

고에너지물리 데이터 그리드 시스템의 구현

(An Embodiment of High Energy Physics Data Grid System)

조기현[†] 한대희^{**} 권기환^{**} 김진철^{**} 양유철^{***}
 (Kihyeon Cho) (Daehee Han) (Kihwan Kwon) (Jincheol Kim) (Yuchul Yang)

오영도^{***} 공대정^{**} 서준석^{****} 김동희^{****} 손동철^{****}
 (Youngdo Oh) (Daejung Kong) (Jun-Suhk Suh) (Donghee Kim) (Dongchul Son)

요약 고에너지 물리학(HEP, High Energy Physics)은 물질의 근본 구성 입자와 상호 작용 연구를 통해 궁극적으로 우주 탄생의 비밀을 밝히는 학문이다. 2007년에 시작하는 유럽입자물리연구소(CERN)의 대형강입자충돌기(LHC, Large Hadron Collider)의 CMS(Compact Muon Solenoid)실험에 참여하는 연구진은 2000여명이나 되며, 생산되는 데이터 양은 연간 수 PetaByte에 달할 예정이다. 그러므로 고에너지물리 실험에서 생산하는 데이터는 기존의 전산자원의 개념으로 처리하는 것이 불가능하다. 그리하여 고에너지물리 분야에서 자료의 계층적 구조(Tier-0, 1, 2) 및 데이터 그리드를 활용하게 되었으며, 이러한 고에너지물리 데이터 그리드 연구는 기존에 수행중인 고에너지물리 실험에도 활용하고 있다. 본 논문에서는 그리드 응용의 한 분야로서 고에너지물리 데이터 그리드에 관한 연구를 보여준다.

키워드 : 데이터그리드, 분산자료, 거대강입자충돌실험, 고에너지물리

Abstract The objective of the High Energy Physics(HEP) is to understand the basic properties of elementary particles and their interactions. The CMS(Compact Muon Solenoid) experiment at CERN which will produce a few PetaByte of data and the size of collaboration is around 2000 physicists. We cannot process the amount of data by current concept of computing. Therefore, an area of High Energy Physics uses a concept of Tier and Data Grid. We also apply Data Grid to current High Energy Physics experiments. In this paper, we report High Energy Physics Data Grid System as an application of Grid.

Key words : Data Grid, Distributed Data, Tier, LHC, High Energy Physics

1. 서론

고에너지 물리학(HEP, High Energy Physics)은 물

질의 근본 구성 입자와 상호 작용 연구를 통해 궁극적으로 우주 탄생의 비밀을 밝히는 학문이다.

2007년에 시작하는 유럽입자물리연구소(CERN)의 대형강입자충돌기(LHC, Large Hadron Collider)의 CMS(Compact Muon Solenoid)실험에 참여하는 연구진은 2000여명이며, 생산되는 데이터 양은 연간 수 Peta-Byte에 달할 예정이다. 그러므로 고에너지물리 실험에서 생산하는 데이터양은 기존의 전산자원의 개념으로 처리하는 것이 불가능하다. 그리하여 고에너지물리 분야에서 데이터 그리드라는 개념을 도입하게 되었다.

고에너지물리 데이터 그리드는 그 정의에 맞게 하드웨어 및 미들웨어 기반 시설을 활용하여 전 세계적으로 분산되어 있는 모든 과학기술 자원인 연구자, 가속기, 검출기, 데이터, 자료 분석용 저장 장치 및 연산 장치 등의 자원을 묶어 가상 기관(VO, Virtual Organization)을 구성하여 하나로 통합하고, 상시 활용하여 연구

This work was supported by grant No. R08-2003-000-10258-0 from the Basic Research Program of the Korea Science & Engineering Foundation / Korea Research Foundation

[†] 정 회 원 : 한국과학기술정보연구원 슈퍼컴퓨팅센터
cho@kisti.re.kr

^{**} 비 회 원 : 경북대학교 고에너지물리연구소 연구원
hanbi@knu.ac.kr
kihwan@knu.ac.kr
jincheol-kim@knu.ac.kr
djkong@fnal.gov

^{***} 비 회 원 : 경북대학교 물리학과
ycyang@fnal.gov
ydoh@fnal.gov

^{****} 비 회 원 : 경북대학교 물리학과 교수
suh@knu.ac.kr
dkim@knu.ac.kr
son@knu.ac.kr

논문접수 : 2005년 9월 3일
심사완료 : 2006년 4월 20일

개발 능력을 혁신할 수 있는 새로운 연구 환경의 패러다임을 구축하는 것이다.

2. 고에너지물리학이란?

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

물질의 근본 구성입자를 연구하기 위하여 관찰할 소스(source)와 검출기가 필요하다. 예를 들면, 테니스공(target)을 관찰하기 위해서는 눈(eye)으로 테니스공을 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, 분자를 관찰하기 위해서는 현미경이 필요하다. 핵 속에 있는 기본입자(target)들을 관찰하려면 가속기(source)가 필요하고 대규모의 검출기가 필요하다.

특히, 고에너지 가속기가 필요한 이유는 드브로이의 불확정성 원리에 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 고에너지 가속기를 이용하여 우주의 초기 단계를 만들어 우주 생성 및 진화에 관한 학문을 연구하는 기초 과학인 것이다.

