

Cloud Computing

Prepared by Dr. Sherif Mostafa

IT381

Lecture 8



Storage Virtualization

What is cloud storage?

What is Storage Virtualization?

Types of Storage Virtualization

How does storage virtualization work

Storage Area Network (SAN)

Network Attached Storage (NAS):

NAS or SAN: How to Choose

What is cloud storage?

Cloud storage is a cloud computing model that enables storing data and files on the internet through a cloud computing provider that you access either through the public internet or a dedicated private network connection.

The provider securely **stores**, **manages**, and **maintains** the **storage** servers, infrastructure, and network to ensure clients has access to the data when they need it at virtually unlimited scale, and with elastic capacity.

Cloud storage removes the need to buy and manage your own data storage infrastructure, giving you agility, scalability, and durability, with any time, anywhere data access.

Why is cloud storage important?

Cloud storage delivers cost-effective, scalable storage eliminate the need to worry about running out of capacity, replacing failed devices, adding infrastructure to scale up with demand, or operating underutilized hardware when demand decreases.

Cloud storage is elastic, meaning you scale up and down with demand and pay only for what you use. It is a way for organizations to save data securely online so that it can be accessed anytime from any location by those with permission.

Why is cloud storage important?

Whether you are a small business or a large enterprise, cloud storage can deliver the agility, cost savings, security, and simplicity to focus on your core business growth.

For small businesses, you no longer have to worry about devoting valuable resources to manage storage yourself, and cloud storage gives you the ability to scale as the business grows.

For large enterprises with billions of files and petabytes of data, you can rely on the scalability, durability, and cost savings of cloud storage to create centralized data lakes to make your data accessible to all who need it.

Why is cloud storage important?

Cost effectiveness

With cloud storage, there is no hardware to purchase, no storage to provision, and no extra capital being used for business spikes. You can add or remove storage capacity on demand, quickly change performance and retention characteristics, and only pay for storage that you actually use. As data becomes infrequently and rarely accessed, you can even automatically move it to lower-cost storage, thus creating even more cost savings. By moving storage workloads from on premises to the cloud, you can reduce total cost of ownership by removing overprovisioning and the cost of maintaining storage infrastructure.

Why is cloud storage important?

Increased agility

With cloud storage, resources are only a click away. You reduce the time to make those resources available to your organization from weeks to just minutes. This results in a dramatic increase in agility for your organization. Your staff is largely freed from the tasks of procurement, installation, administration, and maintenance. And because cloud storage integrates with a wide range of analytics tools, your staff can now extract more insights from your data to fuel innovation.

Why is cloud storage important?

Faster deployment

When development teams are ready to begin, infrastructure should never slow them down. Cloud storage services allow IT to quickly deliver the exact amount of storage needed, whenever and wherever it's needed. Your developers can focus on solving complex application problems instead of having to manage storage systems.

Why is cloud storage important?

Efficient data management

By using cloud storage lifecycle management policies, you can perform powerful information management tasks including automated tiering or locking down data in support of compliance requirements. You can also use cloud storage to create multi-region or global storage for your distributed teams by using tools such as replication. You can organize and manage your data in ways that support specific use cases, create cost efficiencies, enforce security, and meet compliance requirements.

Why is cloud storage important?

Virtually unlimited scalability

Cloud storage delivers virtually unlimited storage capacity, allowing you to scale up as much and as quickly as you need. This removes the constraints of on-premises storage capacity. You can efficiently scale cloud storage up and down as required for analytics, data lakes, backups, or cloud native applications. Users can access storage from anywhere, at any time, without worrying about complex storage allocation processes, or waiting for new hardware.

Why is cloud storage important?

Business continuity

Cloud storage providers store your data in highly secure data centers, protecting your data and ensuring business continuity. Cloud storage services are designed to handle concurrent device failure by quickly detecting and repairing any lost redundancy. You can further protect your data by using versioning and replication tools to more easily recover from both unintended user actions or application failures.

Why is cloud storage important?

With cloud storage services, we can:

- Cost-effectively protect data in the cloud without sacrificing performance.
- Scale up your backup resources in minutes as data requirements change.
- Protect backups with a data center and network architecture built for security-sensitive organizations.

Storage Virtualization in Cloud Computing

Storage virtualization is a crucial aspect of cloud computing, allowing organizations to optimize their storage resources. It abstracts physical storage devices and presents them as a unified logical storage pool, enhancing flexibility, management, and efficiency.

