

# 차세대 인터넷 서비스: Grid Computing

김 일 민\*

## 1. 서 론

최근의 컴퓨터 계산 능력의 발전은 매우 눈부시다. 2001년도의 개인용 컴퓨터의 계산 능력은 1990년도의 슈퍼컴퓨터의 계산 능력과 비슷하다. 2003년도의 개인용 컴퓨터의 하드 디스크 용량으로써 100 GByte가 흔한 것이 되었지만, 1990년대 초의 슈퍼 컴퓨팅 센터의 저장 용량과 비슷한 것이다. 이러한 기술적인 혁신에도 불구하고, 최신 응용 프로그램은 보다 많은 계산 능력과 저장 용량을 요구하고 있다. 예를 들어 1990년대 초의 생물학자는 분자 한 개의 구조를 계산할 수 있는 컴퓨팅 능력을 요구하였지만, 2000년대의 생물학자는 고분자 복합체의 구조를 계산하는 연구를 수행하고 있다[1].

첨단 연구를 위한 컴퓨팅 파워의 수요를 충족시키는 것은 매우 경비가 많이 든다. 특히 순간 최대 컴퓨팅 파워의 요구를 충족시키는 컴퓨팅 서버와 저장 장치를 구매하는 것은 낭비적인 요소가 많다. 그리드 컴퓨팅은 네트워크 컴퓨팅 환경에서 대용량의 데이터와 고속 컴퓨팅이 가변적으로 요구되는 최신 응용 프로그램에 있어서 매우 효과적인 해결책이 될 수 있으리라 생각되고 있다.

그리드(Grid)라는 단어의 어원은 전력망과 유

사함에서 유래하였다. 전력은 콘센트에 플러그를 연결하는 것만으로 어디에서 발전되어 송전되었는지 전혀 의식하지 않고 사용할 수 있다. 이와 같이 그리드의 어원도 지리적으로 분산된 컴퓨터를 결합시켜 그 처리 능력을 전기처럼 발생장소를 의식하지 않고 원하는 만큼 사용할 수 있게 한다는 발상에서 온 것이다[2].

그리드 컴퓨팅은 컴퓨터의 계산 능력과 데이터 저장 능력을 마치 전기, 수도, 통신과 같은 유틸리티(utility) 개념으로 확장하자는 것이다. 예를 들어 생물학자나 천문학자가 데이터 처리를 위해서 방대한 양의 데이터를 2주 동안 처리해야 한다고 하자. 이러한 일시적인 데이터 처리를 위해 값비싼 서버를 구매하는 것이 아니라, 필요한 컴퓨팅 파워만을 사용하고 그에 해당하는 경비를 지불하는 것이 경제적이다. 이것은 마치 필요한 만큼의 전기나 수돗물을 사용하는 것과 유사하다고 볼 수 있다.

그리드 컴퓨팅의 다른 아이디어는 사용되지 않은 자원을 최소한의 경비를 들여 활용하자는 것이다. 새로운 컴퓨터를 도입하기 위해서는 도입 경비뿐 아니라 설치 공간도 필요하다. 일반 사무실의 수많은 개인용 컴퓨터는 네트워크로 연결되어 있으며, 일반적으로 컴퓨팅 능력의 90%는 사용되지 않고 있다. 이렇게 무료로 사용가능한 CPU 계산 능력을 활용하자는 것이 그리드 컴퓨팅의 기본

\* 한성대학교 컴퓨터 공학부

아이디어이다.

일기 예보, 유전자 구조의 분석, 천문 데이터 분석을 위해서는 방대한 데이터 처리가 필요하며, 이를 위해서는 슈퍼컴퓨터라는 값비싼 초고속 컴퓨터를 이용해 왔다. 슈퍼컴퓨터에 속하는 대형 IBM 시스템이나 Cray 제품들이 많이 사용되어 왔다. 그리드 컴퓨팅 기술을 적용하면 저렴한 가격으로 이러한 슈퍼컴퓨터를 대체할 수 있다. 사용되지 않는 네트워크 컴퓨터들의 컴퓨팅 능력을 활용하여 대규모 데이터를 나누어 처리함으로써, 슈퍼 컴퓨팅 파워를 발휘할 수 있다.