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

3. 고에너지물리 데이터 그리드

가속기 및 고에너지 물리학과 관련된 연구는 컴퓨터 역사에서 정보 통신 기술과 많은 관계가 있음을 알 수 있다. 1980년대 이후 고에너지 물리분야의 실험 장치들은 갈수록 복잡해지고 대형화되었다. 요즘 주요 가속기 연구소에서의 실험은 대부분 국제공동으로 수행되는 대규모 기업형의 대형 연구이며, 실험 연구진은 300~2000명의 물리학자와 이에 상당하는 수의 공학자 및 기술자로 구성되어 있고, 5~15년의 기간이 소요되는 장기간의 연구이다. 고에너지 물리 실험에서 공동연구를 하게 되는 까닭은 물리학적 학문 관심이 주된 요소이며, 지리적으로 얼마나 멀리 있는가는 거의 고려하지 않고 있다. 현대의 편리한 교통수단과 의사소통 방법 등이 뒷받침

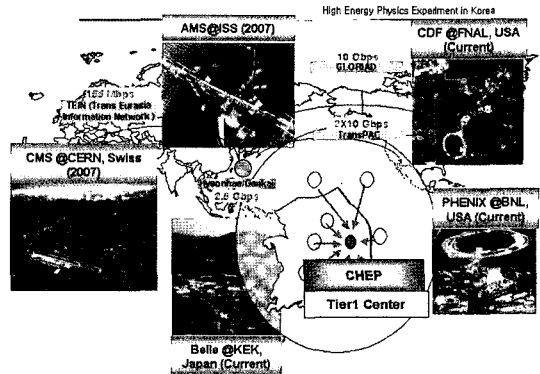


그림 1 한국의 고에너지물리데이터 그리드와 관련된 실험. 미국 페르미연구소의 CDF 실험, 미국 BNL의 PHENIX 실험, 일본 KEK의 Belle 실험 및 2007년 수행예정인 국제우주정거장(ISS, International Space Station)의 AMS(Alpha Magnetic Spectrometer) 실험 및 CERN에서 수행할 CMS 실험 등에 데이터 그리드를 활용하고 있다.

되기도 하지만, 무엇보다도 물리학적 관심과 공동으로 서로 분담하여 연구를 수행할 수 있는 능력을 기반으로 하기 때문이다. 이 연구 수행은 원격지에 있는 실험 장소에서 실험하고, 회의하고, 그 자료를 전송하는 방법에서도 시공간을 초월하여 거의 실시간에 가까운 의사소통 및 자료 교환이 필요하다. 이 필요성에 의하여 1990년 유럽입자물리공동연구소(CERN)의 컴퓨터과학자 버너스-리(Tim Bernes- Lee)가 World Wide Web을 발명하였고, 카이라우(Robert Cailliau)와 함께 URL, HTTP, HTML 등 첫번째 WWW 클라이언트와 서버를 개발하여 전 세계적으로 요즘 우리가 피부로 느낄 수 있듯이 가히 혁명적이라고 할 인터넷 세상을 개척하였다[1].

그리드는 점차 컴퓨터 자원 이용의 효율을 최대한 높이고자 가입자끼리 자원을 공유하면서 이용자의 불편이 없도록 하나의 시스템으로 인식도록 하는 개념으로 출발하였다고 볼 수 있다. 특히, 고에너지물리 데이터 그리드의 연구 수행은 첫째로 대형자료의 수집, 저장, 분석, 특히 원격지에서의 분석 등, 이와 관련하여 고속의 안정된 고도의 네트워크가 필요 불가결하며, 두 번째로 많은 자료의 양을 공유시켜서 사용자가 이용할 수 있도록 하여 분산된 대용량 저장장치를 개발할 수 있게 한다. 세 번째로 모의실험에서도 역시 초고속 연산 능력이 필요하므로 초고속 연산 능력을 제공해 준다. 국내의 고에너지물리 데이터 그리드는 경북대학교 고에너지물리 연구소를 중심으로 서울대, 연세대, 성균관대, 건국대, 이화여대, 전남대, 동신대, 경상대 등이 2002년 3월부터

참여하고 있다.

가장 중심이 되는 연구는 2007년에 수행될 CMS 실험의 데이터 그리드 구축에 관한 연구이며, 현재 수행 중인 미국 페르미연구소의 CDF(Collider Detector at Fermilab) 실험, 미국 브룩헤이븐 국립연구소의 PHENIX 실험, 일본 KEK 실험 및 2007년 수행 예정인 국제우주정거장(ISS, International Space Station)의 AMS (Alpha Magnetic Spectrometer) 실험 등에 데이터 그리드 연구를 수행하고 있다.

3.1 CMS 그리드

CMS 검출기 실험은 스위스 제네바에 있는 유럽입자물리연구소(CERN)의 대형 강입자 충돌기(LHC)를 이용한 4개의 큰 실험(CMS, ATLAS, LHCb, ALICE) 중 하나이다.

CERN에서는 원주가 27km인 7TeV 양성자와 7TeV 양성자를 정면 충돌시킬 가속기인 대형강입자충돌기(LHC)를 2007년 완공 목표로 건설 중이다. LHC 실험에서는 지금까지는 볼 수 없었던 높은 에너지 수준에서 양성자와 양성자의 정면 충돌을 가속화시킴으로써 대폭발 직후의 초기 우주 상태를 재현하여 이전보다 물질의 성분 구조를 더욱 자세히 이해할 수 있을 것으로 기대하고 있다. 특히 한국의 연구진이 참여하고 있는 CMS 검출기 실험의 목적은 질량 근원을 밝히는 데 중요한 역할을 할 힉스(higgs) 입자의 탐색과 새로운 이론인 초대칭 이론에서 예견하는 입자들을 발견하여 물질의 궁극적인 입자와 그 상호작용을 이해하는 것이다. 본 CMS 실험에는 한국을 포함한 36개국 159개 연구기관으로부터 2000여명의 과학기술자가 참여하고 있다[2].

