

그리드 정보 검색 시스템을 위한 동적 가상조직 설계 및 구현

박찬호⁰, 김혁호, 이승하, 김양우, 이필우^{**}

동국대학교 대학원 정보통신공학과, ^{**}한국과학기술정보연구원

{bestchanho⁰, hulegea, lesh915, ywkim}@dgu.edu, pwlee@kisti.re.kr

Grid IR System for Various Dynamic Reconfigurable VO

Chanho Park⁰, Hyukho Kim, Seungha Lee, Yangwoo Kim, Philwoo Lee^{**}

Dept. of Information & Communication Engineering Dongguk University, ^{**}KISTI

요 약

현 그리드 정보검색 시스템은 그리드 환경에 구축이 되어 있지만, 그리드 컴퓨팅의 특징 중 하나인 가상조직을 이용한 자원 관리가 결여되어 있다. 이로 인해, 제공하는 서비스 중 한부분에서만 과부하가 발생을 하여도 시스템 전체에 영향을 미치게 된다. 이러한 문제에 대해 본 논문에서는 그리드라는 분산처리 환경에 정보검색 시스템을 구축하여 보다 안전하고, 유연한 확장성을 갖는 정보검색 서비스를 제공하려 한다. 대표적인 그리드 미들웨어인 글로버스 툴킷(Globus Toolkit)의 가상 조직(VO: Virtual Organization)을 이용하여 정보검색 시스템을 구성해 보고, 가상조직이 갖는 이점인 유연성을 그리드 정보검색에 적용하여, 유연한 그리드 정보 검색 시스템을 제안한다. 그리고 그리드 정보검색 가상 조직 관리 서비스(GIR-VOMS)를 구현하여 시스템 성능을 높여 줄수 있는지 알아본다.

1. 서 론

그리드 컴퓨팅은 차세대 인터넷 인프라로서 과학 기술과 IT 분야 모두의 주목을 받으며 화려하게 부상했고, 지금도 많은 과학 기술 분야와 IT분야에서 점목 시키려는 기술이 시도되고 있다. 물론 현재의 정보 검색 분야를 의심하는 것은 아니지만 정보 검색 분야 또한 시도해야 할 많은 기술들이 남아있다는 걸 우리는 알고 있다. 그래서 정보 검색 분야에도 그리드를 접목시켜 새로운 시도를 하려 한다.

그리드 정보 검색(Grid-IR : Grid Information Retrieval)이란 그리드 컴퓨팅 환경에서의 정보검색을 의미하며, 기존의 정보 검색 시스템에 그리드의 확장성, 효율성, 그리고 보안 등의 장점을 추가하여 더 나은 정보 검색 서비스를 제공한다.

현재 대부분의 검색엔진에서 사용하는 시스템은 검색엔진 전용으로 구성되어 있기 때문에, 과학 계산용이나 온라인 게임 서버용으로 시스템을 일시적으로 전환하여 사용하는 것은 불가능하며, 시스템을 확장해서 작업을 일제히 분배하여 적용하는 것 또한 힘들기 때문에 시스템의 탄력적인 운용을 기대하기 어렵다. 그리고 여러 검색 엔진에서 검색 결과를 취합해 정제된 후 사용자에게 제공하는 기술이 없기 때문에 사용자가 원하는 정보를 정확하게 얻기 위해서는 여러 검색엔진을 이용해야 한다는 문제점을 가지고 있다.

이러한 문제들을 그리드 컴퓨팅의 특징인 가상 조직(VO : Virtual Organization)을 이용하여, 필요에 따라 동적으로 확장 가능하고 높은 보안 정책을 가진 그리드 검색엔진 서비스 모델을 제시하고자 한다. 이러한 개념에 따라 그리드 정보검색은 가상 조직 내에 존재하는 서비스를 필요할 때 필요한 만큼 활용 할 수 있는 온-디맨드(On-demand)한 성격을 갖게 되어 탄력적으로 운용할 수 있을 뿐만 아니라 가상 조직 별로 세분화된 보안 정책을 가질 수도 있어 새로운 패러다임을 제공할 것이다.

이 그리드 미들웨어이다. 그리드 기반 기술에는 OGSA(Open Grid Service Architecture)와 OGS(Open Grid Services Infrastructure)가 있고, 이러한 기술을 갖는 대표적인 그리드 미들웨어로서 GT3(Globus Toolkit 3)가 있다.

