Course: Cloud Computing

Week 3: Data Storage in cloud Computing

Lecturer: Ikwap Flavia Agatha
MSc. Computer Forensic
PHD in IT (Candidate)
University: Kumi University

Lecture learning outcome:

- At the end of the this class you will be able to
- understand cloud data
- Understanding cloud Storage
- Comprehend the Sources of cloud data
- Understand the Cloud storage Architecture
- Understand the Cloud storage Systems/Technologies
- Comprehend the Pros and Cons of Cloud Storage

Introduction to data in the cloud

Types of Data

- Structured data; well organize in form of tables in relational databases and in spreadsheets, it's in a fixed structure format of capturing data
- Unstructured data does not have a pre-defined data model; can be text detailed sentences, data containing dates, numbers, and untagged data representing photos and graphic images. Word processing, documents, presentations, and PDF files are prime examples of unstructured data.

Introduction to data in the cloud

- Semi-structured data; Semi-structured data is not the raw data and is not stored in a conventional database system. It is structured data but is not organized in a rational model like a table or an object-based graph. Semi-structured data contains tags or markers to separate semantic elements
- Quasi-structured data is more of a textual data with erratic data formats. This data type includes web clickstream data such as Google searches.

Big data concept

- Big data has resulted from the flow of huge data into the cloud; it refers to large and complex datasets that are difficult to manage and process

Characteristics that describe big data

- **Volume**: Enormous or massive data, ranging from terabytes to petabytes and beyond coming from various sources like social media applications, databases, management information systems, sensors, machines, transactions, and interactions, resulting into large quantities of data.

Big data concept

- **Velocity**: Big data is generated at high speed, all the time, so to say in real-time. Examples of high-velocity data sources include social media streams, sensor data from IoT devices, and financial market data like forex trade.
- **Variety**: Big data comes in diverse formats and types, including structured, semi-structured, and unstructured data in form of text, images, videos, sensor data, and log files.

