

2008

「Global Convergence 시대를 열자」

ISSN 1226-9182

정보처리학회지

Korea Information Processing Society Review

2008. 3 제15권 제2호

특 집

e-Science

한국 e-Science의 현재와 발전방향

e-Science 국내외 기술 동향

e-Science 문체해결환경 구축을 위한 워크벤치 설계 및 구현

e-Science 워크벤치를 위한 그리드 서비스 통합 기술

e-Science 연구를 위한 협업환경 기술 소개

e-Science를 위한 지식관리 기반구조



사단
법인

한국정보처리학회

Korea Information Processing Society


 사례
발표

고에너지물리 e-Science 연구 환경의 구축 및 활용


 목 차

1. 서 론
2. 사용자 요구사항
3. 사용자에게 제공되는 서비스
4. 모델 시스템 구조
5. 활용 현황

공대정 · 김현우 · 조기현
(경북대학교 · 한국과학기술정보연구원)


 <요 약>

e-Science를 기반으로 한 고에너지물리 연구 환경 구축은, 실제로 실험 데이터가 생산되는 외국의 가속기 연구소에 가지 않고서도, 언제 어디서나 실제 가속기 연구소에서 고에너지물리 연구를 수행하는 것과 같은 연구 환경을 제공한다. 그 구성 요소로서는 1) 데이터 생산(production), 2) 데이터 프로세싱(processing), 3) 데이터 분석(analysis)이 있다. 데이터 생산은 원격 제어시스템(remote control mom)을 구축하여 원격으로 데이터 생산에 참여하는 것이며, 데이터 프로세싱은 그리드 컴퓨팅을 활용한 데이터 처리를 뜻하며 데이터 분석은 EVO(Enabling Virtual Organization) 시스템을 활용하여 공동 협업 환경으로 연구 결과물을 얻는 것이다. 이러한 개념을 고에너지물리 실험의 하나인 CDF 실험에 구현하여 활용한 사례를 보여 준다.

1. 서 론

고에너지 물리학은 물질의 근본구성입자는 무엇인가? 그리고 그 구성 입자들간의 상호작용은

무엇인가?라는 두 가지 질문이 대한 대답을 찾고자 하는 실험물리학이며, 실험입자물리학이라고도 불리고 이론입자물리학과는 서로 상호보완적인 관계에 있다.

고에너지물리 연구는 대규모 가속기와 검출기를 필요로 하며, 검출기 설계, 제작, 신호처리 및 자료 수집, 분석에 이르기까지의 일련의 작업을 가속기가 있는 연구소에서 극세 공동 연구로 수행한다. 또 한 Pbyte급의 대용량의 데이터가 생산되며, 조직은 100여명, 많게는 전세계 분산된 2,000 여명이 동시에 자료를 처리하기도 한다[1].

그러므로 이를 위한 고에너지물리 e-Science 연구는 데이터 처리를 구축하여, 데이터가 생산되는 외국의 가속기 연구소에 있지 않더라도, 언제나 어디서나 실제 가속기 연구소에서 고에너지물리 연구를 수행하는 것과 같은 연구를 하는 것이다.

이러한 연구는 IT 기반의 과학기술 연구 활동(실험, 컴퓨팅, 이론)과 연계하여 융합할 수 있는 환경 구축 및 활용 연구이다. 또한, 슈퍼컴퓨터를 포함한 첨단 연구 장비(사이버 R&D 인프라)

라의 원격실험 및 효율적 활용을 위한 가상실험실 기반의 서비스를 개발하여 연구를 하는 것으로 국가 과학기술 기본계획과도 일치한다[2].

이 논문은 이러한 개념을 미국 페르미연구소 CDF (Collider Detector at Fermilab) 실험에 구현하여 연구에 활용하고 있는 사례를 보여준다.

