클라우드 컴퓨팅을 위한 클라우드 스토리지 기술 분석

박정수* · 배유미** · 정성재***

Technical analysis of Cloud Storage for Cloud Computing

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

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, 스토리지 가상화, 데이터 복제, 데이터 중복제거

Key word

cloud storage, storage virtualization, data replication, data duplication

* 정회원 : (주)매크로임팩트
** 정회원 : 한남대학교 컴퓨터공학과
*** 정회원 : (주)스کم씨앤에스(교신저자, posein@naver.com)

접수일자 : 2012. 11. 05
심사완료일자 : 2013. 04. 22

I. 서 론

2009년에 개최된 심포지엄 아이티 엑스포(ITxpo)에서 가트너는 10대 전략 기술로 클라우드 컴퓨팅을 선정하였다[1]. 클라우드 컴퓨팅 기반의 인프라 구성이 보편화되고, 빅데이터(Big Data)가 새로운 IT의 화두가 되면서 각광받은 기술이 클라우드 스토리지(Cloud Storage)이다.

클라우드 스토리지는 스토리지의 성격을 갖고 있는 특정한 장치의 개념이 아니며, 클라우드 서비스 공급자가 제공하는 환경에서의 계층적인 구조이다. 네트워크의 빠른 성장과 함께 사용자 참여 중심의 웹 2.0과 향후 도래할 웹 3.0 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 전송하고, 저장되어 있는 데이터에 접근하게 되었다. 이로 인해 전세계적으로 데이터 성장 발전과 함께 클라우드 스토리지가 비즈니스 모델로 새롭게 부각받고 있다[4][5]. 클라우드 스토리지는 네트워크 스토리지의 새로운 서비스 개념으로, 클라우드 스토리지 서비스는 저비용, 가용성, 확장성, 보안을 제공해야 한다.

본 논문에서는 클라우드 스토리지의 기술적 동향을 분석하고, 클라우드 스토리지가 가져야할 클라우드 스토리지 핵심 기술에 대해 알아보고, 결론을 맺는다.

II. 본 론

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. 이러한 다양한 클라우드 스토리지 플랫폼을 표준화하기 위해 SNIA(Storage Network Industry Association)에서는 2009년 클라우드 사용자, 서비스 공급자, 개발자 등 140

개 이상의 기업들과 함께 클라우드 서비스 표준으로 CDMI(Cloud Data 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 1]처럼 구성된다.

그림 1. 클라우드 스토리지시스템의 구조
Fig. 1 Structure of Cloud storage systems

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 가상

화로 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 클러스터화로 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 상호 작용을 제공하기 위해서 물리적이면서 논리적 기능을 제공해야 한다.

III. 관련 기술 동향 분석

3.1. 클라우드 스토리지의 기술적 개요

클라우드 스토리지는 초기의 스토리지 기술인 NAS(Network Attached Storage)나 SAN(Storage Area Network)를 스토리지 가상화 기술 등을 이용하여 스토리지 자원을 사용자의 요구에 따라 대규모 확장이 가능하고, 특정 지리적 위치에 구애받지 않고 애플리케이션에 유연한 특징을 보유해야 한다. 특히 대규모의 확장이 가능해야 하는데, 전통적인 NAS 구조를 여러 대의 서버에서 동시에 실행되는 분산 파일 시스템을 이용한 클러스터 NAS(Clustered NAS)를 대표적인 예로 들 수 있다. 클러스터 NAS는 클러스터 노드나 기억 장치에 데이터와 메타데이터(Metadata)를 분산시키는 것이고, 이 때 사용되는 핵심 기술이 분산파일시스템이다.

결론적으로 말하자면 현재 클라우드 스토리지 기술의 핵심은 분산 파일 시스템이라 볼 수 있다. 클라우드 스토리지 관련 기술은 공개 소프트웨어와 상용 소프트웨어로 분류할 수 있는데, 공개 기술로는 하둡(Hadoop), 글러스터FS(GlusterFS), 스위프트(Swift, Open Stack

Object Storage) 등이 있고 상용 소프트웨어에는 IBM의 Scale-out File Services(일명 SoFS)를 비롯하여 Microsoft, EMC, HDS(Hitachi Data System) 등이 있는데, 선형적인 확장을 지원하는 Scale-out 방식과 유니파이드 스토리지(Unified Storage) 방식으로 클라우드 스토리지 시장을 공략하고 있다.

