**Level 3 - Questions**

**Load Balancing**

What are the main functions of a public load balancer?

1. Balancing the incoming traffic load between several front-end web servers that are performing the same function
2. Becomes the entry point to end-users with a single public IP
3. Reduces the load on individual VM and also prevents that a specific VM will become a single point of failure
4. All answers are correct

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A load balancer is a critical building block for scaling out the cloud infrastructure according to the new demand.

**YES**/NO

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

An Application Load Balancer relies only on network layer information for making routing decisions. Meaning, the source, and destination IP as well as the TCP\UDP port numbers.

YES/**NO (it is a Network Load Balancer)**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

An Application Load Balancer can perform content-based traffic routing, and it is mainly used to distribute traffic for web applications.

**YES**/NO

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A load balancer is used in two main locations – facing the internet, which is called an **internal load balancer** or inside our cloud deployment, and in that case it is called a public **load balancer**.

YES/**NO**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Azure Load Balancer is a virtual load balancing service provided by Microsoft Azure, and it is a network load balancer, meaning it is using the networking transport layer (Layer 3 & 4) for load-balancing decisions

**YES**/NO

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**As part of load balancer configuration, we need to configure a backend pool,** which is the group of VMs that are registered to that LB. It is a list of IP addresses of the virtual network interfaces (NICs) that are connected to the load balancer.

**YES**/NO

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A load balancer rule is used to define how traffic is distributed to the VMs. You define the front-end IP configuration for the incoming traffic and the back-end IP pool to receive the traffic, along with the required source and destination port.

**YES**/NO

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A health probe is used to monitor the health of the load balancer, so the load balancer will be able to dynamically adds or removes VMs from the load balancer rotation based on their response to health checks.

YES/**NO (**A health probe is used to monitor the health of backend VMs)

**Azure Availability Sets**

**Resiliency** is the ability of a system, our application to recover from failures and continue to function and stay operational. When we design a system, we should analyze the resiliency of that system while facing different kinds of failure scenarios.

Yes/**No**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A key principle to design a system that will achieve high availability will be to avoid or eliminate any **single point of failure.** A cloud system is based on multiple components, multiple tiers. And as part of the design process, we need to take into account **all kinds of failures** that may happen in a cloud computing environment.

Yes/No

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The scope of a failure in a cloud environment can be related to:

1. A local hardware issue within the datacenter
2. A complete data center
3. A complete region (multiple datacenters)
4. **All answers are correct**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Microsoft Azure Redundancy options are measured in SLAs (Service Level Agreement). SLA describes Microsoft’s commitments for downtime and network disconnection. And it is presented as a percentage. An Availability Zone will provide up to 99.95% SLA.

Yes/No (SLA describes Microsoft’s commitments for uptime and connectivity)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Availability Zones** are used to protect against some localized hardware failures in the data center. We can deploy two or more VMs in a special group called availability Zone.

Yes/No (**Availability Sets** are used to protect against some localized hardware failures in the data center).

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A fault domain is a group of underlying hardware equipment that shares a common power source and shared network connectivity, similar to a rack within an on-premises datacenter. When we create VMs within an **availability set**, the Azure platform will automatically distribute those VMs across several fault domains.

Yes/No

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Placing a few virtual machines into an availability set will also protect our application from any operating system or application-specific failures.

Yes/**No** (It is about physical hardware failures, network outages, or power interruptions in the underline infrastructure)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Microsoft Azure divided the all underline infrastructure into update domains. An update domain is a logical group of underlying hardware that can undergo maintenance or be rebooted at the same time. Meaning, if for example, five physical hosting servers are sitting in the same update domain they may be rebooted at the same time. So the concept will be to distribute our VMs in multiple update domains. In that case, Azure will always perform planned maintenance in a sequential process, meaning only one update domain is rebooted at the same time.

Yes/No

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

We can add an existing VM to a new availability set we just created.

Yes/No (VMs must be created within the availability set to make sure they're correctly distributed across the hardware)

**Azure Scale Sets**

If I have any question or I would like to suggest something, then I can share it using the course dashboard :-)

**Yes**/No

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

In a cloud environment, performance optimization can be done by dynamically adjust the capacity of the underline cloud resources so they will better match the upper layer application demand. A cloud architecture should take into account the performance and scalability of that system. Making sure different components in that system can scale as needed.

**Yes**/No

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Vertical scaling is also called **scaling out or in**. Scaling out is an act of adding more resources to a single instance.

Yes/**No (**Vertical scaling as also called **scaling up or down**. Scaling up is an act of adding more resources to a single instance.)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Which disadvantage is related to vertical scaling:

1. Maximum capacity of the underline hosting environment
2. Cost is not linear when adding more resources to the same instance
3. VM reboot may be needed for updating the OS
4. Answers 1 & 2
5. **Answers 1, 2 & 3**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The next scaling option is horizontal scaling. Also called scaling out or in. Scaling out is about adding additional instances to support the load on our solution. In today's modern applications this is becoming much more popular and cost-effective option compared to vertical scaling.

**Yes**/No

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Autoscaling is the process of manually allocating resources to match the up-to-date performance requirements of our application while taking advantage of the elasticity of a cloud environment like Azure cloud.

Yes/**No (**Autoscaling is the process of dynamically allocating resources)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Autoscaling can be based on:

1. Threshold rules while monitoring metrics
2. Scheduled jobs
3. All answers are correct

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A scale set allows us to deploy and manage a set of identical load-balanced virtual machines like a cluster. We can manually scale the number of VMs in the scale set, or we can define rules to autoscale the number of VMs based on metrics such as CPU, memory demand, network traffic, etc.

**Yes**/No

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A scale set and also an availability set is working without using a traffic load balancer.

Yes/**No (traffic load balancer is a core component for balancing traffic between multiple VMs)**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*