

# 그리드 웹 포탈 동향 분석

김성준\*, 김중권\*, 이상산\*

\*한국과학기술정보연구원 슈퍼컴퓨팅센터

e-mail:sikim@hpcnet.ne.kr

## A Analysis on Grid Web Portal

Sung Jun Kim\*, Joong Kwon Kim\*, Sang San Lee\*

\*Supercomputing Center,

Korea Institute of Science & Technology Information

### 요 약

분산되어 있는 고성능 컴퓨팅 자원이나 저장 장치, 과학계산 장비들을 한데 묶어서 활용함으로써, 자원 활용의 효율성을 극대화하기 위한 그리드 관련 기술이 현재 많은 호응을 얻으면서 관련 연구가 한창 진행중에 있다. 이러한 그리드 환경을 보다 쉽게 활용하기 위한 그리드 웹 포탈의 연구도 같이 진행되고 있는데, 본 고에서는 이미 그리드 웹 포탈서비스를 구축·서비스를 하고 있는 사이트들의 동향을 분석하였다.

### 1. 서론

컴퓨터 및 네트워크의 성능이 비약적으로 향상됨에 따라 이들 자원을 공유하여 효율적으로 사용하고 자 하는 분산 시스템에 대한 연구가 활발히 논의되고 있다. 또한 인터넷의 발전속도가 초기의 예상을 뛰어넘어 급격히 신장하고 있다. 그리드(Grid) 환경은 이러한 배경을 바탕으로 분산 시스템에서 한 걸음 더 나아가 전 세계에 흩어져 있는 컴퓨터 자원을 인터넷망을 통하여 하나의 단일 컴퓨터처럼 활용할 수 있도록 하자는 개념으로 출발하였다[1].

그리드 환경에서 사용자는 개개의 컴퓨팅 자원을 사용하는데 있어서 한번의 컴퓨팅 자원에 대한 로그인 후에 다른 지역의 컴퓨팅 자원에 접근할 때 추가적인 인증 과정을 거치지 않는 단일 인증(Single Sign-on)과 같은 인증서비스, 실제 사용자가 컴퓨팅 자원을 사용할 수 있는지에 대한 권한이 관련된 접근제어(Access Control) 서비스 등과 같은 보안문제에 대한 기반이 뒷받침되어야 효과적이고 안전하게 계산을 수행할 수 있다[2].

이러한 그리드 환경은 사용자에게 많은 자원을 활용할 수 있는 기회를 제공한다. 하지만, 이러한 많은 자원은 한 두가지 자원을 활용하던 사용자에게 자칫 너무 많은 접근점(access point)을 제공함으로써 사용자에게 많은 자원에 대한 개별적인 정보를 사전에

익혀야 되는 부담을 줄 수도 있다. 이러한 이유로 등장한 기술이 그리드 웹 포탈이다.

그리드 웹 포탈이라는 것은 그리드의 여러 자원들을 웹에서 접근할 수 있도록 하는 것이다. 단순한 접근뿐만 아니라, 그리드 자원들의 상세 정보나 특성들을 제공함으로써 하나의 접근점을 통해서 사용자들이 모든 작업을 수행하도록 할 수 있다.

### 2. 관련연구

#### 1) 그리드

인터넷이 보편화되고 컴퓨터 및 네트워크의 성능이 향상되면서 정보의 공유가 보편화되었다. 그 중 WWW를 통해서 인터넷에 존재하는 정보를 본인이 쉽게 사용할 수 있는 기술은 많은 사람들을 흥분시켰고 현재 없어서는 안될 하나의 정보공유 수단으로 자리잡게 되었다.

이러한 WWW의 성공으로 인하여 네트워크로 연결될 수 있는 모든 것을 공유하고자 하는 노력이 진행되어 왔으며 소규모의 지역에서의 cluster 와 워크스테이션을 이용해 병렬처리를 시도한 것에서 벗어나 광대역 통신망을 활용한 이기종의 컴퓨팅 자원을 이용하려는 노력의 일환으로서 'Grid Computing'이 대두되었다.

그리드는 지리적으로 그리고 조직 면에서 흩어져

있는 이종의 정보 자원들을 서로 연결시켜 하나의 가상 컴퓨터처럼 동작하는 체제이다. 고성능 슈퍼컴퓨터부터 개인용 컴퓨터까지 광범하게 이용할 수 있는 컴퓨팅 자원과 수십기가에서 수백테라까지 이용할 수 있는 저장장치, 세계 각지의 연구인력 및 다양한 통신 시스템등 현재 분산된 위와 같은 자원들을 하나의 동일한 인터페이스를 통해 접근하도록 하여 체계적으로 협업시스템을 만들고 고성능의 계산 능력과 데이터 처리를 가능하게 할 수 있는 기술이 그리드이다.

