8SNS COLLEGE OF TECHNOLOGY
DEPARTMENT OF AIML
23CST202- OPERATING SYSTEMS
II YEAR IV SEM AIML-B
UNIT 5 – I/O Systems
TOPIC – Disk Attachment-Stable Storage-Tertiary Storage

The disk attachment is a special feature that allows you to attach a disk (such as a USB flash drive) to your computer's hard drive. Users can use this feature to store files in the cloud or transfer files between computers on their network.

Host-Attached Storage

Host-attached storage (HAS) is a form of internal computer storage that can be attached to a host computer, such as a PC or server. Host-attached devices are often used for backup purposes and can include tape drives, optical drives, hard disk drives (HDDs), solid state drives (SSDs), USB flash drives, and other similar media.

A common example of host-attached storage is the use of an external USB flash drive for data transfer between computers. This type of connection allows users to copy files from one device onto another without having to connect them directly through their operating system's file-sharing functionality - this means that you don't have access rights over your own system if someone else has access!

Host-attached storage is usually very fast, which makes it a popular choice for high-performance applications. It's also relatively cheap compared to network-attached storage (NAS), which is another form of internal computer storage that can be accessed by multiple systems at once.

Host-attached storage systems are often used in data centers because of their high performance and reliability. These devices are usually more expensive than NAS devices, but they offer many benefits over other types of internal computer storage.

Network-Attached Storage (NAS)

A Network-Attached Storage (NAS) is a computer that is connected to a network and shared by multiple users. NAS can also be called a file server, or storage appliance. The term "storage" refers to any device that stores data; this includes hard drives and flashes memory modules.

NAS has its own processor and memory so it can perform many tasks simultaneously without slowing down other devices on the network; this makes it an ideal solution for businesses who need high availability but don't have enough CPU power available at their desktops or laptops. NAS systems are often used as backup solutions for PCs because they allow users to access files from anywhere in the world through remote access services provided by vendors such as Symantec Remote Access Server (RAS).

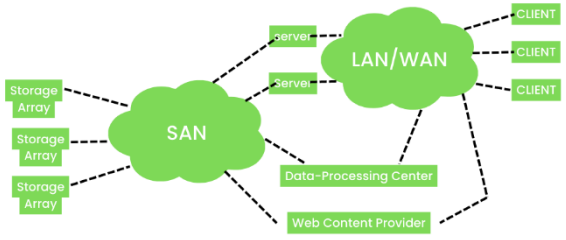
The NAS device is typically installed in the business' server room and shared across the network. It can be configured for both file sharing and print serving, depending on what type of business you run. A NAS system will allow multiple users to access files simultaneously without slowing down network performance; this makes it an ideal solution for businesses that need high availability but don't have enough CPU power available at their desktops or laptops.

Storage Area Network (SAN)

Storage area networks (SAN) are a network of storage devices that can be used to store and retrieve data from a shared central repository. A SAN is usually used for large data storage and retrieval, which requires high availability, scalability, reliability, and performance. The most common type of SAN uses fiber channel adapters to connect the host with its disk arrays.

A block-based storage protocol like Fibre Channel Protocol (FCP) allows multiple devices within the same fabric to communicate with each other in order to share resources such as disks or tape drives. In contrast to other protocols such as iSCSI or Network File System (NFS), FCP does not require any special software drivers on either end because all communication takes place using standard protocols like TCP/IP. Another common type of SAN uses a network called InfiniBand (IB), which is a high-speed serial interconnect that provides better performance than traditional Ethernet networks. When used in conjunction with FCP, IB allows data to be transferred faster and more efficiently between two computers.

A SAN can be implemented in two ways: as an external or internal device. An external SAN is used to connect storage devices to a network, while an internal SAN is integrated into the operating system of a computer. A disk is basically a device that communicates with the host to fetch data or store data in it with the help of I/O devices.



A disk is a physical device that stores data permanently. The disk is a hardware component that is used to store data permanently.

Hierarchical organization of data on disk:

- Sectors are divided into tracks and blocks within each track, each block contains sectors that are aligned on consecutive addresses;
- Tracks are divided into cylinders and heads; cylinder contains heads; heads contain sectors; sectors in every track start at address 0 (zero).

The physical organization of data on disks can be described as follows:
-A disk has multiple concentric tracks. Each track is divided into many sectors. A sector is the smallest unit of storage on a disk – it can hold one or more bytes of data (or nothing at all). Each sector is identified by its starting and ending address on the disk. The range of addresses used by a particular drive is called a cylinder head sector (CHS) addressing scheme.
-In order to find the starting address of a sector, you need to know two things: the cylinder number and head number (or track position). For example, if a disk has 100 cylinders and 40 heads, then each head can read or write data on up to 8 tracks at once.