OGSA는 어플리케이션 통합을 위한 웹 서비스 표준과 리소스 통합을 위한 그리드 기술이 결합된 개방 표준이다.

OGSI는 OGSA 구조가 기반으로 하고 있는 그리드 서비스에 대한 기본 인프라를 정의하고 있다. 여기서는 그리드 서비스들이 만들어져 관리되고 정보를 교환하는 것에 대한 인터페이스와 기술적 묘사를 명확히 정의하고 있다.

그리드에서 가장 널리 사용되는 미들웨어가 바로 글로버스 툴킷인데, 그리드 기반 애플리케이션 구축에 사용될 수 있는 유용한 컴포넌트들의 집합을 이 글로버스 툴킷이 제공하고 있다. 주요 서비스로는 그리드에 연결된 유휴 자원들을 찾아내는 자원 검색 서비스, 할당된 작업들의 처리 순서를 결정하여 분산시키는 스케줄링 서비스, 시스템 안정을 위한 그리드 보안 서비스, 컴퓨팅 자원들을 사용할 때 발생하는 비용 처리를 위한 사용자 계정 서비스 등이 있다.

2.2 가상 조직 (VO : Virtual Organization)

가상 조직이란 자원들의 세트나 정책을 기초로 하여 이익을 위해 협동하는 개인 또는 단체들의 공동 작업할 수 있는 그룹이라는 의미를 갖고 있다.

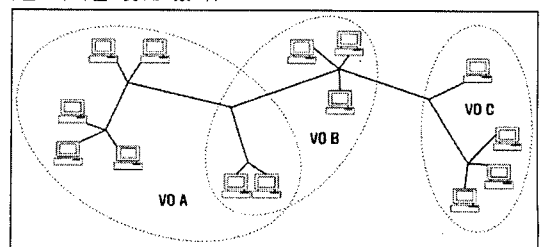


그림 1 가상 조직 개념도

2. 관련 연구

2.1 그리드 미들웨어

첨단 컴퓨팅 환경인 그리드 환경 구축을 가능케 하는 핵심기술

그리드 컴퓨팅을 기존의 컴퓨팅 환경과 비교할 때 가장 두드러진 차이점은 가상의 조직을 구성할 수 있는 것이다. 기존의 인터넷 컴퓨팅 환경은 어떤 물리적으로 나뉜 조직(또는 도메인)과 거기에 속한 개인이 다른 조직 또는 개인과 단순히 정보만을 공유할 수 있었다. 하지만 그리드 컴퓨팅은 물리적으로 다른 조직 또는 개인이 이런 한계를 넘어 가상의 공간에서 새로운 조직을 형성할 수 있도록 한다. 이런 가상의 조직에서 조직 또는 개인은 정보의 공유뿐만 아니라 자신이 가지지 못한 다양한 자원에 대해 사용할 수 있는 권리를 가질 수 있거나 여러 가지 다양한 가상의 조직에 참여할 수 있게 된다.

2.3 GSI

그리드에서는 하나의 클라이언트와 하나의 서버만으로 서비스가 이뤄지는 것이 아니라 여러 곳의 다양한 컴퓨터들이 동시에 사용되는 것을 전제로 하기 때문에 보안 문제가 더욱 복잡해진다. 또한, 그리드 환경에서는 컴퓨터가 서로 다른 기관에 걸쳐 존재하기 때문에 보안 정책도 기관에 따라 달라질 수밖에 없다. 이렇게 그리드 보안에 필요한 여러 기능들을 글로벌스톡킷에서는 GSI(Grid Security Infrastructure)라는 부분에서 구현하고 있다. GSI에서 지원하는 보안 메카니즘은 5가지로 정리할 수 있는데, 그것들은 데이터 비밀성, 무결성, 인증, 허가 그리고 권한 위임이다.

데이터 비밀성이란 송신자와 수신자 이외에는 아무도 이해하지 못하도록 하는 암호기법으로서, 안전한 통신을 위해 반드시 필요하다.

일반적인 보안 시스템에서와 마찬가지로 그리드 환경에서도 무결성은 중요하다. 데이터가 저장되거나 네트워크를 통해 다른 곳으로 전송될 때, 제 3자에게 노출되거나 공격자에 의해 변경·조작되는 것을 방지해야 할 때가 있는데, 글로벌스톡킷에서는 공개키(Public Key) 암호기법을 사용한다.

