

데이터 그리드 웹 서비스 디자인

김복주¹ 오영도¹ 권기환^{0,1} 한대희¹ 조기현¹ 손동철¹ 이지수²

1. 경북대학교 고에너지물리연구소, 2. 한국과학기술정보연구원 슈퍼컴퓨팅센터
{bockjoo, ydoh}@chep12.knu.ac.kr, {kihwan⁰, hanbi, cho, son}@knu.ac.kr, jysoo@hpcnet.ne.kr

A Data Grid Web Service Design

Bockjoo Kim¹ Youngdo Oh¹ Kihwan Kwon^{0,1} Daehee Han¹ Kihyeon Cho¹ Dongchul Son¹ Jysoo Lee²

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

요 약

고에너지 물리연구에서는 입자가속기에서 배출되는 막대한 양의 데이터를 저장, 처리하기 위한 계산능력과 대용량의 저장장치를 확보하는데 어려움을 겪어왔다. 또한, 실제 실험에 참여, 연구, 분석하는 인력 또한 전세계에 흩어져 있어 실험의 효율성을 떨어뜨리고 있다. 이러한 어려움을 해결하고자 차세대 인터넷이라 불리는 그리드에 기반을 둔 데이터 그리드를 활용하고자 노력하고 있다. 현재 한국을 포함한 각국에서 데이터 그리드의 구축을 위하여 노력하고 있으며 테스트베드를 구축하여 실험하고 있다. 이러한 데이터 그리드의 활용도를 높이기 위해서 데이터 그리드를 웹에서 접근할 수 있다면 그 유용성은 배가될 것임이 틀림없다. 이러한 측면에서 현재 경북대 고에너지물리연구소에서 개발하고 있는 데이터 그리드 웹서비스 디자인을 살펴본다.

1. 서 론

고에너지물리 분야는 가속기의 입자충돌을 통하여 검출된 데이터를 분석함으로써 우주의 생성 및 진화를 연구하는 기초과학 분야이다. 고에너지물리연구에서는 입자가속기에서 배출되는 막대한 양의 데이터를 저장, 처리하기 위한 계산능력과 대용량의 저장장치를 확보하는데 어려움을 겪어왔다. 또한, 실제 실험에 참여, 연구, 분석하는 인력 또한 전세계에 흩어져 있어 실험의 효율성을 떨어뜨리고 있다. 이러한 어려움을 해결하고자 차세대 인터넷이라 불리는 그리드를 활용하여 고에너지연구에서 필요로 하는 계산자원과 저장장치, 데이터에 대한 접근성 등을 확보한다는 방안이다.

현재 경북대학교 고에너지 물리연구소에서는 유럽 핵물리연구소인 CERN에서 개발한 EU-DataGrid 소프트웨어를 이용하여 고에너지물리 데이터 그리드 테스트베드를 구축하여 실험하고 있다. 가장 선진적인 그리드의 한 형태인 EU-DataGrid (EDG)는 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. 그림 1은 각각의 관계 및 구축한 그리드 시스템의 구조를 보여주며, 각 노드의 기능 및 역할은 [1,2]를 참고하기 바란다.

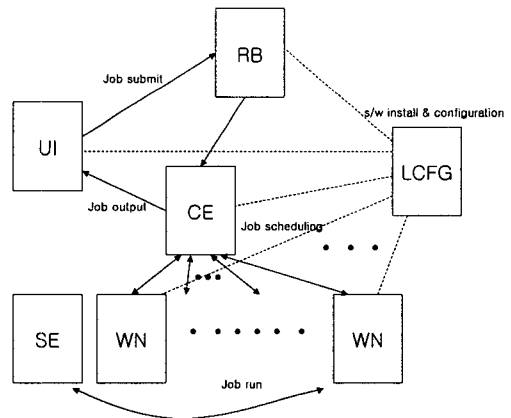


그림 1. EU-DataGrid S/W 구성도

이 글에서는 현재 경북대학교에 구축되어 있는 EU-DataGrid 테스트베드의 활용도를 높이기 위한 데이터 그리드 웹 서비스 디자인에 대해서 살펴보고자 한다.

2. 본 론

2.1 데이터 그리드 웹 서비스 요구 사항

웹을 이용하여 데이터 그리드에 접근하기 위해서는 다음과 같은 사항이 만족되어야 할 것이다.

가. 글로벌스를 바탕으로 구현된 EU-DataGrid에서는 프락시(Proxy)를 생성하여 단 한번의 인증 (single sign-on)으로 사용자에게 허락된 모든 데이터 그리드 자원에 접근할 수 있도록 구현되어 있다. 웹서비스에서는 웹을 통하여 데이터 그리드에 접근하려는 사용자가 데이터 그리드에서 유효한 proxy를 가지고 있는지 확인할 수 있어야 할 것이며, 또한 그렇지 못할 경우에 웹을 통하여 프락시를 생성할 수 있다면 이상적일 것이다.

