

클라우드 컴퓨팅 기반 에뮬레이션 전략을 활용한 전자기록 장기보존 방안 연구

A Study on Long-Term Electronic Records Preservation Using Cloud-Based Emulation Strategy

이봉환(Bong-Hwan Lee)¹, 한희정(Hui-Jeong Han)²,
조철용(Cheolyong Jo)³, 왕호성(Ho-sung Wang)⁴,
양동민(Dongmin Yang)⁵

E-mail: blee@dju.kr, freebirdhj@naver.com, cheolyong@innogrid.com,
kinghosung@gmail.com, dmyang@jbnu.ac.kr

¹ 제1저자 대전대학교 전자·정보통신공학과 교수

² 전북대학교 문화융복합아카이빙연구소 전임연구원

³ ㈜이노그리드 실장

⁴ 국가기록원 기록연구사

⁵ 교신저자 전북대학교 기록관리학과 부교수, 문화융복합아카이빙연구소 연구원



논문접수 2019.7.29

최초심사 2019.8.4

게재확정 2019.11.10

ORCID

Bong-Hwan Lee
https://orcid.org/0000-0003-0088-0530

Hui-Jeong Han
https://orcid.org/0000-0003-2153-5963

Cheolyong Jo
https://orcid.org/0000-0003-1167-4695

Ho-sung Wang
https://orcid.org/0000-0002-1955-7998

Dongmin Yang
https://orcid.org/0000-0002-4029-9372

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

- 본 연구는 "2019년 행정안전부 국가기록원 기록관리 연구개발사업"의 연구비를 지원받아 수행되었음.
- 이 논문은 2018학년도 대전대학교 교내 학술연구비 지원에 의해 연구되었음.

http://ras.jams.or.kr

초 록

전자기록 장기보존의 핵심은 전자기록이 최초로 생성되고 활용되었던 본래의 기능적 속성과 모습 등을 담고 있는 비트스트림을 오랫동안 그대로 유지하는 것이다. 대부분 국내의 주요 아카이브 기관들은 경제적, 기술적인 측면을 고려하여 원본의 비트스트림을 담고 있는 포맷의 안정성, 신뢰성에 따라 새로운 포맷으로의 변환을 허용하는 마이그레이션을 주전략으로 채택하고 있다. 그러나 전자기록 유형이 다양해지고 범위가 확장됨에 따라 마이그레이션 단일 전략은 다양한 문제점들을 발생시키고 있다. 그러므로 본 연구에서는 클라우드 컴퓨팅 기술을 활용하여 전자기록의 본래의 기능적 속성과 모습을 담고 있는 비트스트림을 변경하지 않으면서, 전자기록이 생산·활용된 시스템 및 응용 환경까지 재생산하는 에뮬레이션 전략의 적용 가능성을 검토하고자 한다.

ABSTRACT

The key to the long-term preservation of electronic records is the long-term retention of the bitstream that contains the original functional attributes and appearances of the electronic records that were originally created and utilized. Considering economic and technical aspects, most of the major domestic and foreign archives have adopted a migration strategy that allows conversion to a new format depending on the stability and reliability of the format containing the original bitstream. However, as the type of electronic records varies and its scope continuously expands, several problems arise if a single migration strategy is observed. Therefore, the applicability of the emulation strategy is examined in this paper to reproduce the system and application environment where electronic records are created and utilized. As such, electronic records can be maintained without any changes in the bitstream by utilizing cloud computing technology.

Keywords: 장기보존, 전자기록, 클라우드 컴퓨팅, 에뮬레이션, 마이그레이션

Long-term Preservation, Electronic Records, Cloud Computing, Emulation, Migration

1. 서론

1.1 연구배경 및 목적

디지털 기술의 발전으로 전자기록의 활용도가 급속도로 증가하고 있으며, 이에 따라 기존의 종이기록에서 전자기록 관리체계로 전환되어가고 있다. 디지털 객체로 생성되고 활용되는 전자기록은 종이기록과 달리 객체의 비트스트림이 조금이라도 훼손되면 전자기록을 확인할 수 없을 수도 있다. 현재 전자기록의 생산 및 활용을 위한 관련 기술들은 많이 있으며, 그 기술들이 하드웨어에서 응용 소프트웨어까지 서로 조직적이고 복잡하게 연결되어 있기 때문에 하나의 연결고리라도 끊기거나 이상이 생기면 전자기록을 사용할 수 없게 된다. 또한 전자기록 관련 기술들이 빠르게 발전함에 따라 새로운 기술들도 계속 등장하고 있어 오랜 시간이 지난 후에는 전자기록을 확인할 수 없는 상황이 발생할 수 있다. 이렇듯 기록 자체를 확인할 수 없는 위험한 상황 때문에 전자기록의 장기보존은 종이기록에서 전자기록으로 전환되는 시점에서 해결되어야 하는 가장 중요한 문제이다. 이를 해결하기 위해 아카이브 기관들은 전자기록 장기보존을 위한 다양한 전략을 수립하고 있다.

국내의 주요 아카이브 기관들은 전자기록 장기보존을 위한 주전략으로 마이그레이션(Migration)을 채택하고 있다. 마이그레이션 전략을 통해 대부분의 전자기록에 대한 일괄적인 보존전략을 세울 수 있다. 일반적으로 마이그레이션을 수행하는 디지털 객체의 포맷 변환 기술을 해당 포맷을 개발한 기업에서 기본적으로 제공하

거나 오픈 소스 프로젝트로 무료로 배포되어 왔다. 따라서 마이그레이션 전략은 상대적으로 기술적인 장벽이 낮고 비용이 적다는 이유로 많이 채택되어 왔다. 그러나 마이그레이션 방식은 디지털 객체의 포맷을 더 안정적인 포맷으로 변환하는 방식이기 때문에 비트스트림이 변경될 수 밖에 없고 이로 인해 비트스트림의 손실이 발생한다는 본질적인 단점을 갖고 있다. 전자기록의 유형이 전자문서에서 시청각기록물, 웹기록물, 행정정보데이터세트 등 다양한 형태로 그 범위가 확장되고 있으며, 디지털 객체는 물론 해당 객체를 구동하는 시스템 하드웨어 및 소프트웨어를 포함한 운영 환경까지 보존하는 것을 고려해야 하는 상황들이 생기고 있다. 이러한 상황을 해결하기 위해 장기보존 전략 중 하나인 에뮬레이션(Emulation) 방안이 대두되고 있다. 에뮬레이션 전략은 소프트웨어 기능 및 디지털 객체가 생성되고 활용되었던 예전의 모습을 그대로 현재의 컴퓨팅 환경에서 재현하기 위해 당시 컴퓨팅 운영 환경을 재생산하는 것을 의미한다. 에뮬레이션 전략은 기록물의 생산 당시의 모습 그대로 보존하고 재현할 수 있다는 점에서 기록관리 분야에서 상당한 관심을 받고 있으며 마이그레이션 전략과 함께 지속적으로 논의되고 있다. 다만, 에뮬레이션 전략을 구현하기 위한 경제적인 비용이 높고, 여전히 기술적으로 해결해야 할 난제들이 남아 있기 때문에, 여러 차례 검토 및 연구들이 진행되어 왔음에도 불구하고 적극적으로 활용되지는 못하고 있다.

최근에 기업뿐 아니라 중앙정부부처와 공공기관들까지 자체적으로 서버들을 운영하지 않고, 처음부터 클라우드 컴퓨팅(Cloud Computing)

환경에 서버들을 구축하거나, 기존의 서비스를 제공하고 있던 시스템들을 클라우드 컴퓨팅 환경으로 전환하는 속도가 가속화되고 있다. 클라우드 컴퓨팅 환경에서 사용되고 있는 가상화(Virtualization) 기술은 에뮬레이션의 핵심 기술로 실제 클라우드 전환 사업에 활발히 도입되는 단계까지 올 정도로 성숙된 기술이다. 그러므로 클라우드 컴퓨팅 기술은 장기보존 전략에 에뮬레이션을 적용하는데 걸림돌이 되어 온 여러 가지 문제들을 해결하면서 전자기록을 장기보존하는데 활용될 수 있다.

클라우드 컴퓨팅 기술은 가상화 기술을 기반으로 일반 사용자 또는 기업에게 CPU, 메모리, 디스크 등의 '인프라', 소프트웨어 개발 환경인 '플랫폼', 업무 수행을 위한 '소프트웨어'를 본인의 PC에 설치하지 않고 인터넷 연결을 통해서 서비스를 제공하게 하는 기술이다. 데이터 센터(Data Center)와 같은 대규모 비용이 들어가는 기반시설을 직접 설치·유지 관리하지 않고, 필요한 만큼만 컴퓨팅 자원을 사용할 수 있으며, 서버 구매, 유지보수 비용을 크게 절감할 수 있다는 믿음 때문에 연구 개발되어 왔다.

클라우드 컴퓨팅은 기존 시스템 전환 및 신규 시스템 구축 관점 이외에 장기보존 관점에서도 활용가치가 높은 기술이다. 원하는 사양의 컴퓨팅 자원을 유연하게 제공해 줄 수 있는 클라우드 컴퓨팅의 가상화 기술과 분산처리 기술은 전자기록이 생산·활용되는 시스템 운영 환경을 보존하는 방식인 에뮬레이션 전략에 대한 해결책을 제시할 수 있다. 클라우드 컴퓨팅 기술과 서버 백업 시 사용되는 도구(예, ZConverter, virt-p2v 등)와 결합된다면 전자기록이 생산되었던 시스템과 유사한 운영 환경을 만드는 수

준을 넘어 완전히 동일한 환경을 제공할 수 있기 때문이다.

에뮬레이션은 본래의 기능과 모습을 재현한다는 점에서 기록관리 측면에서 가장 최적화된 방식으로 주목받았지만 경제적인 비용과 기술적인 한계로 도입이 어려웠다. 클라우드 컴퓨팅은 에뮬레이션 방식을 장기보존 전략에 도입하는데 있어 소용되는 경제적인 비용을 낮추고 기술적인 한계를 극복할 수 있는 방안을 제시할 수 있다. 다양한 유형의 전자기록들을 보존하는데 있어 마이그레이션 단일 전략으로는 한계가 드러나고 있는 상황에서 클라우드 컴퓨팅 기술을 기반으로 하는 에뮬레이션 전략의 타당성을 적극적으로 검토하는 것이 필요하다.

이에 본 논문에서는 클라우드 컴퓨팅 기반 에뮬레이션 전략 방안을 연구하여 이를 활용한 전자기록 장기보존 방안을 제시하고자 한다. 또한, 실제 에뮬레이션을 수행할 수 있는 오픈스택(OpenStack)을 활용하여 테스트베드를 구축하고 에뮬레이션 기법을 실험하여 클라우드 컴퓨팅 기반 에뮬레이션 전략의 실제 적용 가능성을 타진하고자 한다.

1.2 연구범위 및 방법

본 연구에서는 클라우드 컴퓨팅 기반 에뮬레이션 전략을 활용한 전자기록 장기보존 방안을 제시하기 위해 가장 먼저 국내외 클라우드 컴퓨팅 기술을 조사 및 분석한다. 조사 및 분석 결과를 바탕으로 가장 현실적이면서 전자기록 장기보존에 적합한 형태의 테스트베드를 구축한다(〈표 1〉 참고). 다양한 전자기록 장기보존을 위한 에뮬레이션 테스트 및 분석을 통해 다

〈표 1〉 클라우드 컴퓨팅 테스트베드 구성

구분		내용
하드웨어 구성	컴퓨팅 서버	2대(CPU: Intel Xeon, 12Core, 2.1GHz/85W)
	스위치	2대(통합스위치: 10G 12Port, 관리스위치: 1G 8Port)
소프트웨어 구성	클라우드 구성	오픈스택(OpenStack) Queens
	가상화	KVM(Kernel-based Virtual Machine)
	에뮬레이터	QEMU(Quick Emulator)
	클라우드 관리	오픈스택잇(OpenStackit)
	시스템(템플릿) 이미지 변환 도구	ZConverter, virt-p2v

양한 전자기록 유형에 대해 에뮬레이션 기반 장기보존 방안을 제시한다.

1.3 선행연구 및 사례조사

국내 전자기록 클라우드 컴퓨팅 기술을 전자 기록관리에 활용하기 위한 연구는 상당히 활발하게 진행되어 왔다. 그러나 대부분의 연구는 기존의 전자기록관리시스템을 클라우드 컴퓨팅 기반 환경으로 변환하는 내용에 초점이 맞춰져 있다. 이승억, 설문원(2017)에서는 기록관리의 환경이 변화함에 따라 우리나라 전자기록관리가 전면적으로 재설계되어야 함을 주장하면서 향후 다루어야 할 정책 영역과 그 방향성을 제시하였다. 새로운 환경변화에 대처함에 있어 클라우드 컴퓨팅은 전자기록관리의 핵심 플랫폼으로 간주되고 있다. 임지훈, 김은총, 방기영, 이유진, 김용(2014)는 대량으로 생산되는 전자기록을 관리하기 위해 비용대비 효과가 뛰어난 클라우드 컴퓨팅을 전자기록관리시스템에 도입하는 방안을 제시했다. 또한, 해외 아카이브 기관의 클라우드 컴퓨팅 활용 사례 및 클라우드 컴퓨팅 환경에 대한 기록물 관리 지침서 등을 기반으로 클라우드 컴퓨팅 도입 방안에 대해 구체적으로는 구성하고, 현재 전자

기록관리시스템의 현황을 분석하여 네 가지 문제점을 도출했다. 이를 고려하여 기록관 업무분야와 기존 시스템의 구성요소를 중심으로 클라우드 컴퓨팅 환경 모델 적용 방안과 전자기록관리시스템 보안 프로세스를 제시하였다. 정예용, 심갑용, 김용(2014)는 전자기록관리시스템 중에서 중앙기록물관리시스템인 영구기록관리시스템의 보존 및 활용 방식을 제안하였다. 영구기록물관리기관의 업무, 자료, 시스템의 특성을 파악하고 이를 기반으로 하는 클라우드 컴퓨팅 기반의 영구기록물관리시스템의 보존 및 활용 모델을 도출하였으며, 하드웨어 및 소프트웨어 자원을 클라우드 컴퓨팅 모델별로 매핑하고 업무 프로세스를 클라우드 컴퓨팅 기반으로 전환하는 방안을 제시하였다. 김기정, 신동수(2018)도 영구기록물관리시스템을 클라우드 컴퓨팅 환경으로 전환하는 방안을 제시하였다. 이 연구는 2017년 국가기록원 연구개발 사업 '차세대 기록관리 모델 재설계 연구'의 일환으로 수행된 연구로 국가기록원 영구기록물관리시스템에 대한 구체적인 조사 및 분석이 이루어졌으며 이를 기반으로 클라우드 환경으로 전환하기 위한 단계적인 방안을 제시하였다. 김주영, 김순희(2019)는 전자기록생산시스템인 업무관리시스템 또는 온-나라 시스템에서 기록관

리시스템으로 전자기록물을 이관할 때 발생하고 있는 디지털 객체의 무결성 훼손 문제를 해결하기 위해 클라우드 저장소를 활용한 논리적 전자기록 이관 방안을 제안하였다.

