

오픈소스 클라우드 컴퓨팅 기반 교육 실습 시스템 구축

윤준원*, 박찬열**, 송의성***

요약

근래 이슈가 되고 있는 클라우드 컴퓨팅은 분산컴퓨팅 환경에서 사용자가 요구하는 컴퓨팅 자원을 최적화하여 유연하고 확장성 있게 지원할 수 있어 각광받는 새로운 패러다임이다. 클라우드 컴퓨팅 환경은 가상화 환경을 구성함으로써 실제적인 구현 및 서비스가 가능해진다.

본 논문에서는 오픈소스 기반의 클라우드 컴퓨팅을 연구하고 대학교 내에서 컴퓨터를 이용한 실습을 수행할 시 요구되는 시스템 환경을 오픈소스 클라우드 컴퓨팅 기반의 환경을 통해 구현하고자 한다. 클라우드 컴퓨팅을 통한 가상화 기반의 실습 환경은 최적화된 자원을 제공할 수 있을 뿐만 아니라 실습 자원 관리의 편리성, 실습 결과에 대한 손쉬운 관리 등의 효율성을 가져올 수 있다. 이로 인해 실습환경 설정에 소요되는 시간을 줄일 수 있을 뿐만 아니라, 교수 입장에서는 실습결과물들을 쉽게 관리 할 수 있게 된다. 또한 다양한 실습환경의 요구사항들을 유연성 있게 적용함으로써 시스템에 대한 활용성 또한 높아지게 된다.

키워드 : 클라우드컴퓨팅, 가상화, 오픈소스

Building the Educational Practice System based on Open Source Cloud Computing

JunWeon Yoon*, ChanYeol Park**, Ui-Sung Song***

Abstract

Recently, cloud computing is being emerged paradigm that a support computing resource flexible and scalable to users as the want in distributed computing environment. Actually, cloud computing can be implemented and provided by virtualization technology.

In this paper, we studied open source based cloud computing and built a educational practice system through cloud computing. Virtualization-based cloud computing provides optimized computing resources, as well as easy to manage practical resource and result. Therefore, we can save the time for configuration of practice environment. In the view of faculty, they can easily handle the practice result. Also, those practice condition reuse comfortably and apply to various configuration simply. And then we can increase capabilities and availabilities of limited resources.

Keywords : Cloud Computing, Virtualization, Open Source

1. 서론

클라우드 컴퓨팅은 가상화라는 기술을 통해 실제적인 구현 및 서비스가 가능해지며 사용자가 요구하는 컴퓨팅 환경을 유연하게 구성할 수 있다[1]. 가상화 기술은 x86 서버에 멀티코어가 일반화되고 집적도가 증가하면서 성능은 점차 증가하였고, 더불어 서버 가상화를 제공하기 위한 오픈소스 또는 상용 프로그램들의 안정성 및 성능에 대한 효용성이 검증됨에 따라 이를 이용한 클라우드 컴퓨팅 활용이 점차 두각을 나타내고 있다. 가상화 기반 클라우드 컴퓨팅의 대표적인

※ 교신저자(Corresponding Author): Ui-Sung Song
접수일:2013년 11월 28일, 수정일:2013년 12월 13일
완료일:2013년 12월 24일

* KISTI 국가슈퍼컴퓨팅연구소 선임연구원
Tel: +82-42-869-0581, Fax: +82-42-869-0569
jwoon@kisti.re.kr

** KISTI 국가슈퍼컴퓨팅연구소 책임연구원

*** 부산교육대학교 컴퓨터교육과
▣ 이 논문은 2013년도 부산교육대학교 교육연구원의 지원을 받아 연구되었음

장점은 분산컴퓨팅 환경에서 사용자가 요구하는 컴퓨팅 자원을 사용자의 요구사항에 맞게 최적화하여 유연하고 확장성 있게 지원할 수 있다는 점이다[2]. 이런 특징을 활용하여 대학 내에서 수행되는 실습환경을 가상화를 통해 구성함으로써 학생들이 수행해야할 시스템 환경을 최적화하여 제공할 수 있을 뿐만 아니라 실습을 진행 관리하는 교수 입장에서 쉽고 투명하게 실습 결과물들을 확인할 수 있다. 또한 다양한 실습 환경을 이미지로 구성하여 사용함으로써 한정된 자원의 활용성을 최대화 할 수 있다.

