

## 고해상도 농업 기후 자료 처리를 위한 클라우드와 온프레미스 비교 분석

박주현\* · 안문일<sup>1,2</sup> · 강위수<sup>3</sup> · 심교문<sup>3</sup> · 박은우<sup>4,5</sup>

<sup>1</sup>에피넷, <sup>2</sup>선문대학교, <sup>3</sup>국립농업과학원, <sup>4</sup>서울대학교, <sup>5</sup>국가농림기상센터  
(2019년 10월 24일 접수; 2019년 12월 16일 수정; 2019년 12월 19일 수락)

### Comparative Analysis on Cloud and On-Premises Environments for High-Resolution Agricultural Climate Data Processing

Joo Hyeon Park<sup>1\*</sup>, Mun Il Ahn<sup>1,2</sup>, Wee Soo Kang<sup>3</sup>, Kyo-Moon Shim<sup>3</sup> and Eun Woo Park<sup>4,5</sup>

<sup>1</sup>R&D Center, EPINET Co., Ltd., Anyang 14056, Korea

<sup>2</sup>Department of Life Science and Biochemical Engineering, Sunmoon University, Asan 31460, Korea

<sup>3</sup>National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Korea

<sup>4</sup>Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea

<sup>5</sup>National Center for Agro-Meteorology, Seoul 08826, Korea

(Received October 24, 2019; Revised December 16, 2019; Accepted December 19, 2019)

#### ABSTRACT

The usefulness of processing and analysis systems of GIS-based agricultural climate data is affected by the reliability and availability of computing infrastructures such as cloud, on-premises, and hybrid. Cloud technology has grown in popularity. However, various reference cases accumulated over the years of operational experiences point out important features that make on-premises technology compatible with cloud technology. Both cloud and on-premises technologies have their advantages and disadvantages in terms of operational time and cost, reliability, and security depending on cases of applications. In this study, we have described characteristics of four general computing platforms including cloud, on-premises with hardware-level virtualization, on-premises with operating system-level virtualization and hybrid environments, and compared them in terms of advantages and disadvantages when a huge amount of GIS-based agricultural climate data were stored and processed to provide public services of agro-meteorological and climate information at high spatial and temporal resolutions. It was found that migrating high-resolution agricultural climate data to public cloud would not be reasonable due to high cost for storing a large amount data that may be of no use in the future. Therefore, we recommended hybrid systems that the on-premises and the cloud environments are combined for data storage and backup systems that incur a major cost, and data analysis, processing and presentation that need operational flexibility, respectively.

**Key words:** Agricultural climate data, Cloud, On-premises, Hybrid, Gluster file system



\* Corresponding Author : J. H. Park  
(parkjh@epinet.kr)

## I. 서 론

농업, 산림, 수문 등의 분야에서 지리정보체계 (Geographic Information System, GIS) 기술을 이용한 고해상도 격자형 기상·기후 자료를 처리하는 연구가 계속되고 있다(Choi *et al.*, 2008; Yun, 2015; Park *et al.*, 2017; Won *et al.*, 2018). GIS 기술을 이용한 격자형 기상자료로서 기상청이 배포하는 동네예보 (Dong-Nae Forecast) 자료는 한반도 영역에 대해 5km 공간해상도로 3시간 마다 갱신되며, KLAPS (Korea Local Analysis and Prediction System) 자료는 동네예보와 동일한 한반도 영역에 대해 5km 공간해상도로 1시간 마다 갱신되는 자료이다(Kim and Yun, 2013). 또한 COMS (Communication, Ocean and Meteorological Satellite), UM (Unified Model), 레이더(Rader) 등의 자료 역시 특정 공간해상도, 시간해상도 및 갱신 주기를 갖는 격자형 자료이다(KMA, 2019). 이러한 기초 자료들은 특히 농촌진흥청과 전라북도농업기술원에서 운영 중인 농업 기상 재해 조기경보 서비스(AGMET, 2019)와 농업 생태계 기후 및 이상기상 변화량 정보 시스템(AGECOCLIM, 2019)에서 필수적으로 이용되고 있다. 농업기상재해조기경보서비스는 각 시군 별로 날씨(최고기온, 최저기온, 강수, 일사, 일조, 풍속 등), 작목 및 품종 별 생육단계 그리고 고온해, 저온해, 가뭄해 등과 같은 기상위험위험지수를 약 30m 해상도의 격자형 자료를 생산하여 서비스하고 있다(Shin *et al.*, 2015). 이러한 사례와 같이 GIS 기술을 이용한 격자형 자료를 생성, 저장, 공유를 원활하게 하기 위해서는 각 자료의 특성에 따라 자료 계산을 위한 다양한 기술적 요소들을 고려해야 하며, 시공간적으로 고해상도의 자료일수록 기술적 요소의 중요성은 증가한다. 예를 들어 10m의 공간적 해상도의 자료는 100m의 공간적 해상도 대비  $10^2$  배의 컴퓨팅 자원을 필요로 한다. 더욱이 연직 방향 또는 시간에 대한 차원이 추가될 경우 차원의 개수에 대해 지수적으로 증가한다.

급속한 컴퓨터 기술의 발달에 힘입어 GIS 기술도 크게 발전하면서 다양한 방법을 이용하여 농업 기후 자료를 처리하고 표현할 수 있게 되었다. 특히 각광받고 있는 클라우드 시스템을 이용한 컴퓨팅 자원의 활용은 기존의 GIS 자료에 기반한 시스템의 개발 및 운용에 대한 접근 방법을 크게 바꾸었으며, 편의성과 확장성 등을 장점으로 대량의 자료를 처리하는데 많이 이용되어왔다(Kim, 2011). 이와 더불어 물리적 서버 전

반을 직접 관리하는 기존의 온프레미스(On-Premises) 시스템은 장기간의 운영 경험에서 비롯된 노하우와 가시성 확보를 강점으로 하여 현업에서 여전히 중요하게 사용되고 있다. 온프레미스는 물리적 서버, 네트워크, 스토리지, 전원과 같은 직접적 자원을 포함하여 항온, 항습, 방진 등의 공조시설과 화재, 침입, 누수, 누전 등과 같은 재난 대비 시설 그리고 이를 운영하기 위한 정책 및 인력 등이 고려되어야 하는 어려움이 있다. 반면 클라우드 시스템은 전산 인프라 운영에 필요한 많은 부분을 고려하지 않고, 어플리케이션 자체에만 집중할 수 있도록 필요에 따라 자원을 임대하고 반환하는 절차를 매우 유연하게 처리할 수 있는 방법을 제공한다. 하지만 자료의 크기, 이동 그리고 데이터 수명 주기 관리 방식에 따라 매우 큰 비용의 차이가 발생하므로 다양한 측면에서 살펴볼 필요가 있다. 본 연구는 대용량, 고해상도의 농업 기상 및 기후 자료를 처리하는 것을 목적으로 하는 비용 최적화된 컴퓨팅 환경의 구성 방법을 찾고자 하였다. 비용, 성능, 편의성, 기존 시스템과의 연계 등을 고려하여 현재 구축된 농업 생태계의 이상기상 변화량 정보 시스템을 기준으로 가장 적합한 전산 기반 시설(Computing Infrastructure) 구성 방법을 알아보았다.

