

고성능 그리드 환경을 위한 자원정보모델에 관한 연구

김희철* · 이강우** · 이용두* · 조세홍*

대구대학교 정보통신공학부* · 동국대학교 정보통신공학부**

요 약

고성능 그리드 환경을 구축을 위해서는 그리드 내의 사용자, 관리자, 서비스, 하드웨어 등에 대한 제반 정보서비스를 제공하는 그리드 정보서비스(Grid Information Service)가 필수적으로 요구된다. 본 논문에서는 그리드 정보서비스의 구조(Grid Information Service Architecture) 설계에 근간이 되는 자원정보 모델(Resource Information Model)에 대하여 체계적인 연구를 수행하였다. 본 연구는 자원요청, 자원탐색, 자원할당 등 자원 스케줄링의 최적화 알고리즘의 개발 및 구현을 보장할 수 있는 자원정보모델의 성격 및 특성에 대한 요구정의(Requirement Definition)의 도출에 초점을 두었다. 본 고에서는 고성능 그리드 정보서비스(GIS)는 엔티티기술(Entity Description)과 자원 상호 간의 관계기술(Relation Description)을 포함한 자원기술(Resource Description), 스케줄링 지원, 자원정보 표현모델과 저장 모델의 독립성, 사용자 측면의 자원기술방식과 시스템 측면의 자원기술방식의 분리에 대한 이슈가 명확하게 반영된 자원정보모델을 기반으로 하여 설계되어야 한다는 점을 명확히 제시한다. 이러한 자원정보모델에 준하여 기존의 대표적인 자원정보모델들을 분석한 후 그 결과를 기술한다.

A Resource Information Model for High Performance GRID Environemnts

Hie-Cheol Kim · Kang-Woo Lee · Yong-Doo Lee · Sae-Hong Cho

For high performance Grid environments, an effective GIS(Grid Information System) should be adopted. In the design of GIS architecture, its grid resource information model provides a key basis. This paper presents our study on the exploration of a high performance grid resource information model. According the exploration, we identified the followings. The resource information model should clearly address the issues of relation descriptions as well as resource descriptions, issues related to scheduling support, the issue of decoupling the expression model of resource information from data repository models, and finally the issue of decoupling user-level resource descriptions from system-level resource descriptions. Based on the proposed conceptual organization of resource information models, analysis result for the existing resource information models are presented

1. 서론

21세기 무한경쟁시대에 지식정보강국으로 도약하고 국가 경쟁력의 확보를 위해서는 기초과학 및 기간산업의 연구력 증강이 그 핵심 요체라 할 수 있다. 현재 기초과학과 산업기술 연구는 고속연산, 대량의 데이터 처리, 첨단 장비의 공유 및 분산협업 등이 필수적으로 수반된다. 지난 20년 간 다양한 컴퓨팅 환경이 개발되어 활용되고 있지만 기초과학 및 산업기술의 연구에 수반되는 고성능 컴퓨팅 파워, 각 연구소마다 소장하고 있는 수십 테라바이트 규모의 각종 데이터의 공유 등의 요구 사항을 만족시킬 수 있는 방법에 관해서는 효과적인 비전을 제시하고 있지 못하다.

최근 이러한 요구를 만족시켜줄 수 있는 수단으로서 GRID 기술이 주목을 받고 있다[1,2]. GRID란 지리적으로 분산된 고성능 컴퓨터, 대용량 DB 및 첨단 장비 등의 정보통신자원을 고속 네트워크로 연동하여 상호공유·이용할 수 있도록 하는 정보통신 서비스로 그 혜택은 기존의 WWW를 능가한다고 평가되어 전 세계적으로 그 인프라를 구축하고자 하는 노력이 매우 활발하게 진행되고 있다.

GRID의 구축을 위해서는 해당 GRID 내의 Computation 자원, Storage 자원, Network 자원, Code repository 및 Catalog 등을 효율적으로 관리할 수 소프트웨어로서 미들웨어가 요구된다. 즉 사용자가 필요한 자원을 미들웨어를 통하여 요구하면 미들웨어는 GRID 내에서 가용한 자원 중에 선택하여 사용자가 요구한 자원을 사용할 수 있도록 예약·할당하며 사용자에게 그 자원을 활용하여 원하는 작업을 할 수 있도록 제반 관리 서비스를 제공한다[2].