인증(Authentication)이란 어떤 요구나 요청이 전달될 때 상대방이 누구인지 확인하는 과정이다. 어떤 서비스가 이뤄지기 전에 먼저 상대방이 누구인지를 알아야 서비스를 받을 자격이 있는 사람인지 아닌지를 판단할 수 있기 때문이다.

허가(Authorization)란 인증된 사용자가 특정한 조직이 가능한지를 판단하는 과정이다. 허가를 하려면 인증된 사용자가 누구인지 알아야 하므로, 당연히 인증 과정이 선행되어야 한다. 그리드에서는 여러 가지 계산 자원이 존재하는데, 사용자가 누구인가에 따라 사용할 수 있는 계산 자원의 종류를 다르게 설정할 수 있어야 할 것이다.

권한 위임이란 그리드 자원이 사용자의 작업을 처리하는 과정에서 사용자를 대신해 또 다른 그리드 자원에 대한 인증을 수행하는 메커니즘이다.

2.4 그리드 정보검색

그리드 정보검색이란 그리드 컴퓨팅 환경에서의 정보검색을 말한다. 그리드 정보 검색 시스템 설계 시 주된 고려사항은 현존하는 대부분의 단일 성격의 정보 검색 기술을 향후 시스템에 끼워써 사용할 수 있도록 하는 것이다. 따라서 그리드 정보 검색도 대부분의 검색 엔진이 갖는 구조를 갖고 있다.

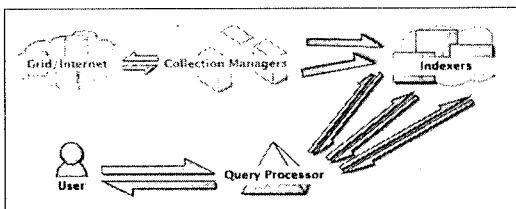


그림 2 그리드 정보 검색 구조도

즉, 웹 크롤러(Web Crawler)라 불리는 검색로봇(Search Robot)이 도큐먼트를 찾으면 그 도큐먼트를 모아서 로컬에 저장하여 관리하는 CM(Collection Manager), 그 도큐먼트를 인덱싱(Indexing) 하기 위한 IS(Index/Search) 그리고, 검색 결과에서 연관된 부분의 추출과 여러 개의 도큐먼트 컬렉션(Document Collection)에 질의를 담당하는 QP(Query Processor)의 구조로 되어 있다.

3. 가상조직 기반 그리드 검색엔진 시스템 설계

그리드 정보검색 시스템은 그리드 컴퓨팅과 정보검색의 조합으로 많은 장점을 갖는다. 이 논문에서는 그 중에서 그리드 컴퓨팅의 특징인 가상조직(VO : Virtual Organization)을 이용하여, 필요에 따라 동적으로 확장 가능하고 높은 보안 정책을 가진 그리드 검색엔진 서비스 모델을 제시하고자 한다.

3.1 다양하고 동적인 가상조직 재구성 시나리오

대한 증권사는 시스템의 성능향상과 증권 정보의 검색 능력을 향상시키기 위해 검색 시스템을 그리드 정보검색(Grid-IR)으로 구성하기로 하였다. Grid-IR로 구성된 대한 증권사의 각각의 시스템 구성요소는 상황에 따라 가상 조직을 확장함으로써 구성요소의 파워를 유연하게 늘리거나 줄일 수 있다.

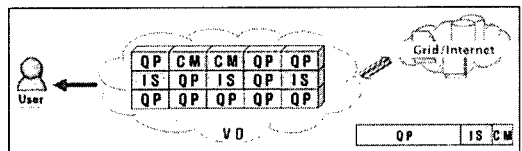


그림 3 가상 조직에서의 Query Processor 확장 예제 모델

첫째, 대한 증권사는 그 날의 시장이 마감하는 동시에 고객들에게 모바일장치로 분석 서비스를 제공하려 한다. 분석서비스는 고객들이 보유하고 있는 증권에 대한 일주일 또는 한 달의 등/폭락 데이터와 예측 그래프를 제공함으로써 고객들이 보다 쉽고 편하게 정확한 데이터를 받아 볼 수 있는 서비스이다. 그러나, 이러한 서비스는 “ $\sum_{i=1}^{n} (C \cdot C^i \text{ 주식보유수})$ ”¹⁾ 이상으로 질의 수행 능력이 필요하다. 이러한 복잡하고 많은 Query에 대해서도 시스템의 고장 없이 빠르고 견고하게 작동하기 위해서는, 가상 조직의 유휴 자원을 Query Processor로 자동 확장 기능이 제공되어야 한다.