What is Storage Virtualization?

Storage virtualization refers to the process of combining multiple physical storage devices into a single logical unit. This allows for simplified management and allocation of storage resources. Users can access these resources without needing to understand the underlying hardware specifics.

There are three main cloud storage types:

- **Block storage**
- **File storage**
- **Object storage**

Each offers its own advantages and has its own use cases.

Types of Storage Virtualization

Block Storage Virtualization:

- Definition: Involves managing raw storage blocks that can be formatted and used by operating systems.
- Use Cases: Ideal for databases and applications needing high I/O performance.
- Example: Virtual SANs (Storage Area Networks) that combine multiple disks into a single pool.

Types of cloud storage - Block storage

Enterprise applications like databases or enterprise resource planning (ERP) systems often require dedicated, low-latency storage for each host. This is analogous to direct-attached storage (DAS) or a storage area network (SAN). In this case, you can use a cloud storage service that stores data in the form of blocks. Each block has its own unique identifier for quick storage and retrieval.

Types of Storage Virtualization

File Storage Virtualization:

- Definition: Allows users to access files stored across multiple locations as if they were on a single server.
- Use Cases: Suitable for collaborative environments where multiple users need to access shared files.
- Example: Network-attached storage (NAS) solutions.

Types of cloud storage - File storage

File-based storage or file storage is widely used among applications and stores data in a hierarchical folder and file format. This type of storage is often known as a network-attached storage (NAS) server with common file level protocols of Server Message Block (SMB) used in Windows instances and Network File System (NFS) found in Linux.

Types of Storage Virtualization

Object Storage Virtualization:

- Definition: Manages unstructured data as objects, each with a unique identifier and metadata.
- Use Cases: Excellent for large-scale storage needs, such as media files, backups, and archival data.
- Example: Amazon S3, where data is accessed via APIs

Organizations have to store a massive and growing amount of unstructured data, such as photos, videos, machine learning (ML), sensor data, audio files, and other types of web content, and finding scalable, efficient, and affordable ways to store them can be a challenge. Object storage is a data storage architecture for large stores of unstructured data. Objects store data in the format it arrives in and makes it possible to customize metadata in ways that make the data easier to access and analyze. Instead of being organized in files or folder hierarchies, objects are kept in secure buckets that deliver virtually unlimited scalability. It is also less costly to store large data volumes.

Types of cloud storage - Object storage

Applications developed in the cloud often take advantage of the vast scalability and metadata characteristics of object storage. Object storage solutions are ideal for building modern applications from scratch that require scale and flexibility, and can also be used to import existing data stores for analytics, backup, or archive.

Types of Storage Virtualization – Key features

Abstraction: Simplifies storage management by hiding the complexities of the underlying hardware.

Resource Pooling: Aggregates storage resources from multiple devices, allowing for efficient resource allocation.

Automation: Supports automated provisioning and management through orchestration tools.

Data Tiering: Moves data between different storage types based on access frequency, optimizing performance and cost.

Types of Storage Virtualization – Benefits

Scalability: Easily scale storage resources up or down according to demand, without downtime.

Cost Efficiency: Reduces capital expenditure on hardware by maximizing existing resources and facilitating pay-as-you-go models.

Disaster Recovery: Offers robust backup solutions and disaster recovery strategies, ensuring data integrity and availability.

Flexibility: Enables dynamic allocation of storage resources, allowing organizations to adjust based on workload changes.

Types of Storage Virtualization – Challenges

Complexity: Requires knowledgeable personnel to implement and manage, as virtualization can complicate storage management.

Performance Overhead: Virtualization can introduce latency, impacting performance, particularly in high-throughput environments.

Vendor Lock-In: Organizations may become dependent on specific cloud providers, making migration to other platforms challenging.

Use Cases in Cloud Computing

Data Backup and Recovery: Virtualized storage solutions enable efficient backup strategies, ensuring data can be quickly restored in case of failure.

Development and Testing: Developers can quickly provision storage resources for testing applications without needing physical infrastructure.

Big Data Analytics: Supports the storage needs of big data applications, allowing for efficient processing and analysis of large datasets.

AI and Machine Learning Integration: Utilizing AI to optimize storage allocation and data management.

Increased Automation: More automated solutions that minimize manual intervention in managing storage resources.

Multi-Cloud Strategies: Enhanced tools for managing storage across multiple cloud environments, reducing vendor lock-in.

How does storage virtualization work?....