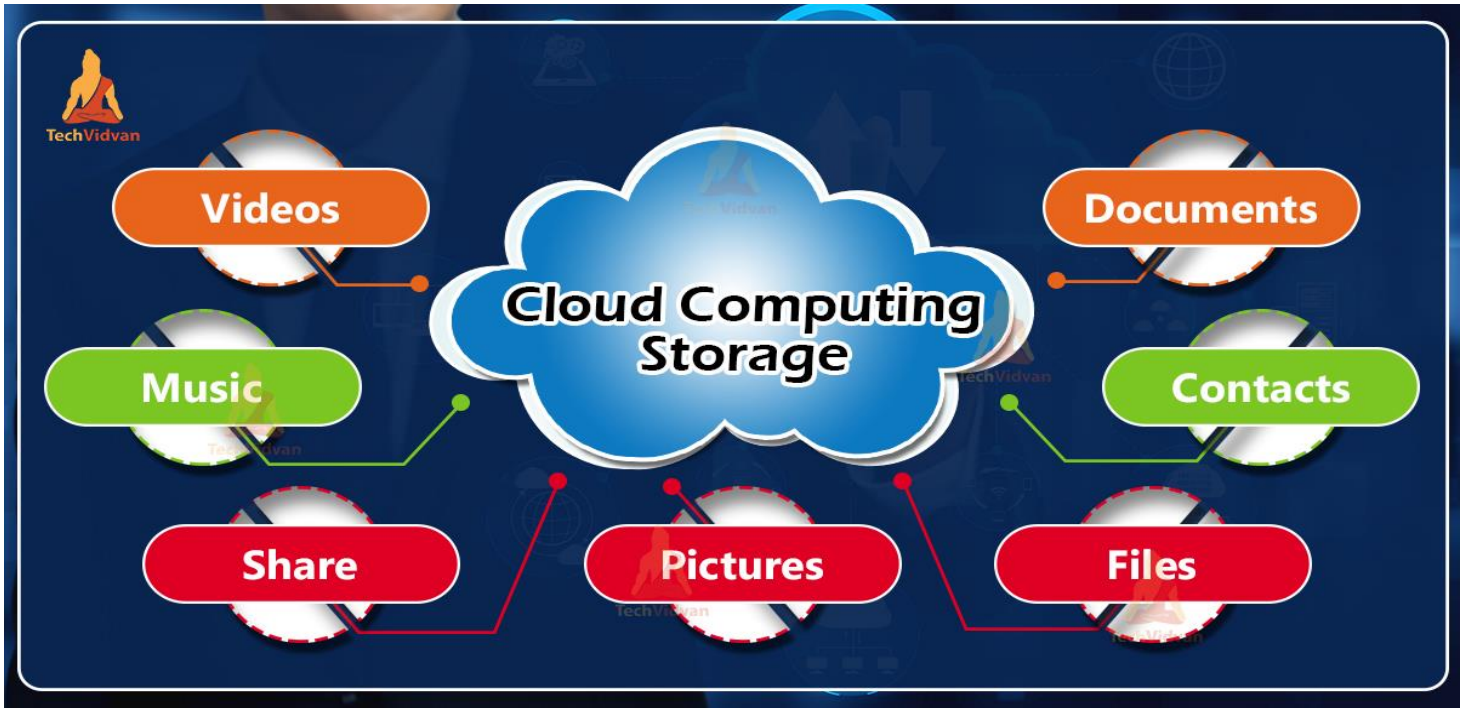
Storage

- **Storage:** is commonly used to define digital devices designed to hold data.
- Data storage system includes;
- **Storage medium** like hard drives, servers, CDs, DVDs, Flash drives, Solid state drives and memory cards
- **Storage device-** the mechanical apparatus that records and retrieves data from a storage medium

Cloud Storage

- **Cloud storage** :Refers to a virtual storage area that can span across many different physical storage devices. When you use cloud storage, some of your files may be on a physical server in New York while other files are on a physical server in California yet you are in Uganda, these include files like Videos, Photos etc

Cloud Storage



- https://i2.wp.com/techvidvan.com/tutorials/wp-content/uploads/sites/2/2021/09

Cloud Storage

- Cloud storage gathers a large number of different types of storage devices through the application software which are based on the functions of the cluster applications, grid techniques, distributed file systems, etc.
- Almost all information accessed through the internet is stored in the cloud; for example while using an internet- based email services like yahoo, Gmail or live; you have access to your email from any part of the world as long as you have internet connectivity, simply the mail data is stored on servers owned by email service provider and not on one's local computer.

Sources of Cloud data:

- Cloud data include all data that is stored, analyzed, and accessed through cloud computing platforms
- Data can be channeled to the cloud through:

Sources of Cloud data:

- **Enterprise Applications**: A number of organizations like educational, health, industries, and business are deploying their applications on cloud. This includes Customer Relationship Management (CRM) systems, Enterprise Resource Planning (ERP) software, Human Resources Management Systems (HRMS), and other business-critical applications. All Data coming from these applications reside in the cloud servers.

Sources of Cloud data:

- **File Storage and Content Management**: Services like Cloud drive, A drive, one drive, iCloud, Dropbox, Amazon S3 Enable users to store, share, and collaborate on files stored.

- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.lemmymorgan.com%2Fonline-data-storage-services%2F&psig

Sources of Cloud data:

- **Mobile Applications**: Most mobile applications like what sup, rely on cloud storage and synchronize user generated data.
- **Data Lakes and Warehouses**: Organizations use cloud-based data lakes and data warehouses to centralize and analyze large volumes of structured and unstructured data from various sources.
- **Backup and Disaster Recovery**: Cloud-based backup and disaster recovery solutions generate a lot of data stored in the cloud.

Sources of Cloud data:

- **Web Applications and Services**: Cloud-based web applications and services produce large amounts of data. This includes data from user interactions, transactions, logs, and other activities. Social media platforms, e-commerce websites and collaboration tools are examples of cloud-based services that generate significant amounts of data.
- **IoT Devices**: The Internet of Things (IoT) incorporates numerous devices that collect and transmit data to cloud platforms for storage. These devices include sensors, wearables, connected appliances, smart meters, and industrial equipment.

Cloud storage Architecture

- **Components**
- **Nodes**: Include physical servers or virtual machines in control of data storage in the cloud. Storage nodes are made of hard disk drives (HDDs), solid-state drives (SSDs), These nodes are organized into clusters or storage pools to form a distributed storage system.

Cloud storage Architecture

-Components

- **Storage Management Software**: Offers functionalities for data organization, replication, authentication, access control, and data security. It includes components for data deduplication, compression, encryption, and RAID (Redundant Array of Independent Disks) for fault tolerance.
- **Data Replication and Redundancy**: Cloud storage systems make backup copies of data across multiple data centers, to ensure data availability and fault tolerance. This redundancy helps in reducing the possibility of data loss due to hardware failures or disasters.

Cloud storage Architecture

-Components

- **Data Access Interfaces**: Cloud storage architectures support various access interfaces, including object storage, block storage, and file storage. Object storage interfaces (e.g., Amazon S3, Azure Blob Storage) are suitable for storing unstructured data like images, videos, and documents. Block storage (e.g., Amazon EBS, Azure Disk Storage) provides raw storage volumes for virtual machines or databases. File storage (e.g., Amazon EFS, Azure Files) offers shared file systems accessible via standard file protocols like NFS or SMB.