본 논문에서는 클라우드 컴퓨팅 기반에서 가상화를 통해 최적화 된 실습 환경을 구축하기 위하여 오픈 소스 기반의 SW를 활용하여 구축할 예정이다. 이를 위해 가상화 기술 및 가상 머신을 관리하는 소프트웨어들에 대해서 조사하였다. 이를 토대로 리눅스 기반 OS상에서 가상 머신들을 생성하고, 가상머신(VM: Virtual Machine)을 관리하는 오픈소스 가령, Eucalyptus, OpenNebula, Nimbus, OpenStack, Abiquo 등과 같은 IaaS (Infrastructure as a Services) 기반의 SW를 설치할 예정이다. 가장 큰 차별점은 가상 머신 위에서 운영하게 될 가상 이미지를 실습환경에 맞도록 최적화하여 제공하는 가상 어플라이언스(Virtual Appliance)로 실습 관리자가 제공하고자 하는 환경(SW, 환경설정)을 이미지로 만들어 복제 사용함으로써 시스템 설정 및 관리가 매우 용이하다. 또한 다양한 실습환경에 맞게 가상 어플라이언스를 생성함으로써 자원 활용성 및 재사용성을 최대화할 수 있다[3].

2. 관련 연구

클라우드 컴퓨팅의 핵심 기술은 가상화 자원을 사용하여 문제를 해결하는 데 있다. 이런 가상머신(Virtual Machine: VM)을 어떻게 구성하고 관리하는가가 가장 큰 이슈 중에 하나이다. 현재의 슈퍼컴퓨터 사용자는 대규모의 시뮬레이션을 수행하는 일부의 응용 연구자들에 국한되어 있을 뿐 아니라 기존의 컴퓨팅 수요 외에도 다양한 형태의 컴퓨팅 수요를 이끌어내기 위한 컴퓨팅 서비스의 다변화가 요구되고 있다[4].

클라우드 컴퓨팅은 다수의 상이한 컴퓨팅 자

원들을 가상화하여 사용자에게 온디맨드 방식으로 제공하는 기술로 자원 효율성의 극대화와 관리비용 최소화라는 장점으로 인터넷 환경 최신 트렌드로 부상하고 있다. 다음은 클라우드의 핵심 기술을 확산시키는데 주도적인 역할을 수행했던 공개 SW, 즉 오픈소스 기반의 IaaS (Infrastructure as a Service) 솔루션 들이다.

▪ Nimbus

Nimbus는 오픈소스 클라우드 툴킷으로 사용자에게 IaaS 클라우드 서비스를 제공한다. 이 프로젝트는 시카고대-아르곤(UChicago Argonne LLC)이 운영하는 아르곤 국립연구소 Kath Keahy 연구팀에 의해서 개발되었으며, 과학자에게 가상머신을 사용하여 방대한 데이터를 처리하는 오픈 소스 클라우드 컴퓨팅 인프라이다. 또한 Nimbus는 Amazon의 EC2와 유사하게 사용자들이 가상머신을 사용하도록 하는 클라우드 컴퓨팅 구조이고, 여러 개의 가상머신들을 가상 클러스터로 결합하여 사용할 수 있는 Nimbus Context Broker와 같은 사용자 레벨의 툴을 제공한다. Nimbus는 Globus.org의 일부인 Workspace Service 프로젝트의 변형이며, Amazon EC2를 이용하여 Xen과 KVM을 지원한다[5].

▪ Eucalyptus

Eucalyptus(Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems)는 리눅스 기반 인프라스트럭처 상에서 IaaS 클라우드 컴퓨팅 서비스를 제공하는 오픈소스 기반 소프트웨어 플랫폼으로 2007년 캘리포니아 대학의 Computing Science에서 연구 프로젝트로 시작하였다.[6] 2007년 VGrADS(Virtual Grid Application Development Software) 프로젝트에서 LEAD(Linked Environment for Atmospheric Discovery) 기후 예측 시스템을 개발 중 LEAD에서 생성되는 대규모의 워크플로를 수행하기 위해 클라우드 환경의 필요성이 제기되었다. 이에 아마존 AWS가 고려되었으며 VGrADS와 AWS 연동을 위한 로컬 클라우드 플랫폼으로 Eucalyptus가 개발되었다. Eucalyptus는 Hypervisor를 위해 KVM 및 Xen을 지원하며 클러스터 관리를 위한 Rocks 클러스터 배포판을 포함하고 있다[7].