In storage virtualization, physical storage hardware is mirrored in a virtual volume. To construct a simple virtual storage environment, multiple physical disks are combined into a grouping that uses a single server. Virtual storage or logical storage blocks are assigned to the same server and help to redirect the input/output (I/O) traffic.

The physical disks are separated from the virtual volume by a virtualization layer that enables operating systems and applications to access and use the storage. Virtual storage software takes the I/O requests and sends them across the overall pool of storage to the appropriate storage devices.

How does storage virtualization work?....

The physical disks themselves are divided into small blocks of data, or objects known as logical unit numbers (LUN's), logical volume (LV), or RAID groups. These blocks are presented to remote servers as a virtual disk. They look just like a physical disk to the server, rather than the collection of storage devices that make up the overall pool of storage in the virtualized environment.

In a more complicated environment, RAID arrays can function as virtual storage. Here, multiple physical drives mimic a single storage device that stripes and replicates data to multiple disks in the background. This process improves I/O performance and protects the data from a failure in any single drive.

How does storage virtualization work?....

The way data is accessed from the physical drives requires an additional step for the virtualization software. Along with creating a barrier between the physical and virtual storage devices, the virtualization software creates a map using metadata that allows the stored data to be located quickly. In some cases, the software creates an algorithm to find the data even more quickly.

How does storage virtualization work?....

Creating virtual storage can be achieved on both block-level and file-level storage environments. Virtualizing a storage area network (SAN) involves adding a translation layer between the hosts and the storage arrays. In this type of storage virtualization, servers are redirected to virtualized LUNs instead of LUNs on the individual storage array. These virtualized LUNs remain on the virtualized device. Virtualizing a NAS involves removing the dependencies between the data accessed at the file level and the location where the files are physically stored.

What are (LV – LUN – RAID)!!

Logical Volumes (LV) are virtual representations of physical storage resources. They allow administrators to manage disk space more flexibly and efficiently.

A Logical Unit Number (LUN) is a unique identifier assigned to a logical unit of storage, typically within a SAN (Storage Area Network). LUNs enable servers to access storage devices over a network.

RAID (Redundant Array of Independent Disks) groups are configurations of multiple physical disks that work together to improve data redundancy, performance, or both. RAID combines multiple disks into a single unit.

What is a Storage Area Network (SAN) ?

It is used for transferring the data between the servers and the storage devices' fiber channels and switches. In SAN (Storage Area Network), data is identified by disk block. Protocols that are used in SAN are SCSI (Small Computer System Interface), SATA (Serial Advanced Technology Attachment).

What is a Storage Area Network (SAN) ?

