

화학계산과 나노물질 계산을 위한 그리드 포털 사이트의 설계 및 구현

문종배*, 김명호*

*숭실대학교 컴퓨터학과

e-mail:comdoct@ss.ssu.ac.kr kmh@comp.ssu.ac.kr

Design and Implementation of Grid Portal Site for Computational Chemistry and Nano-Materials

Jong-Bae Moon*, Myung-Ho Kim*

*School of Computing, SoongSil University

요 약

복잡한 계산 화학 문제의 해결과 신물질 개발을 위한 나노물질 계산 등을 위해서는 고성능 컴퓨팅 파워가 필요하다. 그러나 이러한 계산을 위해 필요한 실험 장비의 가격은 천문학적으로 증가하고 있어 문제 해결의 장애물로 인식되고 있다. 그런데 최근 '그리드(Grid) 컴퓨팅'이라는 새로운 컴퓨팅이 제안되면서 지리적으로 떨어진 여러 자원들을 하나의 슈퍼컴퓨터처럼 사용할 수 있게 되었다. 그러나 그리드 환경을 이용하려면 그리드 미들웨어인 글로버스(globus)를 설치하고 인증기관을 통해 사용자 인증과 호스트인증을 거친 후 응용 프로그램들을 설치하여 수행하여야 한다. 이러한 복잡한 과정을 거쳐야 그리드 환경의 응용프로그램을 실행 할 수 있고, 실행할 수 있더라도 그리드 환경에서 작업을 요청하고 할당하려면 RSL이라는 스크립트 언어를 사용해야하는 번거로운 작업이 필요하다. 따라서 본 논문에서는 복잡한 화학계산과 나노물질 계산을 위한 응용프로그램들을 인터넷 상에서 쉽게 수행할 수 있도록 그리드 환경을 구축하고 웹 인터페이스를 구현한다.

1. 서론

인터넷이 발달하고 컴퓨팅 파워와 네트워크의 성능이 향상됨에 따라 컴퓨터를 이용하는 분야가 다양해지고 있다. 대규모 수치해석이나 유전자분석, 화학계산 등은 슈퍼컴퓨팅 기반의 고성능 컴퓨팅 파워에 대한 요구가 점점 더 증가해가고 있다. 나노기술 역시 실험으로 해석이 어려운 문제를 해결하기 위해 고성능 컴퓨팅 파워를 요구하는 분야이다. 그렇지만 증가하는 장비의 가격은 이러한 요구에 대한 장애물이 되고 있다.

그래서 여러 지역에 흩어져 있는 각종 자원들을 공유하고 협력해서 하나의 슈퍼컴퓨터처럼 쓰자는 개념의 '그리드(Grid) 컴퓨팅'이 도입되었다[3]. 그리드 컴퓨팅이란 지리적으로 분산된 이기종의 고성능 컴퓨터, 대용량 저장장치, 첨단 장비 등의 자원을 고속 네트워크로 연결해 상호 공유하고 이용할 수 있

도록 하는 서비스이다. 이를 가능하게 하기 위해서는 고성능 네트워크와 일관된 인터페이스를 제공하는 미들웨어가 필수적이다. 현재 그리드 서비스를 제공하는 미들웨어들 중 가장 많이 사용하는 것은 Globus Toolkit이다. Globus는 그리드에서 필요한 다양한 모든 서비스들을 반드시 사용해야 하는 것이 아니라 필요한 서비스만 선택하여 사용할 수 있도록 독립적인 요소로써 제안하고 있기 때문이다. 또 기존에 구성되어 있는 각 시스템 및 네트워크의 구성을 바꾸지 않고 각 요소들과 협력하여 그리드 환경을 구축할 수 있다.

그러나 Globus를 이용하여 그리드 환경을 쉽게 구축할 수는 있지만 그리드 환경에 익숙하지 않은 응용프로그램 사용자들이 사용하기에는 아직 여러 복잡한 환경을 가지고 있다. 따라서 본 논문에서는 그리드 환경에서 화학계산과 나노물질 계산을 위한

응용프로그램들을 인터넷을 통하여 일반 사용자들이 손쉽게 사용할 수 있도록 그리드 환경을 구축하고 여러 기능을 통합한 웹 인터페이스를 구현한다.

2. Globus Toolkit

Globus는 분산된 이기종 컴퓨팅 자원들을 하나의 가상 슈퍼컴퓨터처럼 사용하기 위한 대표적인 미들웨어이다. Globus는 크게 보안, 정보 서비스, 자원관리 서비스, 데이터 관리 서비스 등으로 나누어진다 [1, 2, 3].

