Cloud Storage

Dr Mukti Padhya

Assistant Professor

NFSU, Gandhinagar

What is Cloud Storage?

- Cloud storage is a cloud computing model that stores data on the Internet through a cloud computing provider who manages and operates data storage as a service.

- It's delivered on demand with just-in-time capacity and costs, and eliminates buying and managing your own data storage infrastructure.

- This gives you agility, global scale and durability, with "anytime, anywhere" data access.

How Does Cloud Storage Work?

- Cloud storage is purchased from a third party cloud vendor who owns and operates data storage capacity and delivers it over the Internet in a pay-as-you-go model.

- These cloud storage vendors manage capacity, security and durability to make data accessible to your applications all around the world.

- Applications access cloud storage through traditional storage protocols or directly via an API.

- Many vendors offer complementary services designed to help collect, manage, secure and analyze data at massive scale.

Benefits of Cloud Storage

- Total Cost of Ownership.
  - With cloud storage, there is no hardware to purchase, storage to provision.
  - You can add or remove capacity on demand, quickly change performance and retention characteristics, and only pay for storage that you actually use.
  - Less frequently accessed data can even be automatically moved to lower cost tiers in accordance with auditable rules, driving economies of scale.
- Time to Deployment. Cloud storage allows IT to quickly deliver the exact amount of storage needed, right when it's needed.
- Information Management. Centralizing storage in the cloud creates a tremendous leverage point for new use cases.
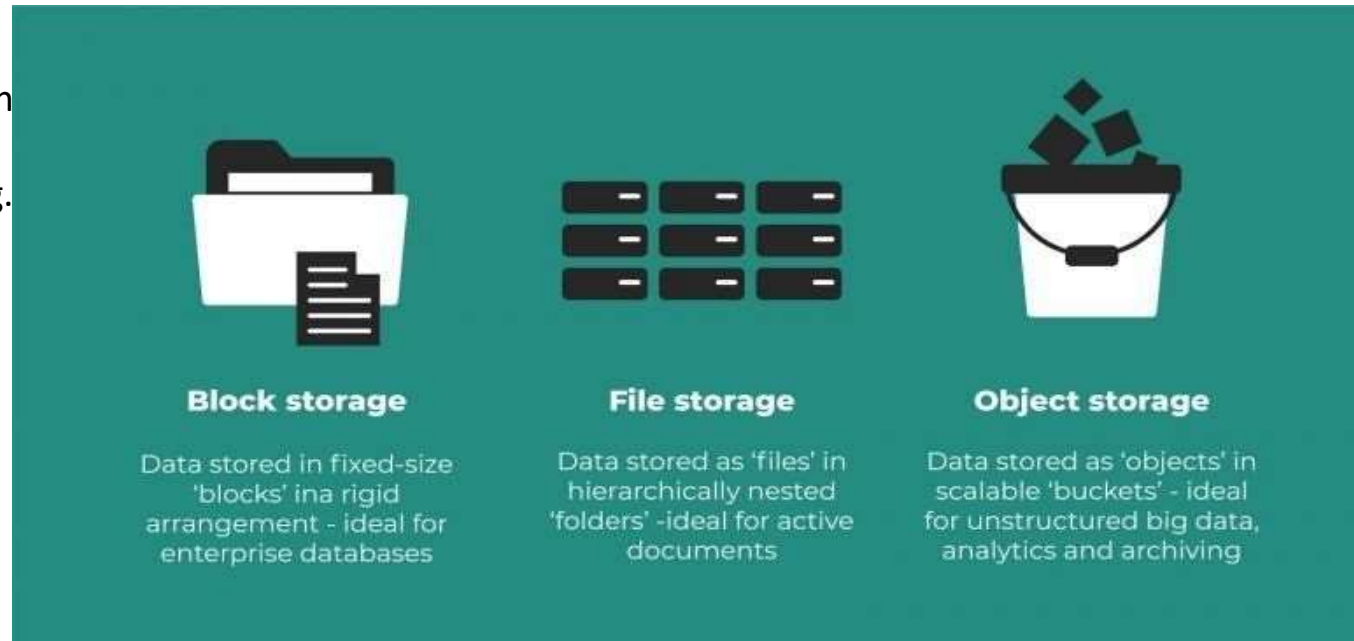
Cloud Storage Requirements

- Durability. Data should be redundantly stored, ideally across multiple facilities and multiple devices in each facility. Natural disasters, human error, or mechanical faults should not result in data loss.

