
MDS를 위한 클러스터 시스템의 PBS 기반 정보 제공자 개발

Development of an PBS-based Information Provider of Cluster System for MDS

조광문

목포대학교 경제통상학부 전자상거래학 전공

Kwang-Moon Cho(ckmoon@mokpo.ac.kr)

요약

최근 그리드 환경 같은 고성능 컴퓨팅 분야에서 널리 보급되고 있는 컴퓨팅 환경은 클러스터 시스템이다. 그러나 그리드 환경에서 클러스터 시스템의 주요한 정보들이 다른 시스템들에게 적절하게 제공되지 못하고 있다. 본 논문에서는 이러한 정보들을 제공하는 체계를 제안하였다. 그리드 환경에서 클러스터 시스템의 정보를 제공한다. 이러한 배경하에서 그리드 정보 서비스의 활용 가능성이 증대될 수 있다.

■ 중심어 : | 그리드 | 정보 제공자 | 클러스터 시스템 | MDS | PBS |

Abstract

The pervasively proliferated computing environment in recent is a cluster system. This system is used in high performance computing area as a Grid system. However, the important information of cluster system in Grid environment does not provided to other systems. In this paper the scheme to provide these information is proposed. The proposed scheme provides the information of cluster system in the Grid environment. Based on this scheme the utilization possibilities of Grid information services are increased.

■ Keyword : | Grid | Information Provider | Cluster System | MDS | PBS |

1. 서론

그리드 환경에서 그리드 정보 서비스는 반드시 필요한 부분이다. 그리드 정보 서비스는 그리드 환경에서 사용 가능한 자원들에 대한 가장 최근의 정보를 사용자나 다른 구성 요소 시스템으로 제공한다. 이는 인터넷에서 DNS(domain name system)의 역할과 유사하다고 할 수 있다. 사용자들은 자신이 속해 있는 가상 조직(VO: virtual organization) 내에서 사용 가능한 자원

(resources)의 정보를 쉽게 찾고 사용할 수 있기를 원하고 있다. 이러한 정보를 제공하기 위한 정보 제공자가 필요하다. 정보 제공자는 자원들에 대한 정보를 수집하고 분류 및 정리한다. 정보 제공자는 이 정보를 사용자나 다른 미들웨어에 제공한다.

자원들은 그리드 환경 내에 분산되어 있고 각 자원마다 관리 방법이 다르기 때문에 각 자원들에 대한 정보의 상태는 자주 변하게 된다. 네트워크의 속도, 각 시스템의 부하, 가용 디스크 공간, 각 CPU의 가용성 등에 대한 정

* 본 논문은 한국과학재단 지역대학우수과학자지원(과제번호: R05-2002-000-01311-0)으로 수행되었습니다.

접수번호 : #041029-001

접수일자 : 2004년 10월 29일

심사완료일 : 2004년 12월 10일

교신저자 : 조광문, e-mail : ckmoon@mokpo.ac.kr

보들은 임의의 시점에 따라 동적으로 변한다. 그리드 사용자들이 이러한 가변적인 그리드 환경에서도 가장 최근의 정보를 알기 원한다는 것은 중요한 사실이다.

그리드 환경에서 자원들의 상태 정보를 사용자에게 제공하는 방법으로 많이 사용되는 것이 MDS(Meta-computing Directory Service)이다. MDS를 구축하는데 글로버스 툴킷(Globus Toolkit)이 사용된다. MDS는 인터넷에서 DNS의 역할과 유사하다. MDS는 그리드 환경에서 자원의 상태 정보를 보유하고, 이를 사용자에게 제공한다. MDS는 이 정보를 저장하고 사용자에게 제공하기 위한 방법으로 LDAP(Lightweight Directory Access Protocol)을 이용한다.