## II. 클라우드 환경

클라우드 컴퓨팅(Cloud Computing)은 인터넷을 통하여 제공되는 어플리케이션들과 이 서비스들을 제공하는 데이터 센터 내의 모든 하드웨어와 소프트웨어를 포함하는 용어이다(Lee, 2018). ‘클라우드컴퓨팅 발전 및 이용자 보호에 관한 법률’ 제2조 1호는 ‘집적·공유된 정보통신기기, 정보통신설비, 소프트웨어 등 정보통신자원을 이용자의 요구나 수요 변화에 따라 정보통신망을 통하여 신축적으로 이용할 수 있도록 하는 정보처리체계’ 라고 정의한다. 클라우드 컴퓨팅 환경은 운영 주체에 따라 크게 프라이빗 클라우드(Private Cloud)와 퍼블릭 클라우드(Public Cloud)로 나눌 수 있다. 퍼블릭 클라우드는 클라우드 컴퓨팅 서비스를 위한 하드웨어 및 소프트웨어 전반이 클라우드 서비스를 제공하는 업체가 소유하고 관리하며 여러 고객이 이를 공유하는 형태를 말한다. 퍼블릭 클라우드는 유지 관리에 대한 아웃소싱, 높은 안정성, 매우 높은 확장성을 특징으로 한다. 그에 반해 프라이빗 클라우드는 클라우드 컴퓨팅 서비스를 위한 하드웨어 및 소프

트웨어 전반의 소유권이 운영 주체에게 있으며 해당 주체가 독점적으로 사용하는 형태를 말한다. 또한 퍼블릭 클라우드 대비 보안 강화, 투명성 향상 등이 특징이며, 물리적 서버를 직접 운영하는 방식 대비 유연성 향상을 특징으로 한다. 하이브리드 클라우드는 온프레미스 또는 프라이빗 클라우드를 퍼블릭 클라우드와 결합하여, 각각의 장점을 취합한 방식이다. 일반적으로 비용 효율성, 보안성, 기존 인프라와의 호환성 등이 특징이다. 본 연구에서는 온프레미스와 명확한 차이가 있는 퍼블릭 클라우드에 대해서만 언급하도록 한다.

퍼블릭 클라우드는 자원을 공유하는 고객의 수가 증가함에 따라 높은 수준의 규모의 경제를 이룰 수 있기 때문에 일정 부분에서는 비용의 장점이 있고, 특히 종량제, 후불제 기반의 요금 정책으로 인해 초기 투입 비용이 온프레미스 대비 크게 낮은 것이 큰 장점이다. 하지만 임대 방식으로 제공되는 가상의 컴퓨팅 자원이 어떠한 특성의 실제의 하드웨어 기반에서 동작하는지에 대한 투명성이 낮고 퍼블릭 클라우드 서비스 제공자(Cloud Service Provider, CSP)가 제공하는 기능을 벗어나는 활동이 요구되는 부분에서는 그 제어가 불가능하다. 퍼블릭 클라우드는 초기 투입 비용에 대한 부담 없이 다양한 사용자 선택 가능 옵션으로 인프라 구성이 가능하며 언제든지 변경 및 반환할 수 있다는 장점이 있다. 자료 저장 측면에서는 CSP 마다 고유한 기능이 제공되며 동일한 CSP 내에서도 사용자의 선택에 따라 10배 이상의 비용 차이가 발생할 수 있다 (Table 1). 대략적인 비용 추정을 위해 농업 생태계 기후 및 이상기상 변화량 정보 시스템을 기준으로 보면 1년간 2TiB 자료 저장에 소요되는 비용에 한하여 최소 가격 옵션에 의해 약 730만원, 최대 1,770만원의 비용이 소요될 수 있다(Table 2). 해당 시스템에서 현재까지 생성한 자료는 1980년도부터 2018년도까지의 일별, 월별, 연도별 자료이며 소규모의 정적 자료 및 기타 자료를 포함하여 약 20TiB이다. 또한 1년당 약 2TiB 이하의 저장공간이 추가로 필요하므로 비용은 시간이 지날수록 증가한다.

CSP에서 제공하는 공유 파일 스토리지는 참조 횟수가 적은 자료를 비용 효율적으로 관리할 수 있는 기능을 제공한다. 이는 일정 기간 동안 참조되지 않은 데이터를 가격이 상대적으로 저렴한 스토리지 클래스로 이동시키는, 데이터 수명 주기 관리 기능을 통해 이루어진다. 이와 같이 자료가 어느 스토리지 클래스에 위치하더라도 POSIX (Portable Operating System Interface) 호

**Table 1.** Costs for storing 1GiB data per month of leading CSPs in Korea

CSP	Storage Name	Cost (KRW/GiB/month)	
		Minimum	Maximum
Amazon AWS	EFS	29.6	355.8
Microsoft Azure	Files	67.4	296.9
Naver NCP	NAS	72.0	72.0
Gabia gCloud	NAS	50.0	50.0

\* 1USD = 1,201 KRW as of Oct. 01. 2019

**Table 2.** Costs for operation and storing 20TiB and an additional 1TiB data per year of leading cloud service providers in Korea and on-premises

CSP	Cost (KRW/year)		
	Operation	Fixed for 20TiB	Additional for 1TiB
Amazon Web Service	1,000,000	7,305,216	365,260
Microsoft Azure	500,000	16,586,342	829,317
Naver NCP	0	17,280,000	864,000
Gabia gCloud	0	12,000,000	600,000
On-Premises	500,000	5,000,000	100,000

\* 1USD = 1,201 KRW as of Oct. 01. 2019

환성을 유지함은 물론 기존의 파일의 접근 방식을 변경할 필요가 없다는 점이 특징이다. 해당 기능을 이용하면 Amazon Web Service (AWS, Amazon.com, Inc의 클라우드 서비스 상품명)의 경우 최대 85%까지 비용 감축 효과가 있으며 Microsoft Azure (Microsoft Corporation의 클라우드 서비스 상품명)도 비슷한 기능을 제공한다. 하지만 데이터가 비용 효율적인 스토리지 클래스로 이동했을 경우 데이터의 요청 횟수, 용량에 따라 추가적인 비용이 발생한다. 데이터 저장 외 추가 예상 비용은 CSP에서 제공하는 기능에 따라 차이가 있다(Table 2).