- Availability. All data should be available when needed, but there is a difference between production data and archives. The ideal cloud storage will deliver the right balance of retrieval times and cost.

- Security. All data is ideally encrypted, both at rest and in transit. Permissions and access controls should work just as well in the cloud as they do for on premises storage.

Types of Cloud Storage

- Object Storage
- File Storage
- Block Storage

Types of Cloud Storage

- File-based storage and block-based storage are well-suited to structured data and continue to work well in certain scenarios.
- But the Internet has changed everything.
- Organizations struggle to manage mounting volumes of web-based, digital content (unstructured data).
- Object-based storage can meet this challenge.

Block storage

Data stored in fixed-size 'blocks' ina rigid arrangement - ideal for enterprise databases

File storage

Data stored as 'files' in hierarchically nested 'folders' -ideal for active documents

Object storage

Data stored as 'objects' in scalable 'buckets' - ideal for unstructured big data, analytics and archiving

Object Storage

- Object storage or object-based storage represents a data storage architecture that allows you to store large amounts of unstructured data in a highly scalable manner.

- Data types include email, images, video, web pages, audio files, datasets, sensor data and other types of media content;

- Object storage offers the preferred method for data archiving, backup, and more or less for storing any type of static content
  - For large enterprises and organizations, storing and managing this unmatched amount of data

- It is cloud-based storage - hosts data on a different device than the one we use to access it.
  - keeps our data safe from any hardware-related problems.

How does it work?

- In object-based storage devices or systems, there are no folders, directories, files, or any hierarchies.

- Instead, these systems store all data in a flat data environment as objects.

- Each object contains the data itself, along with some descriptive information associated with that object, known as metadata, and a globally unique identifier.

Object storage database?

- An object storage database provides a directory of sorts that uses the object's metadata to locate the appropriate data files in a distributed storage system.

- Each object storage group has an object storage database that contains two tables. One table is an object directory and the other table is for the object storage.

- The object directory table contains descriptive information about each object (the metadata).

- This directory keeps track of all objects in the storage hierarchy by recording the collection name identifier, the object name, and other pertinent information.

- The second table in the object storage database is the object storage table, which contains the data content/file itself (the objects).

- The data (fixed digital content such as video and image files or large libraries of documents) sits in the object store, while the metadata (contextual information about the data, including the name ID) resides in a database/object directory table.

Object storage database?

- For example, in IBM's object storage methodology, the object directory table includes three "indexes":
  - The Object Creation Time Stamp
  - The Collection Name Identifier (Name ID), Pending Action Data and Object Creation Time Stamp
  - The Object Name and Collection Name Identifier

How to access object storage?

- The objects in an object-based storage system are accessed via Application Programming Interfaces (APIs).

- This space is managed by the graphic interface (dashboard) offered by the provider directly on their website.

- The dashboard is already integrated with object storage using an API.

- Example S3 API from Amazon.

Benefits of object storage

- Storing/managing unstructured data

- Disaster recovery/availability

- Customizable metadata

- Cloud compatibility

- **Greater data analytics**. Object storage is driven by metadata, and with this level of classification for every piece of data, the opportunity for analysis is far greater.

- **Infinite scalability.** Keep adding data, forever. There's no limit.

- **Faster data retrieval**. Due to the categorization structure of object storage, and the lack of folder hierarchy, you can retrieve your data much faster.