▪ OpenNebula

OpenNebula는 Universidad Complutense de Madrid의 DSA-research Group에서 시작된 오

오픈 소스 기반의 IaaS Management 툴킷으로 물리적 자원들 안에서 가상머신을 동적으로 구성하고 제한당하는 오픈 소스 가상 인프라스트럭처 엔진(open source virtual infrastructure engine)이다. private, public, hybrid cloud를 구축할 수 있으며 Amazon AWS, OGC OCCL, VMware vCloud와 연동이 가능한 공통 인터페이스를 제공한다. Open Nebula는 다양한 네트워킹 및 스토리지 솔루션과 통합될 수 있을 뿐만 아니라 현존하는 데이터센터와 손쉽게 통합되도록 설계되었다. OpenNebula는 Xen, KVM/Linux 및 VMware를 지원하며 libvirt와 같은 요소를 사용하여 관리 및 검사 작업을 수행한다[10].

과학기술 분야의 실험 및 연구 장비의 발전으로 대용량의 데이터들이 생성되면서 이를 아카이브하고, 분석하기 위한요구사항들이 증가하고 있다. 이런 다양한 과학 응용 분석을 위해 클라우드 컴퓨팅은 맞춤형 자원제공서비스를 제공할 뿐만 아니라 유연한 확장성을 가지고 있어 이를 이용한 다양한 연구들이 진행되고 있다[8].

▪ **OpenStack**

OpenStack은 Public과 Private 클라우드를 구축하기 위한 오픈 소스 software로 서버, 스토리지, 네트워크, 가상화 기술들과 같은 리소스들을 모으고, 이들을 제어하고 운영하기 위한 Cloud Operating System이다. 또한, OpenStack은 오픈소스 기반으로 클라우드를 구축하고 운용하고자 하는 오픈소스 개발자, 회사, 사용자들로 이루어진 커뮤니티이다. 오픈스택은 크게 3가지 컴포넌트로 구성된다.

- Swift: Amazon S3와 같은 object/blob 저장소.
- Glance: VM 이미지, 스토리지와 같은 클라우드 자원을 검색하기 위한 컴포넌트
- Nova: Amazon EC2와 비슷하다. 컴퓨팅자원을 할당해서 VM을 만들기 위한 컴포넌트

이들은 각각 독립된 프로젝트로 이들이 서로 협력해서 클라우드 인프라를 구축한다. 클라우드 스택은 하나의 프로젝트로 하나의 통합된 애플리케이션 형태로 제공된다. 이런 형태는 패키징과 배포, 설계가 쉽다는 장점이 있다. 실제 클라우드스택은 쓸 만한 UI와 안정된 기능, 무엇보다 간단하게 설치할 수 있다. 반면 확장이 쉽지가 않는 문제가 있다. 각 모듈들이 같은 도메인에서 서로서로 단단하게 결합되어 있기 때문이다.

반면 오픈스택은 각 컴포넌트들이 매우 느슨하게 연결 되어 있어서, 확장 개발이 용이하다는 장점을 가진다. 물론 느슨하게 연결된 컴포넌트들을 통합해서 하나의 제품으로 만드는 까다로운 작업이 있다. 오픈스택은 클라우드스택에 비해서 통일된 패키지를 만들기 쉽지가 않고, 설치도 매우 힘든 것으로 알고 있다[9].