2.1 보안(Security) 서비스

Globus에서 보안 서비스는 Globus Security Infrastructure(GSI)라고 부르며, 분산 자원들을 공유함에 따라 생기는 보안 문제를 해결하기 위해 필요한 서비스이다. GSI는 사용의 편리성과 더 안전한 보안을 위해 single sign-on을 제공하고 Globus proxy를 이용한다. 사용자는 그리드 환경에 한 번의 인증 과정만을 거침으로써 사용이 허가된 모든 자원들을 사용할 수 있고 분산된 각 자원에 대한 사용자 인증은 프락시(proxy)가 대신 수행한다. 그러나 각 자원 내에서 자원에 대한 허용범위는 각 자원이 제시하는 보안체계를 따른다[5].

2.2 정보(Information) 서비스

Globus에서 정보 서비스를 수행하는 요소를 Metacomputing Directory Service(MDS)라고 부른다. MDS는 그리드 환경 내에 존재하는 자원들의 상태정보를 공유하고 사용자들에게 제공하기 위한 서비스이다. 정보를 저장하고 사용자들에게 제공하기 위해 MDS는 LDAP를 사용한다. MDS는 두 가지 서비스를 제공한다. 첫째는 각 자원의 정보를 수집하는 GRIS, 둘째는 수집된 정보를 통합하는 GIS이다. 이들 정보는 각 자원의 정보를 필요로 하는 어플리케이션 개발자나, resource broker 등에 제공된다.

2.3 자원관리(Resource Management) 서비스

Globus에서는 자원관리를 위해 Globus Resource Allocation Manager(GRAM)이라는 서비스를 제공한다. GRAM은 Globus의 가장 중심이 되는 요소로써 원격지 자원을 사용할 수 있게 하고 분산 자원들을 동시에 사용하게 하면 자원들의 관리의 상이함을 처리한다. 이를 위해 Globus는 resource broker,

co-allocator, local resource manager 등으로 구성되어 있다. 자원 요청을 명시하고 각 요소간 의사소통을 하기 위해 Resource Specified Language(RSL)를 사용한다. 작업이 분산환경에 할당되면 각 작업의 협업을 위해 GRAM 서비스는 DUROC이라는 요소를 사용한다[7].

2.4 데이터 관리

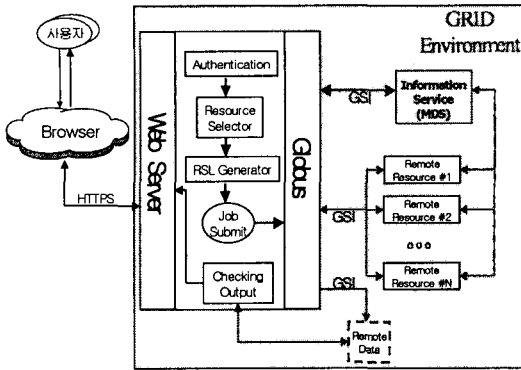
Globus는 데이터 관리를 위해 GASS, GridFTP, Replica catalog 서비스를 제공한다. GASS는 GRAM서비스와 밀접한 관련이 있는 요소로서 원격지에 있는 파일을 사용하여 작업을 처리하기 원하거나 원격지에서 처리한 작업의 결과를 또 다른 저장 장치에 저장하고 싶을 때 사용한다. GridFTP는 그리드 내의 데이터가 대규모 대용량이란 점을 고려하여 고속으로 파일을 전송과 파일의 이어받기를 가능하게 하는 요소이다. Replica catalog는 데이터 그리드를 위해 개발된 것으로 데이터들을 분산 저장 및 관리함으로써 필요할 때에 신속하게 데이터를 사용할 수 있게 하는 기술이다.

이러한 Globus는 그리드 시스템을 구축하기 위한 다양한 서비스들을 독립적인 요소로써 제시하고 있어서 개발자는 필요한 서비스를 선택하여 사용할 수 있고, 기존에 각 시스템 및 네트워크의 관리정책이나 운영 도구들을 수정하지 않고 이것들과 협력하여 그리드 환경을 만들어 주는 장점이 있다.

그러나 사용자가 그리드 환경에서 응용프로그램을 실행하기 위해서는 우선 Globus를 설치하기 위한 디스크 공간과 사용자 계정 등이 필요하고, 또 Globus를 설치하고 사용하기 위하여 필요한 사용자 인증등의 과정을 거쳐야 한다. 그리고 그리드 환경에서 응용프로그램을 자원에 할당하고 수행하려면 RSL을 직접 하여야 하는 번거로움이 있다.

