

분산 환경을 위한 메타정보 시스템의 구축

임재훈⁰, 황석찬, 조재영
승실대학교 컴퓨터학부

{suplim⁰, schwang, choi}@computing.ssu.ac.kr

Design and Implementation of a Meta-Information System for Distributed Systems

Jaehoon Lim⁰, Seogchan Hwang, Jaeyoung Choi
School of Computer, Soongsil Univ.

요 약

기존의 분산 환경은 단순히 네트워크로 연결된 제한된 컴퓨팅 자원을 활용하는 기술이다. 그러나 자원에 대한 집중적인 관리가 이루어지지 않아 계산 집중적인 분산 컴퓨팅 환경에는 적합하지 않다. 분산 자원들을 효과적으로 이용하여 저렴한 비용으로 사용자가 요구하는 최적의 성능을 발휘할 수 있도록 환경을 구축하기 위해서는 자원의 집중적인 관리가 요구된다. 변경되기 쉬운 많은 분산 자원들을 관리하기 위해서는 자원에 대한 세부적인 사항을 기술한 메타정보들을 효과적으로 저장, 관리 등을 할 수 있는 정보 시스템이 필수적이다.

본 논문에서는 분산 컴퓨팅 환경을 위한 자원의 위치나 상태정보 등 자원의 메타 정보를 효과적으로 저장하고 관리하는 메타정보 시스템의 전반적인 구조와 이와 관련된 서비스와의 관계를 고려하여 구현한 메타 정보 시스템을 소개한다.

1. 서 론

인터넷의 비약적인 보급으로 네트워크의 빠른 발전과 고성능의 마이크로 프로세서의 개발로 분산 시스템은 전성기를 맞고 있다. 이러한 분산 환경은 단순히 웹 서비스 제공에 집중되어 있으며 동일한 컴퓨팅 자원만을 사용하는 한계가 있다. 그러나 최근에는 많은 계산량을 요구하는 애플리케이션을 위한 분산 시스템 연구에 관심이 집중되고 있다.

최근에는 물리적 혹은 지리적으로 분산되어 있는 자원들을 이용하여 슈퍼컴퓨터와 비슷한 성능을 가지면서 저렴하게 이용할 수 있는 분산 기술을 개발하고 있다. 이러한 기술은 분산된 자원들을 어떻게 효율적으로 저장하고 관리하며, 변경되기 쉬운 자원에 대한 정확하고 신뢰할 만한 정보를 얻는 것이 중요하다.

분산된 환경에서 발견된 자원에 대한 변경된 정보를 동적으로 제공하기 위한 특별한 서비스가 필요하며, 서로 다른 종류의 플랫폼과 네트워크로 구성되어 있는 환경을 극복하고 필요한 자원을 확보하기 위한 기술이 요구된다. 이를 위하여 분산 자원에 대한 메타정보를 관리할 수 있는 서비스 환경을 필요로 한다. 현재까지 기존의 메타 정보 시스템은 여러 분야에서 다양하게 응용되어 왔으며, 계산 중심의 분산 컴퓨팅 환경을 위해서는 분산되어 있는 이들 자원들을 체계적으로 관리할 수 있는 정보 시스템이 필수적으로 요구된다.

본 논문에서 제시하고 있는 연구는 BT분야의 바이오 인포매틱스(Bioinformatics)[1]의 일환으로 버추얼 스크리닝(virtual screening)[2]기술을 지원하는 분산처리 환경을 위한 정보 서비스를 제공한다. 이러한 계산중심 작업의 High-Throughput을 위한 분산 컴퓨팅 환경에서의 핵심기술로 자원의 정보를 정확

하고 신속하게 이용하며 관리할 수 있도록 보다 효과적인 분산 컴퓨팅 환경을 위한 메타정보 시스템의 설계 및 구현을 제시한다.

2. 관련연구

메타정보(meta-information)란 특정 정보를 기술한 정보로서 특정 자료로부터 추출하여 검색에 사용될 수 있는 정보 형태로 가공된 인덱싱(indexing) 정보를 말한다. 자원이 늘어날수록 자원의 정보를 효율적으로 관리하며 활용하기 위해 메타 정보를 체계적이고 실용적으로 저장, 관리할 수 있는 통합된 정보 서비스를 필요로 한다.