기존 전자기록관리 분야의 클라우드 컴퓨팅 기술 활용 연구는 기존 각 기관들이 보유하고 있는 다양한 전자기록관리의 시스템들을 안정적으로 클라우드 컴퓨팅 환경으로 전환하여 비용과 관리의 효율성을 제고하는 방안을 제시하는데 그치고 있다. 이는 우리나라가 전자기록을 생산·활용·관리·이관·보존하는 온-나라 업무관리시스템과 RMS(Records Management System)를 클라우드 컴퓨팅 기술 기반으로 통합하는 것을 추진하고 있기 때문에 기존의 국내 연구들 역시 클라우드 컴퓨팅 환경으로 전환하기 위한 시스템 통합에 관한 연구가 주로 이루어진 것으로 보인다.

한편, 에뮬레이션 전략에 관한 국내 연구는 다음과 같다. 먼저 김명훈, 오명진, 이재홍, 임진희(2013)는 장기보존 전략으로서 에뮬레이션이 지닌 장단점을 도출한 다음, 에뮬레이션 관련 선진 사례인 CAMILEON, KB, Planets, KEEP 프로젝트를 조사, 분석하였다. 이러한 분석을 기반으로 향후 우리나라 전자기록에 대한 시사점 및 적용 방향을 제시하였다. 국가기록원(2013)은 한국형 에뮬레이션 전략 방안을 제시하였으며, Oracle에서 오픈 소스로 배포하고 있는 VirtualBox라는 VM(Virtual Machine) 에뮬레이터를 기반으로 한국형 에뮬레이션 프로토타입을 개발하여 전자기록 장기보존 전략으로서의 에뮬레이션 방식의 적용 가능성을 점검하였다. 이들 연구는 국내외 에뮬레이션 사례를 구체적으로 조사 및 분석하였고, 에뮬레

이션 전략을 타당성 점검을 위해 상용 에뮬레이터를 활용하여 타당성을 점검하였다는 점에서 의미를 갖는다. 그러나 전자기록의 에뮬레이션 적용 가능성을 검토하고 있지만 기술적인 어려움과 높은 경제적인 비용으로 현재로서는 적용 가능성에 대해 의문을 남기고 있다. 반면, 왕호성, 설문원(2017)은 행정정보데이터세트 기록의 재현과 관련하여 에뮬레이션에 주목하였다. 즉, 에뮬레이션 기법은 가상화, 클라우드 서비스, 원격데스크톱 프로토콜(RDP: Remote Desktop Protocol)을 결합하여 구현될 수 있으며, 이를 서비스로서의 에뮬레이션 전략(EaaS: Emulation as a Service)으로 제안하였다. 또한, 전자기록관리 관련된 문제들을 정보통신 분야와의 협력을 통해서 극복할 수 있음을 언급하고 있다.

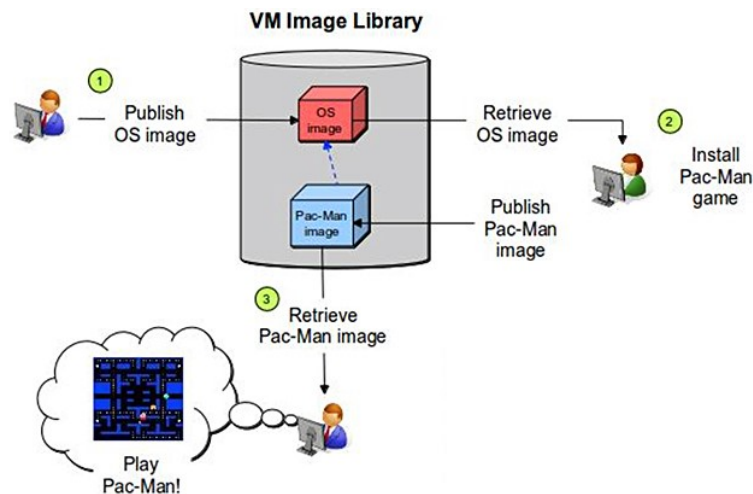
선행연구를 종합해 보면, 에뮬레이션은 다양한 측면에서 전자기록 장기보존 전략으로서의 장점이 있지만, 주전략으로 도입하기에는 경제적인 비용과 기술적인 한계 때문에 실현가능성이 낮다고 여겨졌다. 그러나 상용화 수준으로 기술이 성숙되어 가격 경쟁력을 가지고 있는 가상화, 클라우드 컴퓨팅, 원격접속 기술을 활용한다면 에뮬레이션은 또 하나의 장기보존 전략으로 고려될 수 있을 것으로 판단된다. 이에 본 연구에서 제안하는 클라우드 컴퓨팅 기반 에뮬레이션 전략을 활용한 전자기록 장기보존 방안 연구는 시의적절하다고 판단된다.

미국 카네기 멜론 대학교와 IBM연구소가 공동으로 추진하고 있는 올리브(OLIVE)¹⁾ 프로젝트가 대표적인 클라우드 컴퓨팅 기반 에뮬레이션 기법이다(OLIVE, 2019). 올리브는 Windows 3.1이나 초기 맥(MAC) 컴퓨터의 환경을 가상화

기술로 이용하여 응용프로그램, 디지털 문서 파일, 운영 체제까지 포함한 “가상화 머신 이미지 또는 템플릿 이미지”들을 만들고 클라우드 컴퓨팅 환경 내부에 설치하여 누구나 외부에서 클라우드 컴퓨팅 환경에 접속하여 디지털 문서 파일을 원격으로 열람할 수 있는 서비스를 제공한다.

〈그림 1〉은 올리브 시스템 내에서 어떻게 콘텐츠들(OS, 응용SW 등)이 추가되고 운용되는지를 보여준다. 클라우드 컴퓨팅 기술로 구현된 “VM Image Library”에 사용자들이 OS와 해당 OS에서 동작하는 응용SW(Pac-Man)를 설치 및 업로드하면, 또 다른 사용자는 응용 SW가 탑재된 OS를 클라우드 컴퓨팅 환경에 설치하여 응용SW를 활용할 수 있는 구조를 통해 서비스를 제공한다.

또 다른 사례로 독일의 bwFLA(Baden-Wuerttemberg Functional Long-term Archiving and Access) 프로젝트를 들 수 있다(bwFLA, 2019). 독일의 bwFLA는 클라우드 컴퓨팅 환경 기반의 에뮬레이션 기법을 구현하여 에뮬레이션 전략을 서비스(EaaS: Emulation As A Service)로서 제공하는 것을 목표로 하고 있다. 즉, 복잡한 디지털 기록이 지닌 특성을 장기보존할 수 있으며 본래의 기능 및 모습을 경험하기 위한 가장 좋은 방법은 생성 당시의 응용 프로그램을 사용하는 것이라는 명제에서 출발하여 HW와 OS의 가상화 및 클라우드 컴퓨팅을 통해 구형 응용 프로그램을 실행시킬 수 있으며 일반인에게도 쉽게 접근할 수 있도록 구현하였다. MAAS/Juju, OpenStack로 클라우드 컴퓨팅 환경을 구축하였고, PPC, m68k, Intel-based



〈그림 1〉 올리브의 기술 구현 매커니즘

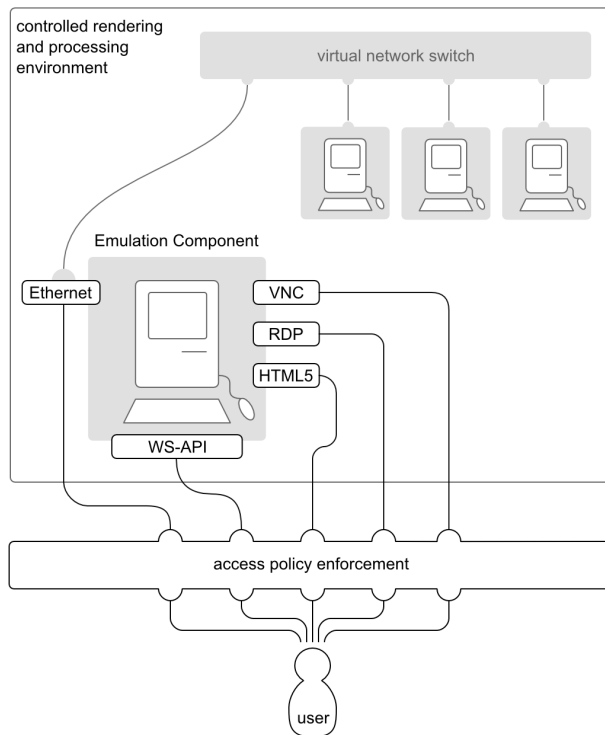
* 출처: Olive Archive, <https://olivearchive.org/about/>

1) OLIVE는 Open Library of Images for Virtualized Execution의 약자로 가상화 실행을 위한 템플릿 이미지를 제공하는 열린 도서관을 의미함.

x86등의 주요 CPU 구조와 OS/2, MS Windows, MacOS 등 주요 OS를 지원한다. 또한, <그림 2>처럼 사용자에게는 VNC,²⁾ RDP, HTML5, WS-API,³⁾ Ethernet 등을 통해 인터넷으로 접근 서비스를 제공한다. 또한, 미국의 스탠포드 대학에서 약 40여 명의 디지털 아키비스트, 프로그래머, IT엔지니어, 학자, 기록관리 실무자들이 공동으로 출범시킨 BDAX(Born Digital Archiving & eXchange) 프로젝트는 2016년도 미국 아키비스트 협회(SAA) 연례회의에서

독일의 bwFLA의 에뮬레이션 기법을 실험하겠다는 계획을 밝혔다(왕호성, 설문원, 2017).

국외에서는 클라우드 컴퓨팅 환경에서의 에뮬레이션 서비스를 구축하고 제공하기 위한 연구 및 개발을 통해 상당한 결실을 맺고 있다. 그러므로 클라우드 컴퓨팅 기술이 성숙되어 있는 현재 시점에서 국내에서도 가능한 신속하게 클라우드 컴퓨팅 기반 에뮬레이션 전략을 활용한 전자기록 장기보존 방안을 연구하는 것이 필요하다.



<그림 2> bwFLA 시스템 구조와 서비스 제공

* 출처: bwFLA, <http://eaas.uni-freiburg.de/>

- 2) VNC는 Virtual Network Computing의 약자로 RFB(Remote Framebuffer Protocol)를 이용하여 원격으로 다른 컴퓨터를 제어하는 시스템임.
- 3) WS-API는 Web Service Application Programming Interface의 약자로 Lua 웹 응용 프로그램에서 웹서버를 추상화한 API임.

2. 이론적 배경

2.1 장기보존 전략

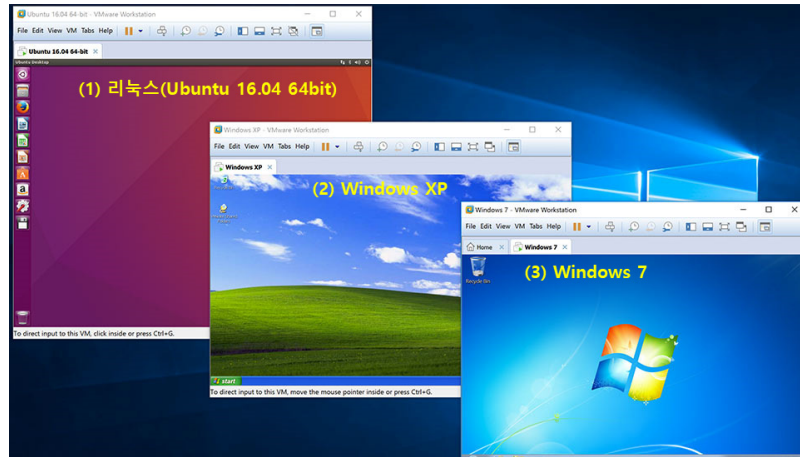
마이그레이션(Migration), 에뮬레이션(Emulation), 인캡슐레이션(Encapsulation)은 대표적인 전자기록의 3대 장기보존 전략이다(DPT, 2006; 임진희, 2013). 먼저 마이그레이션은 기술의 발전, 시스템 노후화로 인하여 운영체제 및 소프트웨어 업그레이드, 시스템 교체 및 업그레이드 등을 수행할 때 전자기록물이 온전하게 접근 및 재현이 가능하도록 기존의 디지털 객체를 안정적인 형식으로 변환하는 전략을 뜻한다. 또한, 현재의 파일 포맷 방식이 새로운 시스템에서 재현이 불가능해 지거나 새로운 버전으로 업그레이드 될 경우에 오랜 시간이 지난 이후에도 재현 확률이 가장 큰 안정적인 파일 포맷으로 변환하는 방식이다. 예를 들어, 온나라 문서 2.0에서 아래아 한글 또는 MS Word로 생성되었던 전자문서가 영구기록물관리시스템으로 이관될 때, 장기보존 되기 전에 문서보존 포맷인 pdf/a 파일로 변환하는 것은 우리나라의 장기보존 전략이 마이그레이션이기 때문이다. 시스템 전체를 유지해야 하는 에뮬레이션 전략보다 경제적인 비용과 기술적인 면에서 장점이 있다. 그러나 안정적인 포맷을 확보하기 위해 전자기록 유형별 그리고 파일 포맷별로 시스템 변경이나 소프트웨어 업그레이드에 대한 지속적인 모니터링이 필요하다. 또한, 마이그레이션은 근본적으로 포맷 변환이 이루어질 때 비트스트림 손실이 발생한다. 이로 인해 일부 본래의 모습이나 기능이 재현이 불가능할 수 있어 원본의 룩앤필(Look and Feel)이 달라질 수 있

으며 결국, 마이그레이션은 진본성을 손상시킬 가능성이 있다는 본질적인 한계를 가지고 있다.

에뮬레이션은 하나의 컴퓨터 안에 전자기록물이 최초로 생산되었을 때와 동일한 형태의 하드웨어 및 소프트웨어로 구성된 가상의 다른 컴퓨터를 실행시키고 그 가상의 컴퓨터에서 전자기록물을 재현할 수 있도록 하는 보존전략이다. <그림 3>은 Windows 10에서 VMWare Workstation Pro를 실행시켜 (1) 리눅스(Ubuntu 16.04 64bit), (2) Windows XP, (3) Windows 7의 세 개의 가상 컴퓨터를 실행시킨 화면을 보여 준다.

