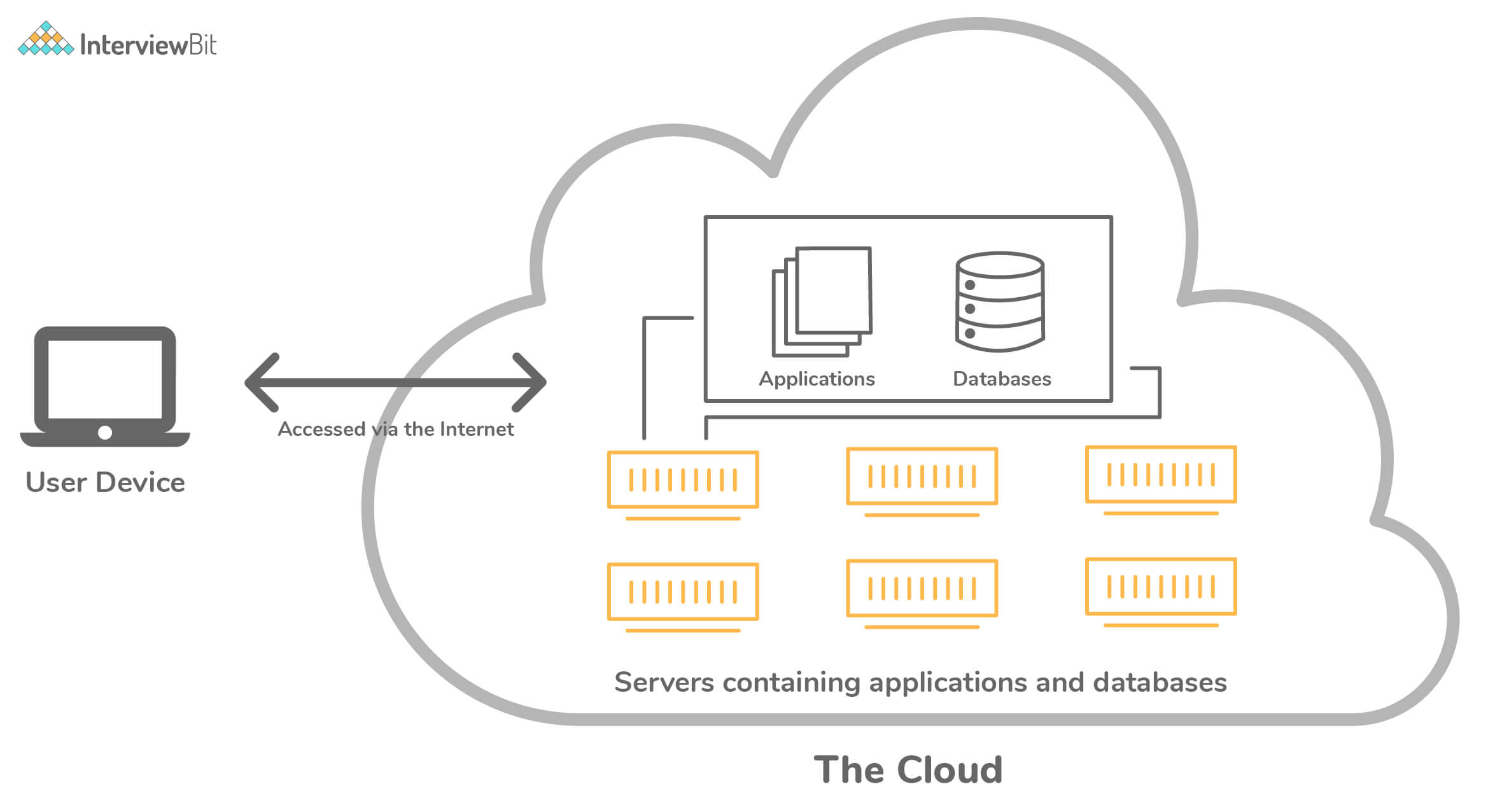
### What do you understand about cloud computing?

* Cloud computing refers to the usage of **computing resources (servers)** on the internet (refers to the term cloud) for the purpose of storing, managing, analyzing, and processing the data. Here, instead of maintaining our own servers, we use the infrastructure provided and maintained by third-party vendors such as Microsoft, AWS, etc, and pay them based on the server usage time duration.
* Cloud computing enhances the speed of execution, ensures flexibility of resources, and easier scalability.
* Cloud computing can be used to attain high fault tolerance and high system availability and this can also be done dynamically as per the infrastructural requirements of the application.