둘째, 대한 증권사는 증시에 영향을 미치는 정치, 경제, 사회, 문화 그리고 국제 정세 등을 분석하여, 고객들이 투자할 주식을 선택할 때 분석된 자료를 서비스 하려한다. 이러한 서비스는 데이터 마이닝(Data Mining) 시스템을 필요로 하는데, 이를 위해 다른 파일 구조와 문서 타입 가지고 있었던 여러 데이터베이스 시스템을 가상화하여 관련된 자료를 모두 가져야 할 필요가 있다. 따라서 고객들이 비교 분석을 요청할 때 실시간으로 제공하기 위해서는 지리적으로 떨어져있고 서로 다른 형태를 가진 Index/Search 자원을 가상화시킬 수 있어야 한다.

셋째, 정보검색 시스템에서 CM은 가장 많은 컴퓨팅 파워를 필요로 한다. 통계에 따르면 현 정보검색 시스템에서 절반 이상의 컴퓨팅 파워가 이 CM에 의해 사용되어 진다고 한다. 정보 수집에 대한 시간적 제약이 약하기 때문에 Query Processor나 Index/Search의 사용률이 낮을 때, 이들의 서비스를 CM으로 전환하여 사용하면 효율성이 높아질 것이며, 처리 시간 또한 줄어들 것이다.

1) C는 고객, Customer

3.2 그리드 정보검색 가상조직 관리 서비스 구조

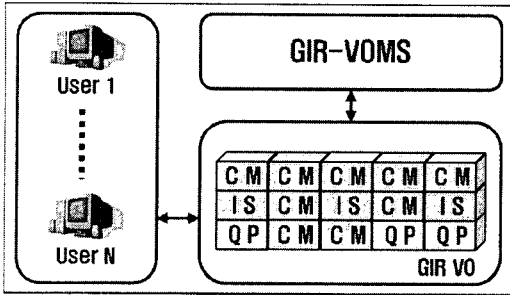


그림 4 시스템 구조

가상조직에 구성된 그리드 정보검색서비스는 GIR-VOMS (Grid Virtual Organization Management Service)에 의해 관리되어진다. 만약 특정 서비스에 과부하가 발생하거나, 더 많은 컴퓨팅 파워가 필요할 경우에는 GIR-VOMS가 이를 감지하여 동적으로 전환하여 관리를 할 수 있다. 또한 그리드 정보검색 시스템의 확장이 필요할 경우에는 새로운 사용자를 해당 가상조직에 참여시킴으로써 가능해진다.

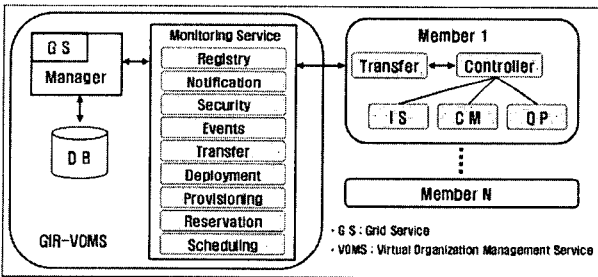


그림 5 GIR-VOMS와 Member의 구조

그림에서 볼 수 있듯이, GIR-VOMS는 그리드 정보 검색 시스템을 모니터링하여, 필요에 따라 가상조직 안에서 동작하고 있는 서비스를 관리하는 구조를 가지고 있다. 모니터링 서비스에는 멤버들의 목록을 담당하는 Registry, 멤버에게 서비스 전환을 요청하는 Notification, 인증/허가를 담당하는 Security, 멤버들이 서비스 전환을 하게 되면 IS/CM/QP의 관계를 재구성하는 Events, 새로운 멤버가 추가되면 설치를 담당하는 Deployment와 Transfer, 과부하가 발생하면 시스템 오류가 발생하기 전에 미리 감지하여 필요한 서비스를 사전설치 하는 Provisioning, 작업 중인 멤버들을 전환하기 위한 Reservation과 Scheduling서비스가 있다.

