

클라우드 컴퓨팅을 위한 클라우드 스토리지 기술 분석

박정수* · 정성재** · 배유미*** · 경지훈**** · 성경*****

*매크로임팩트(주), **(주)스컴씨엔에스, ***한남대학교 컴퓨터공학과, ****(주)시큐브,
***** 목원대학교 컴퓨터교육과

Technical analysis of Cloud storage for Cloud Computing

Jeong-Su Park* · Sung-Jae Jung** · Yu-Mi Bae*** · Ji-Hun Kyung**** · Kyung Sung*****

*Macroimpact, Inc., **Sky Computing C&S, Inc., ***Hannam University. ****Secuve, Inc.,
*****Mokwon University

E-mail : jparks7@naver.com, posein@naver.com, yumidw@hanmail.net, sia2001@secuve.com,
skyys04@mokwon.ac.kr

요 약

클라우드 서비스 공급자가 제공하는 클라우드 스토리지 시스템은 대량의 데이터 저장 및 클라우드 컴퓨팅 처리의 핵심구성요소이다. 구글, 유튜브, 페이스북 등과 같은 대형 벤더에서도 클라우드 스토리지를 사용하여 태블릿, 스마트폰 등 기기종 장치에서 사진 동영상, 문서등을 공유하며, 쉽고 빠르게 네트워크를 통해 데이터의 대량 전송과 접근이 가능하다 이로 인해 전세계적으로 데이터 성장 발전과 함께 클라우드 스토리지가 비즈니스모델로 새롭게 부각받고 있다 클라우드 스토리지는 네트워크 스토리지의 새로운 서비스 개념이며 데이터 구성 스토리지 가상화, 데이터 복제와 중복제거, 보안 등을 포함한 클라우드 컴퓨팅 핵심기술이다 본 논문에서는 클라우드 스토리지에 대해 분석한다.

ABSTRACT

Cloud storage system that cloud computing providers provides large amounts of data storage and processing of cloud computing is a key component. Large vendors (such as Facebook, YouTube, Google) in the mass sending of data through the network quickly and easily share photos, videos, documents, etc. from heterogeneous devices, such as tablets, smartphones, and the data that is stored in the cloud storage using was approached. At time, growth and development of the globally data, the cloud storage business model emerging is getting. Analysis new network storage cloud storage services concepts and technologies, including data manipulation, storage virtualization, data replication and duplication, security, cloud computing core.

키워드

클라우드 스토리지(cloud storage), 스토리지 가상화(storage virtualization),
데이터 복제(data replication), 데이터 중복제거(data duplication)

I. 서 론

2009년에 개최된 심포지엄 아이티 엑스포(ITxpo)에서 가트너는 10대 전략 기술로 클라우드 컴퓨팅을 선정하였다[1].

클라우드 스토리지는 스토리지의 성격을 갖고 있는 특정한 장치의 개념이 아니며 클라우드 서비스 공급자가 제공하는 환경에서의 계층적인 구

조로 네트워크의 빠른 성장과 함께 사용자 참여 중심의 웹2.0 시대에서의 대량의 데이터 저장 및 클라우드 컴퓨팅 처리의 핵심구성요소이다[2][3].

구글, 유튜브, 페이스북 등과 같은 대형벤더에서도 클라우드 스토리지를 사용하여 태블릿 스마트폰 등 기기종 장치에서 사진 동영상, 문서등을 공유하며, 쉽고 빠르게 네트워크를 통해 데이터의 대량 전송 및 저장되어 있는 데이터에 접근하게

되었다. 이로 인해 전세계적으로 데이터 성장 발전과 함께 클라우드 스토리지가 비즈니스모델로 새롭게 부각받고 있다[4][5]. 클라우드 스토리지는 네트워크 스토리지의 새로운 서비스 개념으로 클라우드 스토리지 서비스는 저비용, 가용성, 확장성, 보안을 제공해야 한다. 본 논문에서는 클라우드 컴퓨팅의 핵심구성요소인 데이터 구성 스토리지 가상화, 데이터 복제와 중복제거, 보안 등을 포함한 클라우드 컴퓨팅 핵심기술에 대해 분석한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련연구에 대해 기술하고 3장에서는 이러한 클라우드 스토리지의 핵심기술에 대해 설명하고 마지막으로 4장에서는 결론을 맺는다.