▪ **Amazon AWS**

Amazon Web Services(AWS)는 아마존에서 제공하는 웹서비스를 통칭하며, 웹 서비스 인터페이스를 통해 사용자가 원하는 파워의 규모에 맞게 자원과 시스템 환경을 구성할 수 있다. AWS 서비스로는 EC2, EBS, S3, SQS, Mturk, Alexa, A2S, FPS, DevPay, SimpleDB, CloudFront 등이 다양한 솔루션을 제공하고 있으며 지속적으로 새로운 서비스들을 추가하고 있다. 아마존 AWS 플랫폼 구조는 Xen기반의 EC2 호스트, EC2 매니저, Amazon Simple Storage, S3 관리자로 구성되어 있으며, SOAP과 REST 기반의 AWS 인터페이스를 제공한다. 2010년 7월에 HPC 응용을 위한 Amazon EC2 기반의 계산 클러스터 제공 서비스를 시작했으며 그 외에도 다양한 성능의 인스턴스를 추가 확장하여 제공하고 있다. 클러스터 컴퓨팅, 클러스터 GPU 및 고용량 메모리 클러스터 인스턴스는 고성능 네트워크 기능을 제공하도록 특별히 설계되었으며 프로그래밍 방식을 통해 클러스터에 실행할 수 있으므로 긴밀하게 연결된 노드 간 통신에 필요한 저지연 네트워크 성능을 애플리케이션에 제공할 수 있다[10].

위와 같이 다양한 컴퓨팅 환경에서, 클라우드 컴퓨팅이 가져올 변화에 대해서 살펴보면 독립된 계산 자원을 활용하기 보다는, 이미 존재하는 자원들을 상황과 목적에 맞게 최적화하여 이용할 수 있다는 것이다. 또한 일시적으로 특정 목적에 최적화된 계산 자원을 유연하게 설계하여 제공할 수 있어 다양한 수행 특성에 맞는 컴퓨팅 환경을 유동적으로 제공하여 지원할 수 있게 된다.

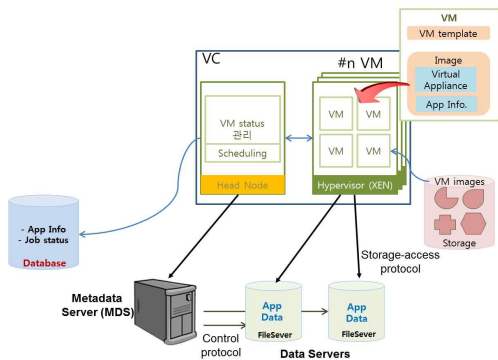
3. 클라우드 기반 실습환경 구축

3.1 가상화 실습 환경 요구사항 분석

클라우드 기반의 교육 실습 환경을 구축하기 위해서는 다음과 같은 기능들이 요구된다.

- ① 다양한 교육용 실습환경이 가상머신을 통해서 구현되기 위해서는 이미지 형태의 가상 어플라이언스들(Virtual Appliance)을 생성하여 저장하게 된다. 가상 어플라이언스는 시스템의 요구사항을 담은 일종의 컨테이너로 의존성이 있는 어플리케이션, 라이브러리, 서비스, 관련 데이터 그리고 운영체제까지 통합된 일종의 소프트웨어 스택이다.
- ② 실습환경이 구동되는 가상머신을 관리하기 위해서는 전체적인 가상 자원의 생성, 초기화, 동작, 대기, 회수 등을 일괄적으로 수행할 수 있는 도구가 필요하다. 관련연구에서 논하였던 오픈소스 기반의 IaaS 도구인 OpenNebula를 사용하여 가상머신들을 제어하였다. OpenNebula는 Xen, KVM, VMware를 지원하며 libvirt와 같은 요소를 사용하여 관리 및 검사 작업을 수행할 수 있다. 본 환경에서는 Apache 라이선스 기반의 유연성과 확장성을 가지는 오픈 소스 툴킷의 특징을 가지고 있어 채택하였다.
- ③ 실습환경이 구성되면 가상 환경으로 제공되는 실습 데이터와 이를 가공한 산출물 데이터를 관리하기 위한 공용의 파일시스템이 필요하다. 또한 실습환경의 규모가 커지게 되면 파일 I/O에 대한 부하가 높아져 실습 데이터를 받거나 산출을 제출할 때 문제가 발생할 수 있다. 이를 위해 병렬파일 시스템에 대한 고려가 필요하다.

