

Storage Technologies

IT403

Prepared by
Dr. Ahmed El-Douh

Lecture 3



Storage Technologies

Lecture 3

IT403

Architecture of Storage Systems

Prepared by
Dr. Ahmed El-Douh

Objectives

Upon completion of this module, you will be able to:

- Introduces the common system architecture of storage systems, the components of storage systems
- Understand the common storage system architectures.
- Grasp the concepts of common storage components.

Outlines

- Architecture of Storage Systems
- Supported Hard Disk Types
- Mechanical Hard Disk Drive (HDD)
- HDD Port Technology
- Solid State Drives (SSD)

Chapter
03

Architecture of Storage Systems



Common Forms of Storage Array Components

Controller and Disk Chassis Separation



Independent
Controller Chassis

+



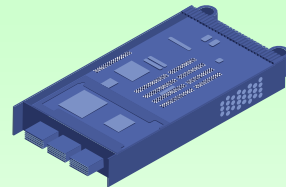
External Disk
Enclosures for
Expansion

=



Storage Array System

Controller and Disk Chassis Integration



Controller Module

+



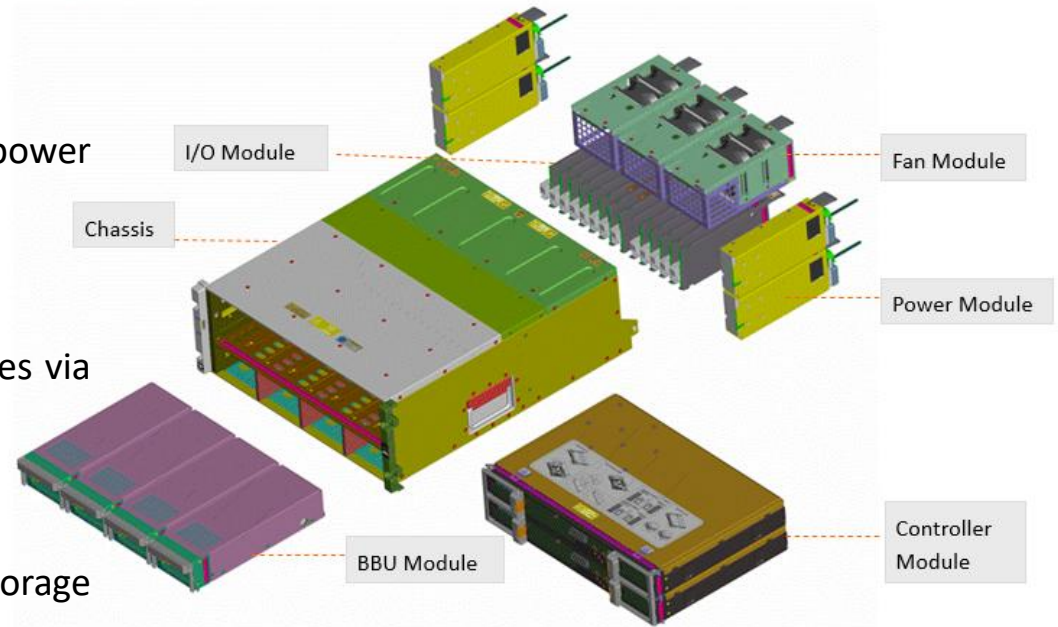
Integrated Controller
and Disk Enclosure

=



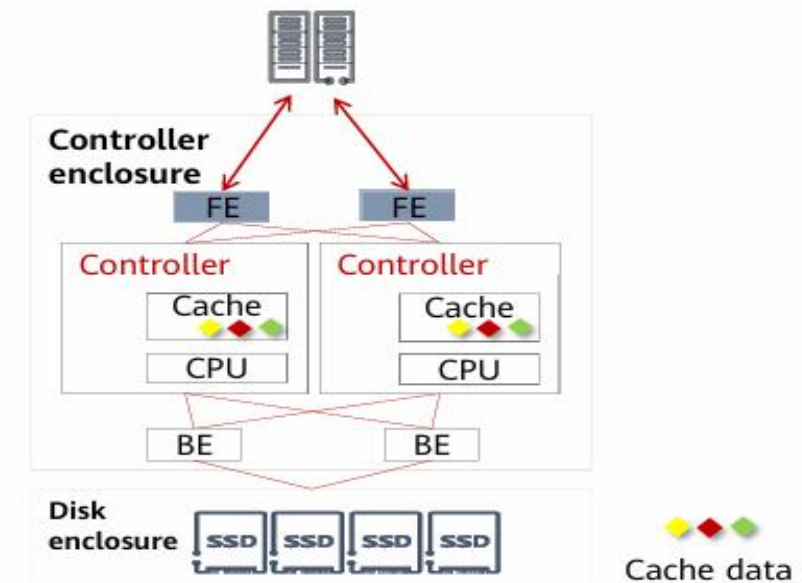
Chassis with Controller and Disk Separation

- The diagram above shows the different components that makes up the controller and disk separated chassis.
- The **main components** consist of the chassis, IO module, fan and power module and the controller and BBU (Backup Battery Unit) module.
- This **chassis** is connected to the external high density disk enclosures via the I/O module.
- **BBU** unit is the backup battery that provides power to the storage controller to save all the I/O cache into the coffer disks in the event of a power failure.