- **Reduction in cost**. Due to the scale-out nature of object storage, it's less costly to store all your data.

- **Optimization of resources**. Because object storage does not have a filing hierarchy, and the metadata is completely customizable, there are far fewer limitations than with file or block storage.

Object storage use cases

- Data recovery & backups – with object storage, you can securely and cost-efficiently store your data

- Analytics – when you have a large amount of data, for example data that is used to train different AIs or to run analytics, then object storage is a great solution to use it as a data lake

- Static content – you can use object storage to serve all the static content for your web app or all the content that's created by your application

What is Amazon S3 Storage?

- Applications developed in the cloud often take advantage of object storage's vast scalablity and metadata characteristics.

- Object storage solutions like Amazon Simple Storage Service (S3) are ideal for building modern applications from scratch that require scale and flexibility

- It can be used to import existing data stores for analytics, backup, or archive.

- Amazon S3 is a storage system for the internet, where you can store and retrieve any amount of data, anytime, anywhere.

- This make web-scaling computing easier for developers, and it also gives them access to the infrastructure that Amazon uses to conduct a global network of websites.

Dr Mukti Padhya : Cloud Sec@MSc_ 2023

File Storage

- File storage organizes and stores data inside a folder.

- Files are named, tagged with metadata (typically the file name, file type, and when it was created and last updated) and organized in folders under a hierarchy of directories and subdirectories.

- A hierarchical storage system like this works well with relatively small, easily organized amounts of data.

- However, as the number of files grows, the search and retrieval process can become cumbersome and time-consuming.

What is Amazon Elastic File System (EFS)?

- Some applications need to access shared files and require a file system.

- This type of storage is often supported with a Network Attached Storage (NAS) server.

- File storage solutions like Amazon Elastic File System (EFS) are ideal for use cases like large content repositories, development environments, media stores, or user home directories.
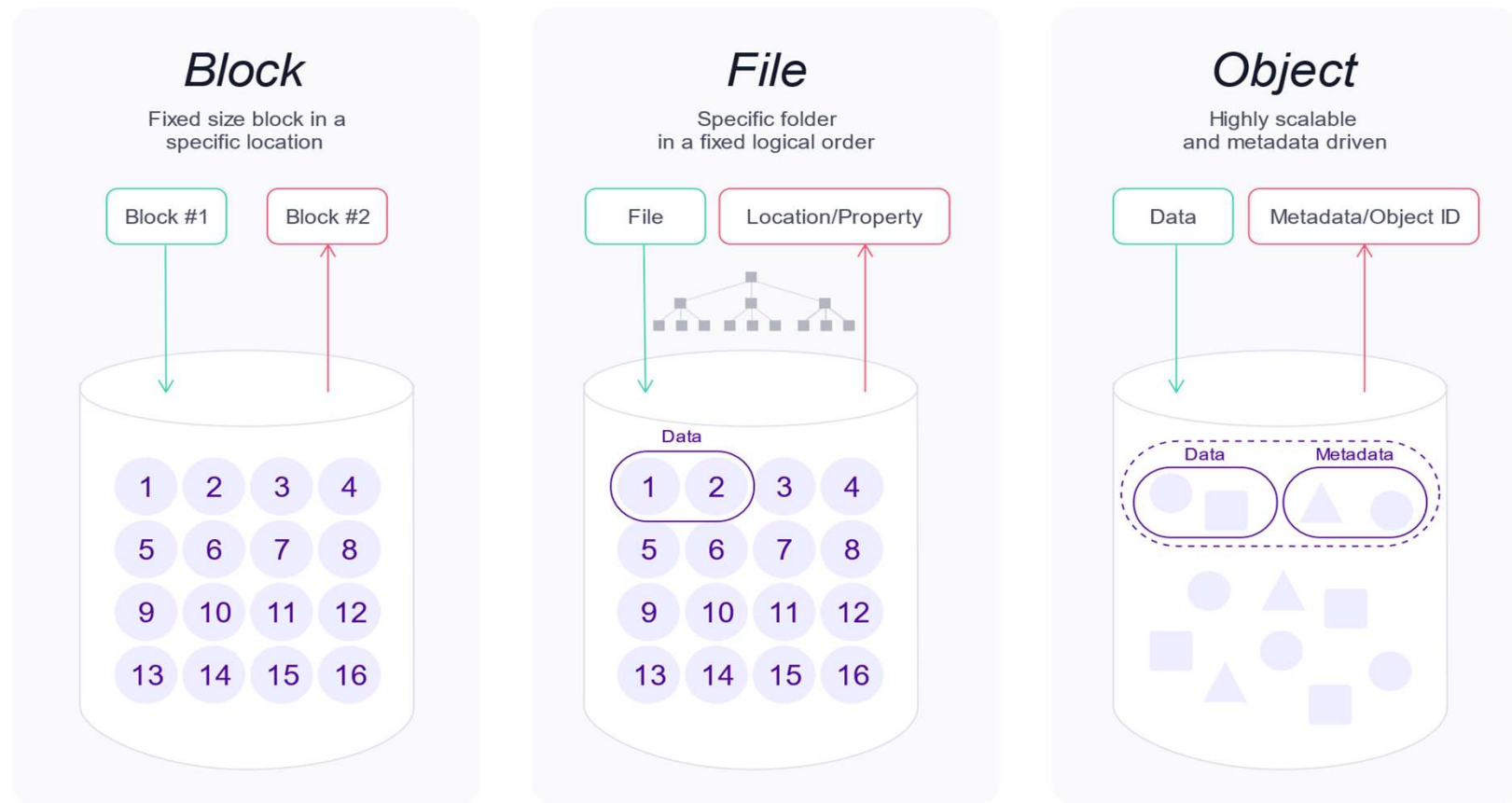
Block Storage