에뮬레이션은 가상의 컴퓨터를 실행시킬 수 있는 에뮬레이터(Emulator)란 소프트웨어가 필요하다(김명훈 외, 2013). VMWare, VirtualBox, Hyper-V 등이 대표적인 상용 에뮬레이터이다. 에뮬레이션 전략은 파일포맷의 버전 업그레이드가 필요 없고 원본 그대로의 모습을 재현할 수 있다는 큰 장점이 있다. 반면, 에뮬레이터와 함께 다양한 운영체제와 응용 프로그램을 담고 있는 대용량 시스템(템플릿) 이미지 파일을 보관하고 유지하기 위한 비용과 가상화 기술을 보유하고 있는 전문 인력 등이 요구된다. 그 외에도 시스템 노후화로 인하여 파일, 에뮬레이터, 시스템 이미지 파일의 저장매체 이관 작업이 여전히 필요하다.

대부분의 아카이브 기관에서는 마이그레이션과 에뮬레이션 중에서 하나를 주전략으로 채택하거나 하나는 주전략, 다른 하나는 부전략으로 사용하고 있다. 소정의, 한희정, 양동민(2018)에서 조사한 5개의 국립 아카이브 기관 중에서 4개 기관이 마이그레이션을 1개 기관에서는 에뮬레이션을 주전략으로 채택하고 있었다.



〈그림 3〉 VMWare Workstation Pro 실행 화면

마지막으로 인캡슐레이션 전략이다. 인캡슐레이션은 관련 메타데이터와 무결성 메시지를 원본 문서와 함께 하나의 개체로 패키징(Packaging)하는 과정이다. 이는 향후 메타데이터를 통해 기록물을 이해하고 신뢰성과 진본성을 제공하는데 도움을 줄 수 있으며, 무결성 메시지를 통하여 기록의 무결성과 진본성을 검증할 수 있다. 인캡슐레이션 전략은 마이그레이션 또는 에뮬레이션 전략을 선택하는 것과 상관없이 적용할 수 있으며 전자기록물의 신뢰성, 진본성을 제공하기 위한 가장 일반적인 방법으로 대부분 아카이브 기관에서 도입하고 있다.

현재 국가기록원 장기보존 주전략은 마이그레이션이다. 마이그레이션은 기능보다는 원본의 내용, 문맥, 외관 등에 대한 특성들을 보존해야 하는 전자문서 유형 장기보존에는 상당히 최적화 되어 있다. 그러나 기록물의 기능을 재현하거나 멀티미디어, 하이퍼텍스트 등의 장기보존에는 마이그레이션 전략이 적합하지 않다. 그러므로 전자문서 이외 전자기록에 대해 에뮬레이션 전략으로 장기보존하는 방안을 모색하

여 마이그레이션과 함께 에뮬레이션을 장기보존 주전략의 하나로 추가하는 것이 필요하다.

2.2 클라우드 컴퓨팅(Cloud Computing)

클라우드 컴퓨팅은 4차 산업혁명 시대의 기반 인프라 기술로 인식되고 있다. 클라우드 컴퓨팅은 언제 어디서나 필요한 만큼의 컴퓨팅 자원을 필요한 시간만큼 인터넷을 통하여 활용할 수 있는 컴퓨팅 방식이다. 예를 들어, 빅데이터 및 인공지능 분야에 뛰어들어 스타트업 기업이 초창기부터 빅데이터의 수집, 저장, 분석을 위한 방대한 컴퓨팅 자원과 인공지능 개발을 위한 슈퍼컴퓨터 등을 자체적으로 구매하고 유지관리 하는 것은 비현실적이다. 이때, 스타트업 기업이 실제 서버나 네트워크 장비를 구매하지 않고, 클라우드 컴퓨팅 서비스를 통해서 필요한 만큼만 비용을 지불하고 컴퓨팅 자원을 사용한다면 엄청난 경제적인 효과를 얻을 수 있을 것이다. 이처럼 자본력이 부족한 중소기업이나 스타트업은 클라우드 컴퓨팅 서비스를

활용하면 대규모 컴퓨팅 자원을 저렴하게 활용할 수 있다. 가트너(Gartner)에서는 2021년까지 공용 클라우드 시장이 매년 17.6%씩 성장할 것으로 예상하고 있으며, IaaS(Infra structure As A Service)와 SaaS(Software As A Service) 분야가 가장 빠르게 성장할 것으로 전망되었다 (강맹수, 2019).

클라우드 컴퓨팅 개념은 1970년부터 조금 다른 의미로 또는 일부 개념으로 언급되었고, 그 실체가 드러난 것은 2000년 이후이다. 특히, 아마존의 AWS(Amazon Web Services)와 마이크로소프트의 애저(Azure)가 대중적으로 사용되면서 클라우드 컴퓨팅 분야는 급격히 발전했다. 대부분 기업과 개인의 민간 분야에서 활발하게 활용되었으며, 최근에는 공공 분야까지 도입되었다. 국가정보자원관리원에서 운영하고 있는 G-클라우드(구, 정부통합전산센터)가 우리나라의 대표 공공 클라우드 컴퓨팅 서비스이며, 중앙부처 및 산하 공공기관까지 각 기관 별로 운영하는 시스템들을 G-클라우드에 이관하고 있다. 행정기관의 클라우드 전환율은 2016년 25.80%, 2017년 32.93%, 2018년 39.96%으

로 점점 늘어나고 있는 추세이다. 또한 클라우드의 주요 특징은 <표 2>와 같이 정리할 수 있다.

클라우드 컴퓨팅 기술은 가상화(Virtualization) 기술과 분산처리(Distributed Processing) 기능이 핵심이다. 가상화는 컴퓨터에서 컴퓨터 자원의 추상화(Abstraction)하는 것을 말하는 상당히 포괄적인 용어이다. 가상이라는 단어에서 알 수 있듯이 실제로 존재하지 않으나 존재하는 것처럼 만들어서 실행하여 보여준다는 의미이다. 클라우드 컴퓨팅에서의 가상화는 CPU, 메모리, 디스크, OS 등을 구매하지 않아 실제로는 즉, 물리적으로 내 손안에 존재하지 않지만 마치 그러한 컴퓨팅 자원이 존재하는 것처럼 가상적으로 소프트웨어를 이용하여 만들어 주는 기술이다. 마법처럼 완전히 새로운 컴퓨팅 자원을 창조해 내는 것이 아니라 이미 존재하고 있는 컴퓨터의 자원을 일부 빌려오는 형식으로 구현한다. 클라우드 컴퓨팅에서의 분산처리 기술은 여러 개의 컴퓨터를 하나로 만들어 주는 역할을 한다. 사용자 측면에서는 클라우드 컴퓨팅을 구성하는 다수 서버에 대한 정보 없이 외부에 드러나 있는 하나의 창구를 통해서

<표 2> 클라우드 컴퓨팅의 특징

구분	특징
접속 용이성	<ul style="list-style-type: none"> • 시간과 장소에 상관없이 인터넷을 통해 이용 가능 • 클라우드에 대한 표준화된 접속을 통해 다양한 기기로 이용 가능
유연성	<ul style="list-style-type: none"> • 갑작스러운 이용량 증가나 이용자 수 변화에 신속하고 유연하게 대응 가능
주문형 셀프서비스	<ul style="list-style-type: none"> • 이용자는 서비스 제공자와 직접적인 상호작용을 거치지 않고, 자율적으로 자신이 원하는 클라우드 서비스를 이용 가능
사용량 기반 과금제	<ul style="list-style-type: none"> • 이용자는 서비스 사용량에 대해서만 비용을 지불 • 전기, 수도물처럼 개인의 사용량에 따라 과금하는 방식
가상화와 분산처리	<ul style="list-style-type: none"> • 하나의 서버에서 다양한 컴퓨팅 자원을 나누어서 쓰는 가상화 기술 • 여러 대의 서버를 하나로 묶어 운영하고, 사용자의 컴퓨팅 자원 요구를 여러 서버에 분산처리함으로써 시스템 과부하를 최소화하는 분산처리 기술

* 출처: (강맹수, 2019)

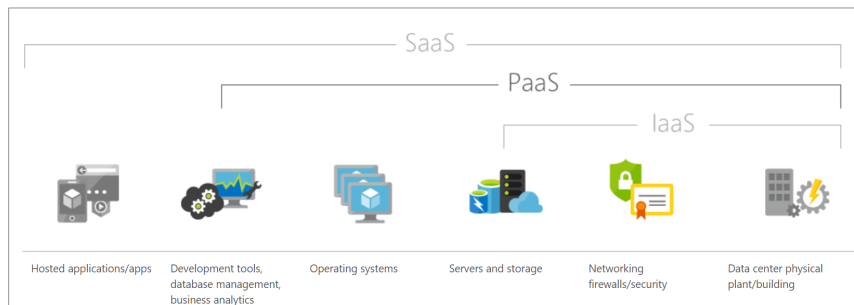
서비스를 제공받을 수 있으며, 관리자 측면에서도 각 서버를 별도 관리하지 않고 하나로 묶어서 감시하고 관할 수 있다. 이를 위해 분산처리 기술은 다수 서버의 컴퓨팅 자원(CPU, 메모리, 디스크, 네트워크 등)을 하나의 Pool로 구성하고, 사용자 요청이 오면 알맞게 요청이 분산되어 수행되도록 로드 밸런싱(Load Balancing) 하여 제공한다.

클라우드 컴퓨팅 서비스는 유형에 따라 IaaS (Infrastructure As A Service), PaaS(Platform As A Service), SaaS(Software As A Service)로 구분된다(〈그림 4〉 참고).

IaaS가 하드웨어라면, PaaS와 SaaS는 소프트웨어를 제공하는 서비스이다. 먼저, IaaS는 CPU, 메모리, 디스크 등 실제로 물리적인 하드웨어를 '구매하지 않고도' 마치 하드웨어를 가지고 있는 것처럼 서비스를 제공한다. 실제 하드웨어는 내 손안에 있지 않고 클라우드에 있으며, 원격지에서 인터넷을 통해 제어하여 동작하게 할 수 있다. 이를 통해서, 일반적으로 서버나 저장소, 네트워크를 필요에 따라 이용할 수 있게 서비스를 제공받을 수 있으며 서버와 저장소, 백업 인프라 구축 시 이용할 수 있다. PaaS는

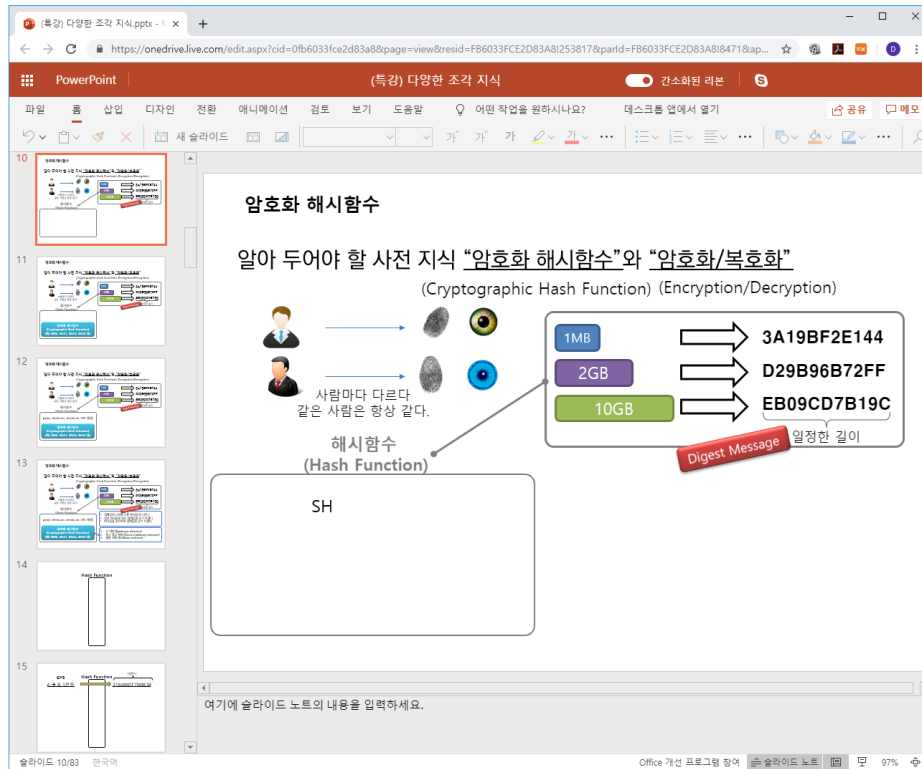
운영체제, 개발 환경을 위한 시스템 소프트웨어, 데이터베이스와 같은 미들웨어 등에 해당하는 플랫폼 수준의 소프트웨어를 제공하는 서비스이다. 이러한 플랫폼을 실제로 본인의 서버나 PC에 '설치하지 않고' 클라우드 컴퓨팅을 통해 설치된 플랫폼 서비스를 제공받을 수 있다. 사용자는 인터넷으로 클라우드 컴퓨팅 서비스에 접속하여 플랫폼을 원격으로 제어하여 실행하고 결과를 볼 수 있다. SaaS는 네트워크를 통해 원하는 소프트웨어를 본인의 서버나 PC에 '설치하지 않고' 클라우드 컴퓨팅을 통해 서비스를 받을 수 있다. Google과 MS에서는 각각 Google Document(Spreadsheet, Presentation 등)와 MS Office(Powerpoint, Excel 등)을 설치 없이 인터넷 브라우저에서 직접 작성하여 저장할 수 있는 서비스를 제공하고 있다(〈그림 5〉 참고).

클라우드 컴퓨팅은 사용자가 실제 실물의 컴퓨터가 없는데도 불구하고 여러 개의 컴퓨팅 환경을 사용할 수 있게 해준다. 이는 다양한 하드웨어를 추상화하는 가상화 기술과 가상화된 하드웨어 위해서 운영체제를 실행해 주는 애플레이터 기술 덕분이다. 이러한 장점으로 기록



〈그림 4〉 제공 서비스 유형에 따른 클라우드 컴퓨팅 서비스

* 출처: Microsoft Azure 홈페이지



〈그림 5〉 Chrome에서 실행시킨 MS Office 365 - 파워포인트

관리 분야뿐 아니라 대부분의 정보화 시스템을 운영하는 기관에서는 클라우드 컴퓨팅 환경으로 시스템들 이전하고 있다. 그러나 현재 가상화 기술이 지금까지 나와 있는 모든 하드웨어를 추상화하지 못한다. 가상화 기술이 지원할 수 있는 CPU 아키텍처 구조들이 제한적이기 때문이다. 예를 들어, 본 연구에서 테스트베드를 구축할 때 사용되는 가상화 기술인 KVM은 Intel 및 AMD에서 개발한 x86 CPU 프로세서를 지원하지만 Sun SPARC, DEC Alpha, IBM S/390 등의 마이크로프로세서를 지원하지 않는다. 그러므로 클라우드 컴퓨팅 환경의 장점을 전자기록의 에뮬레이션 장기보존 방안에 최대한 활용하면서도 한계점을 파악하여 에뮬레이션 전

략으로 해결할 수 없는 부분을 정확히 도출하여야 한다.

