제18회 한국정보처리학회 추계학술발표대회 논문집 제9권 제2호 (2002.11)

MDS 정보서비스를 위한 클러스터 시스템 정보 제공에 관한 연구

강경우, 강윤희, 김도현, 조광문
천안대학교 정보통신학부
e-mail:{kwkang, yhkang, dhkim, ckmoon}@cheonan.ac.kr

A Study on Providing the Information of Cluster System for MDS Information Service

Kyung-Woo Kang, Yun-Hee Kang, Do-Hyun Kim, Kwang-Moon Cho
Div. of Information & Communication Eng., Cheonan University

요 약

최근, 클러스터 시스템은 고성능 컴퓨팅 분야에서 널리 사용하는 컴퓨팅 환경이다. 그렇지만, 클러스터 시스템에 관한 정보가 그리드 환경에서 다른 시스템들에게 제대로 제공되지 못하고 있다. 본 연구에서는 그리드 환경에서 클러스터 시스템 정보를 사용할 수 있게 함으로써 그리드 환경의 확대를 시도한다.

1. 서론

그리드 정보 서비스는 그리드환경에서 중요한 요소로서 분산되어 있는 그리드 환경 내에 존재하는 자원들에 관한 최신의 정보를 사용자 또는 그리드 미들웨어의 다른 요소 시스템에 제공하는 시스템이다.

그리드 내에서 정보 서비스의 역할은 인터넷에서 DNS와 같은 위치와 비교될 수 있다. 사용자는 자신이 속한 VO(Virtual Organization) 내에서 자원들의 정보를 쉽게 검색하고 자신이 필요한 자원을 사용하기 원한다. 이와 같이 정보를 제공하기 위해서는 자원들에 관한 정보를 수집하고 정리해서 사용자 또는 다른 미들웨어에게 제공하는 정보 서비스가 필요하다.

그리드 환경에서 각 자원은 원격지에 흩어져 있고 서로 다른 기관이 관리하기 때문에 각 시점에 따라 상태가 자주 변한다. 네트워크 속도, 각 시스템의 부하, 사용 가능한 디스크 용량, 각 프로세서의 사용 가능 여부 등과 같은 정보는 시간에 따라 변하는 요소들이지만 그리드의 사용자 입장에서는 가장 최신의 정보를 얻을 수 있다는 것이 매우 중요한 부분이다.

전 세계적인 그리드 구축 과제에서 가장 많이 사용되고 있는 Globus Toolkit에서는 사용자에게 자원의 상태 정보를 서비스하기 위해 MDS (Metacomputing Directory Service)라고 하는 요소를 제공한다. MDS는 그리드 내에 존재하는 자원들의 상태 정보를 공유하고 사용자들에게 제공하기 위한 요소로서 인터넷의 DNS와 비슷한 것이다. 정보를 저장하고 사용자들에게 제공하기 위해 MDS는 LDAP(Lightweight Directory Access Protocol)를 이용한다. 정보 서비스를 위해 Globus에서는 두 개의 서버를 제공하는데, 각 자원의 정보를 수집하는 GRIS(Grid Information Service)와 수집된 정보를 통합하는 GIIS(Grid Index Information Service)이다. 이들이 수집하여 제공하는 정보는 각 자원의 구조, 노드 수, 부하 정보, 배치 작업 스케줄러, 네트워크 상태 등이다. 이들 정보는 LDIF(LDAP Data Interchange Format) 형태로 API나 SDK를 통해서 어플리케이션 개발자나, Resource Broker 등에 제공된다.

MDS에는 여러 가지 문제점이 있다. 본 연구에서는 이와 같은 문제들을 해결하기 위한 방안들을 제시한다. 본 연구에서 해결한 문제점은 현재 단일 서버 또는 슈퍼컴퓨터들의 정보는 MDS환경에 잘 전달이 되지만 클러스터 시스템의 정보들은 전달되지

않는 다는 문제가 있다. 클러스터의 컴퓨팅 노드들의 개수가 몇 개인지, 각 노드의 부하는 어떠한지 현재는 MDS환경에 전달되지 않기 때문에 클러스터를 활용하기 원하는 사용자들은 불편을 겪고 있다.

