

# MGrid Portal : 분자 시뮬레이션 그리드 포탈

이종현<sup>o</sup> 김동욱<sup>\*</sup> 이건영<sup>\*</sup> 정갑주<sup>\*</sup> 황선태<sup>\*\*</sup> 박형우<sup>\*\*\*</sup>

<sup>\*</sup>건국대학교, <sup>\*\*</sup>국민대학교, <sup>\*\*\*</sup>한국과학기술원

{lejohn<sup>o</sup>, jeongk<sup>\*</sup>}@imc.konkuk.ac.kr, {sthwang<sup>\*\*</sup>}@cs.kookmin.ac.kr, {hwpark<sup>\*\*\*</sup>}@kisti.re.kr

## MGrid Portal : A Molecular Simulation Grid Portal

Lee Jong Hyun<sup>o</sup> Kim, D.W. Lee, K.Y. Jeong, K.J. Hwang, S.T. Park, H.W.

Department of Computer Science and Engineering, Konkuk University.

### 요 약

그리드 컴퓨팅은 다양한 이기종 분산 컴퓨팅 환경에서 다양한 소프트웨어들에 대한 설치, 운영, 업그레이드, 유지보수 등이 요구된다. 이러한 환경에서 응용 소프트웨어(특히, 레가시 소프트웨어) 지원은 매우 어려운 작업이다. 웹 포탈 기술은 이러한 문제에 대한 효과적 대안이다. 그러나 웹 포탈의 경우 다양한 사용자 인터페이스 제공이 난해하여 긴밀한 사용자와 컴퓨터간 상호작용이 요구되는 응용과학 분야 연구에 적용이 어렵다. MGrid 포탈은 BT/NT분야에서 광범위하게 사용되는 연구기법인 분자 시뮬레이션을 지원하도록 설계된 그리드 포탈 시스템으로 분자 시뮬레이션 작업들에 대한 상세한 모니터링 및 통제 기능을 지원하도록 설계되어 있다.

### 1. 서 론

그리드란 지리적으로 분산되어 있는 각기 다른 조직속에 있는 이질적인 자원들(컴퓨팅 자원, 계측장비, 저장장치등)을 네트워크로 묶어 기존의 하나의 컴퓨팅 자원 또는 소규모의 네트워크로 수행할 수 없는 과학/사회 분야에 활용할 수 있게 만든 분산 환경 구조를 말한다. 그리드 플랫폼은 다양한 소프트웨어, 하드웨어로 구성된다. 이러한 다양한 환경에서 작동될수 있는 소프트웨어를 개발하기란 쉽지 않을 뿐만 아니라 이런 분산환경에서 해당 소프트웨어를 설치하는 것 또한 쉽지 않다. 포탈은 이러한 소프트웨어 개발상, 설치·유지보수상 편리성을 가지고 있다. 또한 임의 장소에서 웹만 가능하다면 사용이 가능하다는 장점을 가지고 있다. 이러한 장점으로 인해 그리드 포탈 시스템이 생겨났다. MGrid 포탈은 BT와 NT분야에서 광범위하게 쓰이는 연구기법인 분자 시뮬레이션을 지원하도록 설계된 그리드 포탈 시스템이다. 이것은 단순한 로컬 리소스 매니저의 작업큐에 쌓인 작업들의 기초 관리정보 모니터링을 하는 타 그리드 포탈이 가지지 못한 어플리케이션 레벨의 모니터링(시뮬레이션 작업 상태확인(시뮬레이션 에너지 그래프 확인, 시뮬레이션 데이터 리스팅))이 가능하다.

이 논문에서는 KISTI K\*Grid 테스트베드[1] 상에서 현재 해당 작업노드의 상태 및 테스트베드 상황 등을 웹상에서 보여주는 것 뿐만 아니라 PSE[2] 를 통해 시뮬레이션 서버에게 작업 제출(Job Submission)한 분자 시뮬레이션 작업의 모니터링과 실험결과를 보여주는 MGrid 포탈을 기술한다. 2절에서는 MGrid Portal의 이해에 필요한 분자 시뮬레이션 그리드에 대해 설명하고 3절에서는 포탈의 전체 구조와 아키텍처에 대해서, 4절에서는 포탈의 구현과정과 상황에 대해서, 5절에서는 그리드 포탈에 관련된 연구에 대해서, 6절에서는 현재까지 소개한 포탈의 결과를 정리하고 향후 과제에 대해 기술한다.

