

# Cloud Computing Platform for GIS Image Processing in U-city

Jong Won Park\*, Chang Ho Yun\*, Shin-gyu KIM \*\*, Heon Y. Yeom\*\*, Yong Woo LEE (Corresponding Author)\*

\*School of Electrical & Computer Engineering, University of Seoul, Korea

\*\*School of Computer Science & Engineering, Seoul National University, Korea

comics77@uos.ac.kr, touch011@uos.ac.kr, {sgkim, yeom}@dclslab.snu.ac.kr, ywlee@uos.ac.kr

**Abstract**— Ubiquitous city (U-city) is a city with ubiquitous information technology that enables citizens to access the converged information anywhere and anytime. A lot of compute power are required in U-city, because large amount of data should be processed in real-time. Cloud computing enables users to use the abstracted and virtualized computing resources and to process huge amount of information without having their own computing power. This paper presents a GIS image processing platform that uses cloud computing. It finds and selects optimal computing resources for applications and run virtual machines to execute the applications. In this paper, we explain the parallel air pollution map generation as a use case for the suggested platform and show how efficiently it processes the massive data.

**Keywords**— Cloud Computing, U-city, PaaS, IaaS, Hadoop

## I. INTRODUCTION

Under the banner of Low-Carbonated Green Growth, the Korean Government proposed u-city which is one of New Growth Engines in 2009.[1] U-city is a smart future city which adopts ubiquitous technologies in the city. Korea has been running long term projects, building more than forty ubiquitous-cities including Seoul, the capital city of Korea, that are one of new economic development paradigm to nation.

U-city consists of fundamental infrastructure, convergence network infrastructure including wired/wireless network, ubiquitous sensor network and integrated urban operation center. Among them, important and difficult point of building u-city is building the integrated urban operation center. The integrated urban operation center collects city-scale context information from various Information and communication technologies (ICT) devices such as sensor devices and processes it to provide the information desired by the user. Because of the amount of information produced in the city, the integrated urban operation center requires large-scale Internet Data Center (IDC) with high performance cluster equipment. However, because u-city is usually built for purpose of the public interest by local government, operating the integrated urban operation center is difficult to earn the money from

providing the u-city service so that it is difficult to have and operate its own IDC.

Cloud computing is a new paradigm and changing IT industry for green IT and low cost. It can make users lease abstracted and virtualized computing resources such as infrastructure, platform and application as a service. [2] Infrastructure as a Service (IaaS) of Cloud computing can make the integrated urban operation center in u-city to process huge amount of information without their own large-scale infrastructure.

This paper introduces the GIS image processing platform based on cloud computing for u-city. In the u-city environment, the 3D image processing with GIS data requires great amount of computational resource because of complex and large amount of spatial information. GIS image processing platform with Cloud Computing is a solution to provide more computing power than the integrated urban operation center has. It consists of OpenNebula[3]/Haizea[4] that is open source cloud computing platforms and Virtual Machine Job Manager (VMJM). It has features that are monitoring status of compute nodes, selecting optimal nodes for GIS processing. We choose the combination of OpenNebula and Haizea due to lease computing resources from other cloud platforms easily and provide API for developer.

In this paper, in order to show how the suggested platform operate, we run a application for generating an 3D air pollution map as one of use cases for the suggested platform. In order to generate the air pollution map for the u-city, data of huge amount sensors should be stored and processed. Because of these reasons, it requires large-scale computing resources so that we implement it on our platform. This application is implemented by using Hadoop that is one of cloud computing runtime technique.

This paper is organized as follows: In section 2, we compare IaaS-style open source frameworks for cloud computing with our work. Section 3 explains how to adopt cloud computing in u-city. Section 4 describes the architecture of GIS image processing platform. Section 5 shows processing of air pollution map application that is a use case of GIS image processing platform. Finally, section 6 gives conclusions and presents future works.

## II. RELATED WORK

Eucalyptus is an open source software framework for cloud computing that supports private and hybrid clouds. It is highly modular with internal components consisting of Web services, which make them easy to replace and expand. It allows users to use external public clouds to form a hybrid cloud. It implements the Amazon's EC2 API [5] and allows interoperability with existing AWS-compatible services and tools. It is also compatible with Amazon's S3 interface by using Walrus that is a storage service that allows users to store persistent data, organized as buckets and objects. Currently it supports VMs that runs on Xen hypervisor [6], KVM [7] and VMware (Enterprise only) [8].

