탈중심분산팜(DeCentralized Analysis Farm)의 구현

한대희[0,1] 권기환[1] 조기현[1] 오영도[1] 손동철[1] 이지수[2]
1.경북대학교 고에너지물리연구소, 2.한국과학기술정보연구원 슈퍼컴퓨팅센터
hanbi@knu.ac.kr[0], kihwan@knu.ac.kr, cho@knu.ac.kr, ydoh@fnal.gov, son@knu.ac.kr, jysoo@hpcnet.ne.kr

The Construction of DCAF (DeCentralized Analysis Farm)

Dae Hee Han[0,1],Kihwan Kwon[1], Kihyeon Cho[1], Youngdo Oh[1], Dongchul Son[1]
Jysoo Lee[2]

1. Center for High Energy Physics, Kyungpook National University, Daegu 702-701, Korea
2. KISTI, Supercomputing Center, P.O. Box 122, Daejeon 305-806, Korea

요 약

표준모형의 힉스입자를 찾는 것을 목적으로 미국 페르미 연구소에서 수행되고 있는 CDF 실험은 전세계에 11개국 55개의 연구소가 참여하고 있다. 테바트론가속기에서 산출되어지는 데이타를 분석하는데는 많은 컴퓨팅 자원이 필요한데, Run IIb 기간동안 생성되는 데이타를 처리하고, 전세계에 흩어져 있는 연구원들이 사용하기에는 페르미 연구소의 분석용 팜(Centralized Analysis Farm)은 자원이 부족하다. 따라서, 실험에 참여하고 있는 여러 나라의 컴퓨팅 자원을 공유할 수 있는 방안으로 DCAF(DeCentralized Analysis Farm)가 개발되었다. DCAF는 Grid 구현을 위한 중간 단계라 할 수 있으며, 궁극적으로 CMS의 Tier-1 환경과의 통합 구축을 목적으로 하고 있다.

1. 서 론

자연을 설명하는 기본 모형인 표준 모형에서 예견된 모든 기본입자는 모두 발견되었으나, 마지막 입자인 힉스입자는 아직 발견되지 않았다. CDF(Collider Detector at Fermilab) 실험은 이 힉스 입자를 발견하는 것을 목적으로 미국 시카고 서부교외에 위치하고 있는 페르미 연구소에서 수행되어지고 있다. 페르미 연구소는 1.0 TeV의 양성자와 1.0TeV의 반양성자의 강입자 충돌 실험을 수행할 수 있는 고에너지 실험을 위한 입자가속기를 갖추고 있으며, CDF 실험은 현재 세계에서 가장 높은 에너지에서의 충돌실험이다.

한국에서는 경북대학교를 비롯한 서울대학교, 성균관 대학교가 CDF 실험에 참여하고 있으며, 전세계 11개국 55개의 연구소 및 학교로부터 500여명이 넘는 학생, 연구원 및 교수들이 참가하고 있다. 따라서, 전세계에 흩어져 있는 공동연구진이 이 데이터를 분석하기 위해서 원격지에서 데이터에 대한 즉각적인 접근과 충분한 고속연산능력을 제공할 수 있는 시설이 필요하다.

페르미 연구소의 파인만 컴퓨터센터(Feynman Computer Center at Fermilab)에 187TB의 Disk 서버와 303대의 Dual CPU를 가진 PC 클러스터로 구성되어 있는 중심분석용 팜(CAF)은 Run IIb 기간 동안에 수집될 데이터를 처리하고, 모의 시늉 데이터를 생성하기에는 자원이 충분하지 못하다.

흩어져 있는 공동연구진이 원격지의 데이터를 분석하고 페르미 연구소의 부족한 자원을 극복할 방안으로 참여하고 있는 여러나라에 산재되어있는 컴퓨팅 자원을 공유하기 위해서 그리드를 도입하였으며, 국제공동연구로 그 첫단계인 탈중심분산팜(DeCenteralized Analysis Farm)을 구축하였다.

2. 본론

2.1 DCAF의 구조

DCAF 시스템에서의 모든 네트워크 연결 과정은 Kerberos 보안시스템 하에서 이루어진다. DCAF의 동작과정을 간략히 설명하자면, 사용자는 사용자 데스크탑에서 CafGui라 불리는 사용자 편의 도구를 이용하여 job을 submit할 수 있다. 그러면, batch 시스템의 하나인 FBSNG[1]를 이용하여 worker node에서 batch job이 돌아가게 되고, 데이터는 CAF Data Server의 하드디스크에 있는 파일을 읽거나 Enstore에 있는 테이프 속의 파일을 dCache로 Hard disk에 dump하여 읽게된다. 결과물은 CafGui에서 지정한 곳으로 전송된다.

