Project Report on Cloud Simulation

Authors:

Karthik Gurram (P.No: 9411108617),
Mounisha Dandu (P.No: 9509229382)

School of computing,
Blekinge Institute of Technology, Karlskrona,
Sweden.
Email: {kagu17, moda17}@student.bth.se

CONTENTS

INTRODUCTION

STATE OF ART

METHODOLOGY

EXPERIMENT SETTINGS

INSTALLATION OF THE CLOUDSIM

CONFIGURATION FOR PART- A

CONFIGURATION FOR PART- B

CONFIGURATION FOR PART- C

OBSERVATIONS AND RESULTS

Part- A

Part- B

Part- C

CONCLUSION

FUTURE WORK

REFERENCES

Introduction

Cloud Computing has grown to be a vital technology for providing secure and scalable computational services. It is presented as an Infrastructure, Platform, or Software as services (IaaS, PaaS, SaaS). Large investments are needed for setting up the experimental setup for analysis of various scenarios. By using a simulation tool, management of a large number of virtualized servers, data centres, different architecture and policies for best utilization of power and resources can be done. Nowadays the increasing demand for the development of cloud products and experimentation of new Cloud technologies has opened the possibility for the utilization of CloudSim simulation tool. CloudReports is a Graphical User Interface, GUI for CloudSim, which uses cloudsim for its simulation. The services offered by it can be implemented in private and public clouds. CloudReports simulation platform, helps the providers and the customers to overcome from inaccurate and abstract evaluations of various other simulation tools.

CloudReports is a tool to simulate different scenarios of Cloud Computing environments. It supports various classes for defining virtual machines, datacenters, users, management policies (scheduling and provisioning) and computational resources

This report mainly focuses on the understanding and observations of various cloud infrastructures. For this, we have considered three scenarios, Part A, Part B and Part C. In the first scenario, Part A, we considered two datacenters with one host each and four customers with two Virtual Machines for each customer. In

Part B, we added four more customers with some changes in configurations for one datacenter . For one customer eight more VM’s were added in Part C.

For all the three scenarios resource utilization, costs and various other parameters were calculated.

State of Art

CloudSim is a simulator for modelling and setting up the experimentation setup for various cloud computing environments for optimizing the performance. Using this tool User can set up various testbeds, design and analyse them.To model the cloud system components such as data centres, virtual machines and configure system architecture and policies and protocols, Cloudsim can be used.

Some of the functionalities of CloudReports are:

- It supports simulation and modelling of large-scale Cloud Computing datacenters and virtualized server hosts.
- It also supports simulation of computational resources, data centre network topologies, and federated clouds.
- The broker provides a list of Virtual Machines to create on hosts and schedule cloudlets.
- It models scheduling of CPU resources in different levels of host and VM.

CloudReports is a simulation tool which uses CouldSim for simulation. It is a GUI, where the users can set up various environments in a cloud [3]. They can add various datacenters and a number of hosts along with its configurations such as RAM, bandwidth, power and number of processing elements. For the customers, they can add a number of virtual machines and set up the configurations such as specifying the scheduling policies, hypervisor, RAM, bandwidth and various other parameters. The output of the CloudReports is a simple HTML page where the user can easily observe the resource and power utilization for both providers and the customers. Users can observe both the overall utilization and the utilization for specific VM's and hosts. The number of VM's in the environment, costs, scheduling algorithms for tasks (cloudlets) and the execution time can be clearly observed.

CloudReports is an easy to use tool for costing infrastructure and utilization for a specific cloud environment.

Methodology

The main aim of this project is to obtain the basic knowledge of cloud simulator, CloudReports and to experience a real-time modelling and maintenance of a data centre. CloudSim helps to simulate the real-time scenarios in datacenters and make the users aware of the estimation of costs and design efficiency.

In this report, the main configuration setup consists of two datacenters with one host each. We setup few customers with certain provisions to submit cloudlets to datacenters. The amount of processing strength available to the task required is specified by the virtual machines scheduling. The scheduling policy applied in this simulation is Dynamic Workload policy and the resource provisioning policy used is simple policy. A stochastic process has been used for the utilization policy for CPU, Bandwidth and RAM. Hence for each cloudlet, the resource allocation is randomly performed.

In the simulation report, 3 different scenarios of cloud environment have been simulated. In the first scenario, we observed the number of successfully executed cloudlets on each of the two data centers by each customer, the number of successfully allocated VMs on the two data centers by each customer, the overall Power Consumption vs Time graph for the two data centers, the debt of each user on both datacenters. In the second scenario, we increase the number of customers to 8 from 4 and try to observe the same aspects as in the first scenario. We also decreased RAM and increased Bandwidth for few customers and observe the changes in debts respectively. In the final scenario, for one of the customers, we increased the number of Virtual machines and try to observe the same aspects as in the first scenario.

Experiment Settings

The experiment has been performed on Ubuntu 16.04 LTS operating system.

Installation of the CloudSim:

1. Clone the CloudReports file from the following link:

https://github.com/thiagotts/CloudReports

2. Go to the directory of CloudReports and follow the instructions by visiting the link.

3. Open the terminal and use the following command to set up the required development environment.
   - apt-get install openjdk-7-jdk
4. Later, go to the directory where the CloudReports zip file is extracted

5. Run the following command to start the Cloud reports Simulator.

   - java –jar CloudReports-master/bin/dist/CloudReports.jar
6. This command runs the CloudReport simulation tool and you can see the simulator as shown in Figure1.

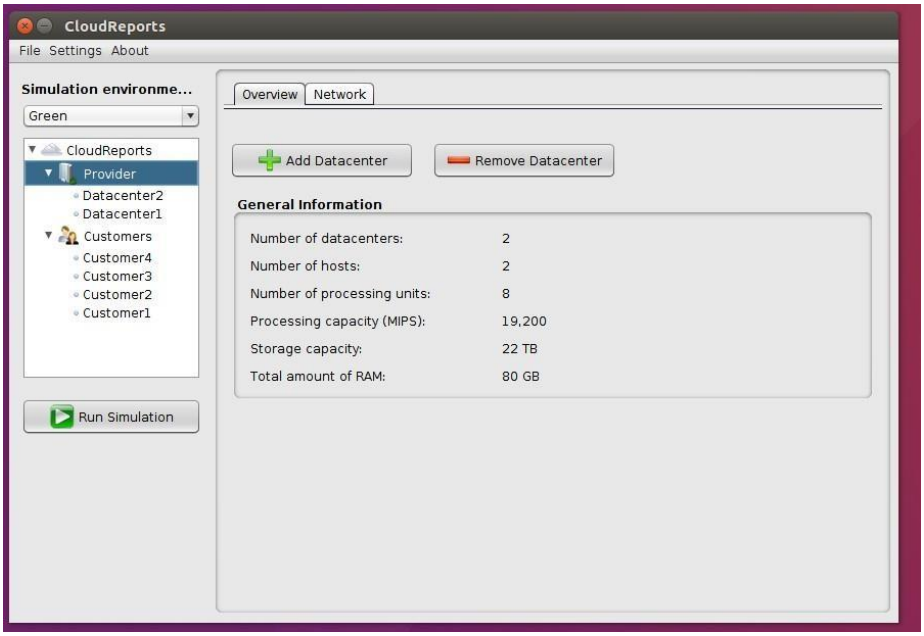

Fig 1: CloudReport simulation tool

Project Specifications:

The project has three scenarios namely Part A, Part B and Part C. Each scenario has different configurations.

Configurations for first scenario, Part A

1. In a simulation environment, select the environment 1.

2. Add two datacenters under providers as Datacenter1 and Datacenter2.

3. Then, add four customers to Customers namely Customer1, Customer2, Customer3 and Customer4.

4. Then, for Datacenter1 give the following parameters:

   Under the General tab, change the parameters as stated below:

   a. Architecture: x86

   b. Operating System: Linux

   c. Hypervisor: Xen

   d. Scheduling Interval: 30

   e. Allocation Policy: Single threshold

   f. Upper threshold: 0.8 (must be set between 0 and 1)

   g. Lower threshold: 0.2 (must be set between 0 and 1)

   h. Enable the VM migrations.

   Under Hosts tab, add one host and change the parameters for Host1 as stated below:

   a. RAM: 40000

   b. BW: 10000000

   c. Storage: 1000000

   d. Max Power: 250

   e. Static power percent: 0.7

   f. VM Scheduling: Time shared

   g. Set the power model to linear

   h. MIPS/PE: 2400

i. Processing elements: 4

j. Power model: Linear

k. Set RAM, BW, and PE Provisioners to Simple

In Costs tab give the parameters as:

a. Processing cost (per sec): 0.1

b. Memory cost (per MB): 0.05 c.

Storage cost (per MB): 0.001 d.

Bandwidth cost (per MB): 0.1

Under Storage Area Network (SAN) parameters are as follows:

a. Capacity: 10000000

b. Bandwidth: 10

c. Latency: 5

After setting up all these configurations, the parameters for the Datacenter1 will be as shown in figure 2.

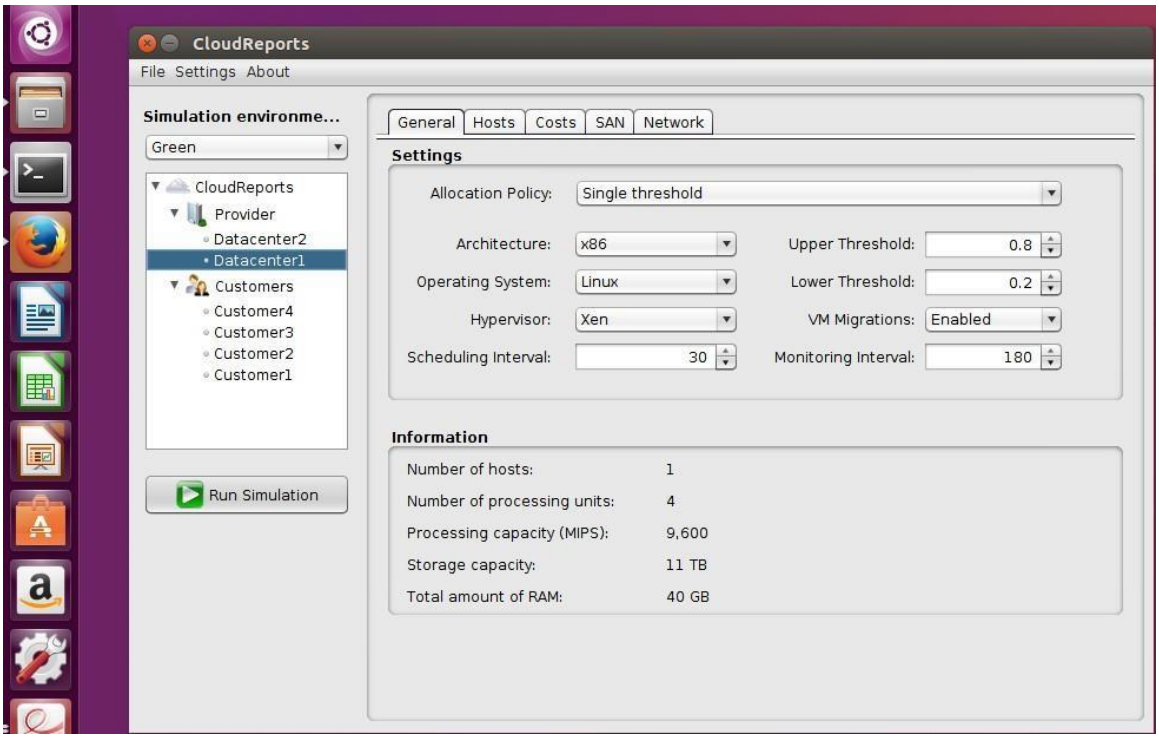

Fig 2: Configuration parameters for Datacenter 1

5. Change the configurations for Datacenter2 similar to that of Datacenter1.

6. Later, select Customer1 and give the configurations as stated below:

   Under Virtual Machines tab, add two Virtual Machines and for each VM provide the following parameters:

   a. Image size: 1000

   d. Proc. Elements: 1

   e. MIPS: 1000

   f. RAM: 512

   g. BW: 100000

   h. Priority: 1

   i. Hypervisor: Xen

   j. Scheduling policy: Dynamic workload.