### III. 온프레미스 환경에서의 하드웨어 수준 가상화

가상화(Virtualization)는 대상이 되는 하드웨어 리소스에 따라 중앙처리장치 가상화, 메모리 가상화, 저장소 가상화, 네트워크 가상화 등으로 나누어 질 수 있다(VIRTUALIZATION, 2019). 하드웨어 수준 가상

화를 위한 도구로는 Microsoft Hyper-V (MICROSOFT, 2019), VMWare vSphere (VMWARE, 2019), Citrix XenServer (CITRIX, 2019) 등으로 대표되는 상용 도구들과 리눅스의 KVM (Kernel-based Virtual Machine) 과 기타 소프트웨어들이 조합된 오픈소스 도구가 있다 (Table 3). Hyper-V의 경우 다른 도구들에 비해 많은 CPU와 RAM을 지원하지만 이것이 성능과 직접적으로 관련된 것은 아니다. KVM의 경우 오픈소스임에도 다른 상용도구에서 지원되는 라이브 마이그레이션(Live Migration)이 지원되는 것이 특징이다. 라이브 마이그레이션은 특정 물리 서버에서 실행되고 있는 VM (Virtual Machine)의 상태를 유지한 채 다른 물리적 서버로 이동하는 기능이다. 또한 KVM은 하드웨어 수준

가상화는 물론 Docker와 LXC와 같은 운영체제 수준 가상화를 호스트 자체적으로 지원하는 것이 장점이다.

VMWare vSphere는 가상화 산업 전반에 가장 많이 사용되고 있는 도구로서 베어 메탈 서버(Bare Metal Server)에 근접하는 강력한 성능, 오랜 기간 동안 사용되어 높은 신뢰성을 확보하고 있다. Hyper-V는 전세계 데스크톱 사용자의 약 81% 이상을 확보하고 있는 Microsoft Windows 운영체제의 사용자 경험에 기반하고 상대적으로 낮은 가격의 라이선스 정책과 최소 2개부터 최대 무한대의 Windows OSE (Operating System Environment) 인스턴스를 생성할 수 있는 특징이 있다. 또한 Windows 10 Professional Edition에 기본적으로 포함되어 있어 해당 운영체제 사용자라면

**Table 3.** Features of four tools for hardware-level virtualization

Feature	Hyper-V 2019	ESXi 6.7	XenServer 7.6	KVM
RAM/Host	24 TB	12 TB	5 TB	12 TB
RAM/VM	12 TB	6 TB	1.5 TB	6 TB
CPUs/VM	240	128	32	240
VM Disk	64 TB	62 TB	2 TB	10 TB
VM Live Migration	Yes <sup>1</sup>	Yes <sup>1</sup>	Yes <sup>1</sup>	Yes
VM Replication	Yes <sup>1</sup>	Yes <sup>1</sup>	Yes <sup>1</sup>	Yes
Native LXC/docker	No	No	No	Yes
Open source	No	No	Yes <sup>2</sup>	Yes

<sup>1</sup> Only supported in commercial version

<sup>2</sup> Partially supported

**Table 4.** Performance of four tools for hardware-level virtualization in terms of CPU, RAM, storage and network

Hypervisor	Vendor	CPU (Calculation speed)	RAM (Read/Write speed)	Storage (Read/Write speed)	Network (Packet Send/Receive speed)
ESXi	VMWare	1.75x	1.26x	1.23x	1.07x
XenServer	Citrix	1.22x	1.22x	0.87x	0.99x
Hyper-V	Microsoft	1.17x	0.91x	1.31x	0.98
KVM	-	1.00x	1.00x	1.00x	1.00x

**Table 5.** Specifications of the physical server used in this study

Component	Specification	Quantity
Barebone	HP DL380P Generation 8	1
CPU	Intel Xeon Processor E3-2667 v2 (25M Cache, 3.30 GHz)	2
RAM	HP 8GB (1x16GB) Dual Rank x4 PC3-12800R	24
Storage	Ultrastar 10 TB (RAID6 using P420i with 1GB FBWC)	12
Network	Intel Ethernet converged Network Adapter X520-DA2	1

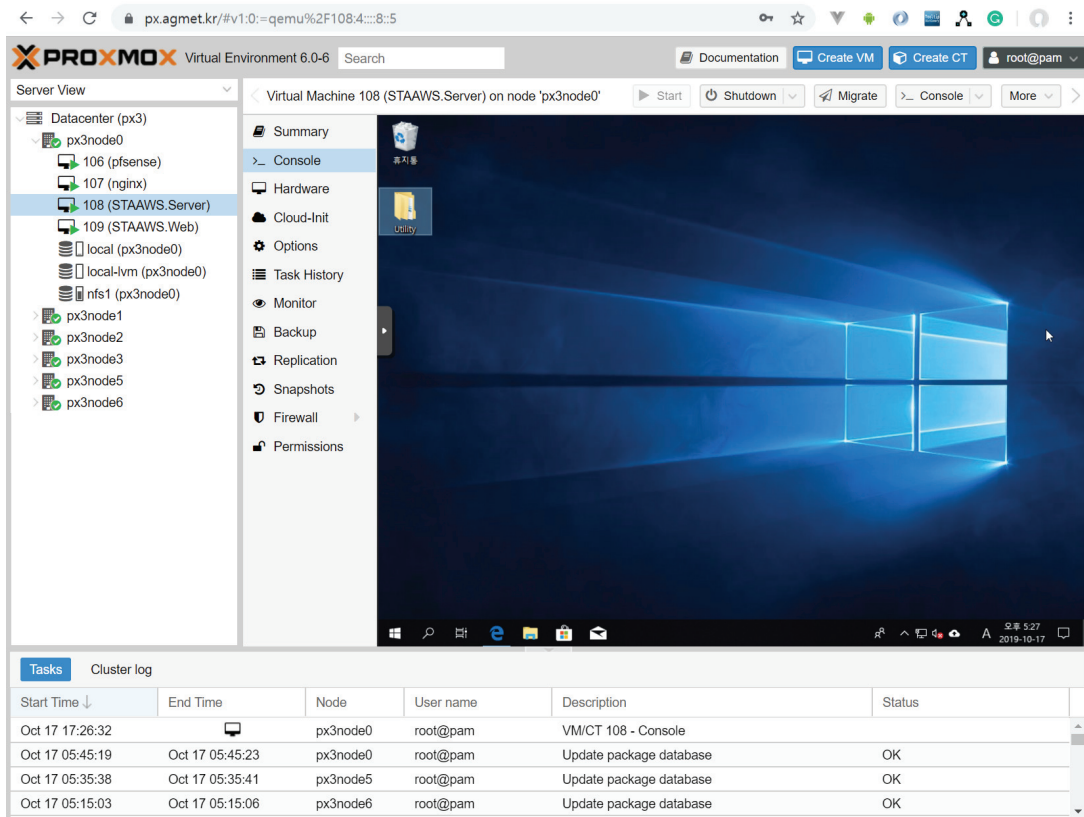
\* Three physical servers were used in this study for on-premises environment.

