

EUDG IMPLEMENTATION IN KOREA FOR HIGH-ENERGY PHYSICS STUDY

Bockjoo Kim[†], Kihwan Kwon, DaeHee Han, Dongchul Son, Kihyeon Cho, Junghyun Kim,
YoungDo Oh, CHEP, Kyungpook National University, Daegu, KOREA
Jysoo Lee, KISTI, Daejon, KOREA

On behalf of the HEP Data Grid Working Group of GFK

Abstract

Researchers in high energy physics (HEP) have had difficulties securing enough computing powers to store and process tons of data produced from accelerators. In addition, the fact that the researchers are distributed over the world compromises the efficiency of the high energy physics researches which are cooperative in nature. Grid, the so-called next generation Internet, is expected to overcome these difficulties. By integrating computing resources distributed over the world, Grid will allow increased accessibility to data, necessary computing powers, and storage space. In this paper, we describe the initial implementation experiences of HEP Data Grid in Korea using European Data Grid (EUDG) software developed by CERN.

INTRODUCTION

High energy physics is a field where researchers study creation and evolution of the Universe by analysing data generated in accelerators through collision of particles. Since in Korea there are not suitable accelerators for such purposes, we often perform coordinated research with Europe, United States and Japan where there exist large accelerators.

Coordinated researches of this kind generates enormous amount of data. For example, CMS and ATLAS experiments to be performed at CERN in 2007 will generate data in scales of peta bytes. Accordingly, researchers on site and at the remote sites need to have facilities that enable them to have immediate access to the data and enough computing power.

However, resources at CERN or Fermilab are not enough to meet all these requirements. Therefore, it is required that resources be shared among countries attending these experiments for analysis. Currently, many countries including Korea are developing Grid infrastructure to realize such kind of resource sharing.

Grid is the so-called next generation Internet which enables sharing resources such as computing power, storage, application software, etc., distributed over the World. Especially, Data Grid makes possible to store and retrieve large amount of data transparently.

CHEP (Center for High Energy Physics) at Kyungpook National University set off to deploy such Data Grid infrastructure to function as a Regional Data Center for

CMS experiment in 2007. This kind of infrastructure can be also utilized for other experiments such as CDF and AMS. In this paper we describe our efforts to deploy HEP Data Grid using EUDG software. Section 2 introduces basic components of EUDG. In section 3, we describe Korean HEP Data Grid based on EUDG. HEP Data Grid web service is briefly introduced in section 4. Then, we conclude in section 5.

EUROPEAN DATAGRID (EUDG)

EUDG is one of Data Grids being developed around the World. EUDG consists of centrally managed machines with multiple functionalities: Resource Broker (RB), Computing Element (CE), Storage Element (SE), Worker Node (WN), User Interface (UI), etc [1]. Each of these is a Linux machine. UI is the interface machine where you can submit jobs. RB is the resource broker, which locates usable computing elements and passes jobs to them. CE distributes jobs through its batch job manager to worker nodes. SE is the place where data comes from for the job. Outputs are also stored in SE. EUDG software structures are shown in Fig.1.

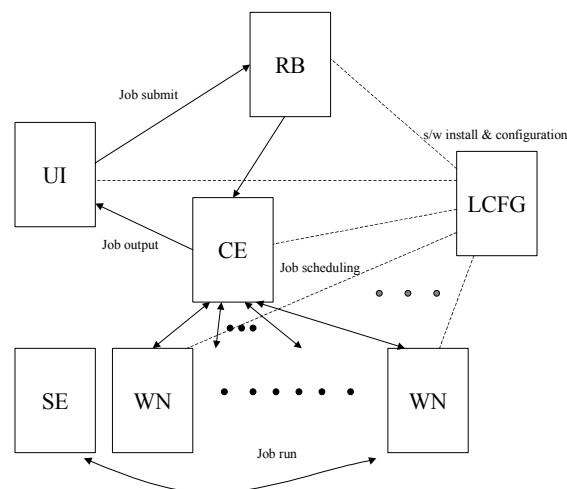


Figure 1: EUDG software structure

HEP DATA GRID IN KOREA

CHEP will be a Tier-1 institute and other universities will be Tier-2 institutes for CMS and other experiments. To this end, CHEP is endeavouring to implement HEP Data Grid system. Currently, we are building such Data Grid based on EUDG software. So far, we have deployed two such EUDG sites successfully, one at the CHEP and

[†]bockjoo@chepl2.knu.ac.kr

the other at Seoul National University (see Fig. 2). We are going to extend this infrastructure to other Tier-2 universities and institutes.

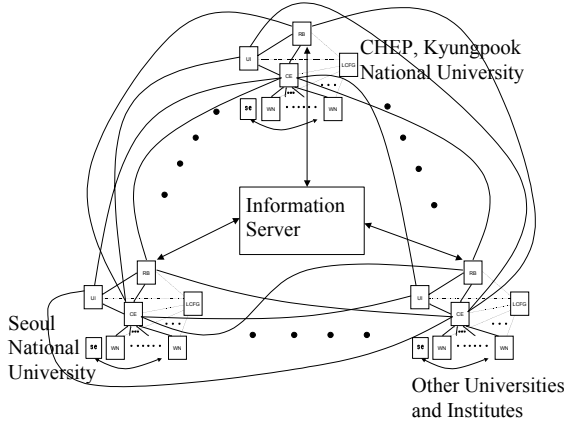


Figure 2: current HEP Data Grid structure

Each site is deployed using LCFG server (see Fig. 1). LCFG server is a local configuration server and we can install any EUDG component by changing the configuration files in LCFG server. However, each site needs to have LCFG server because LCFG uses DHCP (Dynamic Host Configuration Protocol) which usually works in the same network. LCFG is also used for maintenance of the system. For more details, see [2].