3. 화학계산과 나노물질 계산을 위한 그리드 포털 사이트의 설계와 구현

본 논문에서는 구현한 그리드 포털 사이트는 Globus가 설치되어 있지 않은 곳에서도 복잡한 화학 계산이나 나노물질 계산을 위한 응용프로그램들을 Globus에 대해 자세히 모르는 사용자일지라도 쉽게 사용할 수 있도록 웹 인터페이스를 구현한다. 본 논문에서 구현한 포털 사이트의 전체적인 구조는 아래 그림 1와 같다.



(그림 1) 그리드 웹 포털의 전체 구성도

3.1 사용자 로그인 모듈의 설계

사용자는 그리드 응용프로그램을 사용하려면 반드시 그리드 환경에 인증을 받아야 한다. 웹에서 인증을 거친 사용자는 Globus의 GSI 서비스를 통한 프락시를 이용하여 한 번의 인증만을 거쳐서 그리드 내의 필요한 모든 자원을 사용할 수 있도록 해준다. 웹의 로그인 아이디와 그리드 사용자 아이디를 별도 관리하여 시스템에 별도의 계정을 추가하지 않도록 하였고, HTTPS, SSL, Cookies 등을 사용하여 인터넷을 통한 사용자와 데이터들의 보안을 유지할 수 있도록 한다.

3.2 자원선택 모듈의 설계

사용자가 그리드 환경에서 제공되는 여러 응용 프로그램들 중 필요한 것을 선택하고, 이들 응용 프로그램을 실행하기 위한 여러 가지 인자들과 옵션을 선택할 수 있도록 해주는 기능이다.

Globus의 GRAM 서비스에서 MDS를 이용하여 유효한 자원을 자동으로 찾는 기능이 아직 지원되지 않기 때문에 현재 그리드 환경의 유효한 자원을 찾기 위해서는 이 모듈에서 Globus의 MDS를 이용하여 가용한 시스템 자원들을 알아내고 사용자에게 보여주는 기능도 할 수 있도록 한다.

3.3 부하정보 수집 모듈의 설계

자원선택 모듈에서 얻은 여러 지역의 가용한 자원들의 CPU 부하정보를 계속 유지하는 기능을 한다. 이 내용을 바탕으로 하여 자원들의 부하량에 따라 작업의 양을 달리하여 할당할 수 있도록 한다.

3.4 RSL 생성 모듈의 설계

Globus에서 작업을 할당하고 실행하기 위해서는 자원을 RSL 형태로 표현하여 요청하여야 한다. Globus의 내부 명령어를 이용하여 RSL을 생성할 수는 있지만 사용자가 직접 만들기는 번거롭고 원하는 형태의 RSL을 얻기는 힘들다. 그래서 RSL 생성 모듈에서는 자원 선택 모듈에서 선택된 응용프로그램, 자원과 옵션들을 입력받아 분산되어 있는 여러 자원에 작업을 할당할 수 있는 RSL로 바꿔주는 기능을 한다.

부하정보 수집 모듈에서 얻은 각 가용 자원들의 부하정보를 바탕으로 부하가 적은 자원에 많은 작업을 할당하도록 한다

3.5 작업 요청 모듈의 설계

RSL 생성 모듈에서 만들어진 RSL을 Globus의 GRAM 서비스를 이용하여 필요한 자원들에게 할당하는 내부 모듈이다. Globus의 작업 요청 명령어인 'globusrun'을 이용하여 생성된 RSL을 GRAM 서비스에 작업을 요청한다. 사용자가 작업을 요청하고 작업이 수행되는 동안 다른 작업을 할 수 있도록 구현한다.

3.6 전체 구현

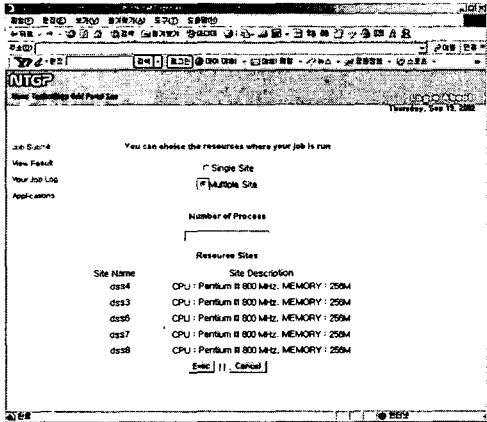
위에서 설명한 모듈들을 구현하기 위해서 Globus의 GRAM, MDS, GSI, GASS등의 서비스를 이용하였고, 기존의 그리드 포털 개발 킷(kit)들을 사용하지 않고 PHP와 C 프로그래밍으로 구현하였다.