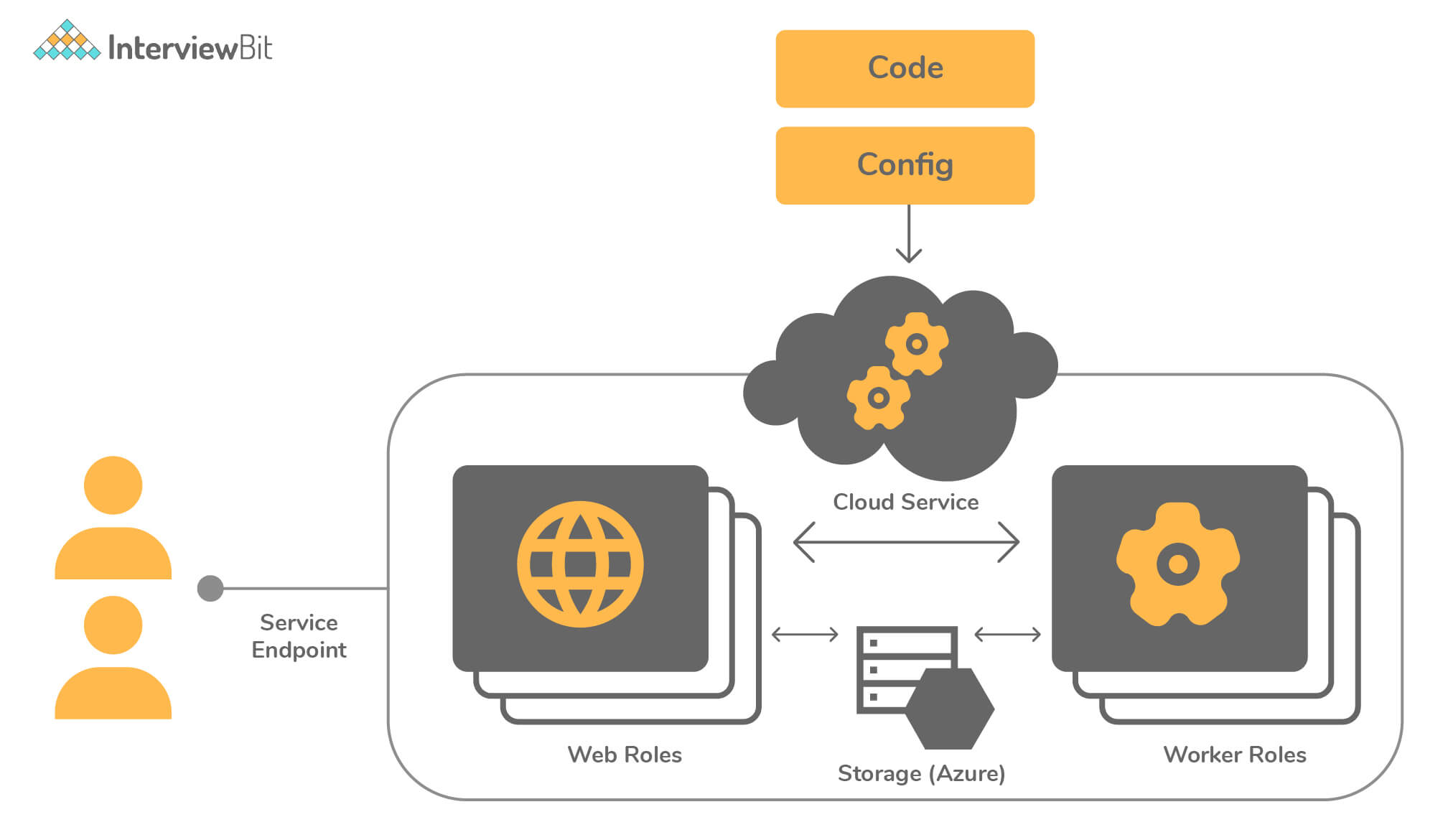


212444+++++

Cloud Computing

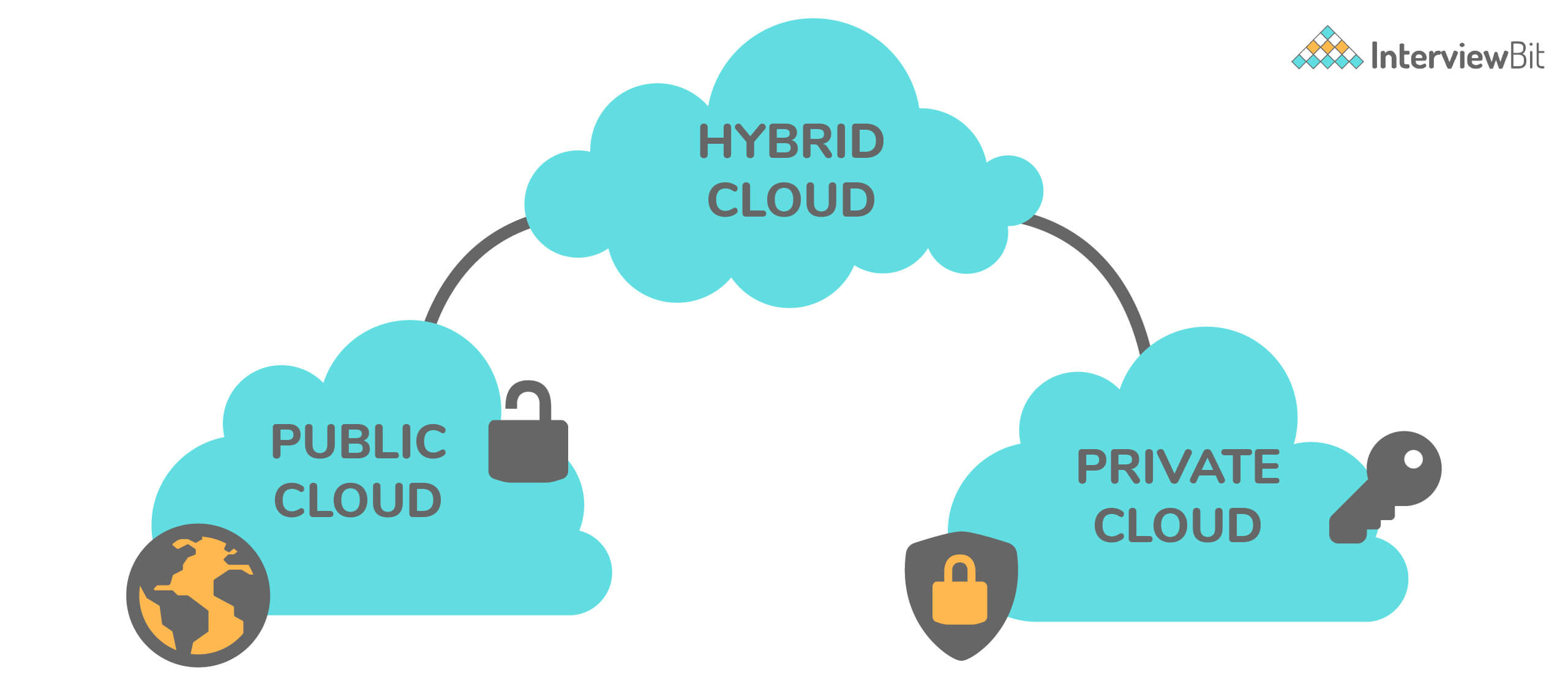
### 2. Can you tell something about Azure Cloud Service?

* Azure Cloud Service is a classic example of a platform as a service (PaaS). This was designed to support those applications which demand high scalability, reliability, and availability all within the constraints of reduced cost of operations. These are hosted on virtual VMs and Azure provides more control over them by letting the developers install the necessary software and enabling them to control remotely.
* Azure cloud services are used for deploying multi-tier web-based applications in Azure by means of creating an instance of cloud service. It is also possible to define multiple roles such as web roles, worker roles, etc for the purpose of distributed processing. Azure cloud services help in the easier and flexible scalability of the application.
* Each role of the cloud service has its own purpose and thereby its own configuration and application files.

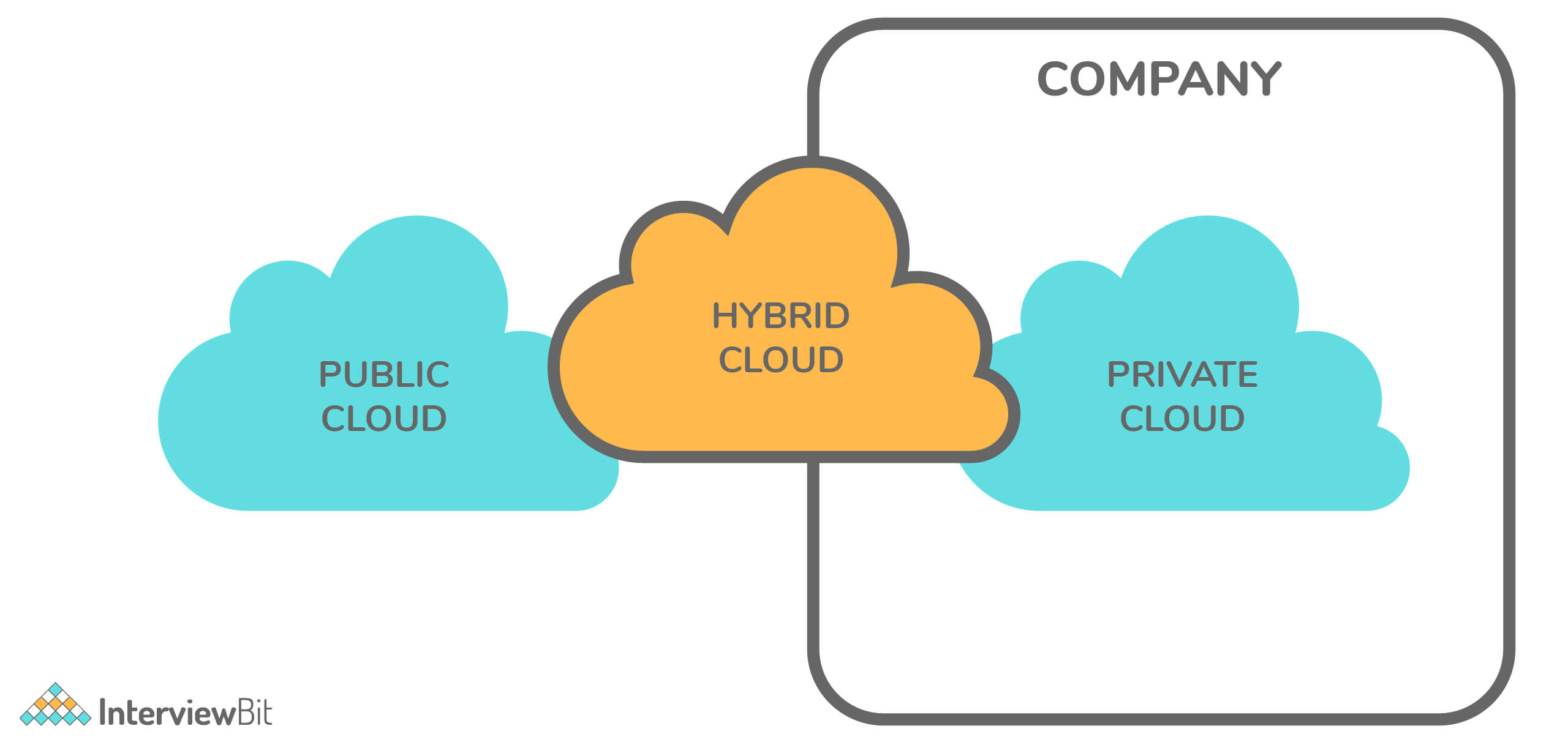
Azure Cloud Service

### 3. What are the various models available for cloud deployment?

There are 3 models available for cloud deployment:

Models For Cloud Deployment

* Public Cloud: In this model, the cloud infrastructure is owned publicly by the cloud provider and there are chances that the server resources could be shared between multiple applications.
* Private Cloud: Here, the cloud infrastructure is owned exclusively by us or exclusive service is provided by the cloud provider to us.
  + This includes hosting our applications on our own on-premise servers or hosting the application on a dedicated server provided by the cloud provider.
* Hybrid Cloud: As the name itself says, this model is the hybrid combination of private cloud and the public cloud.
  + This might include the scenario of using on-premise servers for processing confidential, sensitive data and using public cloud features for hosting public-facing applications.

Hybrid Cloud

Here, we use the best of both worlds to our requirements and advantage.

### 4. Define role instance in Azure.

A role instance is nothing but a virtual machine where the application code runs with the help of running role configurations. There can also be multiple instances of a role as per the definition in the cloud service configuration files.

### 5. How many cloud service roles are provided by Azure?

Cloud service roles comprise a set of application and configuration files. There are 2 kinds of roles provided by Azure:

* Web role: This provides a dedicated web server belonging to IIS (Internet Information Services) that is used for automatic deployment and hosting of front-end websites.
* Worker role: These roles help the applications hosted within them to run asynchronously for longer durations and are independent of the user interactions and generally do not use IIS. They are also ideal for performing background processes. The applications are run in a standalone manner.