### 2. 분자 시뮬레이션 그리드(MGrid)

분자 시뮬레이션 그리드(이하 MGrid)는 BT/NT분야 응용 과학자들이 그리드에 대한 지식이 없어도 그리드를 이용할 수 있는 환경을 구축하는데 그 목적이 있다.

#### 2-1. 광학이성질체 분석

광학이성질체[3]는 신약개발에 중요한 물질로 약 400억 달러 시장이 존재한다고 예측하고 있다. (Technology and Business Review: 2nd Edition, 200) 이러한 광학이성질체는 현재 1000여개의 물질이 발견되었고 선택자는 50여개가 발견된 상태이다. 이 두개의 물질을 스크립트 기반의 Charmm이라는 레가시 소프트웨어를 통해 시뮬레이션한다. Charmm[4]은 입력스크립트를 받아 대용량의 시뮬레이션 데이터를 뽑아낸다. 이 출력된 데이터를 통해 쓰이는 광학이성질체를 판단하게 된다. 그러나 기존연구는 개별물질에 대한 분석에 그치고 있을 뿐만 아니라 단일 워크스테이션상에서 시뮬레이션을 가정한다면 약 2000년 이상이 소요된다. 이처럼 오래 걸리는 응용 연구인 광학이성질체 분석을 MGrid를 통해 대규모로 시뮬레이션한다면 그 기간을 획기적으로 단축할 수 있다.

#### 2-2. MGrid

MGrid는 BT/NT분야의 응용 과학자들이 효율적이고 안정적이며 손쉽게 그리드를 통해 연구를 할수 있는 환경을 제공하는 것을 목적으로 한다. MGrid는 각각 작업을 준비하고 해당자원에 제출하는 사용자 클라이언트격인 PSE(Problem Solving Environment)와 제출된 작업의 제어와 PSE와의 통신을 담당하는 SM(Simulation Manager)과 실제 컴퓨팅 자원에서 Charmm을 수행하는 TEF(Task Execution Framework), 사용자가 제출한 작업의 상태와 정보, 자원의 모니터링을 수행하는 MGrid 포탈로

구성된다.

3. 설 계

3-1. MGrid 포탈 아키텍처

그리드 포탈은 다양한 목적을 가지고 구축될 수 있다. MGrid 포탈은 범용적 목적으로 만들어진 것이 아닌 분자 시뮬레이션을 위한 어플리케이션 포탈이다. 또한 KISTI K\* Grid 테스트베드상에서 얼마나 효율적으로 분자 시뮬레이션 실험환경을 지원하는지에 목적이 있다.

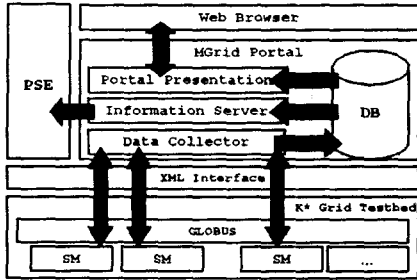


그림1 MGrid 포탈 아키텍처

K\* Grid 테스트베드 상에서 Globus[5]로 묶여있는 각 대학, 기관들의 리눅스 클러스터 또는 슈퍼컴퓨터에는 시뮬레이션을 제어하는 SM(Simulation Manager)이 각각 설치되어 있다.