II. 관련연구

2.1 클라우드 스토리지의 개요

Amazon, EMC, IBM, HP와 NetApp등 많은 클라우드 서비스 공급자들이 클라우드 스토리지 서비스를 하고 있다. 또한, HDFS, GFS, Amazon S3, EMC Atoms, Data ONTAP, HP Upline, Cloud NAS, FileStore처럼 클라우드 스토리지 플랫폼들이 많이 증가하고 있는 추세이다

2009년 클라우드 사용자, 서비스 공급자, 개발자, 브로커에게 클라우드 서비스를 위해서 140개 이상의 기업들과 함께 SNIA(Storage Network Industry Association) 클라우드 스토리지 이니셔티브(Cloud Storage Initiative)에서 클라우드 서비스 표준으로 CDMI(Cloud Storage Management Interface)을 제안을 채택했다[6].

2.2 클라우드 스토리지와 클라우드 컴퓨팅 서비스와의 연관성

클라우드 스토리지는 클라우드 컴퓨팅 환경을 기반으로 하는 클라우드 서비스의 계층적인 구조와 동일한 구조로써 그림 1처럼 네 가지 범주 계층에 속한다.

SaaS(Software as a service) 클라우드 서비스는 클라우드 사용자가 인터넷을 이용하여 필요한 소프트웨어를 이용하는 응용 소프트웨어 계층이고 PaaS(Platform as a service)는 API interface와 스토리지 관리, 스토리지 가상화로 구성된 계층이며, IaaS(Infrastructure as a service)는 서버와 하드웨어 자원과 서버 클러스터 기술을 제공하는 서비스이다[7][8].

DaaS(Data Storage as a service)는 클라우드 스토리지 시스템의 구조로써 데이터 스토리지로 구성되어진 스토리지 서비스 인터페이스에서 데이터 스토리지를 제공하고 추출 전송함으로써 기존 클라우드 사용자에게 다양한 서비스 기능을 제공하고, CIFS, NFS, WebDAV 같은 기존 표준 프로토콜을 이용하기도 한다

클라우드 스토리지는 스토리지 미들웨어 시스템과 분산 파일시스템에 네트워크로 연결된 클러스

터화로 수천대의 스토리지 장치들로 구성되어 사용자에게 클라우드 서비스를 제공하고 SLA(Service Level Agreement), 분산파일시스템 리소스 자원, 네트워크 장치들을 포함하고 있는 구조이다. 사용자에게 더 많은 호환 기능 및 상호작용을 제공하기 위해서 물리적이면서 논리적 기능을 제공해야 한다.

