

## A TEST-BED OF DCAF(DECENTRALIZED ANALYSIS FARM) IN KOREA

Kihyeon Cho<sup>†</sup>, Daehee Han, Youngdo Oh, Dongchul Son, Center for High Energy Physics,  
Kyungpook National University, Daegu 702-701, Korea  
Jysoo Lee, KISTI, Supercomputing Center, Daejeon 305-806, Korea

### Abstract

There is a CAF(Central Analysis Farm) at Fermilab for the analysis of CDF Experiment. However, due to limited resources of CAF, CDF experiment is working on Grid project. As one of this project, we have designed DCAF and constructed DCAF in Korea (KorCAF) at CHEP (Center for High Energy Physics), Kyungpook National University. In order to run remote data at Fermilab in USA, we will use SAM(Sequential data Access via Meta-data) Grid technology

### INTRODUCTION

The objective of the high energy physics (HEP) is to understand the basic properties of elementary particles and their interactions. High energy physics are usually conducted at the major accelerator sites where there are detector design, construction, signal process, DAQ and analysis at a large scale. The size of collaborations is 100 ~ 2000 physicists. Therefore, an area of high energy physics gives birth to the conception of Grid. Grid is next generation of internet which is the virtual organization of heterogeneous resources of computing, storage and even human resources. High energy physicists in Korea are also involved in the activities of the international high energy physics Grid project.

The objective of Data Grid is, therefore, to construct the system to manage and process the data that will be produced in the near future. At this scope, the high energy physics Data Grid is indispensable for the future high energy physics activities. We need to manage mass storage system of hard disks and tapes in a stable state. CPU power may be extendable and access for the data should be transparent.

Recently, CDF (Collider Detector at Fermilab) experiment is to study high energy particle collisions using the world's highest energy particle accelerator whose center of mass energy is 1.96 TeV/c<sup>2</sup>. The goal is to discover the identity and properties of particles that make up the universe and to understand the interactions between those particles. The CDF experiment has started taking data on March 2001 called Run II period. At the CDF experiment, more 600 faculties and students from 55 institutes at 11 countries are involved in this experiment. The experiment has CAF(Central Analysis Farm) system at Fermilab to analyse data. The CAF is a large farm of computers running Linux with access to the CDF data handling system and databases to allow CDF collaborators to run batch analysis jobs [1]. Now, the

CAF system consists stage 1 and 2. The design of CAF considered two things. The first one is to submit jobs from anywhere. The second thing is that job output can be sent directly to desktop or stored on CAF FTP server for later retrieval.

However, in 2005, requirement set by goal at CDF experiment is that 200 simultaneous users should use 10 million events in a day. Therefore, we need more computing power. Finally, CDF experiment produces the concept of CDF Grid. As a first step of Grid, we produce the concept of DCAF (DeCentralized Analysis Farm) based on CAF system. A user can submit a job from anywhere to the cluster either at CAF or at DCAF. Later JIM (Job information Management) environment will decide which cluster they will submit a job automatically. This is a final stage of Grid called SAMGrid. This concept also contains the real data handling system. In order to use real data at Fermilab for DCAF around world, we need the data handling system called SAM(Sequential Access through Meta data). Therefore, for the CDF experiment, we are doing gridification via SAMGrid.

Conclusively DCAF is the early stage of Grid. Korea Group is one of most advanced groups in this Grid project at CDF experiment. Finally we developed and installed DCAF system in Korea.

### CDF EXPERIMENT

CDF is an experiment located at Fermilab has the world's most powerful particle accelerator called the Tevatron whose center of mass energy is 1.96 TeV/c<sup>2</sup>. The Tevatron accelerates protons and antiprotons close to the speed of light. Then, it makes them collide head-on inside the CDF detector. The CDF detector is used to study the products of such collisions.

We try to reconstruct what happened in the collision and ultimately try to figure out how matter is put together and what forces nature uses to create the world around us[2]. The main physics of CDF experiment is as follows: 1) to study QCD at large Q<sup>2</sup>, 2) to enhance the accuracy to the measurement at weak interaction, 3) to find the characteristics of top quark 4) to search for Higgs particles and SUSY(Supersymmetry) particles and 5) to improve the measurement of CKM matrix by studying B physics and Charm physics using lots of data. For this purpose, CDF experiment upgraded detector for 5 years and start to take data again on March 2001. The CDF experiment is currently taking data.

<sup>†</sup>cho@knu.ac.kr

## COMPUTING SYSTEM AT CDF

### Introduction

In order to analyse real analysis of and generate Monte Carlo simulation data, CDF experiment uses officially the CAF(Central Analysis Farm) at Fermilab.