글로버스(Globus)는 정보 제공을 위한 두 가지 서버를 갖고 있다. 하나는 정보를 수집하는 기능을 갖는 GRIS(Grid Information Service)이다. 다른 하나는 수집된 정보를 통합하는 GIIS(Grid Index Information Service)이다. 이 두 가지 서버가 수집하고 제공하는 정보는 각 자원의 구조(structure of each resource), 노드의 개수(number of nodes), 부하 정보(load information), 배치 작업 스케줄러(batch job scheduler), 네트워크의 상태(status of network) 등이 있다. 이 정보들은 LDIF(LDAP Data Interchange Format)의 형태로 API(application program interface)나 SDK(software developer's kit)를 이용하여 응용 프로그램 개발자나 자원 중재자(resource broker)에게 제공된다.

MDS는 몇 가지 문제점을 갖고 있다. 하나의 서버나 슈퍼컴퓨터에 대한 정보는 MDS를 이용하여 잘 관리될 수 있다. 그러나 MDS를 이용한 클러스터 시스템의 정보 관리는 적절하지 못하다. 클러스터 시스템을 구성하고 있는 노드의 개수나 각 노드의 부하 정보들이 MDS를 이용하면 적절하게 제공되지 못한다. 이러한 이유로 인하여 클러스터 시스템을 이용하는 사용자들은 불편을 겪게 된다. 본 논문에서는 클러스터 시스템의 정보를 제공하기 위한 PBS(Portable Batch System) 기반의 정보 제공자를 개발함으로써 이러한 문제점들을 해결하기 위한 방안을 제안하였다.

II. 클러스터 시스템

클러스터는 고속의 네트워크를 통하여 PC와 워크스테이션들을 연결함으로써 고성능과 고 가용성을 제공하는 시스템이다. 클러스터 기술이 갖는 여러 가지 장점들로 인하여 많은 대학이나 연구기관에서 사용하고 있다. 클러스터 시스템은 경제적이고 확장(upgrade)이 용이한 장점을 갖는다. 클러스터 시스템의 운영체제로 리눅스를 사용하는 것은 비용을 절감하고 협업의 가능성을 높일 수 있다는 장점을 갖는다. [그림 1]은 일반적인 클러스터의 구조를 보여준다.

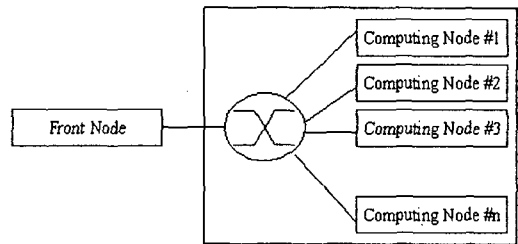


그림 1. 클러스터 구조

클러스터 시스템은 고속의 네트워크를 통하여 동형(homogeneous)의 컴퓨팅 노드들이 연결되어 있다. 사용자는 클러스터의 전위 노드(front node)를 통하여 각 컴퓨팅 노드에 접근할 수 있다. 일반적으로 전위 노드는 외부에서 접근이 가능하여야 하므로 공인(official) IP 주소를 할당받는다. 그러나 다른 컴퓨팅 노드들은 사실(private) IP 주소를 갖는다. 이러한 환경 하에서 글로버스 툴킷은 전위 노드에 설치되어야 한다. 컴퓨팅 노드들은 공인 IP 주소를 갖지 못하기 때문에 외부로부터의 접근이 불가능하다. 따라서 전위 노드가 각 컴퓨팅 노드들의 상태 정보를 MDS에 제공하여야 한다. 각 컴퓨팅 노드의 CPU 부하 정보, 메모리 량, 다른 사용자의 사용 상태 등의 정보들이 MDS로 제공되어야 한다. 이 문제를 해결하기 위한 두 가지 해결책이 있다. 먼저 클러스터를 관리하기 위한 클러스터 관리자(cluster manager)를 사용하는 것이다. 또 다른 방법은 각 컴퓨팅 노드에 정보 수집기(information collector)를 설치하는 것이다. 본 논문에서는 클러스터 관리자를 이용하여 정보를 제공하는 방법을 개발하였다.

