

Method for Calculation of Cloud Computing Server Capacity

Kook-Hyun Choi¹, Yang-Ha Chun², Se-Jeong Park³, Yongtae Shin⁴
and Jong-Bae Kim^{5*}

^{1,2}Department of IT Policy and Management, Graduate School of Soongsil University,
Seoul 156-743, Korea

⁴Department of Computer Science, Soongsil University, Seoul 156-743, Korea

^{3,5*}Graduate School of Software, Soongsil University, Seoul 156-743, Korea

¹khchoi@tsline.co.kr, ²yangha00@yongin.ac.kr,
³sejung90@naver.com, ⁴shin@ssu.ac.kr, ^{5*}kjb123@ssu.ac.kr

Abstract. As cloud computing requires a great deal of service resources, hardware capacity it is an important factor that should be calculated more systematically and accurately. For this reason, research on the hardware capacity calculation technique for SaaS, the foundation of cloud services, is key research that is needed to expand into a variety of projects within the cloud in the future. In this research, the method and criteria for capacity calculation of SaaS-based hardware are presented. It is possible to utilize the results of this research as a guideline for calculating hardware capacity when establishing a cloud computing environment.

Keywords: Cloud Computing, SaaS, Capacity, Calculation, Criteria

1 Introduction

In order to calculate the hardware capacity of cloud computing, it is first necessary to present 3 service models, or the capacity calculation models that are suitable for SaaS (Software as a Service), PasS (Platform as a Service), and IaaS (Infra as a Service) depending on the nature of IT resources. Meanwhile, since SaaS has the same process structure as the service of WEB/WAS, it may be suitable for identifying existing service characteristics and cloud characteristics. Thus, in this research, the method and criteria for calculation of SaaS-based hardware capacity are presented.

2 Literature review

A few studies on the calculation of existing H/W capacity [3, 4] presented the general considerations and procedures for the performance criteria and scale calculation of each server (OLTP server, WEB/WAS server) for H/W scale calculation. However, since the methods presented in the studies are based on such existing computing

environments as client/server and internet/intranet systems, it is difficult to apply them to capacity calculation of cloud computing service.

SaaS is a service model in which cloud computing service providers provide software and services through a network and service users receive the software service after access to and certification of the network [1][7]. The characteristics of SaaS [2, 6, 8] are service continuity, reliability, resource pooling, internet-based service, and virtualization.

3 Hardware Capacity Calculation for Cloud Computing

In this research, the scope for calculation of SaaS-based hardware capacity is limited to CPU and memory.

For calculation of HW capacity, the first process carried out is the process for the direction of system establishment and the investigation of basic data.

Next, as shown in Table 1, the standard loads used by SaaS for DB server (No. of simultaneous users and No. of transactions per user) and WEB/WAS server (No. of simultaneous users and No. of operations per user) are calculated.

Table 1. Determination of Standard Loads for DB Server and WEB/WAS Server

DB Server	
No. of Simultaneous Users	The number of simultaneous users is calculated in consideration of the number of current users and the annual rate of increase (10%) for the next 3 years: $800 * 1.1 * 1.1 * 1.1 = 1065$ persons.
No. of Allocated Virtual Machines	The number of allocated virtual machines is calculated considering that virtual machines are allocated to 90% of the simultaneous users: $1065 * 0.9 = 956$ machines
No. of Transactions per User	The number of transactions generated by a user in a minute ranges from 4 to 5 in general. Based on the investigation of DB server basic data, 4 is applied.
WEB/WAS Server	
No. of Simultaneous Users	Since the number of simultaneous users for the target year should be calculated in consideration of the annual rate of increase, the number of simultaneous users for the target year (after 3 years) = the current simultaneous users * the annual rate of increase (10%) * 3 years. Thus, the number of simultaneous users after 3 years = $800 * 1.1 * 1.1 * 1.1 = 1065$
No. of Allocated Virtual Machines	The number of allocated virtual machines is calculated considering that virtual machines are allocated to 90% of the simultaneous users: $1065 * 0.9 = 959$ machines
No. of Operations per User	According to the basic investigation results, this server provides the web service with frequent transactions and mainly handles application logics. Thus, 6 is applied as the number of operations per user on the basis of the criteria.

For calculation of CPU Size of DB server, the number of transactions per cloud minute, the basic tpmC correction, the correction of cloud peak-time load, the correction of cloud data base size, the correction of cloud application load, the correction of cloud cluster, and the standby rate of cloud system are considered, and the detailed calculation bases and contents are provided in Table 2.