### 6. Why is Azure Diagnostics API needed?

* Azure Diagnostics API helps us collect diagnostic data such as performance monitoring, system event logs, etc from the applications that are running on Azure.
* For the verbose monitoring of the data, Azure Diagnostics has to be enabled for the cloud service roles.
* The diagnostics data can be used for building visual chart representations for better monitoring and also for creating performance metric alerts.

### 7. Define Azure Service Level Agreement (SLA)?

* The Azure SLA is a contract that ensures or guarantees that when two or more role instances of a role are deployed on Azure, access to that cloud service is guaranteed for at least 99.95% of the time.
* It also states that if the role instance process is not in the running state, then the detection of such processes and corrective action for the same will be taken 99.9% percent of the time.
* If the mentioned guarantees are not satisfied at any point in time, then Azure credits a percentage of monthly fees to us depending on the pricing model of the respective Azure services.

### 8. What is Azure Resource Manager?

Azure Resource Manager is a service provided by Azure to provide management and application deployment in Azure.

The resource manager provides the management layer that helps the developer to create, modify or delete the resources in the Azure subscription account. This feature comes in handy when we have requirements like managing access controls, locks, ensuring the security of the resources post-deployment, and organization of those resources.

### 9. What is NSG?

NSG stands for Network Security Group that has a list of ACL (Access Control List) rules which either allows/denies network traffic to subnets or NICs (Network Interface Card) connected to a subnet or both. When NSG is linked with a subnet, then the ACL rules are applied to all the Virtual Machines in that subnet.

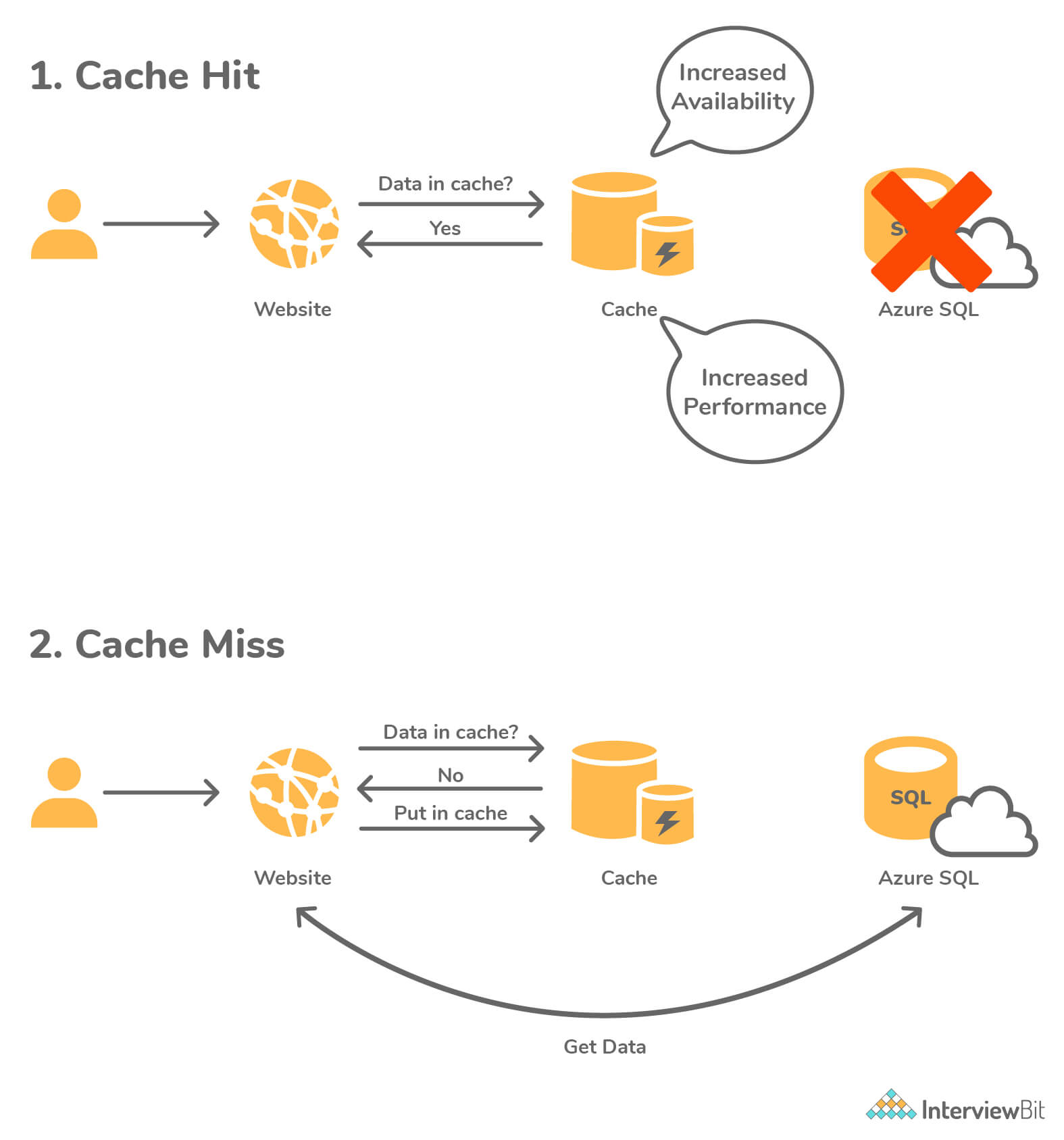
Restrictions of traffic to individual NIC can be done by associating NSG directly to that NIC.

### 10. VM creation is possible using Azure Resource Manager in a Virtual Network which was created by means of classic deployment. True or False?

False. Azure does not support this.

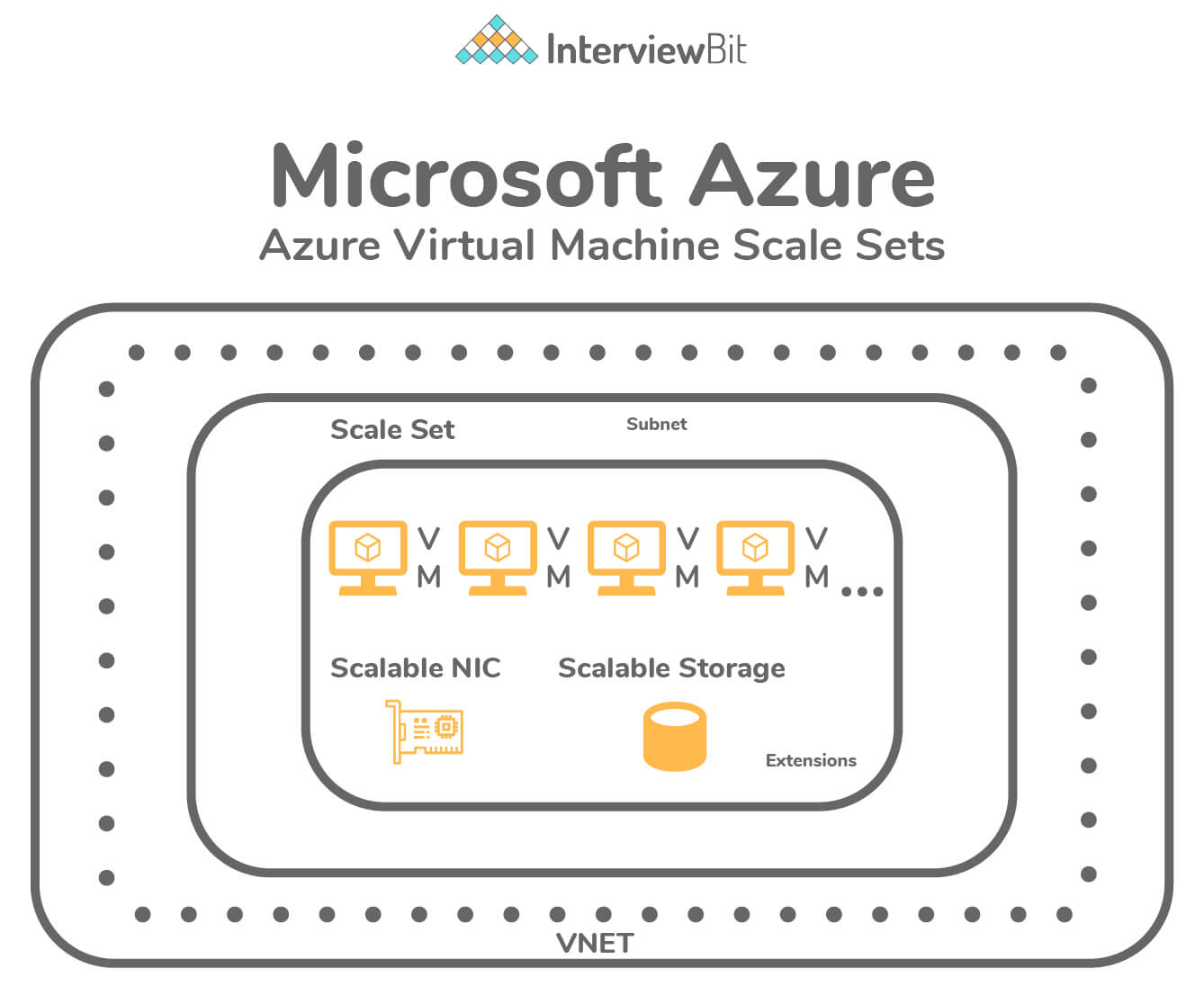
### 11. What is Azure Redis Cache?