DCAF는 크게 head node와 worker node, CAF 부분과 FBSNG부분으로 나눌 수 있다. head node는 worker node에 job을 분배하는 역할을 하고, CAF 부분은 CafGui

에서 입력된 데이터를 FBSNG에 전달하고, 그 결과물을 받아오는 역할을 한다. CafGui에서 job을 submit하면 head node의 submitter가 그것을 받아서 FBSNG의 BMGR(batch manager)에게 전달한다. BMGR은 각 worker node의 lancher에게 전달하고 lancher는 CafExe를 통하여 사용자 어플리케이션을 수행한다. 그림 1에서 이러한 DCAF 시스템의 대략적인 구조를 살펴볼수있다.[2]

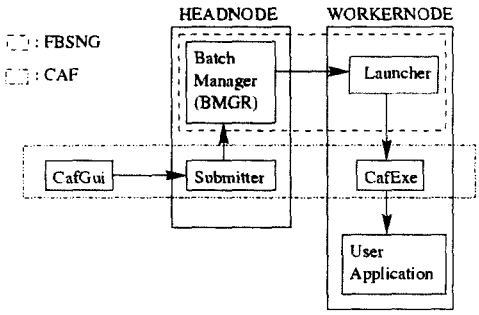

그림 1. DCAF의 구조

2.2 CafGui

DCAF 시스템에서 Job을 submit 하기 위해서는 리눅스 command line 상에서 CafSubmit 명령어를 이용하는 방법과 CafGui라는 GUI툴을 이용하는 방법이 있다. CafGui는 사용자 어플리케이션을 자동으로 압축하여 head node에 전달하기 때문에 훨씬 편리하게 Job을 submit 할 수 있다. 그림 2 에서와 같이 원하는 Analysis Farm을 선택하여 Job을 submit할수 있으며, 결과물은 Job을 submit한 사용자의 데스크탑, 다른 원격에 있는 컴퓨터, 또는 CAF FTP server에 저장하여 필요할 때 FTP로 결과물을 가져올 수 있다. job submit시에 사용자 어플리케이션은 kerberized rcp를 통하여 head node로 복사되고, 이것이 각각의 worker 노드로 분배된다. 프로그램이 종료가 되면 CafGui에서 지정한 위치로 kerberized rcp를 통하여 복사가 된다.

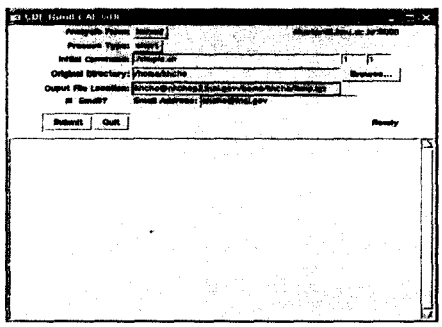

그림 2. CafGui

2.3 SAM

DCAF 시스템에서 데이터 분석시 필요한 원격지에 있는 데이터를 검색하고 전송하기 위해서는 SAM(Sequential Access via Metadata)[3]을 이용한다. 그림 3에서와 같이 head node에서 job을 분배하면 분석에 필요한 데이터 파일의 위치를 찾기 위해서 SAM station에 질의를 한다. 그러면 SAM station은 각 worker node에 파일의 위치를 알려주고, 그 결과를 가지고 dCache를 이용하여 원하는 데이터를 전송 받는 것이다.

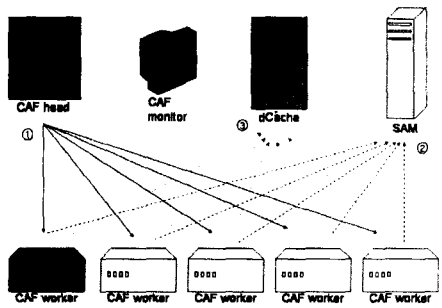

그림 3. DCAF에서의 데이터의 검색 및 전송

2.3 dCache

Enstore는 페르미 연구소의 파인만 컴퓨팅 센타에 있는 테이프 라이브러리 시스템이다. CDF 실험에 관련된 다량의 데이터가 보관되어 있으며, 지역 데이터센터가 구축이 된다면 이 데이타는 분산 보관될 예정에 있다. Enstore의 데이터를 사용해야하는 경우 DCAF에서는 SAM을 통하여 원하는 데이터의 위치를 파악한 후 dCache[4]를 이용하여 그림 4의 흐름도처럼 Enstore 혹은 지역 데이터센터에 있는 데이터를 전송받게 된다.