Nimbus is an open source toolkit (licensed under the terms of the Apache License) that constructs an Infrastructure as a Service (IaaS) for cloud computing focusing experimentation with features for scientific needs and interoperability. It allows a client to dynamically and securely lease remote resources by deploying virtual machines (VMs) on existing physical clusters into the cloud. It support the virtualization based on Xen and KVM. It supports Remote interfaces that are Amazon EC2 WSDLs, Amazon EC2 Query API and Grid community WSRF Storage implementation compatible with S3 REST API. In order to schedule virtual machines, it can configure scheduler like PBS or SGE.

OpenNebula is an open-source project aimed at building the industry standard open source cloud computing tool to dynamically deploy and re-allocate virtual machines on the complexity and heterogeneity of distributed data center infrastructures. It is implemented based on virtualization, storage and network for provision of services on a distributed infrastructure. The underlying virtualization layer supports Xen, KVM, or VMware currently. OpenNebula can also support a hybrid cloud model by using Cloud Drivers to interface with external clouds. In this way, users can be offered computing capacity from a public cloud to meet peak demands or more request than local infrastructure capacity. It provides the XML-RPC API and libvirt interface for developer so that it is easy to modify it for need of developer. As a VM placement policies, libvirt, Haizea or any placement policies are used. We choose OpenNebula due to these reasons: corresponding most to the hybrid cloud definition, integrates transparently external resources in the cloud, easily extended to support other cloud providers and providing various placement policies and reallocating virtual machines [9].

Haizea, an open-source VM-based lease manager, is developed for purposes of the VM scheduler and a drop-in replacement for OpenNebula. [10][11] Because other VM managers use immediate provisioning or best-effort provisioning, VMs must be allocated right away or cannot be allocated, or VMs deploying tasks are queued until resources can be allocated. Otherwise, it offers various leases methodology such as advance reservation leases, best-effort leases, immediate leases, and so on. We adopt combination of OpenNebula and Haizea to our platform for VM management.

TABLE 1. COMPARISON OF OPEN SOURCE CLOUD FRAMEWORK

|                                    | Eucalyptus                                            | Nimbus                                             | OpenNebula                                                                         |
|------------------------------------|-------------------------------------------------------|----------------------------------------------------|------------------------------------------------------------------------------------|
| Who                                | Computer Science Department, University of California | University of Chicago, Argonne National Laboratory | Distributed Systems Architecture Research Group, Universidad Complutense de Madrid |
| Hypervisor                         | Xen, KVM, VMware(Enterprise only)                     | Xen                                                | Xen, KVM, VMware                                                                   |
| Scheduling                         | Round Robin resource selection                        | Round Robin resource selection                     | Rank scheduling using dynamic information                                          |
| Programmable VM Placement policies | No                                                    | No                                                 | Any static/dynamic placement policy                                                |
| Support for Hybrid Cloud           | Partially                                             | Partially                                          | Yes                                                                                |

## III. CLOUD COMPUTING IN U-CITY

"Ubiquitous computing is a post-desktop model of human-computer interaction in which information processing has been thoroughly integrated into everyday objects and activity"[12]. Ubiquitous computing comprises a wide range of technologies which bring computing away from the desktop to the world where people are living and working [13]. It means that existing or being everywhere at the same time, users can access the network anytime, anywhere and they may not necessarily even be aware that they are doing so. It also plays a role as a fundamental technology for u-city which provides various integrated services for smart city.

A u-city is defined as "a city where U-City Services are available without regard to time and location and which is realized by the application of U-City technologies in order to enhance the quality of lives of citizens within and competitiveness of the city itself." [14]. In the u-city, all information are linked, and virtually everything is linked to an information system. As mentioned in the introduction, building the integrated urban operation center is the most important and difficult to serve the integrated services for citizens in u-city. It performs the key function not only to collect and process various information to operate the u-city efficiently but also to provide integrated and incorporated services [15]. The integrated urban operation center needs middleware to meet the mentioned functions [16].

By latest draft of the NIST Working Definition, the definition of Cloud computing is "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal

management effort or service provider interaction.”[17]. It allows users to use computing resources as a service without knowledge of any infrastructure through the internet. It has some characteristics. The consumer can lease as much or as little computing power as resources both up and down as needed. It can make consumer to reduce Initial investment costs. Because it includes virtualization and distributed computing technologies, it scales computing resources up and down easily so that it increases utilization of IDCs. It can also support physical and virtual resources are assigned and reassigned according to consumer demand.

