1) What is Cloud Computing?

Answer: Cloud computing is the delivery of computing services including servers, storage, databases, networking, software, analytics, intelligence and many more other services over the Internet.

2) What is the benefit of cloud computing?

Answer: Below are the few benefits of cloud computing:

  a) Automatic Backup & Software updates
  b) Automation of task
  c) Better Storage
  d) Cost effective
  e) Disaster Recovery & Loss Prevention
  f) Geo Redundant
  g) High Security
  h) Highly Flexible
  i) Low maintenance cost & zero risk failure
  j) Remote Accessibility
  k) Scalability

3) What are the Challenges in maintaining on-prem data storage and servers?

Answer: There are several challenges in maintaining op-prem data storage and servers:

- Security Risks
- Inability to Scale
- Minimal Data Backup & Recovery
- Increased maintenance Costs
- Lack of Access & Mobility for Employees
- Require extra it support
- Require a greater capital investment
- Increase the risk of data loss

4) What are the types of cloud?

Answer: There are the following 4 types of cloud:

  a) **Public Cloud**: Public cloud is open for all to store and access information via the Internet using the pay-per-usage method. In public cloud, computing resources are managed and operated by the Cloud Service Provider (CSP).
  Example: Amazon elastic compute cloud (EC2), IBM SmartCloud Enterprise, Microsoft, Google App Engine, Windows Azure Services Platform.

  There are the following advantages of Public Cloud:
  - Public cloud is owned at a lower cost than the private and hybrid cloud.
  - Public cloud is maintained by the cloud service provider, so do not need to worry about the maintenance.
  - Public cloud is easier to integrate. Hence it offers a better flexibility approach to consumers.
  - Public cloud is location independent because its services are delivered through the internet.
  - Public cloud is highly scalable as per the requirement of computing resources.

- It is accessible by the general public, so there is no limit to the number of users.

There are few disadvantages of Public Cloud as well:
- Public Cloud is less secure because resources are shared publicly.
- Performance depends upon the high-speed internet network link to the cloud provider.
- The Client has no control of data.

b) **Private Cloud**: Private cloud is also known as an internal cloud or corporate cloud. It is used by organizations to build and manage their own data centers internally or by the third party.

Based on the location and management, National Institute of Standards and Technology (NIST) divide private cloud into the following two parts:
- On-premise private cloud
- Outsourced private cloud

There are the following advantages of the Private Cloud:
- Private cloud provides a high level of security and privacy to the users.
- Private cloud offers better performance with improved speed and space capacity.
- It allows the IT team to quickly allocate and deliver on-demand IT resources.
- The organization has full control over the cloud because it is managed by the organization itself. So, there is no need for the organization to depends on anybody.
- It is suitable for organizations that require a separate cloud for their personal use and data security is the first priority.

There are few disadvantages of the Private Cloud as well:
- Skilled people are required to manage and operate cloud services.
- Private cloud is accessible within the organization, so the area of operations is limited.
- Private cloud is not suitable for organizations that have a high user base, and organizations that do not have the prebuilt infrastructure, sufficient manpower to maintain and manage the cloud.

c) **Hybrid Cloud**: Hybrid Cloud is a combination of the public cloud and the private cloud. we can say:
Hybrid Cloud = Public Cloud + Private Cloud
Hybrid cloud is partially secure because the services which are running on the public cloud can be accessed by anyone, while the services which are running on a private cloud can be accessed only by the organization's users.
Example: Google Application Suite (Gmail, Google Apps, and Google Drive), Office 365 (MS Office on the Web and One Drive), Amazon Web Services.

There are the following advantages of Hybrid Cloud:
- Hybrid cloud is suitable for organizations that require more security than the public cloud.
- Hybrid cloud helps you to deliver new products and services more quickly.
- Hybrid cloud provides an excellent way to reduce the risk.
- Hybrid cloud offers flexible resources because of the public cloud and secure resources because of the private cloud.

There are few disadvantages of Hybrid Cloud as well:

- In Hybrid Cloud, security feature is not as good as the private cloud.
- Managing a hybrid cloud is complex because it is difficult to manage more than one type of deployment model.
- In the hybrid cloud, the reliability of the services depends on cloud service providers.

d) **Community Cloud**: Community cloud allows systems and services to be accessible by a group of several organizations to share the information between the organization and a specific community. It is owned, managed, and operated by one or more organizations in the community, a third party, or a combination of them.
Example: Health Care community cloud

There are the following advantages of Community Cloud -
- Community cloud is cost-effective because the whole cloud is being shared by several organizations or communities.
- Community cloud is suitable for organizations that want to have a collaborative cloud with more security features than the public cloud.
- It provides better security than the public cloud.
- It provides collaborative and distributive environment.
- Community cloud allows us to share cloud resources, infrastructure, and other capabilities among various organizations.

There are few disadvantages of Community Cloud as well:
- Community cloud is not a good choice for every organization.
- Security features are not as good as the private cloud.
- It is not suitable if there is no collaboration.
- The fixed amount of data storage and bandwidth is shared among all community members.

5) What are the cloud services?

Answer: There are the following three types of cloud service models -

1. **Infrastructure as a Service (IaaS)**: IaaS is a managed cloud computing services where hardware resources are provided, managed and maintained by the cloud service provider. IaaS helps users to use computing resources such as servers, virtual machine, networking, processing power, storage capacity and many more other services without intensive hardware investments.

   IaaS is beneficial for those customers who want to create cost-effective and highly scalable IT solutions where the expenses and complexities involved in managing hardware resources are outsourced to a service provider.

