

Towards Characterizing Patterns of Cloud Backend Workloads

Haiying (Helen) Shen
Dept. of Electrical and Computer Engineering
Clemson University, Clemson, SC 29631
shenh@clemson.edu

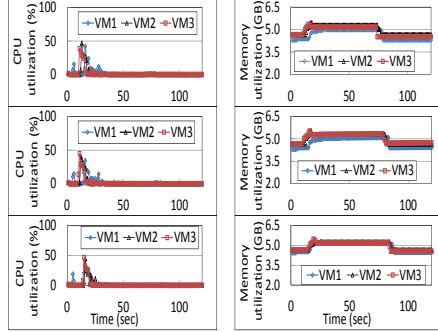
In this report, we study VMs running short-term MapReduce jobs and observe that the VMs running the same job task tend to have similar resource utilization patterns over time. We also study the PlanetLab and Google Cluster VM traces and find that different VMs running a long-term job exhibit similar periodical resource utilization patterns.

In order to predict the resource demand profiles of cloud VMs, we conducted a measurement study on VM resource utilizations. Workload arrives at the virtual cluster of a tenant in the form of jobs. Usually all tasks in a job execute the same program with the same options. Also, application user activities have daily patterns. Thus, different VMs running the same job tend to have similar resource utilization patterns. To confirm this, we conducted a measurement study on both short-term jobs and long-term jobs.

1 Utilization Patterns of VMs for Short-Term Jobs

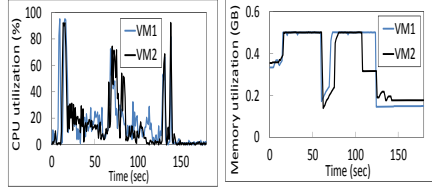
MapReduce jobs represent an important class of applications in cloud datacenters. We profile the CPU and memory utilization patterns of typical MapReduce jobs. We conducted the profiling experiments on our cluster consisting of 15 machines (3.4GHz Intel(R) i7 CPU, 8GB memory) running Ubuntu 12.04. We constructed a virtual cluster of a tenant with 11 VMs; each VM instance runs Hadoop 1.0.4. We recorded the CPU and memory utilization of each VM every 1 second.

We used *Teragen* to randomly generate 1G data, then ran *TeraSort* to sort the data in the virtual cluster. Figures 1(a) and 1(b) display the resource utilization results of three VMs for different generated datasets. Figure 2 displays the resource utilizations of two VMs running *TestDFSIO write*, which generates 10 output files with each file having 0.1GB. Figure 3 displays the resource utilizations of two VMs running *TestDFSIO read*, that reads 10 input files generated by *TestDFSIO write*. From the figures, we can find that the VMs collaboratively running the same job have similar resource utilization patterns. The VMs



(a) CPU utilization (b) Memory utilization

Figure 1: VM resource utilization for *TeraSort* on three datasets.



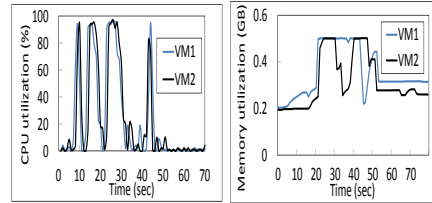
(a) CPU utilization (b) Memory utilization

Figure 2: VM resource utilization for *TestDFSIO write*.

running the same job on different datasets also have similar resource utilization patterns. We repeatedly ran each experiment several times and got similar resource utilization patterns for the VMs, which indicates that VMs running the same job task at different times also have similar resource utilization patterns.

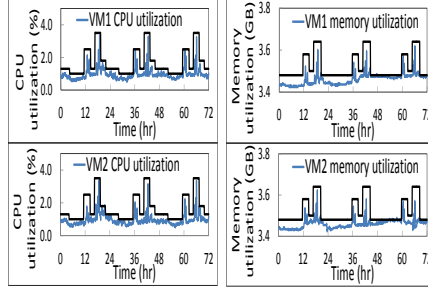
2 Utilization Patterns of VMs for Long-Term Jobs

To study the utilization patterns of VMs for long-term jobs, we used publicly available Google Cluster trace [1] and the PlanetLab trace [2]. The Google Cluster trace records resource usage on a cluster of about 11000 machines from



(a) CPU utilization (b) Memory utilization

Figure 3: VM resource utilization for *TestDFSIO read*.



(a) CPU utilization (b) Memory utilization

Figure 4: VM resource utilization from Google Cluster trace.

May 2011 for 29 days. The PlanetLab trace contains the CPU utilization of each VM in PlanetLab every 5 minutes for 24 hours in 10 random days in March and April 2011. In the Google Cluster trace, we analyzed 700 VMs and found that different VMs running the same job tend to have similar utilization patterns. Also, for a long-term VM, daily periodical patterns can be observed from the VM trace. We randomly chose two VMs running the same job as an example to show our observations. Figure 4(a) shows the CPU utilizations of two VMs every five minutes during three days and Figure 4(b) shows their memory utilizations. We see that both CPU and memory resource demands exhibit periodicity approximately every 24 hours. Also, the two VMs exhibit similar resource utilization patterns since they collaboratively ran the same job. In the PlanetLab trace, we analyzed 900 VMs and also found that they exhibit daily periodical patterns. Figure 5 shows the CPU utilization of a randomly selected VM to show their periodical patterns.

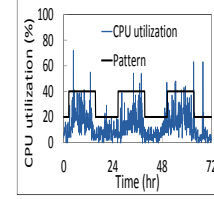


Figure 5: VM resource utilization from PlanetLab trace.

References

- [1] “Google cluster data,” <https://code.google.com/p/google-clusterdata/>.
- [2] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, and R. Buyya, “Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms.” *SPE*, 2011.