

신기술
해설

그리드 컴퓨팅 (Grid Computing)

이 춘 희¹⁾

목 차

- 1. 서 론
- 2. 그리드의 분류
- 3. 그리드의 개념적 서비스 3계층
- 4. 그리드 아키텍처
- 5. 그리드의 인프라스트럭처
- 6. 세계 주요 그리드 프로젝트
- 7. 그리드의 표준화
- 8. 결 론

1. 서 론

그리드 컴퓨팅이란 용어가 우리나라의 과학계에 널리 알려진 것은 정보통신부 주관 하에 그리드 포럼 코리아(Grid Forum Korea)의 출범(2001. 10. 25)이후가 아닌가 생각된다. 지금까지 그리드 컴퓨팅의 거대 과제들은 주로 미국, 유럽 등 세계 선진국을 위주로 수행해 왔었다. 그러나 국내에서도 정보통신부가 2002년부터 2006년까지 435억원을 투자하여 국가 그리드 기반 구축사업(2002~2006) 및 그리드 미들웨어사업(2002~2004)을 본격적으로 수행하게 됨으로써 국제적으로 IT 기술 선진국 대열에 올라서는 계기가 되었고 이들 프로젝트의 성공 여부가 우리나라 첨단 IT 기술의 국제적 평가를 받을 수 있는 좋은 장이 마련되었다고 할 수 있다. 이러한 의미 있는 「그리드 컴퓨팅」의 이모저모를 살펴보기로 한다.

그리드(Grid)라는 단어의 어원은 전력 그리드(전력망)와 유사함에서 유래한다. 전력은 콘센트에 플러그를 꽂는 것만으로도 어디에서 발전되고 있는가를 전혀 의식하지 않고 사용할 수 있다. 이

와 같이 그리드의 어원도 지리적으로 분산된 컴퓨터를 결합시켜 그 처리 능력을 전기처럼 발생 장소를 의식하지 않고 원하는 만큼 사용할 수 있게 한다는 발상에서 온 것이다.

World Wide Web은 네트워크를 통해 문장과 화상 등의 콘텐츠(Contents)를 공유할 수 있게 하였다. 그리드라는 개념은 단순히 콘텐츠만이 아니라 컴퓨터의 모든 기능을 세계가 공유하고자 하는 비전에 바탕을 둔다. 현재 각 국가에서 실제로 그리드 구축이 진행되고 있지만 아직은 이러한 비전을 향하여 제 1보를 내딛은 단계이다. 그리드는 수많은 조직으로부터 형성된 동적인 가상조직(Virtual Organization)으로서 자원을 공유하여 문제 해결을 꾀하는 개념이다[1]. 즉, 그리드는 컴퓨터를 공유함으로써 가상적인 연구소를 만드는 시도라고 말할 수 있다. 근래에 저가로 고속계산을 하기 위해 많은 PC를 병렬로 연결한 클러스터 컴퓨터가 이용되고 있다. 기술적으로 그리드란 이것을 타 기관의 컴퓨터까지 확장하여 클러스터 컴퓨터와 같이 프로그래밍할 수 있게 하는 컴퓨터 네트워크다.

사실 그리드의 개념은 1985년 I-Way 실험에서 처음 시작되었다. I-WAY 프로젝트는 미국의 대

1) 한국과학기술정보연구원 전문연구위원

형 슈퍼컴퓨터센터의 자원을 결합하는 수단으로서 17개의 사이트를 통하여 고성능 컴퓨터와 첨단 가시화 환경을 구축하는 것이다. 이 실험에서는 고속 네트워크가 북아메리카를 가로질러 단시간에 17개 사이트의 첨단 자원을 연결하는데 사용되었다. 이 성공적인 실험 결과로 인해 많은 그리드 연구 프로젝트가 시작하게 되었다. 예를 들면, 미국 국립과학재단(NSF)의 「국가기술 그리드」와 NASA의 IPG(Information Power Grid)는 각각 대학과 NASA 연구자들에게 제공할 그리드 인프라를 구축하는 프로젝트이고, 그 외에 European DataGrid, Particle Physics Data Grid, GriPhyN (Grid Physics Network), NEESgrid(the Network for Earthquake Engineering Simulation Grid) 등 다수의 프로젝트가 시작되었다.

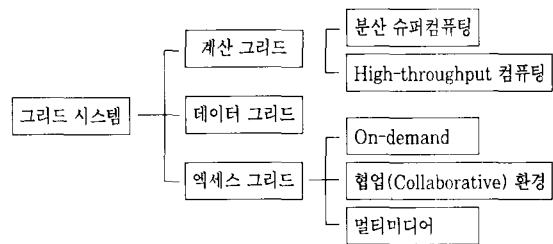
이러한 네트워크가 필요하게 된 것은 과학연구에서 대량의 데이터 처리가 필요해졌기 때문이다. DataGrid의 중심 기관인 CERN의 경우 현재 27Km의 LHC(Large Hadron Colider)라는 실험장치를 건설하고 있는데, 이 장치에서 실험이 실행되면 일초 동안 약 1000조 바이트 (1 Peta bytes)의 데이터가 관측되고 그 중에서 유용한 데이터만 선택해도 일년 동안에 똑같이 약 1 페타바이트의 데이터를 저장해야 한다. 또한 천문학, 환경학, 생물학 등 다른 분야에서도 대량의 데이터가 사용된다. 생물학의 경우 A, T, G, C라는 4문자로 기록된 대량의 염기 서열 데이터를 처리해야 한다. 염기서열은 생명의 설계도이며 현재의 생물학에서 중심적인 역할을 하고 있다. 이러한 이유에서 생물학과 같이 이전에는 IT와 거리가 있었던 분야에서도 강대한 컴퓨팅 파워가 필요하게 되었다.