미국 페르미연구소는 일리노이주 시카고 근교에 있는 국립 입자 가속기 연구소로 양성자 반양성자를 각각 1TeV로 가속시킬 수 있는 네바르론 가속기를 보유하고 있다. 연구소는 1968년 개소한 이래 미국 각 대학과 전 세계에서 모인 천여 명의 물리학자들에 의해 대단위의 실험들이 진행되어 왔다. 이 곳에 있는 가속기는 양성자-반양성자 충돌형 가속기인데, 2008년 유럽입자물리공동연구소(CERN)의 강입자충돌기(LHC, Large Hadron Collider)가 가동될 때까지, 원존하는 세계에서 가장 큰 에너지인 2TeV의 에너지 빔을 생성시켜 서로 입자와 물리현상들을 탐구하는 것이다. 빔은 일종의 에너지 다발인데, 각각의 다발에 있는 양성자-반양성자의 개수는 약 10조 개에 이른다. 이 양성자-반양성자 빔을 충돌시키면, 이 충돌은 곧 이들 입자 내부의 엄청난 파괴로 이어진다. 이러한 파괴에 의해 이들 입자는 다른 입자로 변하게 된다. 연구소의 검출기는 이것을 검출할 수 있도록 충돌지점을 중심으로 입자들을 어휘싸고 있으며, 연구자들은 충돌 후에 검출기로부터 나오는 여러 신호들을 분석해 물리현상을 설명한다[3].

페르미연구소에서 사용하는 검출기의 이름을 CDF (Collider Detector at Fermilab)라 부르며, 이 검출기를 사용한 실험을 CDF 실험이라 부른다. CDF 실험은 미국, 이탈리아, 일본, 영국, 독일, 한국, 대만 등 13개국의 620여 명의 물리학자들이 참여하며 국내에서는 경북대, 서울대, 실용科大, KISTI 및 전남대의 30여 명의 물리학자가 참여하고 있다. CDF 실험은 2001년부터 시작되어 2009년까지 수행될 예정이다.

2. 사용자 요구사항

고에너지물리 연구자는 지리적 위치나 대용량 자료의 크기에 관계없이, 언제나 어디서나 고에너지물리 연구를 수행하길 원한다. 이러한 연구가 고에너지물리 e-Science 연구이며, 그 구성 요소로서는 1) 데이터 생산, 2) 데이터 프로세싱, 3) 데이터 publication이 있다.

데이터 생산은 원격 제이스터널을 활용하여 가속기 연구소에 있지 않더라도 데이터 생산이 참여하는 것이다. 사용자는 언제나 어디서나 데이터 획득이 가능해야 함으로 데이터 획득 수형은 가속기가 있는 페르미연구소에서 수행할 필요는 없다.

CDF 실험에서 충돌횟수는 1초에 760만 번이고 페츠 입자 검출기에 기록된 데이터의 크기는 25~40 MByte에 달한다. 현재까지 Run II에서 모은 데이터 양은 2 Tb-1 이며, 이것은 3.0×10^9 사진 수에 해당하며, 수 PByte의 데이터를 생산하였다. 이러한 방대한 데이터를 처리하고 분석하기 위해서는 대용량의 저장장치, 계산 장치와 네트워크 환경이 구축되어야 하고 데이터 그리드는 그 역할을 제시해 준다[1]. 데이터 프로세싱은 이러한 데이터 그리드를 활용하여 job을 언제나 어디서나 수행하는 것이다.

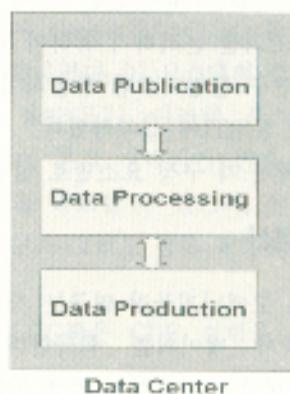


그림 1) e-Science 연구환경 데이터센터 모델

데이터 publication은 연구자들이 공동 협업 환경을 이용하여 그 결과물을 출판하는 것이다. 결론적으로, 이러한 모든 것을 포함하는 데이터 센터를 구축하여 사용자에게 언제나 어디서나 웹 스태프로 연구 환경을 제공한다. (그림 1)은 사용자가 요구하는 e-Science 연구 환경의 데이터 센터 모델을 보여준다.