데이터컬렉터는 DataCollector-SM간 XML통신 프로토콜을 통해 PSE로부터 SM으로 작업 제출된 런타임 정보와 자원 정보, 소프트웨어 정보와 같은 정적인 정보를 30분마다 주기적으로 모으는 작업을 수행하여 데이터베이스에 저장하는 역할을 한다. 이 데이터 컬렉터와 포탈의 glue 코드가 바로 실제 응용 연구자가 원하는 정보의 모니터링을 가능하게 해준다. 인포메이션 서버는 실험자의 작업 제출 대상이 되는 컴퓨팅 자원의 정보를 데이터베이스상에서 가져와 XML문서 형식으로 PSE에게 전달한다. 이 XML문서에는 각 SM이 설치되어 있는 컴퓨팅 자원의 호스트이름, 해당 자원을 소유하고 있는 기관명, 자원의 Cpu, 메모리, 저장용량, 노드갯수등의 자원정보와 사용하고 있는 OS의 종류 및 버전, Job Manager의 종류, 소유하고 있는 소프트웨어의 종류 및 버전, 에이전트 경로등의 정보를 담고 있다. 이러한 정보를 바탕으로 PSE는 작업 제출하려는 컴퓨팅 자원을 선택한다. 포탈 프리젠테이션은 데이터 컬렉터가 각 SM으로부터 수집한 데이터를 저장하고 있는 데이터베이스에서 SQL언어를 통해 데이터를 가져온다. 그런 후에 사용자의 요구에 맞는 형태로 가공하여 웹브라우저를 통해 보여준다.

3-2. 통신 프로토콜

통신 프로토콜은 XML 스키마로 정의되었다. XML은 이질적인 플랫폼으로 구성되어 있는 그리드상에서 서로 다른 응용 어플리케이션을 통합하는데 용이하고 향후 프로토콜의 확장성이 있다는 장점을 가지고 있다. MGrid 통신 프로토콜은 각각 요청과 응답메시지로 구분된다. 자원정보, 현재 수행되고 있는 작업의 정보, 로그정보를 요청하게 되면

SM은 그에 따른 응답메시지를 위의 스키마에 정의된 XML 문서 형식으로 만들어 데이터 컬렉터로 보내게 된다. 데이터 컬렉터는 XML문서를 파싱하여 MGrid 포탈서버 데이터베이스 스키마에 맞춰 각각 저장한다.

3-3. 작업절차

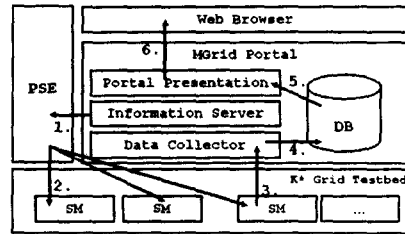


그림3 MGrid Portal 작업절차

1. PSE에서 작업 제출을 위한 컴퓨팅 자원 선택 2. 선택된 컴퓨팅 자원에 작업을 제출 3. 데이터 컬렉터가 주기적으로 SM으로부터 작업에 대한 정보를 수집 4. 데이터베이스의 스키마에 맞게 저장 5. 사용자의 요구에 맞추어 데이터를 가공 6. 해당 웹브라우저 형식에 따라 출력

3-6. 데이터베이스 스키마

MGrid Portal의 데이터베이스는 사용자 정보를 담고 있는 User 테이블(사용자, 패스워드, 이름, 이메일등), 사용자의 작업에 관련한 정보인 JobObj 테이블(Job의 Id, 이름, 작성자, 시작시간, 작업디렉토리등)과 TaskObj 테이블 (Task Id, 시작시간, 수행명령, 상태정보등), 자원을 제공하는 기관정보를 담고있는 Insitute 테이블(조직명, 주소, 전화번호, 이메일등), 자원 정보를 담고있는 Resource 테이블(Cpu타입, 메모리용량, 노드정보, OS정보, 에이전트경로등), 소프트웨어 정보를 담고있는 Software 테이블(소프트웨어 이름, 버전등), 해당 자원의 상태정보를 담고있는 HostStatus 테이블(호스트이름, 기관명, 반응시간, 호스트상태)로 이루어져 있다.