그러나 이러한 필요성에 비해 연구 자금은 한정되어 있어서 각 연구기관이 독자적으로 충분한 컴퓨팅 파워를 갖는다는 것은 어려운 실정이다. 따라서 다른 연구기관이 관리하고 있는 컴퓨터를 이

용하여 분산 컴퓨팅을 수행하는 그리드 컴퓨팅의 필요성이 높아진 것이다. 또한 우수한 연구를 진행하기 위해서는 여러 분야의 전문성을 가진 연구자들이 협력하는 것도 또한 중요하다(CERN의 실험에는 세계에서 수 천명의 연구자들이 참여). 이러한 점에서 우수한 연구를 성공적으로 수행하기 위하여 그리드는 필요한 컴퓨팅 환경을 구축하는 것이다.

2. 그리드의 분류

그리드 시스템을 애플리케이션 특징에 따라 분류해 보면, 많은 양의 계산을 수행하는 어플리케이션, 데이터 전송이 빈번한 어플리케이션, 처리하는 데이터의 양, 여러 다른 분야 전문가들의 협력 연구의 필요성 여부 등에 따라 다음과 같이 분류할 수 있다.



2.1 계산 그리드

많은 자원을 연결하여 계산을 해결할 수 있게 하고, 분산 슈퍼컴퓨팅 관련 그리드는 많은 컴퓨터를 동시에 병렬로 사용하게 하여 작업의 전체 실행시간을 줄일 수 있다. 예를 들면, 기계 항공 분야나 물리분야의 초대형 문제들을 해결하기 위해 사용된다. High-throughput 컴퓨팅은 주어진 시간 안에 가능한 한 많은 작업을 처리하는 것이 목표다.

2.2 데이터 그리드

원격지의 분산된 자료들을 통합하여 분석할 수

있게 해 주는 그리드다. 대표적인 응용분야로는 여러 분산된 곳에서 생성된 자료를 한곳에서 처리하는 형태의 데이터 마이닝이나 분산된 대용량의 데이터를 여러 곳에서 처리해야 하는 고에너지 물리분야, 유전자 정보 처리분야 등이 있다.

2.3 액세스 그리드

분산처리를 필요로 하는 어플리케이션을 위한 그리드다. 액세스 그리드도 역시 사용분야에 따라 나뉠 수 있는데, On-demand 그리드는 사용자 작업의 요구에 따라 동적으로 자원들을 연결할 수 있는 그리드다. 액세스 그리드는 그리드를 통해 인간의 상호작용(Interaction)을 지원하는 자원의 융합이다. 이것은 멀티미디어 디스플레이, 프레젠테이션, 인터랙션 환경, 그리드 미들웨어의 인터페이스, 가시화 환경의 인터페이스 등으로 구성된다. 액세스 그리드를 통해 대규모의 분산 환경 회의, 협업 회의, 세미나, 강의, 교육 및 훈련 등이 지원된다. 액세스 그리드의 설계 요점은 그룹 대 그룹 통신이다. 액세스 그리드에서 노드의 개념은 특히 그리드에 “그룹” 액세스를 제공하는 것이 목적이다.

2.4 협업환경 그리드

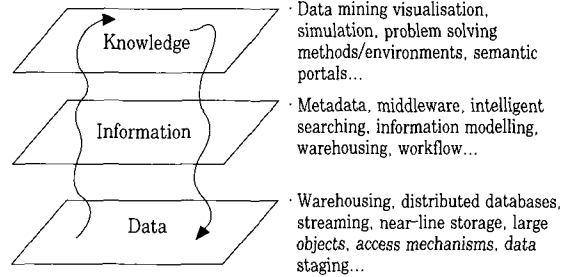
사용자들이 원격지에서 협력하여 작업을 할 수 있게 하는 작업공간을 제공한다.

2.5 멀티미디어 그리드

실시간 멀티미디어 어플리케이션을 위한 인프라를 제공한다. 멀티 미디어 그리드는 다양한 스트림이 동시에 교환되기 때문에 네트워크 QoS가 반드시 보장되어야 한다.

3. 그리드의 개념적 서비스 3계층

Keith Jeffery가 제안한 그리드의 개념적 서비



(그림 1) 개념적 서비스 3계층

스 3계층은 (그림 1)과 같다.

3.1 Data/Computational 그리드

안전한 방법으로 고속 대역폭 및 데이터 스토리지와 관련된 컴퓨팅 파워를 제공하는 그리드의 구조(Fabric)를 형성한다. 다양한 자원은 단일 메타컴퓨터(Virtual Computer)로서 표현된다. 계산 자원의 할당, 스케줄링, 실행방법과 데이터를 프로세싱 자원에 보내는 방법 등을 이 계층에서 다루게 된다.

3.2 Information 그리드

정보를 모든 형식으로 표현, 저장, 공유 및 유지하는 방법으로 처리하여 이기종의 분산 정보에 동종의 액세스를 제공한다. 이 계층은 요구를 충족시키기 위해 워크플로우를 구성하는 툴킷, 메타데이터 액세스, 가시화, 데이터 관리, 장비관리에 대한 툴킷을 포함하여 데이터와 애플리케이션을 잘 융합시킨다. 웹과 현재 미들웨어 기술은 하나의 프레임워크로 통합된다.