* It is an open-source, in-memory Redis cache system provided and maintained by Azure.
* It helps the web applications to improve the performance by fetching data from the backend database and storing it into the Redis cache for the first request and then fetching data from the Redis cache for all subsequent requests.
* Azure Redis Cache provides powerful and secure caching mechanisms by making use of the Azure cloud.

Azure Redis Cache

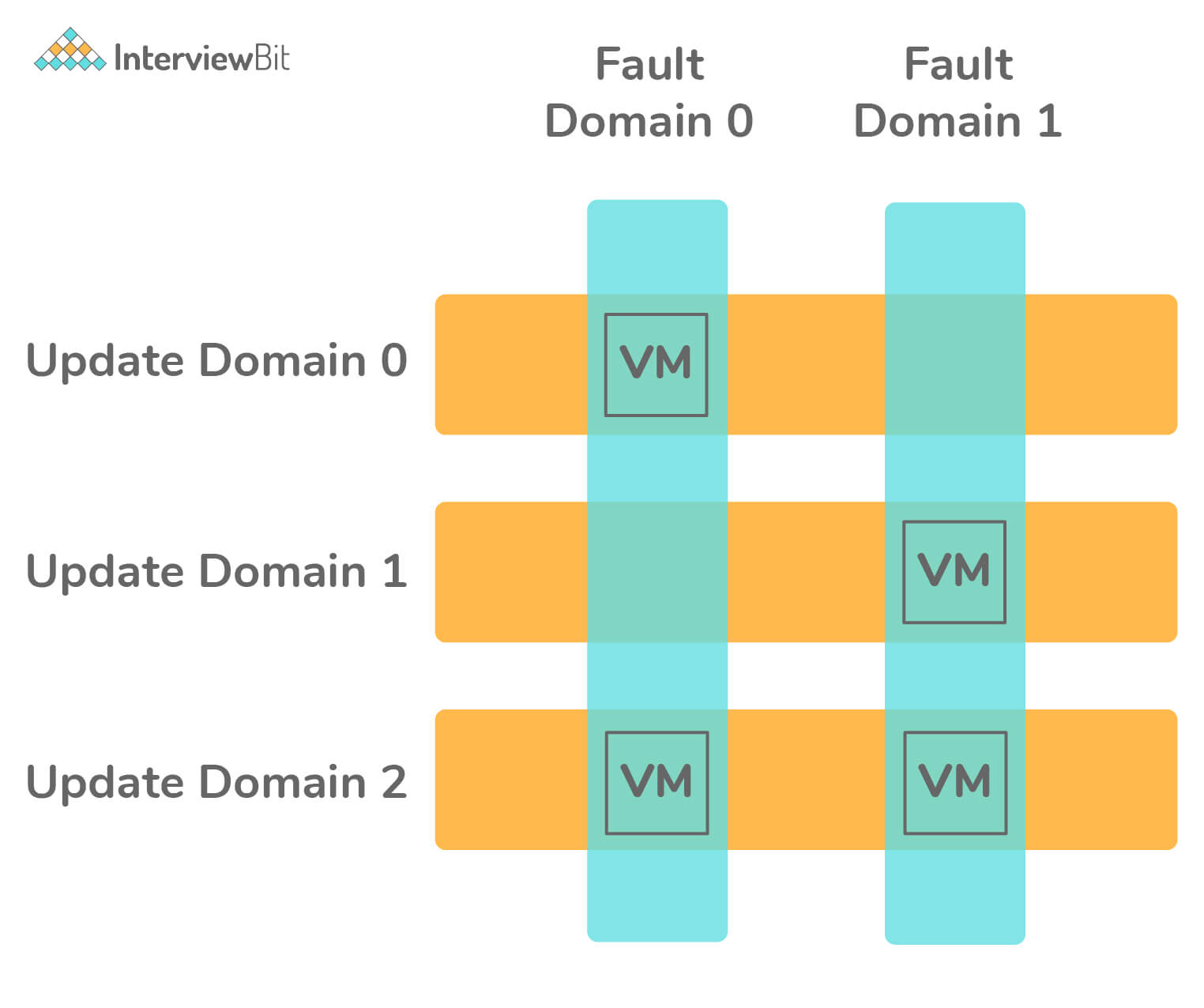
### 12. Define Azure virtual machine scale sets

* These are the Azure computation resources that can be used to deploy and manage sets of identical Virtual Machines (VMs).
* These scale sets are configured in the same manner and are designed to support the autoscaling of the applications without the need for pre-provisioning of the VMs.
* They help to build large-scale applications targeting big data and containerized workloads in an easier manner.

Azure virtual machine

### 13. What do you understand about the “Availability Set”?

* Availability Set is nothing but a logical grouping of VMs (Virtual Machines) that allows Azure cloud to understand how the application was developed for providing availability and redundancy.
* Each VM in the availability set is assigned 2 kinds of domains by Azure:
  + Fault Domain: These define the grouping of VMs that would share a common power source and common network switch. The VMs within availability sets are separated across up to 3 fault domains by default. This separation of VMs in fault domains helps our applications to be available by reducing impacts of network outages, power interruptions, and certain hardware failures.
  + Update Domain: These indicate the grouping of VMs and underlying hardware which are eligible to be rebooted at the same time. Only one update domain can be rebooted at a time, however, the order of reboot does not proceed in a sequential manner. Before the maintenance of another update domain, the previously rebooted domain is given a recovery time of 30 minutes to ensure that the domain is up.
* Azure provides flexibility to configure up to 3 fault domains and 20 update domains for an availability set.

Availability Set

### 14. What are the available options for deployment environments provided by Azure?

Azure provides two deployment environments, they are:

* Staging Environment: This environment is used for validating the changes of our application before making them live into the main environment.
  + Here, the application is identified by means of GUID (Globally Unique Identifier) of Azure which has the URL as: GUID.cloudapp.net
* Production Environment: This is the main environment where our application goes live and can be accessed by the target audience which can be accessed by means of DNS friendly URL: appName.cloudapp.net

### 15. What do you need to do when drive failure occurs?

The following steps need to be performed when the drive failure occurs:

* To make sure that the Azure Storage functions without fail, we need to ensure that the drive is not mounted.
* Replace the drive so that the drive gets remounted and formatted.

### 16. Is it possible to design applications that handle connection failure in Azure?

Yes, it is possible and is done by means of the Transient Fault Handling Block. There can be multiple causes of transient failures while using the cloud environment:

* Due to the presence of more load balancers, we can see that the application to database connections fail periodically.
* While using multi-tenant services, the calls get slower and eventually time out because other applications are using resources to hit the same resource heavily.
* The last cause can be we ourselves as the user trying to hit the resource very frequently which causes the service to deliberately deny the connection to us to support other tenants in the architecture.

Instead of showing errors to the user periodically, the application can recognize the errors that are transient and automatically try to perform the same operation again typically after some seconds with the hope of establishing the connection. By making use of the Transient Fault Handling Application Block mechanism, we can generate the retry intervals and make the application perform retries. In the majority of the cases, the error would be resolved on the second try and hence the user need not be made aware of these errors unnecessarily.

Following is the sample code that can be used for the retry policy. Here, if the connection is not successful, then the action is retried based on the retry policy defined. There are 3 retry strategies - Fixed Interval, Incremental Interval, Exponential Backoff Strategy.

/\*\*\*

\* Class to detect Transient Blocks - Here

\* OperationCancelledException is

\* detected and then the retry strategy is employed.

\*/

internal class AppTransientDetection : ITransientErrorDetectionStrategy

{

bool IsTransient(Exception exception) =>

exception is OperationCanceledException;

}

/\*\*\*

\* Retry Strategy - Here Fixed Interval Strategy is employed and is retried for 5 times.