나. EU-DataGrid에서는 job description file (.jdl)을 사용하여 수행되어질 job을 기술한다. 이 job description file을 자신의 컴퓨터로부터 웹으로 업로드 할 수 있어야 하고, 또한 웹 상에서 jdl 파일을 바로 생성할 수 있다면 더욱 편리할 것이다.

다. job을 던진 후에 자신의 job이 지금 어떤 상태에 있는지 모니터링 할 필요가 있을 것이다. EU-DataGrid에서는 던져진 job의 상태를 살펴볼 수 있는 dgjobstatus라는 명령을 제공한다. 예를 들면, job의 status는 Waiting for Queue, Scheduled, Done, Output Ready 등의 상태를 포함한다. 웹에서 이 명령을 던지고 그 결과를 보여줄 수 있는 메뉴가 존재해야 할 것이다. 또한 던져진 job의 id 를 일일이 손으로 쳐 넣을 필요 없이 자동적으로 추적할 수 있어서 여러개의 job이 던져진 경우, 원하는 job을 선택할 수 있어야 할 것이다.

라. 결과의 도출이다. 간단한 결과를 생성하는 job일 경우, job의 output을 바로 웹상에서 보여 주어야 할 것이다. 하지만, 간단한 결과가 아닌 경우나 웹에서 보다는 output의 위치를 알려주는 경우가 더욱 편한 경우라면, 그 job의 위치를 알려 줄 수 있어야 할 것이다.

2.2 데이터 그리드 웹 서비스 프로토타입

위의 사항을 염두에 두었을 때, 그림 2와 같은 디자인 프로토타입을 생각할 수 있다 (그림에서 WB는 웹 브라우저, WS는 웹 서버를 나타낸다). 클라이언트가 웹서비스에 접속하면 서버는 초기 로그인 페이지를 보여 준다. 사용자가 자신의 프락시에 대한 정보를 입력하면 웹서비스는 UI에 존재하는 사용자의 프락시가 유효한지를 묻게 된다. 이것이 유효하면, 웹서비스는 이제 사용자가 데이터 그리드를 사용할 수 있도록 유럽데이터그리드의 명령들을 웹상에 띄우게 되는 것이다. 웹서비스를 단순하게 하는 것은 데이터 그리드의 UI 머신과만 연동하면 된다는 사실이다.

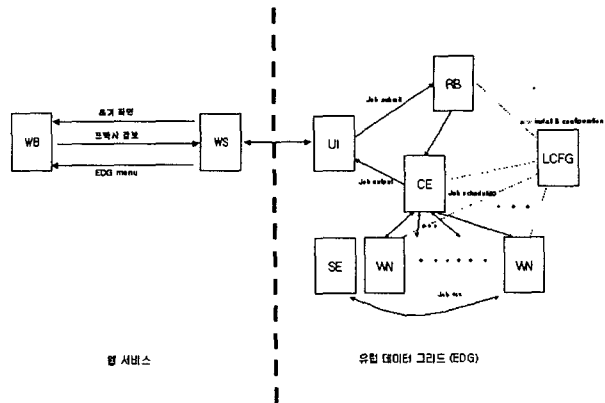


그림 2. 웹서비스 프로토타입

2.3 웹서비스 프로토타입의 구현 (Implementation of Web Service Prototype)

(가) 구현 (implementation details)

그림 2에서 개념적으로 UI와 웹서비스가 분리되어 있지만, 실제 구현상에 있어서는 둘 다 하나의 머신에 존재하는 것이 편리하다. 현재 UI에 아파치 (Apache) 웹서버 [3]를 설치하였다. 그리고 파이썬(Python) 스크립트로 짜여진 메일맨(mailman) 프로그램 [4]을 응용하여 웹서비스를 구현하였다. 또한, 사용자가 자신의 프락시로 로그인 되어있는 동안 세션 쿠키를 사용하여 job status를 저장하기 위하여 MySQL [5] 데이터베이스를 사용하였다.

(나) 웹서비스 메뉴 화면

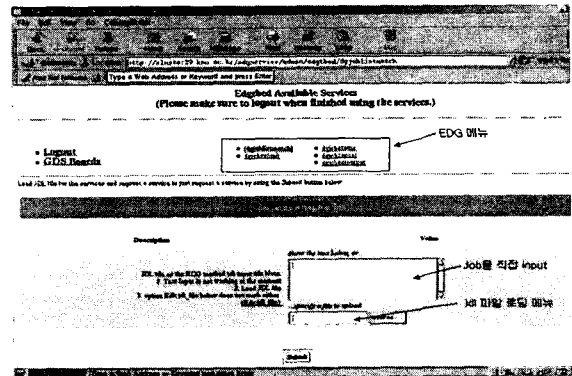


그림 3. EDG Web service 메뉴 화면

그림 3은 사용자가 유효한 프락시(proxy)를 갖고있다는 것이 인증되면, 이제 사용자가 사용할 수 있는 여러 가지 데이터 그리드 메뉴를 보여준다.