그리드라는 용어는 1990년 중반 미국의 슈퍼컴퓨팅센터를 중심으로 고성능의 분산 컴퓨팅 인프라를 구축하는데 비롯하였고, 고성능의 자원, 대용량 정보 및 혁신적인 애플리케이션에 초점이 맞추어진 것이 일반 분산 컴퓨팅과 구별된다[3].

이와 같이 그리드는 대량의 정보와 고성능 컴퓨팅 자원을 효율적으로 이용하기 위하여 차세대 인터넷을 기반으로 구축되고 있으며, 새로운 응용연구의 수행에 큰 변화를 가져올 것으로 기대를 하고 있다.

그리드는 활용되는 애플리케이션의 특징에 따라 크게 다음의 3가지로 나눌 수 있다[4].

(1) 계산 그리드

- 계산 그리드란 많은 양의 계산을 위해 컴퓨팅 자원을 연결하는 환경을 말한다. 특히 분산컴퓨팅 관련 그리드는 작업의 전체시간을 줄이기 위해 가능한 많은 컴퓨터를 동시에 이용하는 것을 의미하며, 보통 기계항공분야와 물리학, 천문학 등에 알맞은 환경이다.

(2) 데이터 그리드

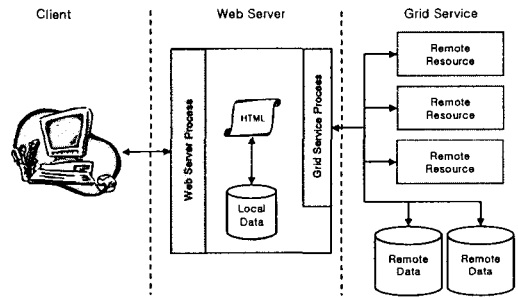
- 데이터 그리드란 여러 곳에 산재되어 있는 분산된 자료를 통합, 관리, 분석 할 수 있는 환경을 말합니다. 보통 처리되는 데이터 양은 수백 테라바이트에서 페타바이트 정도이며, 보통 고 에너지 물리와 유전자 정보처리 분야에 알맞은 환경이다

(3) 액세스 그리드

- 액세스 그리드란 분산처리를 필요로 하는 애플리케이션을 위한 그리드이다. 사용자의 자원요구를 동적으로 할당하고, 원격지의 사람과 같이 협력하여 일을 하며, QoS가 보장되는 분야에 알맞은 환경이다[5].

2) 그리드 웹 포탈

분산된 자원들을 보다 효율적으로 사용하기 위한 그리드 기술이 개발되고 있으나, 사용자 입장에서 분산된 자원의 위치와 현재 상태와 같은 각종 정보를 기존의 텍스트 기반 환경에서는 효율적으로 활용할 수가 없었다. 이와 같은 이유로 제기된 것이 웹을 이용한 그리드 서비스의 제공이었고, 그리드 포탈이라는 웹 기반 그리드 서비스가 탄생되었다. 그리드 포탈은 기존의 웹 기반 기술과 그리드 기술을 접목시킨 것으로 사용자 인터페이스 부분은 서블릿과 같은 웹 기술을 사용하고, 실제 분산된 자원의 활용과 같은 그리드 관련 기술은 그리드 미들웨어와의 통신을 통하여 사용자에게 그리드 서비스를 제공하는 것을 말한다. 그리드 웹 포탈의 전체적인 동작 구조가 [그림 1]에 나타나 있다[6].



[그림 1] 그리드 웹 포탈 구조

사용자는 웹을 통하여 그리드 포탈에 접근을 하게 되고, 웹 환경에서 각종 작업을 실행함으로써 사용자들의 편리성을 높이고, 각 시스템별 어플리케이션 정보나 시스템 정보, 작업관련 문의 및 사용자간 정보교환과 같은 유용한 정보를 단일 접근점을 통하여 사용자를 지원하는데 목적을 두고 있다고 할 수 있다. 웹 포탈 구성시에 적용되는 일반적인 관련 기술들이 [표 1]에 나열되어 있다.

