Roll No. _____ Section _____

National University of Computer and Emerging Sciences, Lahore Campus

	Course:	**Introduction to Cloud Computing**	**Course Code:**	**CS-499**
	Program:	**Computer Science**	**Semester:**	**Fall 2017**
	Duration:	**180 Minutes**	**Total Marks:**	**110**
	Paper Date:	**18-Dec-17**	**Weight**	**40%**
	Section:	**N/A**	**Page(s):**	**3**
	Exam:	**Final**	**Reg. No.**	

Instruction/Notes: Please answer all the questions (including multiple choice) on the answer book provided.

1- The automated ability of a cloud to transparently scale IT resources, as required in response to runtime conditions or as pre-determined by the cloud consumer or cloud provider is called:

a-Resiliency

b-Measured Usage

c-Elasticity (Correct)

d-Multitenancy

Marks 2

2-An architectural model based on a system of predefined scaling conditions that trigger the dynamic allocation of IT resources from resource pools is called:

a- Resource Pooling Architecture

b- Dynamic Scalability Architecture (Correct)

c-Workload Distribution Architecture

d-Redundant Storage Architecture

Marks 2

3- We get the module availability with which of the following formula:

MTTF: mean time to failure

MTTR: mean time to repair

a- (MTTF +MTTR)

b- (MTTF +MTTR)/ MTTF

c- MTTF / (MTTF +MTTR) (Correct)

Roll No. _____ Section _____

Marks 2

4-Google uses (select the correct answer)

a- Global distributed File System(GFS) (Correct)

b- Network Attached Storage (NAS)

Marks 2

5- Which of the following consumes the most power in Warehouse scale computing:

a- DRAM

b- Disks

c- Networking

d-CPUs (Correct)

Marks 2

6- A cloud consumer that requests the installation, configuration, and updating of twentyfive Windows servers with several applications requires that half of the applications be identical installations, while the other half be customized. Each operating system deployment can take up to 30 minutes, followed by additional time for security patches and operating system updates that require server rebooting. The applications finally need to be deployed and configured. Using a manual or semi-automated approach requires excessive amounts of time, and introduces a probability of human error that increases with each installation.

- Explain in detail how you could provide a cloud based system to automate the above mentioned requirement.
- Give details of the cloud architecture you would implement.
- Draw a diagram to help explain your answer.

Marks 30

ANS:

A conventional provisioning process can involve a number of tasks that are traditionally completed

manually by administrators and technology experts that prepare the requested IT resources as per prepackaged

specifications or custom client requests. In cloud environments, where higher volumes of

customers are serviced and where the average customer requests higher volumes of IT resources,

manual provisioning processes are inadequate and can even lead to unreasonable risk due to human

error and inefficient response times.

The *rapid provisioning architecture* establishes a system that automates the provisioning of a wide range of IT resources, either individually or as a collective. The underlying technology architecture for rapid IT resource provisioning can be sophisticated and complex, and relies on a system comprised of an automated provisioning program, rapid provisioning engine, and scripts and templates for on-demand provisioning.

Server Images – These images are similar to virtual server templates, but are used to provision physical servers.

• *Application Packages* – Collections of applications and other software that are packaged for automated deployment.

• *Application Packager* – The software used to create application packages.

• *Custom Scripts* – Scripts that automate administrative tasks, as part of an intelligent automation engine.

• *Sequence Manager* – A program that organizes sequences of automated provisioning tasks.

• *Sequence Logger* – A component that logs the execution of automated provisioning task sequences.

• *Operating System Baseline* – A configuration template that is applied after the operating system is installed, to quickly prepare it for usage.

• *Application Configuration Baseline* – A configuration template with the settings and environmental parameters that are needed to prepare new applications for use.

• *Deployment Data Store* – The repository that stores virtual images, templates, scripts, baseline configurations, and other related data.

The following step-by-step description helps provide some insight into the inner workings of a rapid provisioning engine, involving a number of the previously listed system components:

1. A cloud consumer requests a new server through the self-service portal.

2. The sequence manager forwards the request to the deployment engine for the preparation of an

operating system.