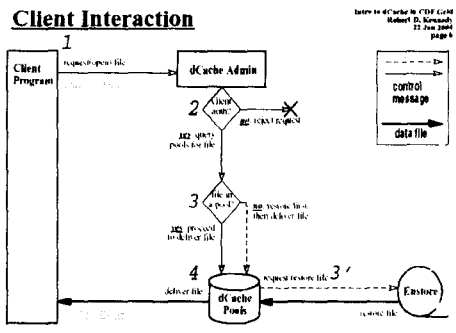

그림 4. dCache client interaction

2.4 fbsWWW

DCAF에서는 fbsWWW[5]를 이용하여 전체 시스템의 모니터링을 할 수 있다. fbsWWW에는 사용자가 job을 submit한 후 job이 실행되는 상태를 확인할수 있을 뿐만 아니라 각 worker node의 상태, 각 worker node가 수행하고 있는 job을 확인 할 수 있으며, 사용자 queue의 상태, 이 DCAF 시스템에서 수행할 수 있는 process type 등을 웹을 통해서 확인 할 수 있다.

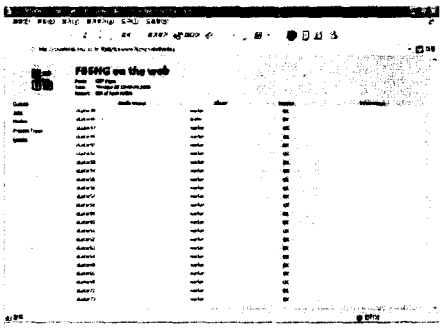

그림 5. fbsWWW

2.5 DCAF의 활용예

DCAF 시스템은 전세계 CDF 그룹 사용자들의 실제 데이터 분석과 몬테카를로 모의 시늉 데이터 생산에 사용될 것이다. 그림 6.은 대표적인 모의시늉의 예로 입자 가속기내의 입자 운동의 궤적을 모의시늉하여, 그 데이터를 CDF RunII EventDisplay[6]로 보여준 것이다.

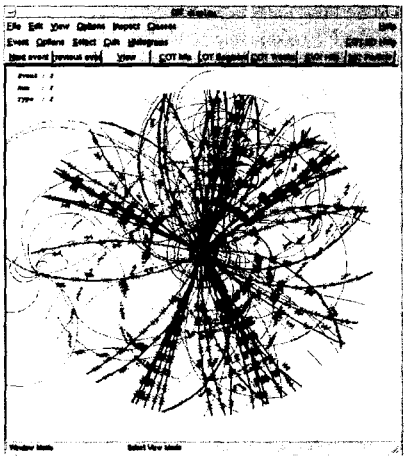

그림 6. CDF RunII EventDisplay

2.6 경북대학교 DCAF의 구성

경북대학교의 DCAF는 총 37 node, 52 cpu로 구성되어 있으며. 각 head node, worker node, sam station, job submission site로 구성되어있다. 2004년 여름 Conference를 위하여 사용자 몬테카를로 모의 시늉 데이터의 생산팜으로 전체 6개의 offine 팜의 CPU power인 1THz의 12%인 약 120 GHz로 늘려 CDF 사용자에 공헌할 것이다.

TYPE	CPU	RAM	HDD	수량
head node	AMD MP2000*2	2G	80G	1
sam station	Pentium 4 2.4G	1G	80G	1
submission site	Pentium 4 2.4G	1G	80G	1
worker node	AMD MP2000*2	2G	80G	4
	AMD MP2200*2	1G	80G	2
	AMD MP2800*2	2G	80G	13
	Pentium 4 2.4G	1G	80G	15
Total	57 CPU	55G	2960G	37

표 1. 경북대학교 DCAF의 구성

3. 결론

경북대학교에 구성된 DCAF 시스템은 앞으로 대량의 모의 시늉 데이터의 생산과 실제 데이터의 분석에 사용될 것이며, 이 데이터는 정밀하게 테스트하여 실험에 참여하고 있는 모든 연구진들이 이용하게 될 것이다. 경북대에서는 앞으로 2006년까지 점차적으로 스토리지 및 Worker node를 증설하여 Run IIb 기간내에 생산되는 데이터를 분석하기 위해서 많은 준비를 할 것이다.

현재 한국외에 대만, 일본, 이탈리아 등 전체 6개의 팜이 운영 및 준비중에 있으며, 앞으로 점차적으로 증가될 것이다. 이렇게 전세계 흩어져있는 분석팜들을 하나로 모아 Grid 시스템을 구현하는 것이 목표라 할수 있겠다. 이러한 DCAF는 Grid 구현을 위한 중간단계라고 할 수 있다. 현재 개발중인 JIM(Job Information Management)이 완성되면 보다 실질적인 Grid의 구현이 이루어질 것이며, 궁극적으로는 CMS의 Tier-1 환경과 통합 구축을 목적으로 하고 있다.

5. 참고문헌

[1] http://www-isd.fnal.gov/fbsng/
[2] http://cdfcaf.fnal.gov/doc/CafSoftware/CafSoftware.html
[3] http://cdfdb-prd.fnal.gov/sam/
[4] http://dcache.desy.de/summaryIndex.html
[5] http://cluster46.knu.ac.kr:8080/
[6] http://www-cdf.fnal.gov/upgrades/computing/projects/display/EventDisplay.html