

탈중앙분산팜(DeCentralized Analysis Farm)의

테스트베드 구축

조기현^{0,1} 오영도¹ 권기환¹ 한대희¹ 손동철¹ 김복주² 이상산³

1. 경북대학교 고에너지물리연구소, 2. 서울대학교 물리학과, 3. 한국과학기술정보연구원 슈퍼컴퓨팅센터
cho@knu.ac.kr⁰, ydoh@fnal.gov, kihwan@knu.ac.kr, hanbi@knu.ac.kr, son@knu.ac.kr, bockjoo@fnal.gov.

sslee@hpcnet.ne.kr

A Test-bed for DCAF (DeCentralized Analysis Farm)

Kihyeon Cho^{0,1}, Youngdo Oh¹, Kihwan Kwon¹, Dae Hee Han¹, Dongchul Son¹, Bockjoo Kim²,
Sangsun Lee³

1. Center for High Energy Physics, Kyungpook National University, Daegu 702-701, Korea
2. Department of Physics, Seoul National University, Seoul 151-742, Korea
3. KISTI, Supercomputing Center, P.O. Box 122, Daejeon 305-806, Korea

요 약

미국 페르미연구소에서 현재 수행되고 있는 고에너지물리 실험의 하나인 CDF 실험에서는 현재 303대의 Dual CPU 클러스터를 이용한 중심분석용팜(CAF, Central Analysis Farm)을 페르미 연구소 내에 구성하여 실제 데이터 처리 및 모의 시뮬 데이터를 생산하는데 사용하고 있다. 그러나 페르미 연구소에서의 중심분석용팜(CAF)은 향후 그 자원이 충분치가 못하므로, 이에 참여하고 있는 여러 나라의 컴퓨팅 자원들을 공유할 수 있어야 한다. 따라서, 한국그룹은 경북대학교 고에너지물리연구소에 있는 PC 클러스터를 활용하여 탈중앙분산팜(DCAF, DeCentralized Analysis Farm)을 국제공동연구로 설계하여 테스트베드를 구축하였다. 이 구성에는 CAF의 기술뿐만 아니라 페르미 연구소라는 원격에 있는 실험 데이터를 이용하여 job을 수행하므로 데이터 전송 기술인 SAM(Sequential data Access via Meta-data) 및 Kerberos의 보안 시스템, 그리드(Grid)를 포함하는 모든 IT 기술의 종합으로 이루어져 있다.

1. 서 론

물질의 근본 구성입자와 그 상호 작용 연구를 통해 궁극적으로 우주 탄생의 비밀을 밝히는 고에너지 물리실험 연구는 한국을 포함, 미국 일본 독일 스위스 등 선진국 연구기관이 서로 협력해 대단위 국제 공동연구로 이루어지는 것이 대부분이다. 왜냐하면, 주로 고에너지 물리는 연구 성격상 물질의 궁극 구조와 작용을 밝히기 위해선 입자 가속기와 검출기 등의 장비를 설치하고 연구하는 것이 필수적이기 때문이다.

고에너지물리실험에서는 검출기에서 생성되는 아날로그(Analog) 전기신호를 디지털화하여, 수 단계 온라인 및 오프라인 필터링(massive parallel processing)을 거쳐 자료분석 컴퓨터에서 자료를 저장하고 사용자 요구에 따라 분석할 수 있도록 제공한다. 이러한 고에너지물리실험 연구소중의 하나인 미국 페르미 연구소는 시카고 서부 교외에 위치하고 있으며 2001년 3월부터 Run IIa 라고 부르는 1.0 TeV의 양성자와 1.0 TeV의 반양성자의 강입자 충돌 실험을 수행할 수 있는 시설과 고에너지 실험을 위한 가속기 사용자 환경을 마련해 주고 있다. 한국,

미국을 비롯한 11개국 55개의 연구소 및 학교로부터 500여명이 넘는 학생, 연구원 및 교수들이 여기에서 CDF(Collider Detector at Fermilab) 실험이라고 불리어지는 강입자 충돌 실험에 참가하고 있다. 한국에서는 경북대학교를 비롯한 서울대학교 및 성균관대학교가 이 실험에 참여하고 있다.

미국 페르미 연구소에서 2005년부터 가동할 CDF Run IIb의 총휘도는 15fb^{-1} 로 현재 Run IIa에서 가동될 데이터(2fb^{-1})의 7 배나 가까이된다. 따라서 원격지에 있는 공동연구진이 이 데이터를 분석하기 위해서 데이터에 대한 즉각적인 접근과 충분한 고속연산 능력을 제공할 수 있는 시설이 필요하나, 현재 페르미 연구소에서의 중심분석용 팜(CAF)은 향후 그 자원이 충분치가 못하므로, 이에 참여하고 있는 여러 나라에서 컴퓨팅 자원들을 공유할 수 있어야 한다. 따라서, 페르미 연구소의 CDF 실험 데이터 저장과 분석을 위해 한국의 컴퓨팅 자원을 이용하는 탈중앙분산팜(DeCentralized Analysis Farm)을 그리드 테스트베드로 구축하였다.