LHC 실험에서는 연간 12~14 PetaByte의 데이터가 생성이 될 것으로 예상하고 있으며, 이것은 일반적으로 사용되고 있는 저장매체인 CD를 예로 들었을 때 1억 2천만 장에 해당하는 양이며, 이 데이터를 분석하기 위해서는 오늘날 가장 빠른 PC 프로세서 7만대가 필요할 것으로 추정된다. 그러므로 LHC실험에서는 기존의 전산개념으로는 자료를 처리하기가 불가능하다. 새로운 개념의 대용량 데이터 분산해석이 필요하며 이와 관련하여 그리드 개념이 절대적으로 필요하다. 이러한 개념으로 고에너지 물리분야에서는 계층적 구조의 Tier-0, 1, 2의 지역데이터센터의 개념을 도입하였다. 그림 2는 CMS 실험에서의 계층적 구조를 보여준다[3].

본 연구의 궁극적인 목표는 CMS실험을 위한 Tier-1 지역데이터센터를 구성하는 것이다. 이를 위한 CHEP (Center for High Energy Physics) 소프트웨어 프레임워크의 서비스 모델을 그림 3처럼 구성할 예정이다.

첫 단계로는 그리드 환경의 기본이 되는 보안 서비스와 웹 포탈, 그리고 소프트웨어 및 하드웨어 인프라를

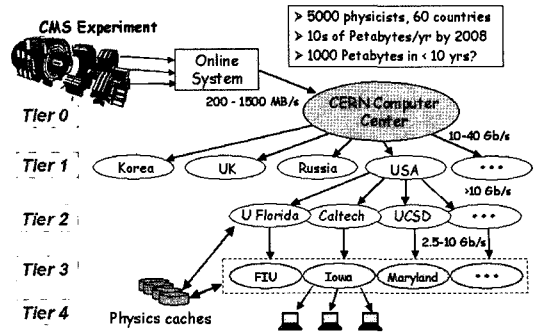


그림 2 CMS 실험에서의 계층적 구조 (Tier-0, 1, 2의 지역데이터센터의 개념)

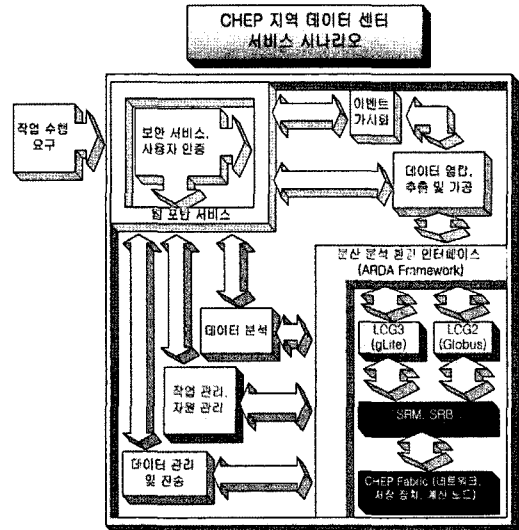


그림 3 고에너지물리 실험을 위한 CHEP 지역 데이터 센터 서비스 시나리오

구축한다. 하드웨어 인프라를 구축하기 위해 과학기술인프라인 네트워크인 과학기술협업연구망(GLORIAD: Global Ring Network for Advanced Applications Development)을 활용한 최적화 프로토콜 연구 및 SRB (Storage Resource Broker) 등의 저장장치 관리 소프트웨어를 연구한다. 한국 고에너지물리 그룹은 궁극적으로 모든 작업을 웹 포탈 상에서 수행할 수 있게 할 예정이다, 기본적인 사용자 서비스를 제공하기 위한 웹 포탈이 구축될 것이다. 고에너지물리 데이터 그리드의 고유 CA(Certification Authority)인 CHEP CA를 구성, 그 보안서비스를 연구하여 한국과학기술정보연구원(KISTI) 및 미국 에너지부(DOE) CA와 상호 인증을 하도록 할 것이다.

두 번째 단계에서는 데이터 분석을 위해 CERN에서

개발한 환경을 CHEP 서비스에 이식하여 통합 분석 분석 환경을 제공할 것이며, 이 통합 환경이 데이터 관리 및 전송 서비스, 작업 및 자원 관리 환경과 연동되도록 할 것이다.

마지막 단계는 기본적인 분산 데이터 분석 환경 구축을 완성하는 단계로, 데이터의 원격 가시화, 서비스, 원격 데이터 분석 및 열람, 가공 서비스 등이 제공될 것이다. 그리고 LCG(LHC Computing Grid)를 조율하고 개선하여 CHEP 데이터 센터의 서비스가 원활하게 통합 되도록 하는 보완 연구가 수행될 것이다. 이로써 CMS 실험의 원격 분석에 활용이 될 것이다.

이를 위한 연구 내용은 크게 네트워크, 저장장치 및 연산자원 등의 세 가지로 구성되어 있다. 구체적인 내용으로는 지역 데이터 센터를 구축하기 위한 기본 전제인 초고속 네트워크에 대한 연구, SRB(Storage Resource Broker)를 이용하여 이기종의 저장 장치에 존재하는 데이터의 관리 연구, 그리고 각각 유럽과 미국에서 추진 중인 그리드 프로젝트인 LCG 및 Grid3/OSG의 적극적 서비스 제공을 통한 활용 등을 포함한다.

3.1.1 네트워크 연구

첫 번째로 지역 데이터 센터를 건설하기 위해서는 네트워크 인프라가 잘 갖추어져야 하며, 또한 이런 네트워크를 잘 이용할 수 있어야 할 것이다.