로그인 모듈에서 로그인 아이디는 시스템 내의 데이터베이스에 저장하여 관리하고 이것을 그리드 환경에서 사용하는 아이디와 연결하여주도록 하였다. 자원선택 모듈에서는 MDS를 이용하여 가용한 자원들의 리스트를 가져오도록 하였다. RSL 생성 모듈에서는 사용자가 선택한 옵션들을 바탕으로 한 지역 또는 여러 지역에 있는 자원들에게 자원을 할당할 수 있는 RSL을 생성하도록 하였다. 작업요청을 한 후에는 다른 작업을 할 수 있도록 하였고, 완료된 작업의 내용이나 에러의 내용은 특정 위치의 파일에 저장하여 작업이 끝나면 이 내용을 화면에 보여주도록 하였다.

4. 실험

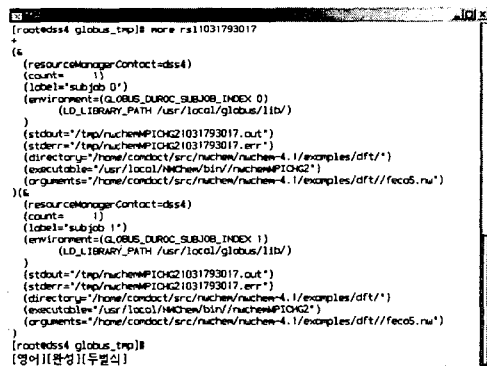
본 논문에서 구성한 사이트에서는 nwchem[8]이라는 화학계산을 하는 응용프로그램을 그리드 환경에

서 수행할 수 있도록 구현하였다. 아래 그림 2는 자원선택 모듈에서 몇 개의 프로세스를 생성하여 응용 프로그램을 수행할 것인지 입력받는 화면이다. 프로세스의 개수를 입력하면 각 자원의 CPU 부하가 적은 자원부터 응용프로그램을 수행하도록 하였다.



(그림 2) 자원선택 모듈 화면

아래 그림 3은 RSL 생성 모듈에서 자동으로 생성한 RSL의 예이다.



(그림 3) RSL 생성 모듈에 의해 생성된 RSL

5. 결론 및 향후 과제

네트워크와 컴퓨팅의 성능 향상으로 다양한 분야에서 슈퍼컴퓨터와 같은 고성능 컴퓨팅 능력을 요구하고 있다. 화학 계산과 나노물질 계산도 이러한 분야이다. 그래서 최근 “그리드 컴퓨팅”이라는 새로운 컴퓨팅 개념이 도입되어 여러 지역에 분산되어 있는 컴퓨팅 자원들을 하나의 슈퍼컴퓨터처럼 사용할 수 있게 되었다. 그러나 그리드 환경에서 일반 사용자들이 응용프로그램을 사용하기에는 복잡한 과정을

거쳐야 하는 불편한 점이 있다. 그래서 본 논문에서는 웹을 이용하여 그리드 환경에 접속하여 고성능 컴퓨팅 파워를 요구하는 화학 계산이나 나노물질 계산을 위한 응용프로그램들을 손쉽게 수행할 수 있도록 포털 개념의 웹 사이트를 구축하였다.

앞으로 더 다양한 응용프로그램들을 위한 RSL 생성 모듈로 수정할 필요가 있다. 그리고 사용자가 수행시키고자 하는 응용프로그램이 어디에서 가장 효율적으로 수행될지 판단할 수 있는 스케줄링 방법이 연구되어야 할 것이다.

참고문헌

- [1] I. Foster, C. Kesselman, "Globus: A Metacomputing Infrastructure Toolkit" Intl. J. Supercomputer Applications, pp, 11(2):115-128, 1997.
- [2] I. Foster, C. Kesselman, "The Globus Project: A Status Report" Intl. J. Supercomputer Applications, pp, 11(2): 115-128, 1997.
- [3] I. Foster and C. Kesselman (eds.) "The Grid: Blueprint for a new Computing Infrastructure" Morgan Kaufmann Publishers, 1998.
- [4] I. Foster, C. Kesselman, G. Tsudik, S. Tuecke, "A Security Architecture for Computational Grids." Proc. 5th ACM Conference on Computer and Communications Security Conference, pp, 83-92, 1998.
- [5] K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman, "Grid Information Services for Distributed Resource Sharing." Proceedings of the Tenth IEEE International Symposium on High-Performance Distributed Computing (HPDC- 10), IEEE Press, August 2001.
- [6] W. Allcock, A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, S. Tuecke, "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets." Journal of Network and Computer Applications, pp, 23:187-200, 2001.
- [7] <http://www.globus.org>
- [8] <http://www.emsl.pnl.gov:2080/docs/nwchem/nwchem.html>