\*/

RetryStrategy retryStrategy = new FixedInterval(retryCount: 5, retryInterval: TimeSpan.FromSeconds(2));

RetryPolicy retryPolicy = new RetryPolicy(new AppTransientDetection(), retryStrategy);

retryPolicy.ExecuteAction(() => {

try {

string commandText = @”USE FEDERATION User\_Federation(ShardId =” + shardId + “) WITH RESET, FILTERING=ON”;

userEntity.Connection.Open();

userEntity.ExecuteStoreCommand(commandText);

} catch (Exception e) {

userEntity.Connection.Close();

SqlConnection.ClearAllPools();

}

});

### 17. Define azure storage key.

* Azure storage key is used for authentication for validating access for the azure storage service to control access of data based on the project requirements.
* 2 types of storage keys are given for the authentication purpose -
  + Primary Access Key
  + Secondary Access Key
* The main purpose of the secondary access key is for avoiding downtime of the website or application.

### 18. What is cspack in Azure?

It is a command-line tool that is used for generating service package files. The tool also helps in preparing the application for deployment in Microsoft Azure or compute emulator.

Every project of cloud service type has the .cscfg file which is basically the cloud service configuration file that is generated by means of cspack tool and is primarily used to store:

* The number of role instances for the deployment of each role in the project.
* The thumbprint of the certificates.
* User-defined configuration and settings.

### 19. What is the best Azure solution for executing the code without a server?

* Azure Functions service can be used for executing the code without a server.
* Serverless Azure Functions are used for simplifying complex orchestration and challenging resolutions. They are meant for being stateless and short-lived.
* They help to connect with other services without the need for hard coding of the integrations thereby making the development process faster.
* It helps the developer to write and concentrate on the business logic code thereby saving time and effort.
* They also provide the features of monitoring and analyzing code performance by means of Azure Application Insights that help in identifying bottlenecks and failure points across the components of the application.

### 20. What would be the best feature recommended by Azure for having a common file sharing system between multiple virtual machines?

Azure provides a service called Azure File System which is used as a common repository system for sharing the data across the Virtual Machines configured by making use of protocols like SMB, FTPS, NFS, etc.

### 21. Is it possible to login to a Linux Virtual Machine without using a password?

Yes, it is possible by making use of the Key Vault mapping to any Admin VM, we can log in to another VM without the need for a password.

### 22. What are the differences between Azure Scale Sets and Availability Sets?

The main difference  between Azure Scale Sets and Availability Sets are given below:

| **Criteria** | **Azure Scale Sets** | **Azure Availability Sets** |
| --- | --- | --- |
| Definition | They are a group of identically configured VMs that are spread across multiple fault domains. | They are the group of discretely configured VMs that are spread across various fault domains. |
| Default Domain | These have 5 fault domains and update domains by default. | By default, these have 3 fault domains and 5 update domains. |
| Workload Type | These are used when there are unpredictable workloads that require the feature of auto scalability. | These are used when there are predictable workload requirements. |
| Configuration Style | Here, the VMs are configured and created in the same manner from the same image. | Here, the VMs are created by making use of different images and configurations. |
| VM Count | The number of VMs can be increased/decreased based on the demand or the pre-defined schedule. | A VM can be added to an availability set only at the time of the set’s creation. |
| Distribution style | Here, the VM scale sets can be distributed across multiple data centers or within a single data center. | Here, the VMs are automatically distributed in a data center. |

### 23. What would happen when the maximum failed attempts are reached during the process of Azure ID Authentication?

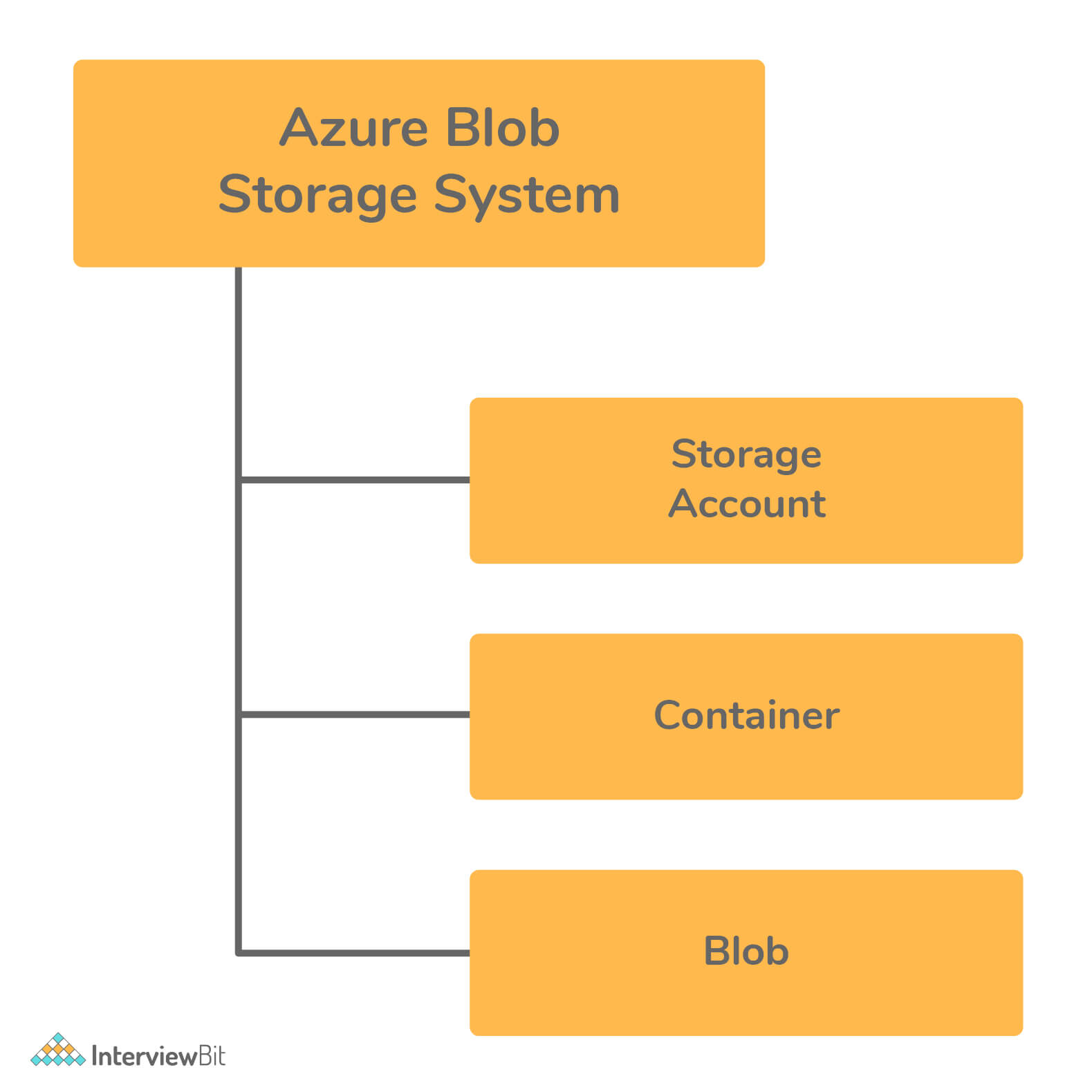
In case of maximum failed attempts, the azure account would get locked and the method of locking is dependent on the protocol that analyzes the entered password and the IP addresses of the login requests.

### 24. Is it possible to get a public DNS or IP address for the Azure Internal Load Balancer?

No! As the name itself says, Azure Internal Load Balancer supports only Private IP addresses, and hence the assignment of a public IP address or DNS name is not possible.

### 25. What is Azure Blob Storage?

* Azure Blob storage is the object storage solution provided by Microsoft for the cloud. Blob stands for “Binary Large Object”. Blob-based storage is used to store massive unstructured data in terms of text or binary format. It is ideal for serving documents/images/audio/video/text directly to browser.
* The data stored in the blob storage is accessible from anywhere in the world. The blobs are tied to user accounts by grouping them into containers. The Azure Blob Service has 3 components:
  + Storage Account: This can be a General Storage Account or Blob Storage Account registered in Microsoft Azure.
  + Container: Container is used for grouping blobs. We can store an unlimited number of blobs in a container. The name of the container should start in lowercase.
  + Blob: A blob is a Binary Large Object like a file or document of any type and size. There are 3 kinds of Blobs supported by Azure:
    - Block blobs: These are intended for text and binary files and can support up to 195GB, i.e up to 50k blocks of up to 4MB each.
    - Append blobs: These are used for appending operations like logging data in log files.
    - Page blobs: These are meant for frequent read/write operations.

