고에너지물리연구를 위한 그리드의 설계

조기현[0,1] 오영도[1] 이재영[1] 김복주[2] 이지수[3]
1.경북대학교 고에너지물리연구소, 2.서울대학교 물리학과, 3.한국과학기술정보연구원 슈퍼컴퓨팅센터
cho@knu.ac.kr[0], ydoh@fnal.gov, intvis@knu.ac.kr, bockjoo@fnal.gov. jysoo@hpcnet.ne.kr

A Grid Design for High-Energy Physics Research

Kihyeon Cho[0,1], Youngdo Oh[1], Jaeyoung Lee[1], Bockjoo Kim[2], Jysoo Lee[3]
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

고에너지물리연구에서 필요로 하는 컴퓨팅 파워와 대용량의 저장공간을 확보하는데는 많은 어려움이 있으며, 실제 실험에 참여, 연구, 분석하는 인력 또한 전세계에 산재되어 있다. 이 같은 어려움을 해결하기 위해 실험에서 필요로 하는 데이터 및 컴퓨팅 리소스를 각 지역별로 산재하여 실험의 효율성, 데이터 백업 및 컴퓨팅 리소스, 데이터 저장공간의 능률적 확보 및 활용을 위해 새로운 컴퓨팅 개념인 그리드를 활용하려는 시도를 하고 있다. 이 글에서는 현재 잘 알려진 그리드 미들웨어인 글로버스 2.0을 기반으로 하여 CERN에서 개발한 EU-DataGrid 소프트웨어를 이용한 그리드 시스템의 설계를 살펴본다.

1. 서 론

고에너지물리학은 거대가속기를 이용하여 전자 혹은 양성자등 입자를 높은 에너지로 충돌시켜 자연의 기본 입자의 성질연구, 입자상호작용의 연구 그리고 새로운 입자의 탐색을 수행한다. 충돌현상을 검출기로 검출하여 디지탈화한 신호를 온라인, 오프라인 필터링을 거쳐 디스크나 테이프에 저장한 후 물리연구에 이용한다. 현재 유럽입자물리연구소(CERN)에서 2007년에 수행할 예정인 고에너지물리실험의 경우 데이터의 양이 연간 수 PB에 달하여 데이터의 저장 및 처리에 현재와는 다른 새로운 방법의 도입이 절실하여졌다. 이에 CERN에서는 그리드의 활용에 주목하게 되었으며, 각 유럽국가의 참여 하에 유럽 데이터 그리드를 구축하고있다.

현재 국제적으로 데이터 그리드 계획에는 미국의 지구기후 시스템 그리드인 Earth System 그리드[1], 고에너지물리학과 핵물리 실험의 연구를 위한 PPDG[2], 가상 데이터(Virtual Data)개념을 연구하기 위한 GriPhyN[3], 그리고 유럽에서 PB급 고용량의 데이터 취급을 위한 European Data Grid(EUDG)[4]등이 형성되어 활동하고 있다.

EUDG는 고에너지물리학, 생물학, 지구관측을 응용분야로 하여 미들웨어의 개발, 기반시설과 테스트베드의 구축을 수행하고 있다. 이 계획은 역할별로 10개의 work package로 구성되어 있다.

(1) 미들웨어의 개발
- Work package 1 : work load management
- Work package 2 : data management
- Work package 3 : monitoring service
- Work package 4 : fabric management
- Work package 5 : mass storage management

(2) 기반시설, 테스트베드 구축
- Work package 6 : test bed integration of released software package
- Work package 7 : Grid network management , information on network performace and reliability

(3) 응용
- Work package 8 : High energy physics
- Work package 9 : Earth observation
- Work package 10 : Biology

위에서 서술한 바와 같이 고에너지물리학은 응용분야의 Work package 8에 속하여 CERN에서 2007년에 수행할 실험인 CMS, ATLAS, LHC-b등을 중심으로 하여 데이터 처리시스템의 그리드화를 수행하고 있으며, 이는 work package 6에서 통합되어 EUDG의 테스트베드로 활용될 수 있도록 하고 있다.