그리드 컴퓨팅의 적용사례로 많이 인용되는 SETI 프로젝트는 우주에서 날아오는 방대한 데이터를 분석하기 위해서 슈퍼 컴퓨팅 파워를 필요로 한다. SETI 프로젝트를 위한 그리드 컴퓨팅은 인터넷에 연결된 컴퓨터의 CPU를 활용하는 것이다. SETI 프로젝트를 위한 그리드에 속하기 위해서는 SETI@home 페이지에 연결하여 클라이언트 프로그램을 설치하면 된다. 이 클라이언트 프로그램은 일종의 화면 보호기이다. 컴퓨터가 일정 시간 사용되지 않으면 화면 보호기가 동작하면서, SETI 사이트에 연결하여 작업을 다운받아 수행하게 된다. 전세계에 퍼져 있는 3백만 컴퓨터를 활용함으로써, 15 Tera-flops의 성능을 낼 수 있다.

지난 10년간 인터넷과 World Wide Web을 통해 멀티미디어 콘텐츠를 공유할 수 있게 되었다. 또한 많은 전자 상거래와 기업 업무가 인터넷을 통하여 이루어지고 있다. 인터넷과 World Wide Web을 그리드 컴퓨팅과 결합시킴으로써, 새로운 정보 인프라를 구축할 수 있다[1]. WWW이 제공하는 정보 검색, 원격 자원 접근 기능에 그리드가 제공하는 확장성, 안전성, 고성능 메커니즘이 결합됨으로써, 그리드 컴퓨팅은 전례없는 수준의 자원 공유, 협동 작업을 제공할 수 있다[3].

## 2. 그리드 컴퓨팅의 정의

그리드 컴퓨팅을 알기 쉽게 정의한다면, 네트워크를 통해 연결된 컴퓨터의 계산 능력을 통합하여 사용할 수 있도록 하는 차세대 디지털 신경망 서비스를 의미한다. 지리적으로 분산된 고성능 컴퓨터, 대용량 저장장치, 첨단 장비 등의 자원을 고속 네트워크로 연결하여 상호 공유하고 적절히 제어하고 사용할 수 있도록 한다[4].

그리드는 과학적 응용을 실행하기 위해서 자원을 공유하고 연동할 수 있는 분산 컴퓨팅 인프라이며, 분산 컴퓨팅 환경의 신기술(disruptive technology)로 간주된다. 또한 표준에 기반한 분산 서비스를 제공하여야 하며, 분산 관리 기능을 제공하여야 한다. 그리드 컴퓨팅은 공개 소스 기술이며, 이를 위한 커뮤니티가 구성되어 있다.

Grid라는 용어는 컴퓨팅 분야에서 많이 사용되고 있다. 예를 들면 다음과 같은 것들이 있다: Computing Grids, Data Grids, Access Grids, Knowledge Grids, Bio Grids, Cluster Grids. 일부 분야에서는 그리드 컴퓨팅이 분산 컴퓨팅과 혼동되는 경우도 많으므로, I. Foster는 다음과 같이 그리드의 요건을 명시하였다[5].

- 중앙의 통제 없이, 분산 자원을 통합하여 제어할 수 있다.
- 표준화된 개방형 프로토콜과 인터페이스를 사용한다.
- 양질의 서비스를 제공한다.

이러한 정의는 매우 광범위한 분산 컴퓨팅 환경을 포함하고 있는 것처럼 보이지만, 클러스터, 네트워크, 네트워크에 연결된 저장 장치 등은 그리드에 속하지 않는다. 즉 컴퓨터 자원이 단순히 네트워크로 연결되어 통신하는 것만으로는 그리드 컴퓨팅에 포함되지 않는다.

IBM, Sun, Oracle, 후지스에서 그릴드 컴퓨팅 분야에 많은 연구와 개발을 하고 있다. Avaki의 CTO Andrew Grimshaw는 그릴드가 분산 자원을 모아서 사용자와 응용 프로그램에게 접근할 수 있도록 하는 모든 것이라고 정의하였다. IBM의 관점에서 그릴드 컴퓨팅은 분산 이질 컴퓨팅 자원은 하나의 가상 자원을 구성하며, 이러한 가상 자원에 대한 동일하고 통합적인 접근 메커니즘이라고 정의하였다[6].

컴퓨터가 네트워크로 연결되기 시작한지는 25년 정도 되었으며, WWW을 사용하기 시작한지는 10년이 되었다. WWW는 컴퓨터 활용에 많은 변화를 가져왔다. 전자 상거래, 정보 검색, 멀티미디어 정보의 교환, 원격 교육 등 많은 분야에 적용되고 있다. 그릴드는 이러한 네트워킹 및 웹 기능에 가상 조직(virtual organization)이라는 개념을 추가하였다. 또한 그릴드는 분산 자원에 신분 확인(authentication), 감시, 협동(coordination), 자원 검색의 기능을 추가하여 보다 효율적으로 작업을 수행하도록 한다. 그림 1은 기존의 웹 서비스와 그릴드 컴퓨팅을 비교 설명한 것이다. 그릴드 컴퓨팅이 기존의 웹 컴퓨팅에 비하여 많은 장점을