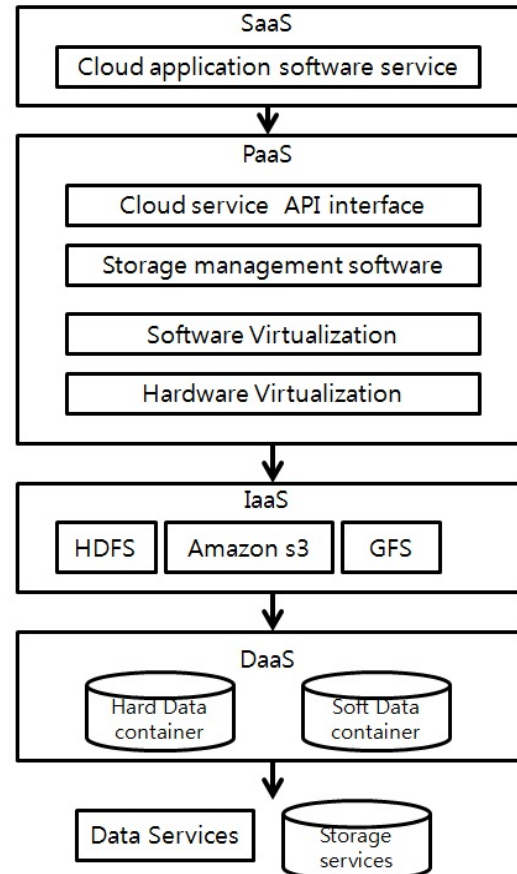


그림 1. 클라우드 스토리지 구조 시스템

III. 클라우드 스토리지 핵심 기술

3.1 데이터 구성

클라우드 스토리지는 파일 스토리지와 블록 스토리지로 두 분류로 나눌 수 있다

파일 스토리지는 파일 레벨 스토리지에 기반을 두고 있으며, 웹 사이트, IPTV, 소셜, 앱 등의 대용량 파일들을 저장하고 스토리지 노드의 제한 없는 확장이 가능하며 특히, Read I/O에 집중하여 공유하는 파일 서비스에 적합하다

블록 스토리지는 엔터프라이즈 응용 소프트웨어에서 운영되는 가상화 환경에서 연동되는 서비스로 고성능의 Read, Write I/O를 제공하지만, 고비용과 대용량 파일 스토리지에는 적합하지 않다

클라우드 스토리지에서 파일 스토리지와 블록 스토리지를 통합 개발하여 파일 스토리지의 장점과 블록 스토리지의 장점을 모두 지원하는 관련 연구와 EMC, NetAPP, IBM, HP 등에서 고확장형 유니파이드 스토리지 솔루션을 제공하고 있다[9][10][11][12].

3.2 스토리지 가상화

가상화 스토리지는 여러 지역에 분산되어 있는 이기종 스토리지들의 논리적인 볼륨 관리 스토리지 관리, 다른 I/O 인터페이스로 스토리지 모니터링 및 배포와 아카이브를 통합하고 관리하기 위해서 스토리지를 가상화해서 스토리지 활용률을 높이고, 스토리지의 RAID0 기술을 활용하여 높은 I/O성능을 향상시킬 수 있으며, 장애가 발생하더라도 계속적으로 서비스를 할 수 있다[13]. 스토리지 가상화 구현에 크게 3가지 방법이 있다. 첫째, 호스트 기반과 둘째, 스토리지 장치 가상화 기반과 셋째, SAN기반이 있다. 현재 추세로 SAN기반에서 가상화 기술이 많이 개발되고 있으며 스토리지 중앙 관리, 용량등을 유연하게 관리할 수 있는 SAN가상화를 많이 사용한다. 객체기반 파일시스템과 클러스터 분산 파일시스템은 대칭형(symmetrical)과 비대칭형(asymmetrical) 구조로 메타데이터와 데이터 통신 채널을 같이 사용하는지와 분리하여 사용하는지에 따라 구분할 수 있다 [14].

3.3 데이터 복제

데이터 복제에는 호스트 기반 복제로 응용소프트웨어, 데이터베이스, 파일시스템, 볼륨레벨, 하이브리드가 있고, SAN 기반 복제로는 디스크 어레이, 가상화 어플라이언스, 캐싱 어플라이언스, 스토리지 스위치가 있고, 백업 시스템 복제로는 블록 레벨 증분 백업, CDP, 중복제거가 있다[15].

데이터 복제 개념은 복제될 스토리지 자원인 소스와 복제 스트림을 수신할 스토리지 자원인 타겟으로 구성이 되며 동기화 방식에는 무손실 방식인 동기식(Synchronous) 복제와 비동기식(Asynchronous) 복제로 나뉘어진다. 동기식 복제는 소스 노드에서는 타겟 노드에서 데이터가 전송되었는지 확인응답(Acknowledge) 후 write I/O를 하고 비동기식 복제는 소스 노드에서 write I/O를 한 후 타겟노드로 전송함으로써 동기식 복제보다 성능저하가 발생하지는 않지만 치명적인 오류로 인해 데이터의 손실이 발생할 수 있다 [16].