However, each institute also uses its PC clusters using its own batch system such as PBS. However, Korean Group developed the DCAF(Decentralized Analysis Farm) besides PBS batch system. The reason is that users at CDF use the PC clusters at DCAF as CAF system. This is a user-friendly concept and this is the first step of gridification for the final CDF Grid.

### CAF ( Central Analysis Farm)

Recently, CDF experiment has CAF(Central Analysis Farm) system to analyse data. Now, the CAF system consists stage 1 and 2. The system consists 303 dual CPUs and 180 Tbyte disks. CDF group is producing MC data and doing analyse using CAF system.

The basic operation paradigm is for users to compile and link their analysis jobs on their desktop. Then they execute their jobs via a distributed batch system of FBSNG(Farm Batch system Next Generation). After process, they receive their job output back on their

desktop or send it to special user scratch space on CAF FTP servers for later retrieval.

A user submits a job using CDF software of Graphic User Interface (GUI). Then, the CAF head node submits a job to work node following by a work node's reading data files which are in hard disk of CAF system or in tape library called enStore.

The result will be sent to one of three places. One is the original desktop where the user submitted a job. Other one is another remote desktop. The last one is CAF FTP server to place the store. The user can retrieve the result any time from CAF FTP server. All of this system is inside of security system called Kerberos.

CDF experiment is going to upgrade CAF system gradually year by year. However, in 2005, 200 simultaneous users should use 10 million events in a day. Therefore, we need around 700 Tbyte disk and 5 THz of CPU by 2005. CDF experiment produces the concept of DCAF (DeCentralized Analysis Farm). This is the bottom-up method of Grid. We start the gridification from real working process step by step.

### DCAF(DeCentralized Analysis Farm)

The concept of DCAF is almost same as CAF. A user does not have to distinguish between CAF and DCAF.

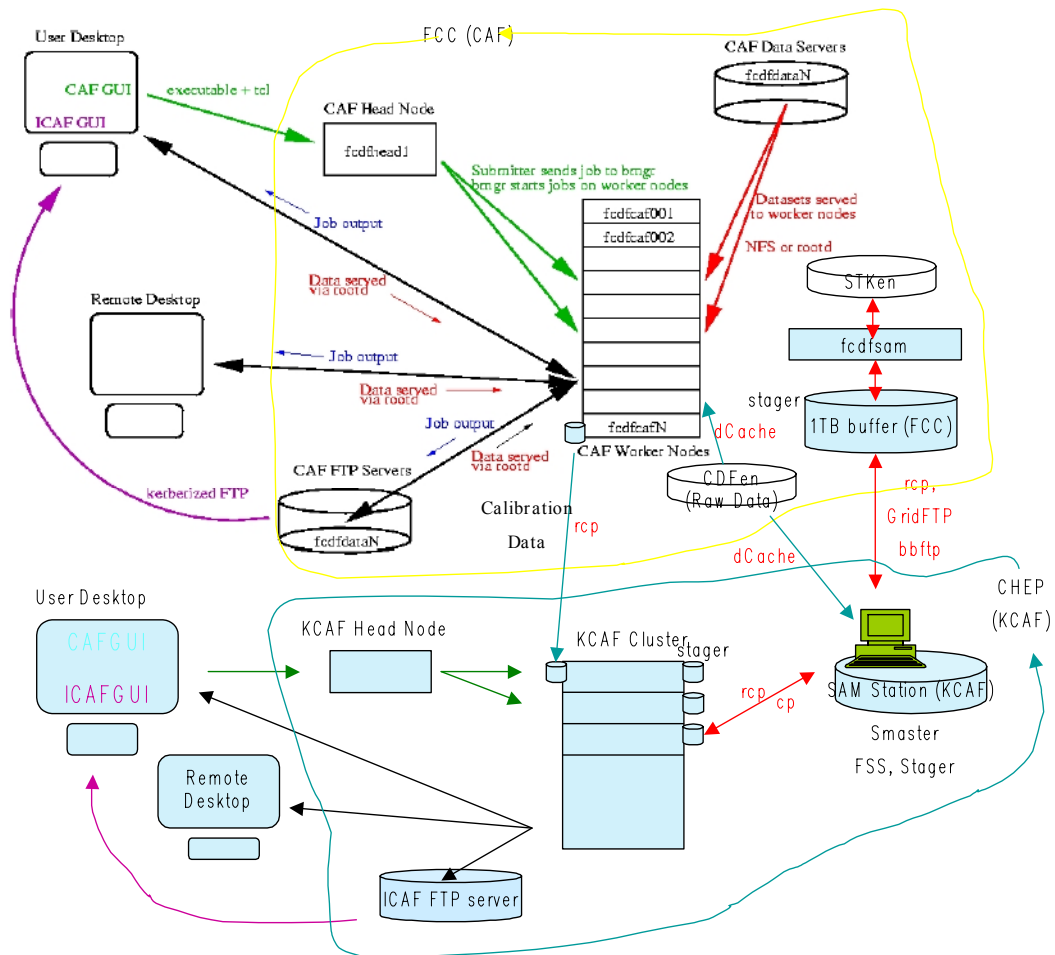


Figure 1. The design of KorCAF. This design includes the CAF at Fermilab. The geographical distance between CAF and KorCAF is around 10,000 km.