4. 구 현

4-1. 개발 환경 구축

펜티엄4 2.4GHz, 512RAM의 리눅스가 설치되어 있는 컴퓨터를 기반으로 그리드에 접근하기 위한 미들웨어인 Globus Toolkit 2.2가 설치되어 있다. K\* Grid에 접근하기 위한 인증서와 인증키는 KISTI CA에서 발급을 받아 사용했다. 또한 그리드 포탈과 인포메이션 서버를 운영하기 위한 아파치 웹서버 1.3.23과 포탈 프리젠테이션을 위한 PHP 4.1.2와 HTML, 그래프 생성을 위한 JFreeChart 0.9.6, 그리드 자원을 접근하여 정보를 수집하는 데이터 컬렉터에게 필요한 자바 Cog[6] 라이브러리를 래핑한 래퍼코드(국민대 분산컴퓨팅 연구실, 2003)와 Java SE 1.4.2가 필요하였다.

4-2. 기능구현

MGrid 포탈은 30분마다 주기적으로 SM이 설치되어 있는 컴퓨팅 자원에 GRAM을 통해 반응시간을 체크하여 5분만

에 접근한다면 Alive상태를, 5분이 지나면 Fail상태를 데이터베이스에 저장하고 포탈 프리젠테이션부에서는 컴퓨팅 자원의 상태를 가공하여 웹브라우저 화면에 보여줌으로써 접근하려는 자원 사용의 가이드라인을 제시한다.

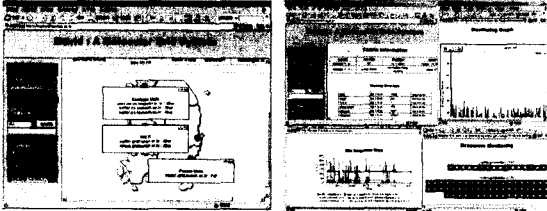


그림4 MGrid Portal

또한 PSE로 작업 제출한 작업의 런타임정보를 사용자에게 보여줄 뿐만 아니라 작업이 수행되고 있는 디렉토리의 리스팅 및 Charmm을 이용한 분자 시뮬레이션 결과데이터를 사용하여 .png 이미지 파일을 GridFTP로 전송해오고 웹상에서 확인할 수 있다. 이러한 파일의 확인은 실제 실험자가 현재 작업의 진행정도를 파악할 수 있을 뿐만 아니라 작업의 데이터가 올바른 것인지 판단하는 실험자의 지표가 되기 때문에 아주 중요하다. 마지막으로 각각의 노드에서 현재 수행하고 있는 작업의 상태를 확인할 수 있다. 노드는 하나의 이미지로 표시되며 현재 동작중인 노드를 웹상에서 확인할 수 있고 현재 수행되는 job과 task의 개수를 알 수 있다.

5. 관련 연구

NPACI HotPage[7] 은 NSF에 의해서 펀딩된 Advanced computing을 위한 국가 파트너쉽인 NPACI[8] 에서 만든 그리드 포탈이다. 이 포탈은 41개의 대학과 연구센터로 이루어진 NPACI의 컴퓨팅 자원을 사용한다. 작업 제출 및 스크립트생성, 작업큐 모니터링, 파일관리, 노드 모니터링등을 지원한다. LAPK Portal[9], GAMESS portal[10] 등도 역시 같은 기능을 지원하는 어플리케이션 포탈들이다. Alliance User Portal[11] 은 NCSA에서 만든 그리드 포탈로 구성원에게 정보서비스와 대화식 서비스를 제공한다. 큐정보, 작업정보, 로드정보 등의 모니터링과 파일전송기능을 제공한다. GENIUS INFN 그리드 포탈[12] 은 INFN프로젝트의 일환으로 생겨났으며 EngineFrame3.2[13] 를 사용하여 개발되었다. 이 포탈은 파일서비스와 작업모니터링, 데이터 서비스, 보안서비스 등을 제공하고 있다. 하지만 이러한 포탈들은 현재 응용 어플리케이션의 통합의 어려움으로 인해 어플리케이션 레벨의 모니터링 정보를 보여 주지 못하고 있다.