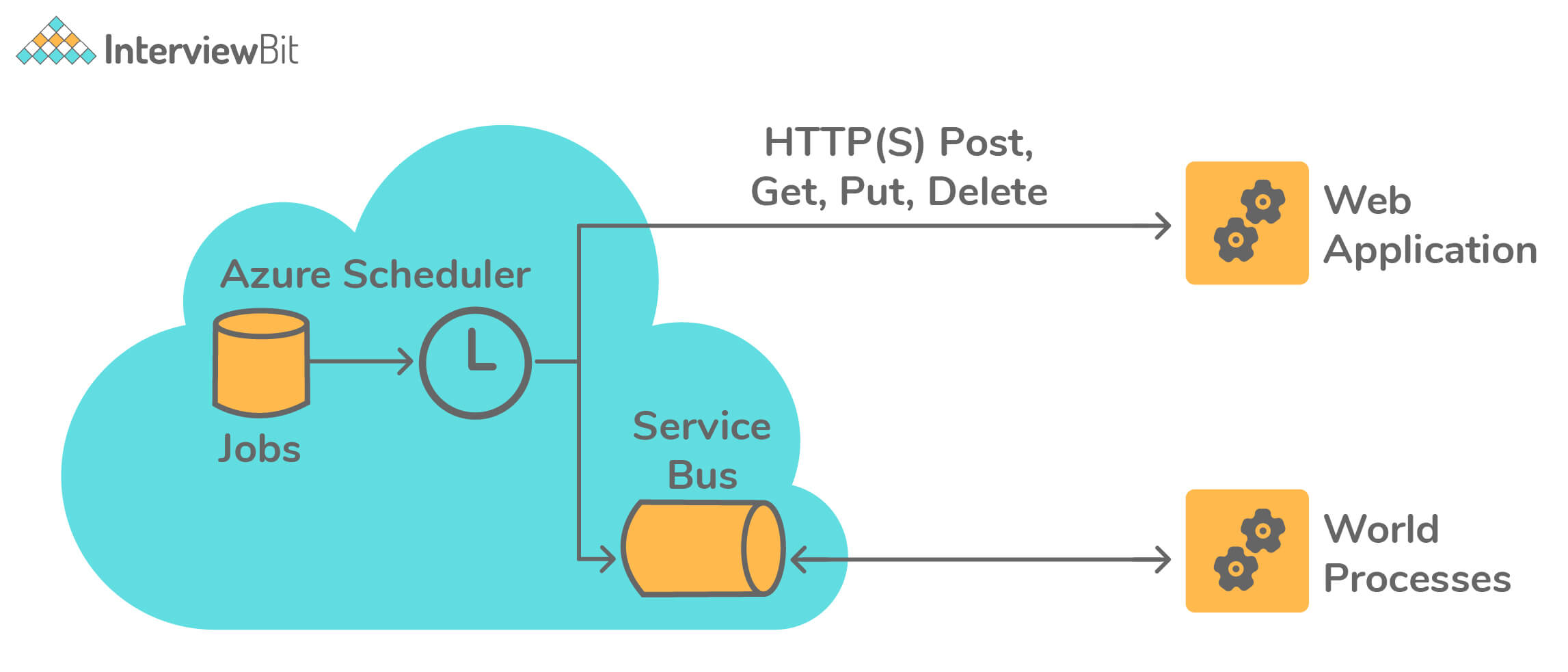
Azure Blob Storage

## Azure Interview Questions For Experienced

### 26. What do you understand by Azure Scheduler?

Azure Scheduler helps us to invoke certain background trigger events or activities like calling HTTP/S endpoints or to present a message on the queue on any schedule.

By using this Azure Schedule, the jobs present in the cloud call services present within and outside of the Azure to execute those jobs on-demand that are routinely on a repeated regular schedule or start those jobs at a future specified date.

Azure Scheduler

### 27. Is it possible to map the Windows machines running on two different port numbers, say 80 and 81, on an IIS Web Server to an Azure Load Balancer?

Yes, it can be done by defining a separate Load Balancer Role in Azure.

### 28. You have an application running on the On-Prem Server and have backup on Azure East US region. Now, On-Prem server application access fails. Is it possible to access the application via the Azure environment?

Yes, it is totally possible by making use of the Site Recovery Service provided by Azure. It is capable of handling fail-over and fail-back scenarios between On-Prem Servers and Azure environments.

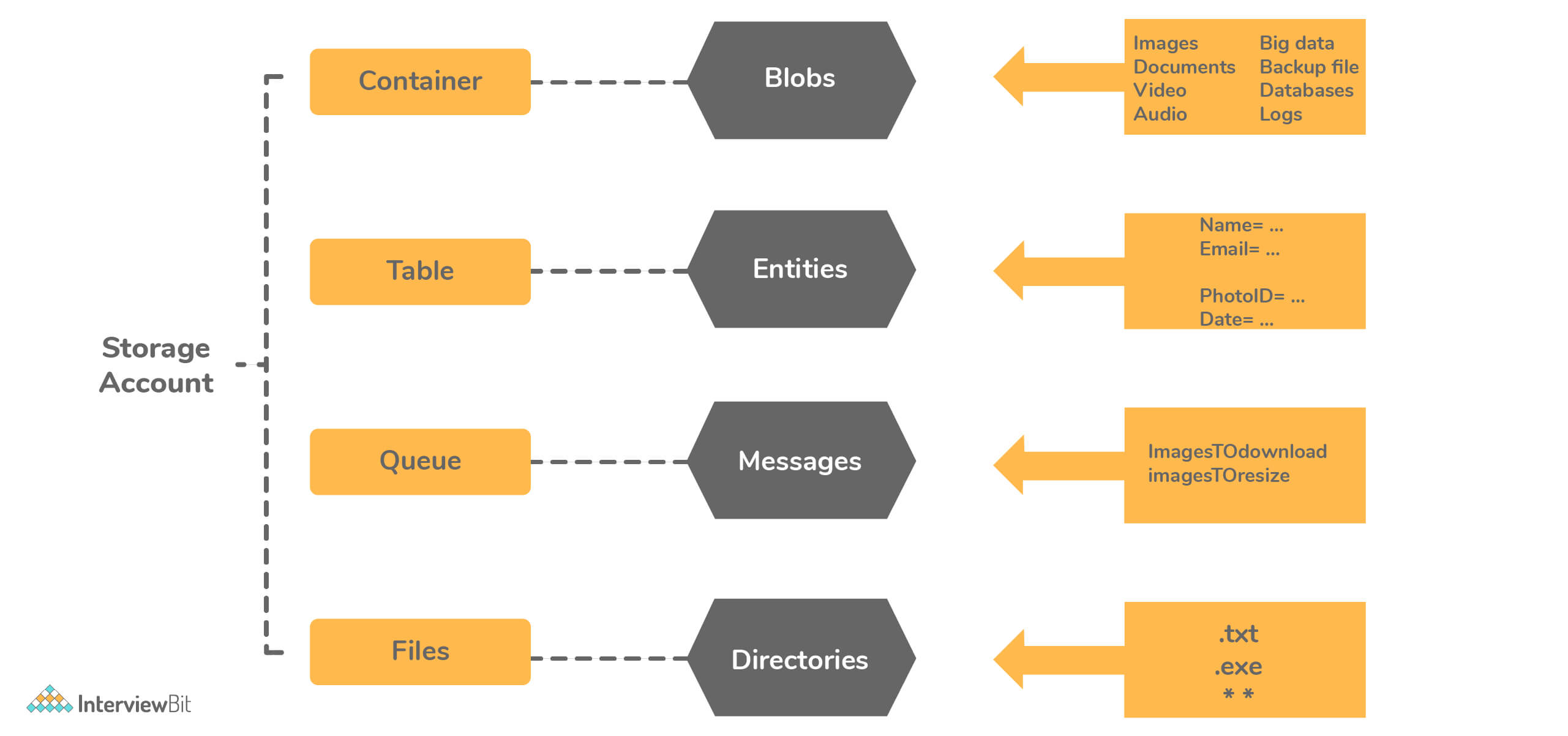
### 29. What feature of Azure can be used to stop the issue of high load on the application in cases of no man support on the flow?

This issue can be stopped by making use of VM Scale sets by defining proper configuration and conditions to provision a new VM whenever the load to the application increases.

* Azure VM Scale Sets lets the developer create and manage a group of VMs that are load balanced. The scale sets can be configured in such a way that the count of VMs can automatically be increased or decreased based on the application demand or based on a pre-defined schedule.
* Usage of Scale Sets ensures high availability of the applications and allows the developers to manage, update and configure large VMs centrally and also help them support the development of large-scale applications supporting big data, big workloads, and compute loads.
* Azure scale sets can support up to 1,000 VMs. If the custom VM images are created and uploaded, then the limit is 600 VMs.