2.3 가상화(Virtualization)

가상화 기술은 실제로 존재하지 않으나 존재하는 것처럼 가상으로 만들어 주는 기술이다. 컴퓨터 분야에서는 실제로 컴퓨터 자원(CPU, 메모리, 디스크 등)을 실제로 구매하지 않아도 자신의 컴퓨터 상에서 실제로 존재하는 것처럼 만들어 준다. 이들은 소프트웨어 기술로 구현되는데, 각자의 컴퓨터에 가상화를 지원하는 소프트웨어(VMware, Virtualbox 등)를 설치하고 Window OS 설치 파일 또는 리눅스 OS

설치 파일을 가상화 소프트웨어와 연결하면 <그림 3>처럼 가상화 소프트웨어에 설치되어 컴퓨터 안에 또 하나의 컴퓨터를 가지게 된다. 여기서 물리적으로 존재하는 컴퓨터(OS)를 호스트 컴퓨터(호스트 OS)라고 하고, 호스트 컴퓨터에 가상화 소프트웨어 도움을 받아 설치된 컴퓨터(OS)를 게스트 컴퓨터(게스트 OS)라고 한다. 예를 들어, 구 버전인 Windows XP만 실행되는 응용 프로그램을 실행시키고 싶은데 현재 본인이 가지고 있는 컴퓨터(호스트 컴퓨터)의 OS는 Windows 10인 경우에 가상화를 지원하는 소프트웨어를 설치하고 Windows XP 설치 파일을 해당 가상화 소프트웨어와 연결하여 설치한 후 가상화 소프트웨어의 도움을 받아 Windows XP를 시작하고 실행을 원하는 응용 프로그램을 실행시킬 수 있다. 새로운 컴퓨터를 갖게 되는 모습을 보이지만 가상화 소프트웨어가 실행되는 호스트 컴퓨터의 컴퓨팅 자원을 나눠 쓰는 것이기 때문에 부모 컴퓨터 이상의 컴퓨팅 자원을 사용할 수 없다.

이처럼 가상화 기술은 호스트 컴퓨터의 컴퓨팅 자원을 사용하는데, 호스트 컴퓨터의 컴퓨팅 자원을 사용하기 위해서는 호스트 컴퓨터의 컴퓨팅 자원을 논리화 및 추상화를 해야 한다. 그래서 가상화를 추상화(Abstraction)로 일컫기

도 한다. 먼저, 논리화는 하나의 컴퓨팅 자원 작은 논리 단위로 쪼개져 있게 만드는 것이고, 추상화는 컴퓨팅 자원과 연관된 복잡하고 물리적인 다른 요소들을 안 보이게 하여 단순화시키는 기술을 얘기한다. 예를 들어, 호스트컴퓨터에 삼성 전자에서 구매한 256GB(DDR4, 2.666MHz) 메모리 2개가 장착되어 있다고 가정하자. 이때, 논리화는 이 메모리를 1GB짜리 512개(2*256개)가 있는 것으로 만들어 주는 기술이고, 추상화는 이 메모리의 제조업체 무엇이고 어떤 기술로 제조됐는지 상관없이 '512GB 메모리(1GB 단위)'로 보여 줄 수 있게 하는 기술이다.

가상화 기술의 논리화 및 추상화를 수행하는 핵심 기능을 하이퍼바이저(Hypervisor)라고 한다. 논리화 및 추상화를 수행하고, 이를 통해 호스트 컴퓨터에서 다수의 게스트 OS를 동시에 실행할 수 있다. 하이퍼바이저는 <표 3>처럼 Type 1과 Type 2로 구분된다.

가상화 기술은 클라우드 컴퓨팅 분야와는 다른 연구 진영에서 연구되어 온 것으로 하나의 호스트 컴퓨터에서 여러 개의 게스트 OS를 실행하기 위해 개발되었다. 여러 개의 호스트 컴퓨터를 하나로 묶고, 여러 개의 게스트 OS를 실행하는 클라우드 컴퓨팅 서비스는 가상화 기술이 필요하고 호스트 컴퓨터와 OS에 최적화

<표 3> 하이퍼바이저 종류

구분	특징	종류
Type I	• OS가 프로그램을 제어하듯이 하이퍼바이저가 해당 하드웨어에서 직접 실행되며 게스트 OS는 하드웨어 위에서 2번째 수준으로 실행됨	MS Hyper-V VMWare ESX/ESXi Redhat KVM
Type II	• 하이퍼바이저는 일반 프로그램과 같이 호스트 운영 체제에서 실행되며 하이퍼바이저 내부에서 동작되는 게스트 OS는 하드웨어에서 3번째 수준으로 실행	MS Virtual PC VMWare WorkStation

된 가상화 기술을 활용하는 구조이다.

이중 KVM(Kernel-based Virtual Machine)은 리눅스 진영에서 구축한 오픈소스 기반 가상화 기술이다. 일단 오픈소스이기 때문에 저작권 문제와 추후 재현하지 못하는 것에 대한 위험이 적다. 또한, Type 1 이기 때문에 직접적으로 하드웨어와 통신하기 때문에 Type 2 보다 실행이 빠른 편이다. 이러한 이유로 본 연구에서는 클라우드 컴퓨팅 환경을 구축하기 위한 가상화 기술로 KVM를 채택한다. 그리고 KVM과 가장 최적화되어 있는 QEMU를 채택하였다. 단, 2007년 이후에 출시된 리눅스 버전에서 하드웨어 가상 머신(HVM: Hardware Virtual Machine) 기능을 지원하는 x86 CPU 프로세서에 설치하여야 한다는 조건이 있다. 그래서 유닉스 계열의 OS가 주로 운영되었던 Sun SPARC, DEC Alpha, IBM S/390 등의 CPU 프로세서를 지원하지 않는다. 예전에 Sun Microsystems에서는 Solaris라는 UNIX 계열 OS를 개발했는데, 이는 x86 CPU 프로세서까지 지원하지 때문에 KVM에서 가상화할 수 있는 OS이다.

2.4 오픈스택(OpenStack)

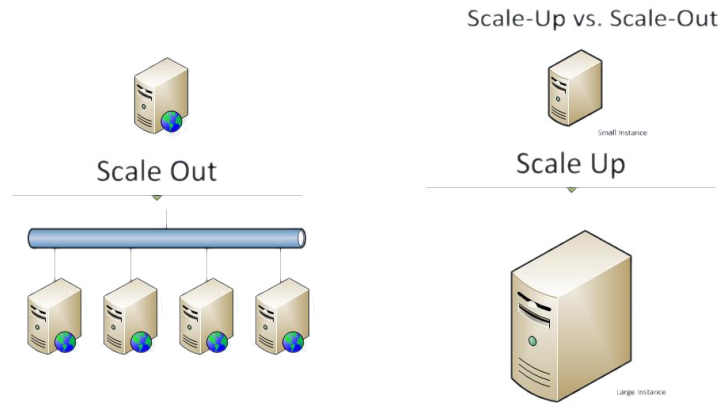
클라우드 컴퓨팅 기술은 가상화와 분산처리가 핵심기능이다. 또한, 클라우드 컴퓨팅 서비스를 제공하기 위해서는 이 2개 핵심기능 이외에 다양한 기능들(사용자 인터페이스, 서비스 프로비저닝, 자원 유틸리티, 보안 및 프라이버시 등)을 포함하여야 한다. 클라우드 컴퓨팅 전문 기관과 기업들은 오픈소스 프로젝트 팀을 구성하여 가상화 기능은 별도로 개발하지 않고 앞서 2.3에서 소개되었듯이 연구되어 온 가상화

기술들 중에서 가장 적합한 기술을 선택할 수 있는 유연한 구조로 설계하고, 분산처리를 포함한 다양한 기능들은 자신만의 방법으로 개발하여 클라우드 컴퓨팅 서비스를 제공하기 위한 플랫폼을 구축하여 왔다.

클라우드 컴퓨팅에서의 분산처리 기술은 여러 개의 컴퓨터를 하나로 만들어 준다. 기존 단일 서버 중심의 컴퓨팅 구조에서는 사용자의 수가 증가하여 서버가 서비스를 제공해야 하는 규모가 커지면 서버의 용량을 늘리기 위해서 기존 서버를 폐기하고 용량이 큰 새로운 서버를 도입하는 Scale-up 방식이었다. 그러나 클라우드 컴퓨팅에서는 분산처리 기능 덕분에 기존의 서버 Pool에 새로운 서버를 추가하는 Scale-out 방식으로 동작한다(<그림 6> 참고).

컴퓨팅, 저장장치, 네트워크 자원 등으로 구성된 대규모 자원 Pool을 제어하고 관리하고, 웹 기반의 사용자 인터페이스를 통해 관리자는 손쉽게 자원을 제어하고 서비스를 관리할 수 있으며, 서비스 이용자는 클라우드 컴퓨팅 서비스를 동적으로 제공받고 관리할 수 있다. 이러한 소프트웨어의 역할이 마치 하나의 컴퓨터에서 자원을 할당하고 스케줄링하는 운영체제와 유사하게 여러 개의 서버에 대해서 하기 때문에 '클라우드 운영체제'라고 불리운다.

대표적인 소프트웨어로는 Cloud.com에서 개발한 CloudStack, UC Santa BarBara에서 개발한 Eucalyptus, 현재는 OpenNebula Systems에서 유지관리하고 있는 OpenNebula, NASA와 Rackspace가 2010년 7월에 시작한 OpenStack(현재는 OpenStack Foundation에서 운영)이 있다. CloudStack을 제외하고 대부분 오픈소스 프로젝트로 운영되고 있으며, 그 중에서도



〈그림 6〉 Scale-up과 Scale-out 개념

* 출처: IDCHOWTO, <https://idchowto.com>

오픈스택은 다른 공개 소프트웨어처럼 기업이 비교적 자유롭게 쓸 수 있다. OpenStack에는 컴퓨터, 저장장치, 네트워크 장비 제조사와 소프트웨어 솔루션 업체 그리고 정보통신 서비스 사업자를 아우르는 수많은 정보통신 글로벌 선도 기업들이 회원사로 등록되어 있어 모든 클라우드 컴퓨팅 오픈소스 프로젝트 중에서 가장 많은 기업이 프로젝트에 참여하고 있다. 이와 더불어 사용자뿐만 아니라 오픈소스에 기여하는 개발자 층도 전 세계적으로 지속적으로 확대되고 있어, 오류 발생 시 빠르고 능동적인 대처가 이루어지고 있다.

2010년 Austin 이라는 이름의 배포판을 시작으로 알파벳 순서에 따라 공식 발표되던 OpenStack은 Diablo 배포판 이후 6개월마다 배포판을 발표하고 있다. OpenStack의 배포과정은 계획, 개발, 사전배포, 공식배포의 4단계로 이루어진다. 매번 새로운 배포판이 공식 발표된 이후에는 차기 배포판에 대한 개발 논의가 Design Summit을 포함해 약 4주간 지속되며, 이 기간 동안 수집된 각종 기능 추가 또는

성능 개선 현안들에 대한 검토를 거쳐 우선순위를 부여하고 전체적인 개발일정 계획을 수립한다. 이후에는 수립된 일정 계획에 맞춰서 선별된 기능 추가 또는 성능 개선 현안에 대한 개발이 동시다발적으로 진행된다(박종근, 최강일, 이상민, 이정희, 이범철, 2013).

OpenStack은 주요 기능이 몇 개의 세부 프로젝트로 나뉘어 개발되고 있다. 초기에는 컴퓨팅 서비스인 Nova, 오브젝트 저장장치 서비스인 Swift, 템플릿 이미지 관리 서비스인 Glance의 3개의 프로젝트로 구성되었다. 계속해서 버전이 업데이트되고 기능이 확장되면서 새로운 프로젝트가 지속적으로 추가 발표되어 점차 그 영역을 확대되고 있다. 예를 들어, 네트워킹 서비스와 볼륨 저장장치 서비스와 같이 Nova에서 함께 제공되던 서비스가 별도의 프로젝트로 분리되거나 인증 서비스 또는 사용자 인터페이스와 같이 독립된 프로젝트가 새롭게 추가되기도 한다.

OpenStack은 라이선스 비용없이 무료로 사용할 수 있는 오픈 소스이며, 많은 회사와 개인

개발자들이 참여하여 프로젝트에 기여하고 있기 때문에 지속적인 업데이트가 가능하다. 그렇기 때문에 OpenStack은 현재 클라우드 컴퓨팅 플랫폼 중에서는 가장 주목 받고 있다. 본 연구팀은 현재 대규모 시스템 시장에서도 꾸준히 시장 점유율을 높이고 있는 Linux 처럼 기본적으로 무료이고, 지속적인 발전으로 완성도가 높아질 가능성이 많기 때문에 OpenStack가 업계 표준이 될 것으로 예상하여, 이 본 연구에서는 클라우드 컴퓨팅 플랫폼으로 OpenStack을 활용한다.

2.5 P2V(Physical-to-Virtual) 변환기

P2V(Physical-to-Virtual) 변환기는 실제로 존재하는 물리(Physical) 서버에서 클라우드 컴퓨팅 환경의 가상(Virtual) 서버로 시스템을 원래의 모습 그대로 이전하는 도구를 의미한다. <그림 7>은 P2V 변환기를 활용하여 물리서버 (<그림 7>에서 Source Server)를 가상서버 (<그림 7>에서 Target Server)로 이전하는 과

정을 보여 준다. P2V 변환기는 기존의 레거시(Legacy) 시스템을 클라우드 컴퓨팅 환경으로 시스템 전체를 이전할 때 주로 사용되고 있다. 먼저 물리서버에 P2V 변환기 소프트웨어를 설치하고, 클라우드 컴퓨팅 환경에 동일한 운영 체제를 가진 가상서버를 생성하고, 네트워크를 통해 물리서버에서 가상서버로 시스템 데이터를 이전하는 과정을 거친다.

<표 4>는 현재 많이 사용되고 있는 P2V 변환기 제품들을 비교하였다. 크게 무료와 유료로 구분된다. 무료 제품은 경제적인 비용이 거의 없다는 장점이 있지만 지원 OS, 지원 클라우드/하이퍼바이저 플랫폼이 제한적이며, 오류 상황이 발생하더라도 해결책을 쉽게 얻을 수 없다. 유료 제품은 경제적인 비용은 들지만 시스템 이전이 발생하는 오류에 대해 기술지원을 받을 수 있기 때문에 안정적으로 시스템 이전을 수행할 수 있다는 장점이 있다. 그래서 본 연구에서는 무료 제품 중에서는 virt-p2v를, 유료 제품 중에서는 국내 기업의 zConverter를 활용하여 실험을 진행한다.



