact-SQL, an implementation of SQL-92 (the ISO standard for SQL, certified in 1992) with some extensions. T-SQL mainly adds additional syntax for use in stored procedures, and affects the syntax of transaction support. (Note that SQL standards require Atomic, Consistent, Isolated, Durable or "ACID" transactions.) MS SQL Server and Sybase/ASE both

communicate over networks using an application-level protocol called Tabular Data Stream (TDS). The TDS protocol has also been implemented by the FreeTDS project in order to allow more kinds of client applications to communicate with MS SQL Server and Sybase databases. MS SQL Server also supports Open Database Connectivity (ODBC).

### 4.5 Functional Requirement

Functional requirement are statement of services the system should provide, how the react to particular inputs and how the system should behave in particular situation. In some cases, it also stated what the system should not do.

### 4.6 Non Functional Requirement

Non-functional characteristics related to interaction which would be important for user interface design. The characteristics are:

- User friendly interface: The Command Prompt mode of cluster Rocks version 4 does not have any user interface. So, the meaningful icon or caption in the access portal makes user easily to traverse through it.
- Accuracy: It expected can perform properly when user submits a job. The outcome is expected logic correct and free errors.
- Reasonable response time: the access portal can retrieve the information within reasonable interval time. Besides, it can produced the result in the minimum time while the job is running within the compute node in the cluster.

- Modularity: Separate information into different modules which have distinct functionality in the access portal as well as cluster. This could make the developers easy to test and maintain.
- Reusability, expandability and maintainability: using the flexible of coding and can be reuse for the purpose of revision upgrade.
- Security: the most important term in the system that moderate level of security protection and access portal should be implemented.

# Chapter 5 – System Design

## 5.1 System functionality design

t enable data-bound controls to be more intelligent at runtime, their design-time counterparts allow the corresponding data-bound control designers to be smarter on the design surface.

### 5.1.1 Context diagram

n process terms, the context contains other processes that provide inputs and outputs to nd from the process in question

### 5.1.2 Data Flow Diagram (DFD)

#### 5.1.2.1 Data flow diagram principle

- The general principle in Data Flow Diagramming is that a system can be decomposed into subsystems, and subsystems can be decomposed into lower level subsystems, and so on.
- Each subsystem represents a process or activity in which data is processed. At the lowest level, processes can no longer be decomposed.
- Each 'process' (and from now on, by 'process' we mean subsystem and activity) in a DFD has the characteristics of a system.
- Just as a system must have input and output (if it is not dead), so a process must have input and output.

# Chapter 5 – System Design

## 5.1 System functionality design

It enable data-bound controls to be more intelligent at runtime, their design-time counterparts allow the corresponding data-bound control designers to be smarter on the design surface.

### 5.1.1 Context diagram

In process terms, the context contains other processes that provide inputs and outputs to and from the process in question

### 5.1.2 Data Flow Diagram (DFD)

#### 5.1.2.1 Data flow diagram principle

- The general principle in Data Flow Diagramming is that a system can be decomposed into subsystems, and subsystems can be decomposed into lower level subsystems, and so on.
- Each subsystem represents a process or activity in which data is processed. At the lowest level, processes can no longer be decomposed.
- Each 'process' (and from now on, by 'process' we mean subsystem and activity) in a DFD has the characteristics of a system.
- Just as a system must have input and output (if it is not dead), so a process must have input and output.

- Data enters the system from the environment; data flows between processes within the system; and data is produced as output from the system

**5.1.2.2 Database flow diagram notation included**

i. Process

- A process transforms incoming data flow into outgoing data flow

ii. External entities/ external agent

- External entities are objects outside the system, with which the system communicates.
- External entities are sources and destinations of the system's inputs and outputs

iii. Data store

- Data store are repositories of data in the system. They are sometimes also referred to as files.

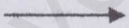
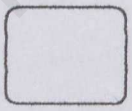
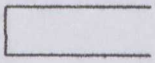
| Symbols | Attribute |
|---|---|
|  | Entity |
|  | Flow of data |
|  | Process |
|  | Data source |