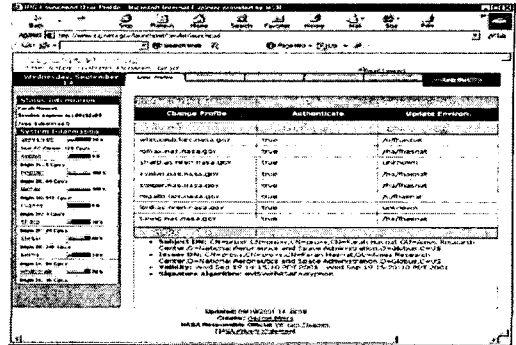
웹 서버	· Netscape (http://www.netscape.com) · Apache (http://www.apache.org)	
운영체제	Windows, Unix, Macintosh, Linux, etc	
프로그래밍 언어	Server	Java, Javascript, PHP, Perl 등
	Client	HTML, javascript
	Protocol	HTTP/CGI/ Servlets/ Applets
보안	· HTTPS, SSL, Encryption · Cookies · Certificate	

[표 2] 웹서버 기술

이러한 그리드 웹 포털을 여러 그리드 관련 프로젝트에서 구축하고 있으며, 구축하기 위한 패키지들을 공개해 놓은 개발 툴들도 많이 있다. 다음은 현재 발표 되어있는 포털 개발툴들이다.

- GridPort Toolkit
- GPDK(Grid Portal Development Kit)
- GridSphere
- Gateway
- CCA

위의 툴들 중에서 GridPort Toolkit과 GPDK가 가장 널리 사용 중에 있으며, NASA와 PACI에서 이를 이용한 포털 사이트를 구축하였다.



[그림 2] IPG Launch Pad 화면

### 3. NASA IPG

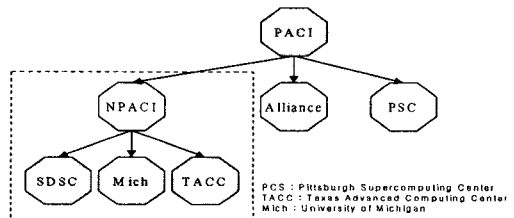
Information Power Grid(IPG)는 NASA의 고성능 계산 그리드이다. 이들의 자원들은 지역적으로 분산된 다양한 기관에 의해서 운영되고 많은 기관들의 연구자들에 의해 공유된다. IPG는 NASA Ames, NASA Glenn, NASA Langly Research Center와의 협업을 제공한다. IPG를 통하여 NASA는 NCSA(National Center for Supercomputing Applications)가 이끄는 NCSA Alliance와 SDSC(San Diego Supercomputing Center)가 이끄는 NPACI(National Partnership for Advanced Computational Infrastructure)와 같은 그리드 환경을 구축하고 있는 다른 몇 개의 국가 기관들과 연결된다[7].

IPG에서는 계산 과학자들, 연구원들과 고성능 컴퓨터 사용자가 다양한 자원에 보다 쉽게 접근할 수 있도록 Launch Pad라는 웹 포털을 구축하였다. 이것은 일반 사용자들이 가장 익숙한 웹브라우저와 같은 형태의 인터페이스를 이용하여 그리드에 접근하는 것을 보다 용이하게 해준다. 또한, 원격 자원에 대하여 안전한 사용자 인증을 수행하고 사용자로 하여금 원격지의 자원 정보를 확인함으로써 자원 활용에 있어서 편리함을 제공하고 있다. 추가적으로, 사용자의 프로파일을 생성함으로써, 수행하고 있는 작업을 모니터링 하거나 결과를 확인 할 수 있는 기능을 제공한다. 아래 [그림 2]는 IPG Launch Pad의 접속 화면이다. 아래 그림에서 보는 바와 같이 각 시스템의 현재 부하량이 표시되고 있다. 사용자는 부하가 적은 시스템에 작업을 할당할 수 있다.

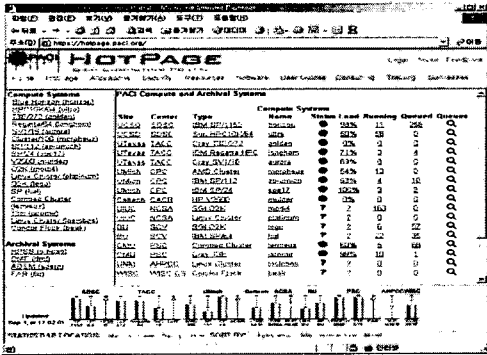
### 4. PACI Hotpage

HotPage는 PACI 계산 그리드의 각 자원들에 대한 정보를 연구자들이 검색이 용이하도록 구성된 사이트이다. HotPage는 기술문서, 동작 상태, 시스템의 부하와 현재 사용량, 큐의 작업등을 포함하고 있다. 또한 HotPage 계정을 가지고 있는 연구자들이 Hotpage를 이용하여 그들의 파일과 데이터에 접근 및 처리하는 것과 작업을 실행, 감시 그리고 삭제하는 것을 가능하게 해준다[8].