3.3 Knowledge 그리드

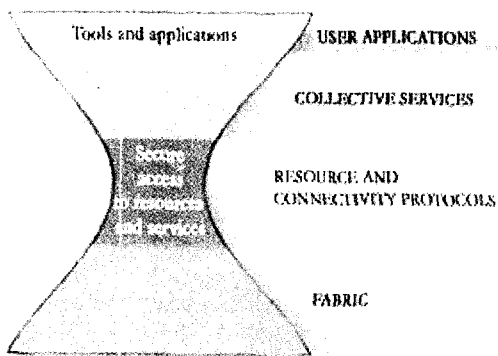
지식 기반의 방법론과 기술을 사용하여 고 수준의 질문에 응답하고, 요구하는 형식에 맞는 답을 보내기 위해 적절한 프로세스를 찾는다. 이 계층은 지식이 요구하는 방법(획득, 사용, 검색, 유지 보수 등)을 지원하기 위해 데이터마이닝, 머신러

닝(Machine Learning), 시뮬레이션, 지능적 포탈, 워크플로우 추론 및 문제 해결 환경(PSEs) 등을 포함하고 있다.

현재 전 세계적으로 그리드 개발을 위한 과제가 국가 슈퍼컴퓨팅센터들을 중심으로 컨소시엄을 구축하여 진행되고 있다. 그 대부분이 계산 그리드와 데이터 그리드를 결합한 하이브리드 형태를 시도하고 있으며 수직적인 분류에서 본다면 데이터 그리드와 정보 그리드에 초점이 맞춰져 있다.

4. 그리드 아키텍처

원격 오퍼레이션을 요구하는데 사용되는 메시지 교환 컨텐츠와 순서를 정의하는 표준 프로토콜은 그리드 시스템의 상호운용에 중요하고도 필수적인 수단이 되었다. 또한 표준 애플리케이션 프로그래밍 인터페이스(API)도 필수적이다. API는 코드 라이브러리에 대해 표준 인터페이스를 정의하고 코드 요소들을 재사용하게 함으로써 그리드 요소를 쉽게 구성하게 한다. (그림 2)는 프로토콜과 API가 그리드 시스템에서의 역할에 따라 분류되는 것을 보여준다.



(그림 2) 그리드의 구조

4.1 Fabric 계층

최하위 계층은 Fabric으로 그리드 사용자가 컴

퓨터, 스토리지 시스템, 카탈로그, 네트워크, 다양한 형태의 센서 등을 공유하고 액세스하기를 원하는 물리적 장비와 자원을 가지고 있다. 자원들은 각자의 메커니즘으로 유지되어도 무관하며, 다양한 기술을 사용해도 되므로 구성에 대해 어떠한 제약도 받지 않는다.

4.2 연결 계층 및 자원 계층

Fabric 상위층은 연결 계층과 자원 계층이다. 이들 계층의 프로토콜은 어느 곳에서나 구현될 수 있어야 하므로 비교적 수가 적어야 한다. 연결 계층은 그리드 특유의 네트워크 트랜잭션에 필요한 통신의 핵심과 인증 프로토콜을 포함한다. 통신 프로토콜은 자원 사이에 데이터 교환을 가능하게 한다. 그리고 인증 프로토콜은 사용자와 자원의 신원을 검증하기 위한 암호로 보안 메커니즘을 제공하기 위해 통신 서비스를 구축한다. 통신을 위한 프로토콜은 IP, DNS, Routing 등으로 이루어지며, 보안에 관련된 사항은 GSI(Grid Security Infrastructure)에서 제공한다. 일정한 인증과 권한설정, Single Sign-on 서비스가 지원되고, 공개키, SSL, X.509, GSS-API를 제공한다.

자원 계층은 통신과 보안 개시, 감시, 자원 공유 실행의 제어를 가능하게 하는 인증 프로토콜을 개발하는 프로토콜을 포함한다. 다른 컴퓨터 시스템에서 같은 프로그램을 실행하는 것은 자원계층 프로토콜을 따른다. 글로벌 툴킷은 연결 소스와 자원 프로토콜 그리고 API에 공통적으로 사용된다.

4.3 수집계층

수집계층은 수집한 자원을 통하여 상호작용을 구현하는 프로토콜과 서비스, API를 포함한다. 그들은 비교적 제한된 자원과 연결 계층으로부터 요소들을 결합하여 이용한다. 수집계층의 요소들은 새로운 자원계층의 요소를 요구하지 않고도 다양한 태스크를 구현할 수 있다. 수집 서비스의 예

로는 자원 발견과 할당을 위해 디렉토리와 브로커링 서비스가 포함된다. 감시 및 진단 서비스, 데이터 중복 서비스도 포함되며 커뮤니티 내에 사람의 행적을 유지하기 위한 멤버십, 정책 서비스는 자원에 액세스를 가능하게 한다.

4.4 애플리케이션 계층

그리드 시스템의 최상위층이 사용자 애플리케이션이다. 이 상위 계층은 다른 계층의 요소들을 구성하고 요구한다. 예를 들면, 수 천개의 독립된 태스크를 처리하는데 필요한 고에너지 물리 분석 애플리케이션은 각각 사건을 포함하는 파일을 입력하여 다음의 기능을 진행할 수 있다.

- Obtaining : 필요한 인증 신임장을 획득한다 (연결계층 프로토콜).
- Querying : 컴퓨터, 스토리지 시스템, 네트워크, 요구한 입력파일 위치의 가용성 등을 결정하기 위해 정보 시스템 및 중복 카탈로그를 질의(수집 서비스)한다.
- Submitting : 계산과 데이터 이동 등과 같은 여러 가지 작업을 시작하기 위해 적합한 컴퓨터, 스토리지 시스템 및 네트워크에 요구한다.
- Monitoring : 여러 가지 계산과 데이터 전송 과정을 감시한다. 과정이 모두 완료되었을 때 사용자에게 알려주고, 실패 조건을 검출하고 응답해 준다(자원 프로토콜).