Therefore, to submit a job and get a result is same. However, in order to run a real data at a local work node, we need date handling system to bring the data from Fermilab. To bring data from Fermilab, we use SAM (Sequential data Access via Meta-data). This process includes bringing data at tape library inside Fermilab by dCache.

On November 2002, we made a test bed and demonstrated at SuperComputing2002 conference. Six institutes - Rutgers, Texas Tech, Toronto, RAL, Imperial College and Kyungpook National University are involved for this demonstration.

## KORCAF (DCAF IN KOREA)

### Introduction

Recently, we succeeded in developing and installing KorCAF (DeCentralized Analysis Farm in Korea) using one head node and three machines of dual CPUs of AMD 2000 (1.6GHz) cluster as work nodes. The Figure 1 is the design of KorCAF system. The current plan is to use DCAF as a MC production farm. In order to product MC data, we need the calibration data that are at Fermilab. For the calibration data, we use Oracle database. All of this system is inside of Kerberos security system which is same as that of CAF. For this system, we get the same system like of Fermilab.

For the future plan, we are going to add more CPUs and to install the SAM station for real data handling system.

### How to run KorCAF

Figure 2 shows the procedure of KorCAF. We use the same GUI of CAF. The difference is only to select the analysis farm for KorCAF.

First we make the CafGui at the submission site. Currently, the submit machines are at Fermilab. The CafGui shows several input information. The first one is to choice an analysis farm. The default analysis farm is CAF. However, we made the KorCaf in Korea. Therefore, we added the KorCAF analysis farm at the GUI. By clicking the choice, we selected the KorCAF(KNU) site. The second line is for the choice of processing type. It is same as CAF. There are several options for process such as test, long, short, medium and long. The third line is the initial command. That is the executable file for a user to run. The fourth line shows the original directory for the executable file. The fifth line shows the output file locations in the machine that should be kerberized. The last line shows the email option. If a user want to receive the result of process as e. mail, then click email 'yes'.

The Figure 2 shows that the job is submitted well and finally we get the result.

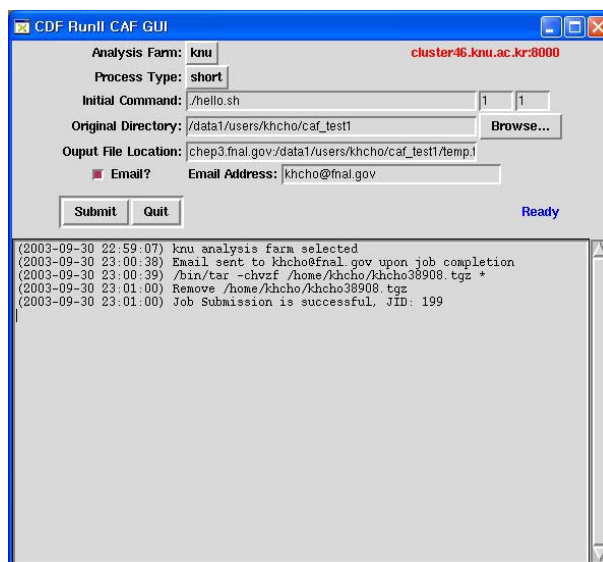


Figure 2: A GUI for KorCAF process.

## RESULTS

We have designed and developed the DCAF system. Finally we installed DCAF in Korea. So, we constructed KorCAF. This is the first DCAF at CDF experiment that is outside of Fermilab. This farm is actually not only for a test-bed but also real farm for a MC production. All the users around world as well as Users of Korean group will use this farm directly.

## CONCLUSIONS

We are going to use the DCAF system to generate lots of MC simulation events. The results will be also store enStore at Fermilab. This system consists PC clusters and security system. This is the beginning of Grid. This will help other experiments for remote site to make analysis farm as a concept of Grid. This construct will improve the concept of the Grid of bottom-up. Later we will include SAM system for the final goal of SAMGrid.

## ACKNOWLEDGEMENT

We acknowledge the assistance of Frank Wurthwein and Mark Neubauer for a CAF code. This work was supported by grant No. R08-2003-000-10258-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

## REFERENCES

- [1] <http://www-cdf.fnal.gov/caf/>
- [2] <http://www-cdf.fnal.gov/>