이전의 메타정보 시스템으로는 LUMP[3], STILE[4] 등의 시스템이 개발되었다. LUMP의 경우에는 이질적인 분산 환경을 위한 일반적인 메타정보를 표현하기 위한 시스템의 설계와 구조를 위한 모델이며 STILE 모델의 경우, JPEG, GIF, PDF, 그리고 RTF와 같은 여러 형식의 문서로 구성된 정보를 주제(subject) 영역으로 분리하고 계층적으로 정보를 구성하여 사용자에게 html형식 같은 웹기반으로 정보를 제공하는데 사용 된다.

메타 컴퓨팅에서 사용되는 정보 서비스의 경우 Globus[5]에서는 MDS(Metacomputing Directory Service)[6]라고 하는 디렉토리 서비스를 이용하여 자원의 메타정보를 저장하고 관리하는 일을 담당한다. MDS는 계층적 구조로 정보를 저장할 수 있으며, 정보에 대한 검색효율을 고려하여 일반적인 관계형 데이터베이스 시스템보다 정보의 양에 관계없이 빠르고 일정한 검색을 보장할 수 있는 LDAP (Lightweight Directory Access Protocol) [7] 서비스를 이용한다.

3. 메타정보 시스템 모델

분산 컴퓨팅 환경을 위해서는 그림 1과 같이 자원을 소유하고 있는 자원 제공자(resource provider)와 자원의 사용을 원하는 자원 소비자(resource consumer) 간의 이벤트 교환이 이루어진다. 이러한 일련의 작업을 위한 서비스의 자원 정보를 저장하여 제공하는 기본적인 Producer-Consumer 모델 [7]을 제시 하고 있다. 또한 분산 환경이 커질수록 효율적으로 자원의 위치와 상태 등의 정보를 제공하기 위하여 자원의 정보를 수집하여 저장하고 관리되어야한다. 이를 지원하는 메타 정보 시스템은 디렉토리 서비스를 이용하여 특정한 영역(domain)에 쉽게 적용 가능하도록 설계되어 있다. 또한 디렉토리 서비스를 제공하기 위해 LDAP라는 기술을 적용하고 있다.

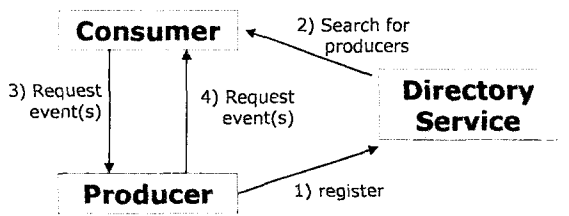


그림 1. Producer-Consumer 모델

LDAP란 IP 프로토콜 기반의 표준화된 소프트웨어 프로토콜이다. LDAP 경우 모든 정보를 디렉토리 서버에 계층적인 방식으로 모든 정보를 표현하며 저장한다. 이러한 특징은 관계형 데이터베이스보다 정보의 양이 증가하더라도 검색에 대한 트랜잭션의 성능이 일정하다는 장점이 있다. 또한 규모가 큰 분산 환경을 위해서는 전화번호부의 yellow page와 유사한 디렉토리 검색 방법을 제공한다. 이것은 일정한 검색 조건만 만족하면 저장된 정보를 쉽게 찾을 수 있는 기능을 제공한다.

또한, 디렉토리 서비스에는 Reference 와 Cache에 의한 인덱스기능과 필터기능으로 정보에 대한 빠른 검색기능과 최적의 질의 접근(query access)서비스를 제공하므로 자원의 메타 정보를 위한 최적의 검색 서비스를 제공할 수 있다.

4. 메타 정보 시스템의 설계 및 구현

메타정보 서비스를 위한 기본적인 정책은 동적으로 변화하는 자원들의 정보를 정확하게 유지하며 이런 메타정보를 필요로 하는 서비스 모듈이나 시스템에 빠르게 제공하는 것으로 검색의 효율성에 중점을 두고 설계되었다.

4.1 메타 정보 시스템의 구성 및 설계

구현된 메타정보 시스템의 핵심인 디렉토리 서비스와 이를 지원하는 서비스에는 자원 발견(Resource discovery), 자료 공개(Resource Publication), 자원 감시(Resource Monitoring)와 자원 선별기(Resource Selector)로 구성되어있다.

