National University of Computer and Emerging Sciences, Lahore Campus

Course:	Introduction to Cloud Computing	Course Code:	CS-499
Program:	BS (Computer Science)	Semester:	Spring 2018
Duration:	60 Minutes	Total Marks:	40
Paper Date:	27-Feb-18	Weight	15%
Section:	A and B	Page(s):	1
Exam:	Midterm	Reg. No.	

Instruction/Notes:

- Answer all the questions
- Read the questions carefully before answering them
- If you are certain that something is unclear, make a *reasonable assumption*, mention it and answer the question
- All questions carry equal marks

---

1. An organization produces 500 GB of data that needs to be processed. The organization owns 20 servers that are each equivalent to an Amazon EC2 instance. An Amazon EC2 instance or the equivalent local server would take 2 hours per GB to process the data. The cost to upload data to Amazon's cloud is $0.10 / GB and the organization has a consistent 20 MB/s connection to the Amazon cloud. The organization needs to decide if it should process the data locally or on the Amazon cloud. Which option is faster? Show how long each option takes.

Solution:

How long will it take to upload the 500 GB data to the cloud?

20 MB take 1 s

1 MB takes 1 / 20 s

1 GB takes 1024 / 20 s

500 GB takes 500 * 1024 / 20 s

500 Gb takes 25,600 s = 7.11 hours

1 GB is computed in 2 hours on one machine

500 GB is computed in 1000 hours on one machine

On 20 machines, 500 GB can be computed in 1000 / 20 = 50 hours

Supposing equal number of instances is used in the cloud, the computation takes 7.11 + 50 hours = 57.11 hours, whereas local computation takes only 50 hours. For the cloud to be quicker:

1000 / n + 7.11 < 50

1000 / n < 42.89

n > 1000 / 42.89 = 23.31

Thus, for the cloud to be faster, at least 24 EC2 instances must be used for approximately 50 hours.

2. A friend owes me some money (with friends like these….), don't ask why. He's not going to pay me back as he doesn't have cash. However, he has 20 brand new high-end servers that cost approximately what he owes me. I can take the servers and call it even. The servers can handle the expected peak workload for an app that I am about to launch. Should I be satisfied with deploying on these servers from my friend, or should I be thinking about a cloud based deployment?

Solution:

The cloud based deployment has several advantages:

- If demand surges, I can quickly respond by scaling out my cloud deployment.
- The redundancy for cloud based resources is provided by the cloud provider, whereas on my own deployment, I would have to manage that myself.
- In this scenario, I don't really need the full control over the hardware and software that requires own deployment.

3. Why is cloud-scale infrastructure provisioning not possible without virtualization?

Solution:

Virtualization enables pre-installed machine images including OS and any software or libraries desired for a certain environment. Thus, provisioning in the presence of virtualization takes the form of creating a virtual machine, connecting it to the machine image and starting it, all within seconds. On the other hand, manual provisioning in a data center with 10s of thousands of servers and potentially multiple machine creation requests per second can't be managed.

4. Oversubscription in a particular tier of a data center network is defined as the ratio of the downlink bandwidth to the uplink bandwidth. For example, if each access switch is connected to 5 servers using 1 Gbps connections and has a 4 Gbps uplink to the aggregation tier, then the aggregation tier has an oversubscription of 5:4. You have to interconnect 24 servers at 1 Gbps each, using 12 port Ethernet switches and the aggregation tier can have an oversubscription of 2:1. Design a two tier network using as few switches as possible to achieve the above target.

Solution:

One 12 port switch at the aggregation tier. Three 12 port switches at the access tier. 8x servers connected to each access switch at 1 Gbps. Each access switch is connected to the aggregation switch using 4x 1 Gps links. Thus, each access switch promises 8 Gbps to the servers while having only 4 Gbps bandwidth to the aggregation tier, making an oversubscription of 8:4 = 2:1

5. What technical challenge is associated with using a fat-tree topology in the data center Ethernet? Suggest two solutions to this challenge.

Solution:

Ethernet computes a spanning tree by pruning several links in a topology, such as a fat tree, that has loops. This would reduce the topology's useful bandwidth significantly. Some solutions to this problem are:

- Use tunneling such as QinQ
- Use some layer 3 technology such as MPLS
- Use some other layer 2 technology

6. How is malicious intermediary threat type different from eavesdropping?

Solution:

Eavesdropping passively observes the packets and copies them. It only affects packet confidentiality. Malicious intermediary also modifies the packets before forwarding them to the destination. It affects both packet confidentiality and integrity.

7. A security bug is discovered in the MySQL Python API. Who (the cloud provider or the consumer) is responsible to fix this in IaaS, PaaS and Saas settings?

Solution:

In IaaS, it is the consumer's responsibility

In PaaS and SaaS it is the provider's responsibility

8. Would a VM acquired on an IaaS model provide consistent performance? Why or why not?

Solution:

Each VM shares resources with other VMs on the same physical hardware. Every VM's workload is likely to vary independently with time. Thus, providing a fixed slice of computing resources to each VM is not possible. This means that each VMs performance is likely to vary over time.