

XMDR을 이용한 데이터 그리드 미들웨어 설계

노선택^o 문석재 엄영현 국윤규 정계동 최영근

광운대학교 컴퓨터 과학과

{pearl, msj1568, class76, ykkook, gdjung, ygchoi}@kw.ac.kr

A Design of Data Grid MiddleWare Using XMDR

Seon-Taek Noh^o, S.J. Moon, Y.H. Eum, Y.G. Kook, K.D. Jung, Y.K. Choi

Department of Computer Science, KwangWoon University

요약

데이터 그리드는 분산 데이터베이스 환경에서 상호간 데이터의 공유 및 교환을 통하여 분산 데이터를 효율적으로 사용하기 위한 기술이다. 최근에는 레거시 시스템간의 협업을 하기 위해 상호 연결의 필요성이 강조되고 있으며 많은 연구가 진행되고 있다. 독립적인 레거시 시스템을 상호 연결하기 위해서는 노드간의 회선 점유율이 증가하고, 데이터의 이질적인 문제점이 발생한다. 이러한 문제점을 해결하기 위해 본 논문에서는 XMDR 기반의 데이터 그리드 시스템을 제안 한다. 본 논문의 데이터 그리드 미들웨어는 ISO/IEC 11179에서 제안한 정보 공유 및 교환을 위한 표준인 MDR과 정보의 효율적인 이용을 위한 온톨로지를 포함하여 확장한 개념인 XMDR(eXtended Metadata Registry)에 기반을 두었으며 분산 플랫폼으로부터 독립적이고 자유로운 이동성을 가지고 있는 이동 애이전트 기술을 사용하였다.

1. 서론

지리적으로 분산되어 있는 시스템에 대한 통합이 이루어지고 데이터의 재사용 필요성이 증가할수록 상호운용성의 확보 문제는 매우 중요해졌다. [9] 이기종 시스템들 간의 정보의 통합이나 교환을 의미하는 상호운용성은 효율적인 분산 데이터의 활용을 위한 데이터 그리드의 한 분야로서, 이에 대한 연구가 활발히 진행되고 있다. 특히 최근 들어 오라클사의 Oracle Database 10g를 선두로 데이터베이스 기반의 데이터 그리드에 대한 많은 연구가 진행되고 있다. [6]

데이터 그리드는 넓게 분산된 환경에서 수백만 기가바이트에 달하는 거대한 규모의 데이터베이스 환경에서의 효율적인 공유 및 분석이 가능하도록 하는데 목적이 있다. 기존의 데이터 그리드에 대한 연구들이 데이터가 어디에 저장되는지에 대한 문제에 초점을 맞춘 반면 최근에는 레거시 시스템에서 독립적으로 운영되는 데이터베이스를 상호 연결하는 필요성이 강조되고 있다. [2] 분산된 데이터베이스 환경에서의 상호운용을 하기 위해서는 네트워크상의 물리적인 문제뿐만 아니라 레거시 시스템에 존재하는 이질적인 데이터의 문제점도 해결해야 한다. [1]

본 논문에서는 이러한 문제점을 해결하기 위해 이동 애이전트 기술과 XMDR를 이용한 그리드 미들웨어를 제안한다. 이동 애이전트는 기준의 원격함수호출(RPC)과 다른 개념인 원격 프로그래밍(Remote Programming)의 개념을 이용하여 애이전트를 목적하는 원격 호스트로 이동시켜 그곳에서 수행하도록 한다. [3] 이는 항상 통신이 연결될 필요 없이 이동 애이전트를 보낼 때나 결과를 받을 때만 연결되게 함으로써 이동 통신 체계에於에서 문제가 되었던 통신 회선의 정유 부담을 감소시킬 수 있다. 또한 정보 공유 및 교환을 위한 표준인 MDR과 온톨로지를 포함한 XMDR 기반으로 설계함으로써 레거시 시스템의 데이터를 통합함으로써 분산 데이터베이스 환경에서 존재하는 노드와 데이터 간의 이질적인 특성을 해결했다.

2장에서는 데이터 그리드, MDR, 이동 애이전트에 대하여 살펴보고, 3장은 이동 애이전트와 XMDR을 이용한 데이터 그리드 미들웨어의 소개와 구현에 대한 설명을 하며, 4장에서 결론을 서술한다.