자원 발견 서비스는 새로운 자원의 정보를 갱신하기 위한 기능

으로 사용 가능한 자원의 메타정보를 주기적으로, 혹은 시스템 부하가 적은 시기에 자원 제공자에게 요청하여 디렉토리 서비스에 항상 자원의 새로운 메타정보가 저장되도록 지원하는 서비스를 제공한다. 자원을 관리하는 호스트나 PC에서 자원에 관한 정보를 통신경로를 통해 전달하게 된다. 이때 자원 발견 서비스에서는 받은 자원의 정보를 LDAP 스키마에 맞는 형식으로 변환하여 저장한다. 이를 위해 항상 자원 발견 서비스는 자원 제공자의 이벤트(혹은 메시지)를 받기 위한 서비스 데몬으로 제공된다.

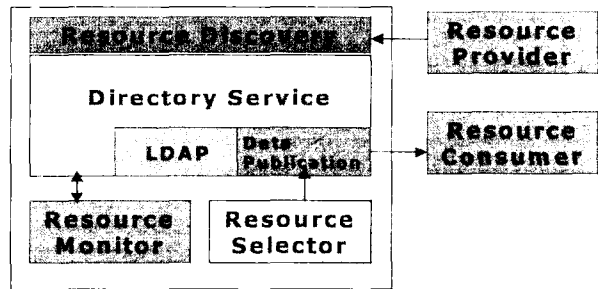


그림 2. 메타 정보 시스템 구조

자원 공개 서비스는 자원 할당 서비스나 로컬 매니저 등과 같이 자원에 대한 정보를 필요로 하는 자원 소비자가 자원에 대한 정보를 요청하면 LDAP서버에 저장되어 있는 적합한 자원의 메타 정보를 찾아 일정한 전달 형식으로 변환하여 제공하게 된다.

자원 감시 서비스는 최적의 자원 관리를 위하여 주기적으로 자원의 상태를 확인할 수 있는 기능을 제공함으로써 자원을 최대한 활용할 수 있도록 점검하는 서비스를 제공한다.

더불어 자원의 증가는 곧 자원에 대한 더 많은 메타 정보의 양을 저장해야 함을 의미한다. 이러한 많은 메타 정보 중에서 애플리케이션에 적합한 자원을 선별하는 일이 매우 중요하다. 그러므로 자원 선별기는 자동적으로 애플리케이션에서 필요한 특성을 확인하여 LDAP 데이터베이스의 검색에 사용할 필터를 구성함으로써 사용 가능한 최적의 자원을 선별하여 준다.

4.2 Producer-Consumer 기반의 메타 정보 시스템

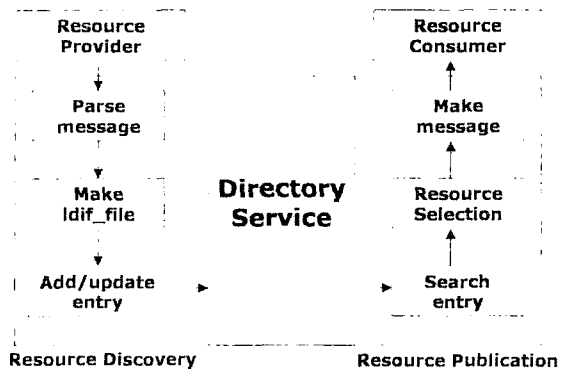


그림 3. Producer-Consumer 모델을 적용한 시스템 구조

시스템 구성을 위한 기본적인 모델의 구성 요소로는 그림 3에서와 같이 자원 발견 서비스와 디렉토리 서비스를 위한 LDAP 데이터베이스, 그리고 저장되어 있는 적절한 정보를 찾아 제공하는 정보 공개 서비스로 구성되어 있다.

자원 제공자는 자원 발견 서비스에 자원의 정보를 보내게 되면 정보 메시지를 파싱하여 데이터를 해독한다. 그리고 LDAP에 정보를 저장할 수 있는 *ldif*파일 형식으로 만들어 추가한다. 자원 소비자의 요청에 의해 LDAP에서 메타 데이터를 검색하고 최적의 자원을 선별하여 전송 가능한 형태로 변환한다.

4.3 LDAP의 적용

분산 컴퓨팅 환경에 맞는 메타정보 서비스를 위해서는 LDAP 서비스를 구성하는 작업이 선행되어야 한다. LDAP에서는 사용자가 직접 스키마를 구성하여 저장될 정보의 기본단위인 엔트리(entry)를 결정할 수 있다. 우선, 버추얼 스크리닝을 지원하기 위한 자원의 메타정보를 저장하기 위해서 그림 4와 같은 LDAP 스키마(schema)구조를 정의한다.