Cloud storage Architecture

-Components

- **Network Infrastructure**: High bandwidth is vital for efficient data access and transfer within cloud storage architectures. Content delivery networks (CDNs) may be employed to cache and deliver frequently accessed data closer to end-users, reducing latency and improving performance.

Cloud storage Architecture

-Components

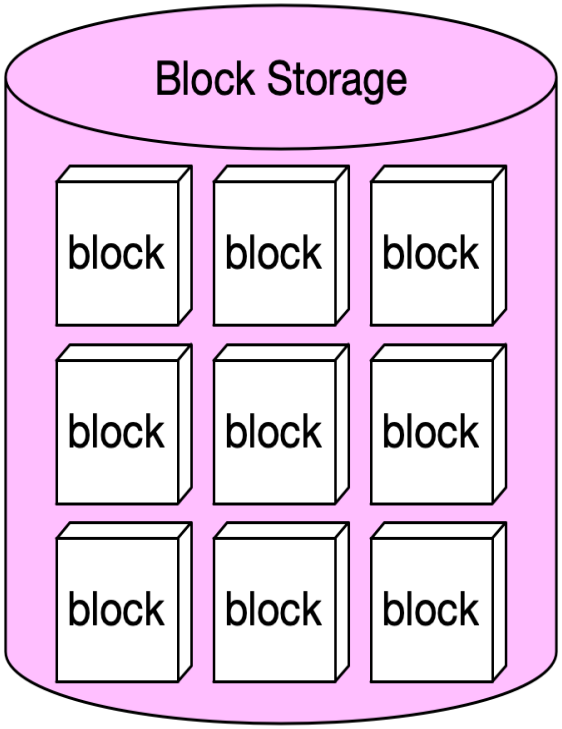
- **Data Tiering** : Cloud storages are built with tiered storage models, where data is categorized depending on its access frequency and performance requirements. Hot data that requires frequent access is stored on high-performance storage tiers, while cold data is moved to lower-cost, slower storage tiers

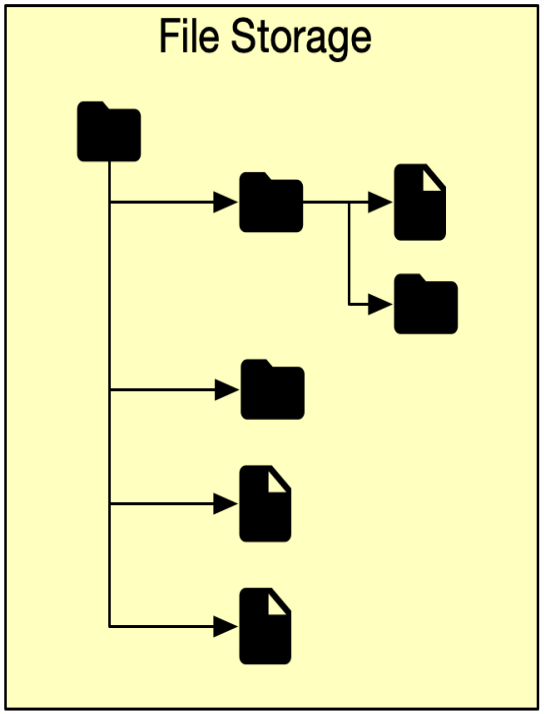
Cloud storage Architecture

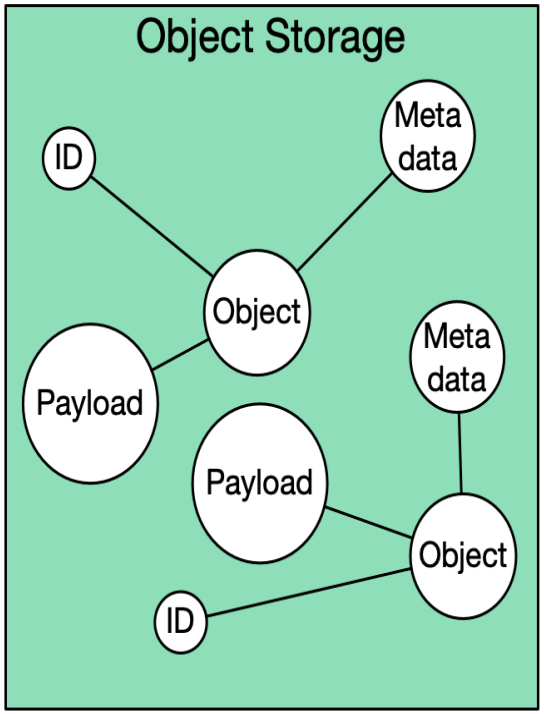
-Components

- **Scalability**: Due to vast data volumes and increasing demand Cloud storage architectures are designed to scale horizontally to be accommodative. Storage nodes can be added and resources allocated dynamically.
- **Security and Compliance**: Cloud storage consists of security measures like data backups, access control, data encryption, authentication and audit trails, Compliance with industry regulations and standards such as GDPR, HIPAA, and PCI DSS is also a consideration in designing cloud storage solutions.