Change the parameters under Utilization Profile as stated below:

a. Broker policy: Round Robin

b. Processing elements: 1

c. Maximum length: 50000

d. File size: 500

e. Output size: 500

f. CPU Utilization model: Stochastic

g. RAM utilization mode: Stochastic

h. Bandwidth utilization model: Stochastic

i. Set the link bandwidth and the link delay between each node to 1.

7. See figure-3 for the Customer1 configurations.

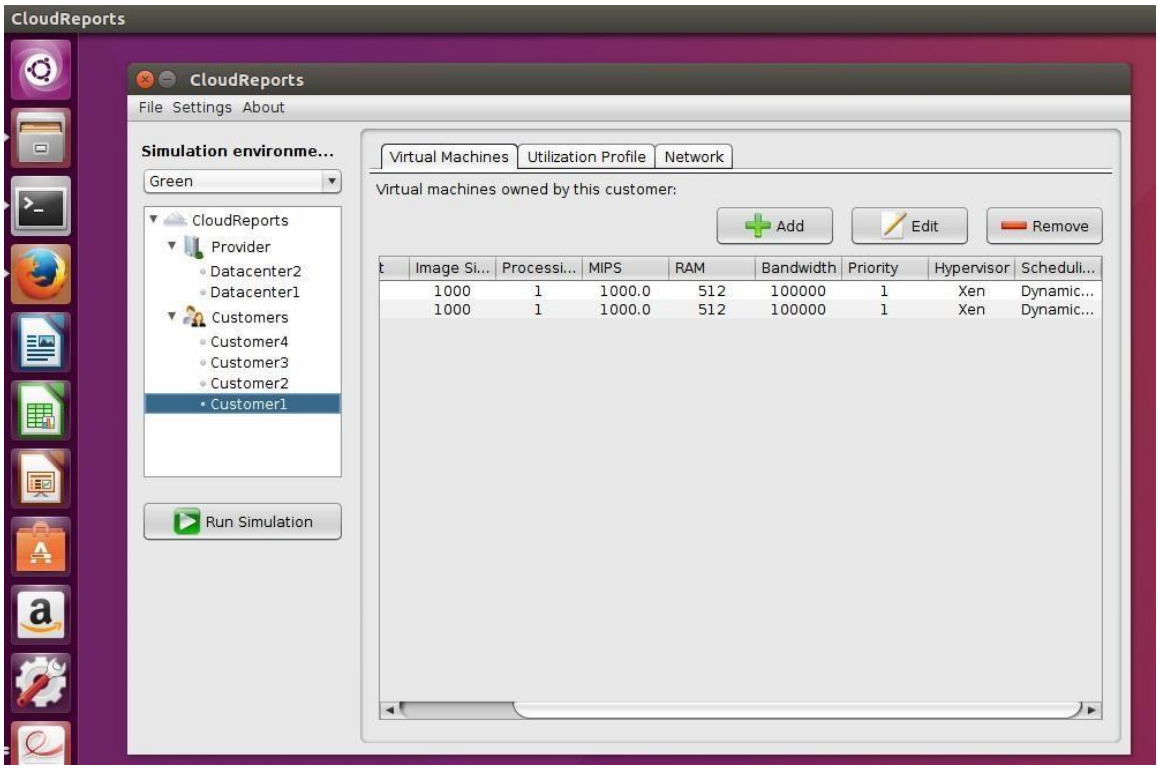

Fig 3: Configuring Customer-1 in Part-A

8. Do the same for Customer2, Customer3 and Customer4.

9. Run the simulation by clicking on Run Simulation.

Configurations for the second scenario, Part B:

1. Part B is an extension for Part A. The configurations will be changed in environment1.

2. For the previous scenario, i.e. Part A, we will add four more customers for environment1.

3. Each customer will have two Virtual Machines as in the case of Part A with the similar configurations.

4. Under Costs tab in Datacenter2, change the parameters as stated below:

   a. Processing cost: 1

   b. Memory cost: 0.5

   c. Storage cost: 0.01

   d. Bandwidth cost: 1

5. Figure-4 shows the Datacenter-2 parameters.

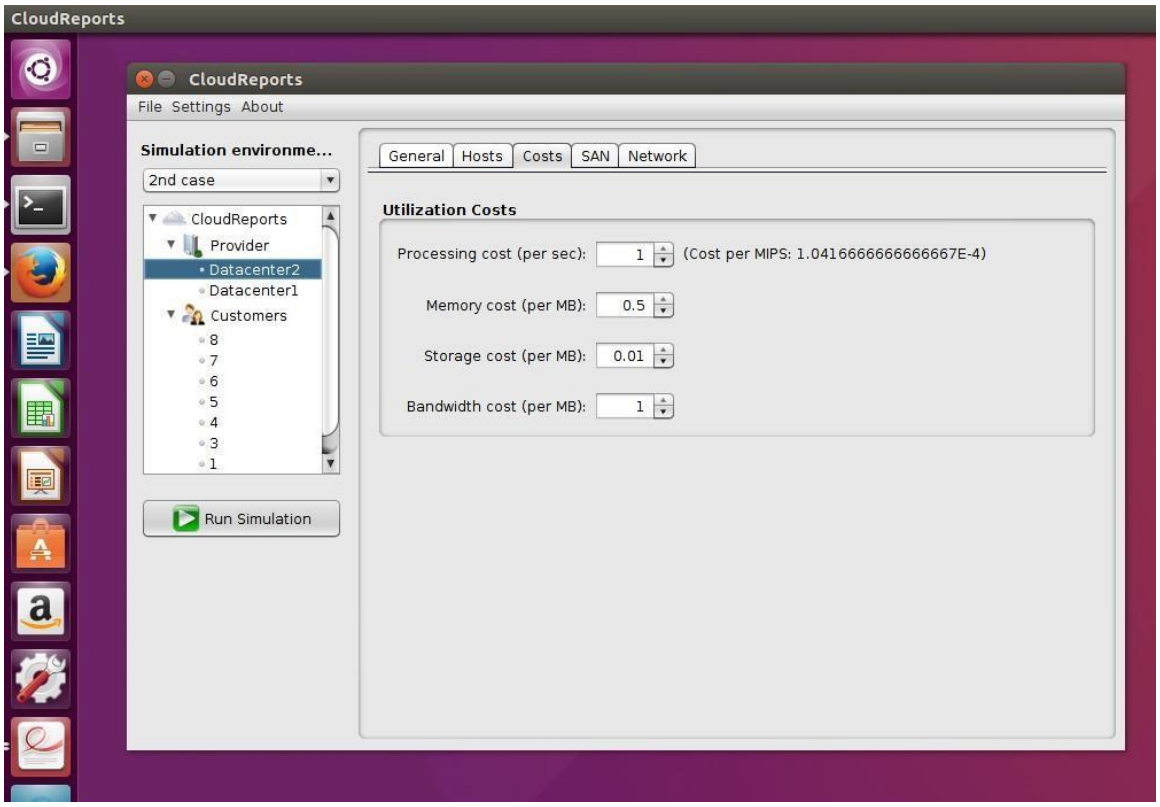

Figure 4: Configuration parameters for Datacenter-2 in Part-B

6. For Customer2, follow the changes as mentioned in figure 5:

   b. VM RAM: 64
   c. VM Bandwidth: 50.

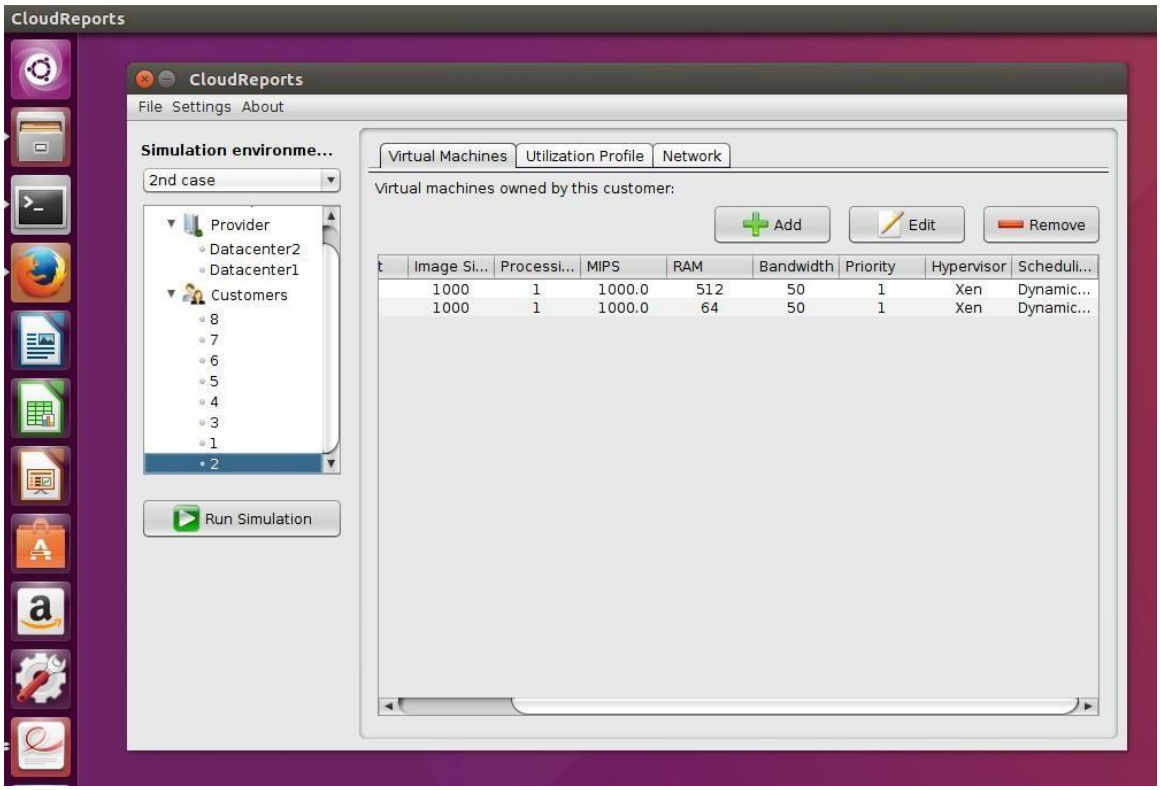

Figure 5: Configuration settings for customer-2 in Part B

7. As shown in the figure-6, for the Customer5 set:

   a. VM RAM: 128

   b. VM Bandwidth: 100.

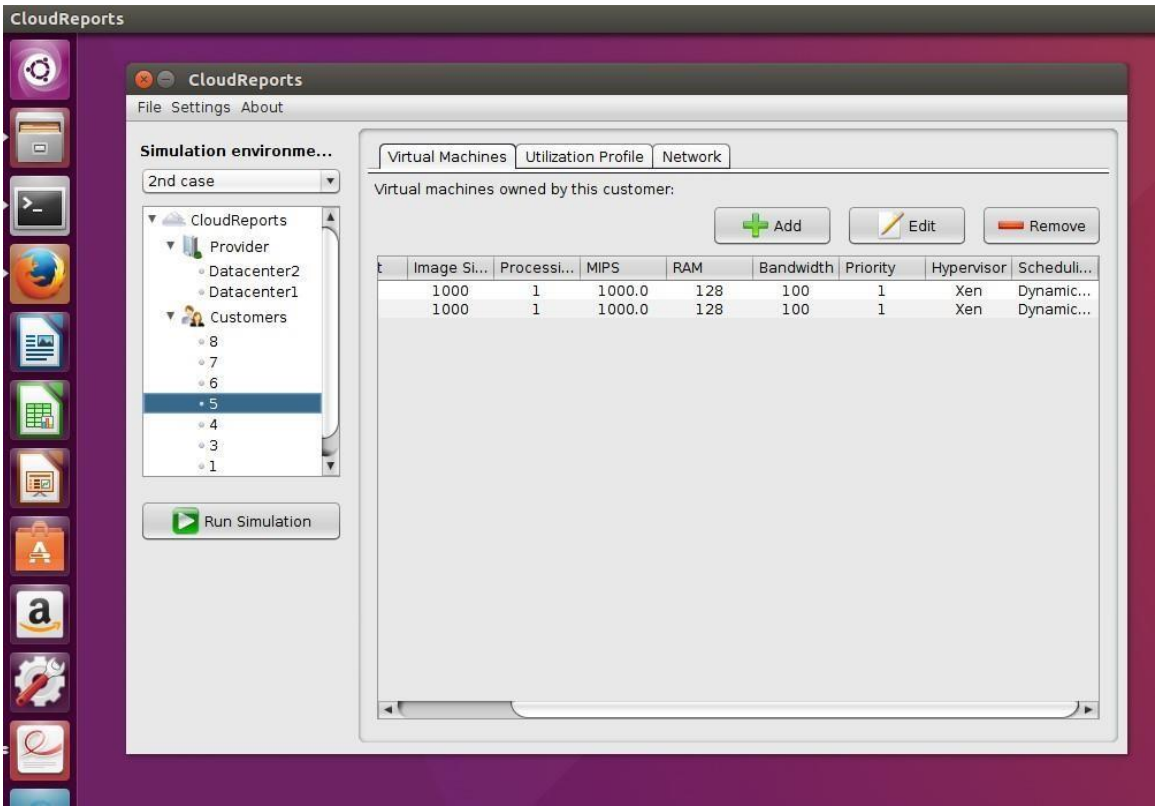

Figure 6: Configuration settings for Customer-5 in Part B

8. For Customer8, under network tab set the link to all customers and to all datacenters as:

   a. Bandwidth=0.1

   b. Latency=1000.

