Cyberinfrastructure for High Energy Physics in Korea

Kihyeon Cho[1], Hyunwoo Kim and Minho Jeung
High Energy Physics Team
Korea Institute of Science and Technology Information, Daejeon 305-806, Korea

E-mail: cho@kisti.re.kr

Abstract. We introduce the hierarchy of cyberinfrastructure which consists of infrastructure (supercomputing and networks), Grid, e-Science, community and physics from bottom layer to top layer. KISTI is the national headquarter of supercomputer, network, Grid and e-Science in Korea. Therefore, KISTI is the best place to for high energy physicists to use cyberinfrastructure. We explain this concept on the CDF and the ALICE experiments. In the meantime, the goal of e-Science is to study high energy physics anytime and anywhere even if we are not on-site of accelerator laboratories. The components are data production, data processing and data analysis. The data production is to take both on-line and off-line shifts remotely. The data processing is to run jobs anytime, anywhere using Grid farms. The data analysis is to work together to publish papers using collaborative environment such as EVO (Enabling Virtual Organization) system. We also present the global community activities of FKPPL (France-Korea Particle Physics Laboratory) and physics as top layer.

1. Introduction

We report our experiences and results relating to the utilization of cyberinfrastrucure in high energy physics. According to the Wiki webpage on “cyberinfrastrucure,” the term “cyberinfrastructure” describes the new research environments that support advanced data acquisition, storage, management, integration, mining, and visulaization, as well as other computing and information processing services over the Internet. Therefore, the cyberinfrastructure in high energy physics consists of infrastructure (computing and network), Grid farms, e-Science and global community such as FKPPL (France-Korea Particle Physics Laboratory) and physics. The hierarchy for cyberinfrastructure from bottom to top is as follows:

- Infrastructure – supercomputing and networks
- Grid
- e-Science
- Community **-** France-Korea Particle Physics Laboratory
- Physics – heavy flavor physics

We will explain the hierarchy of cyberinfrastructure. The basement is supercomputing and network. Next layer is grid farms. The following layer is e-Science and community such as FKPPL (Franc-Korea Particle Physics Laboratory). The top layer is high energy physics experiments. KISTI (Korea Institute of Science and Technology Information) is a government-aided research institute that provides Korean scientists and engineers with scientific and industrial databases, supercomputing,

[1] Corresponding author

research networks. The mission of KISTI is to advance the national information infrastructure by providing the communities involved in advanced research with leading-edge computational resources and networks. In other words, KISTI is the national headquarter of supercomputer, network, Grid and e-Science in Korea. Therefore, KISTI is the best place to work on cyberinfrastructure for high energy physics in Korea.

2. Infrastructure

2.1. Supercomputer

In the infrastructure, there are two components – supercomputer and networks. The first infrastructure is supercomputing. Currently, the CPU at supercomputing center at KISTI is 30 TFlops. We are installing the 4^{th} supercomputer, which will help achieve a computing performance of 322.6 TFlops in various user support programs; this places the centre among top ten providers of computing facilities in the world. The 4^{th} supercomputer will be installed by 2009.

2.2. Network

The next infrastructure is network. KISTI hosts GLORIAD (Global Ring Network for Advanced Applications Development) network. The GLORIAD connects location across the world with a 2.5-10 Gbps network: Daejeon (Korea), Seattle (USA), Calgary (Canada), Chicago (USA), Amsterdam (Netherlands), Stockholm (Five northern European countries), Moscow (Russia), Beijing (China), Hong Kong (China) and Daejeon (Korea). Figure 1 shows the network among KISTI, KEK, Fermilab and CERN.

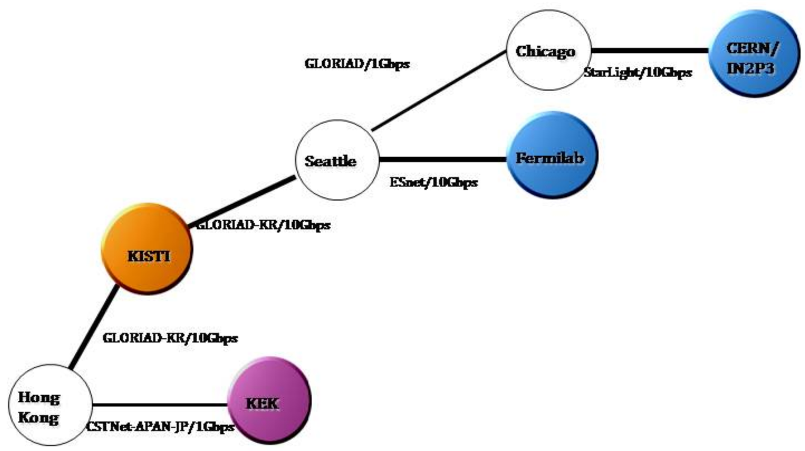

Figure 1. The network among KISTI, KEK, Fermilab and CERN.