<그림 7> P2V 변환기 시스템 이전 과정

* 출처: PlanetVM, <https://planetvm.net/>

〈표 4〉 주요 P2V 변환기 제품 비교

제품명	비용	지원 OS 계열	지원 클라우드/하이퍼바이저 플랫폼
vCenter Converter (VMWare 사)	무료	윈도우/리눅스	•VMware
Disk2vhd (Microsoft 사)	무료	윈도우	•VHD(MS Virtual PC or Hyper-V)
PlateSpin Migrate (Micro Focus 사)	유료	윈도우/리눅스	•VMware vCenter, ESXi, Hyper-V, KVM, Azure, VMware vCloud Director
V2V Converter (StarWind 사)	무료	윈도우	•KVM, VMware, Hyper-V
virt-p2v (Redhat 사)	무료	리눅스	•KVM, Openstack, oVirt, RHV
zConverter (ISA Technologies 사)	유료	윈도우/리눅스	•KVM, Xen, VMware, Hyper-V, Openstack, Azure, AWS, Cloudstack

3. 클라우드 컴퓨팅 기반 에뮬레이션 전략을 활용한 장기보존 방안

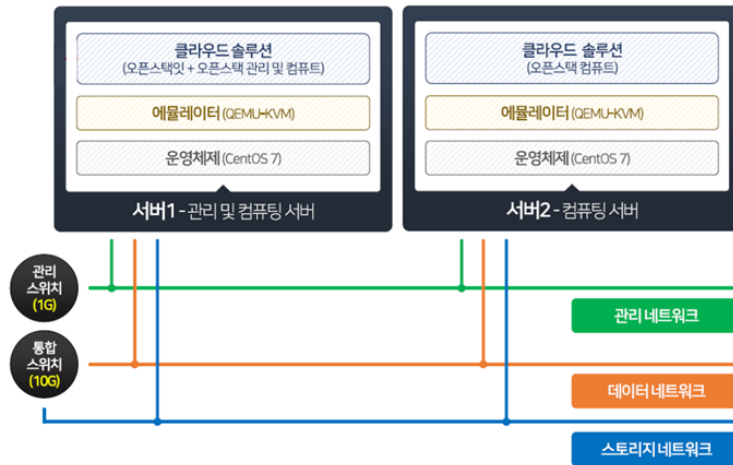
본 장에서는 실제 에뮬레이션 테스트베드를 구축하고 이를 기반으로 실험한 결과를 통해 클라우드 컴퓨팅 기반 에뮬레이션 전략을 활용한 전자기록 장기보존 방안을 제시하고자 한다.

3.1 에뮬레이션 테스트베드 구축

에뮬레이션 테스트베드 구축을 위해서 클라우드 컴퓨팅 환경을 〈그림 8〉과 같이 구성하였다. 테스트베드에는 2개의 서버(〈표 5〉 참고)와 2개의 네트워크 스위치로 이루어져 있다. 기본적으로 클라우드 컴퓨팅을 위해서는 최소 2개의 서버(관리 서버, 컴퓨팅 서버)와 2개의 스위치(관리 스위치, 통합 스위치)가 필요하다. 각 서버의 운영 체제는 CenOS 7(리눅스)를 설치하고, 하이퍼바이저는 KVM를, KVM 상에서 게스트 OS를 실행하는 역할을 담당하는 에뮬레이터는 QEMU를 설치하였다. 그리고 서

버들을 하나의 Pool로 구성하기 위해 KVM과 QEMU와 연동하여 이들을 관리하고 자원을 할당하는 클라우드 운영체제는 OpenStack Queens 버전을 사용하였다. 이들을 관리자 또는 사용자가 관리하는 인터페이스는 〈그림 9〉에서 처럼 (㉠)이 노그리드에서 개발한 오픈스택잇(OpenStackit) 솔루션을 적용하였다.

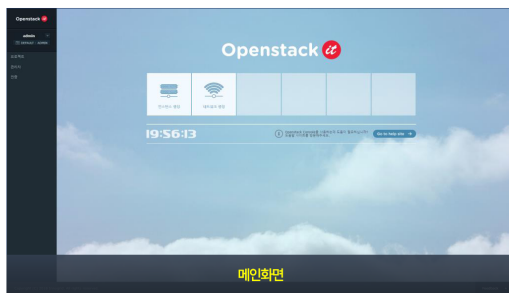
에뮬레이션 테스트를 활용하여 네 가지 실험을 진행하였다. 첫 번째는 클라우드 컴퓨팅에 Windows XP를 설치하고 한컴오피스 97을 설치하고 한컴오피스 97버전의 파일(office97.hwp)을 확인하는 실험이다. 두 번째는 클라우드 컴퓨팅에 MS DOS를 설치하고 보석글 VersionG 1.01을 설치하고 보석글 파일(jewelry.twp)을 확인하는 실험이다. 첫 번째와 두 번째 실험은 전자기록을 에뮬레이션 전략으로 장기보존 주 전략 채택가능성에 대해 검토하기 위한 실험이다. 세 번째 실험은 특정 PC를 백업 도구를 사용하여 템플릿 이미지 파일을 만든 다음, 그 템플릿 이미지 파일을 클라우드 컴퓨팅으로 이관하여 설치하고, 특정 PC의 모습이 클라우드 컴



〈그림 8〉 에뮬레이션 테스트베드 클라우드 컴퓨팅 환경 구성

〈표 5〉 클라우드 컴퓨팅 구성 서버 사양

구분	CPU	메모리	볼륨 스토리지	이미지 스토리지	가상서버 스토리지
서버1(관리, 컴퓨팅)	24 Core	256GB	1TB	1TB	3TB(공유 스토리지)
서버2(컴퓨팅)	24 Core	256GB	-	-	
계	48 Core	512GB	1TB	1TB	3TB



〈그림 9〉 오픈스택의 메인 화면 및 컴퓨팅 서버 조회 화면

퓨터에 재현되는지를 검토한다. 네 번째 실험은 2개의 서버로 이루어진 시스템을 백업 도구를 사용하여 템플릿 이미지 파일을 만든 다음, 그 템플릿 이미지 파일을 클라우드 컴퓨팅으로 이관하여 설치하고, 2개의 서버로 이루어진 시스템이 클라우드 컴퓨터에 제대로 재현되는지

를 검토한다. 세 번째와 네 번째 실험은 마이그레이션으로는 장기보존하기 힘든 전자기록의 기능을 전자기록을 에뮬레이션 전략 방식을 통해서 재현할 수 있는지 여부를 검토하는 실험이다. 〈표 6〉은 네 가지 실험에 대한 개요를 보여 준다.

〈표 6〉 클라우드 컴퓨팅 기반 에뮬레이션 실험 개요

구분	사양	내용	실험방법		
MS DOS 및 보석글 에뮬레이션 실험 (3.2)	OS	MS DOS	클라우드 컴퓨팅 환경에 OS, SW설치 및 실행		
	SW	보석글G 1.01			
	전자기록	jewelry.twp			
XP 및 한컴오피스 에뮬레이션 실험 (3.3)	OS	Windows XP	클라우드 컴퓨팅 환경에 OS, SW설치 및 실행		
	SW	한컴오피스 97			
	전자기록	office97.hwp			
단일 서버 시스템 에뮬레이션 실험 (3.4)	OS	Ubuntu 14.04	zConverter P2V변환기	물리서버에서 클라우 드 컴퓨팅 환경의 가상 서버로 시스템 이전 및 실행	
	SW	MySQL 5.5			
다수 서버 시스템 에뮬레이션 실험 (3.5)	1	OS	Windows Server 2012 Standard R2		zConverter P2V변환기
		SW	WebtoBv4.1, JEUS v6.0		
	2	OS	CentOS 7.6		virt-p2v P2V 변환기
		SW	Oracle 10g		

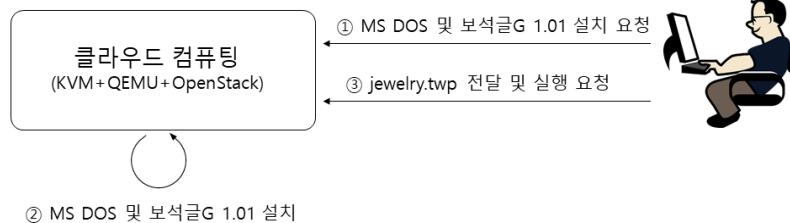
3.2 MS DOS 및 보석글 에뮬레이션 실험

〈그림 10〉과 같이 클라우드 컴퓨팅 환경에 MS DOS를 설치 요청 및 설치 완료하고 전자 문서(jewelry.twp)를 실행하여 jewelry.twp 파일을 보석글G version 1.01을 통해 확인한다.

클라우드 컴퓨팅 환경에 MS DOS 설치 요청 및 설치 완료를 하고 전자문서(jewelry.twp)를 실행하여 jewelry.twp가 잘 실행이 되는지 확인한다. 〈그림 11〉은 MS DOS 설치를 요청하고 클라우드 환경에서 MS DOS가 설치된 후의 화면이다. 〈그림 12〉에서는 MS DOS 컴퓨팅

환경과(왼쪽)과 보석글 파일(오른쪽)을 각각 실행한 모습을 보여 준다.

MS DOS 및 보석글 에뮬레이션 실험을 통해서 구전자문서시스템에서 보석글로 생산된 공문서를 재현하여 확인할 수 있는 가능성을 볼 수 있다. 또한, MS DOS 이외에도 Windows 3.X도 에뮬레이션이 가능하다. 만약, 실시간으로 운영체제와 응용 소프트웨어 설치 요청이 처리가 가능하도록 서비스가 구축된다면 구전자문서시스템에 남아 있지만 확인이 힘든 다양한 전자문서들까지도 재현을 통해 확인할 수 있을 것으로 기대된다.



〈그림 10〉 전자기록 에뮬레이션 실험 과정

118.130.73.35/project/instances/

프로젝트 > 컴퓨터 > 인스턴스

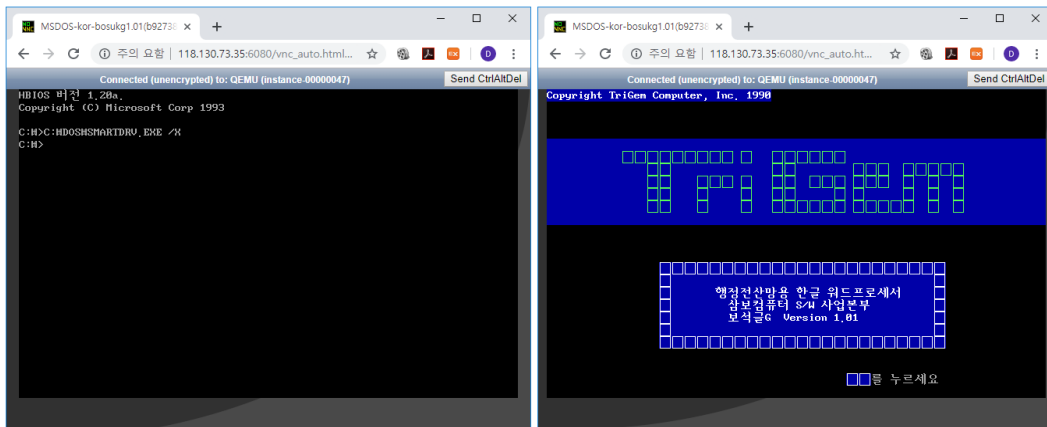
인스턴스

4 항목 표시 업로드한 이미지로 인스턴스 생성 인스턴스 ID = ▾

<input type="checkbox"/>	인스턴스 이름	이미지 이름	IP 주소	사양	키 페어	상태	가용 구역	작업
<input type="checkbox"/>	msdos-kor-qcow	msdos-kor-qcow	10.10.10.10 유동 IP: 192.168.150.36	m1.large	-	Active	nov1	None
<input type="checkbox"/>	**win3	windows_server_...	10.10.10.19 유동 IP: 192.168.150.36	m1.xlarge	-	Active	nov1	None
<input type="checkbox"/>	**p2v_cirros	p2v_linux	10.10.10.19 유동 IP: 192.168.150.38	p2v	-	Active	nov1	None
<input type="checkbox"/>	window_2	windows_server_...	10.10.10.7 유동 IP: 192.168.150.37	m1.large	-	Active	nov1	None

<< < | > >>

<그림 11> 클라우드 컴퓨팅 환경에 MS DOS가 설치된 모습



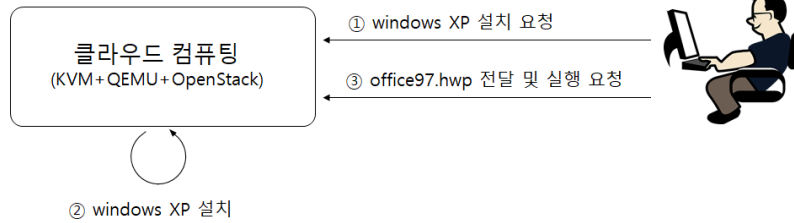
<그림 12> 클라우드 컴퓨팅 환경에 설치된 MS DOS와 보석글G 1.01 실행한 화면

3.3 XP 및 한컴오피스 에뮬레이션 실험

여기서는 <그림 13>과 같이 클라우드 컴퓨팅 환경에 Windows XP 설치 요청 및 설치 완료하고 전자문서(office97.hwp)를 실행하여 office97.hwp가 잘 실행이 되는지 확인한다.

클라우드 컴퓨팅 환경에 Windows XP를 설치 요청 및 설치 완료를 하고 전자문서(office97.hwp)

를 실행하여 office97.hwp가 잘 실행이 되는지 확인한다. <그림 14>는 Windows XP 설치를 요청하고 클라우드 컴퓨팅 환경에서 Windows XP가 설치된 후의 화면이다. <그림 15>에서는 사용자 컴퓨터와(왼쪽)과 클라우드 컴퓨팅 환경(오른쪽)에서 각각 잘 실행되고 있는 모습을 보여 준다.