**Table 5.1: Data flow diagram symbols and attributes**

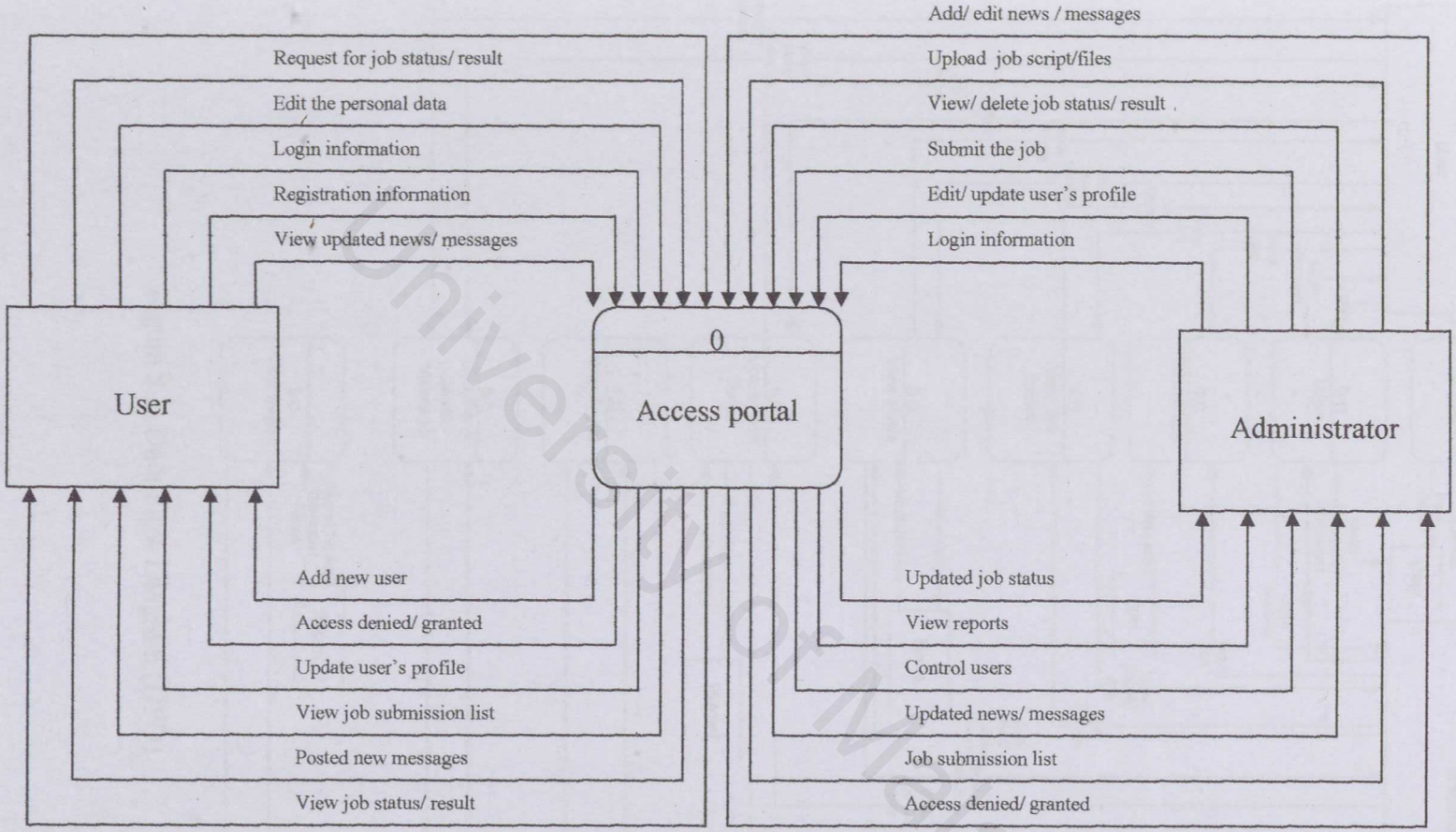



**Figure 5.1: Context Diagram for access portal**

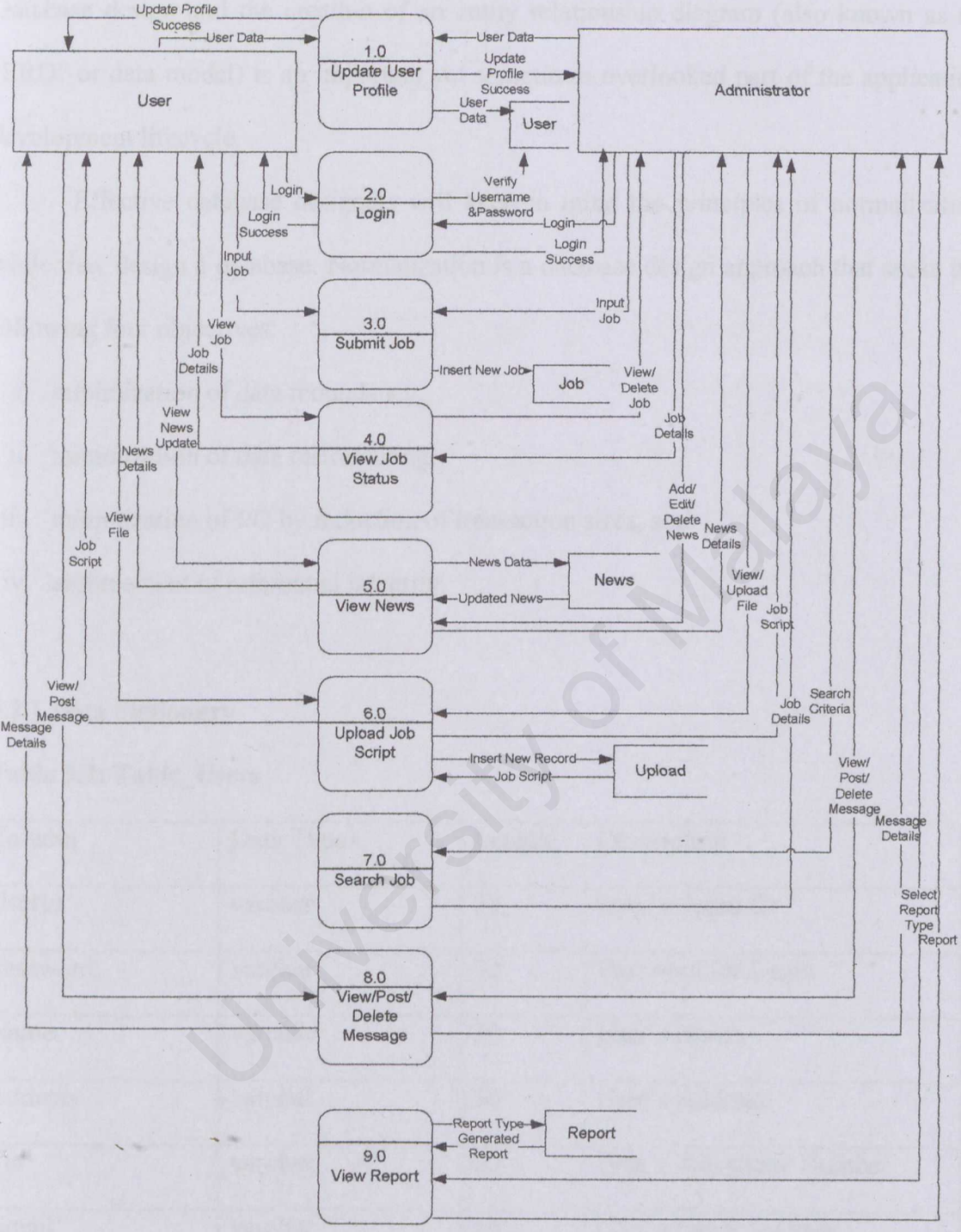



Figure 5.2: Data Flow Diagram (DFD)

## 5.2 Database design

Database design and the creation of an entity relationship diagram (also known as an "ERD" or data model) is an important yet sometimes overlooked part of the application development lifecycle.

Effective database designers will keep in mind the principles of normalization while they design a database. Normalization is a database design approach that seeks the following four objectives:

i. minimization of data redundancy,
ii. minimization of data restructuring,
iii. minimization of I/O by reduction of transaction sizes, and
iv. enforcement of referential integrity.

### 5.2.1 Data dictionary

**Table 5.2: Table_Users**

| Column | Data Type | Length | Description |
|---|---|---|---|
| UserId | varchar | 32 | User's Login ID |
| Password | varchar | 32 | Password for Login |
| Name | varchar | 50 | User's Name |
| Address | varchar | 50 | User's Address |
| Tel | varchar | 50 | User's Telephone Number |
| Email | varchar | 50 | User's Email Address |
| Role | varchar | 50 | User's Role (admin,student,etc) |

Description: This table holds all the information about the administrator and registered users which includes their UserId, Password, Name, Address, Tel, Email and Role. The primary key for this table is the UserId.

**Table 5.3: Table_JobSubmission**

| Column | Data Type | Length | Description |
|---|---|---|---|
| JobId | int | 4 | Job's Id |
| JobName | varchar | 128 | Job's Name |
| JobDescription | varchar | 128 | Job's Description |
| Nodes | int | 4 | Number of Nodes |
| SubmissionDate | datetime | 8 | Job's Submission Date |
| Status | varchar | 16 | Job's Status |
| UserId | varchar | 32 | User's Login ID |

Description: This table holds all the information about the jobs submitted by the users which includes their JobId, JobName, JobDescription, Nodes, SubmissionDate, Status, and UserId. The primary key for this table is the JobId.

**Table 5.4: Table_Message**

| Column | Data Type | Length | Description |
|---|---|---|---|
| MessageId | int | 4 | Message's Id |
| UserId | varchar | 50 | User's Login ID |
| Email | varchar | 50 | User's Email Address |

| Date | datetime | 8 | Message's Posted Date |
|---|---|---|---|
| Subject | varchar | 256 | Message's Subject |
| Message | varchar | 1024 | Message's Content |
| Admin_Reply | varchar | 1024 | Message Reply from Admin |

Description: This table holds all the information about the message posted by the admin and other users which includes their MessageId, UserId, Email, Date, Subject, Message, and Admin_Reply. The primary key for this table is the MessageId.

**Table 5.5: Table_News**

| Column | Data Type | Length | Description |
|---|---|---|---|
| NewsId | int | 4 | News's Id |
| Date | datetime | 8 | News's Uploaded Date |
| News | varchar | 1024 | News's Content |

Description: This table holds all the information about the news uploaded by the admin which includes their NewsId, Date, and News. The primary key for this table is the NewsId.

**Table 5.6: Table_Upload**

| Column | Data Type | Length | Description |
|---|---|---|---|
| UploadId | int | 4 | File's Upload Id |
| Title | varchar | 80 | File's Title |

| | | | |
|---|---|---|---|
| Link | varchar | 80 | File's Link |
| Type | datetime | 80 | File's Type |

Description: This table holds all the information about the file upload by the admin which includes their UploadId, Title, Link, and Type. The primary key for this table is the UploadId.

### 5.2.2 Database diagram

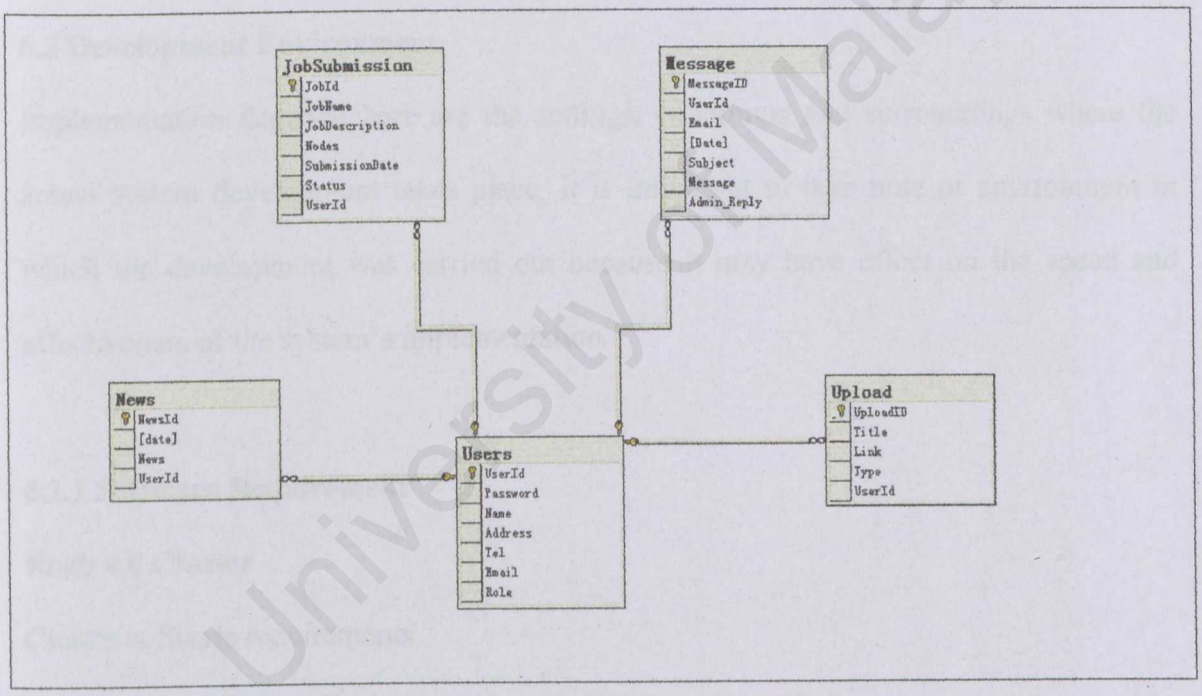



**Figure 5.3 Database diagram**

# Chapter 6 – Implementation

## 6.1 Introduction

An implementation is the practical application of a methodology or algorithm to fulfill a desired purpose. The team builds the components either from scratch or by composition. Given the architecture document from the design phase and the requirement document from the analysis phase, the team should build exactly what has been requested, though there is still room for innovation and flexibility.