As mentioned above, as integrated urban operation center of u-city adopts cloud computing, there are many advantages through cloud computing. They are as follows:

- 1) The integrated urban operation center can lease as much or as little computing power as resources both up and down as needed. It reduces Initial investment costs.
- 2) Because cloud computing supports physical and virtual resources are assigned and reassigned, the integrated urban operation center allocates and reallocates u-city application and compute nodes according to consumer demand.
- 3) Because cloud computing includes virtualization and distributed computing technologies, it manages computing resource so that it increases utilization of IDCs.
- 4) U-city Applications are easily deployed, started, managed through encapsulation. It means that it supports better Security, manageability, and isolation.

#### IV. GIS IMAGE PROCESSING PLATFORM

GIS image processing platform in u-city Processes image with GIS data automatically when user request the job. It loads the virtual machines on the nodes to provide the computing resource, submits the job from request of user and returns the result of the job to user. Also, because number of sensors in u-city is not fixed, it should leases computing resources from external compute nodes. For these functions, GIS image processing platform consists of 4 Layer: Cloud Portal, Cloud Manager, Haizea and OpenNebula. The combination of Haizea and OpenNebula were explained section 2. By using OpenNebula, we can provide the VMs with specific image and a virtual network environment and the various underlying hypervisors to create and control VMs through pluggable Drivers. A separate Scheduler component in OpenNebula makes decision for Virtual Machine placement. We choose Haizea as a Scheduler for Virtual Machine placement. Haizea accesses to information on all requests received by OpenNebula, keep track current and future allocations of VMs and sends the appropriate deployment commands to the OpenNebula Core. Cloud Manager consists of Cloud Infrastructure Manager (CIM) and Virtual Machine Job Manager (VMJM). We implement CIM from OpenNebula Management Console that is a open source for web interface OpenNebula. It provides web interface to be compatible to OpenNebula interface and enable to submit job through cooperating with VMJM. VMJM is a job management system

that performs the function such as PBS [18], Condor-G [19], Nimrod-G [20] and LSF [21] in grid computing. We expands VMJM from Seoul Grid Job Management System (SGJOMS) that supports intelligent Grid job submission and tracing with resource brokering, convenient Grid file management, valuable Grid job accounting, resource monitoring, etc.. [22] VMJM consists of u-city service submitter that submits the job or service on VM and u-city service monitor that monitors VMs through OpenNebula. It provides common solutions for higher level services performing intelligent job management.

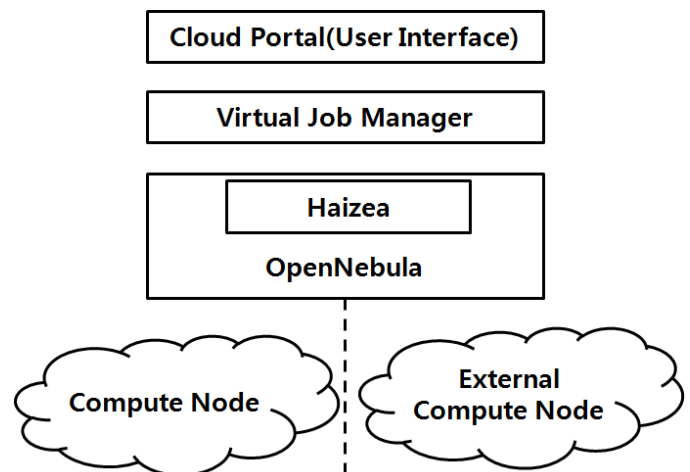


Figure 1. The architecture of GIS Image Processing Platform.

GIS image processing platform is operated as a follows. First of all, if users request the service to u-city portal, u-city portal submit it to VMJM. VMJM assigns job to job queue and request the computing resource suitable for job to CIM. CIM reserves the VMs through Haizea. Haizea schedules VM launching through OpenNebula. OpenNebula launches VMs to optimal Compute Nodes and CIM returns the ready message to execute job to VMJM. VMJM executes the job and after job completion, return job results to u-city portal. VMs are terminated.

When GIS image processing platform is used, the advantage is as follows. 1) It analyses resource of compute nodes, selects the optimal node for requirements and loads virtual machines on the node to provide the computing resource. 2) It leases the computing resources from other cloud computing services if computing resources on the cluster system is not enough to run the applications. 3) It helps the user make air pollution map easily without knowledge to process GIS data.