9. Observe the figure-7 for Part B configurations.

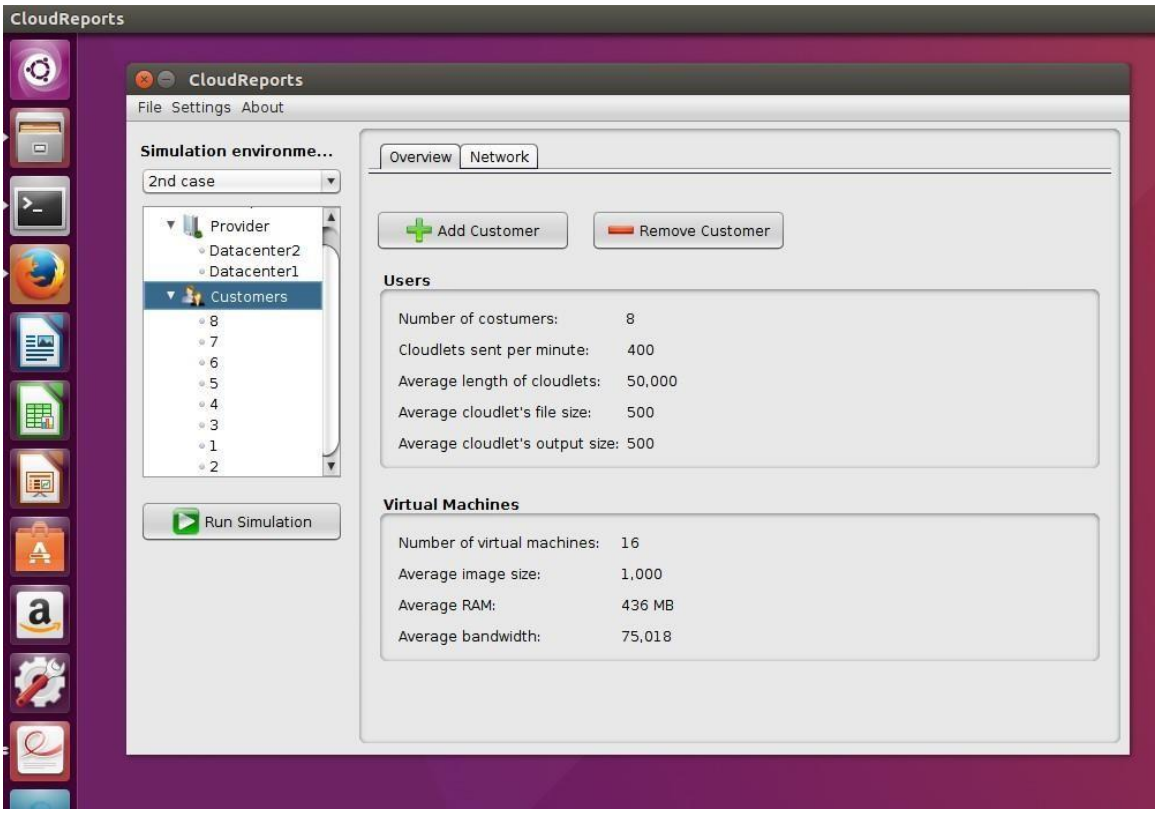

Figure 7: Configuration of Part B

10. Run the simulation process by clicking Run Simulation to observe the output of the simulation tool.

Configurations for third scenario, Part C:

1. Part C is an extension for the previous scenarios.

2. In this scenario set the number of Virtual Machines for Customer4 to ten. Observe the figure 8.

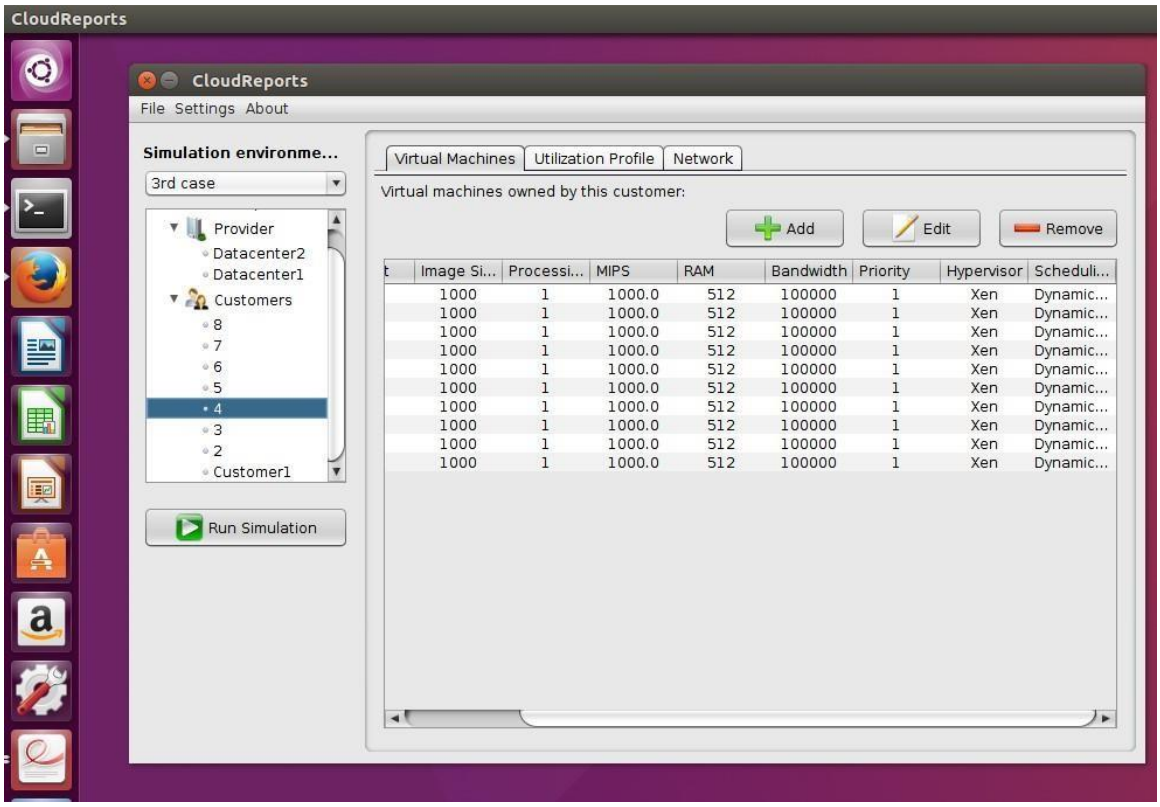

Figure 8: Configuration of Customer-4 in Part-C

3. Now, re-run the simulator tool by clicking on the Run Simulation and observe the output of the simulator.

Observations and Results

Part-A:

After setting up the configuration as described, the simulator tool is run. The following are the observations obtained from the output of simulator tool, and are as described below:

a) The simulation of the configured environment has finished in 4 seconds.

b) The simulation time for the process is 3000 seconds.

c) We observe that two data centres, Datacenter1 and Datacenter2 are successfully created with a single host in it.

d) We observe that there are four customers, each of them having two virtual machines and are using the resources provided by hosts in data centres.

e) The broker policy adopted by the customers is *Round robin*, which is used for submitting a list of VMs to be created as well as it schedules the cloudlets sequentially.

f) The CPU (MIPS) and RAM (MB) vs. Time graph for the Datacenter1 is as shown in the figure-9.

g) The graph represents that the RAM is constant whereas the CPU varies with time.

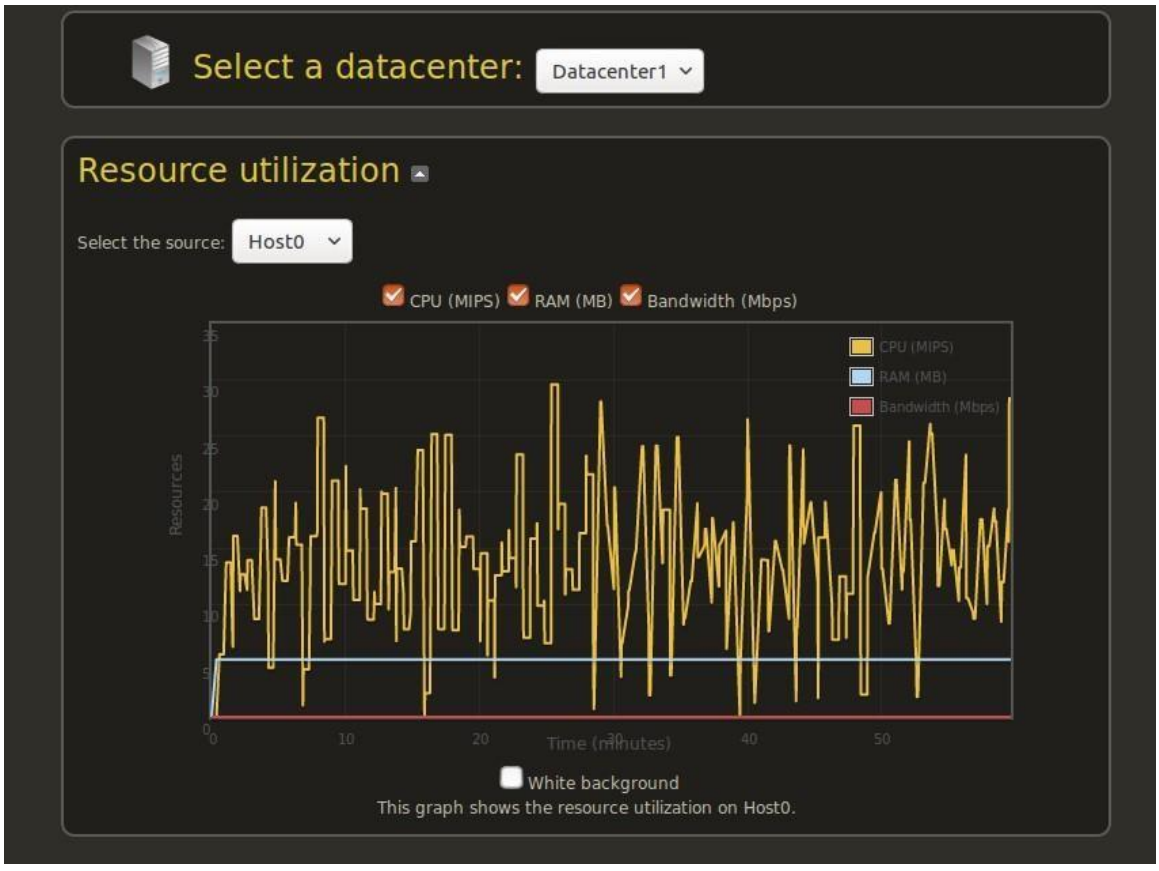

Figure 9: Resource Utilization on Host0 for Datacenter-1

h) The CPU (MIPS) and RAM (MB) vs Time graph for the Datacenter2 is as shown in the figure-10.