## 6.2 Development Environment

Implementation depicted here are the settings, conditions and surroundings where the actual system development takes place. It is important to take note of environment in which the development was carried out because it may have effect on the speed and effectiveness of the system's implementation.

### 6.2.1 Software Requirement

**Rock 4.0 Cluster**

Cluster software requirements

- Cluster software should work with CERN RedHat Linux
- No cluster software with a modified kernel (OpenMosix, Scyld)

Because of a favorable review and positive experiences within CMS NPACI Rocks was chosen as the cluster software instead of OSCAR.

Compilers to install:

- C/C++ and F77/F90 compilers by Intel and Portland Group

Grid tools to install:

- NorduGrid software
- Large Hadron Collider Grid software

Application software to install:

- Compact Muon Solenoid software
- Compact Muon Solenoid production software,
- Molecular dynamics simulation package PARCAS

Platform that contain in Rocks:

- Red Hat® Enterprise Linux® in the Enterprise Edition or CentOS Linux® in Standard Edition
- Ganglia
- Rocks-dist, 411
- mySQL® database
- Multiple MPI libraries (MPICH, MPICH-GM, MVAPICH)
- Platform Lava
- Apache
- Tomcat
- PVFS
- HPL (High Performance Linpack)
- Other miscellaneous tools and applications.

**Access Portal**

Table 6.1 shows the software specification that was needed during the system's implementation. They are essential towards the development of the system. For example, Internet Information Service 6.0 must be installed as web server because the system ultilized a web application as the interface for the user to retrieve and view info from the database. The Microsoft Internet Information Services 5.1 is one of Microsoft products. Hence it is compatible in the .NET environment. Furthermore, since the development is carried out in a .NET environment, thus the Microsoft .NET Framework must be installed before system implementation can be carried out.

The core development tool is Microsoft Visual Studio 2003. All coding behind coding is implemented by using this tool. Besides that dreamweaver MX is used to develop the HTML coding and design the GUL interface.

| **Category** | **Description** |
|---|---|
| Operating system | Window XP |
| Web Server | MS IIS 5.1 |
| Special Requirement | MS .NET Framework 1.1 |
| Development Tool | MS Visual Studio .NET 2003 |
| HTML coding | Dreamweaver |

**Table 6.1: Software Specification**

## 6.3 Approaches to the development of the system

The approaches is all about review of the coding methodologies, convention and the best practices.

### 6.3.1 Coding methodologies

There is study to the different approaches which can used to develop the system is needed before starting the development of the system's implementations. There are two types of methodology.

#### 6.3.1.1 Top down coding methodology

It is a high level specification of the system states functionality. This high level description is then refined to be a sequence or a loop of simpler function or procedures that are then themselves refined, etc. In the top-down model an overview of the system is formulated, without going into detail for any part of it. Each part of the system is then refined by designing it in more detail. Each new part may then be refined again, defining it in yet more detail until the entire specification is detailed enough to validate the model. The top-down model is often designed with the assistance of "dark boxes" that make it easier to bring to fulfillment but insufficient and irrelevant in understanding the elementary mechanisms.

#### 6.3.1.2 Bottom up coding methodology

By contrast in bottom-up design individual parts of the system are specified in detail. The parts are then linked together to form larger components, which are in turn linked until a complete system is formed. Strategies based on his bottom-up information

flow seem potentially necessary and sufficient because they are based on the knowledge of all variables that may affect the elements of the system.

That is mean that it constructed starting with existing primitives of the programming language and gradually more and more complicated features until all the application had been constructed.

#### 6.3.1.3 Advantages and disadvantages

**Advantages**

- Programming team stays focused on the goal
- Everyone knows his or her job.
- By the time the programming starts there are no questions.
- Code is easier to follow, since it is written methodically and with purpose.

**Disadvantages**

- Top-down programming may complicate testing, since nothing executable will even exist until near the end of the project.
- All decisions depend on the starting goal of the project, and some cannot be made depending on how specific that description is.

### 6.3.2 The chosen methodology

After reviewing the approaches, top down methodology is used. The decision is made partly due to characteristic of the chosen programming language for development. The ASP.NET is the usage of object oriented programming practice. Development in object

oriented environment practically envisions an abstract object at the top level before going down specific.

### 6.4 Coding Implementation

To implement the system design into actual coding, it is almost getting started with the coding process in the portal and the cluster rocks 4 is about the configurations. Before that, it is important to come up with set of rules which should be abided by during the coding process as described following sections.

#### 6.4.1 Coding convention

The important attribute of the source codes is the editing convention. It determine the intelligibility of a program. Internal (source code level) documentation, method of data declarations and approaches to statement constructions are the element of coding conventions. Several coding conventions have been employed in the coding process to ensure system consistency, maintainability and readability. These coding conventions are follows:

i. Use meaningful variable names, constant names, procedures names and parameters variable names to self document a program without excessive use of comments.

ii. Plan the layout of the program source code to improve it readability. Each sentence is started on the new line; statements following control structures are indented; white space is used to set off related blocks of codes. These seems insignificant but will help reduce the time needed to understand the program flow in the case where other programmers refer the code.

iii. All variables are declared at the beginning of procedures and declarations are separated from executables statements with a blank line to improve program readability.

iv. Insert comments to document of the program and improve program understandability

v. Group related types of codes together

### 6.4.2 Coding documentation

To begin the coding documentation, it should start with the selection of identifier names and then continue with the line composition of the connecting all separate components. At the end, it should follow by re- organization of the program.

Elements taken to consideration in coding an easy to maintain and enhanced system pre internal documentation, standard naming contentions and standard graphical user interfaces.

To provide a clear guide to programmer for future enhancement, comments are used in coding. The source code should embed in the statements of purpose indicating the function of modules and descriptive comments. This is useful to describe processing function.

Standard naming convention and also a standard usage of graphical user interface components is employed in developing the system. Standard naming conventions provides the programmers with easy identification of variables. A standard usage of graphical user interface components provide users an environment that will not generate much surprise on them.

### 6.4.3 Classification code

In this system, program coding is basically can be divided into a few sections. All these sections are divided based on the functions that performed. They are:

i. Connecting to the Database

The database connection string is stored in Web.Config file.

```
<add key="DBPath" value="Data Source=DING; Initial Catalog=Portal;User=sa;Password=sa;"></add>
<add key="LogPath" value="C:\Log_Portal"></add>
```

Therefore, any reference to the connection string is done like bellows:

```
Dim oConn As SqlClient.SqlConnection
oConn = New SqlClient.SqlConnection(gDBConn)
oConn.Open()
```

By putting the connections string in the Web.Config file can ensures that any future changes to the database such as location, name, and etc will only need to be perform to this file and not to the respective connection string in all the different modules.

ii. Database Operations

A standard database operation can utilize any of the ADO.Net components including the Data Set, Data Adapter and Data View. Example of a sample code that let the new user to register to login into the Mining2 portal is shown below:

```
Dim oConn As SqlClient.SqlConnection

Dim oCmd As SqlClient.SqlCommand

Dim sSQL As String = ""


sSQL = "Insert Into Users (UserId,Password,Name,Tel,Address,Email,Role) Values ('" &

txtuser.Text & "','" & txtpass.Text & "','" & txtname.Text & "', '" & txtTel.Text & "', '" &

txtaddress.Text & "', '" & txtEmail.Text & "','User')"

oConn = New SqlClient.SqlConnection(gDBConn)

oCmd = New SqlClient.SqlCommand(sSQL, oConn)

oConn.Open()

oCmd.ExecuteNonQuery()

oConn.Close()
```

# Chapter – 7 Testing

## 7.1 Introduction

System testing is to check on the performance, security, configurations, sensitivity startup and recovery system testing. Finding errors in an application is a natural result of good testing, but the testing is phase can be hectic, especially if deliverables are behind schedule. To make sure all the errors that testers discover are documented, all have to record in systematic way of recording them.