III. PBS를 이용한 클러스터 관리자

한 사람만이 독점적으로 클러스터 시스템을 사용하지는 않는다. 많은 사람들이 공동으로 클러스터 시스템을 사용하기 때문에 각 사용자들의 작업을 스케줄하기 위한 스케줄러가 필요하다.

클러스터를 위한 잘 알려진 배치 작업 스케줄러에는 LSF, PBS, Condor 등이 있다. LSF는 좋은 성능을 발휘하지만 가격이 비싼 단점이 있다. PBS와 Condor는 무료로 제공되기 때문에 많이 사용되고 있다. 본 논문에서는 PBS를 사용하는 경우에 정보를 수집하는 방안을 연구하였다.

PBS는 큐잉 시스템(queueing system)의 하나이다. PBS는 클러스터를 이용하는 사용자들의 많은 작업들에 시스템 자원을 효과적으로 분배하는데 사용된다. PBS는 다양한 사용 통계를 발생한다. 그러나 이러한 통계들을 분석하기 위한 유틸리티는 적은 편이다.

PBS의 목적은 배치 작업의 실행 시작, 작업의 스케줄, 호스트들에 작업을 할당하는 것 등을 하는 것이다.

배치 시스템에서 각 사이트들은 서로 다른 작업에 사용될 자원의 유형과 양에 대한 사용 정책을 정의하고 그 기준을 세워두어야 한다. 또한 배치 시스템은 각 작업을 완료하기 위하여 필요한 자원들에 접근하는 메커니즘을 제공하여야 한다.

```
[root@joshua ptools]# ./pload -l cluster
joshua: up 33 days, 9 min, 2 users, load average: 0.00,
0.00, 0.00
joshua2:      up 1:46, 0 users, load average: 0.00,
0.00, 0.00
joshua3:      up 33 days, 7 min, 0 users, load
average: 0.00, 0.00, 0.00
joshua4:      up 33 days, 7 min, 0 users, load
average: 0.00, 0.00, 0.00
[root@joshua ptools]#

[root@joshua ptools]# ./pmem -l cluster
joshua: Mem: 254008K av, 192236K used, 61772K free
joshua2:   Mem: 254008K av, 70804K used,
183204K free
joshua3:   Mem: 254008K av, 43508K used,
210500K free
joshua4:   Mem: 254008K av, 41968K used,
212040K free
[root@joshua ptools]#
```

그림 2. "ptools" 명령어와 그 결과의 예

클러스터 관리자로서 PBS를 사용할 때 각 컴퓨팅 노드로부터 정보를 수집하기 위해서는 "ptools"라고 하는 유틸리티가 사용될 수 있다. "ptools"는 PBS에 의하여 관리되는 컴퓨팅 노드들의 상태와 부하 정보를 보여준다. [그림 2]는 "ptools" 명령어와 그 결과들에 대한 예를 보여준다.

IV. 클러스터 관리자를 이용한 정보 제공

그리드 환경에서의 클러스터 자원의 정보는 사용자들의 요구에 따라 제공되어야 한다. 사용자들은 가용한 노드의 개수, 각 노드의 CPU 개수, 각 CPU의 부하 정보 및 성능 정보 등을 요구할 필요가 있다. 정보 제공자는 이러한 정보들을 제공할 수 있도록 설계되어야 한다. 정보를 제공하는 구조는 [그림 3]과 같다.

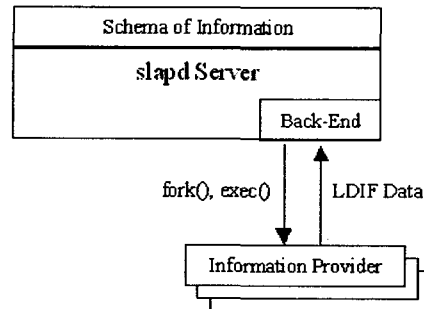


그림 3. 정보 제공자와 slapd

1. 스키마 정의

사용자들에게 새로운 정보를 제공하기 위하여 클러스터 정보에 대한 스키마가 정의되어야 한다. 스키마 파일에는 [그림 4]와 같이 각 정보에 대한 객체(object)와 그 속성(attribute)들이 정의되어야 한다.