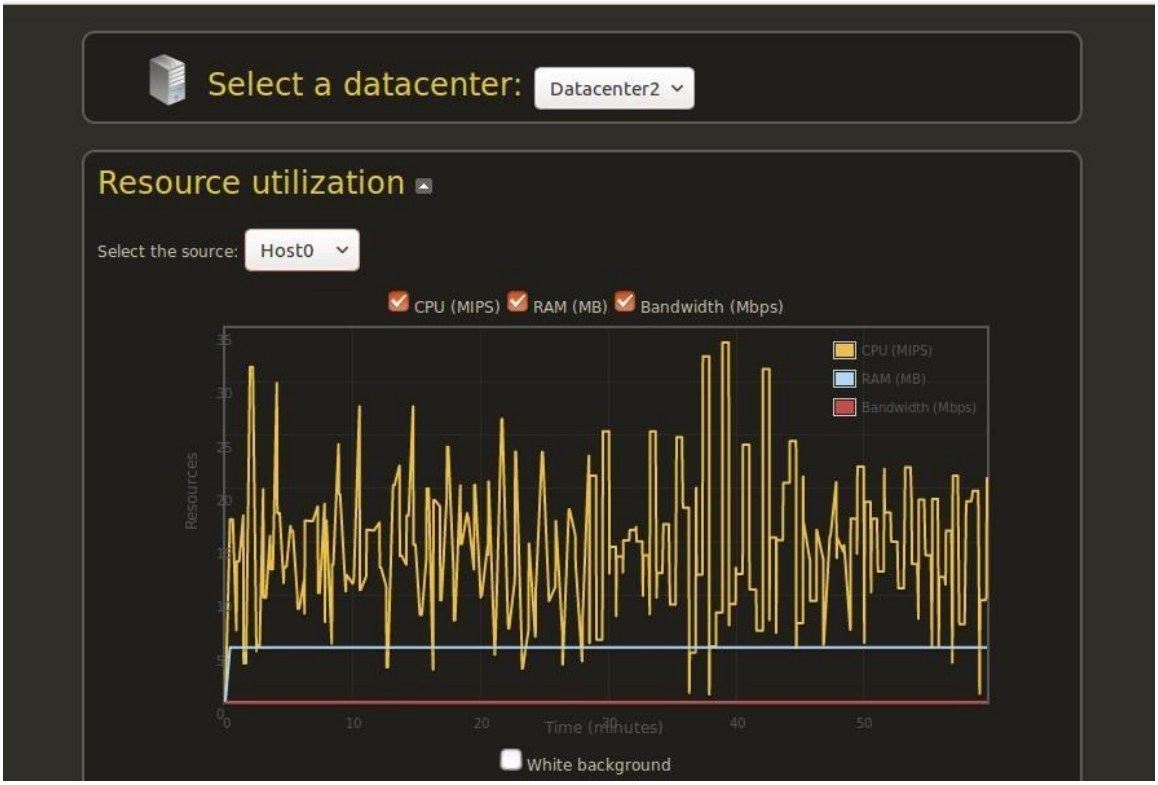

Figure 10: Resource Utilization on Host0 for Datacenter-2

i) The overall Power Consumption vs. Time graph for the Datacenter-1 is as shown in figure-11. The sudden falls in the graph are due to the idle state of Datacentre, where there are no cloudlet requests at that particular time. The power consumption mainly depends upon the cloudlets that are processed by the datacentre at that point of time. At some places, there is a sudden drop in power consumption to zero indicating that no cloudlet was scheduled at that point of time

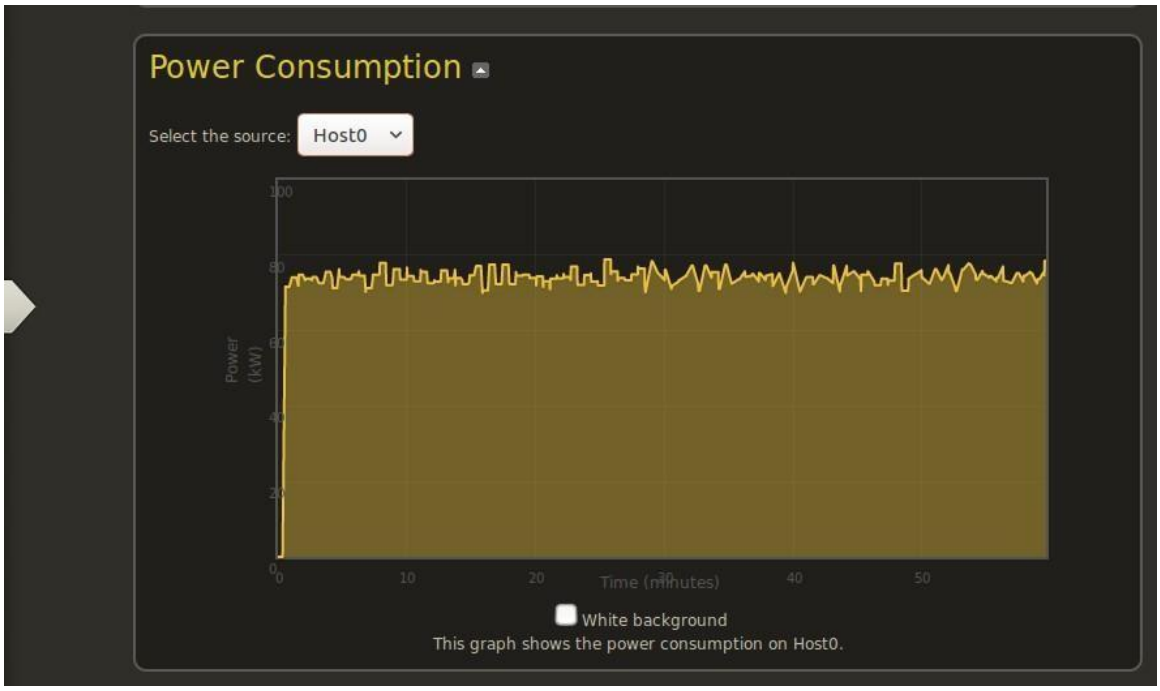

Figure 11: Power Consumption on Host0 for Datacenter-1

j) The overall Power Consumption vs. Time graph for the Datacenter-2 is as shown in figure-12. The sudden falls in the graph are due to the idle state of Datacenter, where there are no cloudlet requests at that particular time.

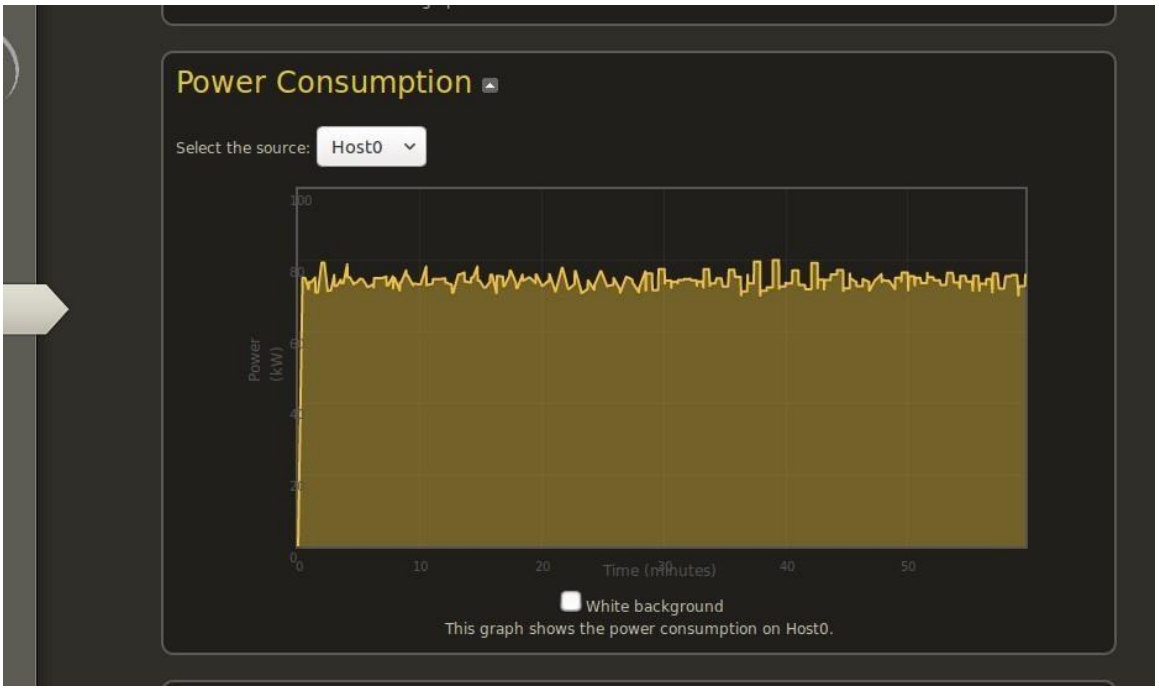

Figure 12: Power Consumption on Host0 for Datacenter-2

k) We observe that in Datacenter 1 and Datacenter 2, four virtual machines are successfully allocated on the host. Each customer has been successfully allocated with one virtual machine in the two data centres as shown in figure-13 and figure-

14.

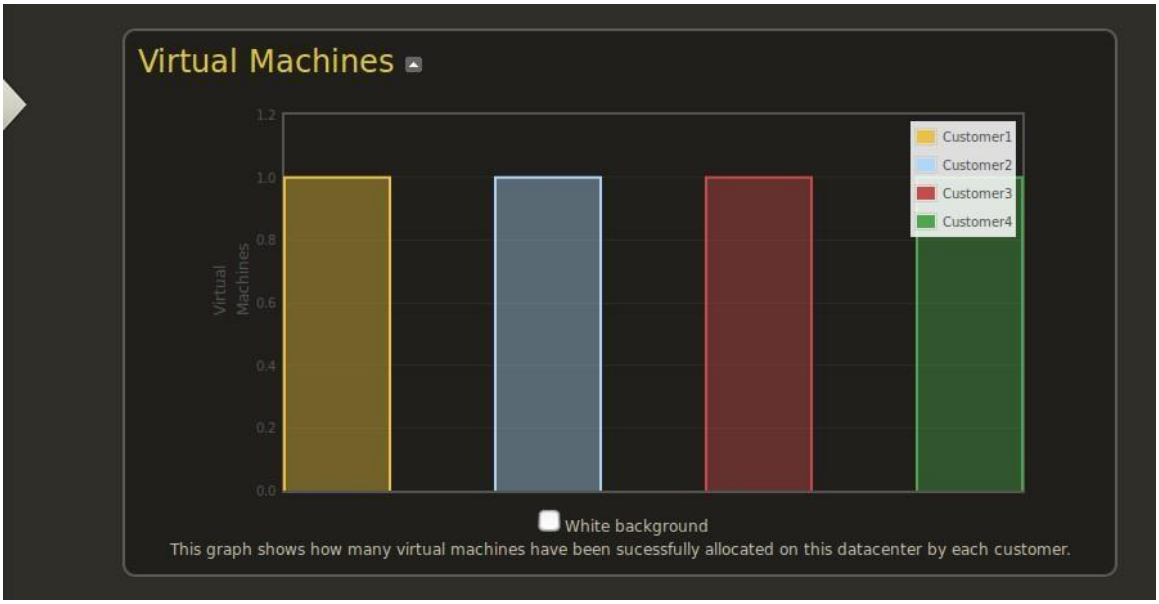

Figure 13: Graph shows how many virtual machines have been successfully allocated in Datacenter1 by each customer.