정부에서는 GLORIAD망 구축 사업을 통해 지구 한 바퀴를 초고속 광통신망으로 연결시켜 주는 프로젝트를 진행 중에 있다. GLORIAD망은 국제 협력을 통한 과학 기술 연구 개발을 위해 한국, 미국, 캐나다, 네덜란드, 러시아, 중국 등 6개국이 콘소시엄으로 참가, 공동 투자하여 지구 전체를 10기가(Gbps)급 환경의 람다(Lambda, 광통신)망으로 연동하는 세계 최초의 글로벌 과학 기술 연구망이다. 2005년 8월 한국-미국 구간은 10Gbps의 망이 개통되었다. 유럽 CERN의 CMS 실험에서 생산되는 연간 PetaByte급의 데이터는 GLORIAD망을 통하여, 유럽, 미국, 한국 등의 Tier-1 기관에 데이터를 전송될 수 있을 것이다[4].

그러므로 이를 활용한 네트워크에 대한 연구가 활발히 이루어지고 있다. 본 연구진은 한국-미국, 한국-유럽 간의 실험을 통해 주어진 네트워크를 최대한 이용할 수 있는 기술을 습득하여 SC2005에서 미국의 Caltech 그룹과 함께 기존의 기록을 깨는 시연을 보였다. 본 시연 동안 미국 시애틀 SC2005 장소로 전송되는 물리 데이터 전송의 최고 대역폭은 131.57 Gbps로 2004년도 101.13 Gbps의 기록을 갱신하였다. Caltech, 유럽입자물리연구소(CERN), 미국 페르미연구소, 스탠포드 선형가속기 연구소(SLAC) 등이 주축이 되어 각 호스트 연구소와 브라질을 포함하는 협력기관들 간에 24시간 내에 475

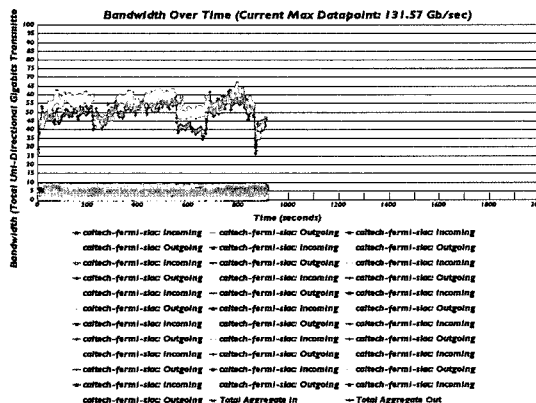


그림 4 미국 시애틀에서 열린 SC2005 대역폭 챌린지에서 미국, 유럽, 남아메리카의 국제 파트너들과 함께 131.57 Gbps라는 기록을 세워 세계기록을 갱신한 순간의 그래프

TByte라는 대량의 물리 데이터를 전송함으로써 2007년부터 시작되는 LHC 실험의 준비가 잘 수행되고 있음을 보여 주었다. LHC 실험에서는 연간 수십 PetaByte의 데이터가 생성되므로 이것을 각 협력 기관에 분산하여 처리할 수 있는 데이터 그리드를 구성하고 있으며, 이것은 필수적으로 초고속 네트워크를 필요로 한다. 본 시연에서는 LHC 실험의 계층적 구조(Tiered Architecture)에 근거하여 최신 기술의 광대역 네트워크 인프라와 그리드 웹 서비스를 이용하고 실시간으로 거대한 물리 데이터의 전송을 요구하는 이벤트 분석 응용을 보여주었다. 이를 위하여 17개의 10Gbps의 람다 웨이브들을 사용하였으며, 모나리사(MonaLisa)라는 툴을 사용하여 광대역 네트워크의 모니터링을 실시하였으며, 분석소프트웨어는 Caltech와 플로리다대에서 개발한 그리드용 분석도구를 사용하였다.

3.1.2 저장장치 연구

두 번째로 이기종의 저장장치에 존재하는 많은 양의 데이터를 관리하는 것이다. 데이터 그리드의 핵심 요소로서 SRB를 이용하여 KISTI 등과 저장 자원을 공유하여 자원 이용의 효율성을 증대시키고 대량의 데이터를 분산시킴으로써 데이터 관리에 대한 연구를 진행하였다. 이러한 SRB 연동 장치는 국내에서는 경북대와 KISTI 및 기초과학지원연구소(KBSI)와 연동하였을 뿐만 아니라 해외로는 일본의 KEK, 호주, 대만, 폴란드의 저장 시스템을 SRB로 연동하여 저장장치를 효율적으로 이용한다.

3.1.3 그리드 연산장치 연구

세 번째로, 연산자원을 그리드로 연동하는 것이다. 고에너지물리 분야에서 그리드 프로젝트는 유럽과 미국을

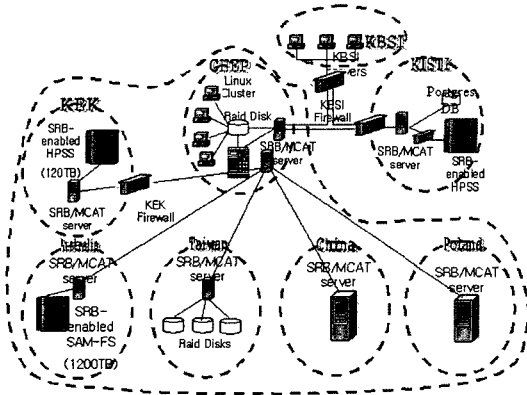


그림 5 SRB로 연동된 모습. 국내적으로는 CHEP, KBSI 및 KISTI가 연동되어 있으며 국제적으로는 일본의 KEK, 호주, 대만, 중국 및 폴란드가 연동되어 있다.

여 운영하고 있다.

3.2 기존 실험의 데이터 그리드

3.2.1 CDF 그리드

CDF 실험은 미국, 이탈리아, 일본, 영국, 독일, 한국, 대만 등 12개국의 800여 명의 물리학자들이 참여하는 대단위의 국제적인 가속기 실험이다. 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 달하며, 2006년도에는 60 MByte에 달할 예정이다. 입자 검출기가 연간 6개월간 작동할 경우 연간 900 TByte정도의 엄청난 데이터가 축적된다. 이러한 방대한 데이터를 처리하고 분석하기 위해서는 대용량의 저장장치 및 계산 장치와 네트워크 환경이 구축되어야 하고 데이터 그리드는 그 해답을 제시해 준다.