3.2. 공개용 기술 동향 분석

3.2.1. 하둡(Hadoop)

하둡(Hadoop)은 대량의 자료를 처리할 수 있는 대용량 컴퓨터 클러스터에서 동작하는 분산 응용 프로그램 환경을 지원하는 공개형 소프트웨어 프레임워크이다 [9]. 하둡은 구글(Google)의 맵리듀스(MapReduce)와 GFS(Google File System)을 기반으로 시작되었다. 현재 하둡은 아파치(Apache)의 하부 프로젝트로 운영되고 있으며, 크게 하둡 커널(Kernel), 맵리듀스, HDFS(Hadoop Distributed File System, 이하 HDFS) 등 3개의 부분으로 구성되어 있다. 하둡은 자바 프로그래밍 언어로 구현되고 있으며, Hive, HBase, Zookeeper 등 아파치 기반의 다른 프로젝트와 연관을 가지고 있다. 하둡은 공개용 클라우드 스토리지 관련 기술중에 가장 많은 서브 프로젝트를 가지고 있고, 높은 확장성, 다양한 기능 제공 등으로 가장 많이 사용되고 있고, 상용 업체에서도 기반 기술로 이용중이다. 하둡과 연관된 프로젝트를 정리하면 다음의 [표 1]과 같다.

표 1. 하둡 관련 프로젝트
Table. 1 Sub-projects of Hadoop

프로젝트명	내용
코어	분산 파일시스템과 일반적인 I/O(직렬화, 자바 RPC, 영속(persistent) 데이터구조)를 위한 컴포넌트와 인터페이스를 제공한다.
에이브로(Avro)	교차언어(Cross-Language) RPC와 영속적인 데이터 스토리지를 위한 데이터 직렬화 시스템이다.
맵리듀스	범용 컴퓨터들의 커다란 클러스터에서 수행되는 분산 데이터 처리 모델과 실행 환경이다.

HDFS	범용 컴퓨터들로 된 커다란 클러스터에서 수행되는 분산 파일시스템이다.
피그(Pig)	아후에 의해 개발되었으나, 현재 하둡 프로젝트의 일부로 진행되고 있다. 피그 프로젝트는 하둡 맵리듀스와의 구동을 위해 고급 언어인 피그 라틴(Pig Latin)을 제공한다. 이는 SQL을 사용한 데이터 조작에 익숙한 개발자들에게 보다 손쉬운 접근 수단이 되고 있다. 또한, 자바 API뿐만 아니라, 인터랙티브(Interactive)한 인터페이스를 제공한다. 피그는 카산드라와 Hbase 데이터베이스에서도 사용가능하다.
HBase	분산 컬럼 지향(Column-Oriented) 데이터베이스로서 스토리지로 HDFS를 사용한다. 맵리듀스를 이용한 일괄처리 방식의 계산과 랜덤 읽기가 가능한 포인트 쿼리 방식 모두를 지원한다.
주키퍼(Zookeeper)	다수 컴퓨터로 분산 처리되는 고가용성 조정(Highly available coordination) 서비스로서 분산 응용 프로그램들을 구축하기 위하여 사용될 수 있는 분산 락(Distributed Lock)같은 프리미티브(Primitive)를 제공한다.
하이프(Hive)	하둡을 기반으로 하는 오픈 소스 데이터 웨어하우스 솔루션으로 페이스북에서 최초로 제시하였다. 단순한 쿼리 구성 기능을 제공하는 웹 인터페이스뿐만 아니라, SQL과 유사한 쿼리 언어를 제공한다. 하이브는 애드혹(ad-hoc, 비정형) 쿼리를 제공하는 점이 가장 큰 특징으로 사용하기 쉬운 인터페이스를 제공하여, SQL에 어느 정도 익숙하다면 전문 개발자가 아니더라도 손쉽게 사용가능하다.
추과(Chukwa)	분산 데이터 수집 및 분석 시스템으로 HDFS에 데이터를 저장하는 수집기(Collector)를 수행하고 보고서를 생성하기 위하여 맵리듀스를 사용한다.

캐스케이딩(Cascading)	캐스케이딩 프로젝트는 하둡 맵리듀스 API로 자바 기반의 사용 편의성 제고를 위한 래퍼(wrapper) 역할을 하는 추상화 계층(abstractton layer)이다. 캐스케이딩은 JVM 기반의 언어(Java, JRuby, Clojure 등)를 사용하여 하둡 클러스터에서 복잡한 데이터 작업 흐름을 작성하고 실행하는데 사용한다. 특히, 맵리듀스를 보다 큰 시스템으로 손쉽게 통합시킬 수 있도록 레이어를 가법게 구성하였다.
------------------	---

3.2.2. 스위프트(Swift:OpenStack Object Storage)