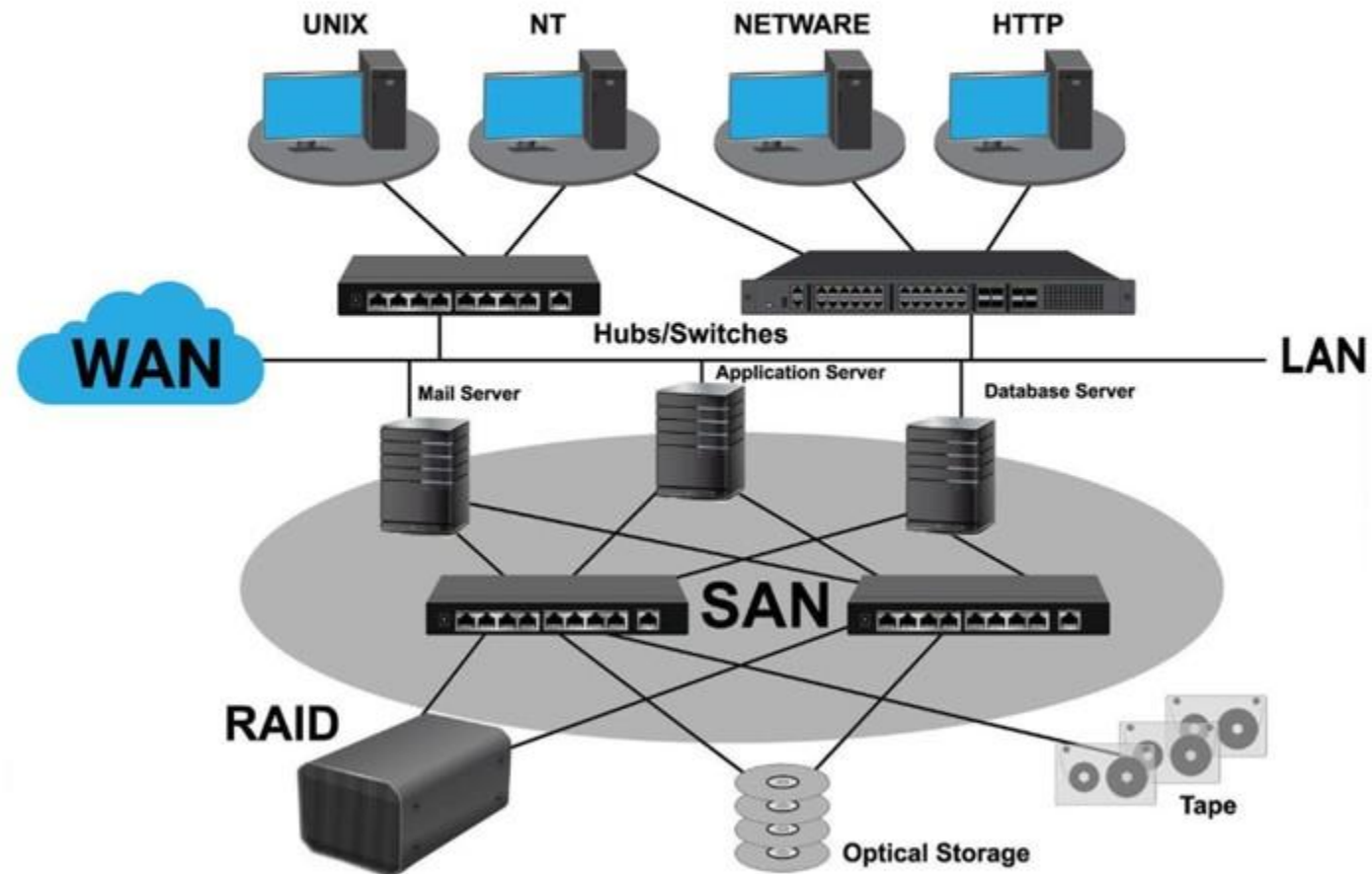
Components of Storage Area Network (SAN):

- Node ports
- Cables
- Interconnect devices such as Hubs, switches, directors
- Storage arrays
- SAN management software

How SAN work?

A Storage Area Network (SAN) is a dedicated, high-speed network that links servers to storage devices, enabling multiple servers to share and access storage resources efficiently. Key components include storage devices, switches, Host Bus Adapters (HBAs), and high-speed cabling. By operating separately from regular networks, SANs avoid traffic congestion and use protocols like Fibre Channel or iSCSI for fast data transfer. They offer improved performance, scalability, and centralized management of storage, making them suitable for demanding applications and extensive storage needs.

How SAN work?



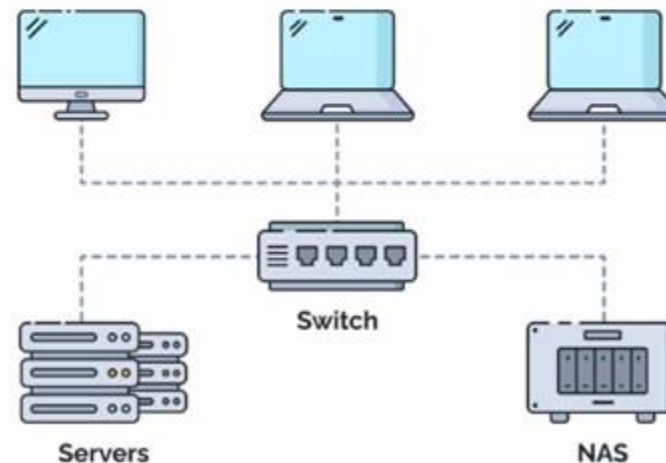
Disaster Recovery: In disaster recovery scenarios, a SAN provides a reliable solution for ensuring data availability and continuity in the event of a site failure or disaster. Effective disaster recovery strategies involve replicating data to offsite locations to safeguard against data loss and ensure business continuity.

SANs facilitate data replication between primary and secondary storage locations. This replication can be synchronous or asynchronous, depending on the organization's recovery point objectives (RPO) and recovery time objectives (RTO). Synchronous replication ensures that data is mirrored in real-time, while asynchronous replication updates the backup data at scheduled intervals.

SANs support disaster recovery by enabling storage to be located at geographically dispersed sites. This geographic separation protects against localized disasters, such as fires or floods, ensuring that data remains accessible even if one site is affected.