(그림 1) 가상화 기반 실습 환경 구성도

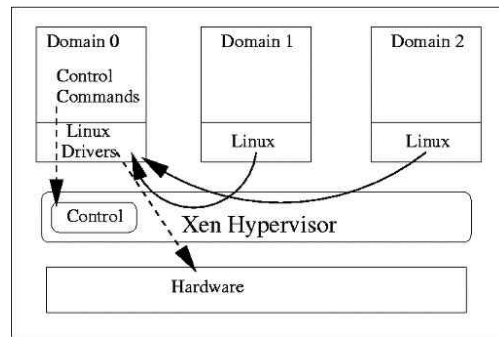


(Figure 1) Practice System based Virtualization

3.2 오픈소스 기반 클라우드 환경 구축

위에서 논한 요구사항을 바탕으로 해당되는 기능들을 오픈소스 소프트웨어를 이용하여 구성하였다. 첫 번째로 논한 가상 어플라이언스를 구성하기 위한 하이퍼바이저(Hypervisor) 즉, 가상머신(VM)으로 Xen을 사용하였다. 대표적인 오픈소스 기반의 가상머신으로써는 Xen, KVM, 등이 있다. Xen은 반가상화(Para Virtualization)라는 게스트 운영체제 일부를 수정하는 방식으로 동작하며 게스트 운영체제는 특권 명령을 실행하려면 하이퍼콜(Hypercall)로 하이퍼바이저에게 서비스를 요청하는 해야 한다. 운영체제에서 어플리케이션이 커널에게 시스템 콜(System Call)로 서비스를 요청하는 방식과 동일하다.

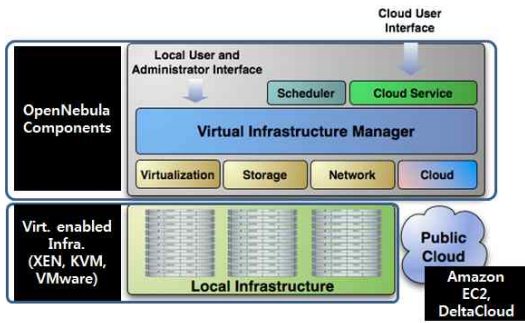
(그림 2) Xen 아키텍처



(Figure 2) Xen architecture

두 번째로 실습환경이 구동될 복수개의 가상 환경(VM)을 관리하기 위해서는 IaaS기반의 오픈소스가 요구된다. 관련연구에서 논하였듯이 대표적인 오픈소스 IaaS 소프트웨어로서는 Nimbus, OpenNebula, OpenStack, Eucalyptus 등이 있으며[11], 본 실험에는 OpenNebula를 설치하여 VM을 관리하였다. 이 툴을 사용하여 VM들을 동작, 종료, 대기, 리셋 시킬 수 있으며 장애가 발생할 시 마이그레이션 기능도 제공된다. 나아가 아마존 E2와 호환되는 API도 구현되어 있다. (그림 3)은 OpenNebula 구성도를 나타내고 있다.