The essence of testing is to:

- Catch as many errors as possible
- Correct the errors
- Track the errors to understand their causes and any patterns that may exist
- Revalidate the stability of the solution, including ensuring that the correction of one error does not lead to the introduction of another error somewhere else

There are four basic concepts:

i. Error detection
   - Identify error through some approaches such as in implementation, walkthrough or others

ii. Error removal
   - Debugging and removing the identified error

iii. Error tracking

- Finding and correct the came of the errors

iv. The regression testing

- Quality is usually appraised by a collection of regression tests forming a suite of programs that test one or more features of the system. A regression test is written and the results are generated. If the results are in error, then the offending bug is corrected. A valid regression test generates verified results.

Testing is executing after the changes or modify on the code. Testing enhances in the integrity of a system by detects deviations in design and errors in the system. Testing aims at detecting error- prone areas. This help in presentation of errors in a system. Testing also add value to the products by conforming to user requirements.

This can help in pin- pointing the goals that the testing process needs to achieved the objectives of testing process have been determined as below:

i. Software reliabilities

- Software reliability is the probability that software will provide failure-free operation in a fixed environment for a fixed interval of time.
- Probability of failure is the probability that the software will fail on the next input selected.
- Measured per some unit of time, whereas probability of failure is generally time independent.

ii. Software quality

- Conduct fundamental research in support of professional software developers, specifically to develop methods and tools that can be used to improve the quality of industrially developed software.
- Critical to the success of the project is the experimental use of these tools and methods for the assessment of software products. Through experimental use of tools and methods, the tools can be improved.

iii. System assurance

- Main purpose of system assurance is to deliver a quality product. Conformance to requirements increase the organization's confidence in the system

iv. Optimum performances and capacity utilizations

- The purpose of testing is to ensure optimum performance and capacity utilization of command system components. The purpose of stress or capacity testing and planning is to make sure that the system is able to perform acceptably at peak usage

v. Price of non- conformance

- It is to detect errors and error prone areas in a system. Testing must be through and well planned.
- A partially tested system is as bad as untested system and the price of an untested and under tested system is high

## 7.2 Types of testing

### 7.2.1 Unit testing

The process of test the procedure, function or the object class of the individual component is known as unit testing. Developers use this approaches to find out the errors in the components. The purpose of the unit testing is to ensure all the components within a system can operate correctly. There are several steps:

i. Develop test cause to show input is properly converted to desire output
ii. Boundary condition are tested to make sure the functions run at boundaries established for limitation or restricting process
iii. Debugging the error to find out the cause of error and fix it

Internal and unit testing can be automated with the help of coverage tools. A coverage tool analyzes the source code and generates a test that will execute every alternative thread of execution. It is still up to the programmer to combine these test into meaningful cases to validate the result of each thread of execution. Typically, the coverage tool is used in a slightly different way. First the coverage tool is used to augment the source by placing informational prints after each line of code. Then the testing suite is executed generating an audit trail. This audit trail is analyzed and reports the percent of the total system code executed during the test suite. If the coverage is high and the untested source lines are of low impact to the system's overall quality, then no more additional tests are required

### 7.2.2 Integration testing

It is performed after all the object component and individual sub modules have posted to local unit tests. To ensure the system that develop meet with the system design specification, a verification process in carried out to test the system. It is to ensure the valid linking and dynamic relationship which establishes between modules in whole system.

Integration testing identifies problems that occur when units are combined. By using a test plan that requires you to test each unit and ensure the viability of each before combining units, you know that any errors discovered when combining units are likely related to the interface between units. This method reduces the number of possibilities to a far simpler level of analysis.

### 7.2.3 System testing

It is designed to reveal bugs not possible to individual components or to interaction between components and modules. System have been tested thoroughly to ensure that the stimulation run smoothly. There were involved functional testing and performance testing for system testing.

System testing falls within the scope of Black box testing, and as such, should require no knowledge of the inner design of the code or logic. It takes, as its input, all of the "integrated" software components that have successfully passed Integration testing and also the software system itself integrated with any applicable hardware system(s). The purpose of Integration testing is to detect any inconsistencies between the software units that are integrated together called assemblages or between any of the assemblages

and hardware. System testing is more of a limiting type of testing, where it seeks to detect both defects within the "inter-assemblages" and also the system as a whole.

# Chapter 8 – System Evaluation

## 8.1 Introduction

This phase measures the effectiveness and efficiency of the instruction. Evaluation should actually occur throughout the entire instructional design process - within phases, between phases, and after implementation. Evaluation may be Formative or Summative.

Formative Evaluation is ongoing during and between phases. The purpose of this type of evaluation is to improve the instruction before the final version is implemented.

Summative Evaluation usually occurs after the final version of instruction is implemented. This type of evaluation assesses the overall effectiveness of the instruction. Data from the Summative Evaluation is often used to make a decision about the instruction (such as whether to purchase an instructional package or continue/discontinue instruction).

## 8.2 Problem encountered and recommendation solutions

The development of the system and solution to overcome them constraint impended on the system. It is important to overcome problem as fast as possible and the best available solutions. The project scheduled should be followed string entry amid problems to ensure the system will be delivered on time.

### 8.2.1 Difficulties in choosing suitable development tools

.NET technology was blessing in some ways but a huge limitation in others. It provide unimaginable possibilities with its language feature, managed code concept and powerful

language capabilities. However, once you declare to plan the system on top of .NET platform, there is no turning back.

The development tools that have been used is Visual Studio .NET 2003. Although it offer many benefits for individual developers and software development but there are not free such as Sun Suite and Jeventor.

### 8.2.2 Difficulty in choosing the suitable programming language

There is variety of languages in Visual Basic .NET such as ASP.NET and C#.NET. With limitation knowledge on the respective programming languages, it was understandable that I was worried I will end up choosing the wrong language for the system.

Besides, I also seek advice from professional programmer who were proficient in all languages. I did also joined online programming languages and asking their opinion from different people to avoid biased judgements in some forum. The online form is extremely helpful and provided many variable in sights on the characteristics.

### 8.2.3 Lack knowledge in the chosen programming language

I did not have any experience before in chosen programming language. There is no sufficient time to learn everything entirely. The lack of full understanding on the concept of the language put me in difficulty. I tried to take alternative solutions to solved them.

### 8.2.4 Lack of knowledge on chosen technology

.NET technology is infancy. The famous is it own "managed library" where a huge collection of classes and methods have been created specially for .NET technology. .NET

programming language can utilize it and eventually any cross usage between the programming is supported.

I never heard of Rock 4.0 before doing this project. It is simple, self-contained, cluster-aware management system that is scalable and upgradeable. It is slowly becoming the de facto cluster package. One of the solutions I take is by trying to ask people who knows and familiar with it for advices and guidance.

**8.3 system strength**

Few characteristic which denotes the strength of the system:

i. Immediate response to user query

- By utilizing the "auto post back" function of the ASP.NET controls, any selection done on the controls is posted back to the server for immediate action. This can minimized response time to the user request and higher satisfaction for the user.

ii. Reduced interaction with the database

- The interface and database reduced with the data set object where records are only retrieved once and coached into data set for sequence operations. After data operations are no longer needed, the records in the dataset object can be updated back to the database. Therefore, records are only retrieved and copied back once.

iii. User friendly graphical interface

- The graphical user interface design is tailored to average user capabilities and promises classiness of usage. Even has been notice that user will have no problem navigating through the web application.

iv. Reusability

- The source codes are organized into reusable classes in an object oriented environment. Interface codes are separated from logic processing clearly. This advocates for easy reuse of the different components of the system or while adding future enhancements to the source codes.

**8.4 system limitation**

**8.4.1 Number of file to be uploaded**

There are some limitation in file uploading in the access portal in which administrator only can upload job script file one for each entry. This can be enhanced by providing options for the administrator can upload more than one file for each entry.

**8.4.2 Security access level**

Basically, there are two security access levels in the access portal which is the administrator and normal user. The administrator have wider access right compared with users. In future, more security access level can be implemented by having more user's categories such as super administrator or team leader. This will help in defining clearer responsibilities and in turn assign more specific access right for each of them.

### 8.4.3 Not multilingual system

Due to time constraint, the system is developed limited to support one language (English) only. However, it can be enhanced to support more languages for global convenient.

## 8.5 Future enhancements

The system should be maintained throughout the lifetime of the system because the user requirement might vary from time to time. Enhancement in the future will extend the usability of the system. Moreover, the system limitation should be improved in order to enhance the system's functionality.