Stable-Storage Implementation:
To achieve such storage, we need to replicate the required information on multiple storage devices with independent failure modes. The writing of an update should be coordinate in such a way that it would not delete all the copies of the state and when we are recovering from a failure we can force all the copies to a consistent and correct valued even if another failure occurs during the recovery.
The disk write operation results to one of the following outcome:

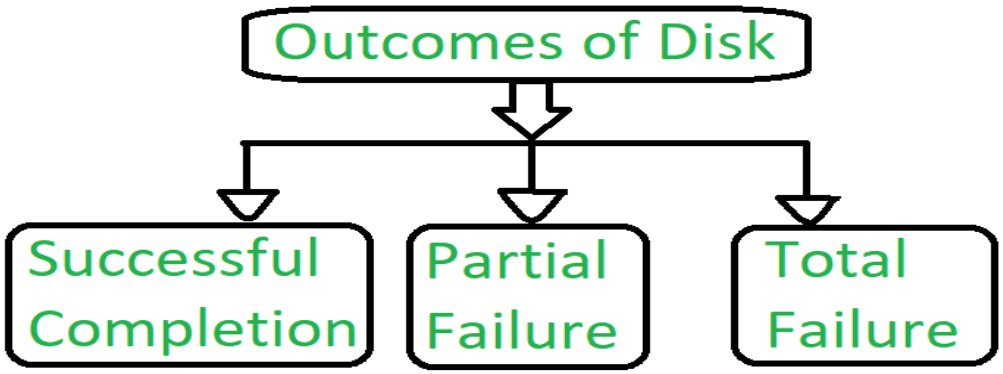


Figure - Outcomes of Disk

1. Successful completion -
   The data will be written correctly on the disk.
2. Partial Failure -
   In this case, failure has occurred in the middle of the data transfer, such that only some sectors were written with the new data, and the sectors which were written during the failure may have been corrupted.
3. Total Failure -
   The failure occurred before the disk write started, so the previous data values on the disk remains intact.

During writing a block somehow if failure occurs, the system's first work is to detect the failure and then invoke a recovery process to restore the consistent state. To do that, the system must contain two physical block for each logical block.
An output operation is executed as follows:

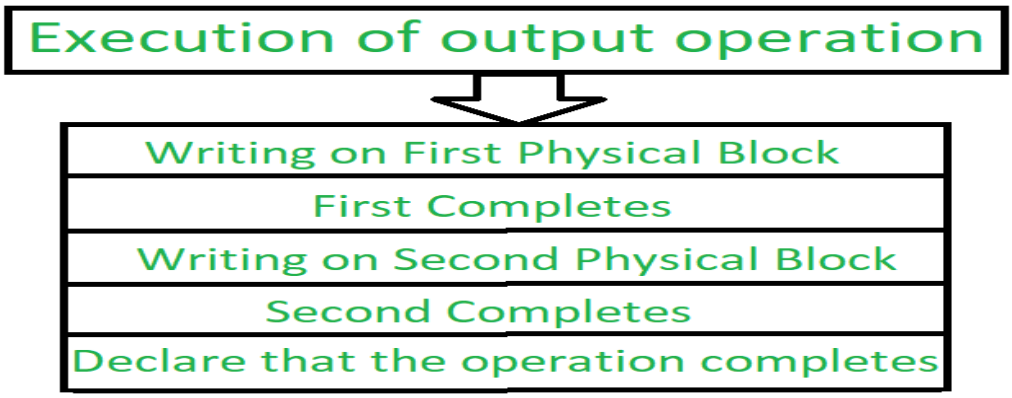


Figure - Process of execution of output operation

1. Write the information onto the first physical block.
2. When the first write completes successfully, perform the same operation onto the second physical block.
3. When both the operations are successful, declare the operation as complete.

During the recovery from a failure each of the physical block is examined. If both are the same and no detectable error exists, then no further action is necessary. If one block contains detectable errors then we replace its content with the value of the other block. If neither block contains the detectable error, but the block differ in content, then we replace the content of first block with the content of the second block. This procedure of the recovery give us an conclusion that either the write to stable content succeeds successfully or it results in no change.

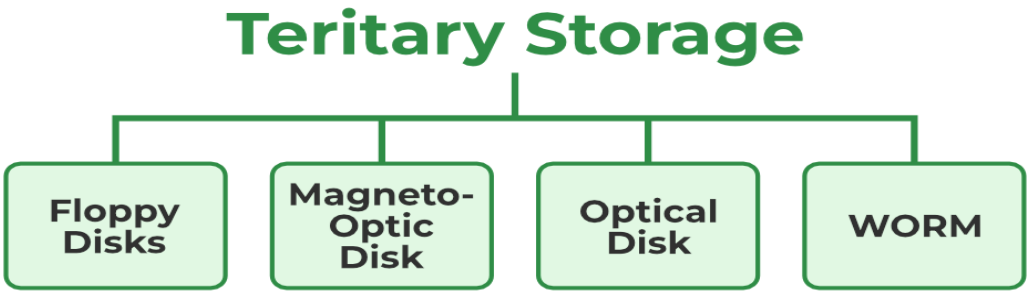
This procedure will be extended if we want arbitrarily large number of copies of each block of the stable storage. With the usage of large number of copies, the chances of the failure reduces. Generally, it is reasonable to simulate stable storage with only two copies. The data present in the stable storage is safe unless a failure destroys all the copies. The data that is present in the stable storage is guaranteed to be safe unless a failure destroys all the copies.