3.3 결과

다음 그림은 하나의 가상조직에 구성된 글로벌 정보검색 시스템으로써 필요에 따라 언제든지 동적으로 정보검색 시스템의 역할을 바꿀 수 있는 가상조직 관리자이다.

이 가상조직 관리자는 그리드 서비스로 등록된 후 서비스 인스턴스를 생성함으로써 실행되어진다. 이때 가상 조직에 구성되어 동작하고 있는 각각 컴퓨터들의 역할 정보를 가지고 온다. 이렇게 함으로써 필요에 따라 등록된 자원들의 역할을 바꾸어 사용하는 다양하고 동적인 가상조직의 재구성이 가능하게 된다. 또한 자원의 역할이 바뀔때 따라 컬렉션관리자(CM), 인덱스/검색(IS) 그리고 질의 처리기(QP)의 관계도 동적으로 재구성되어진다.

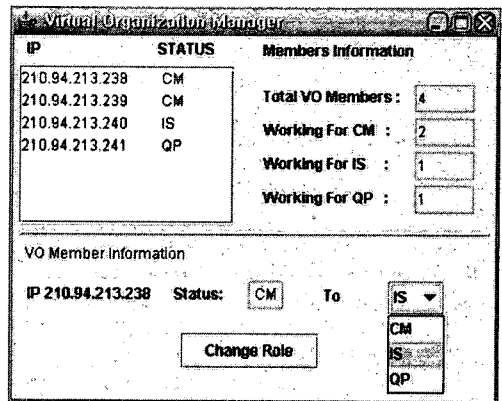


그림 6 등록된 자원의 동적 변경

시스템의 확장이 필요할 경우에는 새로운 컴퓨터를 해당 가상조직에 구성시킴으로써 간단히 해결할 수가 있다. 이것은 매우 유연한 구조를 가지고 있기 때문에 시스템 확장을 위하여 서비스를 멈추지 않아도 되며, 플랫폼 선택도 매우 자유롭다.

4. 결론

본 논문에서는 정보검색시스템에 그리드의 주된 기술인 가상조직을 추가하여, 그리드 정보검색 서비스 환경을 구축하였다. 이를 통해 얻는 이점들을 시나리오를 통해 예상하고, 그것을 실제로 설계하고 구현해 봄으로써 시스템의 성능을 높일 수 있는지 알아보았다. 즉, GIR-VOMS를 통해 동적으로 정보검색 서비스를 관리할 수 있도록 하였다.

향후 연구과제로 본 논문에서 제안한 시스템을 확장하여, 특정 서비스의 과부하 자동 탐지 기능을 추가하여야 할 것으로 보인다. 그렇게 한다면 온 디맨드(On-demand) 환경의 자율컴퓨팅(Autonomic Computing)이 가능해 질것이다. 또한 그리드 컴퓨팅의 보안기술인 GSI(Grid Security Infrastructure)를 이용하여 그리드 정보 검색 시스템에 보안 레벨을 적용할 필요성이 있을 것으로 예상하고 있다.

따라서, 앞으로 기업들이나 정부 기관들은 그들 자신의 검색 엔진을 구현할 때, 더 이상 고가의 서버(Server)나 클러스터(Cluster)를 구입하지 않아도, 그들이 갖고 있는 PC(Personal Computer)로서 고 성능의 검색 엔진을 구축할 수 있을 것으로 예상된다. 그리고 기존의 단일 시스템에서 특정 서비스에 과부하가 생기면, 시스템 성능이 떨어져 결국에는 시스템 오류(System Failure)를 일으키지만 그리드 정보 검색 서비스에서는 특정 서비스에 더 많은 컴퓨팅 파워가 필요할 때, 일시적으로 그 서비스로 전환하여 사용할 수 있어 시스템 오류를 방지할 수 있을 것이라 기대한다.

참고 문헌

[1] N. Nassar and K. Gamiel, "Grid Information Retrieval Architecture" Grid Information Retrieval Working Group, 2004.
 [2] Dr. Bernhard R. Katzy "Design and Implementation of Virtual Organizations" University St. Gallen and Erasmus University Rotterdam
 [3] Radu Prodan and Thomas Fahringer, "From Web Service to OGSA : Experiences in Implementing an OGSA-based Grid Application" Proc. IEEE, 2003