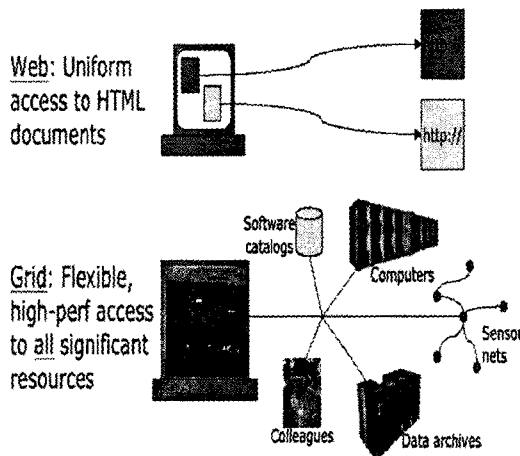


그림 1. 웹과 그릴드의 비교

가지고 있는 것은 사실이지만, 그릴드 컴퓨팅을 활용하여 응용 프로그램을 작성하는 것은 간단하지 않다. 특히 기존의 응용 프로그램을 그릴드 컴퓨팅 환경에서 실행하기 위해서는 많은 수정 작업이 필요하다.

가상 조직은 특정 목적을 위하여 자원을 공유하는 그릴드를 사용하는 조직체의 그룹을 의미한다. 가상 조직은 일반적으로 동일한 기술을 사용한다. 그릴드의 사용자는 자신의 컴퓨터에 로그인하는 것이 아니라 분산 자원이 통합된 가상 컴퓨터에 로그인하게 된다. 사용자의 자원 사용은 자원 공유 정책에 따라 결정된다. 네트워크로 연결된 컴퓨터, 프린터, 파일 서버, 데이터베이스 등이 사용자에게 별도로 보이는 것이 아니라, 통합된 하나의 가상 컴퓨터의 자원으로 보이게 된다(그림 2).

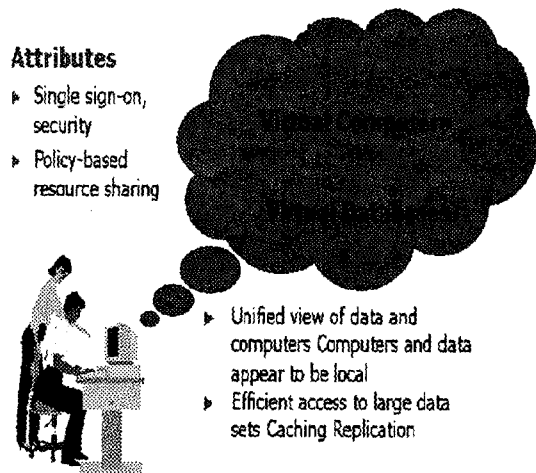


그림 2. 그릴드 = 가상 컴퓨터

### 3. 그릴드의 종류 및 적용 분야

그릴드를 응용에 따라 분류하면 1) CPU 작업 중심의 컴퓨팅 그릴드, 2) 데이터 작업 중심의 데이터 그릴드, 3) 접근 제어 중심의 액세스 그릴드

로 분류할 수 있다.

### 3.1 컴퓨팅 그리드

많은 CPU를 네트워크로 연결하여 컴퓨팅 작업을 수행하는 컴퓨팅 그리드는 작업을 동시에 여러 컴퓨터에 나누어 실행함으로써 작업 전체의 실행 시간을 줄일 수 있다. SETI 프로젝트에 사용되는 그리드 컴퓨팅이나, IBM에서 구현한 ZetaGrid 등이 이에 속한다. 주어진 시간에 가능한 많은 작업을 처리하여 throughput을 증대시키는 것이 목적이다.

### 3.2 데이터 그리드

공유하는 대규모 데이터베이스의 자료를 통합하여 분석할 수 있는 그리드이다. 일반적으로 대규모 컴퓨팅 파워와 데이터 저장 능력을 제공하여야 한다. 분산된 데이터를 분석하는 데이터 마이닝이나 대규모 과학 데이터를 분석하는데 응용될 수 있다. 현재 EU에서는 DataGrid 프로젝트가 진행 중이며, 10의 12승에서 10의 15승 바이트의 분산된 과학 정보를 해석하는 것이 그 목표이다.