Cloud Storage systems







%2Fstorage-systems-overview&psig

Cloud Storage systems

- These are approaches used to store and manage data in cloud computing ecosystems. Enabling remote data storage for clients, the infrastructure is provided by cloud service providers hence offering scalability, flexibility, and accessibility.
- The technologies include:

File Storage:

- Data is stored in one single unit of the information under the hierarchy of folders. Accessible over standard network protocols such as NFS (Network File System) or SMB (Server Message Block). File storage is suitable for applications that require shared access to files and directories, such as file sharing, content management, and home directories. Data retrieval involves the use of a limited amount of metadata that directs to the exact path. Most suitable for structured data, Cloud file storage services include Amazon EFS (Elastic File System), Google Cloud Filestore, and Azure Files.

Benefits of File Storage

- Simple storage system for managing files and folders on drives or
- NAS (Network Attached Storage). Files have names and limited metadata
- Sharing files is easy since files can be centralized on a LAN and can also be stored in (Network Attached Storage); the file is by user based on permissions granted.
- Uses familiar protocols: Operating Systems like Linux and Windows uses SMB (Server Message Block) and NFS (Network File System) and permits to read and write files to servers over the network

Drawbacks of File Storage

- 1. Data Redundancy: File storage is manual and tedious this can result into data redundancy.
- 2. File searching: when large amounts of data are stored, file searching get difficulty, data redundancy worsens the search process.

Object Storage:

- Object storage is a storage architecture that manages data as objects rather than as blocks or files.
- Object storage provides a high level of abstraction to the end-user. The end-user just interacts with the simple interface to store, retrieve and delete objects.
- All the operations are carried through requests and every time an end-user queries object-store all the requests are authenticated (authentication of end-user) and authorized (check privileges and access control to end-user)

Object Storage:

- Examples of object storage services include Amazon S3 (Simple Storage Service), Google Cloud Storage, and Azure Blob Storage.

Object storage approach to storing and referencing data.

Constructs:

- Data: This is the user and application data that requires continuous storage. It can be text, images, videos, binary formats, multimedia, or any other human- or machine-generated content.
- Metadata: This is the data about the data. It includes some predefined attributes such as upload time and file size. The Meta data is stored with the object.

Constructs:

- A universally unique identifier (UUID): This ID is assigned to every object in an OBS system. This UUID allows the object storage system to differentiate objects from one another and is used to find the data without needing to know the exact physical drive, array, or site where the data is

Data Access

- Data access is achieved by using a REST interface over the HTTP protocol, which allows anywhere and anytime access simply by referencing the object ID. The file name of the original data is maintained as a key part of the object metadata, so it can also be used for object access in some object implementations

Benefits of Object Storage

1. Infinite Scalability; Object storage has the capability to handle increasing scalability and extensibility; it works by binding the data and associated metadata in a single container with a distinct identifier, making it easy to add storage servers.

2. Security; Object storage is deployed over private networks and each request from the end-user comes through the authentication and authorization filter, therefore providing improved security.

Benefits of Object Storage

3. Cost Reduction; Scaling Object storage is easy hence reducing on backup costs and time demand.

4. Data recalls at faster rates since metadata and identifiers are used.

Drawbacks of Object Storage

1. Less compatibility with traditional database

Object storage mostly handles unstructured data, this make it very hard to roll back to traditional databases which hold data in a more structured.

3. Can’t be mounted on Operating Systems like a normal disk, it requires some clients and plugins.

Block Storage:

- Block storage is a data storage approach in which each storage block acts as an independent hard drive.
- Block storage provides raw storage volumes that can be accessed and managed as individual blocks by applications or virtual machines. Data is split into equal sized block, with an address tagged to each block and the address is written to an underlying medium, such as disk or flash. No metadata is provided, only the block location is considered.

Block Storage

- The file system or the application accessing the block storage keeps track of the location, when presented to the host cannot be shared unless special clustering or locking software is used.
- Connections to hosts in an enterprise setting are typically over a storage area network (SAN) and Direct attached storage

Block Storage

- Cloud providers who extent block storage services include: Amazon EBS (Elastic Block Store), Google Persistent Disks, and Azure Disk Storage.

