



# 고성능연구망 인프라 활성화를 위한 Grid 구축

장 행 진 | E-mail: hjjang@hpcnet.net.kr  
KISTI 슈퍼컴퓨팅인프라활성화사업 추진단

- I. Grid 개요
- II. 고성능연구망과 Grid 관계
- III. 고성능연구망 중심의 Grid 어플리케이션
- IV. Grid System 분류
- V. Grid 구축을 위한 활동
- VI. 결 론

인터넷의 발생은 슈퍼컴퓨팅센터를 중심으로 구축된 고성능연구망으로부터 시작되었으며, 대부분의 선진국에서 추진중인 차세대 인터넷 프로젝트도 고성능컴퓨팅 자원 활용을 위해 구축된 고성능연구망을 중심으로 추진되었다.

## I. Grid 개요

Grid는 네트워크상에 지리학적으로 분산되어 있는 High-End Resource의 집합인 슈퍼컴퓨터, 정보저장소, 첨단 과학기술장비(가속기, 전자현미경, CAVE) 등의 고성능컴퓨팅 자원을 격자모양의 그물 형태로 상호 연동하여 조직과 지역을 벗어나서 어느 곳에서나 사용할 수 있게 하는 것을 말하며, 이를 또한 Super Internet이라고 부르기도 한다.

(그림 1)은 국가 슈퍼컴퓨팅센터를 중심으로 구축된 미국의 고성능연구망인 vBNS에 지역적으로 분산된 고성능컴퓨팅 자원을 연동하고 연구인들이 지역과 조직을 떠나서도 마치 전기 코드에 전기를 접속시키면

전기를 사용할 수 있는 것처럼 다양한 고성능컴퓨팅 자원을 고성능연구망 환경에서 사용할 수 있는 Computational Grid 모습이다.



그림 1 vBNS 중심의 Computational Grid

## II. 고성능연구망과 Grid 관계

Grid 구축과 고성능연구망과의 관계는 아주 밀접한 관계를 가지고 있으며 상호 의존적으로 발전을 추구하고 있다. 인터넷의 발생은 슈퍼컴퓨팅센터를 중심으로 구축된 고성능연구망으로부터 시작되었으며, 대부분의 선진국에서 추진중인 차세대 인터넷 프로젝트도 고성능컴퓨팅

자원 활용을 위해 구축된 고성능연구망을 중심으로 추진되었다. 또한 Grid 프로젝트도 차세대 인터넷 기반

## 연구전산망 개관



그림 2 캐나다의 C3s Grid 프로젝트

의 고성능컴퓨팅 자원의 활용을 위한 고성능연구망을 중심으로 추진되어지고 있다.

과거에 연구망은 단순히 학술연구 분야 종사자에게 정보전달 수단으로 활용되었으나, 현재의 인터넷보다 100~1000배 빠른 차세대 인터넷이 고성능연구망을 중심으로 발전하면서 고성능컴퓨팅 분야도 새로운 패러다임의 High-End Application 개발을 추구하고 있으며, 이를 활용한 Grid 인프라 구축이 활성화되고 있다.

특히, 고성능연구망이 Grid 구축의 중요한 요소로 등장하게 된 배경을 보면

- 첫째: 대부분의 고성능컴퓨팅 자원이 연구전산망에 연동되어 있음.
- 둘째: 고성능컴퓨팅과 연계된 가속기, 전자현미경, CAVE 등의 고성능 장비들이 차세대 인터넷 기술을 응용하여 원격지에서 이용이 가능함.
- 셋째: 새로운 패러다임의 어플리케이션 및 개발 능력을 보유한 고급 인력이 고성능연구망

을 중심으로 포진되어 있음.

- 넷째: 연구전산망에 접속된 고급인력이 이러한 새로운 패러다임의 응용을 필요로 함.

현재, 주요 국가의 고성능연구망 중심으로 추진중인 Grid 프로젝트 현황을 보면 대부분 국가를 대표하는 고성능연구망을 기반으로 Grid 개발을 추진하고 있다.