Table 2. Results for Calculation of CPU for DB Server

Calculation Results	$\text{CPU(tpmC Unit)} = \text{No. of Transactions per Minute} * \text{Basic tpmC Correction} * \text{Correction of Peak Time Load} * \text{Correction of Data Base Size} * \text{Correction of Application Structure} * \text{Correction of Application Load} * \text{Cluster Correction} * \text{System Standby Rate} = 3,836 * 1.3 * 1.3 * 1.4 * 1.8 * 1.7 * 1 * 1.3 = 36,104 \text{ tpmC}$
---------------------	--

The number of simultaneous users, the system region, the memory required per user, the buffer cache correction, cluster correction, and the system standby rate should be calculated to determine the scale of memory for DB server, and the bases of calculation and the contents of calculation are provided in Table 3.

Table 3. Results for Calculation of DB Server Memory

Calculation Results	$\text{Memory} = (\text{System Region} + \text{No. of Virtual Machines} * \text{Memory Required per User}) * \text{Buffer Cache Correction} * \text{System Standby Rate} = (1408 + 959 * 2) * 1.2 * 1.3 = 5,189\text{MB}$
---------------------	---

Next, for WEB/WAS server, the scale of the calculation target server with 2 tiers is measured in consideration of 8 measurement items, as shown in Table 4.

Table 4. Results for CPU Calculation of WEB/WAS Server

Calculation Results	$\text{OPS} = \text{No. of Virtual Machines} * \text{Correction of Application Load Requested} * \text{Correction of Interface Load} * \text{Correction of Peak-Time Load} * \text{System Standby Rate} * \text{No. of Operations per User} * \text{Cluster Correction} = 959 * 1.1 * 1.1 * 1.3 * 1.3 * 6 * 1 = 11,766$
---------------------	---

The number of simultaneous users, the system region, the memory required per user, the buffer cache correction, and the system standby rate should be calculated in order to calculate the memory scale of WEB/WAS server on the basis of the data investigated at the stage of basic data and work analysis. The calculation bases and the calculation contents are provided in Table 5.

Table 5. Bases for Calculation of WEB/WAS memory

Calculation Results	$\text{Memory (MB)} = \{ \text{System Region} + (\text{Memory Required per User} * \text{No. of Virtual Machines}) \} * \text{Buffer Cache Correction} * \text{System Standby Rate} = \{ 384\text{MB} + (959 * 2\text{MB}) \} * 1.2 * 1.3 = 3,591\text{MB}$
---------------------	---

As shown in Table 6, the results for calculation of the final capacity are obtained by multiplying the calculated value of the target server by the architecture correction value.

Table 6. Results for Calculation of Final Capacity

Division	DB Server		WEB/WAS Server	
	Calculated Value	Decided Value	Calculated Value	Decided Value
CPU	36,104 tpmC		11,766	18,826
Memory	10GB (10,240MB)		8GB (8,192MB)	

4 Conclusion

The research on the calculation of SaaS hardware capacity, the foundation of cloud services, is basic research that is required to expand into various projects within the cloud. Further research will be conducted, using this research as a starting point, on the performance measurement of each service to allocate cloud resources to each customer who uses SaaS.

References

1. TTA PG420. Cloud Computing Glossary of Terms. TTAK.KO-10.0467, 2010
2. TTA PG420. Cloud Computing Threats and Requirements Analysis. TTAK.KO-10.0466, 2010
3. Jonghei Ra, Kwangdon Choi, Haeyong Jung, "The Study on Hardware Sizing Method Based on the Calculating", Journal of Information Technology Services, Vol.5, No.1, pp. 47-59, 2006
4. Jonghei Ra, Kwangdon Choi, "An Exploratory Study on Capacity Sizing Method for Information System : Focus on H/W Sizing in Public Sector", Journal of Information Technology Services, Vol.3, No.2, pp. 9-23, 2004
5. Canturk Isci, Alper Buyuktosunoglu, and Margaret Martonosi. "LONG-TERM WORKLOAD PHASES : DURATION PREDICTION AND APPLICATION TO DVFS". IEEE Micro 25(5):39-51, 2005
6. Cloud Security Alliance. Top Threats To Cloud Computing V1.0. (<http://www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf>)
7. P.Mell and T.Grance. The NIST Definition of Cloud Computing(<http://csrc.nist.gov/groups/SNS/cloud-computing/>)
8. Cloud Security Alliance. Security Guidance for Critical Areas of Focus in Cloud Computing V2.1.(<http://www.cloudsecurityalliance.org/csaguide.pdf>)