There are some suggested future enhancement for the system which described as below:

### 8.5.1 Extent ability of browser

The system should be designed to enable it to be viewed and executed properly in other browser. It has a sizeable share in the browser market.

### 8.5.2 Enhance user interface

User interface should be enhanced from time to time. Multimedia elements such as animated graphic and flash movies can be added to increase its attractive, impressive and interactive. However, complex graphic has its pros and cons. For example, the download time would become longer if there are too much multimedia elements putting in the same page. The access portal needs to be updated frequently and it is vital that the information posted is up to date.

### 8.5.3 support multiple languages

It is possible to enhance this system to support more languages. So that the users can understand and utilize the system.

### 8.5.4 Provide online help documentation

Due to time constraints, only the user manual is attached with the system to guide the users on how to use it. It is more convenient if online help documentation can be provided in this system to give timely to solve any confusion in using the system.

## 8.6 knowledge and experience gained

From beginning of this project until the final documentation, a number of problems are occurred and experiences are learned from there. There are:

### 8.6.1 development tools knowledge

The access portal is using Microsoft Visual Studio .NET 2003, Microsoft SQL server and macromedia Dreamweaver MX as authoring tools. Besides, the mining2 cluster is using Rock version 4.0. The development tools is very powerful and sophisticated. The database application responsible on managing data. The project has helped to develop skills in using these tools.

### 8.6.2 Learn to work in team

This project is a group project which consists of four people. It has provided a chance to learn how to work in team in order to achieve consensus on system design,

development and implementation. This indeed a valuable experience to learn how to arrange and conduct a group project to ensure everything works on schedule.

### 8.7 project conclusion

In conclusion, the access portal and cluster can work well. This is beneficial which it will improve the management of the portal and reduce the works.

There was a lot of knowledge gained throughout the development of this system especially knowledge in data grid, cluster, access portal, programming language and database and so on. Those have been proved to be valuable experience and good start of learning something new.

Lastly, access portal and cluster will still be the forefront technology in shaping the future and thus it is critical to strategically manage it to gain competitive advantage. A well designed and properly implemented access portal can dramatically increase the value. User can gain quickly, division of job submission and timely access to the information they need through the system.

Therefore, there is a need to enhance the system especially in terms of offer wider functionality and flexibility to better fulfill the needs and user requirements that might vary from time to time.

# APPENDIX: User Manual

## How to Start Mining2 Portal

1. Open a web browser (eg: Internet Explorer, Netscape Navigator).

   (A default web page will be opened.)

2. Type "http://localhost/portal" in the address bar and press Enter. (The main page of system will be displayed as shown in Figure 1.)

   **NOTE:** "http://localhost/portal" are used only when the system is in development environment (can be accessed on server side only). The path can be vary depends on where the developer keeps the folder root. If the system is rolled out to live stage (online accessible), URL has to be typed in the address bar instead of the one mentioned.

Address Bar: Type the correct path name here

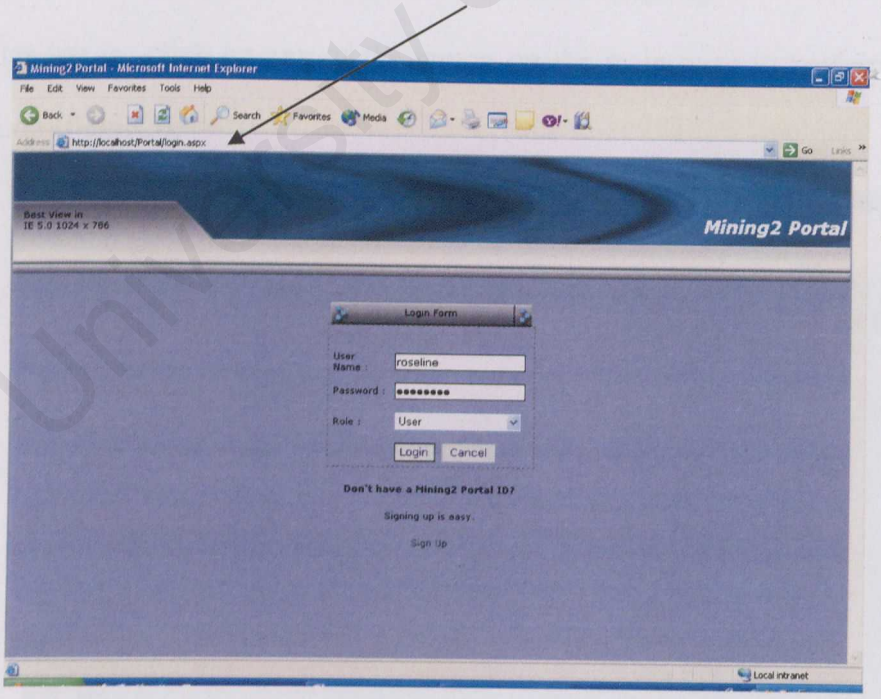



Figure 1: Login page

3. Once log on, login if you are the member. If not, sign up as new user and fill in the details.

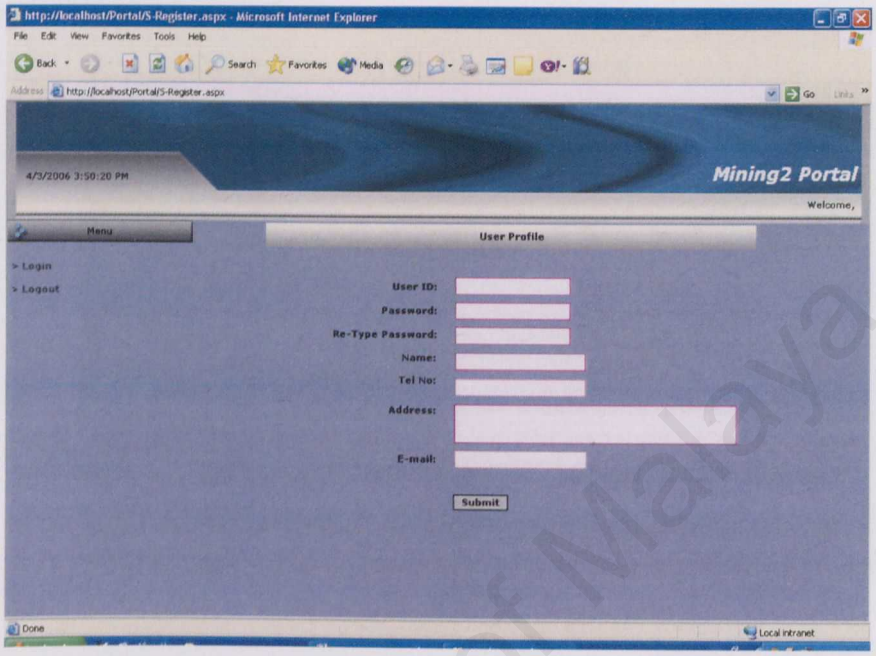



Figure 2: Fill in the details and sign up

4. After log in, click on any of the button on the main page to perform different action (figure 3):

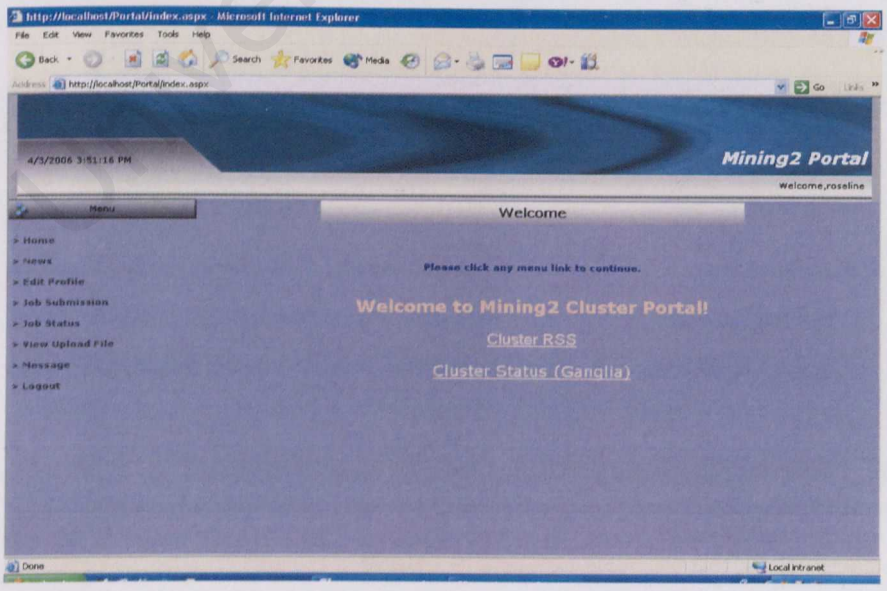



Figure 3: Main page

a. Click on "News" button to view the latest news that post up by administrator (will go to Figure 4)