그러므로 현재 이러한 미들웨어의 기술은 분산, 병렬처리 분야의 연구자들을 중심으로 전세계적으로 매우 활발하게 연구되고 있다[10,11,13]. 그 대표적인 연구 프로젝트로서는 Argonne National Lab.과 USC의 ISI 연구소가 수행하였으며 GRID 미들웨어의 선도 프로젝트라 할 수 있는 Globus 프로젝트가 있다[11]. 현재 본 프로젝트는 중간 결과로서 Globus Toolkit을 제공하고 있으며 이 Toolkit은 전세계 관련 연구기관에서 GRID 시험시스템 구축 프로젝트에 널리 활용되고 있다. Condor 프로젝트는 ClassAD 언어 기반의 프레임워크를 기반으로 University of Wisconsin-Madison에서 개발하여 150여 개의 대학과 기관에 분산 자원관리 서비스를 제공할 수 있는 능력을 갖는 GRID 미들웨어 프로젝트이다[13]. 이는 HRM and SRB 프로젝트 파일공유 및 디스크 캐쉬 기반 기술에 중점을 두고 시작한 미들웨어 프로젝트로서 현재는 GRID 구축에 일반적으로 활용될 수 있는 미들웨어 구축을 목표로 하고 있다.

그리드 정보 서비스(Grid Information Service)는 그리드 내의 사용자, 관리자, 서비스, 하드웨어 등에 대한 제반 정보서비스를 제공하는 그리드(GRID) 미들웨어 내의 소프트웨어 컴포넌트(Component)이다[4,9]. 현재 그리드 정보 서비스(GIS)의 시스템 구조에 관한 많은 연구가 진행되고 있다[3,15]. 이러한 연구는 데이터 모델, 저장 구조, 자원기술 방식 등에 핵심 기술에 관한 세부 연구로서 궁극적으로 실제 그리드 환경에서 활용하기 위한 목표를 지니고 있다.

그리드 정보 서비스는 자원, 서비스, 컴퓨테이션(Computation)의 특성 정의, 탐색, 모니터링(Monitoring)을 포함한 그리드의 다양한 기능들과 직간접으로 상호 연관성을 갖는다[2,9]. 광역

지역분포, 다수의 관리 도메인, 동적 특성, 자원의 불안정성 등 많은 위험 요소들로 인하여 효과적인 그리드 정보 서비스의 구축은 매우 난해하다. 그러므로 그 그리드 정보 서비스는 그 시스템 구조(System Architecture) 설계에 있어 성능, 보안, 확장성, 신뢰성 등의 보장을 위한 충분한 고려가 수반되어야 한다[9]. 그러므로 그리드 정보 서비스는 그 정보 모델 및 시스템 구조가 매우 복잡하며 제공하는 기능 또한 매우 다양하다.

이러한 다양한 기능들 중에서 자원탐색 및 할당(Resource Discovery and Allocation)과 관련된 자원 스케줄링(Resource Scheduling Support) 서비스는 그리드 정보 서비스의 핵심이라 할 수 있다[5,6,7,8]. 자원탐색 및 할당을 효율적으로 수행하기 위해서는 그리드 정보 서비스 시스템은 자원에 대한 정보를 충분히 지니고 있어야 한다. 또한 이러한 정보는 탐색 및 검색, 갱신 작업이 효율적으로 지원할 수 있도록 저장되어야 한다. 이를 위해서는 자원기술기법(Resource Description Scheme), 기술된 자원정보(Resource Information)의 저장방식, 또한 저장된 자원정보에 대한 효율적인 사용자 요청 처리 기법 등 다양한 주제에 대한 체계적인 연구가 필요하다.

위의 주제들은 그 구현과 성능 측면에서 밀접한 관련성을 지니고 있으며 상호 영향을 미치게 된다. 따라서 자원 정보의 기술, 저장, 처리와 관련된 제반 특성을 나타낼 수 있는 포괄적인 개념적 모델을 정립한 후 이를 바탕으로 세부적 주제들에 대한 연구를 수행하는 접근방식이 바람직하다. 본 논문에서는 이러한 모델을 자원정보모델(resource information model)이라고 칭한다.

현재 자원정보모델과 관련하여 다양한 연구가 수행되고 있다.[4,10,11,12,15]. 이러한 연구는

자원정보모델 측면에서 포괄적이며 체계적으로 접근하지 않고 자원정보모델의 특정 구성요소에 초점을 두고 수행되고 있어 고성능 그리드 정보 서비스의 구현에 있어서는 많은 제약을 지니고 있다.

본 논문에서는 이러한 자원정보모델의 개발방향을 제시한다. 먼저 자원정보모델의 구성요소에 대하여 분석·정의하고 이를 기반으로 한 자원정보모델을 설계 방향을 제시한다. 본 자원정보모델을 기준으로 하여 기존 자원정보모델 관련 연구의 내용을 분석하여 그 제한 및 개선 방향을 구체적으로 제시한다:

본 논문은 다음과 같이 구성된다. 2장에서는 자원정보모델(Resource Information Model)의 개념에 관하여 설명한다, 3장에서는 기존의 대표적인 자원정보모델 모델에 대한 분석 결과를 기술하며, 4장에서는 결론으로서 자원정보모델에 준한 그리드 정보서비스의 개발을 향후 연구과제로 제안한다.