### 3.3 액세스 그리드

액세스 그리드는 그리드 접근을 위한 인간 상호 작용을 지원하기 위해서 사용될 수 있는 자원의 집합체이다[8]. 액세스 그리드는 멀티미디어 디스플레이, 출력형식(presentation), 인터랙션 환경, 그리드 미들웨어 인터페이스, 시각화 환경 인터페이스로 구성된다. 액세스 그리드는 대규모 원격 회의, 협동 작업, 세미나, 강의 등을 지원한다. 액세스 그리드 설계의 중요점은 그룹간의 커뮤니케이션이다.

### 3.4 연구 적용 분야

그리드 컴퓨팅을 적용하는 분야는 일반적으로

방대한 컴퓨팅이 필요한 분야이다. 국내의 그리드 연구자들은 그리드의 5개 적용분야를 선정하였다.

- 생명공학 기술 (Bio Technology)
  - 유전자 분석등 방대한 생명 정보를 분석하는데 적용한다.
- 나노 기술(Nano Technology)
  - 물질을 원자, 분자 수준에서 시뮬레이션하고 분석하는데 적용한다.
- 환경 기술(Environment Technology)
  - 대기나 지구 내부 활동의 변화를 예측하는데 적용한다.
- 항공 우주 기술(Space Technology)
  - 기계 항공 분야의 구조 해석이나 유체 역학분야의 시뮬레이션에 적용한다.
- 전통 제조 기술(Traditional Technology)
  - 자동차, 기계, 선박 등의 설계 및 분석에 적용한다.

## 4. 그리드 개발틀: Globus Toolkit

Globus Toolkit은 본래 1995년 IWAY 프로젝트로 개발된 그리드 미들웨어이며, 그리드 시스템을 구축하는데 필요한 기본적인 틀이다. 1998년 Globus 툴킷 1.0이 발표되었으며, 3.0 베타 버전까지 공개되어 있다. 분산 이기종 컴퓨터 자원들을 하나의 가상 컴퓨터 자원처럼 사용할 수 있도록 해준다. 현재 EU DataGrid, NASA Information Power Grid, Grid Physics Network, ASCI DRM Test-bed, GUSTO 등의 그리드 프로젝트이 Globus Toolkit이 사용되고 있다.

Globus 툴킷은 X.509, LDAP, SOAP 등의 표준을 수용하고 있으며, 3.0 버전에는 자원 관리 기능, 레지스트리, 데이터 서비스 기능, 작업 제출 및 감시 기능이 추가되었다. 그림 3은 Globus 툴킷의 구조를 나타낸 것이다. Globus 툴킷은 크게 그리

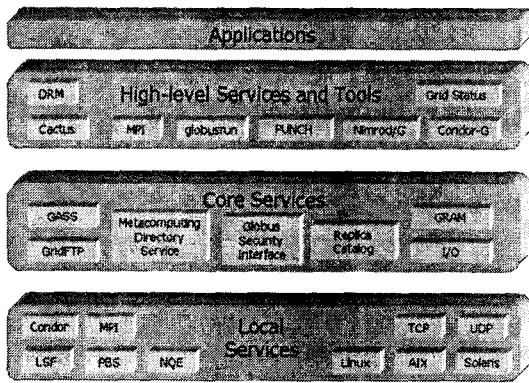


그림 3. Globus 툴킷의 구조

드 보안, 정보 서비스, 자원 관리, 데이터 관리 등으로 나누어진다. 핵심 서비스(Core Services)의 모듈 중에서 GRAM은 Grid Resource Allocation Manager의 약자이며, 그리드 자원 관리를 담당한다. MDS는 Meta Directory Service의 약자이며, 네이밍/디렉토리 서비스를 담당한다. 그리드에 존재하는 자원들의 상태 정보를 공유하고 사용자들에게 제공하기 위한 요소이다. 자원의 이름 및 특성 정보를 제공하기 위해선 MDS는 LDAP 표준을 사용한다. GASS는 Global Access to Secondary Storage의 약자이며, 그리드 컴퓨터의 보조 기억 장치의 접근을 관리한다.

### 5. 그리드 컴퓨팅의 전망

그리드 컴퓨팅의 미래를 전망하는 것은 간단하지 않다. 많은 IT 기술과 마찬가지로 그리드 컴퓨팅은 많은 난제와 불확실성을 가지고 있으며, 앞으로 많은 기술적 발전이 필요한 분야이다. 그리드 컴퓨팅은 분산 자원을 통합하여 관리하여야 하므로, 상호 연동과 표준 프로토콜의 제정이 필수적이다. 많은 그리드 연구자들과 대형 소프트웨어 업체들이 표준의 필요성을 인식하고 표준 제정에 노력하고 있지만, 다양한 집단의 이해관계가

단순하지는 않으리라 생각된다. 그리드를 구성하는 분산 시스템 자원의 보안 문제는 매우 복잡하며 앞으로 해결해야 될 중요 문제 중에 하나이다.