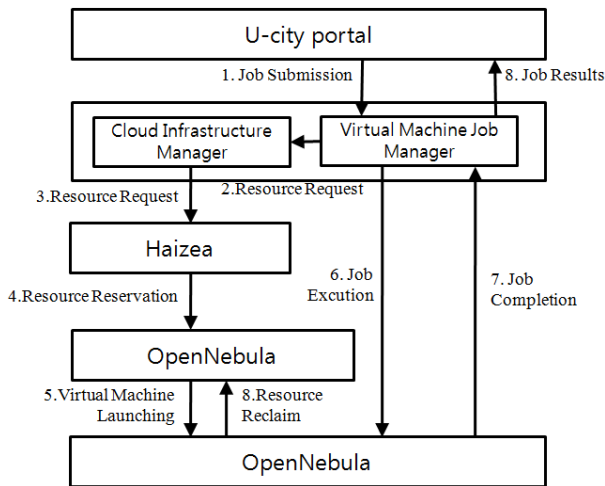


Figure 2. The execution flow of GIS Image Processing Platform.

## V. A USE CASE: GENERATION OF AIR POLLUTION MAP

In this paper, as one of use cases on the suggested platform, we run parallel air pollution map generation application. Because the air pollution map combines noise information with GIS map, it requires a large amount of computing power and cannot be timely done with personal computers. In the reason, the air pollution map is usually made offline mode for long time and not in three-dimension but in two-dimension. However, current cities have high-rising buildings and we need to show difference of the air pollution level on each floor. In consequence, it is important to generate the 3D air pollution map. The 3D image processing with GIS data should deal with complex and large amount of spatial information and requires great amount of computational resource.

Figure 2 shows the execution flow of our platform and processing the generation of air pollution map on our platform. As user in u-city requests the generation of air pollution map to u-city portal, u-city portal sends information of number of VMs, the capacity of VMs and the requested area to VMJM. CIM generates the configuration data for Haizea based on received information and send it to Haizea. Configuration data includes schedule policies such as advance reservation, best effort provisioning and immediate provisioning and duration of VMs. Haizea writes the configuration file based on the received configuration data. After that, Haizea sends start command of VMs based on configuration file to OpenNebula. Because Haizea currently cannot do any image deployment with OpenNebula, we predeploy VM images installed application for generation of air pollution map on a shared NFS filesystem. After VMs are loaded on physical nodes, VMJM sends the start command of application for generation of air pollution map to VMs based on information of requested area from u-city portal. The application for generation of air pollution map is implemented by using Hadoop framework [23]. Hadoop framework is an open source for implementation of MapReduce and has also been successfully used to program data parallel applications. MapReduce [24] is a massively parallel computing programming model, proposed by Google. It lets the

programmer just implement map and reduce function for distributed parallel processing. It provides automatic failure recovery by restarting the failed tasks. It means that allows us to use a simple but convenient cloud computing environment, which eventually allows us to implement parallelism to run our applications. Also, MapReduce gives better support to quality of services such as fault tolerance and monitoring in data intensive parallel applications. Figure 3 shows the overview of the MapReduce execution.

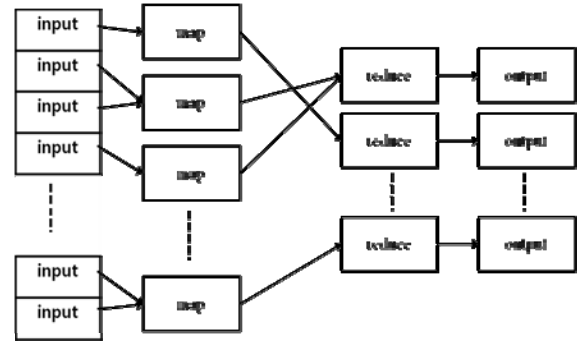


Figure 3. The MapReduce execution overview.

For these reasons, we use the MapReduce to process 3D city modeling and combine the air pollution level on city model with distributed parallel computing. The process of the air pollution map with MapReduce is as shown in figure 4. MapReduce are used twice for generating the air pollution map. We implement map and reduce functions to make city model and combine the air pollution level on city model each other. The output of modelling the city is reused as an input in making of an air pollution map. After this application is terminated, VMJM sends the termination command of VMs to CIM. It shows the user make air pollution map easily without knowledge to process GIS data. The users should just ask the GIS image processing platform to generate the 2D/3D air-pollution maps by selecting the desire area for the map in the u-city portal if they want to visualize the air-pollution. Then, air-pollution maps are generated automatically.