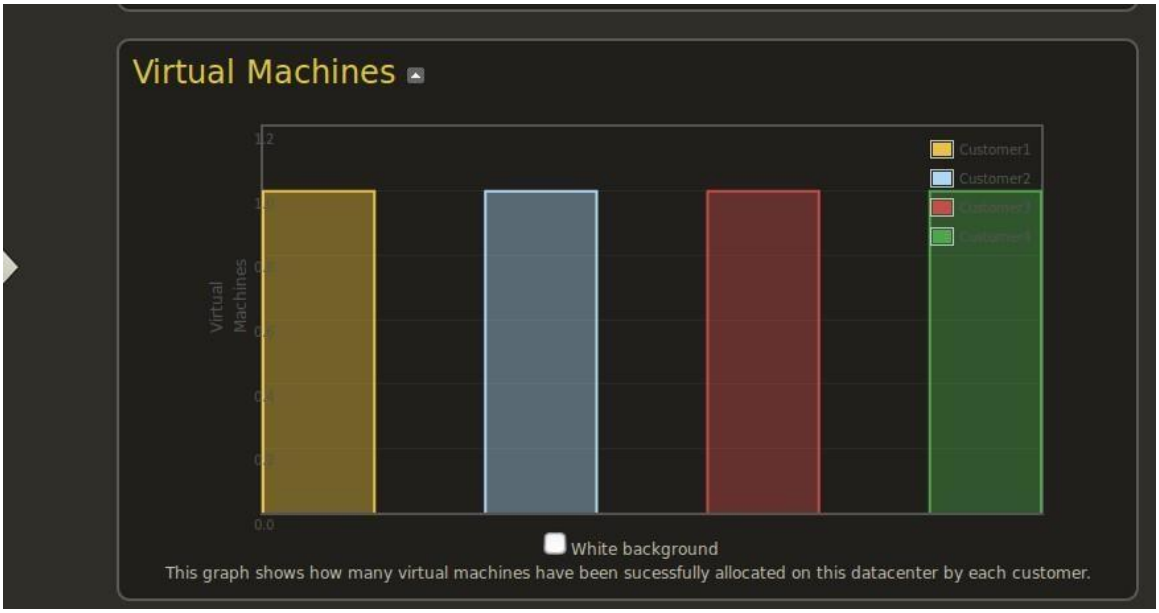

Figure 14: Graph shows how many virtual machines have been successfully allocated in Datacenter2 by each customer.

l) The number of successfully executed cloudlets by each customer on the two data centres on average is more than 35. The number of successfully executed cloudlets are almost same for all the customers because of the same configuration. The figure-15 and figure-16 depict the number of cloudlets executed by each customer.

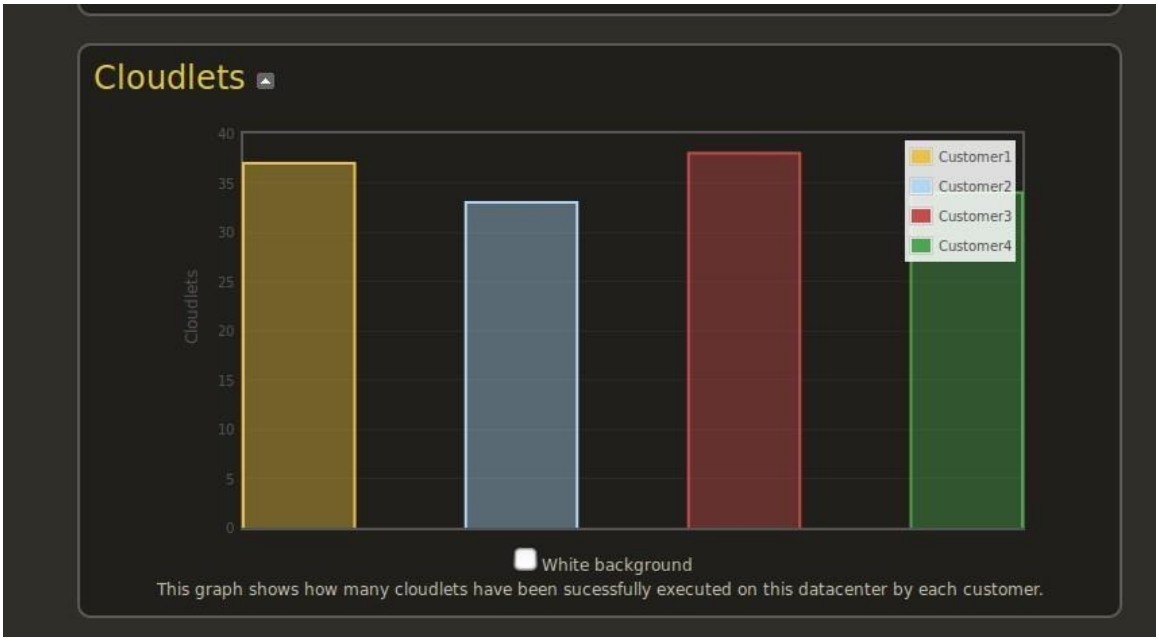

Figure 15: Graph shows how many cloudlets have been successfully executed on Datacenter-1 by each customer.

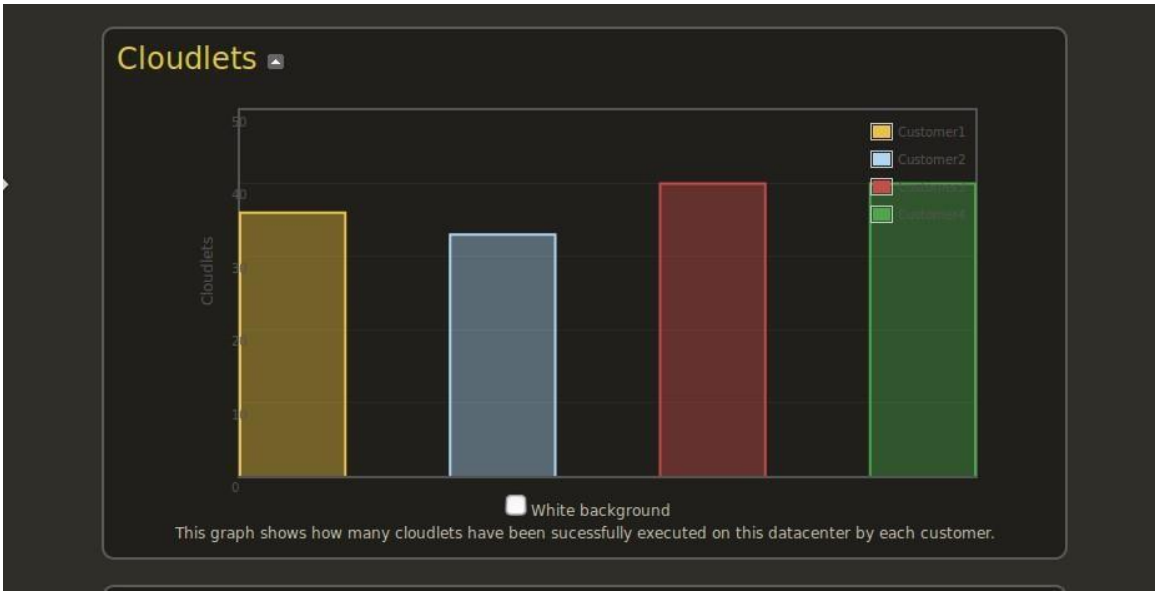

Figure 16: Graph shows how many cloudlets have been successfully executed on Datacenter-2 by each customer.

m) The costs generated by each customer based on their utilization of the datacenters.

a. **Datacenter-1:** The monetary costs generated by each customer based on their resource utilization in the data centre is depicted in the figure-17.

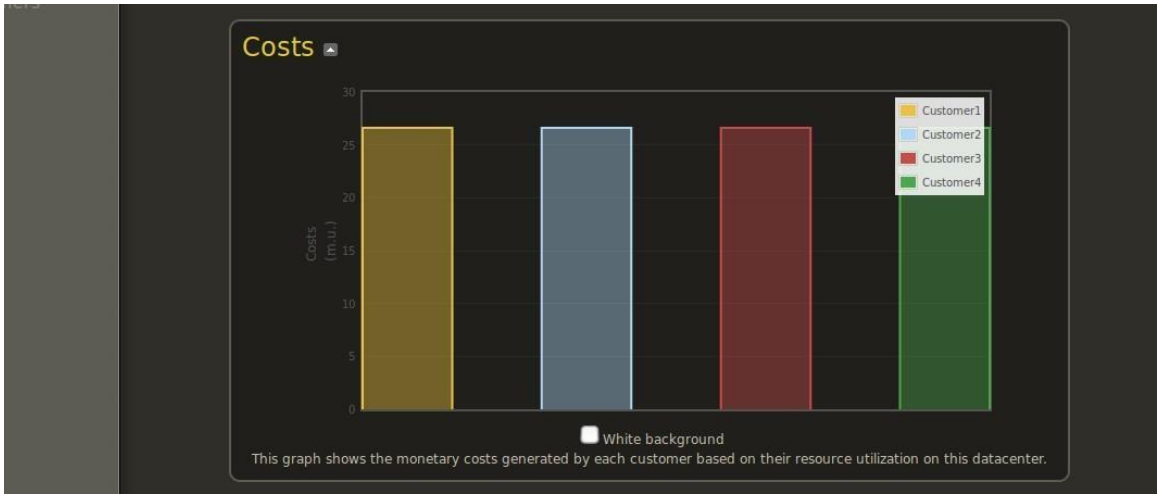

Figure 17: Graph shows the costs generated on Datacenter-1 by each customer.

b. **Datacenter-2:** The monetary costs generated by each customer based on their resource utilization in the data centre is depicted in the figure-18.

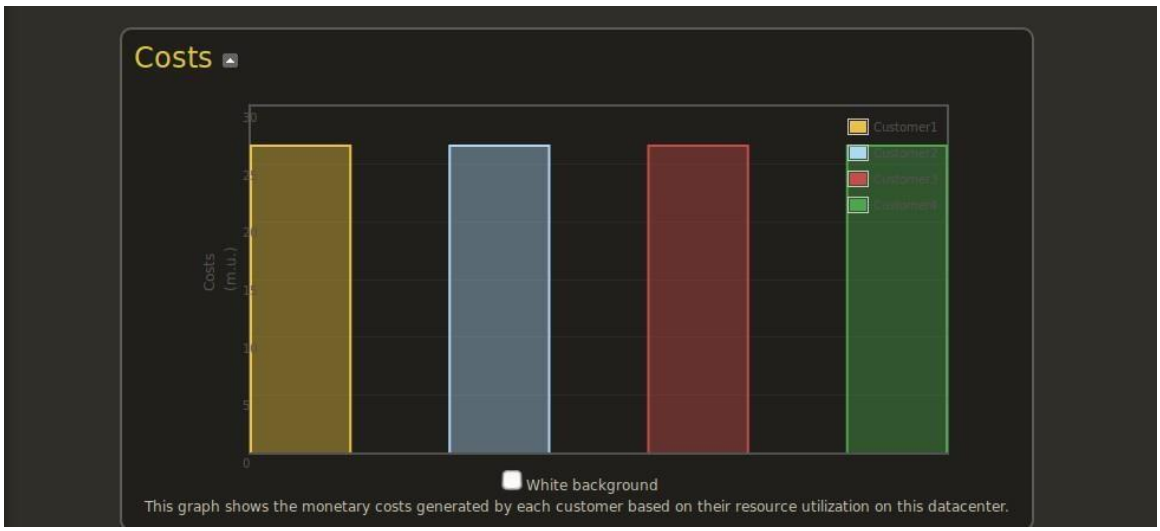

Figure 18: Graph shows the costs generated on Datacenter-2 by each customer.

n) The graphs below shows the CPU (MIPS) and RAM (MB) vs. Time for each virtual machine allocated to each customer.

Customer-1

VM-0: The resource utilization by customer-1 is depicted based on the resources like CPU and RAM utilization. The sudden drops in the figure-19 are due to the idle state at that point of time.

Figure 19: Graph shows the resource utilization of VM-0.

VM-1: The resource utilization by customer-1 is depicted based on the resources like CPU and RAM utilization. The sudden drops in the figure-20 are due to the idle state at that point in time.

Figure 20: Graph shows the resource utilization of VM-1.

VM-0: The resource utilization by customer-2 is depicted based on the resources like CPU and RAM utilization. The sudden drops in the figure-21 are due to the idle state at that point of time.