그럼에도 불구하고 I. Foster[9]는 그리드 컴퓨팅의 성공을 예상하고 있다. 많은 분산 자원을 사용한 웹 기술의 경우에도 많은 기술 표준 및 상호 연동 등에 관한 문제점을 적절히 극복한 경험이 있다. 또한 그리드 컴퓨팅의 기술 표준을 위해서 GGF(Global Grid Forum)이 결성되어 있으며, 매우 현실적인 접근을 하고 있다는 점도 그리드 컴퓨팅의 미래를 밝게 볼 수 있는 이유 중에 하나이다. 그리드 컴퓨팅의 요구사항이 웹 서비스의 요구사항을 포함하고 있으므로, 많은 웹 사용자들이 자연스럽게 그리드 컴퓨팅으로 이동하게 될 것이다. 또한 Globus라는 기본적인 공개 소스 그리드 개발 toolkit이 제공되고 있으며, 이를 사용하는 커뮤니티가 날로 커지고 있다. 현재 IBM, 오라클, Sun, HP, 후지스 등 주요 소프트웨어 기업들이 그리드 컴퓨팅의 가능성을 인정하고 있으며, 자신들의 제품에 그리드 컴퓨팅 개념을 도입하려고 노력하고 있다.

I. Foster[9]는 미래의 분산 컴퓨팅의 수요를 그림 4와 같이 예측하였다. 분산 컴퓨팅 환경이 DCE, CORBA, 그리고 웹 서비스를 거쳐, 그리드 컴퓨팅과 웹 서비스를 통합한 형태로 발전할 것이

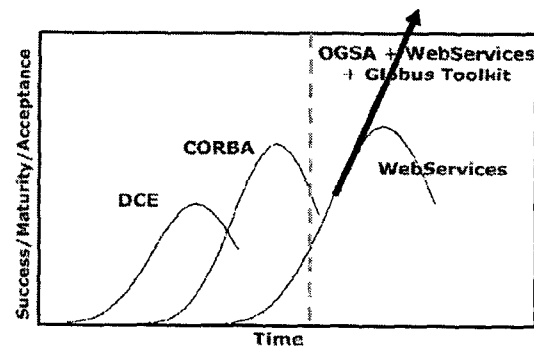


그림 4. 그리드 컴퓨팅의 전망

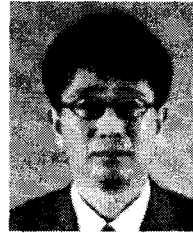
라고 예상하였다. 즉 그리드 컴퓨팅이 차세대 웹 서비스를 대체할 것이라는 전망이다. 이러한 장밋빛 예상이 어느 정도 현실로 이루어 질 것인지, 현재로서는 알 수 없지만, 미래의 컴퓨팅 환경은 웹 브라우저를 사용한 클라이언트의 요청이 그리드 컴퓨팅에 의해 수행될 것은 가능성이 매우 높아 보인다.

## 참 고 문 헌

- [1] Ian Foster. "The Grid: New Infrastructure for 21st Science", *Physics Today*, Feb. 2002.
- [2] 이춘희, "그리드 컴퓨팅", 정보처리 학회지 제10권 1호, 2003년 1월.
- [3] *National Collaboratories: Applying Information Technology for Scientific Research*, National Academy Press, Washington, D.C. (1993).
- [4] Naver 백과사전, "그리드 컴퓨팅"의 정의, <http://100.naver.com>
- [5] Ian Foster, "What is the Grid? A Three Point Checklist", <http://www-fp.mcs.anl.gov/~foster/>
- [6] Paul Korzeniowski. "IBM's Future Strategy: Grid Computing Everywhere", TechNewsWorld,

Aug. 06, 2003.

- [7] Ian Foster, Carl Kesselman, Steven Tuecke. "The Anatomy of the Grid", International J. Supercomputer Applications, 2001.
- [8] "EuroAG - European Access Grid", <http://euroag.accessgrid.org/>
- [9] Ian Foster, "The Grid Computing", OpenGroup Boston Conference, 2003.



김 일 민

- 1984년 경북대학교 전자공학과 (학사)
- 1989년 뉴저지 공과대학 전산과 (석사)
- 1995년 아리조나 주립대 전산과 (박사)
- 1985년~1987년 전자통신 연구원(ETRI) 연구원
- 1996년~1997년 삼성 SDS 교육개발센터
- 1997년 3월~현재 한성대 컴퓨터공학과 교수
- 관심분야: 분산 컴퓨팅, 사이버 교육
- E-mail : ikim@hansung.ac.kr