KISTI is connected with KEK by 1-Gbps network through Hong Kong. KISTI is connected with Fermilab by 10-Gbps network. KISTI is also directly peered to CERN (European Organization for Nuclear Research). The bottle neck of the network between KISTI and CERN is inside of USA which is between Seattle and Chicago. We test the performance of network using Mathis's equation which reads

$$BW \leq MSS / RTT * C / \sqrt{p} \quad (1)$$

where BW is upper bound of TCP throughput, MSS is maximum segment size, RTT is round trip time and p is packet loss. The equation shows that bandwidth depends on round trip time and packet loss. Therefore, in January 2009, we study the packet loss and round trip time between KISTI and KEK, and KISTI and Fermilab, and KISTI and CERN. Figure 2, 3 and 4 shows the packet loss and round trip time from KISTI to CERN, Fermilab, and KEK, respectively. In all cases, packet losses are negligible and routing times are also very stable. However, figure 4 shows that the round trip time between KISTI and KEK is 100 ms, which is longer than expected when we consider the geological distance.

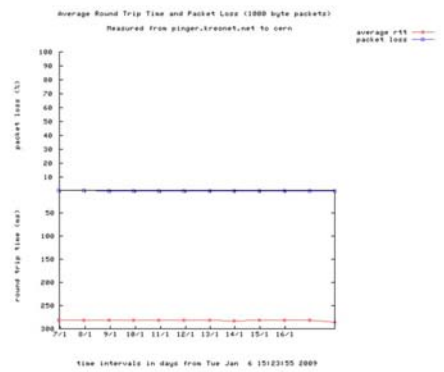

Figure 2. The packet loss and round trip time between KISTI in Korea and CERN in Switzerland.

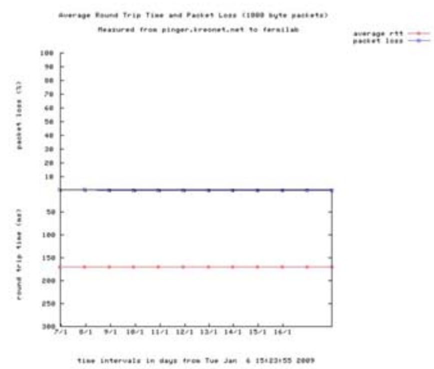

Figure 3. The packet loss and round trip time between KISTI in Korea and Fermilab in the USA.

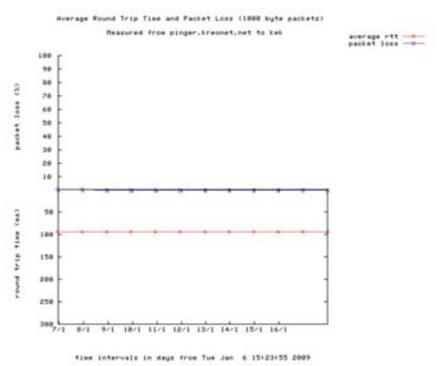

Figure 4. The packet loss and round trip time between KISTI in Korea and KEK in Japan.

3. Grid farms

The next layer is grid farms. Data processing is carried out using a high-energy-physics data grid. The objective of the high-energy-physics data grid is to construct a system to process high-energy-physics data and support high-energy physics community [1]. The LCG (LHC Computing Grid) organization involves a hierarchy of computing centers from CERN labeled as Tiers 1, 2, and 3. There are three LCG farms in Korea. Two are located at KISTI and one is located at KNU (Kyuungpook National University). KISTI runs the KR-KISTI-GCRT01 farm for the ALICE (A Large Ion Collider Experiment) experiment and the KR-KISTI-HEP farm for the CDF (Collider Detector at Fermilab) experiment. Before April 2008, the CDF experiment shares the farm with ALICE experiment at KISTI. Now CDF has its own farm at KISTI which is one of work nodes of Pacific CDF analysis farm. KNU runs the LCG_KNU farm for CMS (Compact Muon Solenoid) experiment. Figure 5 shows the monitoring of LCG farms in Korea.