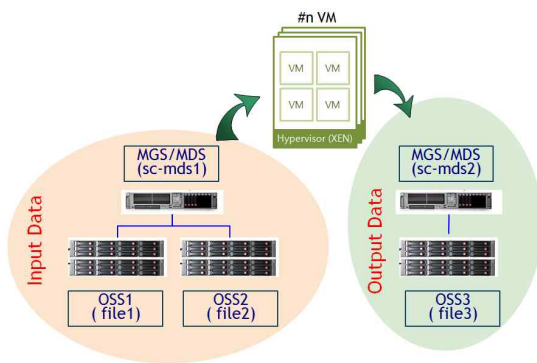
(그림 3) OpenNebula 아키텍처



(Figure 3) OpenNebula architecture

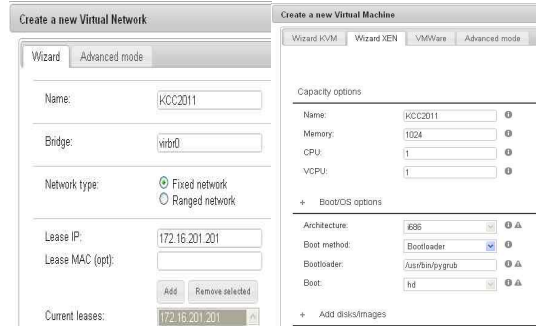
세 번째로 클라우드 기반의 실습환경이 단일 구성에서 벗어나 확장성이 있는 환경으로 확대될 때 발생할 수 있는 데이터 I/O에 대한 요구가 해결되어야 한다. 이를 위해서 대용량 데이터 파일을 처리하기 위한 분산 병렬 파일 시스템 이용하여 데이터를 처리하는 기술이 필요하다. 관련한 SW들로 Lustre, GFS, Dyrad, pNFS 방식들이 있다. 본 논문에서는 실습환경 구축을 위한 파일시스템으로 Lustre를 설치하였다. Lustre는 오픈소스 기반 고성능 클러스터 파일시스템으로 기존의 NFS와 같은 분산 파일시스템에 Object-based Storage Disk 기술을 적용함으로써 성능, 가용성, 확장성 문제를 개선할 수 있다[12].

(그림 4) 가상화 실습 환경 수행을 위한 Lustre 파일시스템 구축



(Figure 4) Lustre filesystem for practice environments

(그림 5) OpenNebula 가상 네트워크 할당 및 VM 생성 화면

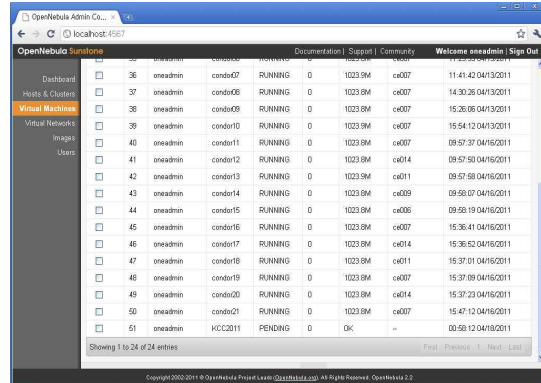


(Figure 5) Set up of virtual network and VM creation on OpenNebula

본 실습환경 구축을 위해서 8코어로 구성된 3대의 노드에서 소규모의 테스트를 수행하였다. 우선 VM은 리눅스 기반(CentOS) 이미지로 생성하였으며 (그림 5)와 같이 오픈소스 기반의 IaaS 도구인 OpenNebula를 이용하여 VM을 생성하였다. 생성된 VM들은 (그림 6)과 같이 모니터링 화면을 통해 상태정보를 파악할 수 있다.

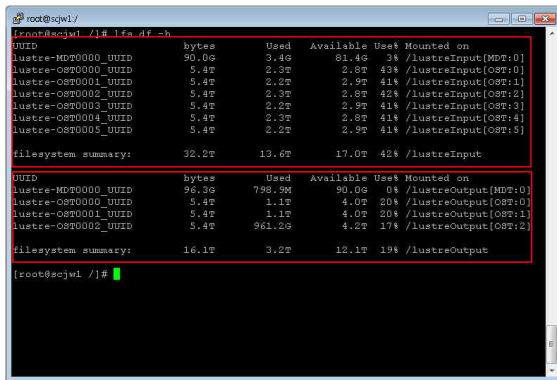
(그림 7)은 Lustre 파일시스템의 설정된 화면이다. 실습 산출물에 대한 보안을 위해 입력 데이터와 출력 데이터를 구분하여 설정하였다. 병렬파일시스템은 데이터를 분산하여 저장하고 가져오면서 발생할 수 있는 파일 I/O에 대한 병목 현상 및 부하 문제를 개선할 수 있다.

(그림 6) OpenNebula 가상 머신 모니터링 화면



(Figure 6) Virtual monitoring screen of OpenNebula

(그림 7) Lustre 병렬 파일시스템 설정



(Figure 7) Configure of Lustre Parallel filesystem

4. 결론

본 논문에서는 클라우드 컴퓨팅의 가상화 속성[13]을 이용하여 다양한 교육 실습 환경에 접목함으로써 사용자가 요구하는 환경을 유연하고 신속하게 제공할 수 있음을 살펴보았다. 빈번하게 실습 환경 제공을 위해 구축해야 할 여러 시스템 환경을 가상 어플라이언스로 이미지화하여 설정함으로써 관리의 부하를 축소시킬 뿐만 아니라 실습 환경의 규모에 대해서도 유연하게 대처할 수 있다. 나아가 실습환경의 다양해지고 규모가 확장되면서 발생할 수 있는 가상화 실습환경의 데이터 입출력 부하를 분산파일시스템을 적용함으로써 해결할 수 있었다.