5. 그리드의 인프라스트럭처

컴퓨터의 속도는 18개월마다 그리고 네트워크의 성능은 9개월마다 두 배가 되는 비율로 네트워크가 컴퓨터의 속도를 앞지른다면 통신은 본질적으로 전혀 문제가 없다. 따라서 통신 집중적인 새

로운 방법으로 계산 자원의 풀링(Pooling), 데이터베이스 또는 장비로부터 원격 컴퓨터까지 대량 데이터의 흐름, 센서간의 연결, 컴퓨터간의 연결, 저장소간의 연결, 인력, 컴퓨팅, 스토리지 등 비용이 적게 드는 협업 환경으로 연결하는 것이다. 통신에 제한이 없고 비용이 들지 않는다면 문제를 해결하기 위해 지역 자원을 사용하는데 제한이 없게 된다.

원격자원은 일반적으로 다른 사람의 소유이고 다른 관리 영역 안에 존재하며 다른 소프트웨어를 가동하고 다른 보안과 액세스 정책을 가지고 있다. 실제로 원격 자원을 사용하려면 여러 단계를 거쳐야 한다. 첫째, 그들이 존재한다는 것을 알아내야 하고, 그 다음은 그들과 액세스하기 위해 교섭을 해야 한다. 그리고 효율적으로 자원을 사용하기 위해 하드웨어와 소프트웨어를 구성해야 한다. 또한 나 자신의 보안, 내가 사용할 원격 자원의 보안과의 타협 없이 이 모든 일이 진행되어야 하고 어떤 작업은 그에 합당한 대가를 지불해야 한다.

이러한 단계를 구현하는 것은 분산 자원에 대한 싱글 Sign-on 지원, 고속의 대량 데이터 전송, 대형 분산 가상 커뮤니티 형성, 커뮤니티 자원의 존재, 상태 그리고 이용 정책에 대한 정보의 유지 관리 등 원격 컴퓨터에서 서비스를 생성하고 관리하는 중요한 태스크에 대한 균일한 메커니즘을 필요로 한다.

오늘날의 인터넷과 웹 기술은 기본적인 통신을 다루지만, 대규모로 그리고 가능한 한 간단히 자원 공유를 안전하게 할 수 있는 인프라와 툴의 제공을 위해서는 Grid가 필요하다.

5.1 그리드 관련 인프라스트럭처

인터넷에 액세스를 하기 위해서는 장비 안에 프로토콜들이 구현되어 있어야 하듯이 그리드 인프라에는 인터넷에 존재하는 것보다 더 많은 기능이

필요하다. 그리드 인프라 기술은 유용하게 사용되기 위해서 광범위하게 전개되어야 하며 간단하고 가치가 높아야 한다. 또한 이것은 고속 데이터 이동, 대규모 데이터의 캐싱, 컴퓨팅의 주문 액세스와 같은 그리드에 힘을 실어주는 자원을 지원하기 위해 존재한다.

튠은 인프라 서비스를 사용가능 하게 한다. 인터넷과 웹 튠은 원격 웹사이트를 액세스하기 위한 브라우저, 전자 메시지를 다루는 e-mail 프로그램, 웹 페이지 위치를 알아내는 검색엔진 등이 있다. 그리드 튠은 자원 발견, 데이터 관리, 계산 스케줄링, 보안 등과 관련된다.

그러나 그리드는 데이터를 공유하고, 분산하고, 자원을 계산하는 것 이상의 힘을 가진다. 과학자들에게 그리드는 새롭고 훨씬 강력한 작업 방법을 제공한다. 그 예는 다음과 같다.

5.1.1 과학적 포털(Science Portals)

과학적 애플리케이션은 컴퓨팅 인프라가 차세대 개발하는 추진력이 되고 있다. 다른 애플리케이션 분야에서 필요로 하고 사용 가능한 공통된 요소를 확인하는 작업은 매우 중요하다. 공통되고 재사용할 수 있는 요소들의 가용성 향상은 다른 애플리케이션 분야에서의 유사한 개발 노력을 가속화할 것이다. 다른 분야들 사이에서 공통된 요소가 일단 확인되면 그것은 과학적 포털(Science Portal)을 구축하는데 사용될 수 있다. 일반적으로 과학적 포털은 특정 과학 분야 또는 다른 어떤 연구 분야와 관련된 정보에 대해 게이트웨이 또는 포털을 정의한다. 과학적 포털은 튠, 데이터베이스, 작업 제출 및 협업 공간에 대한 웹 인터페이스로서 역할을 한다. 또한 웹 브라우저 또는 다른 간단한 "Thin Clients"를 쉽게 다운로드 하여 원격으로 복잡한 패키지를 불러내어 첨단 문제해결 방법을 보다 쉽게 사용하게 한다. 그러한 패키지는 그리드 내에 적합한 컴퓨터에서 원격으로 처리

할 수 있다. 현재 생물학(Biology), Fusion(물리 핵융합), 계산화학, 그리고 다른 분야에서 이러한 포털이 개발되고 있다.

5.1.2 분산 컴퓨팅

고속 워크스테이션과 네트워크는 실질적인 계산 자원을 형성하기 위해 조직의 PC들을 결합시킬 수 있다. Entropia사의 Fight-AIDSAthome 시스템은 AIDS 약품 후보를 분석하기 위해 30,000대 이상의 컴퓨터를 사용하고 있다. 2001년에 수학자들은 "Ng30"이라 명명한 최적화 문제에 있어서의 특정 인스턴스를 풀기 위해 미국과 이탈리아에 걸쳐서 그들의 계산 자원을 공동관리 하였다. 일주일 동안에 협업이 평균 630대의 컴퓨터를 사용하였고 Nug30을 수행하기 위해 최고 1,006대의 컴퓨터, 전체 42,000 CPU/day를 사용하였다. 앞으로 네트워크 성능은 더욱 향상될 것이고, 그리드 기술에 있어서는 통합된 컴퓨팅 자원이 다룰 수 있는 문제의 범위가 증가될 것이다.