**Table 6.** List of commands used to measure the performance of CPU, RAM, storage and network using the virtual operating system of the Ubuntu 18.04 version with 1 CPU, 4 GB of RAM and 256 GB of disk

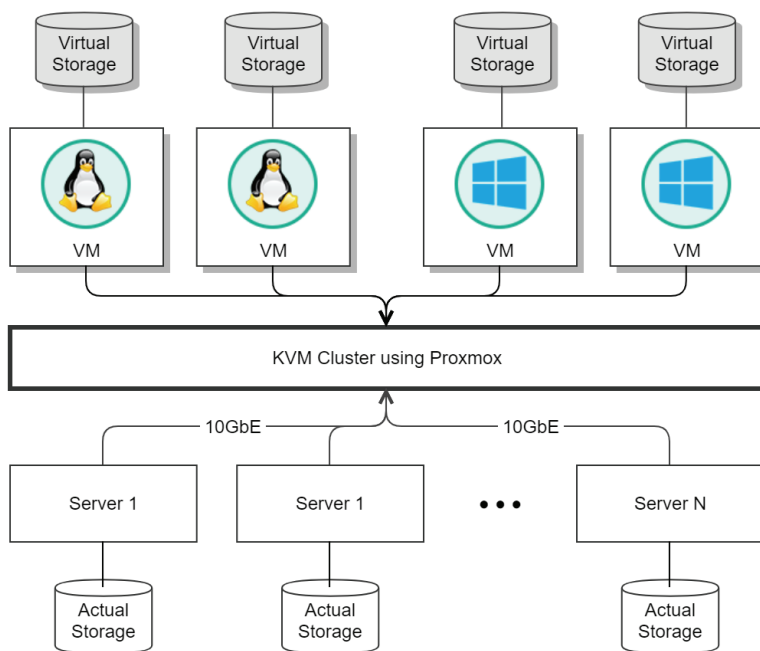
Hypervisor	Command for Measuring Performance on Linux (Ubuntu 18.04)
CPU	sysbench --test=cpu --cpu-max-prime=9999999 --num-threads=1 run
RAM	sysbench --test=memory run
Storage	dd if=/dev/zero of=/dd-test bs=1M count=1000 oflag=dsync
Network	iperf -c iperf-server-ip-address -w 128k

간편하게 경험해 볼 수 있는 장점이 있다. Citrix XenServer는 VMWare vSphere와 더불어 가상화 영역에서 오랫동안 사용되어온 도구이지만 세계 최대 클라우드 서비스 플랫폼인 Amazon AWS가 기존 사용하고 있던 Xen 대신 점차적으로 KVM을 도입함에 따라 향후 전망이 변할 가능성이 있다(ZNNET, 2019). 하지만 여전히 다양한 하드웨어 업체들이 지원하고 있으며 CPU 연산 성능, 메모리 대역폭, 파일 입출력 속도 등은 다른 도구와 비교하여 높은 성능인 것으로 확인된다(Table 4). 또한 Citrix는 기능을 제한하지 않은 XenServer를 무료로 배포하고 향후 추가될 일부

고급 기술과 기술 지원 서비스만 유상 제공하는 라이선스 정책도 병행하고 있으므로 서비스 보증이 필요하면서도 오픈소스 커뮤니티를 이용하고자 하는 경우 장점이 될 수 있다. KVM은 리눅스 커널을 하이퍼바이저로 변환하기 위한 가상화 인프라스트럭처의 하나로써 2007년초 커널 버전 2.6.20의 리눅스 커널 메인 라인에 병합되었다(KVM, 2019). 시스템 운용에 필요한 전반적인 도구를 통합하여 제공하는 상용 도구와는 다르게 KVM은 가상화와 관련된 기본적인 기능만 제공하고 기타 인프라스트럭처 구성에 필요한 사항들은 제 3의 도구들을 이용하여야 한다. 본 연구에 제시된 성



**Fig. 1.** A Screen shot of the Proxmox utility, a hardware-level and operating system-level virtualization tools.



**Fig. 2.** Schematic diagram integrating multiple physical servers into a cluster using Proxmox to run multiple virtual operating systems such as Linux and Windows.

능 측정 자료는 Hyper-V 2019, ESXi 6.7, XenServer 7.6, KVM 5.3.10-1-pve 각각을 동일한 물리 서버에 설치하여 조사하였다(Table 5). 또한 각 하이퍼바이저에 물리 서버 자원의 50%를 할당한 VM을 생성하여 CPU, RAM, Storage, Network의 성능을 측정하였다(Table 6).

오픈소스 가상화 도구 중 PVE (Proxmox Virtual Environment)는 KVM의 기본적 기능 외에 편리한 웹 기반 사용자 인터페이스,고가용성을 위한 클러스터링 (clustering), LXC 기반 가상화 등을 제공한다(Fig. 1). 특히 클러스터링 기능은 여러 대의 서버를 하나의 관리 체제로 묶은 후 각 서버의 운영 상황에 따라 가상의 자원을 동적으로 배치할 수 있도록 하는 기능이다. 특정 물리적 서버의 전원이 꺼지거나 유지보수가 필요하여 운영을 중단할 필요가 있는 경우, 해당 서버에서 구동되는 가상의 운영체제를 적절한 물리 서버에 실시간으로 이주할 수 있다. 이때 해당 운영체제에서 구동 중인 작업에는 아무런 기능적 영향을 주지 않는다(Fig. 2).