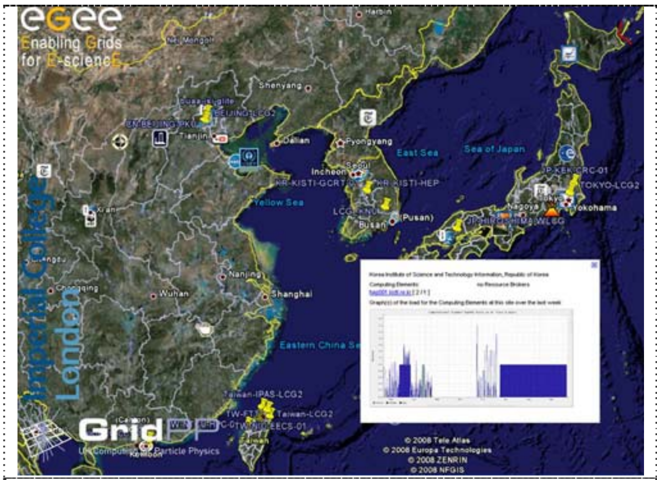

Figure 5. LCG monitoring system. We can see KR-KISTI-GCRT01, KR-KISTI-HEP and LCG_KNU.

University of Seoul also operates the CMS HI (heavy ion) data center. However, the middleware is OSG farm since University of Seoul receives CMS HI data from the Tier1 center at Fermilab in the USA.

For the future plan, figure 6 shows the road map of high energy physics farms at KISTI.

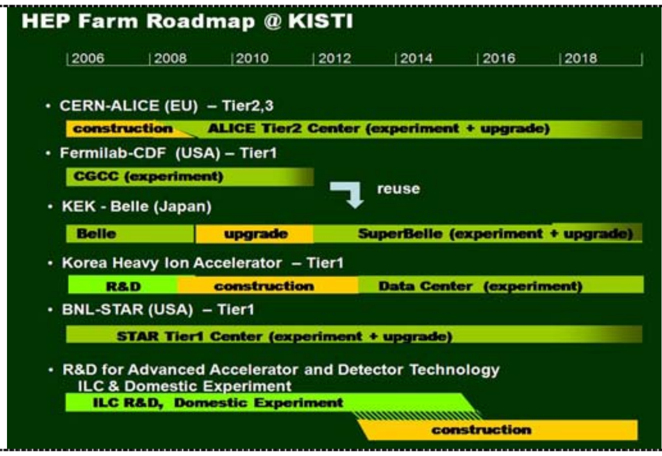

Figure 6. The road map of high energy physics farm at KISTI.

4. e-Science for High Energy Physics

e-Science is a new research paradigm for computationally intensive scientific research and is carried out in highly distributed network environments [2]. e-Science involves the use of immense data sets that require grid technology [2]. The goal of e-Science is to enable researchers to conduct any study at any time and location. Therefore, the goal of e-Science for high energy physics is to study high energy physics anytime, anywhere even if we are not onsite laboratories. Virtual laboratory enables us to research as if we were onsite. The components are data production, processing and analysis. We apply this concept to CDF experiment. At Fermilab, data production is taken by shifts at the main control room. Data processing is done at the Feynman computing center. Data Analysis is done in conference room.

Therefore, we constructed the components at remote site at KISTI. First data production takes both the online and offline shifts anywhere. The online shift is taken through a remote control room, and the offline shift is taken through a SAM (sequential access through meta-data) data handling system. Second, data processing processes data using a high energy physics data grid. Third, data analysis for collaboration around the world is to analyze and publish their results using EVO (Enabling Virtual Organization).The table 1 shows the processes in e-Science compared to those in typical scientific studies at Fermilab [3].