<그림 13> 전자기록 애플레이션 실험 과정

프로젝트 > 컴퓨터 > 인스턴스

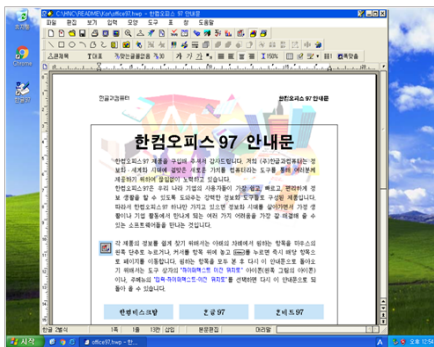
인스턴스

6 항목 표시

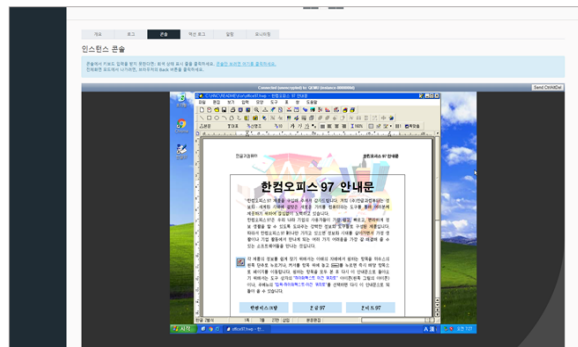
인스턴스 이름	이미터 이름	IP 주소	사양	키 페어	상태	가용 구역	작업	현재 상태	생성된 이후 시간	해상도 정보	작업	
System Test 2	CentOS	10.10.10.12	t1.medium	mykey	Active	af	nova	None	Running	56분	-	스냅샷 생성
System Test 1	WindowsXP-h...	10.10.10.13	t1.medium	-	Active	af	nova	None	Running	1시간, 36분	-	스냅샷 생성
new-instance...	test-snap	10.10.10.10	t1.small	-	Active	af	nova	None	Running	4시간, 48분	-	스냅샷 생성
new-instance	test-snap	10.10.10.3 192.168.150.36	t1.small	mykey	Active	af	nova	None	Running	5시간, 13분	-	스냅샷 생성
Test 2	windows_serv...	10.10.10.15	t1.medium	mykey	Active	af	nova	None	Running	23시간, 31분	-	스냅샷 생성
Test 1	CentOS	10.10.10.17 192.168.150.37	t1.small	mykey	Active	af	nova	None	Running	1일	-	스냅샷 생성

<그림 14> 클라우드 컴퓨팅 환경에 Windows XP가 설치된 모습(OpenStackit 화면)

windowsXP가 설치된 PC에서 실행 중인 한글 97 문서 화면



클라우드 컴퓨팅 환경에서 재현된 windows XP 한글 97 문서 화면



<그림 15> 원본PC와 클라우드 컴퓨팅 환경 가상서버에 설치된 Windows XP와 실행된 한글 문서(office97.hwp)

XP 및 한컴오피스 애플레이션 실험을 통해서도 3.2에서의 실험과 같이 오랫동안 확인이 힘들었던 다양한 전자문서들을 애플레이션 방식으로 확인할 수 있음을 기대할 수 있다.

3.4 단일 서버 시스템 애플레이션 실험

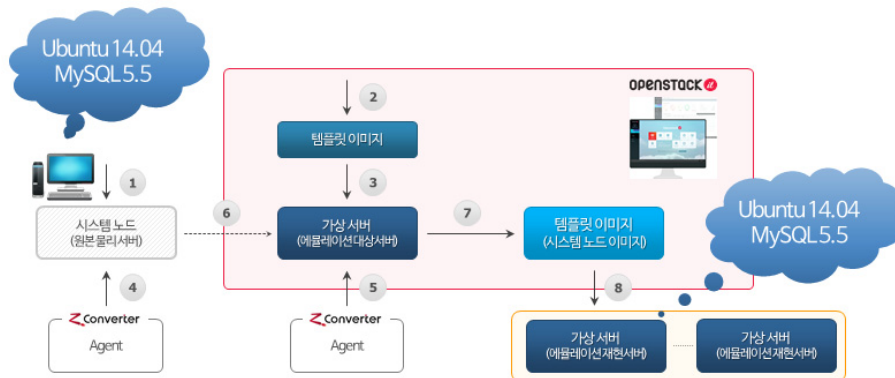
세 번째 실험은 사용자 컴퓨터에서 실제 운영하고 있는 서버를 클라우드 컴퓨팅 환경의

가상서버로 이관하는 실험이다. 실험은 <그림 16>과 같은 절차를 통해 이루어진다. ① 에뮬레이션 대상 원본물리서버(MySQL 5.5가 설치된 Ubuntu 14.04)를 준비한다. ② 원본물리서버와 같은 운영체제(Ubuntu 14.04)로 제작된 템플릿 이미지(설치파일)을 준비한다. ③ 제작된 템플릿 이미지로 클라우드 컴퓨팅 환경에 가상서버를 생성한다. ④ 원본물리서버에 zConverter Agent를 설치한다. ⑤ 가상서버에도 zConverter Agent를 설치한다. ⑥ zConverter로 원본물리서버에서 가상서버로 데이터를 이관한다. ⑦ 가상서버에 정상적으로 이관되었는지 확인하고 백업 및 스냅샷 기능으로 템플릿 이미지(설치파일)을 생성한다. 여기까지가 사용자 컴퓨터의 원본물리서버를 클라우드 컴퓨팅 환경으로 옮기는 과정이다. ⑧ 인터넷을 통해 사용자가 MySQL 5.5가 설치되어 있는 리눅스 확인 요청하면 템플릿 이미지(설치파일)에서 가상서버를 생성한다. 사용자는 원격으로 클라우드 컴퓨팅 환경에 있는 가상서버에 접속하여 MySQL 5.5가 실행되고 있는지 확인할 수 있다.

이 과정은 원본물리서버의 특정 파일이나 응

용 프로그램을 옮기는 것이 아니라 시스템 자체를 옮기는 과정이다. 처음 컴퓨터에 OS를 설치할 때 CD 또는 USB에 파일로부터 설치한다. 즉, 시스템도 하나의 파일로 표현될 수 있다. 템플릿 이미지가 여기에 해당한다. 이러한 개념에 응용하여 시스템 전체를 다시 하나의 템플릿 이미지 즉, 하나의 파일로 만들 수 있으며 시스템 전체를 이관하는데 활용할 수 있는 것이다.

본 연구에서 템플릿 이미지 생성 이후의 과정은 (주)이노그리드에서 개발한 OpenStackit에서 동작시키고 확인할 수 있다. 웹브라우저로 클라우드 컴퓨팅에 접속하면 <그림 17>의 OpenStackit 화면을 볼 수 있고, 프로젝트 → 컴퓨터 → 인스턴스 메뉴에서 클라우드 컴퓨팅 환경에는 원본물리서버와 동일한 가상서버가 설치되었음을 확인할 수 있으며, 클라우드 컴퓨팅 환경의 생성된 가상서버(리눅스 Ubuntu 14.04)을 부팅하여 MySQL 5.5에 접속한 모습과 이전 사용자 컴퓨터에서 부팅되어 MySQL 5.5가 동일하게 실행되고 있는 모습을 <그림 18>에서 확인할 수 있다. 가상서버 접근하는 방법은



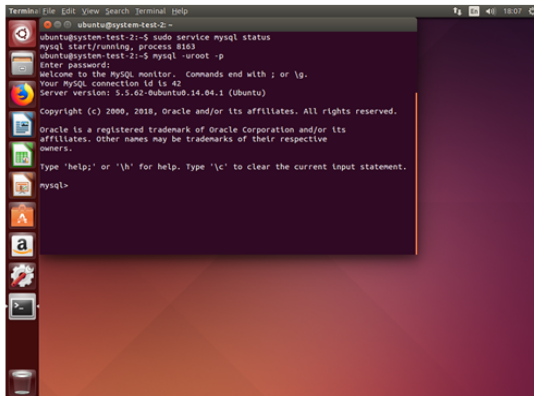
<그림 16> zConverter 기반 전자기록 시스템 에뮬레이션 실험 절차

인스턴스 이름	이미지 이름	IP 주소	사양	키 패어	상태	가용 구역	작업	현황 상태
System Test 2	ubuntu_14.04...	10.10.10.18 유용 IP: 192.168.150.37	t1.medium	mykey	Active	nova	None	Running
System Test 1	Windows09+H...	10.10.10.13	t1.medium	-	Active	nova	None	Running
new-instance...	test-snap	10.10.10.10	t1.small	-	Active	nova	None	Running
new-instance...	test-snap	10.10.10.3 유용 IP: 192.168.150.36	t1.small	mykey	Active	nova	None	Running
Test 2	windows_serv...	10.10.10.15	t1.medium	mykey	Active	nova	None	Running
Test 1	CentOS	10.10.10.17	t1.small	mykey	Active	nova	None	Running

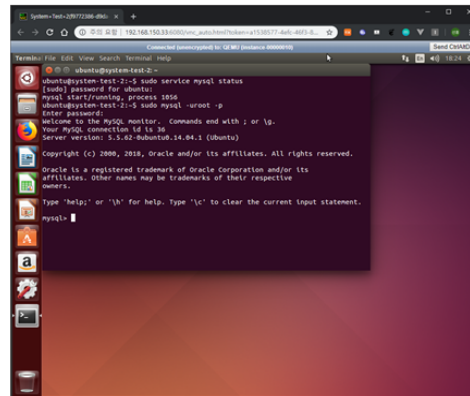
ubuntu_14.04...	10.10.10.18 유용 IP: 192.168.150.37	t1.medium	mykey	Active	nova	None	Running
-----------------	---	-----------	-------	--------	------	------	---------

〈그림 17〉 클라우드 컴퓨팅 환경에 템플릿 이미지 파일로 설치된 리눅스

사용자 컴퓨터(Linux)에서 실행중인 MySQL 접속 화면



클라우드 컴퓨팅 환경에서 재현된 Linux와 MySQL 접속 화면



〈그림 18〉 사용자 컴퓨터와 클라우드 컴퓨팅 환경에 설치된 원본물리서버와 가상서버

OpenStackit을 통해서 콘솔 화면을 통해서 가능하며, 리눅스 계열은 putty와 같은 SSH⁴⁾ 클라이언트로 접속하여 확인할 수도 있다.

이번 실험을 통해서 애플레이션 방식을 통해서 원본물리서버 시스템 전체를 본래의 특성(기능, 모습 등)을 모두 보존할 수 있는지의 가능성을 파악할 수 있었다. 앞서 3.2와 3.3의 실험에서는 전자기록(아래아 한글, 보석글 파일)을 확인할 수 있는 운영 환경에 대한 정보만을 이용하여, 클라우드 컴퓨팅 환경에 동일한 운

영체제 및 구동SW를 설치한 다음 전자기록을 확인하였다. 반면, 이번 실험에서는 전자기록을 포함한 운영 환경 전체를 클라우드 컴퓨팅 환경에 이관하였으며 운영 환경 전체를 보존할 수 있다는 것을 확인하였다. 전자기록의 유형이 전자문서, 시청각기록물, 웹기록물, 행정정보데이터세트 등 다양한 형태로 확대되고 있으며, 디지털 객체는 물론 해당 객체를 구동하는 시스템 하드웨어 및 소프트웨어를 포함한 운영 환경까지 보존하는 것을 고려해야 하는 상황들

4) SSH는 Secure Shell의 약자로 네트워크를 통해 다른 컴퓨터에 로그인하거나 원격 시스템에서 명령을 실행하기 위한 응용 프로그램 또는 프로토콜임.

이 생기고 있는 상황에서 시스템 전체를 이관하는 방식은 전자기록물을 장기보존할 수 있는 좋은 해결책이 될 것으로 기대된다.

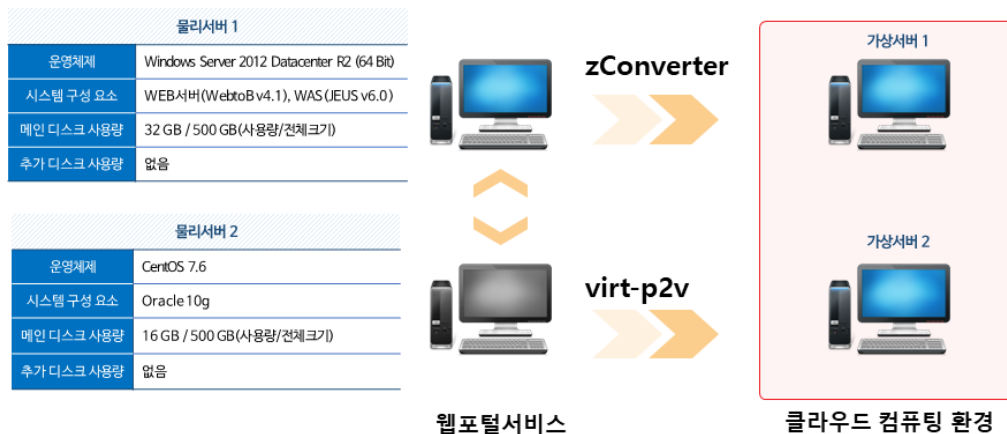
3.5 다수 서버 시스템 에뮬레이션 실험

네 번째 실험은 2대의 서버가 서로 연결되어 기본적인 웹포털서비스를 제공하고 있는 시스템을 클라우드 컴퓨팅 환경의 가상서버로 이전하여 에뮬레이션을 수행하는 것이다. 시스템을 다른 환경으로 이관하면 하나의 서버에 대한 설정, 시스템 연계를 위한 설정, 외부 서비스와의 연계를 위한 설정 등이 필요하며, 설정 과정은 상황에 따라 다양하고 전문성이 필요하다. 이러한 상황을 실험하기 위해 <그림 19>와 같이 웹 포털서비스를 제공하는 물리서버 2대를 각각 zConverter와 virt-p2v를 사용하여 P2V 변환을 수행한 다음 클라우드 컴퓨팅 환경에 가상서버를 생성하는 과정을 통해 에뮬레이션을 수행한다.