#### IV. 온프레미스 환경에서의 운영체제 수준 가상화

하드웨어 수준 가상화는 호스트의 운영체제 커널과 게스트의 운영체제 커널 간의 명령어 번역을 위한 추가적인 절차가 요구되며 이로 인한 성능 상의 오버헤드가 발생하는 단점이 있다. 대안으로 최근에는 운영 체제 수준의 가상화 기술(Operating system level virtualization)이 주목받고 있다(Lee, 2018). 운영 체제 수준 가상화는 운영 체제의 커널이 하나의 사용자 공간 인스턴스가 아닌, 호스트 운영체제의 namespace, cgroup 기능을 이용하여 여러 개의 격리된 사용자 공간 인스턴스를 갖출 수 있도록 하는 서버 가상화 방식이다. 운영 체제 수준 가상화는 가상화된 인스턴스의 모든 명령이 실제 호스트 운영체제의 단일 커널에서 동작하기 때문에 호스트 운영체제의 커널이 손상되면 나머지 가상화 인스턴스들 또한 손상된다는 단점이 있지만, 반대로 단일 커널에서 동작하므로 자원 요구량이 복수 개의 커널을 이용할 때 보다 높지 않은 장점이 있다. 또한 서로 상이한 커널 간의 명령어 번역을 위한 중간 단계가 필요하지 않으므로 일반적으로 네이티브 시스템에 근접한 성능(near-native performance)이 가

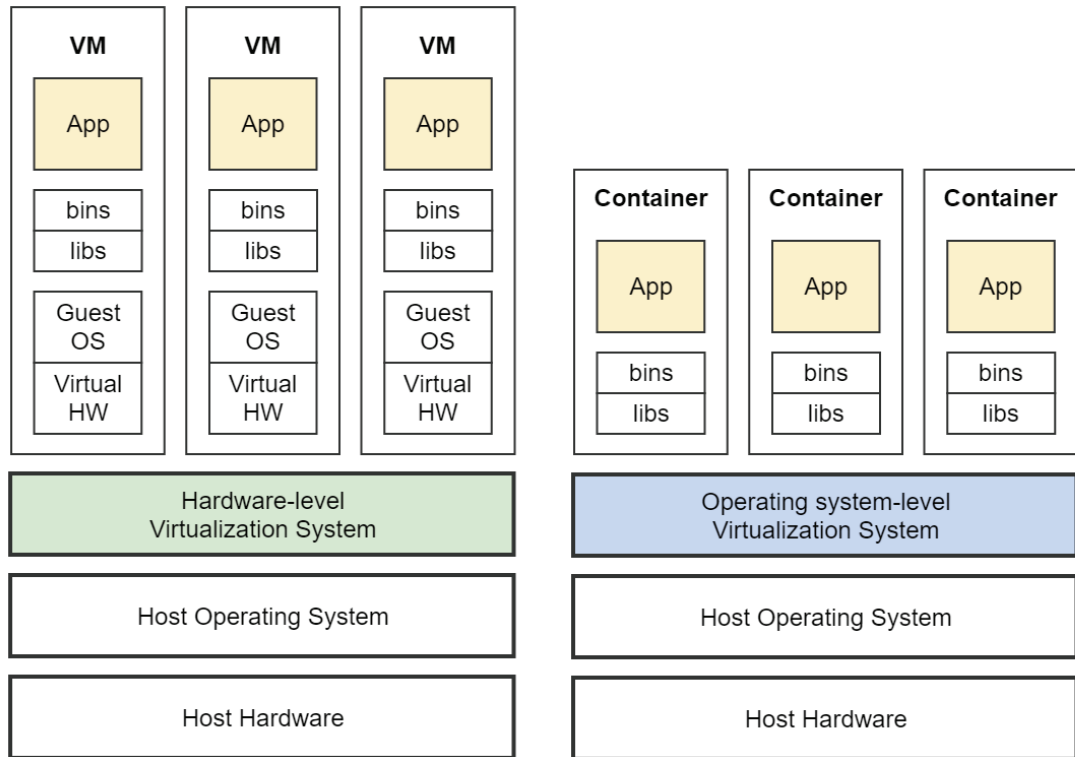


Fig. 3. Comparison illustrations of the hardware-level and the operating system level virtualization.

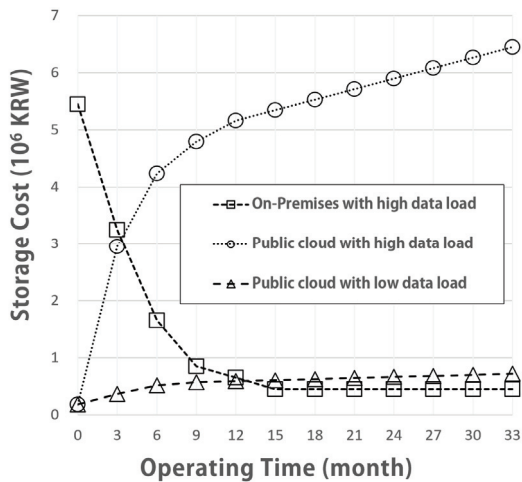
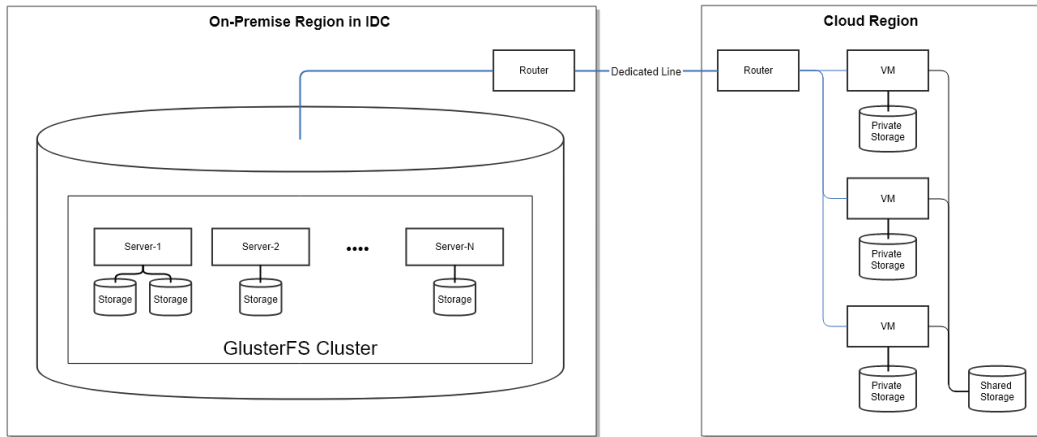


Fig. 4. Changes in storage cost over time for on-premises with high data load, public cloud with high data load and public cloud with low data load.

