

고에너지물리 e-Science 연구환경의 구현

An embodiment of e-Science for High Energy Physics

조기현, 김현우)
한국과학기술정보연구원

Kihyeon Cho, Hyunwoo Kim
KISTI (Korea Institute of Science and Technology
Information)

요약

고에너지물리 e-Science 연구 환경 구축은 데이터가 생산되는 외국의 가속기 연구소에 가지 않더라도, 언제나 어디서나 실제 가속기 연구소에서 고에너지물리 연구를 수행하는 것과 같은 연구 환경을 만드는 것이다. 그 구성 요소로서 1) 데이터 생산, 2) 데이터 프로세싱, 3) 데이터 publication이 있다. 데이터 생산은 원격 제어시스템을 활용하여 원격으로 데이터 생산에 참여하는 것이며 데이터 프로세싱은 그리드 팜을 활용하여 job을 언제나 어디서나 수행하는 것이며 데이터 publication은 EVO(Enabling Virtual Organization) 시스템을 활용하여 공동 협업 환경으로 연구 결과물을 얻는 것이다. 이러한 개념을 고에너지물리 실험의 하나인 CDF 실험에 적용하여 구현하였다.

Abstract

The e-Science for High Energy Physics is to study high energy physics anytime and anywhere even if we are not on-site of accelerator laboratories. The concepts are 1) data production, 2) data processing and 3) data publication. The data production is to do remote control and take shifts remotely. The data processing is to run jobs anytime, anywhere with grid farms. The data publication is to work together to publish papers using collaborative environment such as EVO (Enabling Virtual Organization) system. We apply this concept to high energy physics, especially, CDF experiment and show the results.

I. 서론

인류는 오래 전부터 ‘물질은 도대체 무엇으로 이루어져 있을까? 그 사이의 상호 작용은 무엇일까?’라는 질문을 해 왔다. 이러한 연구를 하는 학문을 고에너지물리학이라고 한다.

고에너지물리 연구는 대규모 가속기와 검출기를 필요로 하며, Pbyte급의 대용량의 데이터가 생산된다. 그러므로 이를 처리하기 위한 고에너지물리 e-Science 연구 환경은 데이터 센터를 구축하여, 데이터가 생산되는 외국의 가속기 연구소에 있지 않더라도, 언제나 어디서나 실제 가속기 연구소에서 고에너지물리 연구를 수행하는 것과 같은 연구 환경을 만드는 것이다.

그 구성 요소로서 1) 데이터 생산, 2) 데이터 프로세싱, 3) 데이터 publication이 있다. 데이터 생산은 원격 제어시스템을 활용하여 원격으로 데이터 생산에 참여하는 것이며 데이터 프로세싱은 그리드 팜을 활용하여 job을 언제나 어디서나 수행하는 것이며, 데이터 publication은 EVO(Enabling Virtual Organization) 시스템을 활용하여 연구자들이 공동 협업을 이용하여 그 결과물을 출판하는 것이다.

이러한 개념을 미국 페르미연구소 CDF 실험에 구현하여 연구에 활용한 사례를 보여준다.

II. 고에너지물리

물질의 근본 구성입자를 연구하기 위해서는 관찰할 소스(source)와 검출기가 필요하다. 예를 들면, 테니스 공(target)을 관찰하기 위해서는 눈(eye)으로 테니스 공을 살펴보면 된다. 이 때 source는 광원이 되고, 테니스공이 target이 되며 검출기(detector)는 눈이 된다. 그러나 관찰하고자 하는 대상마다 소스(source)와 검출기(detector)가 다르다. 은하를 관찰하기 위해서 전파망원경이 필요하며, 행성을 관찰하기 위해서는 망원경이, 세포를 관찰하기 위해서는 현미경이 필요하듯이, 핵속에 있는 기본입자(target)들을 관찰하려면 가속기(source)가 필요하고 대규모의 검출기가 필요하다[1].

특히, 고에너지 가속기가 필요한 이유는 드브로이의 물질파 이론에 의하여, 고에너지는 작은 입자를 관찰할 수 있는 작은 파장을 만들 수 있기 때문이다. 그리고 아인슈타인의 상대성 원리에 의하여 질량과 에너지는 같으므로 고에너지는 질량이 큰 입자를 만들 수 있다. 그러므로 고에너지 물리학 분야는 고

1) 교신저자(hyunwoo@kisti.re.kr)

에너지 가속기를 이용하여 우주의 초기 단계를 만들어 우주 생성 및 진화에 관한 학문을 연구하는 기초 과학인 것이다[1].