다음과 같은 두 가지 종류의 정보 객체가 제공된다.

- 클러스터 정보 : 클러스터 이름, 클러스터 상태, 노드의 개수 등
- PBS 부하 정보 : 가동 시간(uptime), 사용자 수, 평균 부하 등

```

objectclass ( 1.3.6.1.4.1.14305.2.3.2.1001.21
NAME 'PBS_cluster'
SUP 'Mds'
AUXILIARY
MAY ( PBS-cluster-name $ PBS-cluster-state $
PBS-cluster-machines $ PBS-cluster-flops $
PBS-cluster-max-mem $ PBS-cluster-swap $ PBS-cluster-type
)
)
objectclass ( 1.3.6.1.4.1.14305.2.3.2.1001.22
NAME 'PBS_load'
SUP 'Mds'
AUXILIARY
MAY ( PBS-status-uptime $ PBS-status-users $
PBS-status-loadav-one $ PBS-status-loadav-five $
PBS-status-loadav-fifteen )
)
    
```

그림 4. 스키마 정의

이 정보들이 DIT(Directory Information Tree)에 표현되는 방법은 다음과 같다. 스키마 정의에는 각 정보의 이름과 OID(Object Identifier)가 필요하다. 각 클래스와 속성 유형마다 OID가 스키마 내에 존재하여야 한다. 다른 기관과 OID나 이름의 충돌을 피하기 위하여 IANA(Internet Assigned Numbers Authority)로부터 할당받은 번호를 사용해야만 한다.

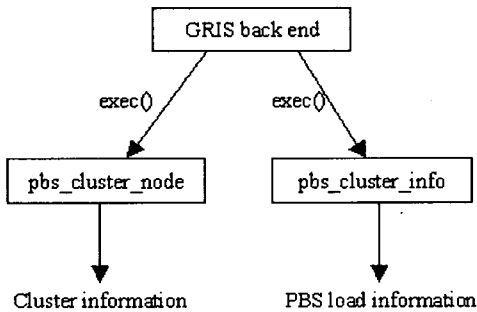


그림 5. 클러스터 정보 제공자의 시스템 구성

2. 정보 제공자의 구현

정보 제공자 프로그램은 표준 입출력 형식에 따라 개발되어야 한다. 이 프로그램은 GRIS의 후위(back end)에서 fork()나 exec()를 이용하여 호출될 수 있다. 이 프로그램의 출력은 LDIF 형태이어야 한다. 정보 제공

자는 쉘 스크립트를 이용하여 개발되었다. [그림 5]는 클러스터 정보 제공자의 시스템 구성을 보여준다.

일반적으로 MDS의 정보 제공자에게 입력은 구성 파일에 지정하거나 실행할 때 질의를 통해 전달된다. MDS는 GRIS 후위(back end)의 구성 파일에 기술된 정보 제공자 프로그램을 호출할 때는 쉘로 작성된 후위(back end)와 같은 입출력 형식을 따른다. 정보 제공자는 명령어에 따른 인수들과 같이 호출된다.

네트워크 정보 제공자는 출력으로 LDIF 객체들을 생성해야 한다. MDS에서 만들어진 LDIF 객체는 OID와 이름 공간 정의에 따라 화면에 출력한다. [그림 6]은 MDS에서 정보 제공자를 이용하여 클러스터의 정보를 수집한 결과의 예이다.

```

[globus@cluster etc]$
/usr/local/globus/libexec/pbs-cluster-infocast
dn: Pbs-Info=load
information,Pbs-Host-name=cluster01,Mds-Host-hn=cluster.h
pnet
.ne.kr,Mds-Vo-name=local,o=grid
objectclass: PBS-load
PBS-status-updays: 186
PBS-status-uptime: 3:49
PBS-status-users: 0
PBS-status-loadav-one: 0.01
PBS-status-loadav-five: 0.02
PBS-status-loadav-fifteen: 0.0
Mds-validfrom: 20021224091222Z
Mds-validto: 20021224091322Z
Mds-keepsto: 20021224091322Z

dn: Pbs-Info=load
information,Pbs-Host-name=cluster02,Mds-Host-hn=cluster.h
pnet
.ne.kr,Mds-Vo-name=local,o=grid
objectclass: PBS-load
PBS-status-updays: 186
PBS-status-uptime: 3:52
PBS-status-users: 0
PBS-status-loadav-one: 0.00
PBS-status-loadav-five: 0.00
PBS-status-loadav-fifteen: 0.0
Mds-validfrom: 20021224091222Z
Mds-validto: 20021224091322Z
Mds-keepsto: 20021224091322Z

.....

[globus@cluster etc]$
    
```