능하다. 하드웨어 수준 가상화에서는 복수 개의 가상화된 운영체제를 동시에 실행하기 위한 논리적 플랫폼인 가상 머신 관리자(Virtual Machine Manager) 또는

하이퍼바이저(Hypervisor)는 운영체제 상에서 혹은 운영체제 그 자체로서 동작한다. 반면 운영 체제 수준 가상화에서는 가상화된 하드웨어 및 운영체제가 존재하지 않고 호스트의 운영체제 커널 상에서 동작하는 독립적인 라이브러리(libs)와 실행(bin)을 비롯한 환경 정보만을 가지고 있다(Fig. 3). 따라서 상대적으로 높은 성능으로 응용프로그램의 운용이 가능하다 (Table 7).

운영체제 수준 가상화 기술은 OpenVZ, FreeBSD Jail, LXC, Linux-VServer 등이 연구되었고 현재는 Docker 기반의 기술들이 가상화 및 클라우드 컴퓨팅 영역에서 가장 주목받고 있다(Son et al., 2019). Docker에서 컨테이너(Container) 라는 것은 하드웨어 수준 가상화 기술에서의 VM(가상머신)과 대응되는 독립된 격리 공간을 갖는 경량 운영 환경이다. 다른 컨테이너 들과는 완전히 격리된 공간을 갖지만 호스트 운영체제의 커널에 종속되어 있는 특징이 있다. 또한 VM과는 다르게 호스트 운영체제의 파일 시스템에 직접 접근할 수 있어 별다른 백업 과정 없이 컨테이너에서 발생하는 파일 데이터를 호스트와 공유할 수 있다.



**Fig. 5.** Schematic diagram of creating the GlusterFS in an the on-premises system within Internet Data Center, connected to a public cloud system through dedicated line.

따라서 VM 대비 파일을 읽고 쓰기 작업에 대한 낮은 오버헤드가 장점이고, 이는 네트워크 입출력 및 Graphics Processing Unit (GPU)와 같은 자원을 공유하는 것에서도 동일하게 적용된다.

### V. 하이브리드

온프레미스 환경의 경우 초기 투입 비용은 크지만 시간이 지날수록 투입된 비용 대비 이용률이 증가하므로 유지 비용이 감소하는 경향을 보인다. 클라우드 환경의 경우 초기 투입 비용은 작지만 종량제 방식으로 이용 요금에 산정되어 총 유지 비용은 시간에 비례하여 증가한다. 연구에 사용된 농업 생태계 기후 및 이상 기상 변화량 정보 시스템은 초기 5TB와 서버1대 구입을 시작으로 하여 데이터 증가에 따라 디스크를 동적으로 추가하였다(On-Premises with high data load). 그리고 온프레미스와 동일한 저장 공간을 AWS상에 사용하고 있다고 가정하고 자료 저장에 소요되는 비용을 산정하였다(Public cloud with high data load). 마지막으로 데이터 사용량이 적고 누적되지 않는 이상적인 시스템을 AWS상에 사용하고 있을 경우에 대한 비용을 산정하였다(Public cloud with low data load). 데이터가 지속적으로 쌓이는 특성이 있는 시스템을 퍼블릭 클라우드 상에서 운영할 경우 온프레미스와는 다르게 지속적으로 증가하는 것을 볼 수 있다(Fig. 4). 시스템이 장기간 운영됨에 따라 지속적으로 쌓이는 최근의 데이터 외에도 거의 참조되지 않았거나 향후 전혀 참조될 가능성이 매우 낮은 데이터 역시 유지해야 하는

**Table 7.** Average time taken by KVM and Docker to process individual operations for virtualization, and the reduction rate between KVM and Docker on the Debian 9.5 Linux OS

Operation	Virtualization Tool		Reduction Rate (%)
	KVM (sec)	Docker (sec)	
Copy	11.2	2.1	81.3
Booting	25.3	1.9	92.5
Shutdown	10.3	2.1	79.6
Destroy	2.8	1.6	42.9

**Table 8.** Estimated costs charged by CSP for inbound and outbound data of 1GiB in a hybrid system using VPN

CSP	Basic Cost	Additional Cost (KRW/GiB)	
		Inbound	Outbound
Amazon Web Service	42,804 (per 1 connections)	Free	83.2
Microsoft Azure	30,000 (100 Mbps, Up to 10 connections)	Free	134.9
Naver NCP	75,000 (10 Mbps)	Free	100.0
Gabia gCloud	Not supported	Not supported	Not supported

\* 1USD = 1,201 KRW as of Oct. 01. 2019



**Table 9.** Basic and additional costs charged by CSP for connecting on-premises environment to cloud environment using dedicated line

CSP	Service Name	Basic Cost (KRW/1month)	Additional Cost for outbound traffic only (KRW/1GiB)
Amazon Web Service	Direct Connect	282,506	48.8
Microsoft Azure	ExpressRoute	490,347	28.1
Naver NCP	Hybrid Cloud Hosting	Not published	Not published
Gabia gCloud	Hybrid Connect	Free	Free

\* 1USD = 1,201 KRW as of Oct. 01. 2019

점은 비용 효율을 저해하는 원인이다. 퍼블릭 클라우드 상에서 사용 후 즉시 반환하거나 사용량에 따른 경매 방식으로 비용을 최소화할 수 있는 연산 장치, 메모리 장치 등과는 달리 저장 장치에 보관되는 파일은 한 번 점유를 하게 되면 삭제하기 전까지 지속적으로 비용을 발생시킨다. 파일의 수명 주기 관리 기능을 통해 Hot, Warm, Cold 방식으로 데이터의 접근 방식, 성능에 따라 차등적 비용을 적용하지만 고해상도의 GIS 자료를 보관하기에는 여전히 부담이 크다(Table 2). 이는 지속적으로 누적되는 대량의 데이터를 시스템의 전체 수명 주기 동안 유지할 필요가 있는 응용 시스템의 경우 클라우드의 다양한 편의 기능에도 불구하고 도입을 쉽게 결정하기 어려운 이유이다.