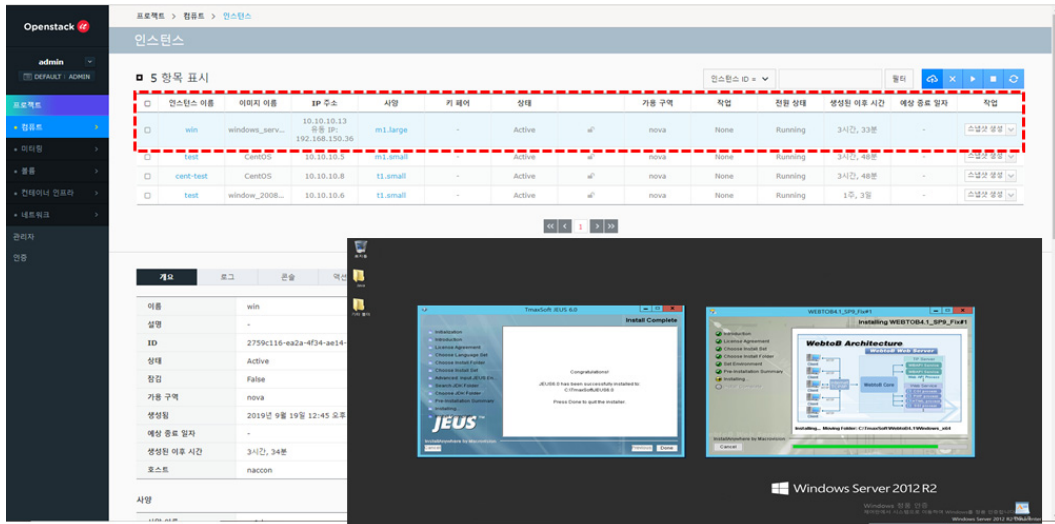
먼저, 물리서버 1은 웹포털서비스의 입구

에 해당하는 곳으로 Web서버와 WAS(Web Application Server)가 설치되어 있다. Web서버로는 WebtoB version 4.1이, WAS로는 JEUS version 6.0이 설치되어 있다. 물리서버 2는 웹포털 서비스의 자료를 보관하는 곳으로 데이터베이스(Oracle 10g)가 설치되어 있다. 물리서버 1은 3.4의 실험에서와 같이 <그림 16>의 에뮬레이션 절차를 통해 물리서버 1에서 가상서버 1로 시스템이 이전된다. <그림 20>은 물리서버 1이 클라우드 컴퓨팅 환경에서 원래의 모습 그대로 가상서버 1에서 실행되고 있는 모습을 보여준다.

물리서버 2는 virt-p2v에 의해서 <그림 21>의 절차에 의해서 클라우드 컴퓨팅 환경으로 이전된다. ① 에뮬레이션 대상 물리서버 2(Oracle 10g가 설치된 CentOS 7.6)를 준비한다. ② 물리서버 2에 virt-p2v를 설치하고, 물리서버 2가 virt-p2v를 통해 부팅하여 물리서버 2 템플릿 이미지는 생성하여 중계서버에 템플릿 이미지를 전송한다. ③ 중계서버는 물리서버 2가 송신한 템플릿 이미지를 수신하고, 중계서버의 virt-p2v



<그림 19> 2대의 물리서버의 클라우드 컴퓨팅 기반 에뮬레이션 과정 개요



〈그림 20〉 클라우드 컴퓨팅 환경 가상서버 1에 설치된 물리서버 1 및 실행화면



〈그림 21〉 virt-p2v 기반 전자기록 시스템 에뮬레이션 실험 절차

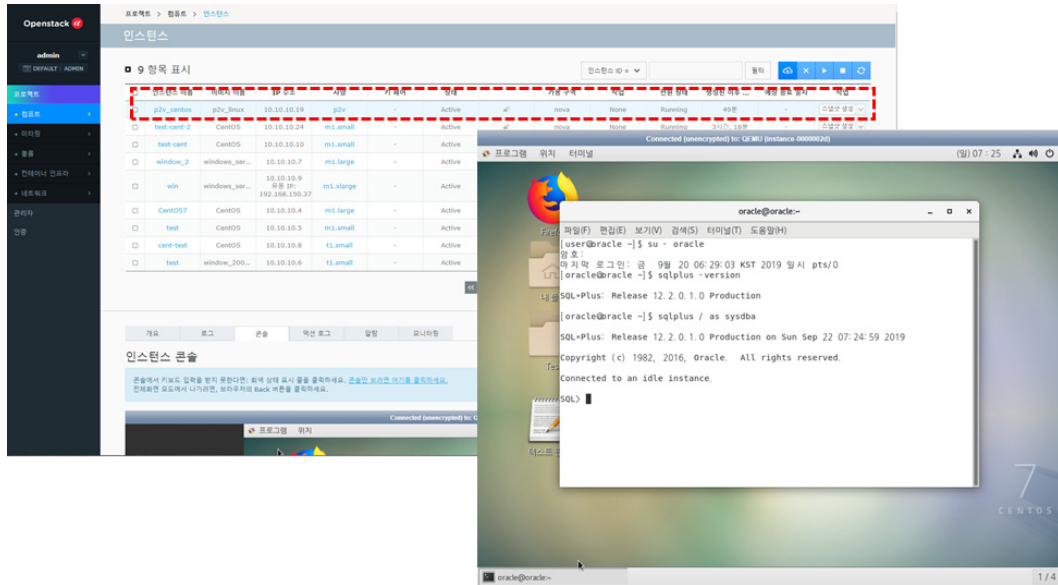
와 virt-managmer를 통해서 해당 이미지를 확인하고 클라우드 컴퓨팅 환경에 적합한 qcow2⁵⁾ 형태의 이미지를 생성한다. ④ 생성된 이미지는 virt-p2v에 의해서 클라우드 컴퓨팅 환경으로 이전된다. ⑤ 변환된 이미지로 클라우드 컴퓨팅 환경에서 가상서버 2를 생성한다.

〈그림 22〉는 물리서버 2가 클라우드 컴퓨팅 환경에서 원래의 모습 그대로 가상서버 2에서 실행되고 있는 모습을 보여준다.

지금까지 2대의 물리서버 1, 2를 클라우드

컴퓨팅 환경으로 이전하여 각각의 서버를 실행하는 데에는 성공하였지만 실제 웹포털서비스에는 접속되지 않았다. 물리서버 1과 물리서버 2가 설정되어 있는 네트워크는 192.168.100.*였고, 물리서버 1의 IP주소는 192.168.100.101, 물리서버 2의 IP주소는 192.168.100.102로 설정되어 있다. 그리고, 2개의 물리서버가 클라우드 컴퓨팅 환경으로 이전되면서 가상서버 1에는 10.10.10.13, 가상서버 2에는 10.10.10.19의 IP주소를 설정하였다. 그래서 사용자가 가상서

5) QCOW2는 가상 머신 디스크 이미지 저장 형식입니다. QCOW는 QEMU Copy On Write의 약자임.



〈그림 22〉 물리서버 2가 클라우드 컴퓨팅 환경에서 에뮬레이션된 가상서버 2 및 실행화면

버 1의 Web서버를 접속하면 WAS를 통해서 가상서버 2의 DBMS에 접속하려고 할 때, 10.10.10.19로 접속하지 않고, 원본물리서버에 설정되어 있는 192.168.100.102로 접속을 시도한다. 그러므로 Web서버/WAS 응용 프로그램에서의 DBMS 접속 정보를 모두 192.168.100.102에서 10.10.10.19로 수정하였고 웹포털서비스는 정상 동작하였다.

이번 실험을 통해서 전자문서를 에뮬레이션으로 확인하는 것을 넘어, 여러 서버로 구성된 시스템에서 생성되고 운영되는 행정정보데이터 세트, 웹기록물까지도 원래의 모습과 기능 그대로를 에뮬레이션 방법으로 보존할 수 있다는 가능성을 확인할 수 있다. 본 실험에서 진행된 2개의 서버로 구성된 시스템은 DBMS와 Web서버/WAS와의 내부 서버들 사이의 연계를 위한 정보만 수정하면 되는 단순한 설정이었다. 그러나 시스템이 복잡해질수록 서버들에 대한 세

부 설정, 특히, 외부 서비스와의 연계를 위한 설정은 더욱 복잡하고 어려운 작업이다. 향후에는 지속적으로 더욱 복잡한 구조의 시스템을 가정하여 테스트베드를 구축하여 실험함으로써 시스템 전체를 에뮬레이션하는 성숙된 기술을 확보해야 한다.

3.6 에뮬레이션 전략을 활용한 장기보존 방안

현재 우리나라는 전자문서, 행정정보데이터 세트, 시청각기록물, 웹기록물 등 다양한 전자기록 유형 중에 전자문서 유형의 전자기록에 대해서는 마이그레이션 전략을 활용하여 체계적인 시스템을 구축하고 있다. 초기에는 전자문서의 비중이 높아졌으나 최근에는 행정정보데이터세트, 시청각기록물, 웹기록물의 비중이 높아지고 있지만 이에 대한 대응은 진행되고 있지만 많은 보완과 연구가 필요하다.

이런 상황에서 에뮬레이션 전략은 마이그레이션 전략을 대체 방안이 아니고 현재 마이그레이션 전략을 해결하지 못하거나 포용하지 못하는 전자기록을 장기보존하는 보완 방안으로 활용해야 한다. 그래서 본 연구에서는 다음과 같은 3개의 장기보존 활용방안을 제시한다.

첫 번째, 아카이브는 유형, 매체, 시대에 관계 없이 생산된 다양한 기록을 수집하며 해당 기록의 진본성·신뢰성·무결성·이용가능성을 보장할 수 있어야 한다. 종이기록은 종이와 필기구의 수명을 고려하여 관리된다면 기록의 4대 속성이 보장될 수 있지만 다양한 매체에 수록된 전자기록은 다르다. 전자기록이 담긴 매체의 수명에 따라 잘 보존한다 하더라도 해당 매체를 읽을 수 있는 기계와 매체에 담긴 파일을 재생할 수 있는 애플리케이션이 없으면 읽고 해석할 수가 없다. 따라서 전자기록은 이용가능성 보장을 위해 별도의 조치가 필요한 것이다.

정부업무관리시스템을 사용하여 표준화된 전자문서를 사용하는 현재와는 달리 PC 도입 초창기에는 (구)전자문서시스템을 이용한 DOS 기반의 워드프로세서들이 공문서로 활용되었다. 에뮬레이션 대상으로 실험한 보석글은 당시 사용되었던 초기 전자문서 중 하나로 현재는 이를 재현할 애플리케이션의 공급이 중단된 상태이다. 또한 당시에 사용되던 OS와 저장매체를 읽을 수 있는 PC도 현재는 찾아보기 어려운 실정이다.

클라우드 컴퓨팅 기반 에뮬레이션 전략은 이처럼 과거에 사용하여 생산된 기록은 존재하지만 이를 재현하여 이용가능성을 보장할 수 있는 기술이 사라진 전자기록에 대해 필요한 전략이다. 1.3의 미국의 올리브와 독일의 bwFLA 역

시 각각 CDROM에 저장된 애플의 Applie II, Macintosh, MS DOS, Windows 3.1 등의 예전 OS에서 예술작품, 예전 프로그램(게임, 브라우저, 워드프로세서 등)의 재현을 위해 추진된 클라우드 컴퓨팅 기반 에뮬레이션 프로젝트다.

두 번째, 행정정보데이터세트는 데이터베이스에 저장되어 있지만, 일반인들이 실제 데이터베이스에 접속하여 데이터를 직접 확인하지 않는다. 대부분 데이터베이스, 웹서버, WAS 등이 외부 응용 프로그램이나 솔루션과 연계하여 전자문서 또는 그래프 등의 형태로 확인할 수 있는 사용자 인터페이스를 제공한다. 예를 들어, 대학 학사정보시스템의 행정정보데이터세트는 데이터베이스 안에 있는 데이터들은 웹솔루션을 통해서 출력되는 재학증명서, 성적증명서, 재직증명서, 수강내역, 강의내역 등의 전자문서 형태로 확인한다. 그렇다고 이러한 전자문서들을 모두 파일이나 종이로 출력하여 보관하는 것은 현실적이지 않다. 대신에 데이터베이스, 웹서버, WAS, 웹솔루션 등 시스템 전체를 클라우드 컴퓨팅 환경으로 이전하여 에뮬레이션 하는 것이 좋은 해결책을 제공할 수 있을 것으로 기대된다.

세 번째, 청와대, 국민청원 등 국가적으로 중요한 홈페이지는 모습, 기능, 데이터 모두 원래의 모습 그대로 보존되어야 한다. 모습은 캡처 화면, 데이터는 덤프파일 등으로 별도로 보관하는 것보다 에뮬레이션 전략을 도입하는 것이 모습, 기능, 데이터 모두를 장기보존하는데 좋은 해결책이 될 수 있다.

이렇게 에뮬레이션 전략은 기존의 마이그레이션 전략의 보완재 역할을 할 수 있으며, 클라우드 컴퓨팅 기술과 접목되어 에뮬레이션 전략이 도

입되면 로컬 환경에서 에뮬레이션 전략에 비해 아래와 같은 3가지의 큰 장점을 가지고 있다.

첫 번째로, 클라우드 컴퓨팅을 활용하는 근본적인 이유에서 장점을 확인할 수 있다. 클라우드 컴퓨팅 구축 허가에 대한 정책적인 결정만 이루어진다면 자체적으로 서버 구매하고 설치하는 것보다 빨리 시작할 수 있으며, 인프라 확장이 용이하고, 공간 및 유지관리 비용 절감할 수 있는 장점이 있다.

두 번째로, 클라우드 컴퓨팅 기술은 가상화, 분산처리 기술과 함께 발전해 왔다. 그러므로 클라우드 컴퓨팅을 구성하는 기술에는 기본적으로 다양한 운영 환경을 재생산하는 에뮬레이션 기술의 핵심이 기본적으로 포함되어 있다. 그러므로 클라우드 컴퓨팅 기술 이외에 에뮬레이션 기술이 별도로 신경 쓸 필요 없이 클라우드 컴퓨팅 기술에 지속적으로 관심을 가진다면 에뮬레이션 기술의 수준은 자연스럽게 높아진다.

세 번째로, 대국민 서비스에 용이하다. 클라우드 컴퓨팅 서비스는 기본적으로 원격에서 접속하는 것이 기본이기 때문이다. 본 연구의 클라우드 컴퓨팅 환경에서도 사용자가 웹 브라우저를 통해서 OpenStack에 접속하면 원하는 사양의 컴퓨터를 클라우드 컴퓨팅 환경에서 생성하여 접속할 수 있다. 이후, VNC, TeamViewer, 원격데스크톱 등 일반인에게도 익숙한 원격 접속 프로그램을 이용해서 시스템에 접근할 수 있다. 또한, 대국민 서비스를 위해서는 지속가능성이 중요하다. 클라우드 컴퓨팅 환경을 구성하는 OpenStack 구성 서버들이 공격당하면 서비스할 수 없다. 이러한 상황은 다른 서비스들에서도 마찬가지이므로, 클라우드 컴퓨팅 환경에서 특정 시스템을 에뮬레이션하는 가상서버에 국

한한다. 해당 가상서버는 템플릿 이미지(설치 파일)에서부터 생성되었고, 하나의 템플릿 이미지로부터 여러 개의 동일한 가상서버가 생성될 수 있다. 만약 하나의 가상서버가 외부로부터 공격받거나 이상 현상이 발생하여도 해당 가상서버를 클라우드 컴퓨팅 환경에서 삭제하고 다시 템플릿 이미지로부터 다른 가상서버를 생성한다면 서비스를 지속할 수 있다. 그러므로 클라우드 컴퓨팅 환경에서 에뮬레이션 전략을 활용한다면 기본적으로 서비스의 지속가능성이 높아진다.