II. 자원정보모델 개념

본 고의 주된 논지는 자원정보모델은 기본적으로 자원기술(Resource Description) 방식, 자원기술의 범위, 스케줄링지원(Scheduling Support) 방식, 기술 데이터 저장(Description Data Store) 방식, 사용자 요청기술(Request Description) 방식에 대한 구체적인 개념을 포함하고 있어야 한다는 점이다. 이러한 논지를 바탕으로 자원정보모델(resource information model)에 대한 체계적인 고찰을 통하여 그 개념적 구성을 아래와 같이 제안한다. 1) 자원기술(Resource Description) 범위는 자원 엔티티기술(Entity Description)과

자원 상호 간의 관계기술(Relation Description)을 필수적으로 포함해야 한다. 2) 자원정보는 최적화 자원탐색 및 자원상태 공지 등의 스케줄링 지원 요소를 제공할 수 있는 스케줄링 파라다임 하에서 설계되어야 한다. 3) 한편 자원정보의 표현모델과 저장모델은 분리하여 그 상호 종속성을 배제하는 것이 필수적으로 요청되며, 4) 사용자 자원요청 대한 높은 가용성과 편의성을 보장하기 위해서는 시스템 측면의 자원기술(Resource Description) 방식과 별도로 사용자 측면의 자원기술방식이 제공되어야 한다.

그리드 정보서비스에서 관리하고자 하는 정보는 그 대상의 범위, 유형 및 성격에 있어 매우 포괄적이며 다양하다. <표 1>은 그리드 정보시스템에서 관리되어야 하는 정보들은 의 예를 보여준다. 향후 그리드가 실용화되는 단계에서는 더 많은 유형의 대상에 대한 정보가 필요될 것으로 쉽게 예상할 수 있다. 이러한 정보를 활용하는 주체는 그리드 사용자, 관리자, 응용프로그램, 자원 스케줄러 등 매우 다양하며 그리드 정보 서비스에 의하여 통합적으로 관리되어 서비스가 제공된다.

위와 같이 그리드 정보 서비스의 대상이 되는 다양한 자원은 그 활용 방식 측면에서 크게 아래와 같은 두가지 유형으로 분류할 수 있다.

- [유형 1] 단순 쿼리를 통하여 정보를 제공하는 자원
- [유형 2] 그리드 응용 프로그램 및 미들웨어에서 탐색, 할당, 모니터링을 통하여 활용하는 자원

[유형 1]에 해당하는 자원으로는 기관, 사용자 정보 등의 정보를 예로 들 수 있다. 한편 [유형 2]에 해당하는 자원은 컴퓨터, 네트워크 등을 들 수 있다. [유형 1]의 자원에 대한 관리는 기본적으로 그리드 응용 프로그램 구동과는 관계없이

그리드 시스템의 유지 및 관리에 주로 활용된다.

<표 1> GIS 관리 정보의 유형의 예[9]

정보유형	정보의 예
기관	자원의 소유주
사람	자원 관리자, 자원 제공자, GIS 관리자
물리적 자원	컴퓨팅 자원, 메모리, CPU, cache, 벤치마크 결과, 모니터링 주기, 부팅 시간, 사용자 수, 시스템 부하
서비스	작업 관리 모듈, 부하 모니터링 모듈 등
통신자원	링크 용량, 스위칭 용량, 에러 비율, 거부 비율
소프트웨어 패키지	BALS, LAPACT
이벤트 생성자, 채널	이벤트 스트림의 생성자 및 전달매체 이벤트 유형 디렉토리
기기	레이다 시스템, 망원경
네트워크	가용한 용량, 예상 지연 시간 네트워크 호스트, 스위치, 라우터
무선기기	무선 호스트, 셀, 중계기
가상 기관	협업 그룹

본 고에서는 자원정보모델에 대한 논의를 명료하게 하기 위하여 그리드 응용 프로그램의 구동과 직접 관련이 있는 자원, 즉 그리드 상에서 탐색, 할당, 모니터링의 대상이 되는 [유형 2]의 자원들에 중점을 두고 자원정보모델을 기술한다. 참고로 [유형 1]의 자원은 기본적으로 그 정보의 저장, 관리, 및 사용자 서비스는 [유형 2]의 자원들과 비교할 때 매우 간단하므로 본 논문의 자원정보모델에 관한 논의에서 제외하여도 무방하다고 사료된다.

본 절의 나머지 소절에서는 자원정보모델의 각 개념적 구성에 관하여 주제별로 그 요구사항(Requirement Definition)을 설명한다.

2.1 자원기술 방식

앞에서 기술한 바와 같이 그리드에서 자원의 행

정적 관리뿐만 아니라 자원탐색, 할당, 모니터링 등 각종 스케줄링 작업을 위해서는 자원에 자원의 성격, 상태를 나타낼 수 있는 기술방식 (Description Scheme)이 요구된다.

