

Hybrid P2P 기반의 데이터 그리드 시스템

이은목¹, 최지연², 윤성민¹, 최영근¹

광운대학교 컴퓨터소프트웨어학과¹, 광운대학교 컴퓨터공학과²

{charjj, yeoni, zumbi, ygchoi}@kw.ac.kr

Data Grid System based on Hybrid P2P

Eun-Mook Lee¹, Ji-Yeon Choi², Sung-Min Yoon¹, Young-Geun Choi¹

Department of Computer Science¹, Computer Engineering²

KwangWoon University

1. 서론

원거리 통신 환경의 발달로 인해 인터넷 사용자가 기하급수적으로 증가하고 있으며 고속화된 인터넷 환경을 기반으로 일반 사용자들도 대용량의 데이터를 다루게 되면서 필연적으로 사용자의 저장 공간의 부족이라는 문제가 발생하게 되었다. 이러한 문제에 대한 기존의 일반적인 해결 방법으로 하드디스크 같은 스토리지를 증설하거나 별도로 네트워크 내의 스토리지를 구입해 사용하고 CD나 DVD를 이용해 데이터를 백업하여 저장하고 있는 실정이다. 하지만 이러한 방법은 추가적인 장비 구입비용과 관리 및 사용성의 문제가 있다. 또한 근거리에 있는 PC들을 연결해서 사용하는 클러스터링 기법이나 지역 네트워크를 활용하는 방안은 공간적인 제약을 가지며 Client/Server 환경의 네트워크를 이용하는 방법은 서버에 부하가 집중되는 문제와 서버의 저장 공간 부족이라는 문제가 여전히 존재한다.

위와 같은 문제를 해결하기 위해 본 논문에서는 다수의 사용자들이 원거리 네트워크를 통하여 저장 공간을 확장시킬 수 있는 Hybrid P2P 기반의 데이터 그리드 시스템을 제안한다.

본 시스템은 다량의 데이터를 네트워크에 연결된 노드 중에 적절한 노드를 찾아서 저장하고 단일 노드에서 수용할 수 없는 용량의 데이터일 경우에는 적절한 크기로 분할하여 분산 저장하는 방식으로 스토리지 자원을 효율적으로 사용할 수 있도록 설계하였다.

2.1 시스템 개요

Hybrid P2P를 이용한 데이터 그리드 시스템은 각 노드 데이터와 디렉터리의 메타 정보를 가지고 있는 인덱스 서버를 갖는다. 그 서버를 중심으로 시스템내의 노드들이 인터넷을 통해 하나의 단일 파일시스템을 구성하여 그리드 커뮤니티를 이룬다. 그림1은 시스템의 구성도를 나타낸다.

그리드 시스템에 의해 커뮤니티가 형성된 후 사용자가 인덱스 서버에게 트랜잭션을 요청하면 서버는 데이터베이스에서 필요한 정보를 읽어와 내부적으로 처리하고 처리 결과를 데이터베이스에 저장한 후 요청한 노드에게 응답 하게 된다. 응답을 받은 노드는 받

은 결과를 가지고 해당되는 다른 노드에게 접속해 P2P 상태로 관련 데이터를 전송하거나 전송 받을 수 있다. 이 과정에서 단일 노드에서 수용할 수 없는 용량의 데이터일 경우에는 다른 노드들에 분산되어

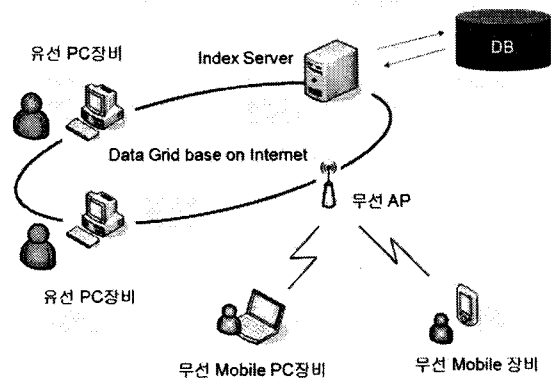


그림 1. 시스템 구성도

저장하게 된다. 분산 저장되어 있는 데이터를 전송 받을 경우에는 데이터를 소유하고 있는 각 노드에서 데이터들을 모아 하나의 데이터로 재구성하여 사용되게 된다.

