Cloud Storage

Cloud storage is making inroads and increasingly becoming quite common in enterprises these days due to the advantages it offers in terms of availability, durability and cost.

Cloud storage solution can either be deployed in the private cloud or accessed over the internet in the public cloud depending upon the sensitivity of the data and compliance requirements. Cloud storage solutions hosted on the public cloud can be easily provisioned since the cloud provider manages the underlying infrastructure supporting the cloud storage solution.

Benefits

●**Availability**– All the data is highly available when needed. Depending upon the type of data, there might be a requirement for the storage solution to be highly available. In some scenarios, where the data is archived or the data itself could be recreated e.g. image thumbnails from the original image, the cost of the storage could be reduced by having a slightly less availability (e.g. 99.9% instead of 99.99%).
●**Durability**– Data is stored redundantly across multiple facilities and physical devices to avoid the loss of data in case of natural disaster or device failure.
●**Cost**– The total cost of ownership forstoring the data in the cloud is less since there is no CAPEX involved in buying thehardware and the inital set-up. Also, there is no operational cost involved in maintaining the hardware.The capacity on the public cloud storage can be added or removed as needed, thus reducing the cost since you don't pay for the 'idle' infrastructure which is provisioned ahead for the future use.

Storage Types

There are primarily three types of storage solution provided by the cloud providers. Each of these storage types provides unique features and functionality for their respective use cases.

Block Storage

Block storage stores the files/data by splitting the data into small chunks known as block which can be accessed by the application using the memory address of the block. The files stored in the block storage can be updated by replacing the data stored in the blocks. This type of storage is analogous to SAN (Storage Area Network).

Enterprises use SAN as a block storage that could be attached to the servers to appear as if the storage volume is running locally. Servers get access to dedicated SAN logical units (virtual hard disk) which are not shared among the servers. There are 2 protocols used to access the SAN storage – Fiber Channel and iSCSI (Internet Small Computer System Interface). iSCSI uses the existing ethernet network to access the SAN and hence has lower cost, however it has lower performance when compared to fiber channel.

Below diagram shows the SAN having Logical Unit Numbers shared among multiple servers.
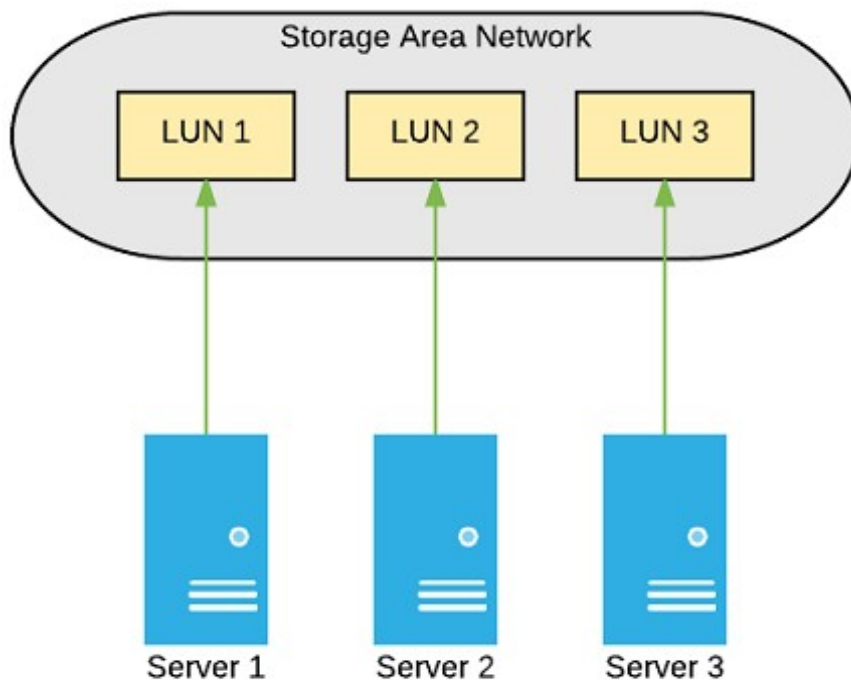
Elastic Block Store (EBS) is the Amazon service for block storage in the cloud. EBS is used as a disk volume when an EC2 instance is provisioned. It is also suitable for running the database. EBS volumes are replicated within the availability zone and can be mounted to only one EC2 instance running within the same availability zone.

File Storage

File storage is also based on the block storage having a file system optimized for serving files with large number of network connections. It is analogous to NAS (Network Attached Storage).

Enterprises use NAS to store files that are shared by multiple systems. NFS protocol is used for accessing the files stored on the NAS share.

Elastic File System (EFS) is the AWS service for storing and sharing the files in the cloud. EFS can be mounted on multiple servers at the same time for sharing the files. It can also be mounted to on-premise servers over the VPN or DirectConnect. EFS is replicated across multiple availability zones within the AWS region.

Object Storage

Traditionally, enterprises have stored the objects (unstructured data) on disks running OS with a file system. This solution can only scale to a certain level due to the limits imposed by the file system. Object storage addresses the limitations of storing the objects in file system.

Object storage is a key-value storage in which the objects are stored as a single entity as opposed to multiple blocks in block storage. Any update to the object requires replacing the full object with a new version. Object storage is not suitable for hosting operating system or running databases. It can however be used for storing the backup or snapshot volumes of the disk or EBS.

Objects are accessed via. the HTTP protocol using RESTful API allowing the application to create, update, copy and delete the objects in the bucket/container. Object storage can store billions of objects and still offer a very good performance due to the underlying key-value storage. Each object is addressed as a URL. Unlike file system, objects are not stored in a hierarchy, however object names can have '/' providing a pseudo-nested hierarchy for organizing the objects within the namespace.

Object consists of the actual file, metadata and the globally unique identifier. The metadata consists of the contextual information like what the data is, its confidentiality and any other data required for accessing the object.

Simple Storage Service (S3) is the Amazon service for storing the objects in the cloud.