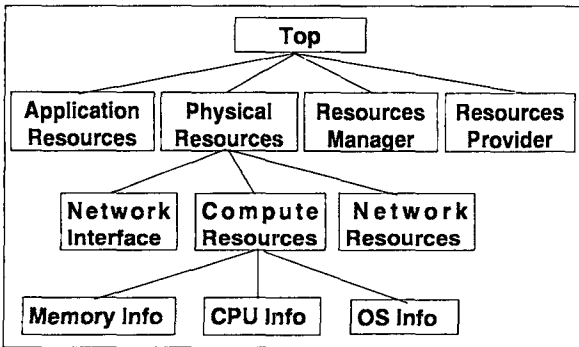


그림 4. 분산 환경을 위한 LDAP Schema 구조

그림 4에서와 같이 분산된 환경에서 자원을 관리하기 위하여 사용할 수 있는 자원들의 목록을 디렉토리로 구성한다. 예를 들어 CPU Info의 정보를 저장하기 위한 LDAP의 objectclass와 관련된 *cpuType*에 대하여 기술한 attribute를 표현하면 그림 5와 같다.[8]

```

objectclass ( 1.1.2.2.1.1.17 NAME 'CPU_Info' SUP
ComputeResources STRUCTURAL
MUST rn
MAY ( cpuClock $ cpuCount $ cpuType $
cpuLoad1 $ cpuLoad5 $ cpuLoad15 ) )

attributetype ( 1.1.2.1.1.38 NAME 'cpuClock'
EQUALITY numericStringMatch
SYNTAX 1.3.6.1.4.1.1466.115.121.1.26 )
    
```

그림 5. objectclass와 attributetype 구성의 예

그리고 이러한 속성과 속성 값으로 검색 필터를 구성한다.

예를 들면, CPU clock(MHz)과 memory (mega) 속성이 원하는 기준에 맞는 지 검사하는 필터의 예를 나타내면 다음과 같다.

(& (cpuClock = 500) (memory >= 64))

이와 같이 CPU의 Clock는 500MHz와 memory는 64Mbyte이상의 의미로 나타낸다.

5. 결론 및 향후연구

본 논문에서는 차세대 인터넷을 위한 준비 단계로서 분산 환경을 위한 메타 정보 서비스의 필요성과 실제 구현을 제시하였다. 이러한 시스템의 구성은 분산되어 있는 기기종의 컴퓨팅 자원을 보다 효율적으로 통합, 관리하고 이용하기 위한 서비스를 제공한다.

본 연구에서 구축한 메타정보 시스템은 데이터의 양에 관계없이 신속하고 정확한 정보를 유지하기 위해 LDAP를 기반으로 구성되었다. 또한 분산된 자원의 메타정보를 수집하여 애플리케이션에 필요한 최적의 자원을 선별하는 기능과 자원에 대한 지속적인 감시와 관리 기능을 제공하고, 새로운 메타 정보의 추가와 삭제가 용이하며 검색을 위한 필터를 쉽게 적용할 수 있도록 구현되었다.

본 연구는 확장된 분산 컴퓨팅을 위해서 보다 많은 추가적인 기능을 필요로 한다. 자원의 안정적인 사용을 위한 자원 예약을 고려하여 원활한 정보를 제공하도록 설계되어야 하며 메타 데이터의 투명성과 안전성을 보장하기 위한 보안을 제공하고, 확장성을 고려하여 원격지의 다른 LDAP 데이터베이스의 메타정보를 일정 기간 캐쉬함으로써 추가적인 검색이 최소화되도록 고려해야 한다.

6. 참고 문헌

- [1] Cynthia Gibas, Per Jambeck, Developing Bioinformatics Computer Skills.
- [2] B. Waszkowycz, T.D.J.Perkins, R.A.Sykes, J.Li, Large-scale virtual screening for discovering leads in the postgenomic era.
- [3] M. Madsen, I. Fogg & C. Ruggles, Libri 44:237-257, 1994.
- [4] M. Madsen et al, Software Practice and Experience, under submission.
- [5] <http://www.globus.org>.
- [6] Heinz Johner, Larry Brown, Franz-Stefan Hinner, Wolfgang Reis and Hohan Westman, Understanding LDAP, IBM, 1998.
- [7] Karl Czajkowski, Ian Foster, Nick Karonis, Carl Kesselman, Stuart Martin, A Resource Management Architecture for Metacomputing System.
- [8] <http://www.openldap.org>.