What is Network Attached Storage (NAS) ?

is a specialized file storage device that connects to a network, enabling multiple users and devices to access and share files from a centralized location. Unlike traditional file servers, NAS systems are designed primarily for data storage and sharing rather than general-purpose computing.



Components of Network Attached Storage (NAS):

- Head unit: CPU, Memory
- Network Interface Card (NIC)
- Optimized operating system
- Protocols
- Storage protocols: ATA (Advanced Technology Attachment), SCSI, FC (Fibre Channel)

Network-Attached Storage (NAS) is a file storage solution that connects to a network, enabling multiple devices to access and share files easily. The NAS device houses storage drives and connects to the network via an Ethernet port. It operates with its own system designed to handle file management and network connections, supporting various file-sharing protocols like SMB, NFS, or AFP. Users can access files from the NAS as if it were a local drive, with features such as data redundancy through RAID, automatic backups, and access controls to manage permissions. The NAS can be managed through a web-based interface or mobile apps, providing a centralized, efficient storage solution for collaborative work and data sharing.

File Sharing and Collaboration:

NAS is perfect for settings where multiple people need to access and work on the same files. It provides a single place where everyone can store and retrieve documents, which makes team projects much smoother for example Small offices, home workspaces or teams collaborating on shared documents or media.

Data Backup and Recovery: NAS devices often include backup features that help you keep your important files safe. You can use them to back up data from different computers or devices to one secure spot example backing up personal files, protecting small business data or planning for disaster recovery.

Media Storage and Streaming with NAS you can store and stream your music, photos and videos across your network. It turns your NAS into a home media hub letting you enjoy your digital library on various devices like setting up a home media server, organizing your digital photo collection, or streaming videos and music to your TV and other devices.

Difference Between Storage Area Network (SAN) and Network Attached Storage (NAS):

SAN	NAS
SAN stands for Storage Area Network.	NAS stands for Network Attached Storage.
In SAN (Storage Area Network), data is identified by disk block.	In NAS (Network Attached Storage), data is identified by file name as well as byte offset.
In SAN (Storage Area Network), the file system is managed by servers.	In NAS (Network Attached Storage), file system is managed by Head unit.
SAN (Storage Area Network) is more costly.	NAS (Network Attached Storage) is less expensive than SAN.

Difference Between Storage Area Network (SAN) and Network Attached Storage (NAS):

SAN	NAS
SAN(Storage Area Network) is more complex than NAS.	NAS (Network Attached Storage) is less complex than SAN.
Protocols used in SAN are SCSI, <u>SATA</u> , etc.	Protocols used in NAS are File server, CIFS (Common Internet File System), etc.
For backups and recovery in SAN, Block by block copying technique is used.	For backups and recovery in NAS, Files are used.
SAN gives high performance in high-speedefforthigh-speed traffic systems.	While NAS is not suitable for that environment which has high speed traffic.

Difference Between Storage Area Network (SAN) and Network Attached Storage (NAS):

SAN	NAS
SAN needs more time and efforts in organizing and controlling.	NAS is easy to manage and provides a simple interface for organizing and controlling.
SAN does not depend on the LAN and uses a high-speed fiber channel network.	NAS needs TCP/IP networks and depends on the LAN.
Mostly used in enterprise environments.	Applications include small-sized organizations high-speed and homes.
It has lower latency.	Compared to SAN, NAS has higher latency.
SAN supports virtualization.	NAS does not support virtualization.

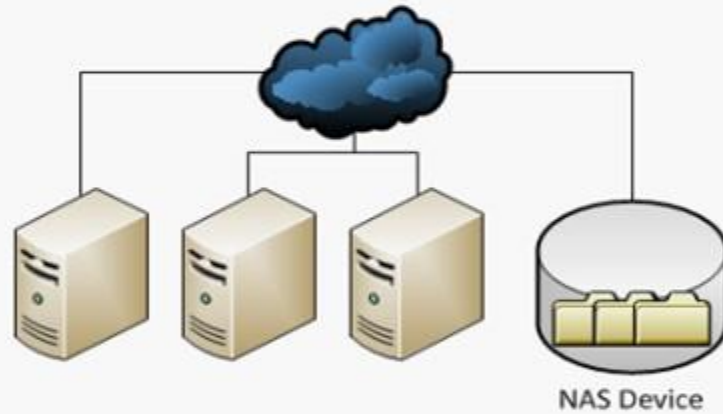
Difference Between Storage Area Network (SAN) and Network Attached Storage (NAS):

Both SAN and NAS each offer distinct advantages depending on your storage needs. SAN is a robust highspeed network that's perfect for large businesses needing centralized storage and the capacity to handle big data loads efficiently. NAS on the other hand is a simpler and more budget friendly option ideal for smaller businesses or home use where easy access and file sharing are the priorities. By understanding what each option brings to the table you can make the best choice for your specific situation ensuring your data is stored and managed smoothly.