   Benefits of IaaS:
   - A typical infrastructure as a service offering saves both time and money as the underlying hardware set up and support is provided by the service provider.
   - Resources are available on demand as and when required so there is no wastage of any unused resources and no delays on adding any resources.
   - Utility-based pricing model i.e., pay only for the resources you actually use.

2. **Platform as a Service (PaaS)**: PaaS is an advanced version of IaaS. Apart from just providing the IT infrastructure, PaaS also provides the computing platform and solution stack as a service. PaaS

lets software developers build custom applications online without having to worry about data storage, data serving, and management.

Benefits of PaaS:
- PaaS makes software development easy even for non-experts as anybody can develop an application through the web browser with just a single click functionality.
- There is no need for the users to upgrade or update the infrastructure as the PaaS service provider handles all the update patches, upgrades, and regular software maintenance.
- PaaS provides location independence as developers in different locations can work together on the same application build.
- There is no need to invest in physical infrastructure or in expertise required to manage it, an example of this would be SAP. The ability to rent virtual IT infrastructure brings in great cost benefits for the users.

3. **Software as a Service (SaaS):** SaaS is a special cloud computing service that incorporates both IaaS and PaaS service offerings. SaaS is a cloud computing service that provides application-level services tailored to diverse business needs such as business analytics, CRM, or marketing automation. SaaS is a cloud computing service offering that provides web-based software applications to customers on-demand. SaaS is a subscription-based offering where users subscribe to software on a monthly basis instead if purchasing it so there are no upfront costs involved. It also provides a provision to the users to end the subscription when it is no longer needed. SaaS is the most familiar cloud computing service offering as users most often interact directly with SaaS applications like Netflix, Gmail, JIRA, Dropbox, or Salesforce.

Benefits of SaaS:
- There is no initial setup cost as the users can make use of the application as soon as they subscribe. There is no hardware cost as well because the processing power is supplied by the service provider.
- Flexible payments as the users pay for the services on a pay-as-you-go model.
- Any updates to the software are automatic and free of charge.
- SaaS provides cross-device compatibility because SaaS applications can be accessed through any internet enabled devices, such as laptop, smartphone, or desktop.
- Enterprises need not engage an IT expert to download the software on multiple systems in the office nor have to worry about the keeping the software up-to-date on every PC.

6) What is Azure Data Factory?

Answer: Azure Data Factory is a cloud-based integration ETL service offered by Microsoft.
that help us to create data-driven workflow for orchestrating and automating data movement and data transformation overcloud.
Data Factory services also offer to create and running data pipelines that move and transform data and then run the pipeline on a specified schedule.

7) What is the component/Building block of Azure Data Factory?

Answer: Below are the component of ADF:

a) **Pipeline**:
- A pipeline is a logical grouping of activities that together perform a task.
- A data factory can have one or more pipelines.
- The Activities in a pipeline define actions to perform on your data

b) **Activity**:
- The Activities in a pipeline define actions to perform on your data.
- Data factory support three types of activities:
  a. Data movement activities
  b. Data transformation activities
  c. Control activities.

c) **Dataset**:
- Datasets identify data within different data stores, such as tables, files, folders and documents.
- Before we create a dataset, we must create a linked services to link our data store to the data factory.

d) **Linked Services**:
- Linked service are like connection strings, which define the connection information needed for data factory to connect to external resources.
- For Example: To copy data from Blob storage to a SQL database, we need to create two linked services: Azure Storage & Azure SQL Database.

e) **Trigger**:
- Triggers are used to schedule an execution of pipeline.
- Pipelines and triggers have a many to many relationships. Multiple triggers can kick off a single pipeline or a single trigger can kick off multiple pipelines.

f) **Integration Runtime**:
- The Integration Runtime (IR) is the compute infrastructure used by Azure Data Factory.
- Following types of integration runtime available:
  a. Azure (AutoresolveIntegration Runtime- Default) -> Cloud Integration
  b. Self-Hosted Integration Runtime -> OnPrem Integration
  c. Azure SSIS -> OnPrem SSIS Integration

g) **Data Flow:**
- Data Flows allow data engineers to develop graphical data transformation logic without writing code.
- Data Flows are executed as activities within Azure Data Factory Pipeline using scaled-out Azure Databrick clusters.

8) How many types of triggers are supported by Azure Data Factory?

Answer: There are four types of triggers available in Azure Data Factory:

- **Schedule trigger**: When we create a schedule trigger, we specify scheduling and recurrence. We can use schedule trigger to limit the number of trigger executions. For example, if a trigger with a monthly frequency is scheduled to run only on day 31, the trigger runs only in those months that have a thirty-first day. We can also use schedule to expand the number of trigger executions. For example, a trigger with a monthly frequency that's scheduled to run on month days 1 and 2, runs on

the first and second days of the month, rather than once a month. If multiple schedule elements are specified, the order of evaluation is from the largest to the smallest schedule setting: week number, month day, weekday, hour, minute.

- **Tumbling window trigger**: Tumbling window triggers are a type of trigger that fires at a periodic time interval from a specified start time, while retaining state. Tumbling windows are a series of fixed-sized, non-overlapping, and contiguous time intervals.
  The tumbling window trigger run waits for the triggered pipeline run to finish. Its run state reflects the state of the triggered pipeline run. For example, if a triggered pipeline run is cancelled, the corresponding tumbling window trigger run is marked cancelled. This is different from the "fire and forget" behaviour of the schedule trigger, which is marked successful as long as a pipeline run started.

- **Event-based trigger**: An event-based trigger runs pipelines in response to an event. There are two flavours of event-based triggers.

  - **Storage event trigger** runs a pipeline against events happening in a Storage account, such as the arrival of a file, or the deletion of a file in Azure Blob Storage account.
  - **Custom event trigger** processes and handles custom articles in Event Grid.