- Block storage offers an alternative to file-based storage—one with improved efficiency and performance.

- Block storage breaks a file into equally-sized chunks of data and stores these data blocks separately under a unique address.
  - One can store the collection of blocks anywhere in the system for maximum efficiency.

- To access a file, a server operating system uses the unique address to pull the blocks back together, assembling these into the file.

- You will gain efficiencies as the system does not need to navigate through directories and file hierarchies to access the data blocks.

- Block storage works well for critical business applications, transactional databases and virtual machines that require low-latency (minimal delay), granular or more detailed access to data, and consistent performance.

What is Amazon Elastic Block Store (EBS) ?

- Enterprise applications like databases or ERP systems often require dedicated, low latency storage for each host.

- This is analagous to direct-attached storage (DAS) or a Storage Area Network (SAN).

- Block-based cloud storage solutions like Amazon Elastic Block Store (EBS) are provisioned with each virtual server and offer the ultra low latency required for high performance workloads.

File vs Block vs Object storage

File vs Block vs Object storage

- File storage is network-attached storage where data is stored in folders.

- When a file needs to be accessed the computer must know the full path to find it.

- A good use case of it is when you have a mix of structured and unstructured data or you want to share the data with many users at once.

- It can store just about anything and you use it any time you access your files from your PC.

File vs Block vs Object storage

- Block storage saves data in raw blocks and different from file storage it can be accessed through a Storage Area Network (SAN).

- Meaning that in a block storage architecture the servers that access the storage can be on different networks as well.

- A good use case for block storage is when the requirements imply a very low latency and a good and consistent performance for I/O operations, such as a database.

File vs Block vs Object storage

- Object storage is a type of architecture where each file is saved as an object and it can be accessed through an HTTP request, usually GET.

- This type of storage is the best fit for scenarios where a lot of unstructured data need to be managed.

- In object storage, each object receives a unique id, that consumers will use to retrieve it and rich metadata that can be from privacy policy up to anything you want.