특히, 캐나다의 경우 지난 수년간 차세대 인터넷 프로젝트를 성공적으로 수행하기 위하여 주로 네트워크 인프라 구축 측면에서 정부 차원의 부재가 이루어졌으나 최근에 와서 이러한 인프라를 중심으로 (표 1)과 같이 국가 보유 컴퓨팅 능력의 중요성을 인식하고 개인, 대학이나 산업체의 C3(Creativity, Connectiveness, and Wealth Creation) 능력을 제공하기 위하여 (그림 2)와 같이 C3.CA 프로젝트를 수행하고 있다.

"C3.ca recognizes that Canada needs capacity and capability computing; therefore, C3.ca will lead in the specification and development of grid technologies to address capacity computing for



Canada C3.ca will define needs to achieve capability computing."

(표 1) 주요 국가의 Grid 프로젝트

국 기	고성능연구망	Grid 프로젝트	중심기관
미 국	VBNS	National Technology/ Grid	NCSA
이탈리아	INFN	Data Intensive Grid	원자력연구소
캐나다	CANet3	C3's Grid	C3.ca
일 본	SINet	National Test-bed Grid	TACC

### III. 고성능연구망 중심의 Grid 어플리케이션

#### 1. Grid 어플리케이션

##### • Distributed Supercomputing

하나의 시스템으로 해결되지 않는 문제들을 해결하기 위해서 지역적으로 분산되어 있는 컴퓨팅 자원들을 고성능망을 중심으로 연동하여 해결함. 여러 개의 컴퓨팅 자원들의 이용을 조화시키는 도구들을 이용하여 컴퓨팅 문제 해결.

- 분야 : DIS(Distributed interactive simulation), Stellar dynamics 등.

##### • Data-Intensive Computing

지리적으로 분산된 저장장보, 디지털 라이브러리, 데이터베이스 등 저장된 정보들을 합성해서 이용하는 것에 중점을 둠.

- 분야 : High-Energy Physics, Digital Sky Survey(astronomical photographic data) Meteorological Forecasting System 등.

##### • High-Throughput Computing

고성능망 기반을 중심으로 유류 컴퓨터프로세서 자원들을 모아서 독립적이거나 상호 연관성이 적은 대량의 작업들에 할당해서 스케줄링함으로써, 단일 컴퓨터

로는 처리하기 힘든 모험과제나 실험과제를 대적함으로써 미래의 기반기술 개발에 활용.

- 분야 : Parameter Studies, Cryptographic Problems 등.

##### • On-Demand Computing

컴퓨팅, 소프트웨어, 데이터 저장소와 같은 자원들을 일시적으로 필요로 할 때 단기간 동안만 이용함으로써 국가보유 자원의 이용을 극대화하기 위한, 고성능망을 이용하여 자원 이용의 비용이 고가이고 지역적으로 근접해 있지 않은 경우에도 효율적으로 이용 가능한 분야.

- 분야 : Medical Instrumentation, Network-Enabled Solver, Cloud Detection 등.

##### • Collaborative Computing

사용자 디스플레이 환경에서 멀리 떨어져 있는 데이터 집합, 시뮬레이션, 제너레이터 등을 고품질의 네트워크상에서 가상 환경 시스템을 통하여 협동연구를 가능케 함.

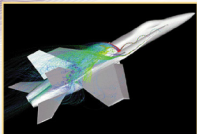
- 분야 : Interactive Scientific Visualization, Education/ Training/Scenario Simulations, Art and Entertainment, Industrial Design 등.

#### 2. 주요 Grid 어플리케이션 개발 사례

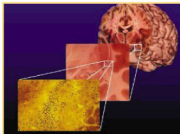
(그림 3)은 NASA의 IPG(Information Power Grid)에서 F18 전투기의 와류(vortex) 가동 시뮬레이션을 비정상 유동해석 Toolkit(UFAT)과 거대규모 비정상 유동 datasets에 사용한 예다. 이러한 시뮬레이션은 제한된 메모리와 소형 워크스테이션 환경에서는 이전에는 할 수 없었던 가시화 작업을 Grid상에서 가능케 하고 있다. 이를 위하여 Origin2000 SMP 시스템 64Node까지 두 시스템에 Grid 환경을 구축하여 활용하였다. 이러한 Grid의 개발을 위하여 고성능망으로 NASA의 NREN을 활용하고 있다.