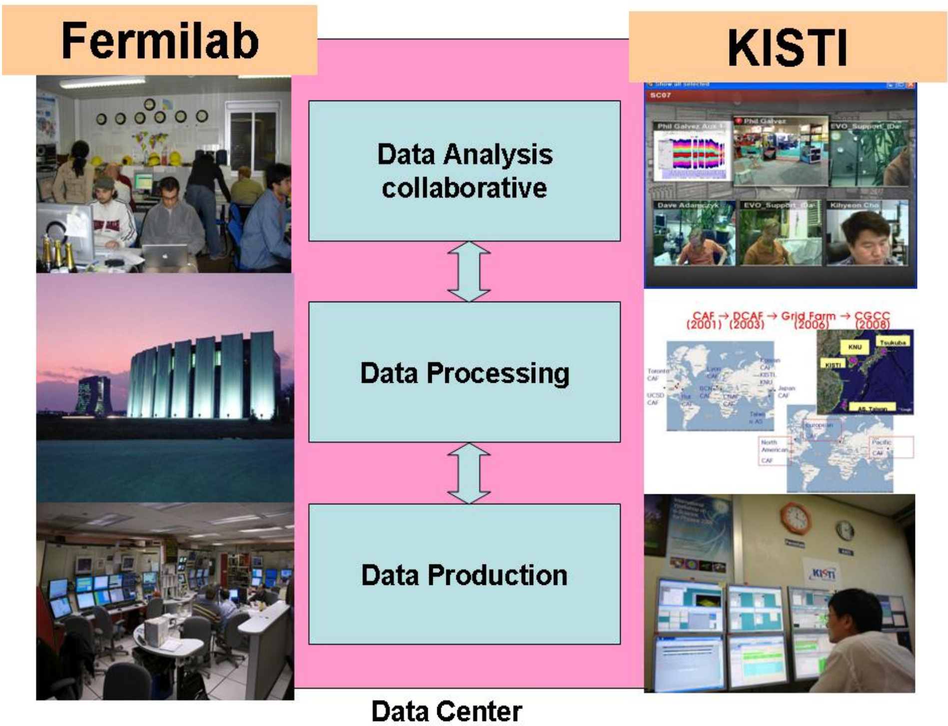

Figure 7. Components of e-Science for high energy physics. The photographs to the left were taken onsite (Fermilab) and those to the right were taken offsite (KISTI).

Table 1. Processes in e-Science compared to those in typical scientific studies at Fermilab [3].

Components	Traditional Science	e-Science
Site	Onsite (Fermilab, USA)	Offsite (KISTI, Korea)
Data Analysis	Conference room (Fermilab)	Enabling Virtual Org. (KISTI)
Data Processing	Central Analysis Farm (Fermilab)	Pacific CDF Analysis Farm (KISTI)
Data Production		
- **Online shift**	CDF Control Room (Fermilab)	Remote Control Room (KISTI)
- **Offline shift**	Off-line shift (Fermilab)	SAM Data Handling shift (KISTI)

4.1. Data Production

Generally, we collect data onsite, where accelerators are located. However, to adhere to the concept of e-Science, we would like to be able to collect data from any location. One method is to use a remote operation center. Currently, we have constructed a remote CDF operation center at KISTI that enables CDF users in Korea to take CO (consumer operator) shifts at KISTI in Korea, not at Fermilab in the USA.

4.2. Data Processing

The increasing luminosity of the Tevatron collider causes the computing requirements for data analysis and MC (Monte Carlo) production to become greater than the dedicated CPU resources that will be available [4]. In order to meet this demand, CDF is investigating various computing methods - CAF, DCAF (Decentralized CDF Analysis Farm), and grids [5].

The regional CDF collaboration of Taiwanese, Korean, and Japanese group built the Pacific CDF analysis farm. The farm is based on the Condor glide-in concept where Condor daemons are submitted to the grid, effectively creating a virtual private batch pool [6]. Thus, submitted jobs and results are integrated and are shared at grid sites. For work nodes, we use both LCG and OSG (open science grid) farms. The head node of the Pacific CDF analysis farm is located at the Academia Sinica in Taiwan. Now, the farm has become a federation of one LCG farm at KISTI in Korea (KR-KISTI-HEP), one LCG farm at the University of Tsukuba in Japan (JP-TSUKUBA-U-03), two OSG farms (IPAS_OSG and TW-NCHC-Formosa2) and one LCG farm (Taiwan-LCG2) in Taiwan. Figure 8 shows the monitoring system of Pacific CDF analysis farm.

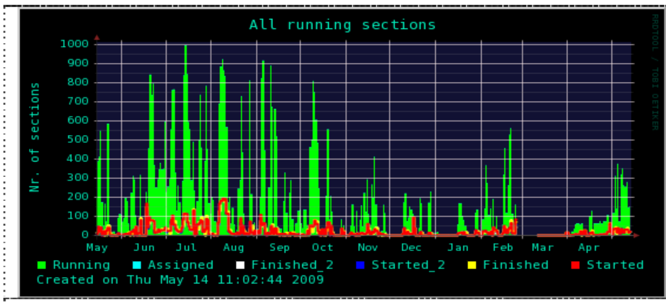

Figure 8. The monitoring of Pacific CDF analysis farm between May 2008 and April 2009.