2. 관련 연구

2.1 데이터 그리드

데이터 그리드는 컴퓨터 네트워크로 연결되어 있는 복수의 조직이 소유한 스토리지 시스템들이 서로 역할을 나누어 데이터들을 분산 보유하고 효율적으로 사용하기 위한 기술이다. 관련된 프로젝트로는 영국 e-사이언스센터나 IBM, ORACLE이 중심이 되어 원격 데이터베이스 접근과 분산통합처리를 행하는 미들웨어를 개발하는 OGSA-DAI(Open Grid Services Architecture Data Access and Integration) 프로젝트가

있다. 데이터 그리드의 연구동향을 분야별로 살펴보면, 독립된 저작자들을 하나의 저장 장치로 통합 및 접근, 속성 기반의 데이터 접근, 에타데이터 마이닝을 포함한 정보 검색 기술, 분산된 데이터들의 위치를 표현하는 방법, 데이터 복제 관리 및 복제된 데이터의 일관성 유지, 데이터 공유 및 보호, 데이터 접근 관리 등이 있다. [7]

2.2 MDR(Metadata Registry)

MDR은 메타데이터 포맷의 다양성으로 인해 발생하는 메타데이터 간 의미와 형식의 이질성 문제를 해결하기 위해 ISO/IEC 11179에 의해 개발되었다. ISO/IEC 11179의 기본 단위는 데이터 요소로서 하나의 데이터에 대한 표현, 이해 및 활용을 위해 필요한 다양한 메타데이터로 구성되어 되어 있다. 이러한 데이터 요소의 집합이 MDR(Metadata Registry)이다. MDR에 의해 관리되는 데이터의 기본 단위는 데이터 요소이다. 하나의 데이터 요소는 하나의 의미를 나타내기 위해서 사용되며, 각 데이터 요소는 유일한 식별자를 가지고 있다. 따라서 MDR은 데이터 요소를 통해 사용자에게 데이터에 대한 유일하고 정확한 의미를 제공함으로써, 데이터 요소간의 의미상 동등성을 확보할 수 있다. [8]

2.3 이동 애이전트

이동 애이전트는 네트워크를 통하여 미리 정해진 이주경로를 바탕으로 이주하거나 자율적으로 이주하며 작업을 수행하는 객체를 말한다. 이동 애이전트는 플랫폼에 독립적인 성격을 가지고 있기 때문에 애이전트 시스템을 이용한 그리드 미들웨어를 설치함으로써 이질적인 분산 데이터베이스 환경에서의 수행에 대한 자율성을 보장 할 수 있다. 또한 작업을 요청하기 위해 이동 애이전트를 전송한 후에 연결을 계속 유지할 필요 없이 요청에 대한 수행 결과만 연결을 하기 때문에 네트워크 통신에서 발생하는 통신 회선 경유에 대한 부담도 덜 수 있다.

3. 데이터 그리드 미들웨어 및 구현

3.1 데이터 그리드 미들웨어 개요

본 논문에서 제안하는 데이터 그리드 미들웨어로 구성할 수 있는 데이터 그리드 커뮤니티는 그림 1과 같다. 기존의 레거시 시스템들의 협업을 위해 형성된 커뮤니티는 파일 공유, 분산 컴퓨팅, 데이터 공유 및 교환, 정보 검색 등을 향으로써 공동 작업을 수행할 수 있다.

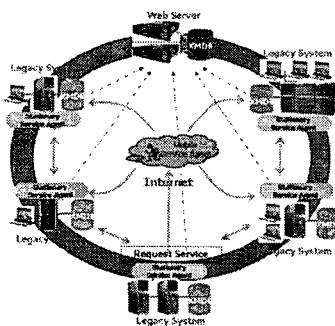


그림 1. 데이터 그리드 커뮤니티 개요

커뮤니티에 속해 있는 노드가 서비스를 요청하게 되면 이동 에이전트를 통하여 각 노드로 필요한 작업 수행을 처리하게 된다. 이후된 이동 에이전트가 수행한 결과 같은 각 노드에 상주하고 있는 상주 에이전트(stationary service agent)가 처리한다. 상주 에이전트는 작업 수행의 특성에 따라 최초 작업 수행을 요청한 노드나 다음 이동 에이전트의 이후 노드로 결과 값을 반환하게 된다. 웹서버는 커뮤니티에 등록된 노드의 정보와 본산 데이터의 상호운용을 위해 정의한 XMDR를 관리한다. 웹서버는 노드들의 자원과 상태 등의 정보를 주기적으로 각 노드들로부터 전달받게 되며 수시로 갱신을 하게 된다. 따라서 웹서버는 이후 경로로 서비스를 요청한 노드가 수행해야 할 작업에 따라 동적으로 경로를 설정해 최적의 이후 경로를 요청한 노드에게 전달 할 수 있다. 그림 2는 웹서버의 노드정보 관리 화면과 이동 에이전트의 이후 경로를 결정하는 화면이다.