2. 클러스터 시스템

"클러스터(Cluster)"란 PC 또는 워크스테이션을 고속 네트워크로 연결하여 고성능 또는 고가용성을 얻을 수 있도록 하는 기술 또는 시스템을 말한다. 클러스터 기술은 상용으로 시중에 판매되는 장비들을 이용하여 구축하기 때문에 적은 비용으로 높은 성능을 낼 수 있어 확장성이나 업그레이드 등의 장점들을 모두 가지고 있기 때문에 최근에 대학교를 중심으로 널리 활용되고 있다. 또한 클러스터의 운영체제로 많이 사용되는 Linux는 코드의 이식비용을 현격히 감소시키고 공동작업의 가능성을 높여 준다. 바로 이러한 특성들이 클러스터를 계산 과학 분야에서 새로운 컴퓨팅 동향으로 자리잡아 가고 있는 이유이다.

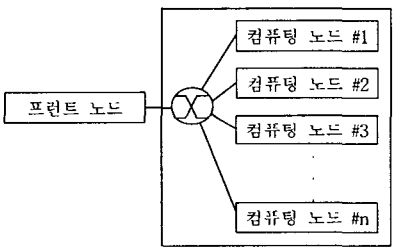

그림 1. 클러스터 구조

클러스터 시스템은 위의 그림과 같이 동종의 컴퓨팅 노드들이 고속의 네트워크로 연결되어 있고 클러스터를 이용하는 연구자는 프런트 노드를 통해서 컴퓨팅 노드에 접근이 가능하다. 일반적으로 프런트 노드는 외부에서 접근해서 사용해야 하기 때문에 공인 IP주소를 부여받지만 컴퓨팅 노드들에게는 공인 IP 주소보다는 사설 IP 주소를 부여한다. 이와 같은 클러스터를 그리드 환경 내에 컴퓨팅 자원으로 사용된다면 공인 IP주소를 부여받은 프런트 노드에 Globus 툴킷을 구축할 것이다. 컴퓨팅 노드들은 공인 IP 주소를 가지고 있지 못하기 때문에 Globus를 구축한다 해도 외부에서 접근이 불가능하기 때문이다. 이와 같이 프런트에 Globus를 구축하면 프런트 노드가 책임져야 하는 또 다른 부분은 컴퓨팅 노드들의 상태에 관한 정보를 MDS에 제공하는 것이다. 각 컴퓨팅 노드들의 CPU부하정보, 메모리양, 다른 사용자들이 사용여부 등과 같은 정보를 모아서 프런트 노드는 MDS에게 제공해야할 의무가 있다. 이와 같은 문제를 해결하기 위해서 두 가지 방안이 있다. 첫 번째는 클러스터를 관리하는 클러스터 관리기를 활용하는 것이고 두 번째는 자체적으로 정보수집기를 컴퓨팅 노드에 심어놓음으로 해결하는 방법이다. 본 논문에서는 클러스터 관리기를 활용한 정보제공 방법을 구현하였다.

3. 클러스터 관리기를 활용한 정보제공

클러스터 시스템은 한사람만이 독점적으로 사용하는 것이 아니고 여러 사람이 사용하기 때문에 사용자 작업을 스케줄을 해 주는 스케줄러가 존재한다. 스케줄러는 또 다른 이름으로 배치작업스케줄러라고 부르는데 고성능 컴퓨팅 자원을 활용하는 작업들은 오랜 시간 수행되기 때문에 자원의 효율성을 위해서 배치처리 방식으로 처리된다.

클러스터를 위해 사용되는 잘 알려진 배치작업스케줄러들은 lsf, condor, pbs 등이 있다. 그런데, lsf는 상용 소프트웨어로써 기능은 막강하지만 비싸다는 문제가 있다. 그래서, 일반적으로는 무료로 사용이 가능한 condor와 pbs를 많이 사용한다. 본 연구에서는 condor를 사용했을 때 정보수집 방안을 다루도록 한다.

Condor는 분산된 워크스테이션들을 통합하여 사용할 수 있게 하는 High Throughput Computing 환경이다. Condor의 개발은 과학기술분야의 계산과학 분야 전문가들이 유휴 컴퓨팅 자원들을 통합하여 사용하고자 하는 필요에 의해 시작되었다. 현재 대부분의 유닉스 시스템 상에 설치가 가능하고, NT 시스템을 위해서는 개발 중에 있다. Condor 환경에서 High Throughput Computing의 성공은 각 컴퓨팅 자원을 소유한 소유주에 달려있다. 각 소유주는 자신의 컴퓨팅 자원을 어떤 상황일 때 외부에 할당할지를 결정할 수 있다. 예를 들어, 소유주가 조금이라도 사용하면 외부사용자가 사용할 수 없는 엄격한 조건을 달 수 도 있고 어느 정도의 부하까지는 허용하는 조건을 달 수 도 있다.