이와 같이 앞으로의 고에너지물리연구에서는, 데이터 및 컴퓨팅 리소스를 각 지역별로 산재하여 실험의 효율성, 데이터 백업 및 컴퓨팅 리소스, 데이터 저장공간의 능률적 확보 및 활용을 위해 새로운 컴퓨팅 개념인 그리

드가 필수적으로 이용될 전망이다. 이 글에서는 현재 잘 알려진 그리드 미들웨어인 글로버스 2.0을 기반으로 하여 CERN에서 개발한 EU-DataGrid 소프트웨어를 이용, 그리드 시스템의 설계를 살펴본다[4,5].

2. 본론

2.1 EU-DataGrid 테스트베드

Linux RedHat을 기반으로 개발되었고, LCFG (Local Configuration), RB (Resource Broker), CE (Computing Element), SE (Storage Element), WN (Worker Node), UI (User Interface) 등으로 구성되어 있으며, 이 각각이 그리드 컴퓨팅을 가능하게 해준다[4]. 그림 1 은 각각의 관계 및 구축한 그리드 시스템의 구조를 보여주며, 표 1 은 각 노드의 기능 및 역할을 설명해준다.

이와 같은 네트웍을 활용한 그리드에서는 보안문제가 크게 나타난다. 익명의 사용자들에게 함부로 데이터 및 컴퓨팅 리소스를 제공할 경우, 실제 실험에 필요로 하는 리소스를 사용하지 못하거나, 시스템의 붕괴 및 데이터 유출의 문제점이 있다. 이것은 그리드의 문제점이 된다. 이 같은 리소스 낭비 및 보안 유지를 위해 여기서는 PKI (PublicKey Infrastructure) 기반의 솔루션을 이용한다. CA (Certificate Authority) 가 존재하고, 이 CA의 인증을 받은 사용자와 노드만이 그리드의 리소스에 접근 이용할 수 있다.

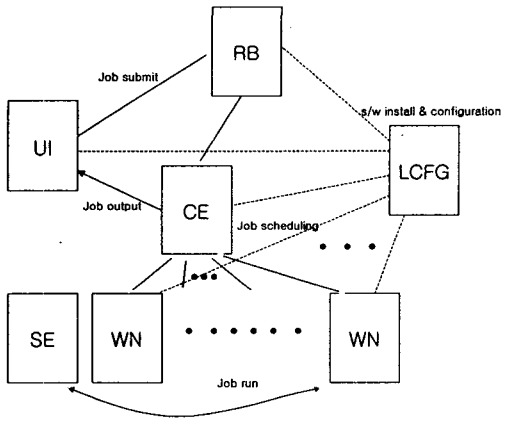

그림 1. EU-DataGrid S/W 구성도

2.1 고에너지물리연구에서의 그리드의 활용

현재, 고에너지물리연구소에서는 이와 같은 시스템을 구축, 실제 실험에 이용하기 위해 준비중에 있고, 이 단계가 완성되면 해외 물리 연구소들과의 리소스 공유를 위해 준비하고 있다. 그림 2 와 같이 각 학교 및 연구소에서 EU-DataGrid 기반의 그리드 시스템을 구축하고, 각 시스템에 대한 정보를 Information Server에서 관리하여 필요한 리소스를 찾아서 실행할 수 있다. RB를 통하여 원격지의 리소스를 활용하여 그 결과를 다시 받아오고, 필요한 데이터를 찾아서 이용한다. 이렇게 구성된

표 1. EU-DataGrid에서 각 노드의 역할

LCFG	한 사이트에서 관리해야 할 컴퓨터가 많을 경우 유용한 개념으로, 실제 그리드 컴퓨팅을 지원하기 위해 많은 수의 PC를 구성한다면 각 PC의 설정을 한 곳에서 통합관리하는 시스템이다. LCFG는 PXE (Pre-boot eXecution Environment), bootp, dhcp를 이용하여 네트워크 상에서 다른 PC에 OS 설치, 설정 변경등을 할 수 있으므로, 다수의 PC를 이용한 그리드 컴퓨팅 환경구축에 용이하다[6].
UI	사용자가 직접 그리드 환경에서 컴퓨팅 및 데이터 조작을 가능하게 하는 사용자 환경을 제공한다.
RB	UI를 통해 들어온 Job을 실행이 가능한 CE를 찾아서 그 Job을 CE로 넘겨준다.
CE	실제 들어온 Job을 Batch Job Manager를 통해 WN에서 컴퓨팅을 수행하고 그 결과를 UI로 되돌려 주는 역할을 한다.
WN	UI에서 보내온 Job을 실제 수행하는 역할을 한다.
SE	대용량 저장 장치를 관리 및 데이터 서비스를 해주는 것으로 UI에서 넘어온 Job을 수행할 때 필요로 하는 데이터 및 결과물을 제공 및 저장하는데 이용된다.