클라우드와 온프레미스를 상호 연계하는 하이브리드를 가장 쉽게 구축할 수 있는 방법은 각 CSP들이 제공하는 VPN (Virtual Private Network) 기능을 이용하는 것이다. VPN은 인터넷과 같은 공중망을 이용하여 두 지점간 연결을 하더라도 전용선으로 연결된 것과 같이 사용할 수 있게 하는 기술로서 전용선보다 훨씬 저렴한 비용을 특징으로 한다. VPN을 이용하여 클라우드 네트워크와 온프레미스 네트워크를 연결하여 상호간의 위치 투명성을 제공할 수 있다. 클라우드에서 VM을 생성하거나 네트워크 파일 시스템을 공유하면 온프레미스 환경에서도 해당 자원에 접근할 수 있다. 하지만 온프레미스에서 클라우드로 들어가는 네트워크 트래픽(Inbound traffic)에 대해서는 요금이 부과되지 않지만, 클라우드에서 온프레미스로 나가는 네트워크 트래픽(Outbound traffic)에 대해서는 상당량의 비용을 부과하므로 데이터 이동량을 고려하여 설계해야 한다(Table 8).

VPN 방식으로 대량의 데이터를 송수신 하는 것은 비용의 문제와 더불어 IP 프로토콜 처리와 데이터의

암호화/복호화에 따른 오버헤드, 공용 인터넷 회선을 사용함에 불규칙한 지연으로 인해 충분한 대역폭을 보장받을 수 없는 문제가 있다. 각 CSP 마다 VPN 방식 외에도 온프레미스 시스템과의 연계를 위해 유사한 관련 서비스를 제공한다(Table 9). CSP의 클라우드 환경을 이용하면서도 지정된 Internet Data Center (IDC)에 물리적 서버를 배치하고, 해당 IDC와 해당 클라우드 간에 이미 연결되어 있는 전용회선을 임대하여 하이브리드 환경을 구축할 수 있다. 본 연구에서는 IDC에 있는 온프레미스 영역에 Gluster File System (GlusterFS)을 구축하여 국내 클라우드 서비스 제공자인 Gambia의 gCloud와 연결하여 테스트를 진행하였다(Fig. 5). GlusterFS는 페타바이트(petabyte) 규모의 확장성 있는 NAS 파일 시스템으로서 다수의 스토리지를 이더넷(Ethernet) 또는 인피니밴드(Infiniband)와 같은 네트워크를 통하여 하나의 커다란 병렬 네트워크 파일 시스템으로 통합하는 도구이다. 이 연구에서는 GlusterFS 구축에는 3대의 물리적 서버를 10Gb 이더넷으로 연결하였다. 각 서버는 평균 약 50 TiB의 저장 공간을 갖고 있고, 1대의 서버를 복제(replica)를 위해 할당하여 약 100 TiB로 구성되었다. 단일 파일의 쓰기 성능은 로컬 파일 시스템의 경우 약 910 MB/s 로 측정되었고 GlusterFS의 경우 673 MB/s 로 측정되었다. 네트워크를 이용하여 파일을 분산 저장함으로써 발생하는 오버헤드로 인해 약 26%의 성능 손실이 있지만 여러 클라이언트가 동시에 입출력 요청을 할 경우 서버 수에 비례하여 성능이 향상되는 장점이 있다. 한편 페타 바이트 규모의 대용량 자료를 클라우드 내부에서 외부로 이동해야 할 요구가 발생할 경우 시간 및 비용적인 측면에서 큰 위협이 될 수 있으므로 대용량 자료의 클라우드 내부 저장은 신중할 필요가 있다.

## VI. 결 론

고해상도의 농업 기후 자료를 처리하는 기반 시스템을 구축하는데 있어 고려될 수 있는 4가지 구성(클라우드, 온프레미스-하드웨어 기반 가상화, 온프레미스-소프트웨어 기반 가상화, 하이브리드)에 대해 알아보고 기존에 구축된 농업기후변화량시스템을 일부 적용하여 성능 및 비용에 대한 효용성을 평가하였다. 다양한 기능, 높은 안정성 및 보안, 낮은 초기 투입 비용, 편리한 접근성 등을 고려한 결과 퍼블릭 클라우드가 가장 좋은 선택인 것으로 확인된다. 하지만 대량의 데이터가 지속적으로 누적됨에 따른 비용 측면을 고려할 경우 다른 구성 방법을 고려하는 것도 필요하다. 특히 저장되는 데이터의 용량 대비 접근 횟수가 적을 경우 클라우드 상의 데이터 저장은 비용 효율이 낮다. 클라우드 서비스 제공 업체들은 비용 효율을 높이기 위한 몇 가지 대안을 제시하지만 데이터의 접근이 빈번할 경우 이 또한 비용 증가의 원인이다. 이에 대해 CSP에서 제공하는 전용선 기능을 이용하여 하이브리드 방식을 이용하여 비용적인 측면에서의 대안이 될 수 있음을 확인하였다. 다만 온프레미스를 위한 하드웨어 및 소프트웨어 인프라를 직접 관리해야 할 경우 이에 대한 충분한 운영 지식과 경험이 요구되며 위험 관리 정책, 재해 복구 정책 보안 정책 등이 고려되어야 한다. 따라서 관련 기술이나 인력을 보유하고 있다면 하이브리드 환경을 구축하는 것이 유리할 것이다. 가변적인 전기료, 공간 사용료, 인건비 그리고 장비비 등을 대략적으로 추정하여 비용적인 측면에서의 일반적인 차이를 제시하는 것에는 어려움이 있다. 따라서 본 연구에서 측정된 수치들은 경험적으로 얻은 수치이므로 주어진 환경에 따라 차이가 있을 수 있으나 4년 이상의 운영 경험을 기반으로 측정한 결과이므로 일반적인 고해상도 농업 기후 자료의 처리 시스템의 구성 방법에도 충분한 도움이 될 수 있을 것이다. 클라우드, 온프레미스, 하이브리드 등과 같은 시스템 운영 환경은 비용적, 기술적 그리고 정책적인 면에서 언제든지 변화할 수 있다. 따라서 구축하고자 하는 시스템의 특성을 고려하여 클라우드와 온프레미스 간의 균형을 유지하고 성능과 비용 효율을 위해 관련 기술의 변화를 지속적으로 수용할 수 있는 전산 운용 체계를 구축하는 것이 연구 개발과 실서비스의 효율성을 높일 수 있다.