Customer-2

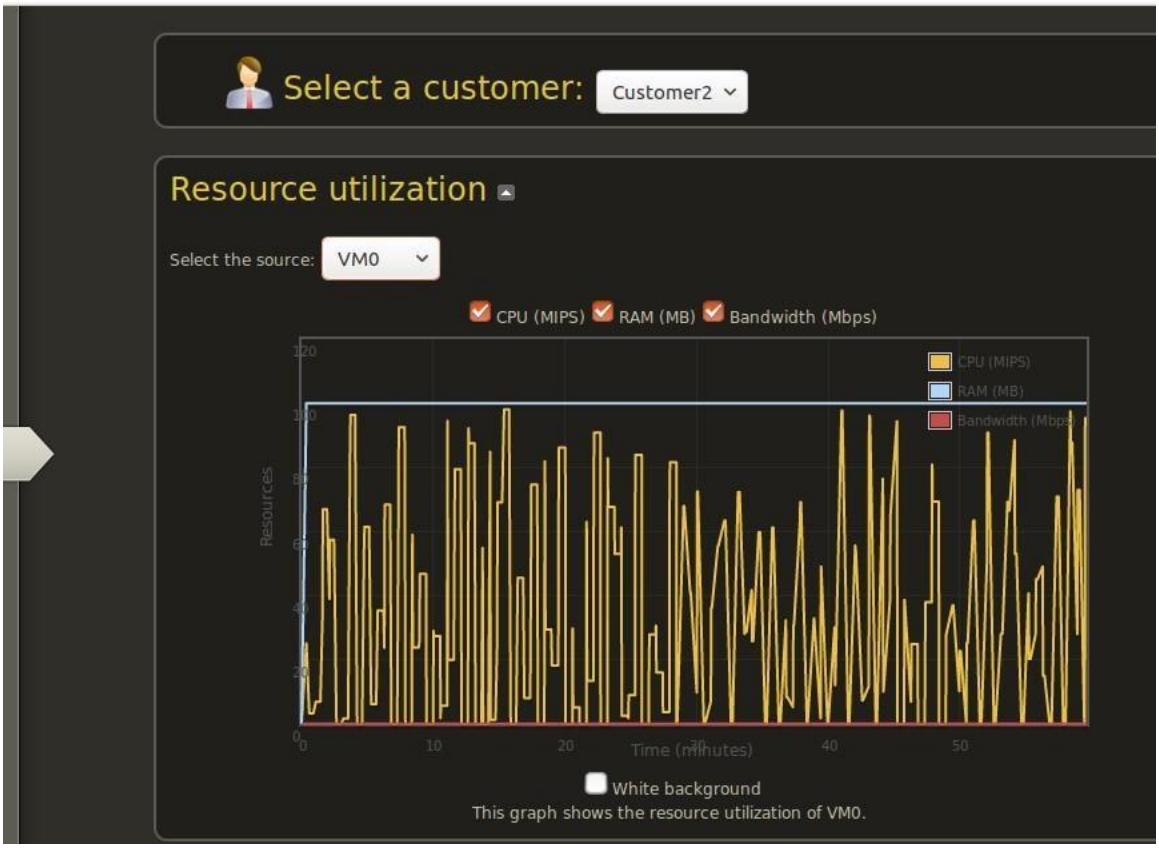

Figure 21: Graph shows the resource utilization of VM-0.

VM-1: The resource utilization by customer-2 is depicted based on the resources like CPU and RAM utilization. The sudden drops in the figure-22 are due to the idle state at that point of time.

Figure 22: Graph shows the resource utilization of VM-1.

Customer-3

VM-0: The resource utilization by customer-3 is depicted based on the resources like CPU and RAM utilization. The sudden drops in the figure-23 are due to the idle state during that point of time.

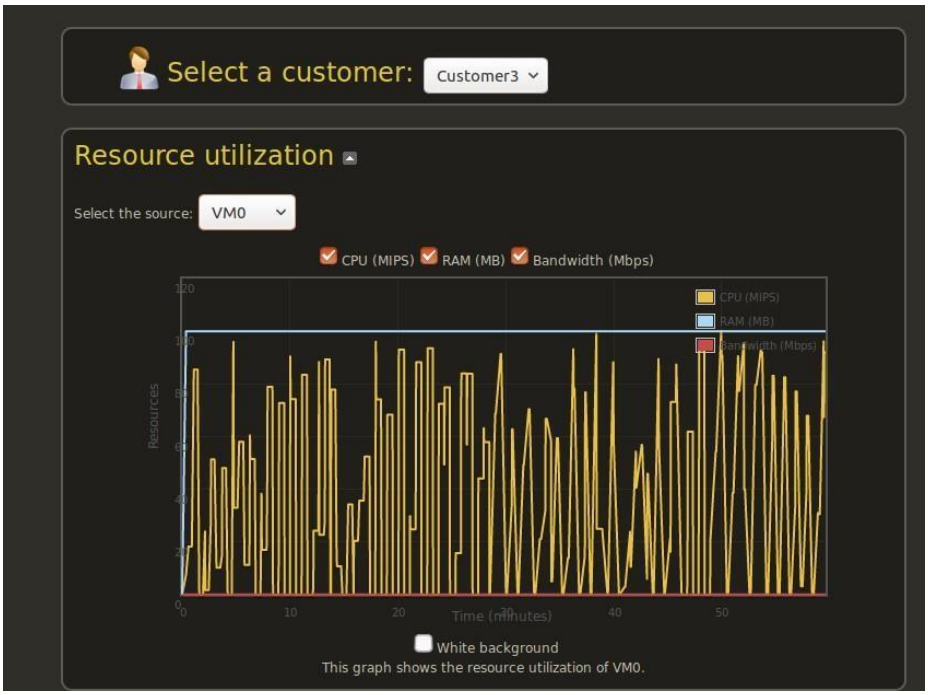

Figure 23: Graph shows the resource utilization of VM-0.

VM-1: The resource utilization by customer-3 is depicted based on the resources like CPU and RAM utilization. The sudden drops in the figure-24 are due to the idle state during that point of time.

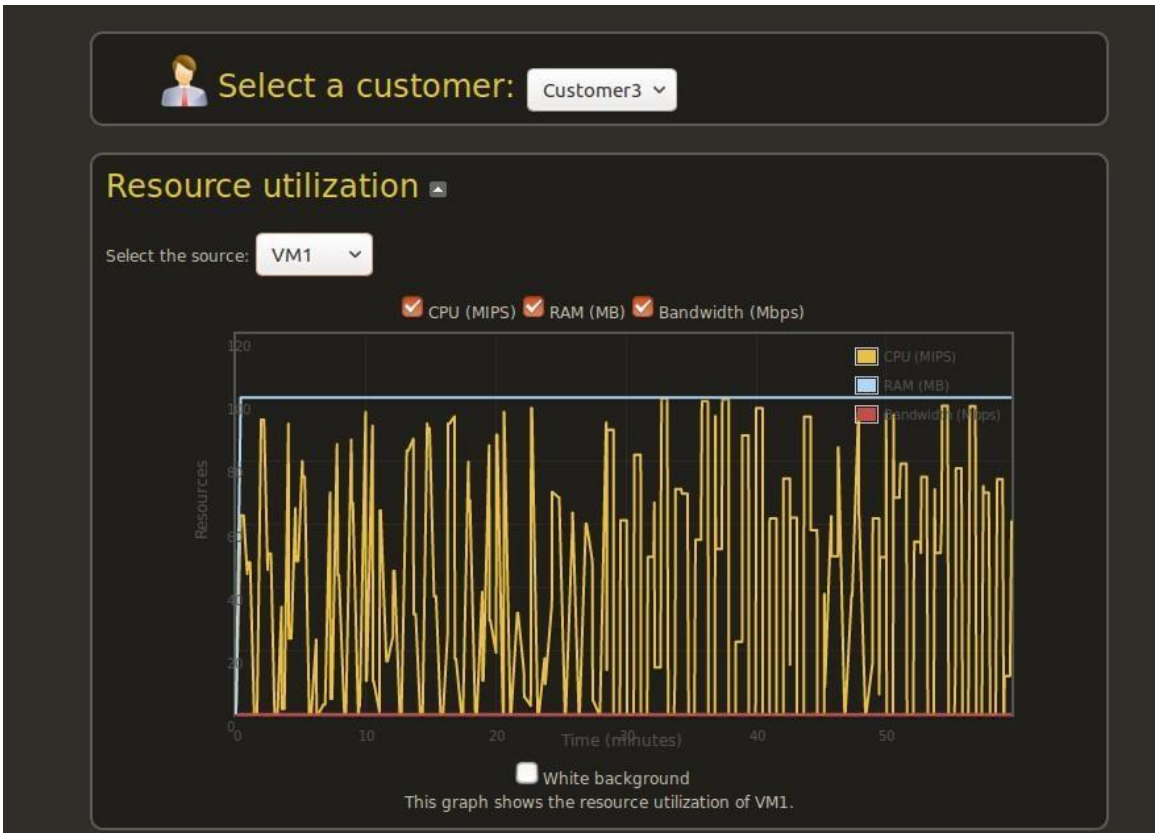

Figure 24: Graph shows the resource utilization of VM-1.

VM-0: The resource utilization by customer-4 resources like CPU and RAM utilization. The sudden drops in the figure-25 are due to the idle state during that point of time.

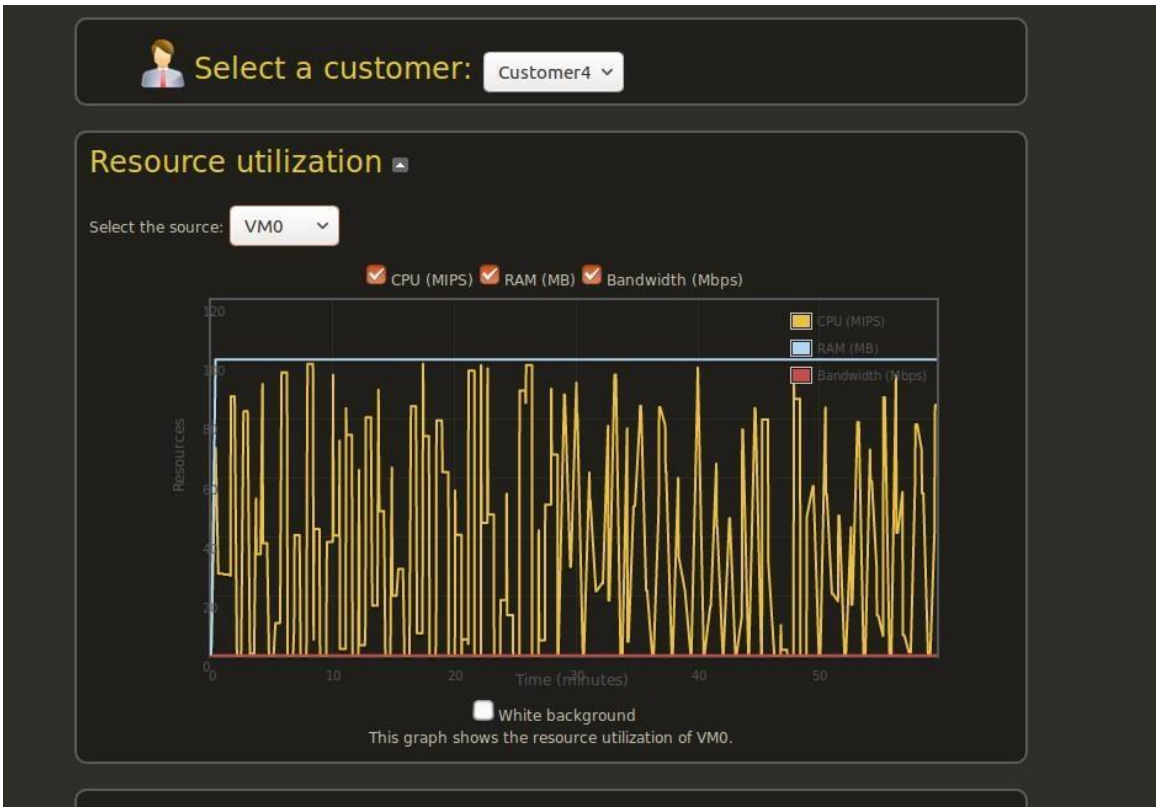

Figure 25: Graph shows the resource utilization of VM-0.

VM-1: The resource utilization by customer-4 is depicted based on the resources like CPU and RAM utilization. The sudden drops in the figure-26 are due to the idle state at that point of time.