5.1.3 대규모 데이터 분석

많은 관심을 모으고 있는 여러 과학 문제들은 대량의 데이터 분석을 필요로 한다. 그런 문제들을 해결하기 위해 분산 컴퓨팅과 스토리지 자원을 사용하는 것은 확실히 대단한 가치가 있다. 더구나, 많은 데이터 분석 절차에서 분산 자원을 효율적으로 사용하기 위한 병렬화는 상당한 타당성이 있다. 예를 들면, LHC와 장래 다른 고 에너지 물리 실험에서 산출되는 페타바이트의 대량의 데이터를 분석하는 일은 중간 결과를 보관하기 위해 수만 대의 프로세서와 수백 테라바이트 디스크 스페이스를 필요로 한다. 여러 기술과 정책적인 이유로 인해 단일 장소에서 이들 자원을 모으는 것은 비현실적이다. 그러나 이런 실험에 참여하는 수백 개의 연구소로부터 모아진 기관 및 국가 자원은 이들이 필요로 하는 자원을 제공할 수 있다. 이들

커뮤니티는 단순히 컴퓨터와 스토리지 이상으로 더 많은 것을 공유할 수 있으며 분석 절차와 계산 결과 또한 공유할 수 있다.

5.1.4 과학 기구(Instrumentation)

망원경, 싱크로트론(Synchrotrons), 전자현미경 같은 과학 기구는 배치(Batch) 프로세싱으로 저장된 가공되지 않은(Raw) 데이터 스트림을 만들어 낸다. 그러나 Quasi-Realtime 분석은 이러한 과학 기구의 능력을 훨씬 더 향상시킬 수 있다.

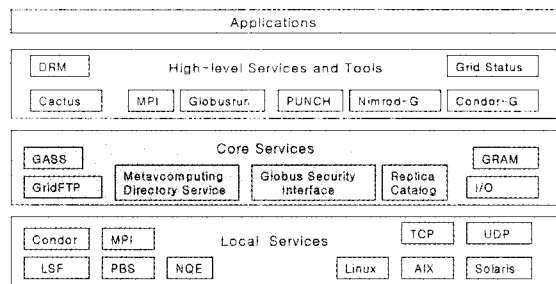
5.1.5 협력 작업(Collaborative Work)

연구자들은 데이터와 컴퓨팅뿐만 아니라 전문 인력도 모으기를 원한다. 그래서 협업 문제 공식 표시, 데이터 분석 등은 중요한 그리드 애플리케이션이다. 예를 들면, 거대한 규모의 수 테라바이트 시뮬레이션을 수행하는 천체물리학자들은 전체 그룹이 실시간으로 같은 시간, 같은 방법으로 그 결과에 대해 토론할 수 있도록 하기 위해 세계 각지의 동료들이 결과에 대해 구체적으로 가시화하기를 요구할 수도 있다. 실제 그리드 애플리케이션에서는 이러한 상황이 자주 발생할 것이다. 예를 들면, Radio 천문학자들은 비슷한 사건을 국제적인 저장소에서 찾아 수행도중에 동료들과 결과를 토론하고 대체 알고리즘을 평가하기 위해 분산 컴퓨팅을 실행한다.

5.2 글로벌스 툴킷

글로벌스 툴킷은 그리드 서비스를 제공하는 미들웨어로서 계산 그리드를 구축하는데 필요한 기본적인 기술이다. 이것은 그리드와 그리드 애플리케이션을 지원하는 커뮤니티 기반, 공개 구조, 공개 소스 집합의 서비스 및 소프트웨어 라이브러리로서 보안, 정보 발견, 자원 관리, 자료 관리, 통신 오류감지, 이식성 등 그리드에서 필요한 다양한 서비스들을 독립적인 요소로 제공한다. 현재

National Technology Grid, European DataGrid, NASA Information Power Grid (IPG), Grid Physics Network, ASCI Distributed Resource Management(DRM) Testbed, GUSTO 등 세계적인 그리드 개발 과제에서 가장 많이 사용되고 있다.



(그림 3) 글로벌스 툴킷 구조

Globus 툴킷은 크게 그리드 보안, 정보 서비스, 자원 관리, 데이터 관리 등으로 나뉘어 진다.

Globus 툴킷에서 보안을 담당하는 부분이 공개 키 방식의 GSI(Globus Security Infrastructure)이며, 이것은 단일 인증방식(Single Sign-on), 통신 보호, 제한된 위임(Delegation) 등을 제공한다. 사용자는 그리드 환경 내에서 한번만 인증을 받으면 프로그램이 계속해서 허용된 자원들을 사용할 수 있고 분산된 각 자원에 대한 사용자 인증은 Proxy가 대신 수행한다. 그러나 각 자원 내에서 자원 사용에 대한 허용범위는 각 자원이 제시하는 보안체계를 따른다.