2010년 7월에 Rackspace Cloud사와 NASA에 의해 시작된 IaaS 클라우드 컴퓨팅 프로젝트인 OpenStack에서는 클라우드 서비스가 한 기술에 종속되지 않고, 데이터와 애플리케이션을 한 클라우드에서 다른 클라우드로 손쉽게 이전할 수 있도록 하는 것을 목표로 한 오픈 소스(Open Source), 오픈 표준(Open Standard)기반의 개방 커뮤니티를 만들었는데 이렇게 탄생한 것이 OpenStack Object Storage이다[10]. OpenStack Object Storage는 스위프트(Swift)라고도 부른다. 스위프트는 데이터가 전 시스템에 걸쳐 골고루 분산되어 있고, REST기반 API를 지원하며 중앙 DB가 없다. 기본 개발 언어는 Python으로 특정 하드웨어에 무관하게 일반사양의 하드웨어를 지원하며 RAID 등의 구조를 필요로 하지 않는다.

파일시스템이 아니라 계층 구조를 지원하지 않으며 다수의 웹 서버들로 구성된다. 사용자 계정에 대응하는 어카운트(account), 디렉토리에 대응하는 컨테이너(container)와 개별 파일에 해당하는 오브젝트(object)라는 논리적인 구조를 갖는다. 어카운트, 컨테이너 및 오브젝트는 각각 다른 개수의 복사본(replica)을 가질 수 있으며 수 페타바이트(petabyte)까지 확장성이 제공된다. 구글 파일시스템이나 아마존 S3가 큰 오브젝트에 최적의 성능을 보이는 반면 스위프트는 큰 오브젝트뿐만 아니라 수 없이 많은 적절한 크기의 파일들 처리에도 아주 뛰어난 성능을 보인다.

3.2.3. 글러스터FS

글러스터(Gluster)는 2005년에 설립된 오픈 소스 스토리지 솔루션기업으로 비정형 데이터 관리를 표준화

하는 확장형(Scale-out) 스토리지 솔루션 기술을 보유하고 있다. 오픈 소스 소프트웨어와 범용 하드웨어를 이용해 스토리지를 단순화하는 기술을 보유하고 있으며, 클러스터의 핵심제품인 클러스터FS는 다수의 범용 스토리지(S-ATA 드라이브, NAS 시스템 등)와 컴퓨터 리소스를 통합해 중앙관리가 가능하고 글로벌 접근성을 제공하는 고성능의 스토리지 풀을 구축할 수 있다. 2011년 10월에는 공개형 리눅스로 유명한 레드햇에 인수되었다[11].

3.3. 상용 기술 동향 분석

IBM, Microsoft, Dell 등을 비롯하여 EMC, HDS, 넷애플 같은 스토리지 전문 업체들은 빅데이터 처리와 데이터의 빠른 증가속도를 따라갈 수 있도록 선형적인 확장을 지원하는 스케일아웃(Scale-out) 기술과 유니파이드 스토리지 기술을 이용하여 클라우드 스토리지 시장을 공략하고 있다. 스케일 아웃 방식은 초기 도입규모를 최소화한 후 데이터 증가에 따라 필요한 만큼 추가로 구입해 확장하는 방식을 말하는데, 현재 스케일아웃 NAS라는 이름으로 시판되고 있다.

표 2. 스케일아웃 상용제품
Table. 2 Scale-out Commercial Product

회사명	특징
IBM	<ul style="list-style-type: none"> Scale out File Service(SoFS)는 높은 확장성을 갖는 그리드 기반의 솔루션 고성능의 공유디스크 클러스터 파일시스템인 GPFS 기술 보유
Microsoft	<ul style="list-style-type: none"> Windows 2012 기반의 스케일아웃 파일 서버 제품 출시
아이실론(EMC)	<ul style="list-style-type: none"> ‘스케일아웃’이라는 아키텍처에 대한 원천 기술을 확보한 업체 방송 및 미디어 시장에서 최적화된 스토리지 시스템을 제공하던 아이실론이 EMC에 인수되어 빅데이터 처리가 중요한 바이오, 헬스케어 분야로 확대 하둠분산파일시스템기반의 엔터프라이즈 NAS 플랫폼인 ‘EMC 아이실론 스케일아웃(EMC Isilon scale-out NAS)’가 대표적인 솔루션