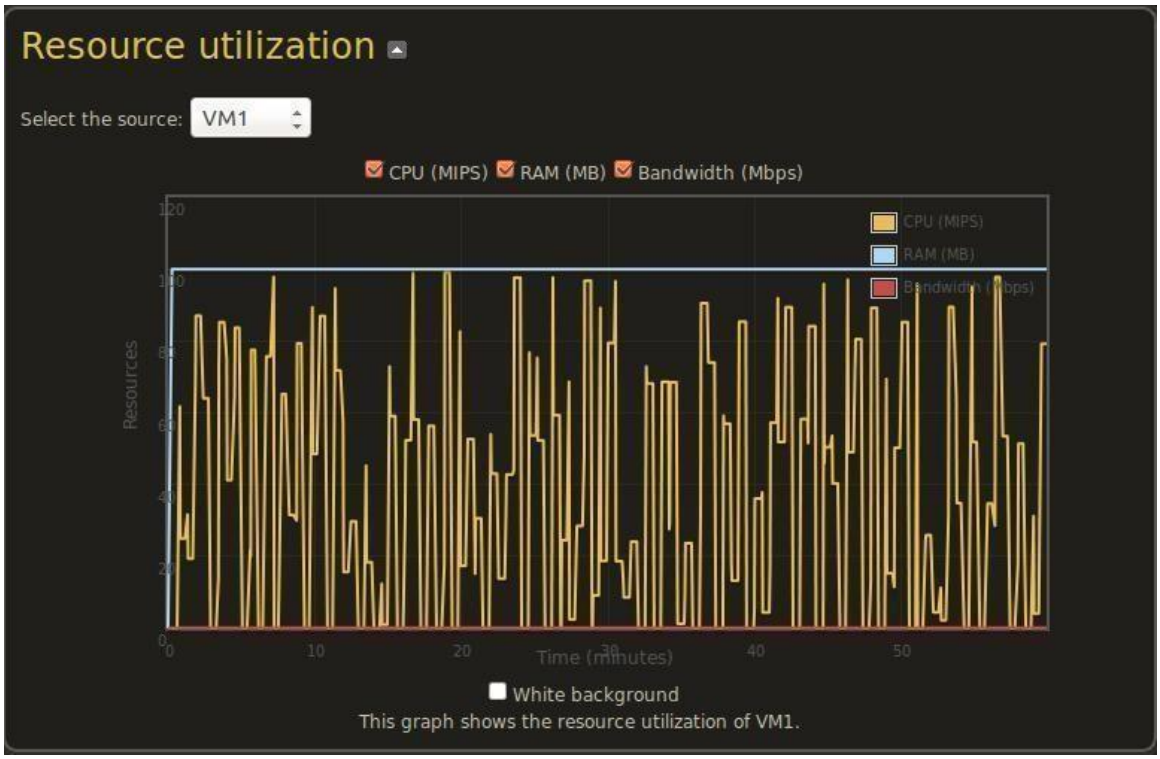

Figure 26: Graph shows the resource utilization of VM-1.

n) The number of successfully executed cloudlets on each VM.

o) The debt for each user and each data centre at the end of the simulation is the same with 26.6.

p) The start times and the finish times of the cloudlets for each resource and virtual machine.

Part-B:

a) The simulation of the configured environment has finished in 9 seconds and the simulation time for the process is 3600 seconds.

b) We observe that two data centres, Datacenter1 and Datacenter2 are successfully created with a single host in it.

c) We observe that there are eight customers, each of them having two virtual machines and are using the resources provided by hosts in data centres.

d) The number of successfully executed cloudlets on each data centre for each customer is as shown in figure-27 and figure-28.

- **Datacenter 1:**

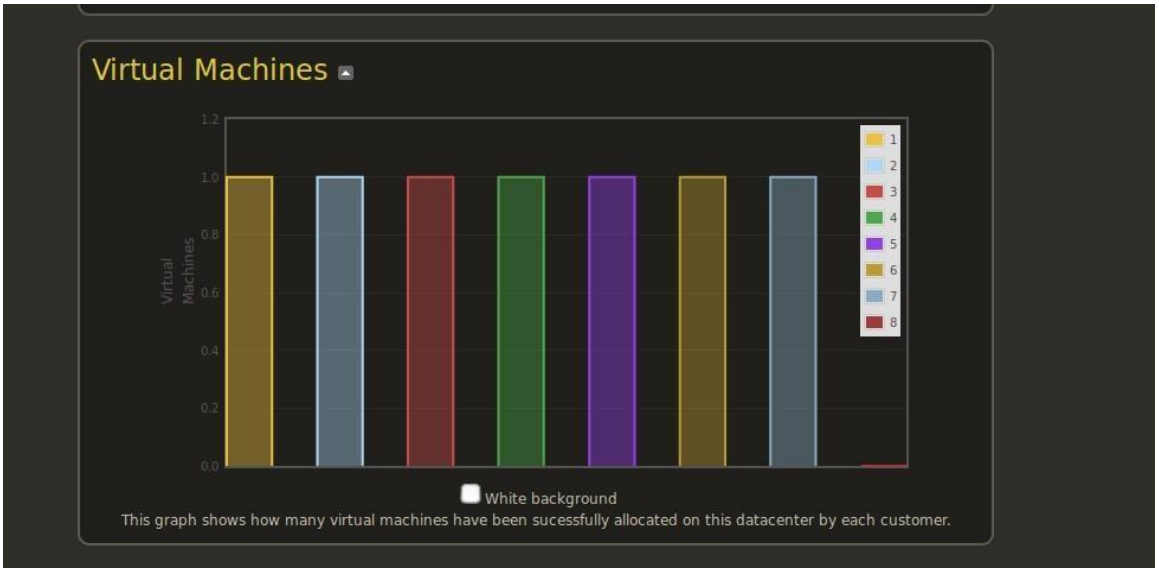

Figure 27: Graph shows how many VMs have been successfully allocated on the Datacenter-1 by each customer.

- **Datacenter 2:**

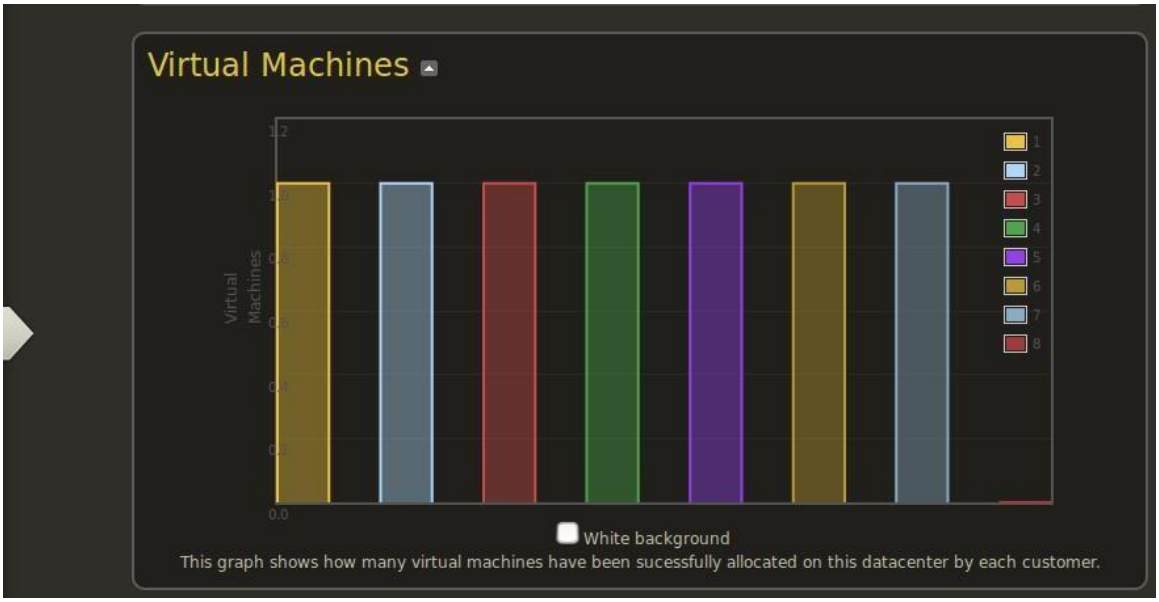

Figure 28: Graph shows how many VMs have been successfully allocated on the Datacenter-1 by each customer.

e) The costs are generated by each customer based on their utilization of the data centre.

- **Datacenter** 1: The costs are same for all of the customers, except for customer2 and customer5. This could be due to the less allocation of resources when compared to others. The costs for customer5 is more than that of customer2 because the virtual machine RAM for all the customers is 512 MB, for customer5 it is 128 and customer2 it is 64. The figure-29 shows in detail information.

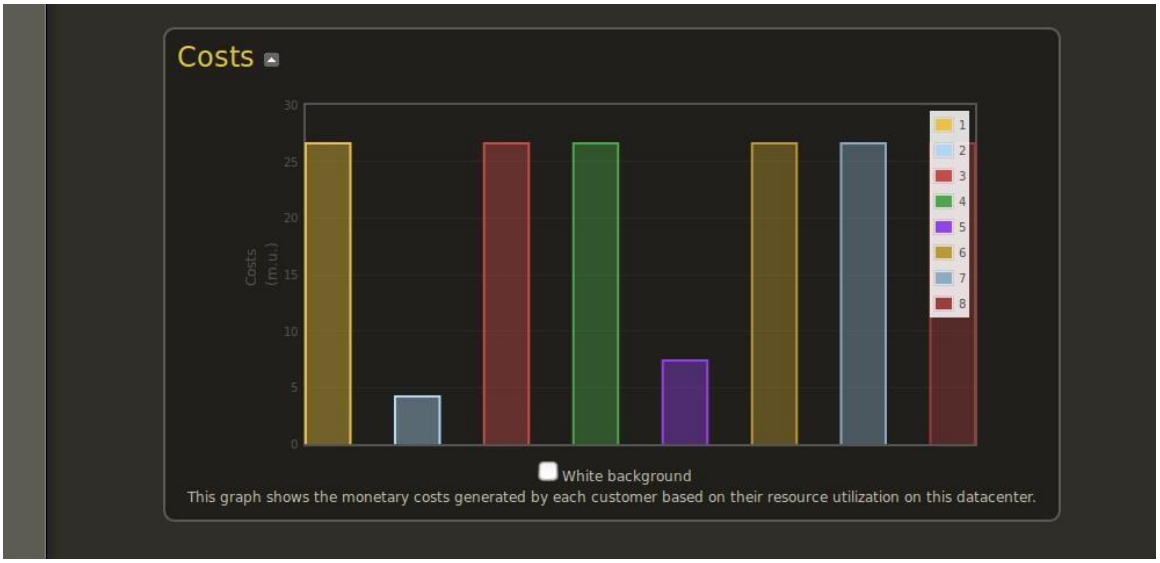

Figure 29: Graph shows the costs generated by each customer on the Datacenter-1.

- **Datacenter 2:** Similarly, the costs are same for all of the customers, except for customer2 and customer5. This can be for the same reason as mentioned above. The figure-30 shows in detail information.