일반적으로 자원기술은 선언을 사용한 방식 (Declarative)과 함수를 사용한 방식(Procedural Scheme)으로 구분할 수 있다.

자원기술방식은 위의 두 방식 중에서 어떤 것을 채택하는 가는 중요하지 않다고 사료된다. 다만 자원기술에 있어 논리적 측면에서 명료하게 이해할 수 있는 구문법(Syntax and Semantics)에 바탕을 둔 자원기술 방식이 필수적이다. 비록 물리적으로는 자원기술 내용이 매우 복잡한 경우에 위의 접근방식은 논리적 측면에서 기술된 내용을 복잡한 물리적 자동변환(Automatic Translation) 도구의 개발을 용이하게 한다.

2.2 자원기술 범위

자원기술에 있어 기술된 내용은 제반 자원에 요구되는 기능들의 성능 및 효율성과 직접 영향을 미친다. 그러므로 자원 기술의 범위의 결정은 위의 자원관리로부터의 요구사항에 대한 명확히 분석에 기초하여야 한다.

일반적으로 자원은 그 유형마다 고유한 성격과 특성을 지니고 있으므로 유형에 따라 기술되어야 하는 내용이 다르게 된다. 한편 자원할당에 있어 자원은 독립적으로 해석하는 것은 많은 제한을 갖는다. 대부분의 경우 다른 자원과의 상관관계로 해석이 된다.

자원기술(Resource Description)의 범위 측면에서 살펴보면 최소한 자원의 자체의 특성 및 성격(이후 엔티티라 칭함)에 대한 기술과 다른 자원과의 관계(Relation)에 대한 기술이 필수적이다.

즉 자원기술의 범위는 엔티티(Entity)와 관계(Relation) 기술을 포함하여야 한다(9).

각 엔티티는 정량 또는 정성적 값을 갖는 속성(Attribute)들의 집합이다. 즉 엔티티 <속성, 값의 집합이다. 각 속성들의 특성(Property)은 즉 읽기전용, 읽기/쓰기 등의 갱신(Update) 측면의 성격, 읽기/쓰기 권한, 데이터 타입 등에 대한 기술이 포함되어야 한다.

관계(Relation)는 다른 자원과의 상관성을 나타낸다. 예를 들면 조합관계(Composition Relation)는 주어진 자원이 다른 자원들의 집합으로 구성되는 관계를 나타낸다. 연결관계(Connectivity Relation)는 두 자원이 상호 네트워크로 연결되어 있음을 나타낸다.

2.3 스케줄링 파라다임

자원의 엔티티의 속성들은 자원에 관한 제반 성격을 나타내는 정보로서의 역할을 갖는다. 이러한 정보를 스케줄링 정보와 비스케줄링 정보로 구분·관리(Decoupled Management)하여야 한다. 일반적으로 스케줄링 정보는 각 자원마다 공통적으로 기술되는 공통 스케줄링 정보(Common Scheduling Information)와 각 자원에 따라 기술되는 개별 스케줄링 정보(Private Scheduling Information)로 구분된다.

공통 스케줄링 정보는 모든 자원에 대하여 공통적으로 기술되어야 하는 정보로서 할당모드 및 할당단위, 할당 후 선점형/비선점형 여부 등을 그 예로 들 수 있다. 자원의 할당모드(Allocation Mode)란 자원의 성격이 항상 배타적 할당(Exclusive allocation)되어 사용되는 경우와 이와는 반대로 여러 응용 프로그램에서 공유(Shared)하여 사용되는 것을 나타낸다. 한편 할

당단위(Allocation Unit)에 따라 전체할당(Dedicated Allocation)과 부분할당(Partial Allocation)으로 구분할 수 있다. 전체할당이란 주어진 자원의 할당 시에 그 자원 전체가 할당의 단위가 되는 것을 나타내며 부분할당이란 그 자원의 일부분을 할당할 수 있는 특성을 나타낸다. 위의 할당모드 및 할당단위에 대한 정보는 모든 자원에 대하여 공통적으로 기술되어야 하는 공통 스케줄링 정보라 할 수 있다.

한편 개별 스케줄링 정보(Private Scheduling Information)의 예로서 사용 가능한 시간과 할당 대상 등에 관한 정보가 있다.

이러한 스케줄링 정보는 실제 자원 등록자가 기술하는 부분과 자원관리모델에서 자동으로 첨가하는 속성으로 구분할 수 있다.

2.4 자원정보 표현모델과 저장모델의 구별

주어진 자원기술(Resource Description) 방식은 그 방식에 준하여 기술된 데이터의 표현 및 저장방식의 구조에 큰 영향을 미친다.