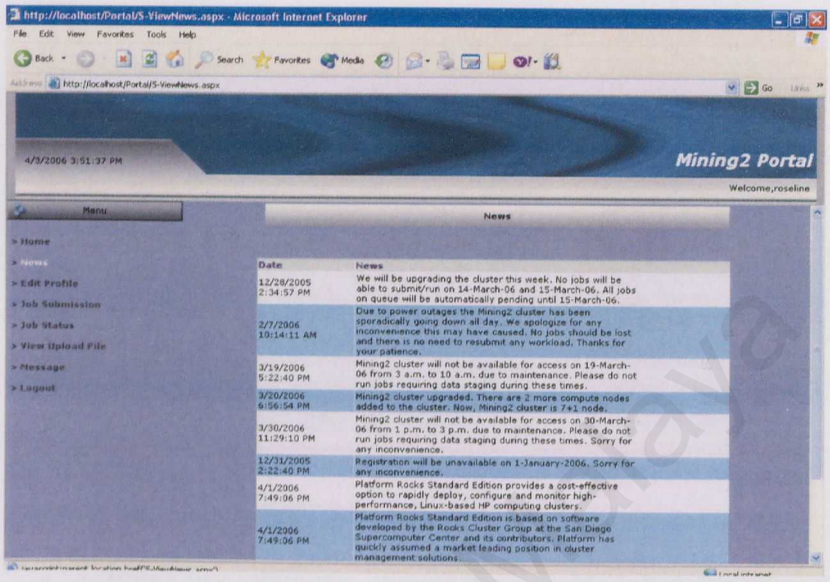



Figure 4: News page

b. Click on "Edit Profile" button to update or edit own profile (will link to Figure 5)

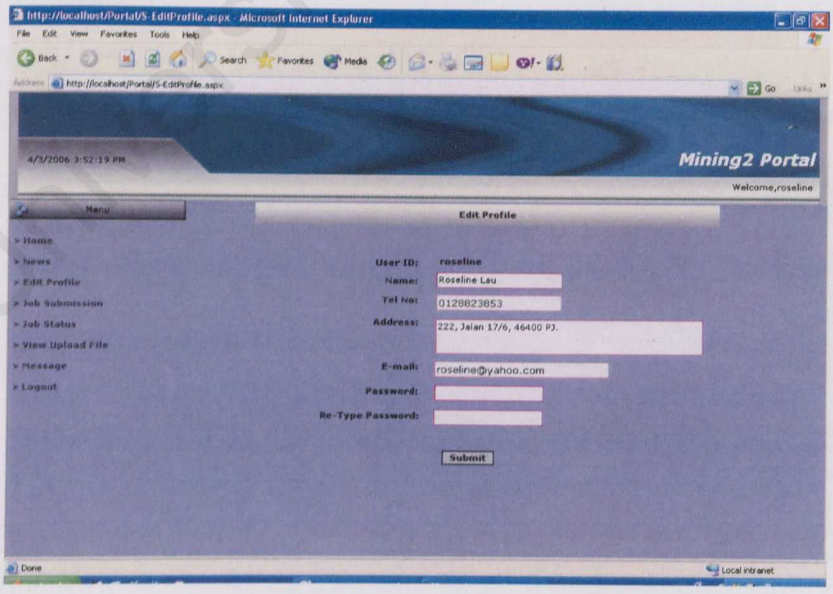



Figure 5: Edit profile

c. Click on "Job Submission" button to submit new job to the access portal (will traverse to Figure 6)

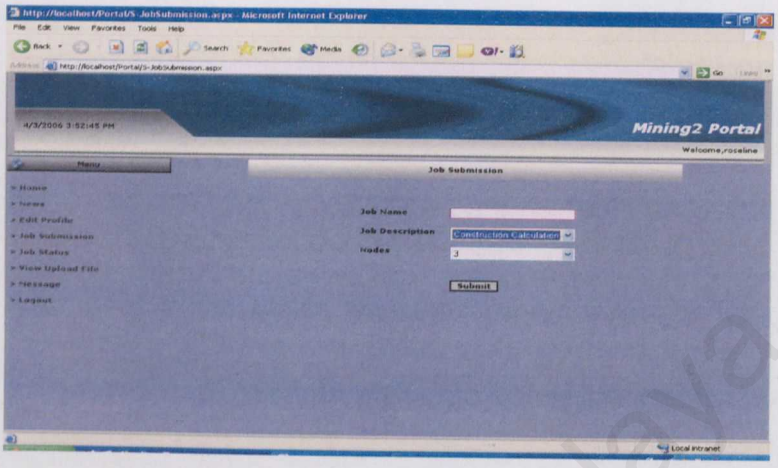



Figure 6: Job submission

- Enter any job name, choose the suitable job description and choose the numbers of nodes required. Then click "submit".

d. Click on "Job Status" button to view ONLY own job that have been submitted whether it is pending, running or complete. View result if only the job is complete. (will link to Figure 7)

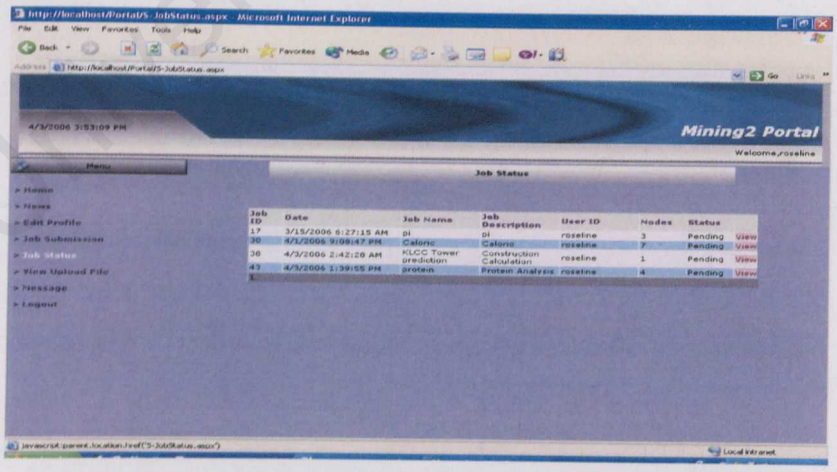



Figure 7: Job Status

- User able sort all the attributes that available in the column

e. Click on "View Upload File" button to view uploaded file by administrator (will traverse to Figure 8)

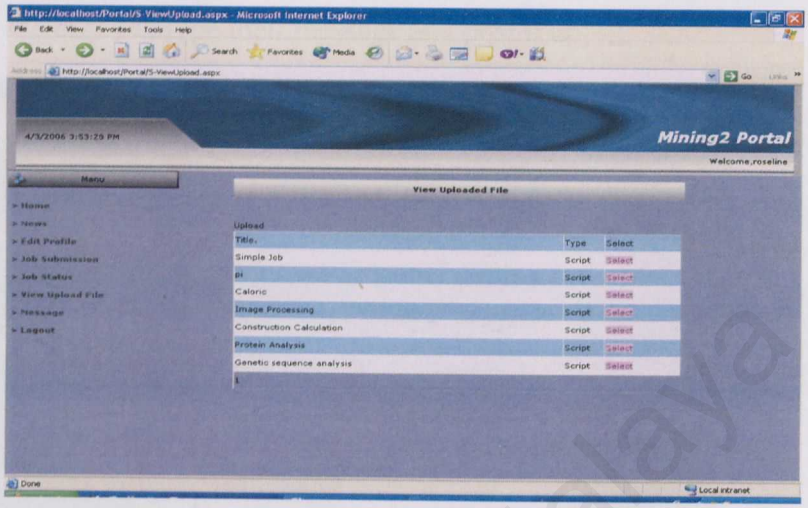



Figure 8: View uploaded file

- Click "Select" to view the file that wish to retrieve
- The file contain either is job script or user manual that uploaded by administrator

f. Click on "Message" button to post up message to contact the administrator or view on replied messages. (will traverse to Figure 9)