HDS & 효성인포메이션	<ul style="list-style-type: none"> HIS-Scaleway는 슈퍼컴퓨팅 환경에서 사용되는 클러스터링 기법을 적용한 제품으로 고성능 병렬 파일시스템을 탑재 일반 NAS가 가지고 있던 성능과 확장성의 한계를 극복한 스케일 아웃 NAS 게이트웨이
Dell	<ul style="list-style-type: none"> 스케일아웃 NAS 제조사인 엑사넷 인수를 통해 시장 참여
넷애플	<ul style="list-style-type: none"> 하둠 인프라를 이용하여 페타바이트급까지 확장가능한 ‘넷애플 E-시리즈 플랫폼’ 출시
DDN	<ul style="list-style-type: none"> 스케일 아웃 방식의 스토리지 솔루션 전문 기업인 DDN은 지리적으로 분산된 빅데이터 애플리케이션 환경에 적합하고, 수백 개의 장소에서 20페타바이트 이상 확장할 수 있는 ‘WOS 2.0’ 출시
카디날정보기술	<ul style="list-style-type: none"> 하둠 기반의 분산 파일 시스템을 이용한 확장형 네트워크 스토리지(NAS)인 ‘모나드 네트워크 스토리지’ 출시

유니파이드 스토리지(Unified Storage)는 분산된 환경에서 파일 및 블록을 중앙 집중식으로 관리하고, 멀티프로토콜 파일시스템을 공유하고 확장할 수 있다. DART(Data Access in Real Time) 운영체제에서 NFS 프로토콜과 CIFS 프로토콜의 파일락메커니즘을 사용하여 윈도우, 리눅스 클라이언트간에 파일을 효과적으로 공유할 수 있다. 또한, 블록 어플리케이션에서는 iSCSI 프로토콜을 사용하고, 스토리지 전용 네트워크인 화이버채널(Fibre Channel)을 사용하여 높은 대역폭과 지연시간을 줄일 수 있다.

[그림 2]는 고확장형 유니파이드 스토리지 플랫폼으로 하나의 스토리지 박스에서 어플리케이션 지원을 위한 FC/FCoE, iSCSI, NFS, CIFS의 프로토콜을 통합 지원하고, FC, SAS, SATA, Third Party Arrays의 통합된 스토리지 기술을 지원할 뿐만 아니라, Resource Pooling, Backup, D/R(Disaster Recovery), Archive의 통합된 데이터 클래스를 지원한다[12][13]. 현재 이 제품은 EMC, HDS, 넷애플 등 스토리지 전문업체에서만 공

급하고 있으며, 대응할 만한 공개용 소프트웨어는 없는 실정이다.

그림 2. 유니파이드 스토리지 플랫폼
Fig. 2 Unified storage platform

IV. 클라우드 스토리지 핵심 기술

4.1. 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 장점의 서비스를 모두 지원하는 관련된 연구

와 EMC, NetAPP, IBM, HP 등에서 고확장형 유니파이드 스토리지 솔루션을 제공하고 있다[14][15].

4.2. 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 RAID-0 기술을 활용하여 높은 I/O 성능을 향상시킬 수 있으며, 장애가 발생하더라도 계속적으로 서비스를 할 수 있다[16].

스토리지 가상화 구현에 크게 3가지 방법이 있는데, 호스트기반, 스토리 장치 가상화 기반, SAN 등을 들 수 있다. 현재 추세로 SAN기반에서 가상화 기술이 많이 개발되고 있으며 스토리지 중앙 관리, 용량 등을 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[17].

4.3. 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, 중복 제거가 있다[18].

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[19].

4.4. 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.

데이터 중복 제거는 증분 백업에서 큰 차이점을 가진다. 일반적인 증분 백업은 변경된 파일 전체를 백업하는 반면 데이터 중복 제거는 변경된 파일의 데이터를 중복 제거하여 백업을 수행한다. 따라서 데이터 중복 제거는 변경된 파일에는 중복된 데이터가 존재하지 않게 된다[20]. 데이터 중복 제거의 핵심은 효율적인 데이터 스토리지 감소와 데이터 인스턴스 유지이다[21]. 일반적인 데이터 중복 제거는 데이터 압축 기술을 사용하여 중복 데이터를 제거한다. 하지만 이는 데이터 블록 또는 파일이 위치한 동일한 스토리지 시스템에서 압축 알고리즘을 사용하여 중복 제거를 해야하는 단점이 있다[22]. 데이터 중복 제거를 간단히 표현하면 [그림 3]과 같다.

그림 3. 데이터 중복제거
Fig. 3 Data Deduplication

중복 제거 소프트웨어에서 생성된 백업 데이터 스트림 인덱스를 저장소에 저장하여 데이터 블록 유무를 판단하고 디지털 서명(fingerprint)을 사용하여 오버헤드 없이 중복된 데이터를 찾을 수 있도록 한다. 이때 중복된 데이터는 원본 데이터 세그먼트의 포인터를 링크한다.

