Storage type

Local Disk Storage:

kem chho palty

1.
   1. Each compute node in an HPC cluster often has local storage directly attached to it.
   2. **Characteristics:** Fast access times, suitable for temporary or scratch space.
   3. **Usage:** Often used for storing application binaries, temporary files, or intermediate results during computations.
2.

Network File Systems (NFS):

3.
   1. NFS is a common choice for providing shared storage across multiple compute nodes in an HPC cluster.
   2. **Characteristics:** Provides centralized storage accessible from all compute nodes over the network.
   3. **Usage:** Suitable for storing shared data, user home directories, and software installations that need to be accessible cluster-wide.
4.

Parallel File Systems (e.g., Lustre, GPFS):

5.
   1. These are high-performance, distributed file systems designed to handle large-scale data storage and access in HPC environments.
   2. **Characteristics:** High throughput and scalability, optimized for parallel access across many nodes simultaneously.
   3. **Usage:** Ideal for storing large datasets, scientific simulations, and applications that require high data throughput and low latency.
6.

Object Storage (e.g., Amazon S3, Swift):

7.
   1. Object storage provides scalable and durable storage solutions typically used for storing large amounts of unstructured data.
   2. **Characteristics:** Scalable, durable, and often used for storing backups, archival data, and data accessible via APIs.
   3. **Usage:** Useful for storing large datasets, backups, and data that needs to be accessed programmatically.

Considerations:

-

Performance Requirements: Different storage types offer varying levels of performance, scalability, and accessibility. Choose the type of storage based on the performance requirements of applications and the volume of data being handled.

kem chho palty

-
-

Data Management: Ensure compatibility and integration between storage solutions and the workload management system (e.g., Slurm) to optimize data access and management across the cluster.

-
-

Data Protection: Implement appropriate data protection mechanisms (e.g., RAID configurations, backups) depending on the criticality of data and the risk of data loss.

In the context of Storage and Backup Management (SBM), Here are the key components of a disk drive:

1.

Platters:

2.
1. Platters are the circular disks inside the hard drive where data is stored. These are typically made of glass or aluminum and coated with a magnetic material that allows data to be written and read using magnetic heads.
3.

Spindle:

4.

1. The spindle is the central axis around which the platters rotate. It connects the platters to the motor, which spins them at high speeds (usually measured in revolutions per minute, RPM).

5.

Read/Write Heads:

6.

1. Read/write heads are electromagnetic devices located on an actuator arm. They read data from and write data to the platters by changing the magnetic polarity of the disk's surface.

7.

Actuator Arm:

kem chho palty

8.

1. The actuator arm is responsible for positioning the read/write heads over the correct location on the platters to access or store data. It moves rapidly across the platter surfaces under the control of an actuator motor.

9.

Actuator Motor:

10.

1. The actuator motor controls the movement of the actuator arm and read/write heads. It ensures precise positioning of the heads to access specific data tracks on the platters.

11.

Controller Board (PCB):

12.

1. The controller board, or PCB (Printed Circuit Board), manages the overall operation of the disk drive. It includes components like the disk controller, cache memory, and interface connectors (e.g., SATA, SAS) for communication with the computer system.

13.

Cache Memory:

14.

1. Cache memory is a small amount of high-speed volatile memory (DRAM or SRAM) located on the disk drive's PCB. It stores frequently accessed data and improves read/write performance by reducing latency.

15.

Firmware:

16.

1. Firmware is embedded software stored on the disk drive’s PCB. It controls the drive’s operations, manages data access, error handling, and interfaces with the computer system’s operating system.

17.

Interface Connector:

kem chho palty

18.

1. The interface connector provides the physical connection between the disk drive and the computer system’s storage controller. Common interfaces include SATA (Serial ATA) and SAS (Serial Attached SCSI).

////////////////////////////////

FC PROTOCOL STACK

The Fibre Channel (FC) protocol stack is a layered architecture that defines how data is transmitted and managed across Fibre Channel networks. It consists of several layers, each responsible for different aspects of communication and data handling. Here's an overview of the FC protocol stack layers:

1. FC-4 (Upper Layer Protocols):

- **Purpose:** The FC-4 layer defines the upper layer protocols that determine how higher-level data and commands are encapsulated and transmitted over the Fibre Channel network.
- **Examples:** Various upper layer protocols can operate at this level, including SCSI (SCSI FC Protocol, FCP), IP (FCIP), and others.
- **Functionality:** Each upper layer protocol adapts its specific application data into FC frames for transmission over the network. For example, FCP encapsulates SCSI commands and data for storage devices.

2. FC-3 (Common Services):

- **Purpose:** The FC-3 layer provides common services that are independent of the specific upper layer protocol in use.
- **Functionality:** It includes functions such as encryption (FC-SP - Fibre Channel Security Protocols), class of service (CoS) management, and virtual fabric configuration.
- **Encryption (FC-SP):** Ensures data security by providing authentication and encryption mechanisms to protect data transmitted over the FC network.