2.2 시스템 구성

본 시스템은 그림2와 같은 Layered Architecture로 구성되어 있다. 가장 하위 레이어에 서로 다른 파일 시스템이나 네트워크 환경에서도 동작할 수 있도록 해당 로컬 시스템에 접근 할 수 있는 추상화된 인터페이스를 제공하는 Infrastructure Layer가 있다. 그 위로 그리드 시스템의 핵심 서비스들을 담당하는 Grid Service Layer가 어플리케이션에서 발생하는 트랜잭션 요청처리, 메타 데이터 관리, 실행 등을 처리한다. 그리고 최상위에서는 실질적인 서비스용 Application Layer가 존재한다.

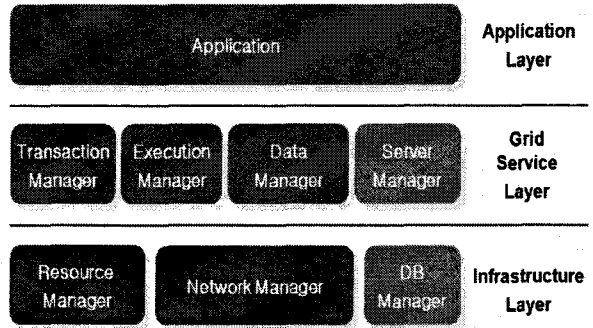


그림 2. 어플리케이션 아키텍처

3. 시스템 비교

본 논문에서 제안하는 시스템은 네트워크에 있는 다수의 스토리지를 연결해 데이터 그리드를 이용한 커뮤니티를 구축하고, 중복데이터를 제거 및 분산처리 하여 자원을 효율적으로 사용하는데 집중하였다.

기존의 P2P 서비스들이 단순히 "파일공유"를 위한 서비스를 제공하기 위해 '1:1' 통신을 제공하는 반면에 본 논문의 시스템에서는 사용자가 최초의 그리드 커뮤니티 참여시에 시스템에 제공하는 자원을 서버에 등록하여 다 같이 공유하는 방법을 통해 여러 노드들 간의 '불특정 다수 : 1'의 관계를 형성하며 "스토리지 자원"을 공유한다. 이를 통해 부가적인 파일 공유 등의 서비스가 가능하도록 하였다.

4. 결론

본 논문에서 제안한 시스템은 그리드 시스템을 개인 사용자 환경에서 사용가능하도록 하였으며 대용량 데이터 수요증가에 따른 공간 부족 문제를 하드웨어가 아닌 소프트웨어 측면에서 해결할 수 있는 방안을 제시하였다. 또한 Hybrid P2P 기술을 사용함으로써 기존의 Server/Client 구조에서 문제되었던 서버의 부하 문제를 시스템에 참여한 각 노드들에게 분산시킴으로 해결하였다.

본 시스템은 높은 비용에 비해 제대로 활용되지 않고 있는 각 노드들의 스토리지와 컴퓨팅 자원을 활용함으로써 자원 효율성을 높였다. 하지만 수많은 노드가 그리드에 연결되어 동시에 단일 데이터를 요청 했을 때의 로드 밸런싱 처리와 분산 저장된 데이터를 가진 사용자가 접속하지 않았을 경우 이를 사용할 수 없는 등의 중요한 문제가 남아있다.

아직 해결되지 않은 본 시스템의 문제점들을 해결하고 데이터의 저장뿐만 아니라 각종 응용 프로그램이나 데이터베이스까지 모두 공유하는 등의 시스템 개선을 통해 더욱 다양한 분야에서 새로운 서비스들을 창출해 낼 수 있을 것이라 기대한다.