그리드 시스템은 국지적으로 제한된 리소스를 효율적으로 이용할 수 있으며, 데이터의 분산 관리도 용이하다. 하지만, 그리드 구조는 기본적으로 네트웍의 빠른 성능을 요구하므로 이와 같은 인프라를 구축하는 데도 노력

중이며, 데이터 관리 (EU-DataGird 에서는 아직 이 부분에 대한 기능이 미약함) 를 위한 효과적인 수단도 요구되어진다. 고에너지연구소를 포함한 국내의 각 연구기관은 그리드 환경에 필요한 현재의 컴퓨팅 환경을 조사하였고, 국내의 각 연구기관과 국제 공동 실험 연구소간의 네트워크 속도 성능을 검사하였으며, 각 연구 기관의 컴퓨팅 환경을 구축하였다. 이 리소스를 이용하여 실제 대용량의 데이터 및 컴퓨팅을 현 네트웍을 이용하여 얼마나 효율적으로 그리드 환경에서 처리할 수 있는지 실험할 계획이다.

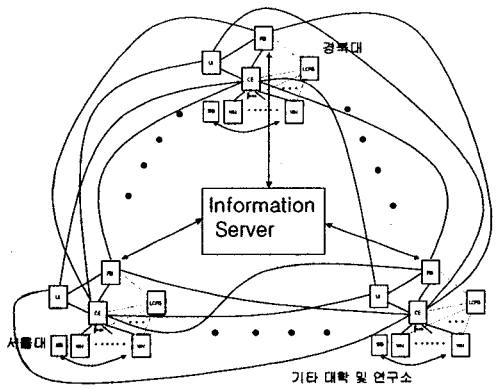

그림 2. 고에너지물리연구에서 활용을 위한 시스템 구조

또한 향후, 적용할 실험은 CDF, AMS, CMS, PHENIX, K2K 와 같은 국제 공동 실험들로서, 고에너지물리 분야에서의 그리드 테스드베드 구축에 따른 그리드 컴퓨팅의 활성화와 미국, 일본, 유럽 등과의 국제 공동 연구를 위한 기반 조성 및 국제 공동 연구 수행을 통한 국내 연구기반의 국제화 및 첨단기술을 습득을 위한 시스템으로 자리잡을 것이다.

3. 결론

서방 선진국 모두 고에너지 실험을 위한 가속기 설비를 보유하여 이를 활용한 실험을 활발히 진행중이며, 고에너지 물리 데이터 그리드를 활용하여 데이터 전송과 분석에 연구를 활발히 하고 있다. 한국에서도 고에너지 물리 데이터 그리드를 이용하여 고에너지 실험의 데이터의 전송 및 결과 분석연구에 직접 참여함으로서의 이 분야에의 실질적인 공헌도를 높여 국제적인 위상을 더 높인다.

또한, 데이터 그리드 환경은 High-End 기반의 정보기술 (IT)로 나노기술(Nano Technology, NT), 우주항공기술(Space Technology, ST), 환경 기술 (Environment Technology, ET), 문화기술 (Culture Technology, CT), 생명기술 (Bio Technology, BT) 등의 모든 차세대 핵심 기술 개발 및 응용에 필요한 요소이다. 고에너지물리 데이터 그리드 구축 연구를 통하여 얻어비는 기술과 경험은 이러한 분야의 발전에 충분히 기여할 수 있다.

4. 참고문헌

[1] http://www.earthsystemgrid.org/
[2] http://www.ppdg.net/
[3] http://www.griphyn.org/
[4] http://eu-datagrid.web.cern.ch/eu-datagrid/
[5] http://www.globus.org/
[6] http://www.lcfg.org/doc/