이에 본 연구진은 2002년도에 SAM(Sequential Access through Meta-Data) Grid 프로젝트에 참여하여 페르미연구소의 D0 실험과 함께 SC2002에서 시연하였다. 그리고 페르미연구소 내에 있는 파인만 컴퓨터 센터의 중심분산팜(CAF, Central Analysis Farm)으로 데이터를 처리하고 있으나, 용량의 한계가 있으므로 CDF Grid의 일환으로 한국 그룹과 미국의 샌디에이고(UCSD) 그룹은 각 지역별로 데이터를 처리하는 탈중심 분산팜(DCAF, DeCentralized Analysis Farm)을 구상하였다. 2003년 본 연구진은 CDF 그룹 내 처음으로 실제 연구자가 사용할 수 있는 탈중심분산팜을 연구 개발하여 설치하는 데 성공하였으며, 2004년에 이러한 구성의 개발 노하우를 타 연구기관에 전수하여 지금은 대만, 이탈리아, 캐나다의 토론토 대학, 미국의 럿거스대학 및 샌디에이고 대학에도 탈중심분산팜이 설치되어 있으며, 탈중심분산팜은 전체 CDF실험 CPU 처리량의 50%에 해당한다. 경북대학교에 설치된 탈중심분산팜에는 12개국 62개 기관 800여명의 물리학자들이 실제로 작업을 수행하고 결과를 얻고 있다. 이러한 기술을 SC2003 및 SC2004에 시연함으로써 그 기술을 국내외에 한층 더 알릴 수 있었으며, 향후 LCG 또는 OSG 팜과 통합하여 그리드화할 예정이다.

3.2.2 Belle 그리드

한국의 Belle 실험 그룹은 대만, 오스트레일리아 등과 함께 Belle 그리드를 연구 중에 있다. 한국 그룹은 경북대 및 성균관대에 구성된 글로벌스 환경위에 테스트베드를 활용하여 분산된 작업을 수행하였으며, 그동안 구축해 온 KBDG(Korea Belle Data Grid)를 실제 Belle 실험의 데이터 분석에 적용하여 분석하여 나온 결과를

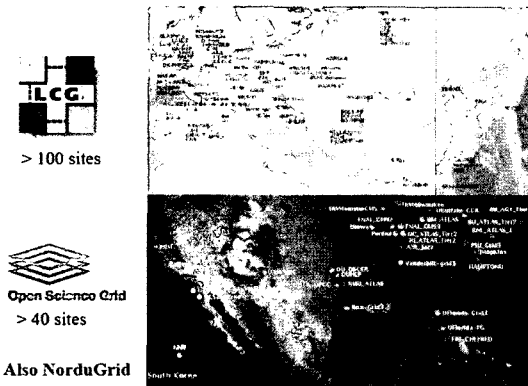


그림 6 고에너지물리분야의 전 세계적으로 중요한 그리드 팜. 한국의 경북대학교(KNU) 고에너지물리연구소는 LCG(LHC Computing Grid) 팜 및 OSG(Open Science Grid) 팜을 동시에 갖고 있다.

중심으로 양대 산맥을 이루고 있다. 유럽을 중심으로 LCG(LHC Computing Grid) 프로젝트가 진행되고 있으며, 미국을 중심으로 Grid3에 바탕을 둔 OSG(Open Science Grid) 프로젝트가 진행되고 있다. 한국은 두 개의 프로젝트에 동시에 참가하여 두 개의 팜을 동시에 설치하여 운영하고 있다.

Grid3 프로젝트에 2003년부터 미국이 아닌 유일한 사이트로 한국의 경북대학교 고에너지물리연구소가 참여하였으며 2004년에는 경북대학교 고에너지물리연구소의 84 CPU의 이중운영시스템(dual operating system)으로 구성된 Grid3 팜을 전 세계 30여개 연구기관의 다양한 분야의 과학자들에게 서비스를 제공하여 실제 활용하였고, 2005년부터는 OSG(Open Science Grid)로 확대하

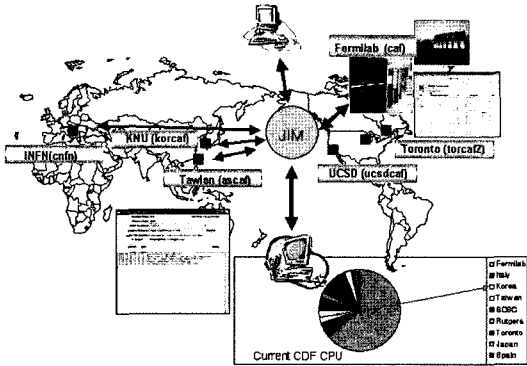


그림 7 CDF 실험의 전세계에 분산된 탈중심분산팜의 개념도. 오른쪽 하단은 각 연구기관의 CPU 기여도를 보여 준다.

비교하였다.

본 연구를 위하여 KISTI의 클러스터에 KBDG 프로그램을 설치하였다. KISTI의 클러스터를 사용한 이유는 우선 시스템 안정성 및 네트워크 등이 잘 구비되어 있기 때문이다. KISTI의 그리드 접속환경에 맞게 장비를 설정하고, KISTI에 접속 분석에 사용한 MDST 파일들을 전송하였으며, KISTI의 주 컴퓨터에 Belle 라이브러리 등 분석에 사용될 소프트웨어를 설치하였다.