4.3. Data Analysis
For the data analysis collaboration, we hosted the EVO server systems at KISTI so that high-energy physicists in Korea can use it directly without using reflectors in the USA. Therefore, the EVO server at the KISTI enables global collaborations for performing analysis and easily publishing results together [7]. For the data analysis collaboration, we constructed EVO servers at KISTI. When users in Korea use EVO servers at KISTI, the routing time is reduced by 60 ms with no congestion from the network inside of the USA, which provides a very stable research environment [7].

5. France-Korea Particle Physics Laboratory
The next layer is global community. One of them is FKPPL (France-Korea Particle Physics Laboratory) was established in March 2008. The objective of FKPPL is to carry out joint cooperative activities or "joint research projects", under a scientific research program in the fields of high energy physics experiments, e-Science, and related technologies recommended by its steering committee. The projects are Grid computing, ILC (International Linear Collider) calorimeter, Bioinformatics, ILC electronics, ALICE and CDF as listed in table 2.

For this project, we provide LCG grid farm called "FKPPL VO" farm. The nodes of UI (User Interface), VOMS (Virtual Organization Management Service) and WMS (Workload Management System) are located in Korea and the nodes of SE (Storage Element) and CE (Computing Element) are located in France. More than 5,000 CPU are available for both sides.

Within the framework of this project, KISTI and LPHHE-IN2P3 developed joint research on CDF as an e-Science application. Research focused on CDF grid technology and heavy flavour physics.

Table 2. The list of projects on FKPPL (France-Korea Particle Physics Laboratory).

Programs	France (IN2P3)	Korea (KISTI)
Co-directors	Vincent Breton (LPC-Clermont Ferrand)	Ok-Hwan Byeon (KISTI)
Grid Computing	Dominique Boutigny (CC-IN2P3)	Soonwook Hwang (KISTI)
ILC Detector R&D	Jean-Claude Brient (LLR-Ecole Polytechnique)	Jongman Yang (Ewha University)
Bioinformatics	Vincent Breton (LPC-Clermont Ferrand)	Doman Kim (Chonnam National Univ.)
ILC Electronics	De La Taille Christophe (LAL-IN2P3)	Jongseo Chai (Sungkyunkwan Univ.)
ALICE	Pascal Dupieux (LPC-Clermont Ferrand)	Yongwook Baek (Kangnung National Univ.)
CDF	Aurore Savoy-Navarro LPNHE / IN2P3-CNRS	Kihyeon Cho (KISTI)

6. Physics
The top layer is physics. Until 2008, we concentrated on the environment for high energy physics. Since 2009, we have been probing the standard model and searching for new physics using e-Science. We study heavy flavour physics through e-Science in the CDF experiment. Heavy flavour physics is an important element in understanding the nature of high energy physics. There are three known generations of quark doublets. However, the origins of families are unknown in the standard model.

Only the charged-current electro weak interaction can change flavours in the standard model. Electroweak eigenstates are not mass eigenstates, which are introduced in the CKM (Cagbibbo-Kobayashi-Maskawa) matrix [8]. Therefore, precision measurement of heavy flavor physics provides confirmation of CKM theory and a clue for new physics beyond the standard model.

7. Conclusions

Cyberinfrastructure consists of infrastructure (computing and network), Grid farms, e-Science, community and physics. We apply this concept to high energy physics in Korea. KISTI leads cyber-infrastructure for high energy physics in Korea and global communities.

8. References

[1] Cho K 2007 *Computer Physics Communication* **177** 247
[2] Cho K 2007 *International Journal of Computer Science and Network Security* **7** 49
[3] Cho K and Kim H 2008 Heavy flavor physics through e-Science *Preprint* hep-ex/0810.1616
[4] Neubauer M 2003 *Nuclear Instrumentation and Methodology* A **502** 386.
[5] Cho K 2007 *Proc. Int. Symp. on Grid Computing (Taipei)* vol 1 (Springer) p 41
[6] Sffligoi I 2007 *Computer Physics Communications* **177** 235
[7] Cho K 2008 *Journal of the Korean Physical Society* **53** 1187
[8] Kobayashi and M Maskawa T 1973 Progress in Theoretical Physics 49 652