3.4 데이터 중복제거

중복제거는 중복된 파일을 탐지해 중복된 파일을 삭제하여 유일(unique)한 데이터를 시스템에 저장하는 기술로 스토리지 시스템 내에 중복 데이터를 감소함으로써 스토리지 공간의 활용과 대역폭을 높일 수 있다. 중복제거는 점진적 백업과는 완전히 다른 개념이다. 점진적 백업은 파일이

변경될 때만 백업을 하기 때문에 데이터 변경이 없다면 중복된 데이터가 많이 존재한다[17]. 중복제거는 데이터 압축 기술을 사용하여 압축 알고리즘으로 중복 데이터를 제거하지만 데이터 블록 또는 동일한 파일 위치에 있는 스토리지 시스템에서 알고리즘을 사용하여 제거하는 단점이 있다. 중복제거의 핵심은 효율적으로 데이터 스토리지 감소와 데이터 인스턴스를 유지해야한다[18]. 중복제거의 기술은 백업환경에서와 백업데이터 스트림에서의 데이터 블록 및 백업대상 타겟지역에 기록된 데이터 블록으로 나누어져 있다. 중복 제거소프트웨어는 인덱스를 생성하여 저장소에 저장하여 데이터 블록 유무를 판단하고 디지털 서명(fingerprint)으로 오버헤드 없이 중복된 데이터를 찾을 수 있고, 데이터가 중복 되었다면 원본 데이터 세그먼트의 포인터를 링크한다[19]. 데이터 중복 제거를 간단히 표현하면 그림 2와 같다.

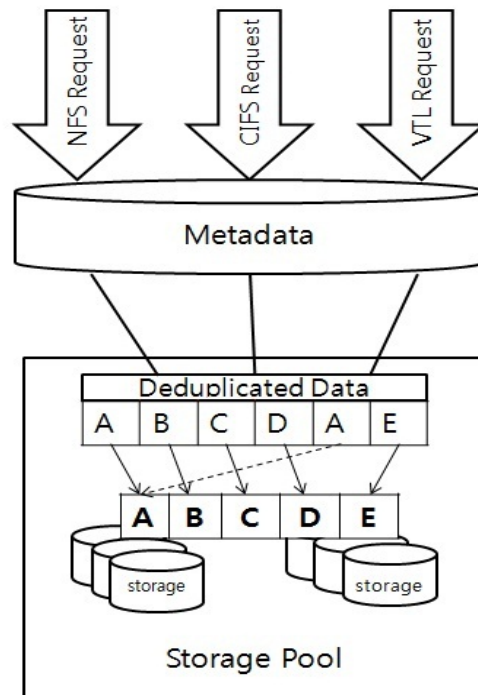


그림 2. 데이터 중복제거

3.4 스토리지 보안

클라우드 스토리지는 네트워크를 통해 스토리지 서비스를 이용하기 때문에 네트워크를 통한 클라우드 사용자의 많은 데이터를 포함할 수 있으므로 클라우드 특성상 가상화 센터와 클러스터 서버들은 하나의 지역에 국한되어 있지 않고 여러 지역으로 분산되어 있어 데이터가 여러 지역에 분산되어서 잘못된 접근으로 인해 개인 정보 보호 문제가 있기 때문에 반드시 클라우드 서비스 공급자로 접속을 해야 한다[20]. 또한, 클라우드 네트워크 환경에서 실행되는 안전 위험 관리와

데이터 관리, 데이터 위치, 데이터 분리, 데이터 복구, 감사 및 서비스 연속성을 위해 관련된 업체에서 연구과제로 직면하고 있다

IV. 결 론

클라우드 스토리지는 수직적으로 급증하는 데이터를 효율적으로 통합 관리할 수 있을 뿐만 아니라, 클라우드 스토리지 시스템에서는 높은 확장성과 가용성을 제공하여 클라우드 사용자에게 자원 관리와 비용 절감 등 많은 서비스 이점을 갖고 있다.