고에너지 물리 실험은 검출기 설계, 제작, 신호처리 및 자료 수집, 분석에 이르기까지의 일련의 작업을 가속기가 있는 연구소에서 국제 공동 연구로 수행한다. 적게는 100여명, 많게는 전 세계에 분산된 2000여명이 동시에 자료를 처리하기도 한다 [1].

III. 미국 페르미연구소 CDF 실험

미국 페르미연구소 CDF 실험은 미국, 이탈리아, 일본, 영국, 독일, 한국, 대만 등 13개국의 620여 명의 물리학자들이 참여하는 대규모의 국제적인 가속기실험이다. 그 목적은 각각 1 TeV의 에너지로 가속시킨 양성자와 반양성자들을 정면 충돌시켜 새로운 입자와 물리현상들을 탐구하는 것이다. 실험은 2001년부터 시작되어 2009년까지 수행될 예정이다. 충돌횟수는 1초에 760만 번이고 매초 입자 검출기에 기록된 데이터의 크기는 25~40 MByte에 달한다. 현재까지 Run II에서 모은 데이터 양은 2 fb^{-1} 이며, 이것은 3.0×10^9 사건 수에 해당하며, 수 PByte의 데이터를 생산하였다. 이러한 방대한 데이터를 처리하고 분석하기 위해서는 대용량의 저장장치 및 계산 장치와 네트워크 환경이 구축되어야 하고 데이터 그리드는 그 해답을 제시해 준다[1].

CDF 실험의 컴퓨팅 모델은 CAF(CDF Analysis Farm)의 구성에 두고 있으며, 분산된 자원을 활용하여 자료를 처리한다. 그 첫 번째 단계는 DCAF(Decentralized CDF Analysis Farm)이며 그 다음 단계는 그리드를 활용하는 것이다. 국내에는 경북대학교에서 DCAF를 구축하였으며[2], KISTI 그리드 팜을 구축하여 CDF job을 수행할 수 있도록 하였다.

IV. 고에너지물리 e-Science의 구성 요소

1. 데이터 생산

일반적으로 고에너지물리 연구는 가속기가 있는 on-site에서 데이터를 획득한다. 그런데 e-Science의 개념을 적용하여 언제 어디서나 데이터를 획득하는 시스템을 구축한다. 그 방법은 원격 제어 시스템을 이용하는 것이다. 그 예로서 미국 페르미연구소에는 유럽입자공동물리연구소(CERN)의 LHC (Large Hadron Collider) 실험을 위한 ROC(Remote Operation Center)가 있다.

이와 유사하게 KISTI에서 CDF 실험을 위한 원격제어실을 구축하고 있다. 이 시스템이 완성이 되면 국내 CDF 사용자들이 미국 페르미연구소에 상주하지 않고도 한국 KISTI에서

CO(Consumer Operator) 근무를 할 수 있게 된다.

2. 데이터 프로세싱

그림 1은 고에너지물리 e-Science(e-HEP)의 데이터 프로세싱의 아키텍처를 보여준다. e-HEP의 데이터 프로세싱을 위한 주 전산자원은 네트워크와 컴퓨팅이다.



▶▶ 그림 1. e-HEP 데이터 프로세싱 아키텍처

먼저 네트워크를 설명하면 KREONET(Korea Research Environment Open NETwork)은 과학기술부의 지원을 받아 KISTI에서 운영하는 국가 R&D 네트워크이다. KREONET은 GLORIAD(Global Ring Network for Advanced Applications Development) 망에 연결되어 있다. GLORIAD 망은 홍콩, 한국, 미국 사이에 10 Gbps로 연결되어 있으며 KISTI는 CERN과 직접 10 Gbps 네트워크로 연결이 가능하다[3].

두 번째 전산자원은 컴퓨팅이다. 현재와 미래의 고에너지물리 활동은 대규모 데이터 처리가 필요하므로 데이터 그리드와 대용량 저장장치는 불가피하며, CPU는 확장 가능하여야 한다. 데이터는 투명하여 사용자인 고에너지물리 연구자가 어디에 있더라도 실제 데이터 저장 장소를 알 필요가 없어야 한다[4].

미들웨어 자원으로는 LCG(LHC Computing Grid) 팜을 운영하고 있으며, ALICE 실험과 CDF 실험에서 가상기관(VO, Virtual Organization)을 구축하여 활용하고 있다. KISTI에 EGEE 미들웨어인 LCG를 설치하여 공식적인 인정을 받았다. 현재는 30 노드(30 CPU)와 2 TByte 저장장치를 두고 있다. 2007년 말 KISTI의 LCG 팜은 50노드(800 Core processes)와 50 TByte로 확대할 예정이다.