자원정보 표현모델은 크게 계층모델(Hierarchical Model), 객체 지향 모델(Object Oriented Model), 관계형 모델(Relational Model), 네트워크 모델(Network Model)로 구분할 수 있다 [10,11,12,14]. 각 모델은 자원기술데이터의 표현에 대하여 표현 범위, 저장구조의 효율성, 스케줄링 지원의 효율성 등의 측면에서 상호 장단점을 지닌다.

위의 모델들은 표현 측면에 있어 상호 다른 특성을 지니고 있다. 상호 계층모델은 계층적 구성을 갖는 자원들의 조합을 매우 자연스럽게 표현할 수 있는 모델이라는 장점을 갖는다. 객체 지향 모델은 상속(Inheritance) 특성을 이용하여 기술데

이터의 표현의 확장 능력이 매우 우수하다. 관계형 모델은 다양하고 복잡한 자원들 간의 관계를 표현하는데 매우 적합한 장점을 갖는다. 네트워크 모델은 자원들 간의 다양하고 복잡한 관계를 적절하게 표현할 수 있다는 점에서 관계형 모델과 유사하며 관계형 모델이 정규적인 관계 표현에 적합한 반면 네트워크 모델은 비정규적 관계에 대한 표현에 적합하다.

한편 위의 모델들은 자원기술 데이터의 저장 구조 및 자원탐색 최적화 알고리즘 등의 데이터 처리 측면에서 상호 다른 특성을 지니고 있다. 계층 모델은 레퍼럴(Referral)을 활용하여 LDAP 기반으로 구현이 가능하지만 데이터의 갱신 및 조인(Join) 기능의 부재 등의 쿼리 표현의 제약 그리고 자원탐색을 위한 최적화 알고리즘 구현에 있어 난해성 등의 단점을 지니고 있다[4]. 한편 관계형 모델과 객체 지향 모델은 기존의 상용 데이터베이스를 활용하여 매우 효율적으로 데이터를 저장·관리할 수 있지만 그리드 자원탐색에 있어 데이터베이스 쿼리(Query)가 조건에 부합하는 모든 결과(All Solution)를 찾으므로 초래되는 성능저하의 단점을 갖는다[14]. 한편 네트워크 모델은 데이터 저장 및 데이터 처리에 곧 바로 부합하는 상용 데이터베이스가 없다. 그러므로 이 모델은 기존 상용 데이터베이스를 활용하기 위한 저장 기술이 개발하거나 독립적인 데이터 저장 기술을 개발하여야 하는 단점을 갖는다. 한편 네트워크 모델은 비결정적 단일 결과(Non-deterministic single solution)를 검색 및 각종 유형의 최적화 탐색 알고리즘을 가장 유연하게 개발할 수 있는 기반을 제공한다.

현재 그리드 정보 서비스에 대한 기술은 매우 초보적인 수준으로서 전체적인 구현의 가능성(Feasibility) 검증 단계의 연구에 집중하고 있

다. 즉 성능 측면에서의 최적화 기술 부분에 대한 연구는 아직 착수되고 있지 않다. 그러므로 향후 자원탐색(Resource Discovery), 여러 자원의 동시 할당(Co-allocation), 자원 예약(Reservation) 등에 대한 성능 및 효율 측면의 최적화 알고리즘에 대한 연구 결과가 없이는 위의 각 모델에 대한 장 단점을 정확히 도출할 수 없다[4].

기존의 자원정보모델은 자원정보의 표현모델과 자원정보 저장모델에 대한 구별이 없이 주로 데이터 저장 모델에 중점을 두고 개발되고 있다. 이러한 접근방식은 기본적으로 자원기술 범위 및 내용에 대한 제한을 초래할 수 있으며 나아가 특정 데이터 저장 모델을 가정함으로써 향후 그 데이터 저장 모델에 종속되어 위에서 지적한 바와 같이 자원탐색 등 각종 최적화 알고리즘의 개발에 한계에 직면할 수밖에 없게 된다.

그러므로 자원정보모델에서는 자원정보의 표현모델과 저장모델을 상호 독립적인 개념으로 구분할 필요가 있다. 이러한 자원정보모델 상에서는 자원정보 표현모델은 저장모델에 대하여 추상화(Abstraction)된 형태를 갖게되어 주어진 자원정보 표현모델이 다양한 자원정보 저장모델에의 매핑을 통한 구현의 용이성, 이식성 및 성능을 제고할 수 있다.

2.5 사용자 자원기술과 시스템 자원기술의 분리

시스템 측면에서 보면 스케줄러는 자원탐색, 할당 등의 작업을 효율적으로 수행을 위하여 논리적으로 기술되어 저장된 다양한 자원 정보를 사용한다. 앞에서 논의한 바와 같이 자원에 대하여 기본적인 사양, 스케줄링 측면의 특성, 다른 자원과의 관계, 갱신이 필요한 정보의 갱신 방법 등을 논리적으로 표현하기 위해서는 매우 난해한 구문법

(Syntax & Semantics)을 갖는 기술방식이 요구된다.