2. 본론

2.1 중심분석팜 (Central Analysis Farm)

지금 페르미 연구소의 CDF 실험에서는 자료 분석과 모의 시뮬 데이터를 위하여 파인만 컴퓨터센터 (Feynman Computer Center at Fermilab)에 187TB의 Disk 서버와 303대의 Dual CPU를 가진 PC 클러스터로 두 단계 (Stage 1과 Stage 2)의 중심 분석팜 (Central Analysis Farm)을 구성하였다. 이 시스템은 주로 MIT 그룹이 구성하였으며, CDF 실험의 연구진은 이 CAF를 이용하여 실험 자료를 처리하고, 모의 시뮬 데이터를 생산하고 있다.

여기서 사용자는 사용자의 데스크탑에 있는 사용자 편의 도구인 CafGui라 불리는 CDF 소프트웨어 패키지를 이용하여 job을 submit하게 된다. 그러면 CAF head node에서 이 job을 Fermilab 소프트웨어 중 하나인 FBSNG라는 batch 시스템을 이용하여 work node에서 batch job을 돌리게 된다. 이 때 데이터는 CAF Data Server의 하드디스크에 있는 파일을 읽거나 Enstore에 있는 테이프 속의 파일을 dCache로 Hard disk에 dump하여 읽게된다. 결과물은 세 군데로 보낼 수 있다. 첫 번째는 job을 submit한 사용자 데스크 탑, 두 번째는 다른 원격에 있는 데스크탑이며, 세 번째는 CAF FTP server에 저장하여 필요할 때 FTP로 결과물을 가져오는 것이다. 이러한 모든 과정은 Kerberos 라는 보안시스템 하에서 이루어진다. 이 CAF 시스템은 연차적으로 CPU와 하드디스크를 확장할 계획이다.

그러나, Run IIa(2001.3 - 2004) 와 Run IIb(2005 - 2006 이후) 로 나누어지는 실험 기간 중 Run IIa의 경우 7.6MHz의 충돌이 발생하고 세 단계에 걸친 필터링 후 약 20Mbyte/s의 데이터를 수집하는 반면, Run IIb에서는 Run IIa보다 7 배 이상의 데이터를 획득한다. Run IIb 기간 동안에는 3 PBytes/year 보다도 더 많은 양을 수집할 것이다. 따라서 이러한 데이터를 다루기 위하여 페르미 연구소 밖의 여러 나라에 있는 컴퓨터 자원들을 공유할 수 있어야 한다. 따라서, 탈중심분산팜(DeCentralized Analysis Farm)기교등을 이용한 그리드 기술(Grid Technology)은 분산 컴퓨팅 자원(Distributed Computing Resource)을 공유할 수 있는 필수적 요소이다.

2.2 탈중심분산팜(DeCentralized Analysis Farm)

DCAF에서 job을 submit하고 batch job을 돌리는 것은 CAF와 개념이 같다. 그러나 원격지인 미국 페르미연구소에 있는 실험 데이터 파일을 가져오는 것

은 SAM(Sequential data Access via Meta-data)을 이용하여 미국 페르미 연구소의 Enstore에 있는 Tape속의 파일을 실제의 원격지 팜에 dCache로 가져오게 된다. 2002년 11월에 이러한 DCAF의 개념을 SAM Grid에 접목하여 CDF와 D0 그룹에서 Testbed를 구성하였다. 이 Testbed 구성에는 CDF 실험에서 페르미연구소이외에 6개 연구기관이 참가하였는데, 미국의 Rutgers, Texas Tech대학, 캐나다의 Toronto대학, 영국의 RAL 및 Imperial College 와 한국의 경북대학교가 참가하였다. 그림 1은 CDF와 D0 그룹이 테스트 베드를 구성하여 SAM 그리드의 모니터링을 위한 초기 홈페이지를 나타낸 것이다[1].