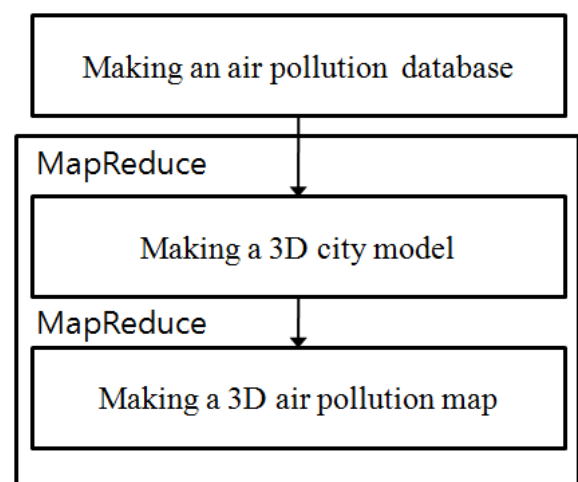


Figure 4. The process of the 3D air pollution map.

For the performance evaluation, we used a ten nodes cluster, where 8 nodes had Dual Core Intel processor and 2 nodes had Quad Core Intel Processor and each node had 4 GB memory. Each node of the cluster was connected through a giga-bit Ethernet switch. The JVM version 1.6.0\_20 was used for Hadoop. For the air pollution map area, we selected Cheonggyecheon, Seoul, Korea. Figure 5 shows the performance.

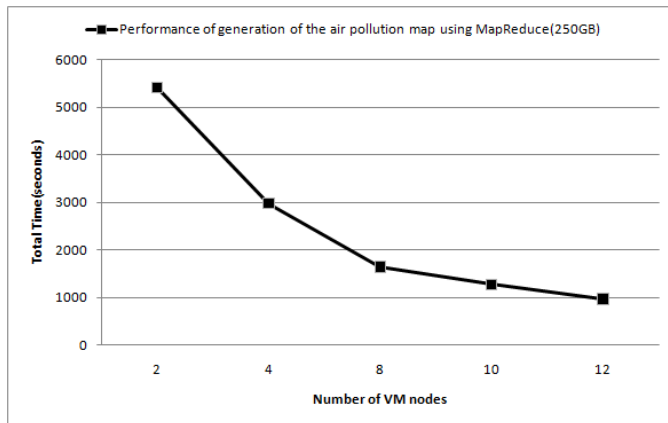


**Figure 5.** Performance of generation of the air pollution map using MapReduce

## VI. CONCLUSIONS

This paper proposed how to adopt Cloud Computing in the u-city. It shows new possibility that the integrated urban operation center reduces operating costs by using Cloud Computing. As an example of applying cloud computing to u-city, this paper explained Cloud Computing Platform for GIS Image Processing in U-city. It had several advantages as follows: 1) Reservation of VM compute nodes, 2) monitoring VM nodes, 3) submission the job on VM nodes 4) leasing VM nodes from external other cloud providers. We run application for generation of air pollution map as a use case on the platform to show this platform is operated. We will leave the performance evaluation of the application for generation of air pollution map on the platform as a future work.

## ACKNOWLEDGMENT

This study was supported by the Seoul Research and Business Development Program (10561), Smart (Ubiquitous) City Consortium and Seoul Grid Center. We would like to give thanks to, Mr. Cheol Sang Yoon, Mr. Eui Dong Hwang, Mr. Kyoung Kyu LEE and the staffs of Seoul Grid Center and the members of Smart (Ubiquitous) City Consortium for their contribution to this research.