3. 사용자에게 제공되는 서비스

CDF 실험에서 사용자 요구에 의하여 사용자에게 제공되는 서비스는 1)언제나 어디서나 데이터 생산을 할 수 있는 원격제어시스템 2)대용량의 자료를 언제나 어디서나 처리할 수 있는 Pacific CAF(CDF Analysis Farm), 3)협업연구 활동으로 데이터 처리 결과물을 출판하는 EVO (Enabling Virtual Organization) 시스템이 있다.

CDF 실험의 컴퓨팅 모델은 CAF(CDF Analysis Farm)의 구성에 두고 있으며, 분산된 자원을 활용하여 자료를 처리한다[4]. 그 첫 번째 단계는 탈중앙분산팜(DCAF, Decentralized CDF Analysis Farm)이며 그 다음 단계는 그리드를 활용하는 것이다. 국내에는 경북대학교에서 DCAF를 구축하였으며[5], 최근에는 KISTI 그리드 팜을 활용하여 CDF job을 수행할 수 있도록 하였다.

또한 전세계 13개국이 분산된 620여명의 물리학자들이 그 결과를 논의하기 위하여 EVO 시스템이란 협업연구 환경시스템 활용하며 그 서비스를 국내 최초 KISTI에 설치하였다.

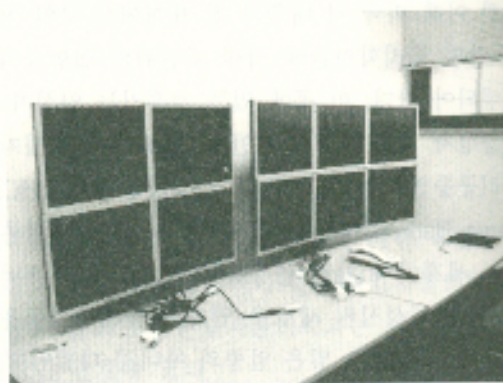
다음 세션에서 각 구성 요소별로 설명을 한다.

3.1 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 유럽입자물리연구소(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. (그림 2)는 현재 KISTI내에 구축 중인 원격제어실의 모니터링 시스템을 보여준다.



(그림 2) 원격제어실의 모니터링 시스템

3.2 데이터 프로세싱

현재와 미래의 고에너지물리 실험은 대규모 데이터 처리가 필요하므로 데이터 그리드와 대용량 저장장치는 불가피하며, CPU는 확장 가능하여야 한다. 데이터는 두명하여 사용자인 고에너지물리 연구자가 어디에 있더라도 실제 데이터 저장 장소를 알 필요가 없어야 한다. 미국 페르미연구소의 CDF 실험의 데이터 분석은 현재 그 방대한 크기의 데이터 양의 원활한 처리를 위해, 세계 각지에 분산되어 있는 CDF 구성 기관들이 보유하고 있는 전산자원과 저장장치를 그리드 개념을 이용해서 처리하는 방식으로 진화하고 있다. KISTI는 미국 페르미연구소와 10Gbps의 GLORIAD망으로 연결되어 있으며

[6]. 파일 전송을 위한 SAM (Sequential Access thru meta-data) Data Handling 시스템을 공동 연구 하고 있다.

3.3 데이터 분석

데이터 분석(Analysis)은 전세계 분산된 고에너지물리 연구자들이 협업 연구 환경을 통하여 연구 결과물을 분석하여 그 결과를 일반에 공개하도록 하는 시스템을 일컫는다. KISTI는 EVO (Enabling Virtual Organization) 시스템을 한국 최초로 유지하여 국내 연구자들에게 제공한다. 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.