KEK의 테스트베드와 KISTI의 그리드 테스트베드와의 물리 실험 결과를 비교하여 논문을 작성하였다 [5]. Belle 실험의 대표적인 붕괴모드인 $B^0 \rightarrow J/\psi K_s$ 을 기존의 Belle 머신과 그리드 환경으로 구축된 KISTI 머신을 사용하여 같은 결과를 얻었다. 이로써 지금까지 연구해 온 그리드 시스템의 데이터 안정성이 검증되었다. 기본적으로 한 개의 노드를 사용하였을 때 Belle 머신의 사양이 더 우수하나 Belle 머신의 경우 여러 연구원들과 같이 사용하였기 때문에 작업 시간이 더 오래 걸렸다. 그리드 환경에서의 작업의 안정성을 검증함으로써 그리드 환경 하에서 belle software의 활용방안을 제시하고, 기존 전산자원을 그리드화하여 자원의 효율적인 운영을 도모할 수 있을 것이다.

3.2.3 AMS 그리드

한국의 AMS 실험 그룹에서 2004년 1월부터 시작된 모의 시능 데이터를 bbftp를 이용하여 스위스 CERN으로 전송하였다.

한편, OpenSSI (Single System Image) 구조를 연구하여 몇 개의 클러스터를 하나의 클러스터로 인식되는 연구를 하였다. 그림 8은 AMS실험의 모의 시능 데이터 생산의 개요도를 보여 준다.

3.2.4 그리드 자원 연동 연구(Interoperability)

Grid3팜, 탈중심분산팜 및 CMS 클러스터와 상호작동

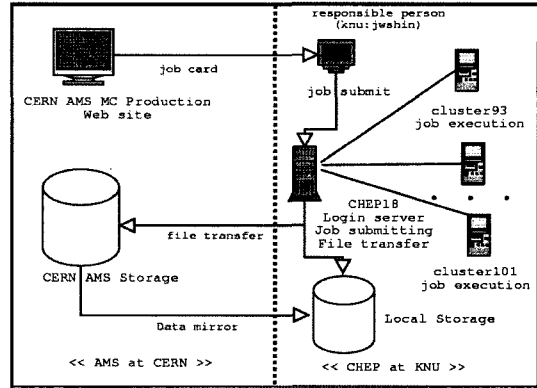


그림 8 AMS실험의 모의시능 데이터 생산의 개요

(interoperability)을 연구하여 condor 배치 시스템의 환경 변수를 통하여 자원의 이용도를 모니터링하고 그 상황에 따라 능동적으로 CMS와 탈중심분산팜의 워크노드를 Grid3 팜에 제공하는 이중적인 배치 시스템 구조로 자원 효율성을 극대화하는 연구도 수행하였다.

경북대 고에너지물리연구소는 다양한 물리실험에 참여하고 있으며, 클러스터 자원 또한 각 실험을 위해서 분배되어 있고 여러 개의 그리드 테스트베드를 가지고 있다. 예를 들면 Grid3는 주로 CMS 실험의 그리드 테스트베드이며, DCAF(Decentralized CDF Analysis Farm)는 주로 CDF실험에 사용된다. 하지만, 이런 것은 자원의 효율성 측면에서 만족스러운 환경이 아니며 진정한 의미의 그리드라 볼 수 없을 것이다. 그러므로 각 실험의 자원을 연동하여 효율을 극대화하는 것이 이 연구의 목적이다. 각 실험별 그리드 테스트베드들을 연동하여 자원의 사용을 극대화하는 방법에는 다음과 같은 것들이 있을 수 있다.

첫째, 여러 개의 배치 시스템(e.g., Condor, PBS, LSF)을 동일한 리소스에 설치하여, 서로 경쟁함으로써 자원 이용을 극대화하는 방법이다. 둘째, 하나의 배치 시스템을 사용하되 다중 인터페이스(Grid3, DCAF, LCG 등의 헤드노드)와 연동하여 작업요청을 모든 인터페이스에서 받아들이는 것이다.

경북대학교는 첫 번째 방법을 이용하여 자원의 효율성을 높이고 있다. 예를 들면, Grid3의 경우 84 CPU를 할당하고 있지만, 순수한 Grid3 노드는 오직 3대뿐이며, DCAF와 CMS 클러스터의 자원을 포함하고 있다. 이 경우 DCAF와 CMS 로컬 작업이 우선 순위를 가지며, 이 자원들이 유힘할 때 Grid3의 작업이 수행된다. 만약 Grid3 작업이 수행 중일 때 CMS 작업이 수행된다면, Grid3 작업은 Suspended 모드로 들어가게 된다. 앞으로의 계획은 경북대의 모든 자원을 LCG(LHC Computing Grid)의 자원으로 활용하는 것이다. 현재 LCG2가 설치

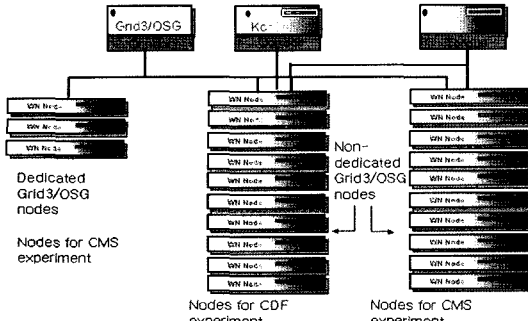


그림 9 자원 연동(Interoperability)의 개요도. CDF 실험의 DCAF 워크노드를 Grid3의 워크노드로도 활용하였다. 향후 DCAF 워크노드는 LCG의 워크노드로도 활용할 계획이다.

되어 있으며 기존의 Batch 시스템은 LCG의 워크노드로 활용 중이며, 향후 모든 자원들이 LCG2에 포함되어 작업 효율을 높일 것이다. 현재 Grid3/OSG와 LCG는 상호 작업을 수행하기 위해 규약을 설정하고 있다.