## REFERENCES

- [1] Prime Minister's Office, Ministry of Education, Science and Technology, Public Relations Officer (2009) The 29th National Science & the third Presidential Council for Future & Vision Press Releases.
- [2] Buyya, R., Chee Shin Yeo, Venugopal, S., Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities, 10th IEEE International Conference on High Performance Computing and Communications, 2008. HPCC '08., DOI: 10.1109/HPCC.2008.172, pp. 5-13
- [3] OpenNebula Homepage [online], Visited in October 2010, Available from: <http://opennebula.org/>.
- [4] B. Sotomayor, K. Keahey, and I. Foster, "Combining batch execution and leasing using virtual machines," in HPDC '08: Proceedings of the 17th International Symposium on High Performance Distributed Computing. ACM, 2008, pp. 87-96.
- [5] Amazon EC2 Homepage [online], Visited in October 2010, Available from: <http://aws.amazon.com/ec2/>
- [6] Christopher Clark, Keir Fraser, Steven Hand, Jakob Gorm Hansen, Eric Jul, Christian Limpach, Ian Pratt and Andrew Warfield "Live Migration of Virtual Machines" In Proceedings of the 2nd Symposium on Networked Systems Design and Implementation (NSDI '05), May 2005, Boston, MA.
- [7] I. Habib. Virtualization with kvm. Linux J., 2008(166):8, 2008.
- [8] VMware Homepage [online], Visited in October 2010, Available from: <http://www.vmware.com/>
- [9] Borja Sotomayor, Ruben S. Montero, Ignacio M. Llorente, Ian Foster, "Virtual Infrastructure Management in Private and Hybrid Clouds," IEEE Internet Computing, vol. 13, no. 5, pp. 14-22, Sep./Oct. 2009, doi:10.1109/MIC.2009.119
- [10] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, and A. Roy, "A distributed resource management architecture that supports advance reservations and co-allocation," in Proceedings of the International Workshop on Quality of Service, 1999, DOI:10.1.1.27.7478, pp. 27-36.
- [11] W. Smith, I. Foster, and V. Taylor, "Scheduling with advanced reservations," in IPDPS '00: Proceedings of the 14th International Symposium on Parallel and Distributed Processing. IEEE Computer Society, 2000, p. 127.
- [12] M. Weiser, "The Computer for the Twenty-First Century", Scientific American, 1991, pp. 94-101.
- [13] Vipul Hingne, Anupam Joshi, Tim Finin, Hillol Kargupta, Toward as Pervasive Grid, International Parallel and Distributed Processing Symposium, 2003(IPDPS '03), DOI: 10.1109/IPDPS.2003.1213380, 2003.
- [14] Ministry of Land, Transport and Maritime Affairs, Decree about building a ubiquitous city, This Decree enter into force on June 9, 2009, 9770.
- [15] Ho Lee, Sang and Hoon Han, Jung and Taik Leem, Yoon and Yigitcanlar, Tan (2008) Towards ubiquitous city : concept, planning, and experiences in the Republic of Korea. In: Yigitcanlar, Tan and Velibeyoglu, Koray and Baum, Scott (Eds.) Knowledge-Based Urban Development : Planning and Applications in the Information Era. IGI Global, Information Science Reference, Hershey, Pa., pp. 148-169.
- [16] Hae-Sun Jung; Jong-kwun Baek; Chang-Sung Jeong; Yong-Woo Lee; Phil-Doo Hong;. Unified Ubiquitous Middleware for U-City, International Conference on Convergence Information Technology, 2007(ICCIT '07), DOI: 10.1109/ICCIT.2007.356, pp. 2374-2379, 2007.
- [17] NIST, the latest draft of the NIST Working Definition of Cloud Computing, Available from: <http://csrc.nist.gov/groups/SNS/cloud-computing/>, Version 15 of the document, dated 8-19-09.
- [18] PBS Homepage [online], Visited in October 2010, Available from: <http://www.pbspro.com>.
- [19] J.Frey, T.Tannenbaum, I.Foster, M.Livny, S. Tuecke, "Condor-G: A Computation Management Agent for Multi-Institutional Grids", Cluster Computing, 5(3):237-246, 2002.
- [20] D.Abramson, D.Giddy, J.Foster, and L.Kotler, "High Performance Parametric Modeling with Nimrod/G: Killer Application for the Global Grid?", International Parallel and Distributed Processing Symposium, 2000.
- [21] Ming Q. Xu.: Effective Metacomputing using LSF MultiCluster, 1st International Symposium on Cluster Computing and the Grid, ISBN:0-7695-1010-8, 2001.
- [22] Yong Woo Lee, Phil Doo Hong, "A Job Management System for a Large Scale Grid System", Seventh International Conference on Computer and Information Technology, pp. 291-294, 2007.
- [23] Hadoop Homepage [online], Visited in October 2010, Available from: <http://hadoop.apache.org/>
- [24] Jeffrey Dean and Sanjay Ghemawat. "MapReduce:Simplified Data Processing on Large Clusters." In Communications of the ACM, Volume 51, Issue 1, pp. 107-113, 2008.