MDS(Meta Directory Service)는 정보서비스를 수행하는 요소다. 그리드내에 존재하는 자원들의 상태 정보를 공유하고 사용자들에게 제공하기 위한 요소로서 인터넷의 DNS와 유사하다. 정보를 저장하고 사용자들에게 제공하기 위해 MDS는 LDAP를 이용한다. 정보 서비스를 위해 Globus에서는 두 개의 서버를 제공하는데, 그것은 각 자원의 정보를 수집하는 GRIS(Grid

Resource Information Service)와 수집된 정보를 통합하는 GIIS(Grid Index Information Service)이다. 이들이 수집하여 제공하는 정보는 각 자원의 구조, 노드 수, 부하 정보, 배치작업 스케줄러, 네트워크 상태 등이다. 이러한 정보는 어플리케이션 개발자나, 자원 브로커 등에게 제공된다.

자원 관리를 담당하는 부분은 GRAM이라 한다. GRAM 프로토콜은 계산 자원에 대해 신뢰성 있고 안전한 원격 생성 및 관리를 수행한다. 원격 계산에 대한 인증, 권한 부여, 위임 작업을 하기 위해 GSI를 사용한다. 서비스 생성은 규모가 작고 신뢰성이 있는 게이트키퍼(Gatekeeper) 프로세서가 관리하며 GRAM 리포터(Reporter)는 내부 계산의 신원 및 상태에 관한 정보를 모니터링하고 출력한다.

또한 글로벌스에서는 데이터 관리를 위해 GASS, GridFTP, 중복 카탈로그를 제공한다. GASS는 GRAM과 밀접한 관련이 있는 요소로서 원격지에 있는 파일을 사용하여 작업을 처리하기 원하거나 원격지에서 처리한 작업의 결과를 또 다른 저장장치에 저장하기를 원할 때 사용한다. GridFTP는 그리드 내의 데이터가 대규모, 대용량이란 점을 고려하여 고속의 파일 전송과 파일의 이어받기를 가능하게 하는 요소이다. 중복 카탈로그는 데이터 그리드를 위해 개발된 것으로 데이터들을 분산 저장 및 관리함으로써 필요할 때 신속하게 데이터를 사용할 수 있게 하는 기술이다.

1998년 그리드 미들웨어인 Globus 툴킷[1] 1.0이 발표된 후 현재는 2.2까지 출시되었다. 현재의 그리드에서 표준은 존재하지 않지만 Globus 툴킷에서 이루어지고 있는 작업들이 암묵적으로 그리드가 이루어져야 할 요소 기술들로 인식되고 있다. 2002년 초반에 발표된 Open Grid Services Infrastructure(OGSI : Grid Service Specification[2]를 중심으로 이루어진다)를 바탕으로 한 Globus 툴킷 3.0은 2002년 말에 알파

버전이 나올 예정이다. Globus 툴킷 3.0은 웹 서비스를 기반으로 사용자가 좀 더 손쉽게 사용할 수 있는 형태를 갖출 것으로 예상된다.

6. 세계 주요 그리드 프로젝트

이미 주요 선진국에서는 그리드 컴퓨팅 프로젝트와 관련된 연구가 활발히 진행되고 있다. 오늘날 그리드 프로젝트를 추진하고 있는 대부분의 어플리케이션은 높은 대역폭과 대량의 데이터 및 계산 사이클이 필요한 대규모의 복잡한 문제를 풀고 있는 큰 조직에서부터 시작한다. 대규모의 e-Science가 그리드 컴퓨팅 개발의 최초 동기였다.

그리드 프로젝트는 다양하고 일반적인 범주에 의해 크게 표준화 사업, 어플리케이션 지향, 그리드 인프라스트럭처 지향 등으로 나눌 수 있다. 여기에는 세계적으로 중요한 몇 개의 그리드 컴퓨팅 프로젝트를 간략하게 소개한다.

6.1 TeraGrid

미국은 5,300만 달러를 투자하여 TeraGrid를 구축하고 있다. 이 프로젝트는 NCSA(국립슈퍼컴퓨터응용연구소), SDSC(샌디에고 슈퍼컴퓨터센터)등 4개의 연구소를 40~80Gbps(처음에, 40Gbps까지)로 연결하여 13.6 Teraflops의 계산 능력을 실현하는 것으로서 그 스토리지는 450 Terabytes에 달하며 2003년 4월에 완성할 예정이다. 즉, TeraGrid는 여러 개의 초당 수십 기가비트 속도의 광(Optical) 네트워크 백플랜으로 연결된 일리노이와 캘리포니아의 주요 시설을 가진 국가 분산 고성능 기술 구조 (Fabric)라고 할 수 있다.

6.2 European DataGrid

지난 8월 (2002. 8.12)에 DataGrid 미들웨어의 버전 1.2가 전세계에서 가장 크고 가장 복잡한

인프라 중의 하나인 제품 테스트베드에 성공적으로 진출하였다. 유러피안 데이터 그리드는 유럽 공동체가 투자한 프로젝트로서 널리 분산된 과학 커뮤니티가 수백 테라바이트(Terabytes)에서 수 페타바이트(Petabytes)까지 대규모의 데이터베이스를 공유하여 집중적으로 계산하고 분석할 수 있는 차세대 컴퓨팅 인프라를 구축하는 프로젝트다.

스위스의 CERN과 이론 물리학자들이 중심이 된 그리드로서 일본의 TACC와도 공동 연구중이며 고성능 컴퓨팅, 병렬 컴퓨팅, 분산 컴퓨팅, 스토리지 등의 분야를 집중적으로 연구하고 있다. 스위스, 이태리, 영국, 일본 등 10개국에서 참여하고 있으며 특히 국제적인 테스트베드 인프라로 활용할 계획이다.