본 논문에서는 IT 인프라의 핵심인 클라우드 컴퓨팅을 위한 클라우드 스토리지 기술에 대해 살펴 보았다. 여러 관련된 대형 벤더에서는 스토리지 용량과 확장이 가능하게 융합하여 스케라블(scalable) NAS, 또는 Scale-out NAS라는 기술 서비스를 제공하여 클라우드 스토리지 시장에 빠르게 진입하여 공략하고 있다. 국내에서도 스토리지 인프라에 관련된 기업들이 외산 제품과 경쟁하고 있고, 국내 기술을 바탕으로 연구 개발하여 고확장형 유니파이드한 스토리지 솔루션으로 IT 자원 활용성을 극대화하고 기업의 ROI(Return on Investment, 투자자본에 대한 수익률)를 대폭 개선하여 국가 기술 경쟁력 향상 할 것으로 기대한다.

참고문헌

[1] Gartner, "Gartner identifies the top 10 strategic technologies for 2011", <http://www.gartner.com>, 2010.

[2] Robison S., A Bright Future in the Cloud. Financial Times, March4, 2008.

[3] Fu-yi group to talk about cloud computing pan development path, in November, 2008

[4] Jun-wei Ge, Yong-long Deng, Yi-qiu Fang, "Research on Storage Virtualization Structure in Cloud Storage Environment", Multimedia Technology (ICMT), IEEE Conference Publications, pp1-4, 2010.

[5] Steve Lesem. Cloud Storage and The Innovator's Dilemma . <http://cloudstoragestrategy.com/cloud-ecosystem/> , July 19, 2009.

[6] SNIA CLOUD Storage Summit. 2009. <http://www.snia.org/events/wintersymp2009/cloud/>. Held at the WINTER SYMPOSIUM 2009.

[7] 배유미, 정성재, 장래영, 성경, 소우영, "클라우드 컴퓨팅의 국내 현황과 전망", 한국정보통신학회, pp277-280, 2011.5.

[8] Jeff O'neal, "클라우드를 위한 스토리지 인프라", <http://www.netapp.com/kr>, 2009.

[9] EMC, <http://www.emc.com/>

[10] NetApp, <http://www.netapp.com/us/>

[11] IBM, <http://www.ibm.com/>

[12] HP, <http://www.hp.com/>

[13] 김영철, 차명훈, 이상민, 김영균, "클라우드 컴퓨팅에서 스토리지 가상화 기술 동향, 전자통신동향분석, 24권, 4호, pp69-78, 2009.

[14] 차명훈, 이상민, 김 준, 김영균, 김명준, "대규모분산 파일 시스템 환경의 메타데이터 관리", [ETRI] 전자통신동향분석, 제22권, 제3호, 2007.

[15] Jacob Farmer, "A Crash Course in Wide Area Data Replication", SNIA, 2009.

[16] Qinlu He, Zhanhuai Li, Xiao Zhang, "Analysis of the key technology on cloud storage", FITME, IEEE Conference Publications2010, pp.426-429, 2010.

[17] Mark W. Storer Kevin Greenan Darrell D. E. Long Ethan L. Miller. 2008. Secure Data Deduplication. StorageSS'08, October 31, 2008, Fairfax, Virginia, USA. 2008.

[18] ZHU, B., LI, K., AND PATTERSON, H. Avoiding the disk bottleneck in the Data Domain deduplication file system. In Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST) (San Jose, CA, USA, February 2008), pp. 269-282, 2008.

[19] Thomas Rivera, "Understanding Data Deduplication", SNIA, 2009.

[20] Cloud computing development status, <http://hi.baidu.com/lmc625263041/blog/item/122aeb1d0a492bd309a50276b.html> 2009-03-01 16:17