### 30. What are the types of storage services apart from blob storage provided by Azure?

Azure provides overall 4 types of storage services - Blob Service, Table Storage, Queue Storage, and File Storage Services as shown in the figure below:

Types of Storage Services

* Azure Table Storage: This type of storage lets user deploy their applications with semi-structured data and a NoSQL-based key-value store.
  + This is used when there is a need for applications that follow a flexible schema of data.
  + Table Storage focuses on enterprise-level data and follows strongly consistent models.
  + The data is represented in terms of Entities grouped under tables.
* Azure Queue Storage: This storage provides a message queue system for handling large workloads by letting users develop and build flexible and modular applications.
  + This storage ensures that the application becomes less prone to failure of individual components and is scalable.
  + With the help of message queues, it provides the queue monitoring feature for helping the application to ensure the user demands are met.
* Azure File Storage: This storage type provides features of file sharing that are accessible using SMB (Server Message Block) Protocol. The data in this storage is protected by HTTPS and SMB 3.0 Protocol.
  + They are used for improving the performance and capabilities of on-premise applications.
  + The OS deployments and hardware management is taken by Azure itself.

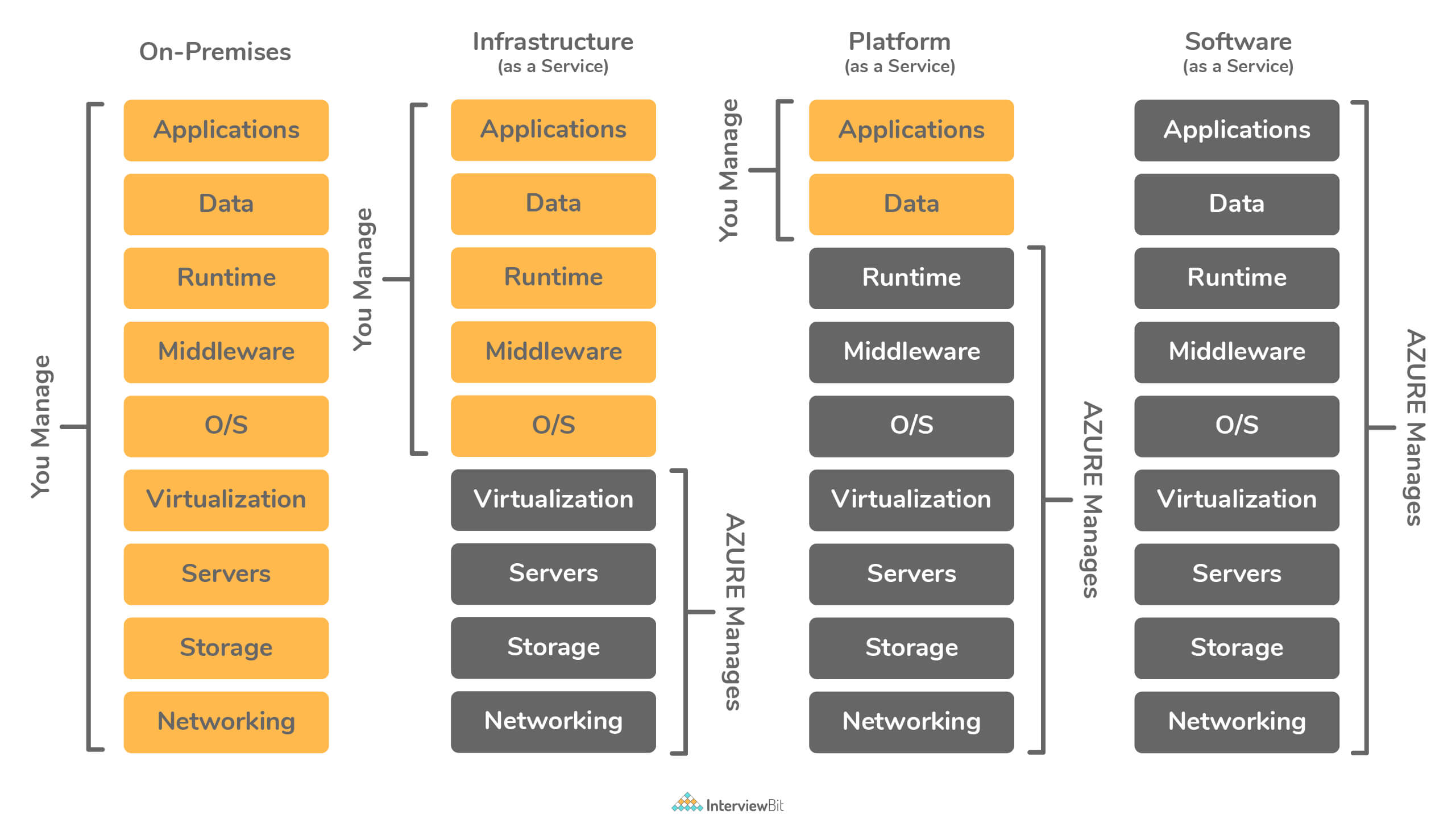
### 31. What are IaaS, PaaS and SaaS?

**IaaS:** This stands for “Infrastructure as a Service” which provides a set of capabilities like OS, network connectivities, etc which are at the infrastructural level and are delivered as pay per use policy. The infrastructure is used for hosting applications. Examples include Azure VM, VNET, etc.

**PaaS:** PaaS stands for “Platform as a Service” which is mostly about underlying infrastructure abstraction to the developers for enabling quicker development of the applications without the need for worry about hosting management. Examples include Azure web apps, Storage services, cloud services, etc.

**SaaS:** SaaS stands for “Software as a Service” and are those applications which are delivered using the service delivery model where the applications are simply consumed and used by an organization. These applications are generally mobilized by making the organization pay for their usage or through ads. Examples include applications like Office 365, Gmail, SharePoint Online, and so on.

The following table shows the difference between the On-Prem Service, IaaS, PaaS, and SaaS services. We can observe that as we go right, the level of control the developer or the user has over the application reduces.

IaaS, PaaS and SaaS

### 32. What are the differences between the Azure Table Storage and the Azure SQL service?

The main difference between Azure Table Storage and Azure SQL Service is given below:

| **Table Storage Service** | **Azure SQL Table** |
| --- | --- |
| This follows a NoSQL type of storage on Azure. | This follows the relational storage structure on Azure. |
| The data is stored in key-value format and is referred to as Entity. | The data here is stored in rows and columns combination in the SQL table. |
| The data schema is not enforced for storage. | The data schema is enforced for storing data and if the schema violation occurs, then it results in an error. |
| The relationship between tables is not possible. | Relationships between tables are defined by means of the foreign keys. |
| The partition and row key combination are considered unique for each entity. | Uniqueness can be defined by the user by means of a primary key or unique key. |
| This service can be used for storing log information or diagnostics data. | This service is widely used for transaction-based applications. |

### 33. Consider a scenario where an application front end hosting is done on Azure but the customer needs the database hosting to be done on on-premise server due to security concerns. What are the ways to handle the connectivity in Azure for this scenario?

* **Possibility 1:** Azure VNET based “Point to Site” service can be a correct choice for this scenario of connecting one on-premise DB to an Azure-hosted app. “Point to Site” is valid for cases where the count of resources to be connected via VPN is very limited.
* **Possibility 2:** In case there is a large number of resources for connection, then “Site to Site” or “Express routes” are the other options that could be considered.
  + There might be chances that using “Site to Site” might lead to network latency as VPN due to these work only via Internet (public infrastructure). In such cases, “Express Routes” are used as it provides dedicated leased line for overcoming latency issues.
* **Possibility 3:** In case the customer is not willing to work via VNET, then Windows Communication Foundation (WCF) service can be developed and hosted on-premise which would have CRUD operations meant only for the database hosted on-premise. This works by means of using the “Service bus relay” option for developing communication between the Azure-hosted app to the WCF service for database access.

### 34. What are the differences between the Azure Storage Queue and the Azure Service Bus Queue?

The main difference between Azure Storage Queue and the Azure Service Bus Queue is given below:

| **Azure Storage Queue** | **Azure Service Bus Queue** |
| --- | --- |
| Here, the FIFO (First In First Out) ordering is not guaranteed. | The FIFO order is guaranteed for the messages by means of sessions. |
| Sessions are not supported. | Sessions that are messaging level are supported here. |
| Here, only the “At Least Once delivery” model is supported. | This supports “At least once”, “Atmost once” and “Exactly once” delivery models for the messages. |
| There is no automatic detection of duplicates here. | Automatic duplicate detection is supported here. |
| Does not support dead lettering. | Supports dead lettering. |
| The size of the message is 64KB. | The size of the message is 256KB. |
| Supports one-to-one delivery of messages. | Supports both one to one and one-to-many deliveries of messages. |
| The transaction is not supported. | The transaction is supported here. |
| This queue supports only batch receive. | This supports both batch send and batch receive of messages. |
| The behavior of receiving messages is non-blocking. | The behavior can be either blocking or non-blocking based on the configuration. |

### 35. What are the possible causes of the client application to be disconnected from the cache?

There can be 2 possible causes:

* Client-side causes:
  + The application might have been redeployed.
  + The application might have just performed a scaling operation.
  + The client-side networking layer has been changed.
  + There might be transient errors in the client or the network between the client and the server.
  + Another possible reason could be the bandwidth threshold limits have been crossed.
* Server-side causes:
  + It might occur if the Azure Redis Cache service itself might undergo a failover from the primary to the secondary node.
  + The server instance where the cache was deployed might have undergone patching or maintenance.