사용자 측면에서 보면 그리드의 응용프로그램은 스케줄러에 필요한 자원 요청을 통하여 해당 자원을 할당받아 사용한다. 즉 자원요청을 위한 기술방식이 요구된다. 이러한 사용자 자원요청에 대한 자원기술방식은 위의 시스템 측면에서의 자원기술방식을 그대로 활용할 수 있다. 그러나 이러한 접근방식은 앞에서 지적한 바와 같이 시스템 측면의 자원기술방식이 매우 복잡하므로 사용자측면에서의 자원요청이 매우 복잡하고 난해한 작업이 된다.

따라서 시스템 측면의 자원기술방식과 사용자 자원요청에 대한 기술방식의 상호 분리가 필요하다. 이러한 분리는 자원기술 측면에서는 자원기술에 중점을 둔 효율적인 기술방식을, 자원요청 측면에서는 사용자의 편의성을 최고를 보장할 수 있는 기술방식을 각각 도출할 수 있다. 이러한 분리는 사용자 자원요청 기술방식을 시스템 내의 자원기술방식으로 효율적으로 변환할 수 있는 기법에 대한 고려도 함께 요구된다.

III. 기존 자원정보모델 분석

3.1 기존 자원정보모델 소개

CCC의 RSD[10]에 채택되고 있는 자원정보모델은 그리드 자원을 그래프로 표현하는 계층적 구조와 네트워크 구조를 조합한 표현모델을 기반으로 한다. RSD 기반에서는 자원그래프는 노드, 포트, 가지(Edge)로 구성된다. 노드는 컴퓨팅 자원을 등을 나타내며 재귀적(Recursively)으로 표현할 수 있다. 포트는 통신을 위한 게이트웨이 또는 네트워크 접속장치(Adapter) 등을 나타낸다. 가

지(Edge)는 노드와 노드간의 통신회선을 나타낸다. 자원관리와 자원요청을 위한 사용자 API, 그래픽 편집기 그리고 자원의 텍스트 기술을 위한 RSD(Resource and Service Description) 언어(Language) 제공하고 있다.

Globus의 자원정보모델에서는 RSL(Research Specification Language)을 기반으로 하는 사용자 자원기술방식을 제공하며 자원 데이터 표현과 저장 모델은 LDAP 기반의 계층적 트리 모델을 기반으로 한다[11].

Condor의 자원정보모델은 ClassAd를 자원기술방식으로 제공하고 있으며 기술된 자원에 대하여 즉 자원탐색 등의 스케줄링 관련 각종 효율적인 알고리즘에 대한 연구결과를 산출하고 있다[12,13].

관계형 접근방식의 자원정보모델은 자원기술(Description)에 있어 관계기술(Relation Description)의 중요성을 인식하여 관계형 데이터 모델을 채택하여 자원기술을 수행하고자 하는 접근방식이다[14,15]. 현재 이 접근방식은 구현가능성(Feasibility)의 검증 단계에있으며 실제 구현을 통한 가시적인 성능 평가에 대한 결과는 아직 산출되지 않고 있다.

3.2 자원기술범위 및 방식

CCC의 RSD 자원기술방식은 단지 노드와 노드간의 통신회선 연결 관계와 재귀적 노드로 표현되는 자원의 계층 구조의 관계만을 기술할 수 있다. 그러므로 자원에 대한 엔티티기술(Entity Description) 매우 구체적이지만 관계기술(Relation Description)에 있어서는 제한성을 지니고 있다. Globus의 RSL 자원기술방식은 기본적으로 사용자 측의 기술방식이며 엔티티기술은 매우 구체적

이지만 관계기술에 대해서는 제한성을 갖는다. Condor ClassAd 자원기술방식은 엔티티기술은 매우 구체적이지만 관계기술에 대한 개념은 제공하지 않는다. 관계형 접근방식의 자원에 대한 엔티티기술(Entity Description)과 관계기술(Relation Description)을 매우 구체적으로 기술할 수 있는 장점을 갖는다.

〈표 2〉 기존 자원정보모델 분석요약

정보유형	CCC RSD	Globus RSL	Condor ClassAd	Relational GIS
관계기술	△	△	x	o
스케줄링 패라다임 고려	x	x	△	x
표현모델과 저장모델의 구분	o	x	o	x
사용자 기술방식	x	o	x	x

3.3 스케줄링 패라다임

일반적으로 기존의 대부분의 자원정보모델에서는 스케줄링에 대한 구체적인 고려를 하고 있지 않으며 자원기술에 있어 스케줄링 관점에서 자원을 해석한 스케줄링 정보와 이에 해당하지 않는 정보 구분에 대한 개념이 지원되지 않는다.