3. The deployment engine uses the virtual server templates for provisioning if the request is for a virtual server. Otherwise, the deployment engine sends the request to provision a physical server.

4. The pre-defined image for the requested type of operating system is used for the provisioning of the operating system, if available. Otherwise, the regular deployment process is executed to install the operating system.

5. The deployment engine informs the sequence manager when the operating system is ready.

6. The sequence manager updates and sends the logs to the sequence logger for storage.

7. The sequence manager requests that the deployment engine apply the operating system baseline to the provisioned operating system.

8. The deployment engine applies the requested operating system baseline.

9. The deployment engine informs the sequence manager that the operating system baseline has been applied.

10. The sequence manager updates and sends the logs of completed steps to the sequence logger for storage.

11. The sequence manager requests that the deployment engine install the applications.

12. The deployment engine deploys the applications on the provisioned server.

13. The deployment engine informs the sequence manager that the applications have been installed.

14. The sequence manager updates and sends the logs of completed steps to the sequence logger for storage.

15. The sequence manager requests that the deployment engine apply the application's configuration baseline.

16. The deployment engine applies the configuration baseline.

17. The deployment engine informs the sequence manager that the configuration baseline has been applied.

18. The sequence manager updates and sends the logs of completed steps to the sequence logger for storage.

7- Write code in Apex to create an account which should have the following name: 'Habib Bank Ltd'.

Write SOQL (Salesforce Object Query Language) which will return the Account you just created with the name 'Habib Bank Ltd'. You can use the name in the SOQL.

Marks 20

ANS:

Account creation:

```
Account a = new Account(name = 'Acme');
insert a;
```

Query Account:

```
List<Account> aa = [SELECT Id, Name FROM Account WHERE Name = 'Habib Bank Ltd'];
```

8- Describe and explain what the following mechanisms are and how are they used in cloud computing:

a- Hypervisor

b- Load Balancer

c- Multi-Device Broker.

Marks 30

ANS:

Hypervisor:

The *hypervisor* mechanism is a fundamental part of virtualization infrastructure that is primarily used to generate virtual server instances of a physical server. A hypervisor is generally limited to one physical server and can therefore only create virtual images of that server. Similarly, a hypervisor can only assign virtual servers it generates to resource pools that reside on the same underlying physical server. A hypervisor has limited virtual server management features, such as increasing the virtual server's capacity or shutting it down. The VIM provides a range of features for administering multiple hypervisors across physical servers.

Load Balancer

A common approach to horizontal scaling is to balance a workload across two or more IT resources to increase performance and capacity beyond what a single IT resource can provide. The *load balancer* mechanism is a runtime agent with logic fundamentally based on this premise.

Beyond simple division of labor algorithms (Figure 8.5), load balancers can perform a range of specialized runtime workload distribution functions that include:

• *Asymmetric Distribution* – larger workloads are issued to IT resources with higher processing capacities

• *Workload Prioritization* – workloads are scheduled, queued, discarded, and distributed workloads according to their priority levels.

• *Content-Aware Distribution* – requests are distributed to different IT resources as dictated by the request content

Multi-device Broker:

An individual cloud service may need to be accessed by a range of cloud service consumers differentiated by their hosting hardware devices and/or communication requirements. To overcome incompatibilities between a cloud service and a disparate cloud service consumer, mapping logic needs to be created to transform (or convert) information that is exchanged at runtime.

The *multi-device broker* mechanism is used to facilitate runtime data transformation so as to make a cloud service accessible to a wider range of cloud service consumer programs and devices.

9-What are Warehouse Scale Computers? Give an Architectural overview of WSC.

Marks 10

ANS:

Warehouse Scale Computers:

they belong to a single organization, use a relatively homogeneous hardware and system software platform, and share a common systems management layer. Often, much of the application, middleware,

and system software is built in-house compared to the predominance of third-party software running in conventional datacenters. Most importantly, WSCs run a smaller number of very large applications (or Internet services), and the common resource management infrastructure allows significant deployment flexibility.1 The requirements of homogeneity, single-organization control, and enhanced focus on cost efficiency motivate designers to take new approaches in constructing and operating these systems.