Condor의 중요한 특징은 ClassAd 기법에 있다. 각 사용자는 자신의 작업을 수행하기 위해 적합한 컴퓨팅 자원을 찾기 원한다. ClassAd는 각 컴퓨팅 자원에 대해 프로세서 타입, 메모리 양, 운영체제 타입 등 모든 가능한 특징에 관한 정보이다. 또한, 사용자는 자신의 작업이 부동소수점 연산에 있어 나은

컴퓨터에서 수행되길 원하거나 특정한 컴퓨터에서 수행되기를 원할 수 있는데 이와 같은 모든 것도 ClassAd에서 기술할 수 있다. 사용자는 자신의 작업을 위한 ClassAd를 기술한다. 그러면, Condor는 사용자 작업에 관한 ClassAd와 각 컴퓨팅 자원의 ClassAd를 일치시킴으로 적당한 자원들을 찾는다. 각 컴퓨팅 자원 관리자는 자신의 자원에 대해 다른 사용자의 접근을 명시함으로 제한할 수 있다. 특정 시간에만 사용 가능하게 한다거나 특정한 사용자 그룹에만 사용을 허용한다거나, 어떤 작업을 더 선호하는지도 기술 할 수 있다.

Condor가 클러스터관리기로 활용될 때 각 컴퓨팅 노드들에 관한 정보를 수집하기 위해 사용할 수 있는 명령어는 condor_status이다. condor_status는 condor가 관리하는 컴퓨팅 노드들의 상태와 부하정보 등을 보여준다. 다음은 condor_status 명령어를 사용한 예이다. 아래에서 볼 수 있듯이 각 컴퓨팅 노드들의 평균부하와 메모리의 양을 얻을 수 있다. 이 정보를 기반으로 정보제공자는 MDS에 부하정보와 현재 다른 사용자가 사용하는지, 메모리크기 등을 알려줄 수 있다.

```
% condor_status
Name              OpSys Arch   State     Activity LoadAv Mem  ActvtyTime
node1@vulture.c LINUX INTEL Owner     Idle     0.020  128  0+00:57:13
node2@vulture.c LINUX INTEL Claimed   Busy     1.006  128  0+01:16:03
node3@vulture.c LINUX INTEL Claimed   Busy     0.978  128  0+03:32:53
node4@vulture.c LINUX INTEL Claimed   Busy     1.001  128  0+02:21:07
                     Machines Owner Claimed Unclaimed Matched Preempting
  INTEL/SOLARIS26 4        0     4         0         0       0
            Total 4        0     4         0         0       0
```

그림 2 condor_status 결과

4. 클러스터 관리기를 활용한 정보제공

그리드 환경 내에 클러스터 자원들에 관한 정보들은 사용자의 요구에 따라 제공되어야 한다. 사용자는 클러스터 내에 사용 가능한 노드수, 노드 내에 CPU수, 이들의 load정보, 각 CPU의 성능 등에 관한 정보를 필요로 한다. 이와 같은 정보를 제공하기 위해 정보 제공자를 설계하고 구현해야 하는데, 구조를 그림으로 나타내면 다음과 같다.

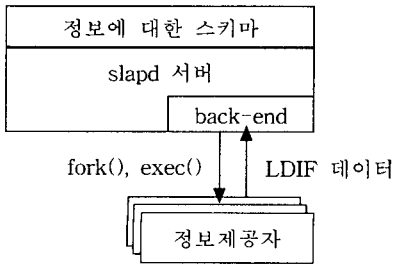

그림 3 정보제공자와 slapd 관계

4.1 스키마 정의

새로운 정보를 사용에게 제공하기 위해서 제공될 클러스터 정보에 대한 스키마를 먼저 정의해야 한다. 스키마 파일에는 각 정보에 대한 객체와 속성들을 정의해야 한다. 클러스터 정보에 대한 객체들과 속성들을 객체중심으로 요약하면 다음과 같다.

```
objectclass ( 1.3.6.1.4.1.14305.2.3.2.1001.1
NAME 'CondorCluster'
SUP 'Mds'
AUXILIARY
MUST Condor-Cluster-name
MAY ( Condor-Cluster-machines $ Condor-Cluster-owner $
                    Condor-Cluster-unclaimed )
 )
objectclass ( 1.3.6.1.4.1.14305.2.3.2.1001.2
NAME 'CondorHost'
SUP 'Mds'
AUXILIARY
MUST Condor-Host-name
 )
objectclass ( 1.3.6.1.4.1.14305.2.3.2.1001.3
NAME 'CondorLoad'
SUP 'Mds'
AUXILIARY
MAY ( Condor-status-loadav $ Condor-status-cpubusyt $
              Condor-status-state $ Condor-status-acty )
  )
objectclass ( 1.3.6.1.4.1.14305.2.3.2.1001.4
NAME 'CondorHostInfo'
SUP 'Mds'
AUXILIARY
MAY ( Condor-host-opsys $ Condor-host-arch $
    Condor-host-cpuf $ Condor-host-mips $
    Condor-host-ncpus $ Condor-host-maxmem $
    Condor-host-disk )
  )
```

제공되는 정보객체는 크게 3가지이다.