CCC, Globus, 그리고 관계형 접근방식의 자원정보모델에서는 최적화 자원탐색 등의 제반 스케줄링 패라다임(Paradigm)에 대한 고려 부재로 스케줄링과 관련된 최적화 기술 개발에 심각한 제한을 받게된다. Globus의 경우를 살펴보면 최적화 자원 탐색(Optimized discovery service) 서비스를 위해서는 쿼리에서 조인(Join) 기능이 필수적이지만[4] 현재 LDAP 쿼리는 조인(Join) 기능을 제공하고 있지 않으므로 Globus는 스케줄

링 패라다임에 제한을 지니고 있다. 관계형 접근 방식의 경우, 상용 데이터베이스에서 단일 결과(Non-deterministic single solution)를 산출할 수 있는 쿼리 지원 부재로 인한 과부하로 초래되는 성능저하 문제 및 자원탐색 알고리즘에 대한 최적화 알고리즘의 개발에 문제에 대한 해결책이 필요하다.

Condor의 자원정보모델은 제한적이지만 스케줄링 패라다임에 대하여 고려를 하고 있다. 본 모델에서는 최적화 자원 탐색(Optimized discovery service) 서비스를 위한 매치메이커(Matchmaker) 알고리즘과 검색의 성능향상을 위한 인덱싱 알고리즘에 연구 결과를 제공하고 있다.

3.4 자원정보 표현모델과 저장모델의 구별

CCC의 RSD와 Condo의 ClassAd 자원정보모델은 자원정보의 물리적 저장모델과 독립적인 자원정보 표현모델을 제공하고 있다. CCC RSD의 경우 자원의 사양 및 관계를 기술할 수 있는 자원 기술 모델과 기술언어를 제공하므로 특정 저장모델에 종속되지 않는다. 한편 Condo의 ClassAd의 경우는 비록 매우 단순한 형태지만, 자원기술은 스키마(Schema)를 기반으로 하고 있으므로 특정 데이터 모델에 종속되지 않는다.

Globus와 관계형 접근방식의 자원정보모델에서는 자원정보에 대한 표현모델과 저장모델 간의 구별을 두고 있지 않다. Globus의 경우 LDAP 기반의 계층적(Hierarchical Directory Tree) 구조의 데이터모델과 저장구조에 종속된다. 관계형 접근방식의 경우에는 관계형 데이터베이스 스키마 표현 및 저장모델에 종속된다.

3.5 사용자 자원 요청 기술방식

Globus 자원정보모델은 사용자 자원기술체계와 시스템 자원기술체계를 분리하여 제공한다. Globus 자원정보모델에서는 사용자 측면에서는 자원기술을 위하여 RSL(Resource Specification Language)을 제공하고 있으며 시스템 자원기술 체계로서 LDAP의 X.500 데이터 기술방식의 제공한다.

RSD, Condor, 관계형 접근방식의 자원정보모델은 사용자 요청을 자원기술과 시스템 측면의 자원기술을 구분하여 제공하지 않는다. RSD의 경우 비록 그래픽 편집도구를 제공하지만 자원할당을 위한 사용자 측면에서 구체적인 자원기술체계를 할 수 없다. 실제로 RSD API는 주변노드 정보 탐색, 노드 내용 보기 등 사용자가 자원요청시에 시스템 자원기술 내용을 이해가 선행되어야 한다. Condor의 경우는 사용자 자원기술체계와 시스템사용체계 모두 ClassAd Schema를 사용한다. 한편 관계형 접근방식은 사용자 자원기술 체계 사용하는 SQL 문장 작성을 위해서는 시스템 자원기술 체계인 관계형 데이터베이스 스키마를 정확한 이해를 필요로 한다.

IV. 결 론

본 고에서는 그리드 정보 서비스에서 요구되는 자원정보모델(resource information model)에 대한 체계적인 고찰을 통하여 그 개념적 구성을 제시하였다. 이 자원정보모델의 개념적 구성을 바탕으로 기존의 대표적인 자원정보모델들을 분석한 결과 기존 모델은 대부분 위와 같은 요구사항을 충분히 만족시키지 못하고 있음을 보였다. 본 자원정보모델의 개념은 고성능 그리드 정보 서비스

시스템의 설계 및 구현에 있어 필수적으로 반영되어야 한다고 사료되며 향후 고성능 그리드의 실현을 위해서는 제안하는 자원정보모델의 개념을 기반으로 하여 자원기술언어, 자원정보 표현기법, 자원정보 저장기법, 사용자 요청 기술기법에 관한 구체적인 연구가 요청된다.

감사의글

본 논문은 2001년 대학기초연구지원사업(2001-011-3)에 의하여 지원을 받았음.