3.7 클라우드 컴퓨팅 기반 에뮬레이션 전략의 한계점

IT기술은 항상 빠르게 발전하고 있으며 특히 시장 및 자본이 집중되는 분야에서는 더욱 그러하다. 클라우드 컴퓨팅 기술 CPU 프로세서 기술에 의존적이며, “현재 시점에서” 가장 많이 사용하고 있는 CPU 프로세서 기반으로 발전되어 있다. 현재 가장 많이 사용되고 있는 CPU 프로세서 구조는 Intel 및 AMD의 x86 CPU 프로세서이다. 클라우드 컴퓨팅 기술은 이들 CPU 프로세서를 중심으로 발전되었기 때문에 대부분의 가상화 기술은 x86 기반이다. 그래서, SPARC CPU 프로세서에 동작하는 HP-UX의 유닉스 계열 OS는 현재 원래 모습 그대로의 에뮬레이션은 어려운 실정이다. 현재 오래되었으면서 중요한 서버들의 OS에는 유닉스 계열이 많이 존재한다. 그러나 유닉스 계열의 OS는 현재 시장이 없어지고 있는 실정으로 IT 기술과 자본이 외면하고 있는 추세이다. 전자기록 장기보존에는 가장 발전되고 안정적인 IT 기술을 활용하여

야 하는데 IT 자본이 외면하고 있는 IT 기술이 성숙되기를 기다릴 수는 없다고 판단된다. 그러므로 유닉스 계열 OS의 경우에는 시스템이 보존되어야 할 필수 속성들을 도출하여 유닉스 계열 OS를 가장 유사한 리눅스 계열 OS로 전환하는 U2L(Unix-to-Linux)을 활용하여 원본과 동일 기능과 유사한 외관을 보여 줄 수 있는 방법이 좋은 해결책이 될 수 있다.

CPU, 메모리, 디스크, 네트워크 등 컴퓨팅 하드웨어 자원을 추상화하는 기술을 가상화 기술이며, 여러 서버들의 가상화된 컴퓨팅 하드웨어 자원을 하나의 컴퓨팅 자원 Pool로 묶고 기술을 분산처리 기술이다. 그리고 컴퓨팅 자원 Pool 상에서 사용자의 요청에 따라 알맞은 OS를 컴퓨팅 자원에서 설치하여 컴퓨터를 생성하는 것을 애플리케이션 기술이라고 한다. 본 연구에서는 가상화, 분산처리, 애플리케이션 기술로 각각 KVM, OpenStack, QEMU를 활용하고 있는데 각각의 기술은 지원하는 CPU 프로세스 종류, OS 등에 제한을 가지고 있다. 또한, 클라우드 컴퓨팅 환경에 신규로 서버를 생성하지 않고 기존의 레거시 서버를 클라우드 컴퓨팅 환경으로 전환할 때 활용되는 P2V 변환기인 zConverter도 지원하는 <표 7>과 같이 OS

종류에 제한이 있다. 이러한 기술들도 대부분의 Windows 계열과 리눅스 계열의 OS를 지원하고 있다. 그러므로 이들이 지원하지 않는 OS의 경우에는 지원하는 OS로 시스템 변환을 하는 U2L과 같은 방안이 효과적인 차선을 제공할 수 있다.

4. 맺음말

전자기록 장기보존의 핵심은 전자기록이 최초로 생성되고 활용되었던 본래의 기능적 속성과 비트스트림을 오랫동안 그대로 유지하는 것이다. 그러나 대부분 국내외의 주요 아카이브 기관들은 경제적, 기술적인 측면을 고려하여 원본의 비트스트림에 변경을 허용하는 마이그레이션을 주전략으로 채택하고 있다. 그러나 전자기록 유형이 다양해지고 범위가 확장됨에 따라 마이그레이션 전략으로 진본성과 무결성을 유지하는데 한계를 드러내고 있다.

그러므로 본 연구에서는 클라우드 컴퓨팅 기술을 활용하여 전자기록의 비트스트림을 변경하지 않고 유지할 수 있도록, 전자기록이 생산·활용된 시스템 및 응용 환경을 재생산하는 애플

<표 7> zConverter 지원 OS

구분	Windows	Linux
버전	Win 2012 R2 64bit Win 2008 R2 64bit Win 2008 ENT 64bit Win 2008 ENT 32bit Win 2003 64bit Win 2003 32bit	CentOS 4.8 32bit, CentOS 5.8 64bit, CentOS 6.3 32bit, CentOS 6.3 64bit, CentOS 6.4 32bit, CentOS 6.4 64bit, Ubuntu 10.04 32bit, Ubuntu 10.04 64bit, Ubuntu 12.04 32bit, Ubuntu 12.04 64bit SUSE Linux Enterprise 11 SP2 32bit SUSE Linux Enterprise 11 SP2 64bit

* 출처: Zconverter, <http://www.zconverter.co.kr/index.php/product/Migration/>

레이션 적용 가능성을 검토하고자 한다. 실제 에뮬레이션 테스트베드를 2개의 서버와 2개의 스위치로 구성하였으며, 하이퍼바이저는 KVM, 에뮬레이터는 QEMU를 사용하였으며, 이를 바탕으로 클라우드 컴퓨팅 환경을 구성하는 클라우드 운영체제는 OpenStack Queens 버전을 사용하였다.

네 가지 실험을 진행하였다. 첫 번째와 두 번째 실험은 재현하고자 하는 전자문서가 구동될 수 있는 환경을 클라우드 컴퓨팅 환경에 조성한 다음 전자문서를 재현하는 실험이었다. 연구개발 사업을 통해 클라우드 에뮬레이션 환경을 구축하여 검증한 결과, DOS 및 Windows XP 환경에서 각각 보석글 워드프로세스와 한컴오피스 애플리케이션을 그대로 재현할 수 있었다. 이는 기록의 내용·구조·맥락에 더해 전자적 특성인 당시의 그대로의 외형과 기능을 재현함으로써 에뮬레이션이 기록의 속성을 보장하는 방안이 될 수 있음을 확인한 것이다. 그리고 세 번째와 네 번째 실험은 하나 이상의 서버들로 구성된 시스템 자체를 백업 도구를 활용하여 템플릿 이미지를 생성하여 클라우드 컴퓨팅 환경에 가상서버들을 설치하여 원본물리서버와 동일하게 실행하는 것을 실험하여 확인하였다. 이 실험을 통해서 전자기록의 유형이 다양한 형태로 확대되고 있으며, 디지털 객체는 물론 해당 객체를 구동하는 시스템 하드웨어 및 소프트웨어를 포함한 운영 환경까지 보존하는 것을 고려해야 하는 상황들이 생기고 있는 상황에서 시스템 전체를 이관하는 방식은 전자기록물을 장기보존할 수 있는 좋은 해결책 제시할 것으로 기대할 수 있었다. 또한, 클라우드 컴퓨팅 기반 에뮬레이션 전략은 클라우드 컴퓨팅

본래의 장점과 안정적으로 대국민 서비스를 운영할 수 있다는 장점도 있다.

그래서 에뮬레이션 전략은 마이그레이션 전략을 대체하는 방안이 아니고 현재 마이그레이션 전략을 해결하지 못하거나 포용하지 못하는 전자기록을 장기보존하는 보완 방안으로 활용할 수 있으며 오랫동안 확인하지 못했던 전자기록, 다양한 시스템 요소들이 연계되어 전자기록이 생성되는 경우, 국가적으로 중요한 홈페이지의 경우를 클라우드 컴퓨팅 기반 에뮬레이션 전략을 도입해야 한다.

그러나 현재 가장 x86 CPU 프로세스 기반으로 하는 클라우드 컴퓨팅 기술에 집중되어 있기 때문에 x86 이외에 다른 CPU 프로세스를 기반으로 OS는 직접 클라우드 컴퓨팅 환경에서 에뮬레이션 전략을 적용하는 것이 어렵다. 이러한 경우는 원본과 동일 기능과 유사한 외관을 보여줄 수 있으면서 클라우드 컴퓨팅 환경이 지원하는 OS로 변환하여 에뮬레이션 전략 도입하는 방법이 좋은 해결책이 될 수 있다.

지금까지 에뮬레이션은 마이그레이션 비해 비용이 높고 구현이 어려운 장기보존 전략으로 평가받아 왔다. 하지만 상용 에뮬레이터의 가격이 상대적으로 낮아지고 클라우드 컴퓨팅 기술이 발전하면서 가치 있는 중요 전자기록물에 대해 선별적으로 적용할 만한 기술로 진화되고 있다. 진화하는 디지털 환경에서 데이터세트, 트위터, 유튜브 등 국가 아카이브의 기록관리 대상은 급격히 증가하고 있다. 이에 대응하여 국가 아카이브 기술의 변화 추이를 지속적으로 모니터링하여 다양한 기록관리 전략을 확보할 필요가 있다.

참 고 문 헌

- 강맹수 (2019). 클라우드 컴퓨팅 시장 동향 및 향후 전망. 산은조사월보, 1(758), 54-71.
- 국가기록원 (2013). 행정기관 전자기록물 재현기술 연구 및 프로토타입 개발 완료보고서. 대전: 국가기록원
- 김기정, 신동수 (2018). 클라우드 컴퓨팅 환경 영구기록물관리 시스템 구축 방안 연구. 한국기록관리학회지, 18(3), 49-70. DOI: <https://doi.org/10.14404/JKSARM.2018.18.3.049>
- 김명훈, 오명진, 이재홍, 임진희 (2013). 전자기록 장기보존 전략으로서의 에물레이션 사례 분석. 기록학연구, (38), 265-309.
- 김주영, 김순희 (2019). 클라우드 저장소를 활용하여 기록생산시스템에서 기록관리시스템으로 전자기록물을 이관하는 방안에 관한 연구. 한국기록관리학회지, 19(2), 1-24.
DOI: <http://doi.org/10.14404/JKSARM.2019.19.2.001>
- 박종근, 최강일, 이상민, 이정희, 이범철 (2013). OpenStack 클라우드 네트워크 기술 분석. 전자통신동향분석, 28(5), 122-132.
- 소정의, 한희정, 양동민 (2018). 국외 전자기록물의 장기보존 정책 비교 분석. 한국기록관리학회지, 18(4), 125-148. DOI: <http://doi.org/10.14404/JKSARM.2018.18.4.125>
- 왕호성, 설문원 (2017) 행정정보 데이터세트 기록의 관리방안. 한국기록관리학회지, 17(3), 23-47.
DOI: <http://doi.org/10.14404/JKSARM.2017.17.3.023>
- 이승연, 설문원 (2017). 전자기록관리정책의 재설계에 관한 연구. 기록학연구, (52), 5-37.
- 임지훈, 김은충, 방기영, 이유진, 김용 (2014). 클라우드 컴퓨팅 기반의 전자기록관리시스템 구축방안에 관한 연구. 한국기록관리학회지, 14(3), 153-179.
DOI: <http://doi.org/10.14404/JKSARM.2014.14.3.153>
- 임진희 (2013). 전자기록관리론. 서울: 선인.
- 정예용, 심갑용, 김용 (2014). 클라우드 컴퓨팅 기반 중앙기록물관리시스템 설계 및 적용에 관한 연구. 한국비블리아학회지, 25(4), 209-233. DOI: <http://doi.org/10.14699/kbiblia.2014.25.4.209>
- DPT (2006). 전자기록의 유형별 보존기법(=Digital Preservation Testbed. From digital volatility to digital permanence). (이미화, 현문수 공역). 서울: 한국국가기록연구원(원전 발행년 2003).
- OLIVE (2019). Retrieved October 19, 2019, from <https://olivearchive.org/>
- bwFLA (2019). Retrieved October 19, 2019, from <http://eaas.uni-freiburg.de/>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

DPT (2006). From digital volatility to digital permanence(=Digital Preservation Testbed. From

- digital volatility to digital permanence). (Lee, Mi Hwa & Hyun, Moon Soo). Seoul: The Research Institute for Korean Archives and Records.
- Jung, Ye Yong, Shim, Gab Yong, & Kim, Yong (2014). A Study on Design and Application of Central Archives Management System Based on Cloud Computing. *Journal of the Korean Biblia Society for Library and Information Science*, 25(4), 209-233.
DOI: <http://doi.org/10.14699/kbiblia.2014.25.4.209>
- Kang, Maeng Soo (2019). Cloud Computing Market Trends and Future Challenges. *KDB Monthly*, 1(758), 54-71.
- Kim, Ju young & Kim, Soon Hee (2019). A Study on Transferring Electronic Records from Record Production System to Record Management System Using Cloud Storage. *Journal of Korean Society of Archives and Records Management*, 19(2), 1-24.
DOI: <http://doi.org/10.14404/JKSARM.2019.19.2.001>
- Kim, Ki Jung & Shin, Dong Soo (2018). A Study on the Archives Management System in Cloud Computing. *Journal of Korean Society of Archives and Records Management*, 18(3), 49-70.
DOI: <https://doi.org/10.14404/JKSARM.2018.18.3.049>
- Kim, Myoung hun, Oh, Myung Jin, Lee, Jae Hong, & Yim, Jin Hee (2013). An Analysis of Cases of Emulation for Long Term Electronic Records Preservation Strategy. *The Korean Journal of Archival Studies*, (38), 265-309.
- Lee, Seung Eok & Seol, Moon Won (2017). A Study of Redesigning Electronic Records Management Policies. *The Korean Journal of Archival Studies*, (52), 5-37.
- Lim, Ji Hoon, Kim, Eun Chong, Bang, Ki Young, Lee, Yu Jin, & Kim, Yong (2014). An Application Method Study on the Electronic Records Management Systems based on Cloud Computing. *Journal of Korean Society of Archives and Records Management*, 14(3), 153-179.
DOI: <http://doi.org/10.14404/JKSARM.2014.14.3.153>
- National Archives of Korea (2013). Research and prototype development of electronic records reproduction technology. Daejeon: National Archives of Korea.
- Park, Jong Keun, Choi, Kang Il, Lee, Sang Min, Lee, Jeong Hee, & Lee, Bum Chul (2013). Analysis of OpenStack Cloud Networking Technology. *Electronics and telecommunications trends*, 28(5), 122-132.
- So, Jeong Eui, Han, Hui Jeong, & Yang, Dong Min (2018). A Comparative Analysis of Long-Term Preservation Policies in Foreign Electronic Records. *Journal of Korean Society of Archives and Records Management*, 18(4), 125-148.
DOI: <http://doi.org/10.14404/JKSARM.2018.18.4.125>

Wang, Ho Sung & Seol, Moon Won (2017). A Study on Managing Dataset Records in Government Information Systems. *Journal of Korean Society of Archives and Records Management*, 17(3), 23-47. DOI: <http://doi.org/10.14404/JKSARM.2017.17.3.023>

Yim, Jin Hee (2013). *Electronic Records Management*. Seoul: Seon-in.