4. 연구 결과

4.1 고에너지물리 데이터 그리드 서비스 현황

전국대, 서울대, 성균관대, 전남대, 동신대, 경북대, 연세대, 포항공대, 고려대, 이화여대 등 10여개 기관에서 약 140여명의 사용자(중복 제외)들이 CMS, CDF, Belle, AMS, PHENIX 및 고에너지물리 이론 분야에 그리드 테스트베드를 사용하고 있다.

국외로는 총 12개국 70여개 기관의 900여명(중복 제외)의 물리학 및 타 분야 과학자들이 경북대학교 고에너지물리연구소에 설치된 그리드 팜을 활용 중이다. 구체적으로 살펴보면, 일본 KEK의 물리학자들이 Belle 팜을, 스위스 CERN, ETH, 독일 Aachen공대 등의 유럽 물리학자들이 AMS 팜을 이용하고 있다. 미국 등 30여개 기관 100여명의 물리학자, 지구과학자, 천문학자들이 Grid3/OSG 팜을 활용하고 있고, 12개국 62개 기관의 800여명의 물리학자들이 CDF 그리드 팜을 활용하고 있으며, 스위스 CERN 등의 2000여명의 물리학자들이 LCG 팜을 활용하고 있다.

4.2 자료 처리 결과

자료처리 결과의 한 예로서 본 논문에서는 CMS 실험을 위한 LCG팜으로 모의 시뮬한 결과를 보여주었다. LCG팜은 전 세계 CMS 실험 그룹 사용자들 뿐만 아니라 ATLAS, ALICE, LHCb 실험 등의 실제 데이터 분석과 모의 시뮬 데이터 생산에 사용될 예정이다. 그림 10은 LCG 팜을 이용하여 QCD jet 붕괴반응의 대표적인 모의시뮬을 한 결과로서 붕괴된 입자 운동 궤적을

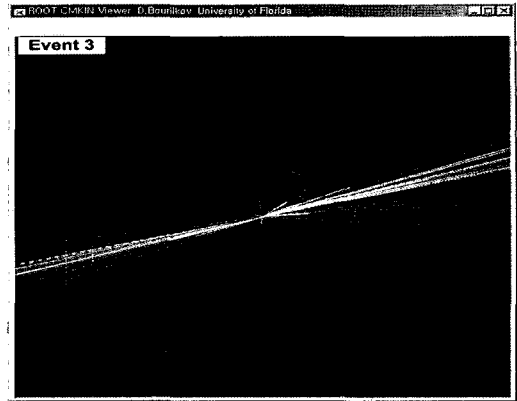


그림 10 LCG 팜을 이용한 모의시뮬 데이터의 생산 결과

CMKINView라는 소프트웨어로 보여준 것이다.

5. 결론

고에너지물리 분야는 전 세계적으로 그리드를 활용한 연구가 진행 중이며, 국내에도 그 연구가 활발히 진행되고 있다. 현재 총 12개국 70여개 연구기관의 900여명의 물리학자, 지구과학자, 천문학자, 생물학자, 전산학자 등 다양한 분야의 과학자들이 경북대 고에너지물리연구소의 그리드 팜을 활용하여 그리드로 연구를 진행하고 있다. 향후 고에너지물리 데이터 그리드를 통해 가장 직접적으로 활용될 분야는 유럽의 CERN에서 추진하고 있고, 한국에서 중점적으로 참가하는 CMS 실험이다.

고에너지물리 데이터 그리드 연구는 한국의 고에너지물리 연구 환경을 개선하고 국제적인 연구 성과를 얻을 수 있는 기회를 넓힐 수 있을 것이다. 우리 연구진은 데이터 그리드의 실제 구현과 활용으로 물리학 분석 등의 대부분의 실험과정을 그리드로 수행하고 있다. 특히, 외국 연구소를 직접 방문하지 않더라도, 국내에서도 외국과 똑같은 환경으로 연구 활동을 할 수 있다. 국내에는 대학원생을 포함하여 약 200여명이 관련 연구를 수행하게 되며, 아시아권 뿐만 아니라 전 세계적으로 2000여명의 물리학자들이 관련 연구를 수행하여 그리드를 활용한다. 그리드의 활용으로 국내의 고에너지물리 연구진이 학계에 기여하는 바가 크게 증가할 것이며, 이는 곧 국내 연구의 선진화와 세계 물리계에서 주요한 역할을 담당할 수 있을 것으로 기대한다.

한편, Tier-1급 지역 데이터 센터는 유럽 CERN과 연결된 국제 우주 정거장에서의 AMS 실험 과학 분석 센터의 지역 자료 센터 구축에도 활용될 수 있다.

한마디로 그리드는 디지털산업 분야에 일대 혁명을 가져올 뿐만 아니라, 우주 만물을 구성하는 기본 입자의 비밀을 밝혀내는 고에너지 물리학의 훌륭한 도구가 되

고 있다. 고에너지물리 실험은 순수 학문에 속하지만 실험 결과를 처리하고 분석하게 해 주는 데이터 그리드 기술은 산업적 측면에서도 파급효과가 엄청날 것이다[6].

결론적으로 고에너지물리분야는 그리드 응용 분야의 중요한 분야의 하나이며, 국가 그리드 기반의 하드웨어 및 미들웨어 구축을 활용함으로써 그 시너지 효과를 가지며, 정보통신기술을 이용하여 과학기술 연구 환경의 획기적인 개선을 한 중요한 사례가 될 것으로 기대할 수 있다. 이러한 그리드 구축사업은 반드시 필요하며 고에너지 물리학에서 대용량 데이터의 분산 해석 응용환경을 구축하기 위해서 반드시 데이터 그리드 연구는 필수적이다.