(다) dgjoblistmatch

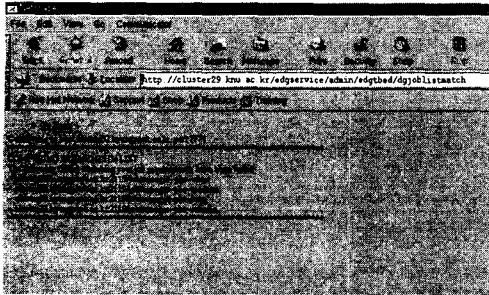


그림 4. dgjoblistmatch 메뉴 실행 결과

dgjoblistmatch메뉴는 jdl 파일에 묘사된 job을 실행할 수 있는 컴퓨팅 자원을 찾아 주는 역할을 한다. 그림 4에서 여러 가지 pbs queue들이 나타남을 볼 수 있다.

(라) dgjobsubmit

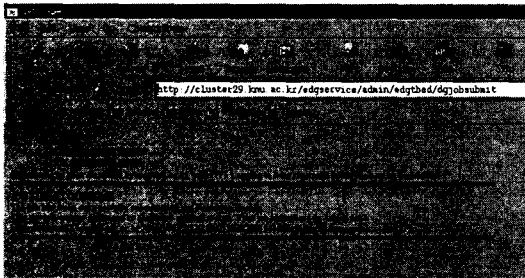


그림 5. dgjobsubmit 메뉴의 실행 결과

dgjobsubmit 메뉴는 실제로 job을 데이터 그리드에 던지는 것이다. 이것의 결과로서 던져진 job의 jobId가 리턴된다.

(마) dgjobstatus

던져진 job의 jobId를 이용해서 그 작업의 현재 상태를 볼 수 있다. 그림 6은 dgjobstatus 메뉴를 이용해서 던

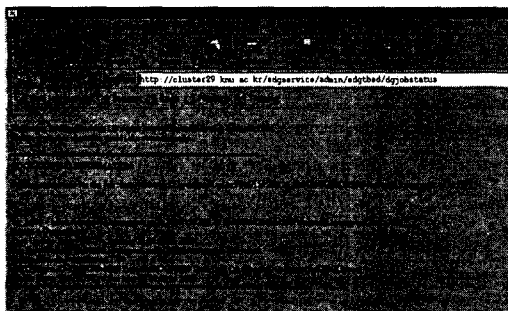


그림 6. dgjobstatus 메뉴 실행 결과

져진 작업의 상태를 체크했을 때 벌써 작업이 끝나고 결과가 준비되었다는 OutPutReady의 상태를 나타내고 있다.

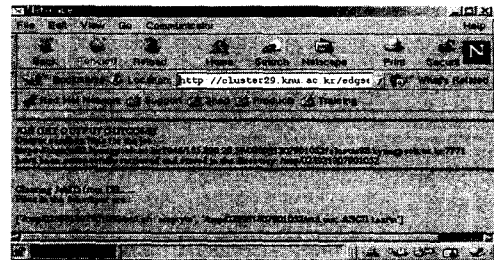


그림 7. dgjobgetoutput 메뉴의 실행 결과

(바) dgjobgetoutput

이 메뉴는 결과가 어디에 저장되어 있는지 위치를 알 수 있게 해 준다.

3. 결론

데이터 그리드의 웹서비스는 데이터그리드를 인터넷에서 이용할 수 있게 함으로서 그 활용도를 높이는 데 목적이 있다. 사용자는 자신의 운영체제에 상관없이 언제 어디서나 job을 체크하고 submit 할 수 있을 것이다. 경북대학교 고에너지물리연구소에서는 독자적으로 웹서비스 디자인을 수행하여 초기단계의 서비스를 가동중이며 앞으로 좀 더 사용하기 편리한 서비스를 제공하는 것을 목표로 하고 있다. 아직 요구 조건을 다 만족시킨 것은 아니며, 현재 웹서비스는 프락시 서버를 사용하지 않기 때문에 테스트베드에 접근하기 위해서는 사용자의 로컬 머신으로 프락시를 다운로드해야하는 한계가 있다. 이런 문제점은 프락시 서버를 도입하거나 그런 종류의 서비스를 이용함으로써 해결할 수 있다. 결론적으로, EU-DataGrid의 UI와 연동되어 웹상에서 job을 load할 수 있고, 던질 수 있다는 사실은 더 나은 웹서비스를 위한 기초가 마련되었다는 것을 의미한다.

4. 참고문헌

- [1] 조기현 외. 고에너지물리연구를 위한 그리드의 설계, 2002년 10월, 정보과학회
- [2] <http://eu-datagrid.web.cern.ch/eu-datagrid/>
- [3] <http://www.apache.org>
- [4] <http://www.gnu.org/software/mailman/mailman.html>
- [5] <http://www.mysql.com>