Difference Between Storages

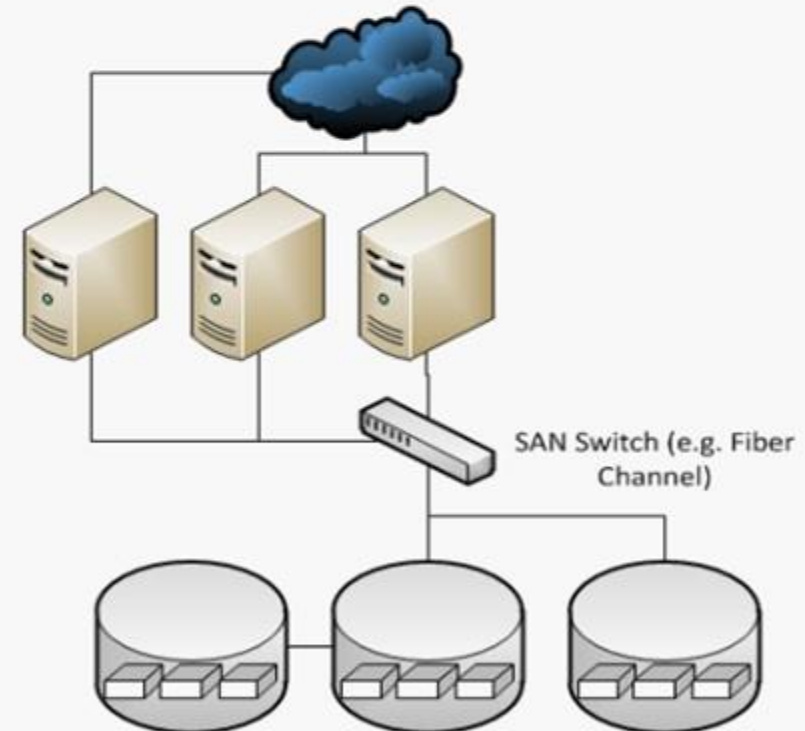
Network Attached Storage

- Shared storage over shared network
- File system
- Easier management



Storage Area Network

- Shared storage over dedicated network
- Raw storage
- Fast, but costly



Direct Attached Storage (DAS) the common easiest method to implement. A computer, laptop, or other computing device that has its own storage directly connected is considered to be direct-attached storage.

These devices can be hard drives, flash drives, or any other type of storage that is connected to the computer and not over a network.

The connection used is commonly an ATA, SATA, or SCSI interface connecting the storage media to the motherboard of the computer.

Network-attached storage (NAS) is a file-level access of the data across a network. In a NAS configuration, files are sent over the network rather than blocks of data as in storage area networks (SANs).

The data is not stored on a local computer, as with direct-attached storage, but over a LAN as network-attached storage.

Files are sent over the LAN using a protocol such as TCP/IP. The data structures are dependent on the operating system used.

Storage Area Network (SAN) is a very high speed, highly redundant, and completely dedicated to storage traffic.

When a server accesses storage over a SAN network, the SAN must be completely lossless and highly available. The most common dedicated store networking technology is Fiber Channel.

Typically, two SAN networks are configured for redundancy, and the two networks operate in parallel to each other and are not interconnected. This architecture offers two paths from the server to the storage arrays, with a primary network and a standby ready to take over should the primary failure.

Choosing SAN or NAS depends on your organization's requirements and capabilities.

SAN: is ideal for high-performance, mission-critical applications in large enterprises with the resources to support it.

NAS: is better for cost-effective, easy-to-manage file storage in smaller organizations, departments, or environments primarily focused on file sharing and collaboration.

Storage virtualization in cloud computing plays a vital role in modern data management strategies. By allowing organizations to abstract and manage their storage resources more efficiently, it provides significant benefits in terms of scalability, flexibility, and cost-effectiveness. As technology evolves, the capabilities and applications of storage virtualization are expected to expand further, making it an essential component of cloud infrastructure.



Thank You