Because waiting for disk writes to complete is time consuming, many storage arrays add NVRAM as a cache. Since the memory is non-volatile it can be trusted to store the data in route to the disks. In this way it is considered as a part of the stable storage. Writing to the stable storage is much faster than to the disk, so performance is greatly improved.

Data that is not commonly accessed and typically not required for daily use is stored in tertiary storage. Tertiary storage is often slower and less expensive than primary and secondary storage is frequently used for data archiving and long-term storage.

Tertiary storage units are widely employed for offsite storage or for the long-term retention of volumes of data that are rarely accessed. Tape libraries, optical jukeboxes, and cloud storage are a few examples of tertiary storage systems. Data is kept on magnetic tapes, which are affordable, and long-lasting, but slower to access than other forms of storage, in tape libraries. In general, optical jukeboxes are faster than tape libraries but have a shorter lifespan since they store data on optical discs like CDs or DVDs.

Data storage on remote servers that are maintained by a third party and are accessible online is referred to as "cloud storage." Because it is made for rarely accessed data and does not need to be as quick or dependable as main and secondary storage, tertiary storage is typically less expensive and slower than primary and secondary storage. Data archiving and long-term retention, as well as data backup and recovery, are frequently done using tertiary storage.



Features

There are several key features of tertiary storage:

- Low cost: Because tertiary storage is intended for rarely accessed data and does not have to be as quick or dependable, it is typically less expensive than primary and secondary storage.
- Large storage capacity: Tertiary storage devices are made to hold a lot of data, usually between terabytes and petabytes.
- Offsite storage: Tertiary storage systems are frequently used for offsite storage, which can add security and safeguard against data loss due to disasters or other unforeseen circumstances.
- Slow access: Tertiary storage is not designed for frequent use, hence it often accesses more slowly than main and secondary storage.
- Storage for the long term: Tertiary storage is frequently used to store data for the long term that is not in use but must be kept for regulatory or compliance reasons, or for data archiving.
- Data backup and recovery: Tertiary storage is frequently used for data backup and recovery because it offers an affordable and dependable way to store data that might be required in the event of data loss or corruption.
- Large storage capacity: Tertiary storage offers significantly larger storage capacity compared to primary and secondary storage, making it ideal for storing large amounts of data that may not fit in primary or secondary storage.
- Cost-effective: Tertiary storage is typically more cost-effective than primary and secondary storage, as it is designed for large-scale data storage and is available in high-capacity devices.
- Easy accessibility: With tertiary storage, data can be easily accessed and retrieved as needed, even if it is not currently being used. This is because tertiary storage operates at a slower speed than primary and secondary storage.

- Improved data backup and recovery: Tertiary storage provides a convenient backup solution for critical data and enables easy data recovery in case of a failure or data loss in primary or secondary storage.
- Long-term data preservation: Tertiary storage is designed for long-term data preservation, making it ideal for archiving data that is not frequently used but must be kept for regulatory or historical purposes.
- Scalability: Tertiary storage can be easily scaled up or down to meet changing storage requirements, making it a flexible and adaptable solution for organizations of any size.

Applications

- Backup and Recovery: Tertiary storage is commonly used to store backups of critical data to protect against data loss due to hardware failure or other forms of data corruption.
- Archiving: Tertiary storage can be used to store large amounts of historical data that is not frequently accessed but still needs to be preserved for regulatory, legal, or business reasons.
- Digital Preservation: Tertiary storage is used to store and preserve valuable digital content such as historical documents, audio and video recordings, and photographs.
- Big Data Analytics: Tertiary storage systems can store large amounts of raw data that can be processed and analyzed for insights and decision-making.
- Cloud Storage: Tertiary storage is a component of cloud storage solutions, where data is stored remotely and accessed over the internet.
- Data Warehouses: Tertiary storage is used to store large amounts of structured data for business intelligence and data analysis.
- Data Lakes: Tertiary storage is used to store raw and unstructured data for later processing and analysis.

Limitations

- Data saved on tertiary storage is not always accessible because retrieving data from tertiary storage takes longer than from primary or secondary storage.
- Tertiary storage is not designed for regular use, hence it often takes longer to access than main and secondary storage.
- Data kept on tertiary storage may be hard to access because it may be stored offsite and require specialist equipment, which can make it harder to recover data quickly.
- Since it often necessitates the use of off-site storage facilities and specialist technology, retrieving data from tertiary storage can be costly.
- Data loss due to physical deterioration or other problems may occur in tertiary storage devices like tape libraries because of their limited lifespan.
- Because tertiary storage is not designed for active users and may not have the same level of protection against data loss or corruption as primary and secondary storage, it may not offer the same level of data security as those two storage types.