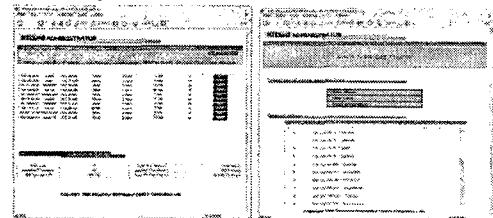


그림 2. 데이터 그리드 시스템의 웹 관리화면

3.2 XMDR(eXtended Metadata Registry)

XMDR은 ISO/IEC11179에서 제시한 정보 교환을 위한 표준인 MDR과 온톨로지와 포함한 확장된 개념이다. 본 논문에서는 데이터 그리드를 구성하는 레거시 시스템들의 데이터베이스를 기반으로 MDR를 구성했다. 구축된 MDR은 ISO/IEC 11179-3에서 제안한 데이터의 속성 명세를 따른다. 온톨로지는 위치 온톨로지와 표준 온톨로지로 나누어 구축하였다. 그림 3은 본 논문에서 정의한 XMDR의 구축과정이다.

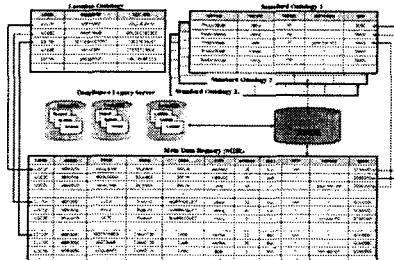


그림 3. MDR과 온톨로지 구축과정

분산 데이터의 효율적인 접근을 위한 위치 온톨로지는 위치 식별을 위한 LOCID와 위치 정보인 LOC_URL, 데이터베이스 정보를 속성으로 갖는다. 데이터의 공유 및 교환을 위한 표준 요소를 설정하는 표준 온톨로지를 구축한다. 표준 온톨로지는 협업을 위한 공유를 위하여 그룹

으로 구성하고, 메타데이터의 의미적인 내용과 데이터 요소의 구문을 위한 표준을 제시한다. 각 레거시 시스템에서 사용하던 데이터의 이질성을 극복하기 위하여 표준 용어와 자료형, 크기, 표현 방법에 대하여 제시한다.

3.3 데이터 그리드 미들웨어 구조

본 논문에서 제안하는 데이터 그리드 미들웨어는 4계층으로 구성된다. 각 계층은 관리 모듈과 응용 계층으로 이루어져 있으며 관리 모듈은 각각 네트워크 계층, 시스템 계층, 서비스 계층으로 이루어져 있다. 그림 4는 데이터 그리드 미들웨어의 구조도이다.

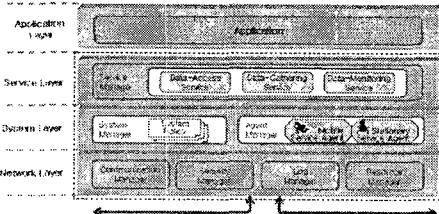


그림 4. 데이터 그리드 미들웨어 구조

데이터 그리드 미들웨어의 관리모듈에 구성되어 있는 기본적인 기능과 설명은 다음과 같다.

- 통신 관리 : 노드 간의 통신 형성 관리를 담당하는 모듈이다. 본 미들웨어는 P2P 기술[5]을 바탕으로 구성되어 클라이언트와 서버의 역할을 동시에 수행하는 시스템이다. 미들웨어가 구동됨과 동시에 웹서버에 자신의 자원 정보를 전송하게 되며 웹서버로부터 받는 서비스와 수행에 필요한 에이전트를 전송하고 전송받는 부분을 관리하게 된다.
- 자원 관리 : 노드의 프로세서, 메모리, 네트워크와 하드 디스크의 정보 및 데이터베이스의 정보를 감시한다. 그림 5는 노드의 정보를 감시하는 부분을 나타내는 데이터 그리드 미들웨어의 화면이다.