- 클러스터 전체적인 정보: 클러스터 이름, 클러스터 내의 전체 노드의 수, 현재 사용자가 사용하는 노드 수, 사용하지 않고 있는 노드 수
- 각 노드의 정적인 정보: 운영체제, 구조, FLOPS, MIPS, 노드에 있는 CPU수, 메모리양, 하드 디스크양
- 각 노드의 동적인 정보: 평균 부하정보, cpu의 busytime, 상태정보

이와 같은 정보들은 DIT에서 어떻게 표현되어야 할지를 결정한다. 이것은 스키마의 정의와 OID, 정보에 대한 이름 등이 이에 해당한다. 이때 OID는 각

클래스와 속성 타입에 대한 스키마에 존재해야 한다. MDS를 위한 ISI의 이름 공간은 IANA(Internet Assigned Numbers Authority)와 같이 등록된 OID 부분공간이다. ISI는 사용자 정의 정보 제공자들을 위해 OID들을 관리하기 위해 OID 부분공간을 활용한다. 다른 기관과 OID 및 이름의 충돌을 피하기 위해 IANA에서 부여한 번호를 이용한다.

4.2 정보제공자 구현

입력과 출력 형태를 표준 형식에 맞춰서 정보 제공자 프로그램을 작성한다. 이 프로그램은 GRIS의 back end에서 fork()나 exec()를 이용하여 호출이 가능해야 하고 출력은 LDIF 형태로 되어야 한다. 정보 제공자를 작성하기 위해 쉘 스크립트를 이용하여 구현하였다. 정보제공자의 시스템 구조를 개략화하면 다음과 같다.

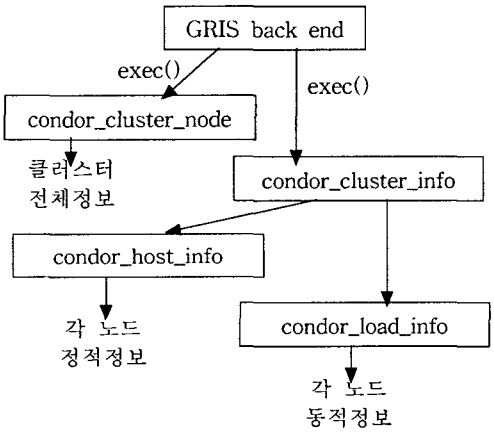

그림 4 클러스터 정보제공자 시스템 구성도

5. 결론

클러스터 시스템은 일반적으로 프런트 노드가 시스템 내의 모든 노드를 대표하며 MDS에 클러스터 내의 자원 정보들을 제공해야 한다. 프런트 노드가 MDS에 정보를 제공하기 위해서는 컴퓨팅 노드들의 정보를 모집할 수 있어야 하는데 본 연구에서는 이를 위한 방안으로 클러스터 관리 시스템인 Condor, PBS의 도움을 받는 방안을 제시하였다. Condor에서는 condor_status라는 명령어가 컴퓨팅 노드들의 평균 부하, 메모리 양, 다른 사용자가 사용하는지 여부 등을 알려준다. condor_status가 제공해 주는 정보는 클러스터 정보 제공자에 의해 다시 MDS에 LDIF 형식으로 제공된다.

참고문헌

[1] "Creating New Information Providers," MDS 2.1 GRIS Specification Document, USC/ISI, May, 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, C. Kesselman, "Globus: A Metacomputing Infrastructure Toolkit" Intl. J. Supercomputer Applications, 11(2):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, C. Kesselman, "Grid Information Services for Distributed Resource Sharing." Proceedings of the Tenth IEEE International Symposium on High-Performance Distributed Computing (HPDC-10), IEEE Press, August 2001

[6] http://www.cs.wisc.edu/condor

[7] http://www.gridforum.org

[8] http://www.gridforumkorea.org

[9] http://www.globus.org/