SDSC 슈퍼컴퓨팅센터에서는 NPACI 프로젝트 차원에서 Globus 기반의 Metasystems 구축, Data-

## 연구전산망 개관



〈그림 3〉 PG의 F18 전투기 워류 영상



〈그림 4〉 Refining and Linking Brain Data

Intensive Computing, Neuroscience, Earth Systems Science, Engineering Grid 프로젝트를 수행하고 있는데, 그림 4은 Data-Intensive Computing 측면에서 UCLA와 Washington University에 있는 인간의 뇌세포 데이터를 분석하고 통합하여 3D 기술을 이용하여 가시화한 이미지다. 특히 이를 수행하기 위하여 vBNS상에서 UCLA에서 Washington University까지 45Mbps의 전송 속도로 연동되어 다양한 통합작업 및 외과 수술적인 작업을 수행함으로써 프로젝트 수행이 가능하였다.

이와 같이 Grid 구축 요소인 컴퓨터와 컴퓨터, 컴퓨터와 고성능 장비, 사람과 사람 및 고성능 장비들을 그물처럼 연동하고, 이를 중심으로 고성능 어플리케이션을 개발하기 위하여서는 고품질의 네트워크 환경과 새로운 패러다임의 데이터베이스 및 인적 자원이 필요하다.

그래서 선진국에서는 이러한 요구를 충족시켜 줄 수 있는 네트워크 환경은 고성능연구망 환경이 적합하다고 판단하고 이를 중심으로 Grid 구축을 발전시켜 나가고 있는 추세에 있다.

### IV. Grid System 분류

#### 1. Computational Grid

Grid는 크게 Computational Grid와 Data Intensive Grid 그리고 Access Grid로 아래 그림과 같이 분류할 수 있는데, Computational Grid는 분산된 컴퓨팅 자원을 한 대의 고성능 컴퓨터처럼 사용할 수 있게 하는 것이고, Data Intensive Grid는 대용량의 정보를 생산하는 장비들과 고성능컴퓨터 및 사용자를 중심으로 구성된 것을 말한다. 아울러 Access Grid는 Computational Grid나 Data Intensive Grid에서 생성된 정보를 원거리에서 떨어져 있는 사람들과 Group 간의 공유할 수 있는 협업 연구 환경을 제공하는 것이다.

특히 Computational Grid에 속하는 시스템들은 고성능컴퓨팅을 위해서 많은 자원들을 연결하여 해결할 수 있도록 하는 것을 말한다. 현재 구축중인 Grid 프로젝트들은 대부분이 Computational Grid에 속한다. 이

(Grid System 분류)





들 중에서도 분산 슈퍼컴퓨팅 권인 Grid는 작업의 전체 수행시간을 줄이기 위해서 많은 고성능컴퓨터들을 동시에 이용하는 것을 말하며, 주로 초대형 문제들을 해결하기 위해서 사용된다. 반면에 High-Throughput 컴퓨팅은 주어진 시간 안에 가능한 많은 작업을 처리하는 것을 목표로 한다.

아울러 Computational Grid를 구축하는 데 있어서 가장 중요한 것은 미들웨어인데, 많은 대학과 연구기관에서 미들웨어를 개발하고 있으며 그 중에서 대부분이 Globus Metacomputing Toolkit(GMT)라는 미들웨어를 사용한다. Globus 기술은 분산된 이기종 컴퓨팅 자원들을 하나의 가상 슈퍼컴퓨터처럼 사용할 수 있게 하기 위한 미들웨어이다. 이것은 Grid를 위해 요구되는 서비스와 기능들을 제공하기 위한 도구들이다. 이들 서비스는 보안, 자원관리, 자원할당, 통신 등으로 이루어져 있다. Grid는 다양한 형태의 어플리케이션들과 프로그래밍 패러다임들을 지원하며, GMT는 불 개발자나 어플리케이션 개발자가 자기에게 맞는 서비스를 쉽게 찾을 수 있도록 서비스들의 볼로써 지원한다.