6. 결론 및 향후 계획

웹 포탈의 단점인 제한된 사용자 인터페이스, 다양한 사용자 상호작용 등의 제약에도 불구하고 소프트웨어 개발편리성, 설치, 유지보수의 편리, 쉬운 접근성면에서 그리드 포탈은 중요하다. MGrid 포탈은 KISTI K\* Grid 테스트베드를 통한 안정적 효율적 분자 시뮬레이션 환경을 사용자 친화적인 웹 포탈 모델로 구현했다. 이 MGrid 포탈은 작업의

모니터링과 결과전송으로 그리드 기반 분자 시뮬레이션을 가능케 한다. 또한 그리드를 통한 분자 시뮬레이션을 시스템의 관점(작업이 PBS[14] 를 통해 누구의 계정으로 수행되고 있는가)에서 본 것이 아닌 실제 응용 사용자의 관점에서 어플리케이션 레벨의 모니터링(시뮬레이션 작업 상태 확인 (에너지 그래프))이 포탈을 통해 이루어진다. 또한 작업결과를 언제 어디서든 웹을 통해 글로벌하게 접근하여 볼 수 있을 뿐만 아니라 다운로드할 수 있다. 이는 타 그리드 포탈에서 볼수 없는 큰 장점이다.

GridSphere[15] 는 ASC[16]와 GPKD[17] 프로젝트의 경험을 토대로 포탈 개발자들이 쉽고 빠르게 포탈을 개발하기 위한 Portlet[18] 기반의 통합 프레임워크를 제공한다. GridSphere는 역할기반 접근제어와 사용자 인증, 써드 파티 포틀렛과 호환, 코드와 프리젠테이션 분리등의 장점이 있다. 분자 시뮬레이션만에 특화된 기능을 지원하는 MGrid Portal은 로직과 프리젠테이션이 섞여있어 새로운 기능을 추가할 때 어렵고 가독성면에서도 좋지 않다. 또한 실제 응용 사용자에 따른 커스터마이징이 적어 사용자의 요구에 부응하지 못한 면이 있다. 또한 K\* Grid 테스트베드로 접근하기 위한 Glue 코드는 새로운 서비스 추가시 서비스 로직을 만드는 것 외에 그 서비스에 대한 Glue 코드를 새로 개발해야 하는 부담을 가지고 있다. 때문에 향후 GridSphere의 포틀렛 서비스 아키텍처를 이용하여 코드(서비스)와 프리젠테이션(포틀렛)을 분리, 각 응용 사용자에 따른 포틀렛만을 바꿈으로써 기존의 서비스를 통해 분자 시뮬레이션 뿐만 아니라 다른 응용분야를 다룰 수 있는 통합 포탈의 개발을 수행할 것이다. 또한 서비스와 포탈간 범용적 인터페이스의 개발을 병행하여 분자 시뮬레이션에 국한된 것이 아닌 좀더 다양한 어플리케이션을 지원할수 있는 그리드 응용 어플리케이션 포탈을 개발할 것이다.

참고사이트

[1] KISTI K\*Grid Testbed, <http://testbed.gridcenter.or.kr>  
 [2] PSE, <http://pse.cs.kookmin.ac.kr/>  
 [3] 광학이성질체, <http://www.chiral.com/>  
 [4] Charmm, <http://www.charmm.org/>  
 [5] Globus project, <http://www.globus.org/>  
 [6] Java Cog, <http://www.globus.org/cog>  
 [7] HotPage, <http://hotpage.npaci.edu/>  
 [8] NPACI, <http://www.npaci.edu/>  
 [9] LAPK Portal, <https://gridport.npaci.edu/LAPK/>  
 [10] GAMESS Portal, <https://gridport.npaci.edu/GAMESS/>  
 [11] Alliance User Portal, <http://aup.ncsa.uiuc.edu/>  
 [12] GENIUS INFN Portal, <https://genius.ct.infn.it/>  
 [13] EngineFrame3.2, <http://www.engineframe.com/>  
 [14] PBS, <http://www.openpbs.org/>  
 [15] GridSphere, <http://www.gridsphere.org/>  
 [16] ASC, <http://www.ascportal.org/>  
 [17] GPKD, <http://doesciencegrid.org/projects/GPKD/>  
 [18] Portlet, <http://jcp.org/en/jsr/detail?id=168>