## 적 요

GIS 기반의 농업 기후 자료의 처리 및 분석 체계의 유용성은 클라우드, 온프레미스, 하이브리드 구조와 같은 컴퓨팅 인프라의 신뢰성, 가용성에 영향을 받는다. 현재는 정보 기술 산업에서 클라우드 컴퓨팅의 시대라고 할 수 있을 만큼 클라우드와 관련된 기술이 확산되어 있으나, 장기간의 운영 경험으로 누적된 다양한 참조 사례를 볼 때 온프레미스 기술이 클라우드 기술 보다 유리한 경우도 있다. 또한 클라우드 환경의 경우 초기 비용이 온프레미스와 비교하여 저렴하지만 사용 방법에 따라 매우 높은 비용이 부과될 가능성이 있다. 따라서 각 시스템의 특성에 맞는 적절한 구성법이 고려될 필요가 있다. 본 연구에서는 농업 기후 자료 처리 및 분석 체계에 이용가능한 일반적인 컴퓨팅 플랫폼 4개를 소개하고 대량의 자료 처리 및 저장의 특성을 갖는 응용 시스템을 적용하여 각 플랫폼의 장단점을 비교 분석하였다. 현재로서는 대량의 농업 기상 및 기후 데이터를 필요로 하는 시스템은 비용상의 이유로 퍼블릭 클라우드로의 이주가 불가능함을 확인하였다. 향후 참조될 가능성이 높지 않은 대용량 자료를 클라우드 상에 유지해야 하는 점이 주요 원인이다. 따라서 가장 높은 비용의 저장 및 백업 부분을 클라우드 대신 온프레미스에서 운용하고, 자료의 분석 및 처리 그리고 표출 부분과 같이 유연성이 요구되는 부분은 클라우드에서 운용하는 것이 합리적이다.

## 감사의 글

본 연구는 농촌진흥청 국립농업과학원 농업과학기술 연구개발사업(과제번호: PJ0122932019)의 지원에 의해 수행되었음.

## REFERENCES

- AGMET, 2019: [https://www.agmet.kr\(2019.12.11\)](https://www.agmet.kr(2019.12.11))  
 AGECOCLIM, 2019: [https://agecoclim.agmet.kr\(2019.12.11\)](https://agecoclim.agmet.kr(2019.12.11))  
 Armbrust, M., A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, 2009: Above the clouds: A berkeley view of cloud computing. *Department Electrical Engineering and*

- Computer Sciences, University of California, Berkeley, Rep. UCB/EECS 28(13).*
- Choi, Y. S., K. T. Kim, and J. H. Lee, 2008: Development of grid based distributed rainfall-runoff model with finite volume method. *Journal of Korea Water Resources Association* **41**(9), 895-905.
- CITRIX, 2019: [https://www.citrix.com/downloads/citrix-hypervisor/\(2019.10.01\)](https://www.citrix.com/downloads/citrix-hypervisor/(2019.10.01))
- Kim, D. J., and J. I. Yun, 2013: Improving usage of the Korea Meteorological Administration's digital forecasts in agriculture: 2. Refining the distribution of precipitation amount. *Korean Journal of Agricultural and Forest Meteorology* **15**(3), 171-177.
- Kim, C. Y., Y. C. Kim, Y. C. Kim, S. M. Lee, Y. K. Kim, and D. H. Seo, 2014: Performance enhancement of distributed file system as virtual desktop storage using client side SSD cache. *Korea Information Processing Society* **3**(12), 433-442.
- Kim, K. S., 2011: Impact assessment of climate change by using cloud computing. *Korean Journal of Agricultural and Forest Meteorology* **13**(2), 101-108.
- KMA, 2019: [http://www.kma.go.kr\(2019.10.01\)](http://www.kma.go.kr(2019.10.01))
- KVM, 2019: [https://linux-kvm.org\(2019.12.10\)](https://linux-kvm.org(2019.12.10))
- Lee, K. P., 2018: Where would you build your big data analytics platform? - Comparative analysis between cloud and On-Premise. *Journal of Research Methodology* **3**(2), 101-117.
- MICROSOFT, 2019: [https://www.microsoft.com/en-us/evalcenter/evaluate-hyper-v-server-2019\(2019.10.01\)](https://www.microsoft.com/en-us/evalcenter/evaluate-hyper-v-server-2019(2019.10.01))
- Park, J. H., Y. S. Shin, S. K. Kim, W. S. Kang, Y. K. Han, J. H. Kim, D. J. Kim, S. O. Kim, K. M. Shim, and E. W. Park, 2017: Speed-up techniques for high-resolution grid data processing in the early warning system for agrometeorological disaster. *Korean Journal of Agricultural and Forest Meteorology* **19**(3) 153-163.
- PROXMOX, 2019: [https://www.proxmox.com/en/proxmox-ve\(2019.10.01\)](https://www.proxmox.com/en/proxmox-ve(2019.10.01))
- Shin, Y. S., J. H. Park, S. K. Kim, W. S. Kang, K. M. Shim, and E. W. Park, 2015: An operational site-specific early warning of weather hazards for farmers and extension workers in a mountainous watershed. *Korean Journal of Agricultural and Forest Meteorology* **17**(4) 290-305.
- Son, S. S., J. A. Park, and Lee, K., Y., 2019: Technology trend to enable cloud computing services in HPC environments. *Communications of the Korean Institute of Information Scientists and Engineers* **37**(10), 17-24.
- VIRTUALIZATION, 2019: [https://en.wikipedia.org/wiki/Virtualization\(2019.10.01\)](https://en.wikipedia.org/wiki/Virtualization(2019.10.01))
- VMWARE, 2019: [https://www.vmware.com/kr/products/esxi-and-esx.html\(2019.10.01.\)](https://www.vmware.com/kr/products/esxi-and-esx.html(2019.10.01.))
- Won, M. S., K. C. Jang, and S. H. Yoon, 2018: Development of the national integrated Daily Weather Index (DWI) model to calculate forest fire danger rating in the spring and fall. *Korean Journal of Agricultural and Forest Meteorology* **20**(4), 348-356.
- Yun, J. I., 2015: A feasibility study of a field-specific weather service for small-scale farms in a topographically complex watershed. *Korean Journal of Agricultural and Forest Meteorology* **17**(4) 317-325.
- ZDNET, 2019: [http://www.zdnet.co.kr/view/?no=20171108115328\(2019.10.01.\)](http://www.zdnet.co.kr/view/?no=20171108115328(2019.10.01.))