이런 방법을 성공시키기 위해서 GMT 서비스의 기능을 구분하였고 API들을 제공하며, 다음과 같은 요소 시스템들로 구성되어 있다.

- GRAM: 자원할당 및 프로세스들을 관리하는 도구
- Nexus: 서로 다른 통신 시스템간에 통신할 수 있게 해 주는 통신 서비스
- GSI: 사용자 인증을 위한 보안 서비스
- MDS: 분산된 자원들의 연결상태와 자원에 관한 정보 제공
- HBM: 시스템 요소들 및 프로세스들의 동작 상태를 알려주는 도구
- GASS: 병렬 또는 직렬 인터페이스를 이용하여 원

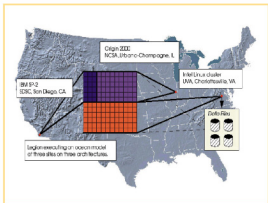


그림 5) vENS실험망의 대략 모양의 그래프

- 격지 데이터들 접근할 수 있게 해 주는 도구
- GEM: 실행 파일들을 만들고 찾아주는 도구
- GARA: 자원들을 미리 예약하고 필요할 때 할당해 주는 도구

위의 <그림 5> 미국의 고성능연구망인 vENS에 연동된 SDSC 및 NCSA를 중심으로 대양 모델링을 위하여 구축된 대표적인 Computational Grid 시스템의 하나며, 슈퍼컴퓨터와 슈퍼컴퓨터를 연동하여 작업을 분산 처리하기 위하여서는 고성능연구망 환경의 지속적인 네트워크 대역폭을 필요로 하는 분야 중의 하나다.

## 2. Data Intensive Grid

Data Intensive Grid는 원격지에 분산된 자료들을 통합하여 분석할 수 있게 해 주기 위한 시스템이다. 대표적 사용분야는 자료가 분산된 여러 곳에서 생성되고 한곳에서 처리하는 형태의 Data Mining과 한곳에서 처리된 자료를 분산된 사이트에서 처리하는 형태로 분류할 수 있는데, 현재 유럽을 중심으로 추진중인 Data Grid는 후자에 속하며, 미국을 중심으로 추진중인 Data Mining은 SDSC 등 NPACI 프로젝트에 참

## 연구전산망 개관

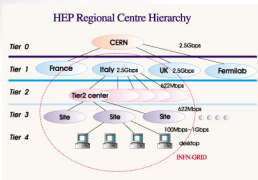
여러는 협력기관 중심으로 개발되고 있다. <그림 6>은 스위스 CERN에서 생성된 이론물리학 분야의 대용량 정보를 이탈리아로 가져와서 이탈리아의 고성능 연구망인 INFN을 통하여 각 사이트에 전송, 이론물리학자들이 연구를 수행하는

대표적인 사례 중의 하나다. 향후 CERN을 중심으로 구축된 Data Intensive Grid는 전세계 10여개 국가에 30개 슈퍼컴퓨팅 시스템을 대상으로 GMT를 설치하고 CERN에서 생성된 수십 테라바이트의 데이터를 분산처리하기 위한 Global Metacomputing Grid를 구축할 예정이다.

### 3. Service Grid

Service Grid는 On Demand Grid, Access Grid, Multimedia로 분류할 수 있는데, On Demand Grid는 사용자 작업이 요구에 따라 동적으로 자원들을 연결하는 것이 가능한 Grid다. 사용자는 작업 처리중에 결과를 좀더 정밀하게 보기 위해 더 많은 자원을 요구할 수 있다. Access Grid는 사용자들이 원격지에서 협력하며 일을 할 수 있게 하는 하나의 작업공간을 제공한다. 멀티미디어 Grid는 실시간 멀티미디어 어플리케이션을 위한 인프라를 제공한다. 멀티미디어 Grid는 다양한 스트림이 동시에 교환되기 때문에 QoS가 보장되어야 한다.