4. 포탈 시스템 구조

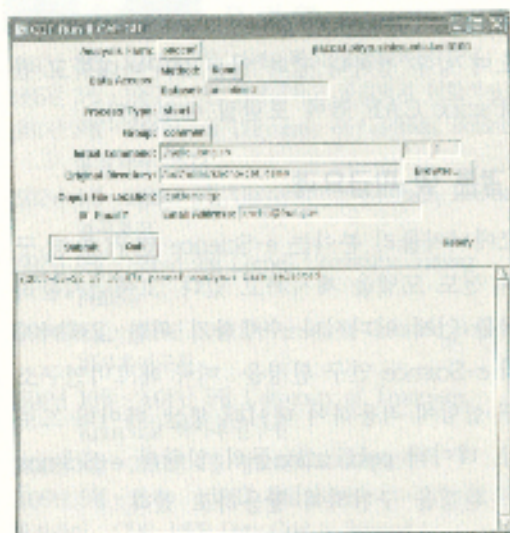
최근 CDF 실험에서는, 미국 위스콘신 주립대에서 개발된 Condor 라는 컴퓨터 클러스터 관리 프로그램에서 최근 개발한 Glideins라는 개념을 도입하였다. 이 개념의 핵심은 기존에 그리드화 되어 있는 각각의 기관들에 최소한의 변화만을 요구해서 그 기관을 Condor의 워크노드(worker node)로써 작동하게 만들 수 있다는 것이다.

이 개념을 아시아 대륙양 지역의 기존 CDF 구성 기관들에 전산자원을 접목시킨 것이 바로 Pacific CAF(CDF Analysis Farm)이다. Condor 시스템의 구조에 따르면 워크 노드 들을 관리하는 헤드노드(head node)가 필요한데, 대만 Academia sinica라는 기관이 헤드노드를 맡다, 그 헤드노드의 관리하여 그리드로 변동될 수 있는 워크노드를 구축 하고 있다. 여기에 참여하는 연구기관으로 대만의 Academia Sinica, 일본 쓰쿠바대학, 한국 KISTI 및 경북대 등이 있다.

현재 이 목적으로 KISTI 내 40개의 리눅스용 클러스터를 활용하고 있으며, 향후 그 노드를 확

장할 예정이다. 경북대는 현재 그리드화 되기 전 단계인 탈중심분산망을 구축하여 CDF 그룹이 활용하고 있으나, 향후 그리드화하여 Pacific CAF의 워크노드로 연동할 예정이다.

(그림 3)은 사용자가 job을 submit하는 CAFGUI(CDF Analysis Farm Graphic User Interface)를 보여준다.



(그림 3) CAF GUI

5. 활용 현황

CDF 실험에는 여러가지 컴퓨터 프로세싱 방법이 활용되고 있다. KISTI는 그 중에 LCG 방식을 활용하여, Globus와의 Condor를 사용하는 시스템을 구축하였다. 이 시스템의 장점은 아주 유연하다는 것이다[7]. 이러한 시스템의 구현은 LHC 실험이 본격 시작되기 전에 CDF 실험을 위한 CPU를 확보할 수 있으며, 전 세계 고에너지물리 연구가 그리드 망을 활용하는 추세에 부합된다.

KISTI LCG판은 CDF 실험을 위한 대만의 ASGC(Academia Sinica Grid Center)의 OSG(Open Science Grid) 판 및 LCG 판과 연동되어 Pacific CAF를 구성한다. Pacific CAF의

헤드노드는 ASGC에 있으며, KISTI LCG 폼은 Pacific CAF의 워크노드로 활용된다. 구체적으로, 한국의 KISTI (KR-KISTI-GRCT-01) 등과 대만의 4개의 LCG 및 OSG (Open Science Grid) 폼(IPAS_OSG, Taiwan-LCG2, Taiwan-NCUCC-LCG2 및 TW-NIU-EECS)이 연동되어 있다.

KISTI LCG 폼은 2007년 5월부터 본격적으로 서비스를 시작하여 성공적으로 작업을 수행하고 있으며 향후 경북대 등과 일본 쓰쿠바대학교 등이 Pacific CAF 폼에 포함될 예정이다.