### 36. How can a VM be created by means of Azure CLI?

az vm create `

--resource-group myResourceGroupName `

--name myVM --image win2016datacenter `

--admin-username AzureuserNAME `

--admin-password AzurePASSWORD

types of integration runtime?

There are three types of integration runtimes offered by Data Factory:

**1.Azure integration runtime**.

2.Self-hosted integration runtime.

3.Azure-SQL Server Integration Services (SSIS) integration runtime.16-Aug-2022

# 1.Integration runtime is the compute infrastructure used by Azure Data Factory (ADF) to provide various data integration capabilities across different network environments.

# 2.azure self hosted integration runtime

# The self-hosted integration runtime is the compute infrastructure that Azure Data Factory uses to provide data-integration capabilities across different network environments.

# 3.Azure SSIS integration runtimes in the portal

# In the Azure Data Factory UI, switch to the Manage tab and then switch to the Integration runtimes tab on the Connections pane to view existing integration runtimes in your data factory.

# Select New to create a new Azure-SSIS IR and open the Integration runtime setup pane.

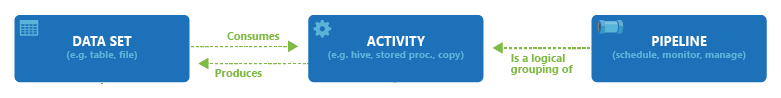
### Azure Blob Storage

## Store and access unstructured data at scale

Azure Blob Storage helps you create data lakes for your analytics needs and provides storage to build powerful cloud-native and mobile apps. Optimise costs with tiered storage for your long-term data and flexibly scale up for high-performance computing and machine learning workloads.

**What is Activity in Azure Data Factory?**

The activity is the task we performed on our data. We use activity inside the Azure Data Factory pipelines. ADF pipelines are a group of one or more activities. For ex: When you create an ADF pipeline to perform ETL you can use multiple activities to extract data, transform data and load data to your data warehouse. Activity uses Input and output datasets. Dataset represents your data if it is tables, files, folders etc. Below diagram shows the relationship between Activity, dataset and pipeline:



An Input dataset simply tells you about the input data and it’s schema. And an Output dataset will tell you about the output data and it’s schema. You can attach zero or more Input datasets and one or more Output datasets. Activities in Azure Data Factory can be broadly categorized as:

1- Data Movement Activities

2- Data Transformation Activities

3- Control Activities

**DATA MOVEMENT ACTIVITIES :**

**1- Copy Activity:** It simply copies the data from Source location to destination location. Azure supports multiple data store locations such as Azure Storage, Azure DBs, NoSQL, Files, etc.

To know more about Data Movement activities, please use below link:

[Pipelines and activities in Azure Data Factory - Azure Data Factory | Microsoft Docs](https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities#data-movement-activities)

**DATA TRANSFORMATION ACTIVITIES:**

**1- Data Flow:**In data flow, First, you need to design data transformation workflow to transform or move data. Then you can call Data Flow activity inside the ADF pipeline. It runs on Scaled out Apache Spark Clusters. There are two types of DataFlows: Mapping and Wrangling DataFlows

**MAPPING DATA FLOW:** It provides a platform to graphically design data transformation logic. You don’t need to write code. Once your data flow is complete, you can use it as an Activity in ADF pipelines.

**WRANGLING DATA FLOW:**It provides a platform to use power query in Azure Data Factory which is available on Ms excel. You can use power query M functions also on the cloud.

**2- Hive Activity:** This is a HD insight activity that executes Hive queries on windows/linux based HDInsight cluster. It is used to process and analyze structured data.

**3- Pig activity:** This is a HD insight activity that executes Pig queries on windows/linux based HDInsight cluster. It is used to analyze large datasets.

**4- MapReduce:**This is a HD insight activity that executes MapReduce programs on windows/linux based HDInsight cluster. It is used for processing and generating large datasets with a parallel distributed algorithm on a cluster.

**5- Hadoop Streaming:** This is a HD Insight activity that executes Hadoop streaming program on windows/linux based HDInsight cluster. It is used to write mappers and reducers with any executable script in any language like Python, C++ etc.

**6- Spark:**This is a HD Insight activity that executes Spark program on windows/linux based HDInsight cluster. It is used for large scale data processing.

**7- Stored Procedure:**In Data Factory pipeline, you can use execute Stored procedure activity to invoke a SQL Server Stored procedure. You can use the following data stores: Azure SQL Database, Azure Synapse Analytics, SQL Server Database, etc.

**8- U-SQL:** It executes U-SQL script on Azure Data Lake Analytics cluster. It is a big data query language that provides benefits of SQL.

**9- Custom Activity:** In custom activity, you can create your own data processing logic that is not provided by Azure. You can configure .Net activity or R activity that will run on Azure Batch service or an Azure HDInsight cluster.

**10- Databricks Notebook:**It runs your databricks notebook on Azure databricks workspace. It runs on Apache spark.

**11- Databricks Python Activity:** This activity will run your python files on Azure Databricks cluster.

**12- Azure Functions:** It is Azure Compute service that allows us to write code logic and use it based on events without installing any infrastructure. It stores your code into Storage and keep the logs in application Insights.Key points of Azure Functions are :

1- It is a Serverless service.

2- It has Multiple languages available : C#, Java, Javascript, Python and PowerShell

3- It is a Pay as you go Model.

To know more about Data Transformation activity, use below link:

[Pipelines and activities in Azure Data Factory - Azure Data Factory | Microsoft Docs](https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities#data-transformation-activities)

**3- Control Flow Activities:**

**1- Append Variable Activity:**It assigns a value to the array variable.

**2- Execute Pipeline Activity:** It allows you to call Azure Data Factory pipelines.

**3- Filter Activity:** It allows you to apply different filters on your input dataset.

**4- For Each Activity:** It provides the functionality of a for each loop that executes for multiple iterations.

**5- Get Metadata Activity:**It is used to get metadata of files/folders. You need to provide the type of metadata you require: childItems, columnCount, contentMDS, exists, itemName, itemType, lastModified, size, structure, created etc.

**6- If condition Activity:** It provides the same functionality as If statement, it executes the set of expressions based on if the condition evaluates to true or false.

**7- Lookup Activity:** It reads and returns the content of multiple data sources such as files or tables or databases. It could also return the result set of a query or stored procedures.

**8- Set Variable Activity:** It is used to set the value to a variable of type String, Array, etc.

**9- Switch Activity:**It is a Switch statement that executes the set of activities based on matching cases.

**10- Until Activity:** It is same as do until loop. It executes a set of activities until the condition is set to true.

**11- Validation Activity:** It is used to validate the input dataset.

**12- Wait Activity:**It just waits for the given interval of time before moving ahead to the next activity. You can specify the number of seconds.

**13- Web Activity:** It is used to make a call to REST APIs. You can use it for different use cases such as ADF pipeline execution.

**14- Webhook Activity:** It is used to to call the endpoint URLs to start/stop the execution of the pipelines. You can call external URLs also.

## Difference Between Gen 1 vs Gen 2

| **ADLS GEN 1** | **ADLS GEN 2** |
| --- | --- |
| With Azure Data Lake Gen 1, you could only store and process unstructured data like text and binary files. | But with Azure Data Lake Gen 2, you can also store structured data such as JSON documents, tables, or CSV files in your account. |
| With Azure Data Lake Gen 1, you could only process data using Hadoop MapReduce scripts. | But with Azure Data Lake Gen 2, you can also use Apache Spark and Apache Hive to process structured data in your account. |
| With Azure Data Lake Gen 1, you could only store your data in perpetuity. | But with Azure Data Lake Gen 2, you have the option of setting up a time-to-live policy that automatically deletes your data after a certain period of time. |
| With Azure Data Lake Gen 1, you pay per gigabyte of storage used. | But with Azure Data Lake Gen 2, you pay only for the amount of data processed by your job scripts. |
| With Azure Data Lake Gen 1, you could only store up to 1 TB per account. | But with Azure Data Lake Gen 2, there’s no limit on how much data you can store in one place—as long as it doesn’t exceed the available storage capacity of your subscription. |
| Azure Data Lake Gen 1 didn’t have any built-in security features. | But with Azure Data Lake Gen 2, you can use Kerberos authentication to protect your data from unauthorized access |