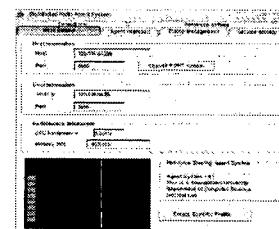


그림 5. 데이터 그리드 미들웨어의 자원관리

- 보안 관리 : 데이터 그리드 커뮤니티에 구성된 노드들은 인증된 에이전트만으로 통신하여 작업을 수행하게 된다. 노드의 인증 정보는 웹서버를 통해 인증받게 되며 수행을 담당하는 이동 에이전트의 인증처리는 노드부분에서 직접 관리하게 된다. 본 논문에서 제안하는 데이터 그리드 미들웨어는 JAVA로 이루어져 있기 때문에 JAVA에서 제공하는 보안 모듈을 사용하였다.
- 시스템 관리 : 협업에 관한 시스템 정책을 관리하는 모듈이다. 자원 관리자에 의해 수집된 노드의 정보들을 웹서버에 전송하여 작업 수행에 필요한 이후 경로를 설정할 수 있도록 한다.
- 에이전트 관리 : 자신이 작업 요청하는 에이전트와 다른 노드로부터 요청받은 에이전트를 처리하는 모듈이다. 서비스 관리자에 의해 형성된 서비스 에이전트의 작업수행, 라이프사이클, 에이전트의 이후 경로 관리, 작업우선 순위 관리, 결함 허용 정책을 관리한다. 그림 6은 에이전트 관리부분 중 에이전트의 작업 우선순위 관리를 나타내는 화면과 서비스 에이전트의 작업 수행을 선택하는 화면이다.

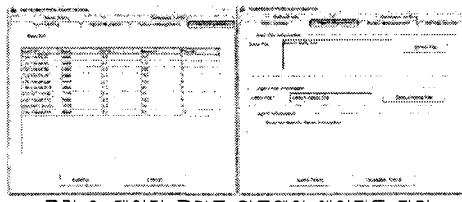


그림 6. 데이터 그리드 마일웨어 에이전트 관리

• 로그 관리 : 데이터 그리드 마일웨어를 통해 발생한 노드의 상태 정보, 에이전트의 송수신 정보, 결함등의 모든 이벤트를 기록, 관리, 담당 한다.

• 서비스 관리 : 데이터 그리드 커뮤니티에서 수행하는 작업을 형성하는 서비스를 관리한다. 데이터 그리드 마일웨어에서 제공하는 서비스는 데이터 접근 서비스 (DAS:Data Access Service), 데이터 수집 서비스(DGS:Data Gathering Service), DMS(Data Monitoring Service)이다. 본 서비스를 제공하기 위해서는 3.2절에서 설명한 XMDR을 이용한다.

4. 성능 평가

본 논문에서 제안한 데이터 그리드 마일웨어의 성능 측정을 위해 다음과 같은 실험을 수행하였다.

1) 실행 환경

- Windows Server 2003 / Windows XP professional : 7대
- Linux Fedora Core4 : 3대
- MS-SQL, Oracle8i
- 웹서버 구축 언어 : JSP

2) 실행 방법

- 특정 자산에 대한 검색(자산번호, 자산명, 단기동)
- 조건에 맞는 검색 평균 응답시간
- 에이전트 최종 도달 시간 측정

10개의 노드로 데이터 그리드 커뮤니티를 구성하여 최초 2개의 노드 수행에서 측정된 처리시간을 기준으로 노드를 두 개씩 증가시키며 측정하였다. 본 데이터 그리드 마일웨어에서 사용하는 에이전트의 성능 측정을 위한 것이므로 각 노드에서 검색되는 데이터의 크기는 같다고 가정하고 측정하였다. 결과는 그림 7-1과 그림 7-2와 같다.