6.3 NASA IPG

NASA(National Aeronautics & Space Administration)의 IPG는 NASA의 과학기술 커뮤니티의 대규모 분산 자원을 사용하여 실제로 문제를 해결할 수 있는 계산 능력을 증가시키는 것이 목적이다. 이것은 각 NASA 센터의 자원을 배치하고 통합 관리하기 위해 인프라와 서비스를 개발하는데 초점을 둔다. IPG를 통해 NASA는 SDSC가 선도하는 NPACI (National Partnership for Advanced Computational Infrastructure)와 NCSA (National Center for Supercomputing Applications)가 선도하는 NCSA Alliance를 포함하여 여러 다른 국가 기관과 그리드 구축 작업에 참가하고 있다.

IPG를 구축하기 위한 전략은 첫째, 요구분석으로 어플리케이션 사용자들의 요구사항을 그리드의 서비스로 일반화시키는 것이다. 둘째, 그리드 서비스 중 이미 구현되어 있는 것과 개발해야 할 것을 구분하고 그리드 서비스의 타당성을 위해 대표적인 어플리케이션 작업을 구성한다. 셋째, 필요한 서비스와 인프라를 정의하고 구현한다.

6.4 ASCI

ASCI (Accelerated Strategic Computing Initiative) 계산 그리드 프로젝트는 SNL (Sandia National Lab.), LANL(Los Alamos National Lab.), LLNL(Lawrence Livermore National Lab.)을 중심으로 ASCI 프로젝트를 성공적으로 수행하기 위하여 구축된 프로젝트다. 이것은 3개 연구실에서 분산된 슈퍼컴퓨팅 자원을 이용할 수 있도록 메타 컴퓨팅 시스템을 구축하는 것이며 2010년까지 개발 완료할 예정이다.

이와 같은 학술적인 그리드 프로젝트 외에 상용화를 위한 그리드의 개발도 진행 중이다. 특히 대학과 연구기관을 중심으로 형성의 진로 예측, 외계인 생명체 찾기, 소수 찾기, 암이나 에이즈 정복, 인간 유전자 해독, 단백질 구조 분석, 신약 개발, 주가 추이 분석, 투자 위험 관리 등의 용도로 사용되고 있다. 미국 암 협회와 국립 암 연구재단, 영국 옥스퍼드 대학, 인텔 등은 백혈병 치료제를 개발하기 위한 프로젝트를 진행하고 있으며, 미국 암 연구 센터도 파라본과 함께 암 치료제 개발에 사용할 그리드를 구축했다.

또한 미국의 JP모건은 주가 추이를 예측하고, 투자 위험을 관리하기 위해 소규모의 그리드를 이용하고 있다. 데이터시냅스가 구축한 이 솔루션은 회사 내의 100여대 컴퓨터를 사용하고 있으며, 향후 정확도를 높이기 위해 컴퓨터의 수를 늘릴 계획이다.

한편 NASA는 그리드를 이용해서 외계의 생명체를 찾고 있다. SETI@home이라고 알려진 이 프로젝트는 전파망원경으로 수집된 데이터를 인터넷을 통해 분석하는 방식으로 이뤄진다. 현재 미국에는 엔트로피아, 데이터시냅스, 패러본컴퓨테이션, 유나이티드디바이스 등 10여개 업체가 슈퍼컴퓨팅 ASP 사업을 펴고 있다. 또한 국내에서도 글로벌 인터넷 비즈니스(GIB)가 이러한 사업을 시작할 계획이라고 밝히고 있다.

일본에도 일본 원자력 연구소를 중심으로 구 과학기술청 산하의 연구소 컴퓨터를 연결하는 ITBL(Information Technology Based Laboratory)이 있으며 SuperSINET으로 연결된 대형계산기센터에 Globus를 도입하고 있다.

아시아를 중심으로 하는 APGrid(Asia Pacific Grid)는 자원 공유, 그리드 기술개발, 협업 구축, 새로운 애플리케이션 개발에 상호 기술사용 등 상호 협조체제를 강화한다. 특히 APGrid의 주된 목표는 APGrid 테스트베드 구축이다. 현재 한국(KISTI), 일본, 싱가포르, 중국, 호주, 미국, 캐나다 등 15개국에서 25 기관이 참여하고 있고 그 참여 수는 계속 증가하고 있다.

그리드 컴퓨팅에서는 방대한 데이터와 애플리케이션을 네트워크를 통하여 처리하기 때문에 프로세서 파워보다 이를 효과적으로 운용할 수 있는 소프트웨어 인프라가 훨씬 더 중요하다. 따라서 현재 그리드 컴퓨팅에 관한 연구 또한 강력한 분산 처리 능력을 가진 소프트웨어 개발과 광 대역 네트워크 구축에 초점을 두고 있다.

7. 그리드의 표준화

그리드 컴퓨팅을 보급하기 위해서는 표준화가 상당히 중요하다. 지난 2월, 캐나다 토론토에서 개최된 GGF4에서 IBM과 Globus 프로젝트 그룹은 그리드와 웹 서비스를 통합한 OGSA(Open Grid Services Architecture) 모델을 발표하였다. 미국 알곤국립연구소를 중심으로 미 정부, 연구기관, IBM, 마이크로소프트, 시스코 시스템즈 등 많은 조직이 참여하는 글로벌 프로젝트는 공개 분산 컴퓨팅 구축에 필요한 프로토콜, 소프트웨어, 개념의 개발을 6년 가까이 계속하고 있다.

현재 대학 및 연구기관을 연결하는 그리드에서는 Globus라는 소프트웨어(미들웨어)가 사실상

표준으로 되어 있다. EuroGrid 등 몇몇을 제외하고는 모두 Globus를 이용하고 있다. Globus는 공개 소스이고 누구나 무료로 사용할 수 있다. 그러나 Globus는 기본적인 기능을 제공할 뿐이고 완전한 솔루션을 제공한다고는 할 수 없다. 아직까지 필요한 그리드 미들웨어의 각 요소들이 모두 개발되어 매끄럽게 동작하고 있지 못하기 때문이다.