Benefits of Block Storage

1. High Performance: Block storage provides very low latency and fast retrieval, commonly used in database servers to handle queries simultaneously.

2. Flexibility and Scalability; its easy to add new block storage from time to time hence enable scaling. Block can moved with easy between servers by swapping destination routes.

4. Bootability in OS: Operating systems can be easily booted directly from blocks using SAN (Storage Area Network)

Drawbacks of Block Storage

1. Metadata limitation: Block Storage uses a constrained amount of metadata, much lesser than other types of storage. This can affect application storage retrieval and searching operations because applications cannot recognize the location and may have to perform a scan on a large number of blocks.

2. Server Binding; Block storage embedded into the server, limiting it from accessing nodes at the same time therefore requiring a software which adds over head to the system.

3. Cost; SAN (Storage Area Network) requires heavy investment and highly trained individuals for maintenance.

Distributed File Systems:

- A distributed file system enables programs to store and access remote files exactly as they do local ones, allowing users to access files from any computer on a network. They also distribute data across multiple storage nodes in a cluster to provide high availability, fault tolerance, and scalability. These systems use distributed data protocols and replication mechanisms to ensure data integrity and reliability

Distributed File Systems:

- Examples of distributed file systems used in cloud storage include Hadoop Distributed File System (HDFS)-(The Hadoop Distributed File System-which is designed to store very large data sets reliably, and to stream those data sets at high bandwidth to user applications) and GlusterFS.

Google File System

- The architecture of the file system is based on a master node, which contains a global map of the file system and keeps track of the status of all the storage nodes, and a pool of chunk servers, which provide distributed storage space in which to store files

Google File System

- Files are logically organized into a directory structure but are persisted on the file system using a flat namespace based on a unique ID. Each file is organized as a collection of chunks that are all of the same size. File chunks are assigned unique IDs and stored on different servers, eventually replicated to provide high availability and failure tolerance.

Centralized file system

Aneka storage service offers:

- A centralized file storage- for computing intensive applications, Compute-intensive applications mostly require powerful processors and do not have high demands in terms of storage, which in many cases is used to store small files that are easily transferred from one node to another
- It provides distributed applications with the basic file transfer facility and abstracts the use of a specific protocol to end users and other components of the system,

Centralized file system

- which are dynamically configured at runtime according to the facilities installed in the cloud. The option that is currently installed by default is normal File Transfer Protocol (FTP). To support different protocols, the system introduces the concept of a file channel that identifies a pair of components: a file channel controller and a file channel handler used to upload, download, or browse files.

Cloud Storage verses traditional storage					
	Traditional Storage	Cloud Storage		Traditional Storage	Cloud Storage
Architecture	Servers, Storage Area Network (SANs), Network-attached storage (NAS)	Distributed Systems, multiple data centers – internet access	Security	Organizations manage security	Cloud providers implement security
COST	Large investments	Pay as you go	Scalability	Limited to capacity of on premise Hardware	Can provide additional storage on demand
Management	In-House IT Team	Cloud providers manage	Accessibility	Limited to organizational users	Can be accessed anywhere with internet connection

Advantage of Cloud Storage:

1. Off-Site Backup:

2. File Accessibility:

3. Security of Data:

5. Ease of Management:

6. Cost-effectiveness:

Disadvantage of cloud storage:

1. Dependent on a third party:

2. Dependent on Internet Speed:

4. Loss of Control:

Next week

- Next Week Lecture will cover;
- Networking in the Cloud

Reference

- Buyya, R. (2013). *Mastering Cloud Computing.* United States of America: MK Publishers.
- Huang, L. (2012). Data Storage Technology and Its development based on Cloud Computing. *Proceedings of the 2012 2nd International Conference on Computer and Information Application* (pp. 1-4). China: Atlantis Press, Paris, France.
- Kulkarni, G. (2012). Cloud Storage Architecture. *nternational Conference on Telecommunication Systems, Services, and Applications* (pp. 1-6). Pune India: IEEE.
- Liu, K. (2012). Research on Cloud Data Storage Technology and Its Architecture Implementation. *International Workshop on Information and Electronics Engineering*, 1-5.
- Velte, A. T. (2010). *Cloud Computing: A Practical Approach.* New york: MC Graw Hill.
- Zehau, C. (2013). A Data Storage and Management Scheme in Cloud Storage Model . *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering* (pp. 1-4). Wuhan: Atlantis Press, Paris, France.