4.5. 스토리지 보안

클라우드 스토리지는 네트워크를 통해 스토리지 서비스를 이용하기 때문에 네트워크를 통한 클라우드 사용자의 많은 데이터를 포함할 수 있어야 한다. 클라우드 특성상 가상화 센터와 클러스터 서버들은 하나의 지역에 국한되어 있지 않고, 여러 지역으로 분산되어 있어 잘못된 접근으로 인해 개인 정보 보호 문제가 있기 때문에 반드시 클라우드 서비스 공급자로 접속을 해야 한다[23]. 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.

V. 결 론

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

본 논문에서는 IT 인프라의 핵심인 클라우드 컴퓨팅을 위한 클라우드 스토리지 기술에 대해 살펴 보았다. 공개형 기술과 여러 관련된 대형 벤더에서는 스토리지 용량과 확장이 가능하게 융합하여 스케일러블(scalable) NAS, 또는 Scale-out NAS라는 기술 서비스를 제공하여 클라우드 스토리지 시장에 빠르게 진입하여 공략하고 있다. 또한, 전문 스토리지 업체에서는 스케일아웃 NAS 방식이외에 유니파이드 스토리지를 가지고 클라우드 스토리지 시장을 추가로 공략하고 있다. 클라우드 스토리지는 빅데이터와 웹 3.0 시대에 대비하기 위한 핵심 기술이라고 할 수 있으며, 스케일 아웃 NAS와 고확장형 유니파이드 스토리지 솔루션이 주축을 이룰 것이라 사료된다.

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, March 4, 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 춘계학술대회 논문집, pp. 294-297, 2011년 5월.
- [8] Jeff O'neal, "클라우드를 위한 스토리지 인프라", <http://www.netapp.com/kr>, 2009.
- [9] Hadoop Project, <http://hadoop.apache.org>
- [10] Openstack, <http://www.openstack.org>
- [11] GlusterFS, <http://www.gluster.org>
- [12] EMC, <http://www.emc.com/>
- [13] NepApp, <http://www.netapp.com/us/>
- [14] IBM, <http://www.ibm.com/>
- [15] HP, <http://www.hp.com/>
- [16] 김영철, 차명훈, 이상민, 김영균, "클라우드 컴퓨팅에서 스토리지 가상화 기술 동향", ETRI 전자통신동향분석, 제24권, 제4호, pp. 69-78, 2009년 8월.
- [17] 차명훈, 이상민, 김 준, 김영균, 김명준, "대규모 분산 파일 시스템 환경의 메타데이터 관리", ETRI 전자통신동향분석, 제22권, 제3호, pp. 154-165, 2007년 6월.
- [18] Jacob Farmer, "A Crash Course in Wide Area Data Replication", SNIA, 2009.
- [19] Qinlu He, Zhanhuai Li, Xiao Zhang, "Analysis of the key technology on cloud storage", FITME, IEEE Confrence Publications 2010, pp. 426-429, 2010.
- [20] 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.
- [21] 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.
- [22] Thomas Rivera, "Understanding Data Deduplication", SNIA, 2009.
- [23] Cloud computing development status, <http://hi.baidu.com/lmc625263041/blog/item/122aeb1d0a492bd309a50276b.html> 2009-03-01 16:17

저자소개

박정수(Jeong-Su Park)

2007년 2월 청주대학교 컴퓨터 공학과 공학학사
2012년 3월~현재 숭실대학교 정보과학대학원 소프트웨어공학 석사과정

2007.8~현재 (주)매크로임팩트 팀장
※관심분야: 리눅스, 운영체제, 클라우드 컴퓨팅, SW테스트, 빅데이터

배유미(Yu-Mi Bae)

2005년 2월 한남대학교 컴퓨터멀티미디어학과 학사
2007년 8월 한남대학교 정보기술학과 석사

2013년 2월 한남대학교 컴퓨터공학과 박사
2013년 3월~현재 한남대학교 컴퓨터공학과 겸임교수
※관심분야: 리눅스, 정보보호, 서버가상화, 멀티미디어, 웹디자인, 클라우드 컴퓨팅

정성재(Sung-Jae Jung)

1998년 2월 한남대학교
컴퓨터공학과 학사

2003년 8월 한남대학교
컴퓨터공학과 석사

2011년 2월 한남대학교 컴퓨터공학과 박사

2005년 3월~2010년 2월 한남대학교 국제IT교육센터
전임강사

2012년 9월~현재 (주)스컴씨엔에스 기업부설연구소장
※ 관심분야: 리눅스, 정보보호, 서버 가상화, 운영체제,
클라우드 컴퓨팅