6. 결론 및 파급효과

고에너지물리 분야는 e-Science 연구 환경 구축의 선도 모델을 제시하고 있다. 고에너지물리 연구를 언제 어디서나 수행하기 위한 고에너지물리 e-Science 연구 환경을 미국 페르미연구소 CDF 실험에 적용하여 데이터 생산, 데이터 프로세싱, 데이터 publication 등의 일련의 e-Science 연구 환경을 구현하여 활용하고 있다.

그 파급효과로서 사용자 중심의 커뮤니티에서 요구한 데이터 생산, 프로세싱, publication까지의 원스탑의 데이터센터로 연구자에 제공함으로써 연구 생산성 향상 및 타 분야 과학의 e-Science 연구의 시범 모델이 된다. 또한 CDF 실험에서 구축한 e-Science 연구 환경은 국내 사용자 뿐만 아니라, 전세계 620여명의 사용자들이 활용함으로써 국제 공동 CDF 실험 연구에 한국이 기여하여 국제 협력 및 기술 개발의 우수성을 보여준다.

참고문헌

- [1] 조기현 등, "고에너지물리 데이터 그리드 시스템의 구현", 정보과학회논문지, 제33권, 제7호, pp.330-336, 2006.
- [2] 과학기술부, 제2기 (2008-2012) 과학기술기본계획.
- [3] 김동희, "페르미연구소의 고에너지물리 CDF 한국그룹" 경북대 신문 (2003.12.01)
- [4] M.Neubauer et al., "Computing for RunII at CDF", Nucl. Instr. and Meth, A479, 117 (2002).
- [5] 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. 8, No. 3, pp. 49-54, 2007.
- [6] Kihyeon Cho, "Cyberinfrastructure in Korea", Computational Physics Communication, Vol. 177, No. 1-2, pp. 247-248, 2007.
- [7] Igor Sffigoi, "CDF Computing", Computer Physics Communications, Vol. 177, No. 1-2, pp. 235-238, 2007.

저자약력



김대정

1988년 경북대학교 물리학과 학사
 2000년 경북대학교 물리학과 석사
 2006년 경북대학교 물리학과 박사
 2008년 2월 ~ 2009년 12월 기초과학지원연구원 방문연구원
 2009년 5월 ~ 2009년 9월 미국 페르미연구소 초청연구원
 2006년 ~ 현재 경북대학교 물리학과 박사후연구원
 관심분야 : CDF, CMS, Data Processing
 이 메 일 : djkim@hpc.knu.ac.kr



소기연

1981년 2월 ~ 1985년 2월 연세대학교 물리학과 석사
 1985년 3월 ~ 1987년 2월 연세대학교 물리학과 이학석사
 1991년 8월 ~ 1996년 5월 University of Colorado, Boulder
 물리학과 Ph.D.
 1996년 1월 ~ 1999년 5월 University of Colorado, Boulder,
 연구조교
 1992년 9월 ~ 1996년 9월 Cornell University, Visting
 Fellow
 1996년 5월 ~ 1997년 10월 University of Colorado Boulder,
 박사후연구원
 1997년 10월 ~ 2001년 5월 University of Tennessee,
 Knoxville 박사후연구원
 2001년 6월 ~ 2009년 4월 경북대학교 초빙교수
 2006년 5월 ~ 현재 KISTI 책임연구원
 관심분야 : CDF, HEP Data Grid, e-Science
 이 메 일 : cho@kisti.ac.kr



김연우

1994년 고려대학교 물리학과 학사
 1996년 고려대학교 물리학과 석사
 2006년 고려대학교 물리학과 박사
 2008년 ~ 2009년 텍사스주립대 박사후연구원
 2009년 ~ 2007년 성균관대학교 BK21 연구교수
 2007년 ~ 현재 KISTI 책임연구원
 관심분야 : CDF, systems programming
 이 메 일 : hyunwoo@kisti.ac.kr