참고문헌

- [1] I. Foster and C. Kesselman, "The Grid: Blueprint for a New Computing Infrastructure", Morgan-Kaufmann, 1999.
- [2] I. Foster, et al., "The Anatomy of GRID Architecture", J. of Supercomputing Applications, 2001
- [3] K. Czajkowski et al., "A Resource Management Architecture for Metacomputing Systems", In Proc of 4th Workshop on Job Scheduling strategies for Parallel Processing, pp 62-82, 1998
- [4] K. Czajkowski, et al., "Grid Information Services for Distributed Resource Sharing", In Proc. of 10th IEEE International Symposium on High Performance Distributed(HDPC-10), 2001
- [5] M. Harchol-Balter, T. Leighton, and D. Lewin, "Resource discovery in distributed networks", ACM Symposium on Principles of Distributed Computing, May 1999, pp. 229-237.
- [6] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in Wireless sensor networks", ACM Mobicomm, 1999, pp. 174-185.
- [7] H. Casanova, M. Kim, J. S. Plank and J. Dongarra, "Adaptive scheduling for task farming with Grid middleware", International Journal of Supercomputer Applications and High-Performance Computing, pp. 231-240, Volume 13, Number 3, Fall 1999.
- [8] L. A. Hall, "Approximation algorithms for scheduling", in Dorit S. Hochbaum (ed.), "Approximation algorithms for NP-hard problems", PWS Publishing Company, 1997.
- [9] B. Plale, et al., Key Concepts and Services of a Grid Information Service, GWD-GIS-018
- [10] A. Keller, A. Reinefeld, "Anatomy of a Resource Management System for HPC Clusters", Annual Review of Scalable Computing, vol 3, pp 1-31, 2001
- [11] I. Foster and C. Kesselman, "The

Globus project: A status report", 1998 IEEE Heterogeneous Computing Workshop (HCW '98), 1998, PP. 4-18.

- [12] M. J. Litzkow, M. Livny, and M. W. Mutka, "Condor - A hunter of idle workstation, "8th International Conference on Distributed Computing Systems, 1988, pp.104-111.
- [13] R. Raman, M. Livny, and M. Solomon, "Matchmaking: Distributed resource management for high throughput computing", 7th IEEE International Symposium on High Performance Distributed Computing, 1988, pp. 28-31.
- [14] W. Dinda, B. Plale, "A Unified Relational Approach to Grid Information Services", Grid Forum Informational Draft GWD-GIS-012-1
- [15] W. Fisher, "Relational Model Approach to Grid Information Services", Grid Forum Informational Draft GWD-GIS-012-1



김희철

1983년 연세대학교 전자공학과 졸업(공학사)

1991년 Univ. of Southern California (Computer Eng. M.S.)

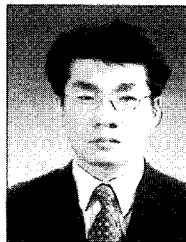
1996년 Univ. of Southern California (Computer Eng. Ph.D.)

1983년 - 1988년 (주)삼삼전자 주임연구원

1996년 - 1997년 (주)삼성SDS 수석연구원

1997년 - 현재 대구대학교 정보통신학부 조교수

관심분야 : 병렬처리, 컴퓨터구조, 컴파일러



이강우

1985년 연세대학교 전자공학과 졸업(공학사)

1991년 Univ. of Southern California (Computer Eng. M.S.)

1997년 Univ. of Southern California (Computer Eng. Ph.D.)

1998년 - 현재 동국대학교 정보통신학부 조교수

관심분야 : 병렬처리, 컴퓨터구조, 컴파일러



이 용 두

1975년 한국항공대학교 통신학과 졸업(공학사)

1982년 영남대학교 대학원 전자공학과(공학석사)

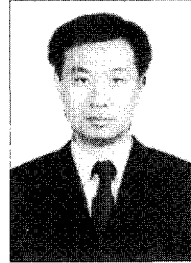
1995년 한국항공대학교 대학원 전자공학과(공학박사)

1982년 - 현재 대구대학교 정보통신공학부 교수

1981년 - 1982년 (일)동경대학 전자공학과 객원 교수

1991년 - 1993년 Univ. of Southern California 교환교수

관심분야 : 컴퓨터구조, Internet 응용 기술



조 세 흥

1983년 연세대학교 3년 수료

1991년 캘리포니아주립대 (California State Univ.)

컴퓨터 공학 이학사

1996년 애리조나주립대

(Arizona State Univ.) 컴

퓨터 공학 석사, 멀티미디어 전공

1999년 애리조나주립대 컴퓨터 공학 박사, 멀티미디어 전공

1999년~현재 대구대학교 정보통신공학부 교수

관심분야 : 멀티미디어 시스템, 멀티미디어 응용프로그램 개발, 가상현실 시스템, 인터넷 응용프로그램 개발, 원격교육, 가상대학 등