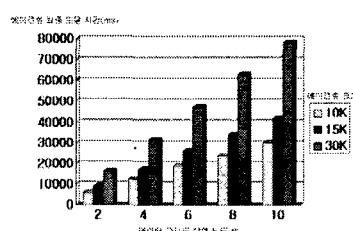


그림 7-1 에이전트 크기에 따른 최종 도달시간

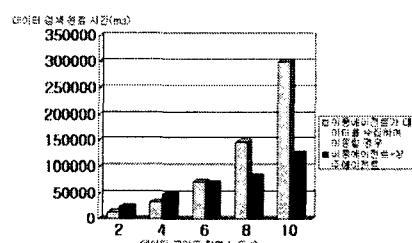


그림 7-2 에이전트를 이용한 데이터 검색 완료시간

웹서버에서 받은 이주표만을 가지고 이동하게 되는 에이전트의 최소 크기는 10K이다. 그림 7-1은 5K 크기의 메타데이터를 가지고 이동하는 에이전트와 20K 크기의 메타데이터를 가지고 이동하는 에이전트가 모든 노드를 거쳐 최초 전송한 노드로 돌아오게 되는 시간을 나타내고 있다. 이동 에이전트는 단지 검색정보에 관한 메타데이터를 목적지 노드에 전달만하고 다음 이주 노드로 이동하기 때문에 검색 데이터량이 증가해도 각 노드간 전송 시간의 차이는 항상 일정하다. 그림 7-2는 이동 에이전트가 각 노드에서 검색된 데이터를 포함하여 최초 요청한 노드로 도달하는 방식과 이동 에이전트는 검색정보에 관한 메타데이터를 목적지 노드에 전달하고 각 노드에서 검색된 결과값은 상주에이전트를 이용하여 전송하는 방식을 비교하여 나타낸 결과이다. 이동 에이전트가 검색 결과를 포함하여 각 노드로 이동할 경우에는 노드 수가 증가할수록 수행시간이 데이터량 누적에 비례하여 증가하는 것을 확인할 수 있다. 이는 각 노드에서 최소 한 개 이상의 이동 에이전트가 수행된다는 특성을 감안할 때 비효율적인 수행결과를 가져올 수 있다. 본 논문에서 제안한 데이터 그리드 마일웨어는 이동에이전트와 상주에이전트를 상호적으로 사용하여 각 노드와의 작업을 비동기적으로 수행함으로써 요청에 대한 수행시간 단축과 각 노드간의 통신 부담을 최소화 할 수 있다.

5. 결론

분산된 데이터베이스 환경의 데이터 상호운용을 목적으로 구성한 데이터 그리드는 각각 다른 환경의 레거시 시스템의 독립성과 자율성을 보장해야 한다. 따라서 본 논문에서 제안한 데이터 그리드 마일웨어는 에이전트 시스템을 이용하여 네트워크의 회선 경유 부하를 줄임과 동시에 플랫폼 독립성을 보장하였다. 또한, 데이터의 공유 및 교환을 위한 표준으로써 XMDR을 사용하여 데이터의 투명성과 자율성을 보장하도록 하였다. 또한 커뮤니티에 속한 노드들의 자원을 실시간으로 관리하고 수행을 위한 이동 에이전트의 경로 설정을 통적으로 구성해서 노드에게 서비스하는 웹서버를 이용함으로써 수행 작업의 효율성을 높이고 노드의 작업 부하를 줄일 수 있다.

향후 연구로 데이터 그리드 커뮤니티에서의 데이터 상호운용을 위한 메타데이터에 대한 연구와 웹서비스의 연동과 웹서버의 포탈화를 통해 확장된 서비스 제공을 위한 연구가 필요하다.

참고 문헌

- [1] Amit Sheth, Kartic Ramakrishnan and Christopher Thomas,"Semantics for the Semantic Web: The Implicit, the Formal and the Powerful ",Int'l Journal on Semantic Web & Information Systems, 1(1), 1-18, Jan-March 2005
- [2] DataGrid <http://eu-datagrid.web.cern.ch/eu-datagrid>
- [3] Gifford D. and Stamos J., "Remote evaluation," ACM Trans. on Programming Language and Systems, Vol. 12, No. 4, pp 537-565, 1990.
- [4] ISO/IEC JTC 1/SC 32, "ISO/IEC 11179: Specification and standardization of data elements, Part 1-6" 2003
- [5] Karl Aberer, Philippe Cudré-Mauroux, Anwitana Datta, Zoran Despotovic, Manfred Hauswirth, Magdalena Punceva, Roman Schmidt, "P-Grid: A Self-organizing Structured P2P System", SIGMOD Record, 2003
- [6] Matthew Hart, Scott Jesse, "Oracle Database 10g: High Availability with RAC, Flashback & Data Guard", Oracle Press, 2005
- [7] Sheau-Yen Chen, et al."SRB Tutorial 2002". 2002, NPACI All Hands Meeting, March 2002, <http://www.sdsc.edu/srb/Tutorials/>
- [8] 김진관, 김중일, 정동원, 백두권 "메타데이터 이질성 해결을 위한 MDR 기반의 애시지 변환 시스템". 정보과학회 논문지 : 데이터베이스 제31권 제3호, 2004.6
- [9] 정의석 "상호운용성 확보를 위한 지식정보 표준화 동향". ITFIND 주간기술동향 제1230호 2006년 1월 25일