Hotpage는 과학 계산 처리를 보다 쉽게 하기 위하여 설계되었다. Hotpage를 통하여 사용자 매뉴얼과 각종 기술문서들을 쉽게 접근 가능하다. 또한 Hotpage를 사용하면, 시뮬레이션을 실행할 때, 작업 또는 큐의 상태를 모니터링 함으로써 어떤 시스템에서 작업을 실행시키는 것이 가장 빠르게 결과를 받아 볼 수 있는지를 결정할 수 있다. Hotpage에서는 파일을 업로드하여 편집하고 해당 작업을 임의의 시스템에서 실행시킬 수 있으며, 실행의 결과를 다운로드 하거나 저장장치에 저장할 수도 있다. 다음 [그림 3]은 PACI Hotpage의 구성도이다. 그림에서 보여지는 바와 같이 PACI hotpage는 NPACI의 hotpage를 포함하며, NPACI hotpage를 기본으로 하여 확장한 그리드 포털이라고 할 수 있다.



[그림 3] PACI Hotpage 구성도



[그림 4] Hotpage 화면

위의 [그림 4]는 Hotpage의 접속화면을 보여주고 있다. 그림에서 각 시스템별로 시스템의 상태, 부하율, 실행되는 작업의 개수와 대기작업의 개수 등 자세한 정보가 표시되고 있다. 사용자는 이 정보 등을 참고하여 자신이 원하는 작업을 수행할 수 있다.

5. 비교 분석

아래 [표 2]는 IPG Launch Pad와 PACI Hotpage의 구성 및 서비스 내용을 비교한 것이다. 두 개의 포탈들은 구성기관의 성격이 서로 상이하기 때문에 어느 한쪽이 반드시 좋은 것이라고는 할 수 없다. 앞에서 설명한 바와 같이 IPG는 일반 연구원들이 아닌 NASA 산하의 연구원들이 동일 기관내에서 서로 지리적으로 떨어져 있는 여러 시스템의 자원들을 활용하는 데 목적을 두고 있는 반면에, PACI Hotpage는 PACI 프로그램에 의해 지원을 받는 여러 학교와 연구소들의 자원을 다른 연구자들에게 시스템을 보다 쉽고 편리하게 활용할 수 있도록 구축되었다.

	IPG Launch Pad	PACI Hotpage
개발툴	GSDK	GridPort Toolkit
Protocol	HTTP	HTTPS
작업 실행	O	O
작업 모니터링	O	O
파일 전송	O	O
파일 업로드	O	O
작업 파일 수정	X	O
시스템 정보(Load)	O	O
큐 상태 정보	X	O
구성 기관	NASA 산하 연구소	PACI 프로그램 참가기관들
구성 시스템	9개	22개

[표 3] IPG Launch Pad와 PACI Hotpage 비교

6. 결론

그리드 관련 기술은 계속 발전되고 있으며, 제한된 자원을 보다 효율적으로 활용할 수 있는 기술이라는 점에서 많은 호응을 얻고 있는 상태이다. 다만, 단일 자원이 아닌 여러 자원을 사용한다는 점에서 자칫하면 사용자들에게 사용의 부담을 줄 수 있기 때문에, 웹 포탈이라는 방법을 통해서 보다 쉽게 사용자들이 공유 자원을 활용할 수 있다.

본 논문에서는 현재 구축되어 있는 서로 다른 그리드 웹 포탈을 비교해 보았다. 이러한 비교를 통해서 현재 그리드 웹 포탈 기술이 어디까지 발전되고 있는지를 알 수 있다. 이를 통해, 향후 그리드 포탈을 구축시에 참고할 수 있는 기준을 정할 수 있으며, 보다 사용자에게 편리한 시스템을 구축하는데 도움이 될 것이다.

참고문헌

- [1] 윤찬현, 심은보, "그리드 구조 및 연구동향", 『한국정보과학회 정보과학회지』 제 20권 2호, 2002. pp.11-15
- [2] Randy Butler Von Welch, Douglas Engert, Ian Foster, Steven Tuecke, John Volmer, Carl Kesselman, "A National-Scale Authentication Infrastructure", IEEE, December. 2000. pp.60-66
- [3] Ian Foster, Carl Kesselman, Gene Tsudik, Steven Tuecke, "A Security Architecture for Computational Grids", 5th ACM Conference on Computer and Communications Security. 2000.
- [4] 김동균, 이필우, 황일선, "그리드 컴퓨팅" 『한국정보과학회 정보과학회지』 제 20권 2호, 2002. pp.5-7
- [5] 강 경우, 박형우, "그리드 연구개발 동향" 『한국정보과학회 정보과학회지』 제 20권 2호, 2002. pp.23-33
- [6] <http://sizzle.kookmin.ac.kr/research/gridcom.htm>
- [7] [http://www.ipg.nasa.gov/aboutipg/aboutipg\\_index.htm](http://www.ipg.nasa.gov/aboutipg/aboutipg_index.htm)
- [8] <https://hotpage.paci.org/about/about.html>