**AWS Solution Architect Associate**

**What is AWS?**

AWS stands for amazon web services is a cloud provider. They provide you with servers and services that you can use on demand and scale easily.

**AWS CLI:**

Command : aws configure, aws iam list-users

|  |  |
| --- | --- |
| Commands | usage |
| aws configure | To connect your local machine to your aws account. We should give our secret credentials and default regions. |
| aws iam list-users | To see all iam users. |

**What is IAM Role?**

IAM role is an IAM entity that defines a set of permissions with credentials for making requests to AWS SERVICES, and will be used by an AWS Service.

**What is an IAM Policy?**

JSON documents that define a set of permissions for making requests to AWS services, and can be used by IAM Users, User groups, IAM roles.

**What is EC2 instance?**

EC2 instance is a virtual machine running on a physical host.

SSH: Secure Shell. Default port is 22.

command to connect to EC2 instance: ssh -i “pem file name” ec2-user@<public IPv4 addres>

**What is main difference between dedicated hosts and dedicated instances?**

With Dedicated Host the physical server is basically yours. It does not change, **it's always the same physical machine for as long as you are paying.**

Dedicated Instance does not work like this. Your instance runs on some dedicated hardware. It’s not lockdown to you. If you stop/start instance, you can get some other hardware somewhere else. Basically, the hardware is "yours" (you are not sharing it with others) for the time your instance is running. You stop/start it, you may get different physical machine later on (maybe older, maybe newer, maybe its specs will be a bit different), and so on. So your instance is moved around on different physical servers - whichever is not occupied by others at the time.

**What is EC2 User Data?**

EC2 User Data is used to bootstrap your EC2 instances using a bash script. This script can contain commands such as installing software/packages, download files from the Internet, or anything you want.

**What is compute power?**

The ability of a computer to perform work, often considered in terms of the number of instructions that can be carried out in a given time. It is also referred as processing power.

**What are EC2 Instance types?**

We can use different types of EC2 instances that are optimised for different use cases.

General Purpose: General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads. These instances are ideal for applications that use these resources in equal proportions such as web servers and code repositories. Ex: t2.micro

Compute Optimized: Great for compute-intensive tasks that require high performance processors. Batch processing workloads, Media transcoding, High performance web servers, High performance computing (HPC),Scientific modelling & machine learning, Dedicated gaming servers.

Memory Optimized: Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory. High performance, relational/non-relational databases, Distributed web scale cache stores, In-memory databases optimized for BI (business intelligence), Applications performing real-time processing of big unstructured data.

Storage Optimized: Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications. High frequency online transaction processing, relational and non-relational databases, Cache for in-memory databases (for example Redis), Data warehousing applications, Distributed file systems.

**What is higher bandwidth?**

Having a higher bandwidth means you will be able to achieve a higher data transfer rate which in turn leads to shorter download times. This is especially significant when downloading large files.

**What is Low latency?**

Low latency describes a computer network that is optimized to process a very high volume of data messages with minimal delay (latency). Simply we can say higher data transfer rate with minimal delay.

**What are different placement groups?**

Cluster: For critical applications.

Spread: For distributed applications.

Partition: For high performance applications.

**What is Public IPv4 and Private IPv4?**

Public IP is accessible around the globe and private IP is only accessible between instances in virtual private cloud(VPC).

**What is Elastic Network Interfaces(ENI)?**

ENI is a logical component in VPC and they are what gives EC2 instances access to the network.

**What is EC2 Hibernate?**

EC2 Hibernate is used to preserve the in-memory RAM state. When hibernate is enabled all and you instructed to stop-hibernate instance then Hibernation saves the contents from the instance memory (RAM) to your Amazon Elastic Block Store (Amazon EBS) root volume.

1. The EBS root volume is restored to its previous state
2. The RAM contents are reloaded
3. The processes that were previously running on the instance are resumed
4. Previously attached data volumes are reattached and the instance retains its instance ID
5. So that it's ready to be resumed to the desired state whenever needed.