3. FC-2 (Network Services):

- **Purpose:** The FC-2 layer manages the framing and flow control mechanisms necessary for reliable data transmission over Fibre Channel.
- **Functionality:** Includes functions like framing, flow control, and error detection/correction.

- **Framing:** Defines how data from the upper layers is encapsulated into Fibre Channel frames for transmission.
- **Flow Control:** Manages the rate of data transfer to prevent data loss and ensure reliable delivery.
- **Error Detection/Correction:** Includes mechanisms such as CRC (Cyclic Redundancy Check) to detect and optionally correct errors in transmitted data.

4. FC-1 (Transmission Protocol):

kem chho palty

- **Purpose:** The FC-1 layer specifies the physical and electrical characteristics of the Fibre Channel interface.
- **Functionality:** Includes encoding, signaling, and transmission of Fibre Channel frames over physical media (fiber optic cables, copper cables).
- **Encoding:** Converts digital data into a format suitable for transmission over the physical medium.
- **Signaling:** Defines the electrical signaling method used for transmitting data across the physical medium (e.g., NRZ - Non-Return to Zero).

5. FC-0 (Physical Interface):

- **Purpose:** The FC-0 layer defines the physical characteristics and specifications of the Fibre Channel interface.
- **Functionality:** Includes connector types, fiber types (e.g., single-mode, multi-mode), signaling rates (e.g., 1Gbps, 2Gbps, 4Gbps, 8Gbps, 16Gbps, 32Gbps), and cable distances.
- **Connector Types:** Specifies the types of connectors used to connect Fibre Channel devices, such as LC, SC, or MPO/MTP connectors.
- **Fiber Types:** Defines the types of optical fiber that can be used, influencing the maximum distance and data rates supported by the FC link.

Summary:

The Fibre Channel protocol stack is a comprehensive architecture that provides a layered approach to ensuring reliable, high-performance data transmission over Fibre Channel networks. Each layer contributes specific functionalities and services to manage different aspects of data communication, from physical transmission characteristics (FC-0) to upper-layer protocol encapsulation (FC-4). Understanding the FC protocol stack helps in designing, deploying, and troubleshooting Fibre Channel-based storage networks effectively.

Storage Replication

Definition: Storage replication involves creating and maintaining duplicate copies of data on separate storage systems or locations. The primary purpose of storage replication is to ensure data availability, disaster recovery, and high availability of critical applications and data.

Key Aspects:

- **Data Redundancy**: Replication creates redundant copies of data across multiple storage devices or locations, ensuring that if one copy becomes inaccessible or corrupted, another copy can be used.
- **Synchronous vs. Asynchronous**: Replication can be synchronous (immediate replication) or asynchronous (delayed replication), depending on the required recovery point objectives (RPO) and performance considerations.
- **Disaster Recovery**: It plays a crucial role in disaster recovery strategies by providing failover capabilities. In case of a primary storage failure or disaster, the replicated data can be accessed from the secondary location.
- **Use Cases**: Commonly used in enterprise environments for critical applications, databases, and virtualized environments where continuous data availability and minimal downtime are essential.

kem chho palty

Hierarchical Storage Management (HSM)

Definition: HSM is a data storage technique that automatically moves data between different storage tiers based on predefined policies and access patterns. The primary objective of HSM is to optimize storage resources by migrating less frequently accessed data to lower-cost storage tiers while keeping frequently accessed data on faster and more expensive storage media.

Key Aspects:

- **Storage Tiering**: HSM typically involves multiple storage tiers, such as high-performance disk storage, lower-performance disk arrays, and even tape or cloud storage.
- **Policy-driven**: Data movement between storage tiers is governed by policies that consider factors such as access frequency, file size, age of data, and business importance.
- **Efficiency**: HSM helps optimize storage costs by ensuring that expensive, high-performance storage resources are used efficiently for active data while less critical or rarely accessed data is stored on cheaper storage.
- **Archival and Compliance**: It facilitates data archiving and compliance with retention policies by automatically moving data to appropriate storage tiers as per regulatory requirements.
- **Use Cases**: Often implemented in environments with large volumes of data, such as scientific research, healthcare, media, and financial services industries, where managing data growth and storage costs is a concern.

Relationship Between Storage Replication and HSM

While both storage replication and HSM involve managing data across different storage environments, they serve complementary but distinct purposes:

- **Data Protection vs. Optimization**: Storage replication focuses on data protection and availability by creating redundant copies, whereas HSM focuses on optimizing storage resources and cost-efficiency by tiering data based on usage patterns.
- **Disaster Recovery vs. Storage Efficiency**: Replication ensures data availability and disaster recovery readiness, while HSM optimizes storage resources and improves overall efficiency by moving data to the most appropriate storage tier.

In summary, storage replication ensures data availability and disaster recovery readiness through redundancy, while HSM optimizes storage resources and cost-efficiency by automatically moving data between different storage tiers based on predefined policies. Both are critical components of modern data storage and management strategies, each addressing distinct aspects of data protection, availability, and efficiency.

kem chho palty

$$

Backup media (LTO)
Explained in video - https://www.youtube.com/watch?v=07fc9SrJfWI

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$