3. 데이터 Publication

데이터 Publication은 전세계 분산된 연구자들이 협업 연구

환경을 통하여 연구 결과물을 출판하는 것이다. KISTI는 EVO 시스템을 한국 최초로 유치하여 국내 연구자들에게 제공하고 있다. 한국 연구자들은 EVO 시스템의 한국 서버를 사용하여, 기존의 미국 서버를 활용할 때 보다 지연속도를 60msec 이상 줄였으며, 한국-미국간의 네트워크 혼잡도 없이, 안정적인 협업 연구를 수행한다. 그림 2는 EVO 시스템이 KISTI 서버와 연동되고 있음을 보여준다.



▶▶ 그림 2. EVO 시스템의 모니터링 시스템

IV. 결 과

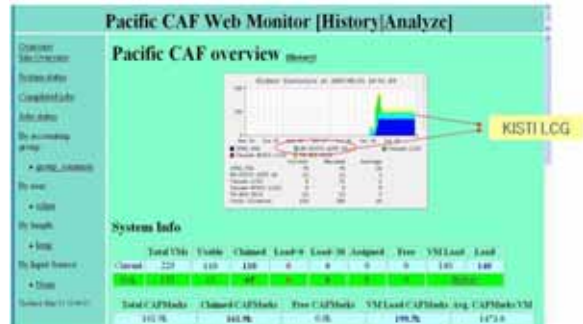
CDF 실험에는 여러가지 컴퓨팅 프로세싱 방법이 활용되고 있다. KISTI는 그 중에 LCG 팜을 활용하여, Globus위의 Condor를 사용하는 시스템을 구축하였다. 이 시스템의 장점은 아주 유연하다는 것이다[5]. 이러한 시스템의 구현은 LHC 실험이 본격 시작되기 전에 CDF 실험을 위한 CPU를 확보할 수 있으며, 전 세계 고에너지물리 연구가 그리드 팜을 활용하는 추세에 부합된다.

KISTI LCG 팜은 2007년 5월부터 본격적으로 서비스를 시작하여 2007년 9월 30일 현재까지 submit된 job 수는 9,866이며, 성공한 job의 수는 9,485로 96%의 성공적인 job 수행률을 보이고 있다.

KISTI LCG팜은 대만의 ASGC(Academi Sinica Grid Center)의 OSG(Open Science Grid) 팜 및 LCG 팜과 연동되어 있으며, 이 연동된 팜을 "Pacific CAF"라 부른다. Pacific CAF의 헤드노드는 ASGC에 있으며, KISTI LCG 팜은 Pacific CAF의 워크노드로 활용된다. 구체적으로, 한국의 KISTI (KR-KISTI-GCRT-01) 팜과 대만의 4개의 LCG 및 OSG (Open Science Grid) 팜(IPAS_OSG, Taiwan-LCG2, Taiwan-NCUCC-LCG2 및 TW-NIU-EECS)이 연동되어 있다.

그림 3은 KISTI 워크노드를 포함한 Pacific CAF의 웹 모

니터링 시스템을 보여준다. 향후 경북대 팜과 일본 쓰꾸바대학교 팜이 Pacific CAF 팜에 포함될 것이다.



▶▶ 그림 3. Pacific CAF 웹 모니터링 시스템

V. 결 론

고에너지물리 분야는 e-Science 연구 환경 구축의 선도 모델을 보여주고 있다. 고에너지물리 연구를 언제 어디서나 수행하기 위한 고에너지물리 e-Science 연구 환경을 미국 페르미연구소 CDF 실험에 적용하여 데이터 생산, 데이터 프로세싱, 데이터 publication등의 일련의 e-Science 연구 환경을 구현하였다.

■ 참고 문헌 ■

- [1] 조기현 등, "고에너지물리 데이터 그리드 시스템의 구현", 정보과학회논문지, 제33권, 제7호, pp.390-398, 2006.
- [2] M.Neubauer et al., "Computing for RunII at CDF", Nucl. Instr. and Meth. A479, 117 (2002).
- [3] Kihyeon Cho, "Cyberinfrastructure in Korea", Computational Physics Communication, Vol. 177, No. 1-2, pp. 247-248, 2007.
- [4] Kihyeon Cho, "A test of the interoperability of grid middleware for the Korean High Energy physics Data Grid system", Int. Journal of Computer Science and Network Security, Vol. &, No. 3, pp. 49-54, 2007.
- [5] Igor Sfligoi, "CDF Computing", Computer Physics Communications, Vol. 177, No. 1-2, pp 235-238, 2007.