대부분의 Grid가 추구하는 고성능컴퓨팅 자원은 지역적으로 분산되어 있으며 사용자 그룹 및 어플리케이션



[그림 6] CERN Data intensive Grid

도 다양화되고 분산되어 있다. 특히 인터넷 및 네트워크 엔지니어링 기술의 발달로 원거리에서 실시간으로 슈퍼컴퓨터에서 해석된 결과를 과학적 가치의 장비나 소프트웨어를 활용하여 가상공간에서 사용자 그룹간에 협업

연구나 정보를 공유할 수 있는 Access Grid는 슈퍼컴퓨팅 응용 환경을 중심으로 구축하고 있는 추세다.

현재 Access Grid는 미국 NSF가 지원하는 슈퍼컴퓨팅 및 정보 인프라 National Technology Grid를 효율적으로 관리 지원 할 목적으로 슈퍼컴퓨팅 및 멀티캐스팅 기술과 고성능 망기술 등을 이용하여 그룹간 영상회의, 교육·훈련, 슈퍼컴퓨터 사용자 지원 등 협업연구 환경을 제공할 목적으로 구축하고 있으며, 이러한 인프라 및 관련 기술이 전세계적으로 확산되고 있는 추세다. 특히 유럽의 경우 이와 유사한 형태의 협업연구 환경을 국가 슈퍼컴퓨팅센터들 중심으로 구축하고 국가나 지역간의 슈퍼컴퓨터 사용자 그룹을 중심으로 협업연구 환경을 구축하고 연구시간 단축 및 기술개발의 최적화로 미국의 항공우주산업을 따라잡은 성공 사례가 주목된다.

(그림 7)에서 보는 바와 같이 원거리 Access Grid는 실시간으로 슈퍼컴퓨터에서 해석된 결과를 연구 그룹이나 연구원들이 정보를 공유하고 협업 연구함으로써 연구기간의 단축 및 최적화로 빠른 시일에 새로운 연구 결과를 창출할 수 있다.

### V. Grid 구축을 위한 활동



(그림 8)에서 보는 바와 같이 Grid 프로젝트는 미국을 중심으로 '98년도부터 시작되어 급속히 발전, 유럽 중심의 EGrid와 일본 중심의 AFGrid로 활성화되고 있으며 캐나다는 C3.ca Grid를 추진중이다. 특히 캐나다는 차세대 인터넷 발전이 HPC 발전에 달려 있다고 판단하고 카나리안과 협력하여 그 동안의 구축된 차세대 인터넷 인프라를 중심으로 Grid 프로젝트를 활성화시키고 있다.

아울러, Grid 구축 기술을 개발하고 표준화하기 위하여 미국을 중심으로 Gridforum이 구성되어 지난 10월까지 5회 개최되었으며 내년 2월부터는 Gridforum과 EGridforum, AFGrid를 통합하여 아래의 같이 Global Gridforum을 발족할 예정이다.



(그림 7) Access Grid



(그림 8) Global Gridforum

## VI. 결 론

향후 Grid 구축은 많은 HPC 분야에 적용되어 Global Grid 인프라로 발전할 것으로 예상됨으로 우리 나라도 정부 차원에서 공공 슈퍼컴퓨팅 센터를 중심으로 국가 메타컴퓨팅 Grid 구축이 시급하다. 특히, Grid 구축 기술을 확보하기 위하여 연구전망을 고도화하고 이를 중심으로 선별적으로 Grid 어플리케이션을 확장하여 나가는 것이 바람직하다

고 생각되며, 아울러 대외적으로 국제 협력 프로젝트를 추진하여 현재 정부에서 구축하여 놓은 APPI Test-bed나 향후 구축될 TEIN(Trans-Eurasia Information Networks)를 활성화함으로써 국가 고성능컴퓨팅 자원을 중심으로 한 Grid 구축 활성화가 가능하다고 생각된다. <end>

### ■ 참고문헌

- (1) <http://www.gridforum.org/>
- (2) <http://www.globus.org/>
- (3) <http://gridweb.com.chy/grid/>
- (4) The Grid: Blueprint for a New Computing Infrastructure
- (5) <http://www.eagrid.org>
- (6) <http://www.c3.ca/>