참 고 문 헌

- [1] <http://www.cern.ch/>
- [2] <http://cmsinfo.cern.ch/Welcome.html/>
- [3] Paul Avery, "U.S. Grid Projects: Grid3 and Open Science Grid," International ICFA workshop on HEP Networking & Digital Divide Issues for Global e-Science, May 23-28, 2005, Daegu, Korea.
- [4] <http://www.gloriad-kr.org/>
- [5] 김선민, 성균관대 석사학위논문, "Data Grid 환경에서의 Belle 실험의 $B^0 \rightarrow J/\psi K_s$ 붕괴연구 및 데이터 검증", (2004.8).
- [6] 헤럴드경제신문, 과학 한국의 미래 '영사이언티스트', "그리드 아시아 포탈로 육성", (2003.11.29).



조 기 현

1981년 3월~1985년 2월 연세대학교 물리학과 이학사. 1985년 3월~1987년 2월 연세대 대학원물리학과 이학석사. 1991년 8월~1996년 5월 미국 University of Colorado, Boulder 대학원 물리학과 Ph.D. 2006년 5월~현재 한국과학기술정보연구원 슈퍼컴퓨팅센터 선임연구원. 2001년 6월~2006년 4월 경북대학교 고에너지물리연구소 교수. 1996년 10월~2001년 5월 미국 University of Tennessee, Knoxville 박사후연구원. 1996년 5월~1996년 8월 미국 University of Colorado, Boulder 박사후연구원. 1992년 9월~1995년 9월 미국 Cornell University 방문연구원. 1992년 1월~1996년 5월 미국 University of Colorado, Boulder 연구조교

한 대 회

1991년 3월~1998년 2월 경북대학교 물리학과 졸업. 1998년 3월~2000년 2월 경북대학교 물리학과 석사 졸업. 2000년 3월~2001년 9월 라이선스피아 개발팀 사원. 2001년 10월~2002년 10월 (주)클리오드 개발팀 사원. 2002년 11월~현재 경북대학교 고에너지물리연구소 전산연구원

권 기 환

1989년 3월~1995년 8월 경북대학교 영어영문과 졸업. 1996년 9월~1998년 4월 캐나다 사스카치완 대학교 컴퓨터사이언스 졸업. 1998년 9월~2001년 11월 알버타 대학교 컴퓨터 사이언스 석사졸업. 2002년 11월~2005년 12월 경북대학교 고에너지물리연구소 전산연구원



양 유 칠

1996년 경북대학교 물리학과 입학. 2003년 경북대학교 대학원 물리학과 석사과정 입학. 2005년 경북대학교 대학원 물리학과 박사과정



오 영 도

1986년 3월~1990년 2월 경북대학교 물리학과(이학사). 1990년 3월~1993년 2월 경북대학교 대학원 물리학과(이학석사) 1994년 3월~1999년 8월 경북대학교 대학원 물리학과(이학박사). 2000년 1월~2001년 6월 성균관대학교 박사후 연구원 2001년 7월~2005년 7월 경북대학교 박사후 연구원. 2005년 9월~2005년 12월 경북대학교 외래교수. 2005년 11월~현재 포항공과대학 포항가속기연구소 책임연구원



공 대 정

1990년 경북대학교 물리학과 입학. 1998년 경북대학교 대학원 물리학과 석사과정 입학. 2000년 경북대학교 대학원 물리학과 박사과정 입학. 2006년 경북대학교 고에너지물리연구소 박사후 연구원 재직 중



서 준 석

1984년 3월~1990년 2월 서울대학교 천문학과 천문학(이학사). 1987년 5월~1988년 10월 육군 상병제대. 1991년 10월~1996년 12월 독일 Bonn University 물리학(Diplom Physiker). 1995년 12월~1996년 12월 독일 Max-Planck Institut fuer Radioastronomie 연구원. 1997년 1월~1999년 10월 독일 Bonn University 핵/입자물리학(Dr. rer. nat). 1997년 1월~1999년 10월 독일 Institut fuer Strahlen- und Kernphysik 연구원. 1999년 11월~2000년 7월 독일 Institut fuer Strahlen- und Kernphysik 박사후 연구원. 2000년 7월~2002년 7월 미국 Brookhaven National Laboratory 박사후연구원. 2002년 8월~2000년 12월 독일 Helmholtz-Institut fuer Strahlen- und

Kernphysik 박사후연구원. 2003년 9월~2006년 2월 경북대학교 대학원 물리학과 교수. 2006년 5월~현재 경북대학교 고에너지물리연구소 교수



김 동 회

1978년 3월~1983년 2월 서울대학교 물리교육과 이학사. 1984년 9월~1990년 4월 미국 시라큐스 대학원 물리학과 Ph.D. 1985년 9월~1990년 4월 시라큐스 대학원 연구조교. 1990년 4월~1992년 12월 Fermi National Accelerator Laboratory 박사후연구원. 1993년 3월~1997년 3월 경북대학교 자연과학대학 물리학과 조교수. 1997년 8월~2002년 3월 경북대학교 자연과학대학 물리학과 부교수. 2000년 8월~2001년 7월 Fermi National Accelerator Laboratory 객원교수. 2002년 4월~현재 경북대학교 자연과학대학 물리학과 교수.



손 동 철

1971년 3월~1975년 2월 서울대학교 물리학과(이학사). 1977년 8월~1983년 8월 미국 University of Maryland, College Park 물리학 Ph. D. 1998년 12월~현재 경북대학교 고에너지물리연구소 소장. 1996년 2월~1998년 8월 경북대학교 전자계산소 소장. 1994년 9월~1996년 2월 경북대학교 기획연구부 실장. 1985년 3월~현재 경북대학교 물리학과 교수. 1982년 8월~1985년 1월 미국 Columbia University Research Scientist