CHEP and Seoul National University are connected each other by sharing RB, and jobs can be submitted from either sites. When a job is submitted to the Data Grid, time synchronization is critical because user certificate cannot be authenticated if time difference is significant between the machines. Therefore, it is desirable for LCFG server to be configured so that each node can be synchronized to a time server.

LDAP (Lightweight Directory Access Protocol) server is used to manage virtual organizations (VO) in the Data Grid. Each user should belong to at least one VO to be able to submit jobs to the Data Grid. One LDAP server is running on one of our machines to manage VOs in Korea. We also created a new VO in addition to default VOs in EUDG, which is CDF VO for CDF experiments and users belonging to CDF VO can submit jobs to machines with CDF application software.

Currently SE is not well implemented and we are trying to incorporate SRB (Storage Resource Broker) [3] into our system. HEP Data Grid based on EUDG needs more intensive tests by each experimental group to better identify problems which can be occurred when a large amount data are produced.

Furthermore, EUDG framework was referenced to build data grid testbed for Belle Experiment inside Korea. Most of EUDG components and concepts are realized using globus toolkit and those are setup properly to run simulation code under Belle experiment specified environment.

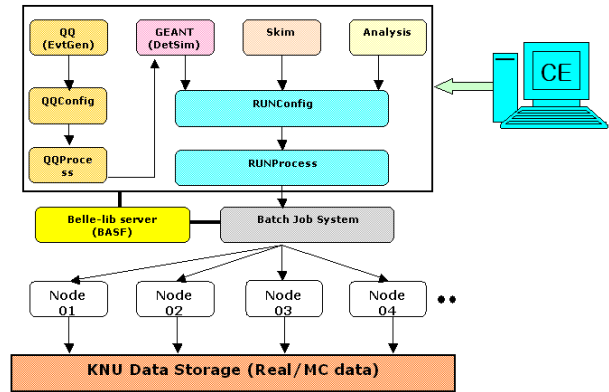


Figure 3: The structure of CE for Belle experiment

HEP DATA GRID WEB SERVICE

HEP Data Grid web service is constructed to increase utilization of HEP Data Grid. One of the greatest advantages is that users can submit jobs from the Internet without regard to their machine platform. Since the web server needs to interact with UI machine, only implementation of web service is not complicated.

When a user's proxy is verified, he can access Korean EUDG web interface (see Fig. 4). The interface has basic EUDG menus such as dgjoblistmatch, dgjobsubmit, dgjobstatus, dgjobcancel and dgjobgetoutput. With dgjoblistmatch, you can find any CE available at that time. Then, you can submit jobs with dgjobsubmit. Currently, dgjobgetoutput just indicates the location of the output in the UI machine. Job description file should be uploaded from the UI machine because it cannot be edited on the fly from the web interface at the moment.

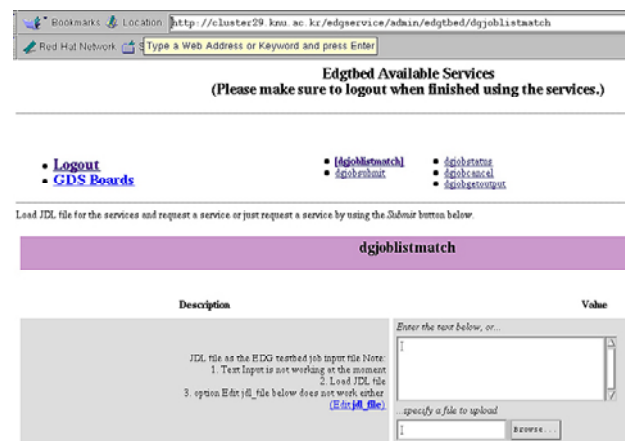


Figure 4: EUDG Web Service Interface

CONCLUSION

Data Grid may be a solution for problems such as lack of computing power and storage in high energy research. Many countries are working towards construction of Data Grid infrastructure and Korea is not an exception. CHEP at Kyungpook National University is currently establishing HEP Data Grid system to function as a

Regional Data Center for CMS and other HEP experiments. Although many other kinds of Data Grid systems are tried, currently HEP Data Grid system is implemented based on EUDG software developed at the CERN. Two such EUDG sites are established and they are connected each other sharing RB for resource sharing. Also, for increased utilization of the HEP Data Grid, EUDG Web Service was developed so that users can submit their jobs from anywhere. In the future, Korean sites should be connected with EUDG sites in other countries and job submission to each other should be tested to fully utilize computing resources dispersed over the World.

REFERENCES

- [1] <http://eu-datagrid.web.cern.ch/eu-datagrid>
- [2] <http://www.lcfg.org/>
- [3] <http://www.npaci.edu/DICE/SRB/>