그림 1. DCAF지역의 SAM 그리드의 모니터링을 위한 초기 홈페이지

2.3 한국의 탈중심분산팜 (KCAF, DeCentralized Analysis Farm in Korea)

한국 그룹은 경북대학교 고에너지물리연구소에 있는 클러스터 중 6대의 Dual CPU인 AMD 2000 (1.6GHz)를 이용하여 DCAF를 설계 제작하여 execution site로 test-bed를 구성하였다. 그림 2는 한국 그룹이 디자인한 DCAF 시스템의 구성도 이다. 우선 페르미연구소에서 submit한 job이 job broker에 의해 경북대학교에서 수행하도록 되도록 만들었으며, 경북대학교도 submission 사이트를 구축할 예정이다. 10,000 km 떨어진 미국 페르미연구소에 있는 실험 데이터 파일을 가져오는 것은 SAM (Sequential data Access via Meta-data)을 이용한다. SAM을 이용하여, 미국 페르미연구소의 Enstore에 있는 Tape속의 파일을 직접 디스크로 가져와서 job을 수행하는 것을 dCache라고 부르며, DCAF에서는 실제 데이터를 돌리기 위하여 이 기술이 필수적이다. 모의 시뮬 데이터를 생산하기 위하여 Calibration Data가 필요한데 실제 모의시뮬을 위한 calibration

data base는 Oracle database를 사용한다. CAF와 두 유럽 또는 미국에 있으므로 미국의 페르미연구소와

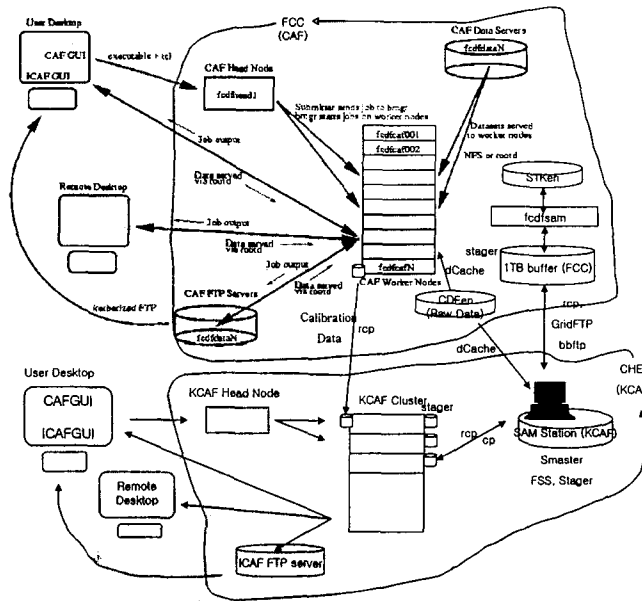


그림 2. 경북대학교 고에너지물리연구소에 구성된 DCAF의 구성도

마찬가지로 이러한 모든 과정은 Kerberos라는 보안 시스템 하에서 이루어진다. 향후, 이 시스템은 한국, 일본 및 대만을 묶는 CDF 아시아 지역의 사용자에게 공헌 할 뿐만 아니라 전 세계적으로 CDF의 다른 사용자들에게도 공헌하게 된다.

3. 결과

페르미 연구소에서 job을 경북대학교의 work node에 job을 수행함으로써 다음과 같은 사실을 알 수 있었다.

가. NFS vs. local

우리는 클러스터에 NFS로 mount된 directory와 local home에 directory 사이의 차이로 인해서는 아무런 문제가 발생하지 않음을 알 수 있었다.

나. NIS vs. 사용자

우리는 NIS 와 local 사용자 사이에 큰 문제가 없음을 알 수 있다.

다. 한국과 페르미연구소 사이의 network의 connection은 아주 느리며 submission site에서 submit한 job이 idle로 오랫동안 있음을 알 수 있다. 이 기간 동안에는 status와 관계하여 알 수 있는 유용한 정보가 필요하였다. 이것은 모든 사이트에서 공통적이었으나, 특히 약 10,000Km의 원거리에서 상대적으로 느린 connection(17Mbps)에 있는 페르미 연구소와 한국 사이에 문제가 두드러졌다. 한국을 제외한 다른 노드는 모

인접하며 지금 현재 유럽과 미국사이의 TeraGrid 프로젝트로 bandwidth가 상대적으로 빠르다.

4. 결론

한국의 탈중심분산망(DCAF) 시스템은 많은 양의 모의 시뮬 데이터를 만들어 내는 것과 실제 데이터를 분석하는데 이용될 것이다. 만들어진 데이터는 정밀하게 테스트되어 문제가 없을 경우 실험장소인 페르미 연구소의 파인만 컴퓨터 센터로 전송되어 STKen이라는 스토리지에 저장되어 실험에 참여하고 있는 모든 연구진들이 이용하게 된다.

이 것은 현재의 PC 클러스터와 Meta Data를 통한 파일 Access 및 보안까지 포함한 그리드의 초기 단계로 종합 구성되어 있다. 이러한 구성은 국제 공동연구를 수행하고 있는 다른 실험의 원격 사이트의 Analysis Farm을 만드는 것에도 도움을 줄 것이다. 그리고 이러한 알고리즘은 Data Grid로서 글로벌 한 데이터를 사용하는 다른 여러 분야에서도 활용할 수 있을 것이다.

5. 참고문헌

[1] <http://samadams.fnal.gov:8080/prototype>