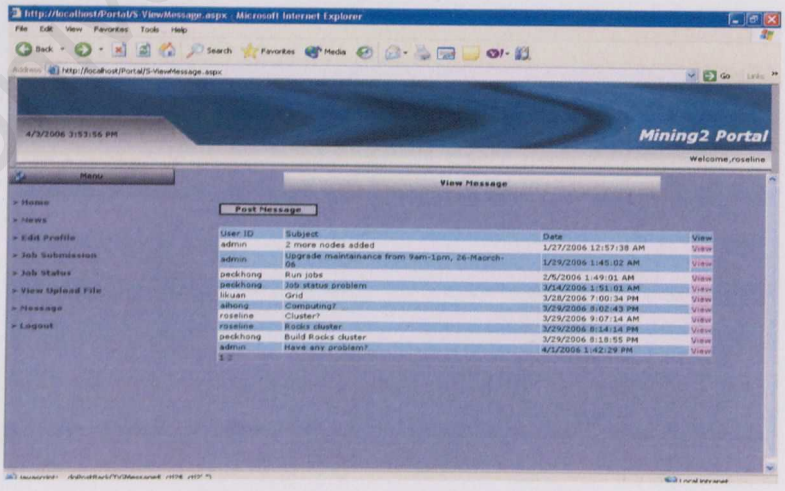



Figure 9: View messages

- Click "View" to see the replied messages or click "Post Message" to add post new message (will link to figure 10)
- user can type their message and post it up

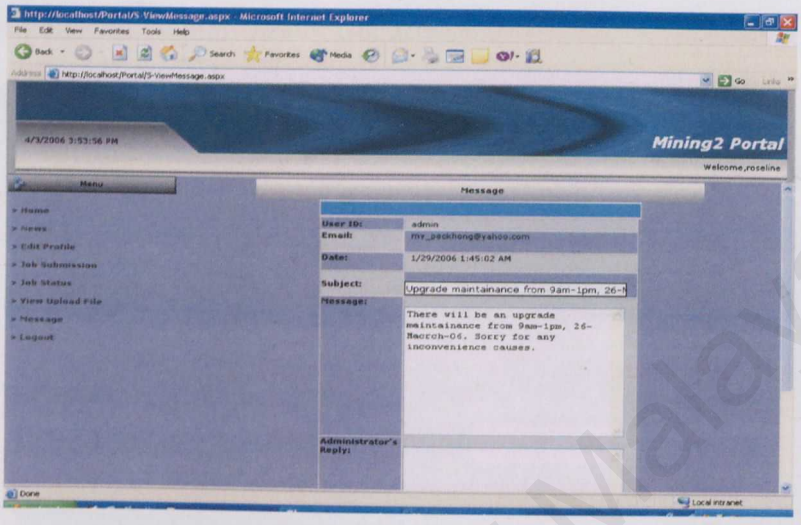



Figure 10: Post messages or view replied messages

g. After finished log on, log out and it will return to Figure 1, log in page

## APPENDIX B: Administrator Control

Administrator has the same log in page as normal user. Administrator has to log in with "Admin" role. The access right for administrator definitely widely than normal user.

Below are the pages that administrator can access:

i. Can edit and delete the news uploaded

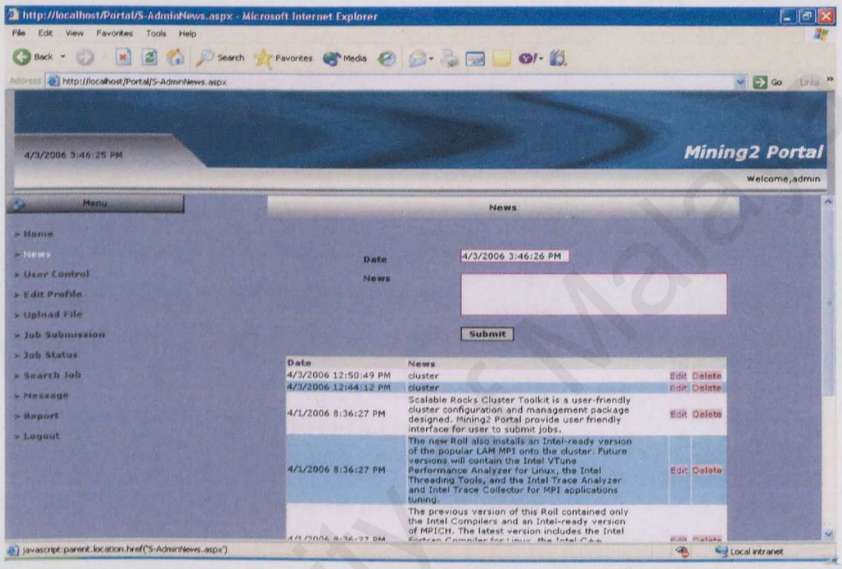



Figure A: Edit and delete news

ii. User control to delete or update user's profile

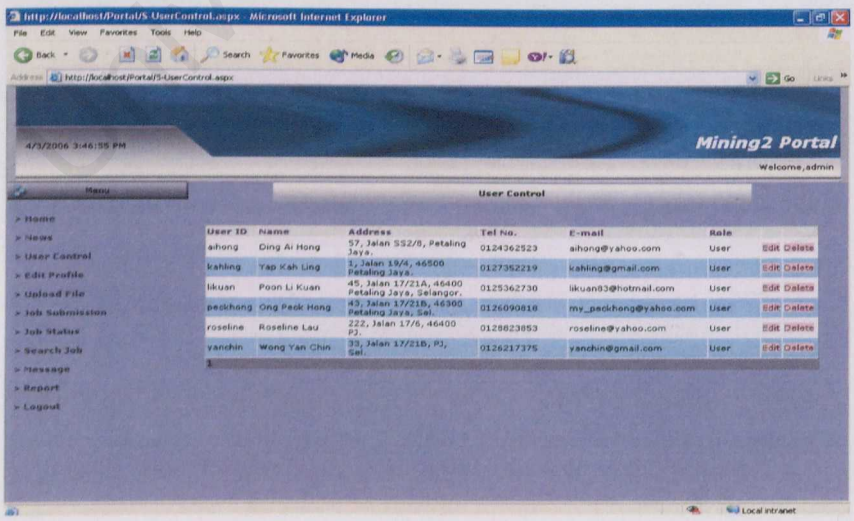



Figure B: User control

iii. Able to upload file, for example job script

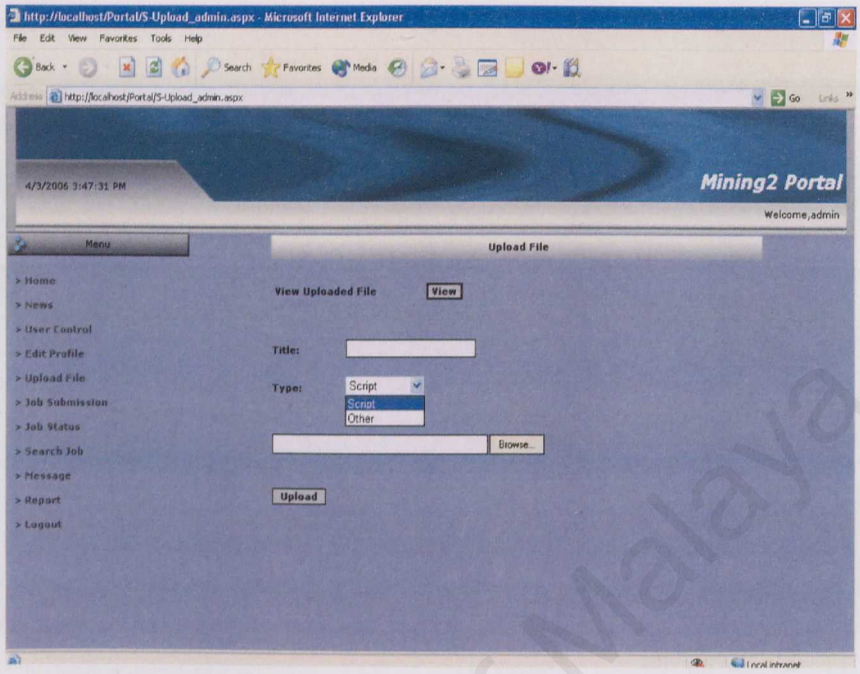



Figure C: Upload file

iv. Can delete submitted job by user in job status page

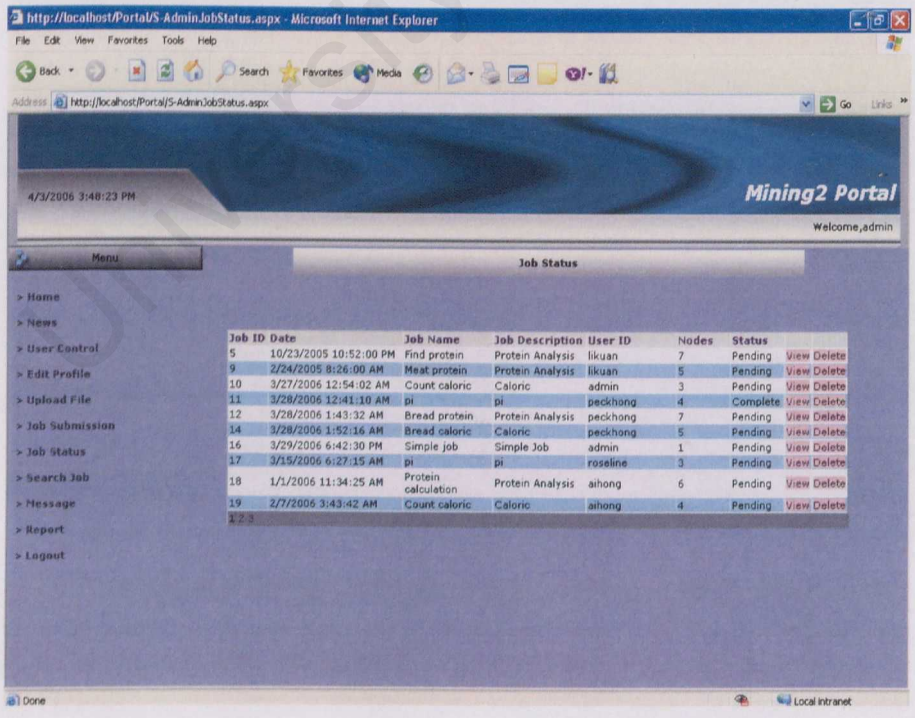



Figure D: Job status

v. Can view reports: User- job report, user reports, nodes reports, job reports and job status reports

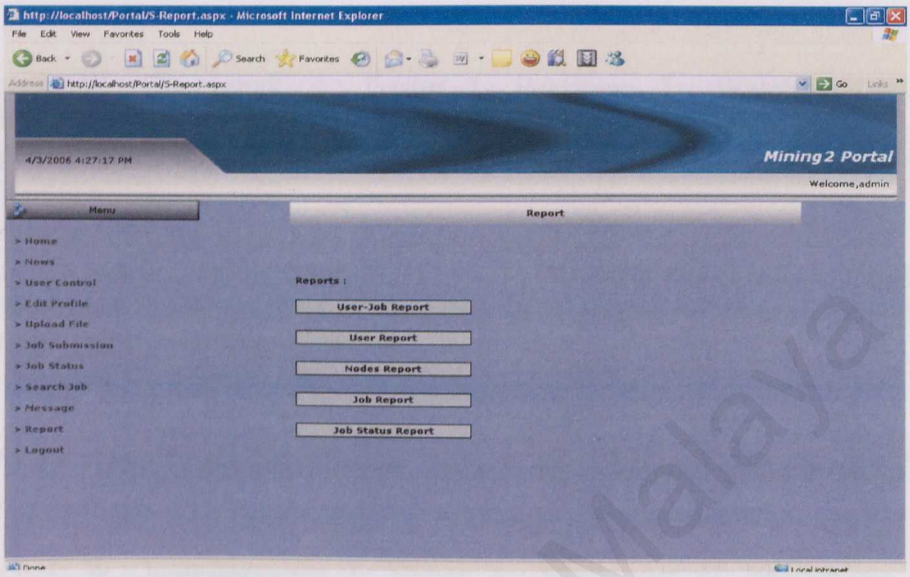



Figure E: Various kind of Reports

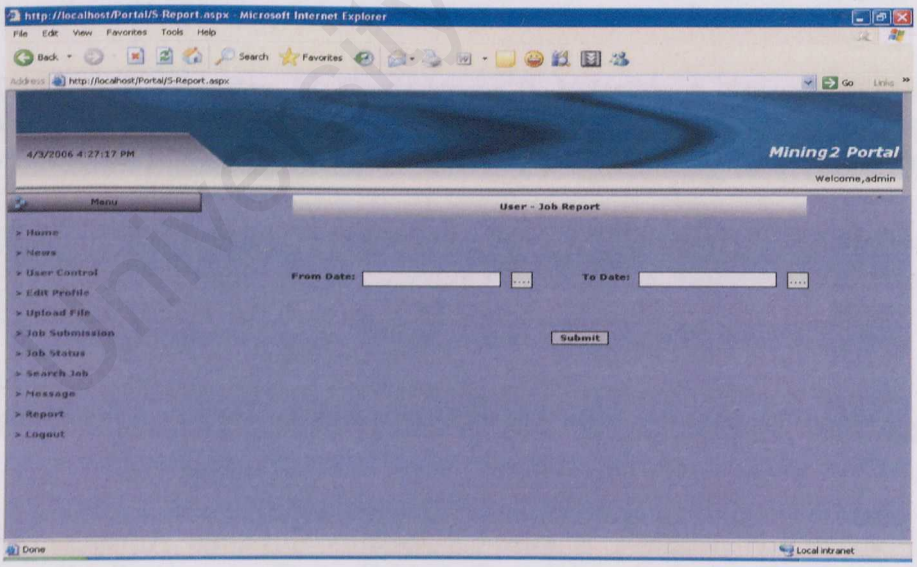



Figure F: Retrieved Time Range for User- Job report

- The reports can view by choosing from certain range of time that required

# REFERENCES

**Journals/ Documents**


1. Welchi, V.; Siebenlist, F.; Foster1, I.; Bresnahan, J.; Security for Grid Services, [online], Available: http://www.globus.org/alliance/publications/papers/GT3-Security-HPDC.pdf

2. Allcock, W.; Chervenak, A.; Foster, I.; Pearlman, L.; Welch, L.; Wilde, M,; Globus Toolkit Support for Distributed Data-Intensive Science, [online], Available: www.globus.org/alliance/ publications/papers/Globus.CHEP01.pdf

3. Bosio, D.; Casey, J.; Frohner, A.; Guy, L.; Next-Generation EU DataGrid Data Management Services, 2003. Computing in High Energy Physics (CHEP 2003), La Jolla, California, March 24 - 28

4. NPACI Rocks Cluster Distribution, Version 4.0.0 Edition: Users Guide; 2005, UC Regents.

5. Hoschek, W.; Jean-Martinez, J.; Samar, A.; Stockinger, H.; Stockinger, K.; Data Management in an International Data Grid Project. 1st IEEE/ACM International Workshop on Grid Computing (Grid'2000). Bangalore, India, Dec 17-20, 2000.

6. Adrian Fish ReDRESS Project Review: Survey of Portals and Related Software. (University of Lancaster, February 2004)

7. S. Krishnan, R. Bramley, M. Govindaraju, R. Indurkar, A. Slominski, D. Gannon, J. Alameda, D. Alkaire The XCAT Science Portal Proceedings SC2001, Nov. 2001, Denver. Extended version in Scienti_c Programming 10(4) (2002) 303-317

8. D. Gannon, M. Christie, O. Chipara, L. Fang, M. Farrellee, G. Kandaswamy, W. Lie, B. Plale, A. Slominski, A. Saarangi, Y. Simmhan, Building Grid Services for User Portals, GGF Workshop on Grid Services, Chicago, (Oct. 8th 2003)

9. K. Holtman, Clustering and Reclustering HEP Data in Object Databases, Proc. of CHEP'98,Chicago,USA.http://kholtman.home.cern.ch/kholtman/chep1/art1.web.html

10. H. Stockinger, A. Samar, B. Allcock, I. Foster, K. Holtman, B. Tierney. "File and Object Replication in Data Grids." Proceedings of the Tenth International Symposium on High Performance Distributed Computing (HPDC-10), IEEE Press, August 2001

**Sample Application**

1. PANDA GRID Portal: http://panda.physics.gla.ac.uk/
2. GENIUS Web Portal http://egee.cesnet.cz/en/user/gen/
3. PACI HotPage Grid Computing Portal: https://hotpage.paci.org/

**Web site**


http://deyalexander.com/

http://www.its.monash.edu.au/

http://egee.cesnet.cz/en/user/submission-GUI.pdf.

http://prisms.ices.utexas.edu/rocks-documentation/ 3.2.0/faq-installation.html

http://www.umt.edu/portalproject/articles.htm

http://prisms.ices.utexas.edu/rocks-documentation/3.2.0/

http://www-itg.lbl.gov/Akenti/.

http://www.globus.org/rearch/papers/ogsa.pdf.

www.w3schools.com/aspnet/default.asp

http://esc.dl.ac.uk/Papers/Portals/portals.pdf

www.microsoft.com/net/

www.aspnetfaq.com/

http://dast.nlanr.net/Projects/GridPortal/

http://www.liferay.com/products/index.jsp

http://www.lib.ncsu.edu/eresources/mylibrary/

http://mis105.mis.udel.edu/ja-sig/uportal

www.learnasp.com/

www.microsoft.com/sql/