Architectural Overview:

STORAGE

Disk drives or Flash devices are connected directly to each individual server and managed by a

global distributed file system (such as Google's GFS [58]) or they can be part of Network Attached

Storage (NAS) devices directly connected to the cluster-level switching fabric. A NAS tends to be

a simpler solution to deploy initially because it allows some of the data management responsibilities

to be outsourced to a NAS appliance vendor. Keeping storage separate from computing nodes also

makes it easier to enforce quality of service guarantees since the NAS runs no compute jobs besides

the storage server. In contrast, attaching disks directly to compute nodes can reduce hardware

costs

NETWORKING FABRIC

Choosing a networking fabric for WSCs involves a trade-off between speed, scale, and cost. As of

this writing, 1-Gbps Ethernet switches with up to 48 ports are commodity components, costing

less than $30/Gbps per server to connect a single rack (including switch port, cable, and server

NIC). As a result, bandwidth within a rack of servers tends to be homogeneous. However, network

switches with high port counts, which are needed to tie together WSC clusters, have a much different

price structure and are more than ten times more expensive (per port) than commodity rack

switches. As a rule of thumb, a switch that has 10 times the bisection bandwidth often costs about

100 times as much.

STORAGE HIERARCHY

A server consists

of a number of processor sockets, each with a multicore CPU and its internal cache hierarchy,

local shared and coherent DRAM, a number of directly attached disk drives and/or Flash-based

solid state drives. The DRAM and disk/flash resources within the rack are accessible through

the first-level rack switches (assuming some sort of remote procedure call API to them), and all

resources in all racks are accessible via the cluster-level switch. The relative balance of various resources

depends on the needs of target applications. Flash-based storage in particular is a recent

addition to the WSC storage picture and its use in the real world varies widely. The configuration

used below assumes an order of magnitude less Flash capacity than traditional spinning media since

that is roughly the relative cost per byte difference between these two technologies.

QUANTIFYING LATENCY, BANDWIDTH, AND CAPACITY

For

illustration we assume a system with 2,400 servers, each with 16 GB of DRAM and four 2 TB

disk drives (we leave Flash out of this scenario due to its highly varying deployment models, and

address Flash later in this section). Each group of 80 servers is connected through a 1-Gbps link to

a rack-level switch that has an additional eight 1-Gbps ports used for connecting the rack to the

cluster-level switch (an oversubscription factor of 5). Network latency numbers assume a TCP-IP

transport, and networking bandwidth values assume that each server behind an oversubscribed set

of uplinks is using its fair share of the available cluster-level bandwidth. For disks, we show typical

commodity disk drive (SATA) latencies and transfer rates.

10-Information Risk Management is the act of aligning exposure to risk and capability of managing it with the risk tolerance of the data owner. Give five recommendation for 'Information Risk Management'

Marks 10

ANS:

√ Adopt a risk management framework model to evaluate IRM, and a maturity model to assess the effectiveness of your IRM model.

√ Establish appropriate contractual requirements and technology controls to collect necessary data to inform information risk decisions (e.g., information usage, access controls, security controls, location, etc.).

√ Adopt a process for determining risk exposure before developing requirements for a Cloud Computing project. Although the categories of information required to understand exposure and management capability are general, the actual evidential metrics gathered are specific to the nature of the cloud computing SPI model and what can be feasibly gathered in terms of the service.

√ When utilizing SaaS, the overwhelming majority of information will have to be provided by the service provider. Organizations should structure analytical information gathering processes into contractual obligations of the SaaS service.

√ When utilizing PaaS, build in information gathering as per SaaS above, but where possible include the ability to deploy and gather information from controls as well as creating contractual provisions to test the effectiveness of those controls.

√ When utilizing an IaaS service provider, build information transparency into contract

language for information required by risk analysis.

√ Cloud service providers should include metrics and controls to assist customers in

implementing their Information Risk Management requirements.