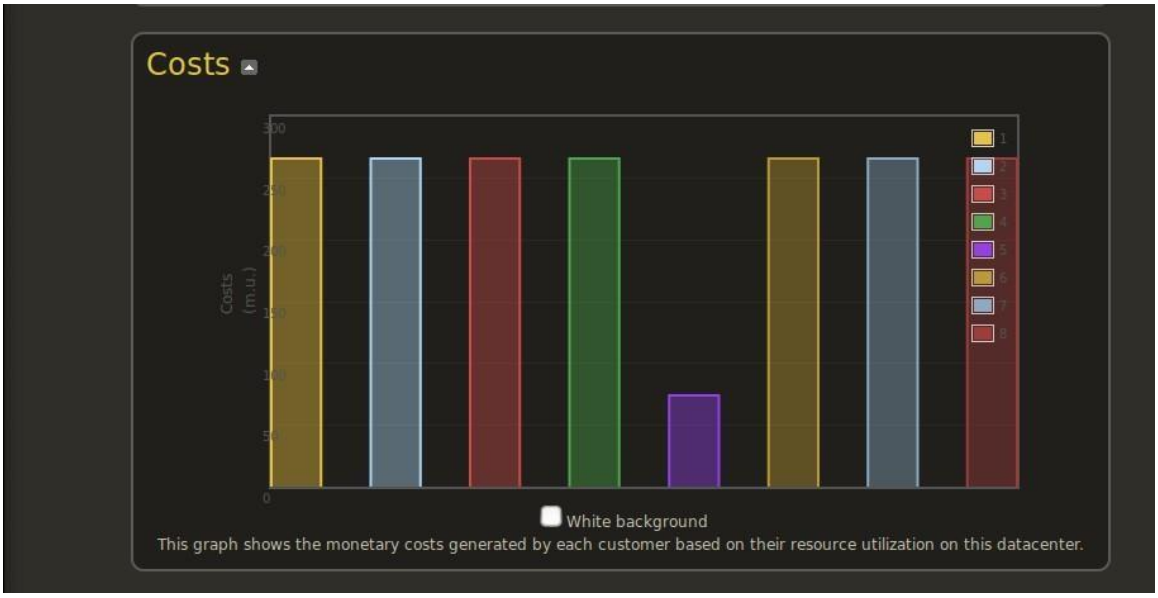

Figure 30: Graph shows the costs generated by each customer on the Datacenter-2.

f) The number of cloudlets successfully executed in Datacenter1 for each of the customers on an average is more than 35 and for customer8 is zero. This is because the bandwidth is very low and the latency is very high for customer8.

g) Similarly, the number of cloudlets successfully executed in Datacenter2 for each of the customers is more than 35 and for customer8 is zero. This can be because of the same reason as mentioned above.

h) The number of successfully executed cloudlets on each virtual machine by each of the customers is more than 35, except for customer8 for which the cloudlets executed are zero.

i) The debt of each user for each data centre is shown in figure-31. The debts of Datacenter-2 increased by 10 times of Datacenter-1 because the costs for processing, storage, memory and bandwidth in Datacenter-2 are 10 times of Datacenter-1. So the debts increased respectively.

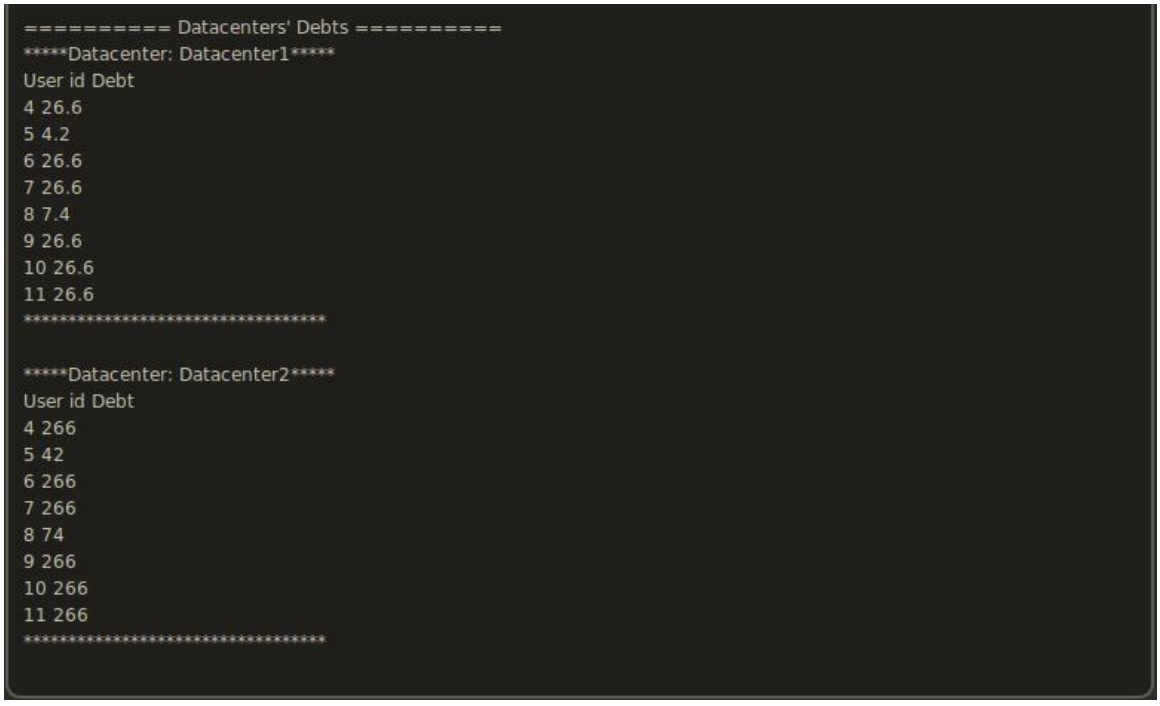

Figure 31: Debts of Datacenter-1 and Datacenter-2

Part -C:

a) There are 12 virtual machines running in Datacenter1 and Datacenter2.

b) When compared to Part-B configuration, the costs of customer4 increased in both Datacenter1 and Datacenter2 because the resources allocation is more (i.e., the number of virtual machines increased from 2 to 10) in Part-C.

c) From the date of start and finish times of the cloudlets, we observe that the scheduling of cloudlets on the virtual machines of Customer 4 follows *Stochastic Process.*

d) The allocation of each VM by a cloudlet overtime on the two datacenters is described below.

- **Datacenter-1:** The number of successfully executed cloudlets by each customer on average is more than 35 and for customer 8 is zero due to less bandwidth and high latency. The number of successfully executed cloudlets for customer-4 is more because it has a number of virtual machines allocated. The figure-32 depicts the number of cloudlets executed by each customer.

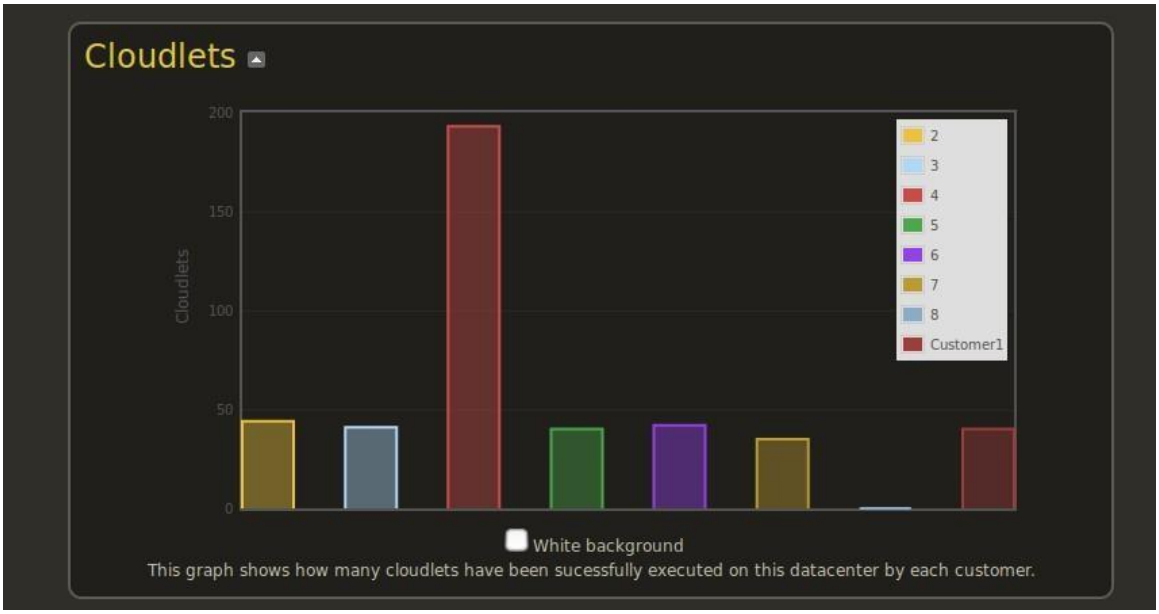

Figure 32: Graph shows how many cloudlets have been successfully executed on Datacenter-1 by each customer

- **Datacenter-2:** The number of successfully executed cloudlets by each customer on average is more than 35 and for customer 8 is zero due to less bandwidth and high latency. The number of successfully executed cloudlets for customer-4 is more because it has a number of virtual machines allocated.

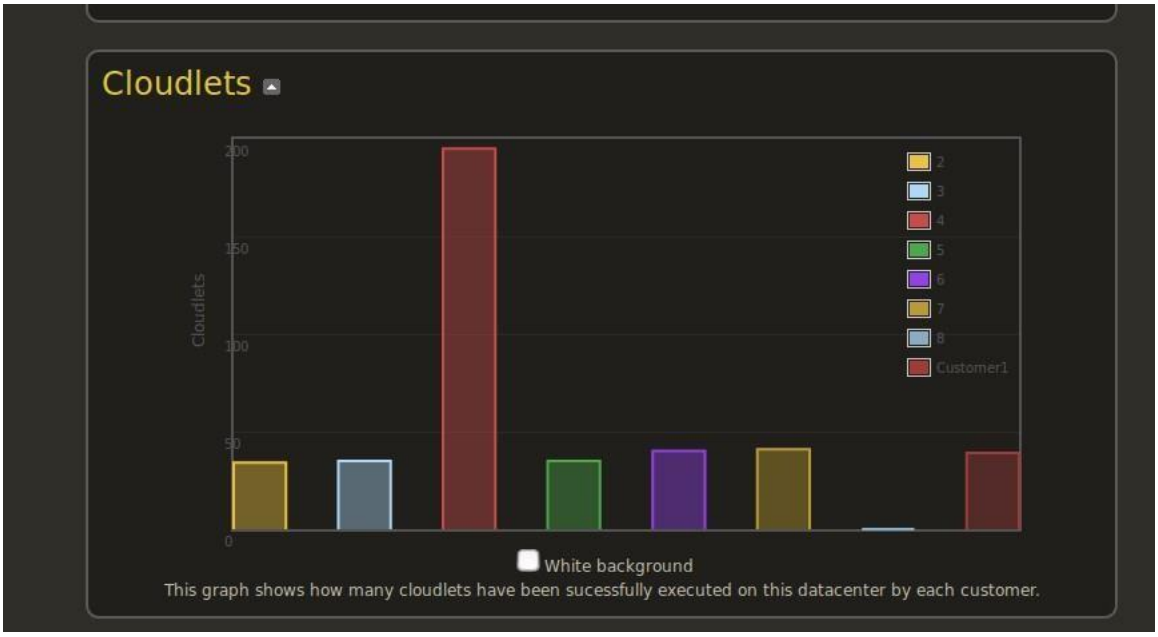

Figure 33: Graph shows how many cloudlets have been successfully executed on Datacenter-1 by each customer

Conclusion:

From the above-obtained results, we conclude that, as the utilization of CPU is set to stochastic, the cloudlets on each datacentre are executed randomly with respect to time and the sudden drops in the graphs show that the datacentres are idle at that particular time. In the first scenario, Part A, we see that the resource utilization and power consumed are zero when the datacentres are idle. In the second scenario, Part B, we observe that the debts of Datacenter-2 increased by 10 times of Datacenter-1 because the costs for processing, storage, memory and bandwidth in Datacenter-2 are 10 times of Datacenter-1. Also, we observed that the cloudlets allocated are zero for low bandwidth and high latency. In the third scenario, Part C, with the increase in the number of VM allocated there is an increase in the debts of the customer and cloudlets allocated also.

Future Work:

1. Observing the type of Distribution for the CPU (MIPS) & RAM vs time plot for the data centre. Since the distribution varies according to the resources allocated, determining the type of distribution for this kind of resource allocation helps to properly allocate resources to the customers in the data centres.
2. Also one needs to explore about stochastic processes and their implementations in various scenarios for a better understanding.

References:

[1] "The CLOUDS Lab: Flagship Projects - Gridbus and Cloudbus." [Online]. Available: http://www.cloudbus.org/cloudsim/. [Accessed: 22-Jan-2018].

[2] Ip. S. Ranbhise and K. K. Joshi, "Simulation and Analysis of Cloud Environment," *Simulation*, vol. 2, no. 4, 2014.

[3] T. T. Sá, R. N. Calheiros, and D. G. Gomes, "CloudReports: An Extensible Simulation Tool for EnergyAware Cloud Computing Environments," in *Cloud Computing*, Z. Mahmood, Ed. Springer International Publishing, 2014, pp. 127–142.