그림 6. 수집한 클러스터 정보의 예

V. 결론

클러스터 시스템에서 전위 노드가 시스템의 모든 노드를 대표하며 MDS에 자원 정보를 제공한다. 전위 노드는 컴퓨팅 노드들로부터 정보를 수집하여 MDS에 제공할 필요가 있다. 본 논문에서는 클러스터 관리자로서 PBS를 사용하는 방법을 제안하였다. PBS에서는 클러스터 시스템의 정보를 제공하기 위하여 "ptools" 유틸리티를 사용한다. 이 정보들은 LDIF 형태로 변환되어 MDS로 제공된다.

이렇게 됨으로써 MDS를 이용한 클러스터 시스템의 정보 관리가 적절하게 이루어질 수 있다. 클러스터 시스템을 구성하고 있는 노드의 개수나 각 노드의 부하 정보들을 적절하게 제공할 수 있게 됨으로써 클러스터 시스템을 이용하는 사용자들은 불편을 줄일 수 있게 된다.

향후에는 LSF나 Condor와 같은 다양한 클러스터 관리자를 위한 정보 제공자가 개발되어 통합되어야 할 필요가 있다.

for Distributed Resource Sharing," Proceedings of the Tenth IEEE International Symposium on High-Performance Distributed Computing (HPDC-10), IEEE Press, 2001.

[6] K.W. Kang, Y.H. Kang, D.H. Kim, K.M. Cho, and H.W. Park, "A Development of a Globus-based Metacomputer," CIC '02, 2002.
 [7] Y.H. Kang, K.W. Kang, D.H. Kim, K.M. Cho and H.W. Park, "Resource Discovery Scheme in P2P Grid Environments," The 2002 International Conference on Security and Management, 2002.

참고 문헌

[1] MDS 2.1 GRIS Specification Document, "Creating New Information Providers," USC/ISI, 2002.
 [2] Y. Tanaka, et al., "Performance Evaluation of a Firewall-compliant Globus-based Wide-area Cluster System," 9th IEEE HPDC 2000, pp.121~128, 2000.
 [3] I. Foster, and C. Kesselman, "Globus: A Metacomputing Infrastructure Toolkit," Intl. J. Supercomputer Applications, Vol.11, No.2, pp.115~128, 1997.
 [4] I. Foster and C. Kesselman (eds.), The Grid: Blueprint for a new Computing Infrastructure, Morgan Kaufmann Publishers, 1998.
 [5] K. Czajkowski, S. Fitzgerald, I. Foster, and C. Kesselman. "Grid Information Services

저자 소개

조 광 문(Kwang-Moon Cho)

정희원



- 1988년 2월 : 고려대학교 컴퓨터학과(이학사)
- 1991년 8월 : 고려대학교 컴퓨터학과(이학석사)
- 1995년 8월 : 고려대학교 컴퓨터학과(이학박사)

- 1995년~2000년 : 삼성전자 통신연구소 선임연구원
 - 2000년~2005년 : 천안대학교 정보통신학부 조교수
 - 2005년~현재 : 목포대학교 경제통상학부 전자상거래학 전공 전임강사
- <관심분야> : 전자상거래, 콘텐츠 유통, 모바일 콘텐츠, 데이터베이스, 그리드 컴퓨팅