이러한 상황에서 OGSA 모델 발표는 그리드 개발 연구에 가장 많은 활동을 하고 있는 GGF에서, 그리드 표준화에 많은 영향을 주고 있는 Globus 프로젝트 그룹과 웹 서비스 표준화를 주도하고 있는 IBM이 협력하여 발표했다는 점에서 많은 주목을 받고 있다.

실제 Sun, Microsoft 등에서도 OGSA 모델에 많은 관심을 가지고 협력과 투자 계획을 발표하였다. 따라서 OGSA의 각 구성 요소들이 바로 표준화될 가능성이 상당히 높으며, 그리드 각 구성 기술영역 워킹그룹, 연구그룹의 표준화 활동에 직접적인 영향을 줄 것으로 예상된다.

세계에 있는 컴퓨터를 하나의 그리드로 통일하려면 어떤 규격을 보급할 필요가 있다. 한가지 가능성은 Globus가 기업에도 정착되는 것이다. 인터넷은 본래 대학간을 연결하는 네트워크가 사실상 표준이 되고 상용으로 확대된 것이다. 마찬가지로 그리드에서도 현재 Global Grid Forum에서 진행하고 있는 표준화의 작업을 성공적으로 받아들이는 것이다.

그리드도 인터넷과 같이 세계의 모든 컴퓨터를 연결할 수 있는 그리드 네트워크로 성장될 것으로 기대된다.

8. 결 론

그리드 컴퓨팅이란 물리적으로 분산된 가상조직 상호간에 자원을 공유하고 문제를 해결하는 솔루션으로 네트워크에 연결된 많은 컴퓨팅 자원을 활

용하여 마치 하나의 컴퓨터를 사용하듯 대규모의 처리 능력을 실현하기 위한 아키텍처다. 한편, 그리드 환경은 웹서비스 환경으로 변화하고 있다. 웹서비스 기술과 그리드 기술을 통합한 것이 OGSA(Open Grid Service Architecture)이며, 이는 차세대 인터넷 모델이 될 것으로 보고 있다. OGSA에서는 그리드 서비스에 중점을 두고 있으며, 계산 자원, 저장 자원, 네트워크, 프로그램, 데이터베이스 모든 것들을 서비스로 표현한다.

코헨그룹은 제4차 GGF(Global Grid Forum)에서 인터넷 교통량이 그리드 상용화에 힘입어 앞으로 몇 년 동안은 2배씩, 2005년에서 2008년 사이는 1.5배, 2008년에 이르면 인터넷 교통량의 90%를 그리드 컴퓨팅에 바탕을 둔 P2P(Peer-to-Peer)나 S2S(Server-to-Server)가 차지할 것으로 예측하였다. 한편, 가트너에서는 2006년까지 과학기술 관련 계산을 수행하고 있는 기업의 약 5%가 그리드 컴퓨팅을 사용하게 될 것이라고 예측하고 있다. 앞으로 그리드 컴퓨팅은 현재의 그리드 컴퓨팅 이미지와는 상당히 다르게 변화될 것이고 그 적용 범위는 훨씬 광범위하게 확대될 것이다.

우리나라도 정보통신부의 기본계획 수립 하에 올해부터 향후 5년간 435억원을 투자하여 2005년까지 그리드 핵심 기술을 확보하기로 하였다. 금년에 40억 정도가 투입되어 서울대, KAIST, 포항공대, 표준과학연, 항우연 등 50여 기관이 국가 그리드 구축사업 및 그리드 미들웨어 기술 개발 등에 관한 협동연구를 활발하게 진행하고 있다.

그리드 컴퓨팅이라는 목표 지향은 모두가 인정하는 바이지만 컴퓨터 자원이 전기와 같이 사용할 수 있게 되기까지는 10년 정도의 기간이 필요할 것이고 그때가 되면 그리드 컴퓨팅은 오늘날과는 상당히 다른 모습으로 변화될 것이다. 차제에 우리나라의 그리드 컴퓨팅 개발 사업의 성공도 크게 기대해 볼만하다.

참고문헌

- [1] Ian Foster, Carl Kesselman, Steven Tuecke, "The Anatomy of the Grid", International J. Supercomputer Applications, 2001
- [2] Ian Foster, Carl Kesselman, Jeffrey M. Nick, Steven Tuecke, "The Physiology of the Grid", <http://www.globus.org/reseatch/papers/ogsa/pdf>
- [3] Carole GOBLE, David DE ROURE, "Semantic Web and Grid Computing", <http://www.semanticgrid.org/>
- [4] "Grid Computing Projects", IEEE distributed systems ONLINE
- [5] GREGOR VON LASZEWSKI AND IAN FOSTER, "GRID INFRASTRUCTURE TO SUPPORT SCIENCE PORTALS FOR LARGE SCALE INSTRUMENTS", Mathematics and Computer Science Division ANL.
- [6] "The Grid: A New Infrastructure for 21st Century Science", PHYSICS TODAY.org 2002. 2

저자약력



이춘희

1985년. 2월. 연세대학교 전자계산전공 (공학석사)

2000년. 2월. 아주대학교 컴퓨터공학과 (공학박사)

1968년 - 1996년 KIST/시스템공학연구소 책임연구원

1996년 - 1998년 ETRI 컴퓨터·소프트웨어연구소 책임연구원

2002년. 현재 한국과학기술정보연구원 전문연구위원

관심분야 : 분산컴퓨팅, 실시간컴퓨팅, 슈퍼컴퓨팅, 그리드
컴퓨팅

이메일 : chlee@hpcnet.ne.kr