향후, 다양한 실습환경을 보다 용이하게 관리하기 위해 단일 환경을 묶어 클러스터 형태로 구성하여 제공하는 가상 클러스터(Virtual Cluster) 기술을 적용이 필요하다. 이를 이용해서 여러 실습환경을 보다 쉽게 클러스터 단위로 제공할 수 있다. 또한 본 실습환경에 적용된 오픈소스 소프트웨어 대한 성능평가(가상자원 관리 및 병렬처리 부하, 디스크 I/O, 작업처리 방식 등)가 필요하며 이에 부합하는 적합한 오픈소스 평가가 요구된다.

References

[1] Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb. A taxonomy and survey of cloud computing system

s. In Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC, NCM '09, pages 44-51, 2009.

[2] J. Diaz, G. v. Laszewski, F. Wang, and G. Fox, "Abstract Image Management and Universal Image Registration for Cloud and HPC Infrastructures" IEEE Cloud. 2012.

[3] Daniel J. Abadi, "Data Management in the Cloud: Limitations and Opportunities", In IEEE DE Bulletin, vol 32(1), pp 3-12, Feb 2009.

[4] B.Sotomayor, R. S. Montero, I. M. Llorente, and IFoster, "Virtual Infrastructure Management in Private and Hybrid Clouds," IEEE Internet Computing, vol. 13, pp. 14-22, Sep-Oct 2009.

[5] Nimbus Project Web Page. <http://www.nimbusproject.org>

[6] Eucalyptus Web Pages (Open Source). <http://open.eucalyptus.com>

[7] Nurmi, D.Wolski, R. Grzegorzczak, C. Obertelli, "The Eucalyptus Open-Source Cloud-Computing System", CCGRID '09. 9th IEEE/ACM International Symposium, Cluster Computing and the Grid, 2009. pp.124 - 131, 2009.

[8] OpenNebula Web Page. <http://www.opennebula.org>

[9] OpenStack Home Page. <http://www.openstack.org>

[10] Amazon Elastic Compute Cloud (EC2). <http://aws.amazon.com/ec2>

[11] Peter Sempolinski and Douglas Thain, "A comparison and critique of eucalyptus, opennebula and nimbus", In Proceedings of the 2010 IEEE Second International Conference on Cloud Computing Technology and Science, CLOUDCOM '10, pp.417-426, 2010.

[12] Lustre a Network Clustering FS, <http://www.lustre.org>

[13] Jong-Hei Ra, "Qualitative Study on Service Features for Cloud Computing "Journal of Digital Contents Society Vol. 12 No. 3, pp.319-327 Sep. 2011.



윤 준 원

2004년 : 고려대학교 대학원 컴퓨터학과(이학석사)
2011년 : 고려대학교 대학원 컴퓨터교육학과 박사 수료
2005년~현재 : KISTI 국가슈퍼컴퓨팅연구소 선임연구원

관심분야 : 그리드 컴퓨팅, 분산 컴퓨팅, 결합포용시스템, 클라우드 컴퓨팅, 슈퍼컴퓨팅



박 찬 열

1995년 : 고려대학교 대학원 컴퓨터학과(이학석사)
2000년 : 고려대학교 대학원 컴퓨터학과(이학박사)
2002년~현재 : KISTI 국가슈퍼컴퓨팅연구소 책임연구원

관심분야 : 그리드 컴퓨팅, 분산 컴퓨팅, 결합포용시스템, 슈퍼컴퓨팅



송 의 성

1997년 : 고려대학교 컴퓨터학과(학사)
1999년: 고려대학교 대학원 컴퓨터학과(이학석사)
2005년: 고려대학교 대학원 컴퓨터학과(이학박사)

2006년~현재: 부산교육대학교 컴퓨터교육과 교수
관심분야: 컴퓨터교육, 교육용로봇교육, 컴퓨터네트워크, 스마트러닝