Controller

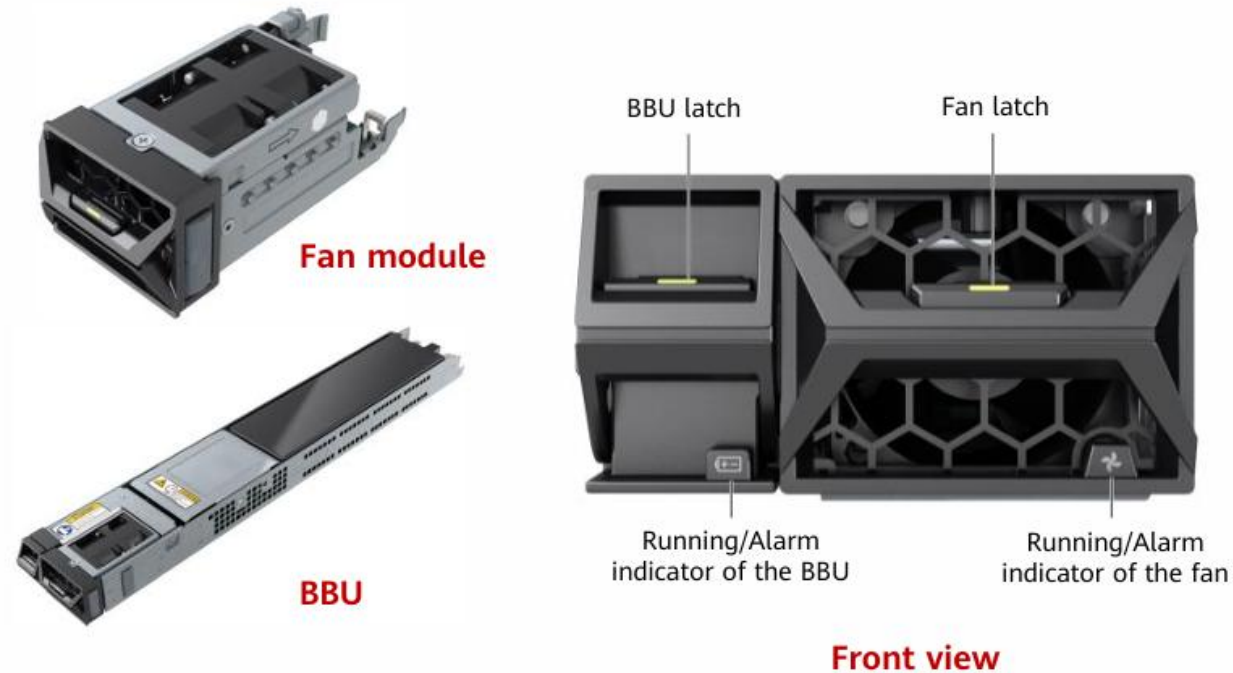
- A controller is the core component of a storage system. It processes storage services, receives configuration management commands, saves configuration data, connects to disks, and saves critical data to coffer disks.



Controller

- **The controller** is a core module that processes services in a storage system.
 - The CPU and cache on the controller work together to process I/O requests from the host and manage RAID of the storage system.
 - Each controller has multiple built-in disks to store system data. If a power failure occurs, these disks also store cache data. Disks on different controllers are redundant of each other.
- **Front-end (FE)** ports are used for service communication between application servers and the storage system, that is, for processing host I/Os.
- **Back-end (BE)** ports connect a controller enclosure to a disk enclosure and provide channels for reading/writing data from/to disks.
- A **cache** is a memory chip on a disk controller. It provides fast data access speed and functions as a buffer between the internal storage and external interfaces.
- An **engine** is a core component of a development program or system on an electronic platform. It is usually the support part of a program or a set of systems.

BBU and Fan Module



- **Fan** modules dissipate heat from the system, allowing the controller enclosure to operate normally at maximum power.
- **BBUs** supply power to the storage system in the event of an external power supply failure to protect data in the storage system.

BBU (Backup Battery Unit)

- When the external power supplies are normal, the BBUs are standing by. Faulty BBUs can be isolated without affecting normal running of the storage system.
- The BBUs allow the system to write cache data to the built-in disks of the controllers upon a power failure to prevent data loss.
- When the external power supply recovers, storage software recovers data from the built-in disks of the controllers into the cache.
- Battery capacity is regularly tested through charging and discharging to detect issues early and ensure reliable data backup during power failures.

Coffer Disk

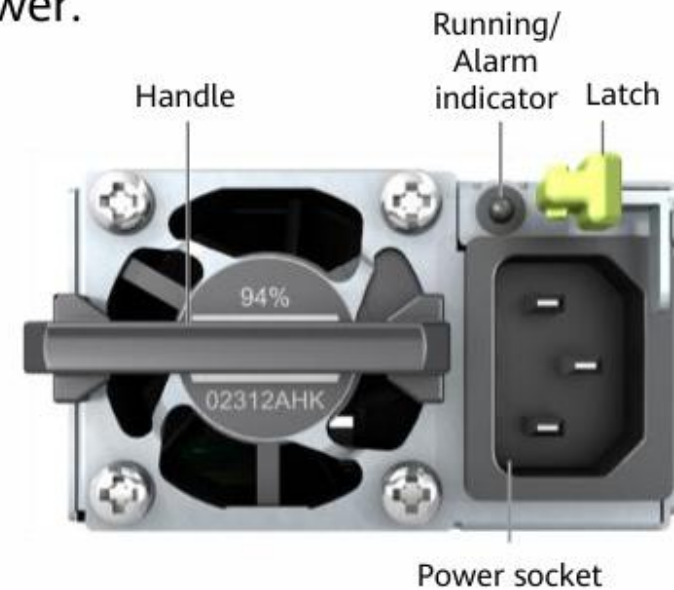


Note: Huawei OceanStor Dorado V6 is used as the example.

- **Coffer disks** are used to store user data, system configurations, logs, and dirty data in the cache in the event of an unexpected power outage.
- There are **built-in coffer disks** and **external coffer disks**.

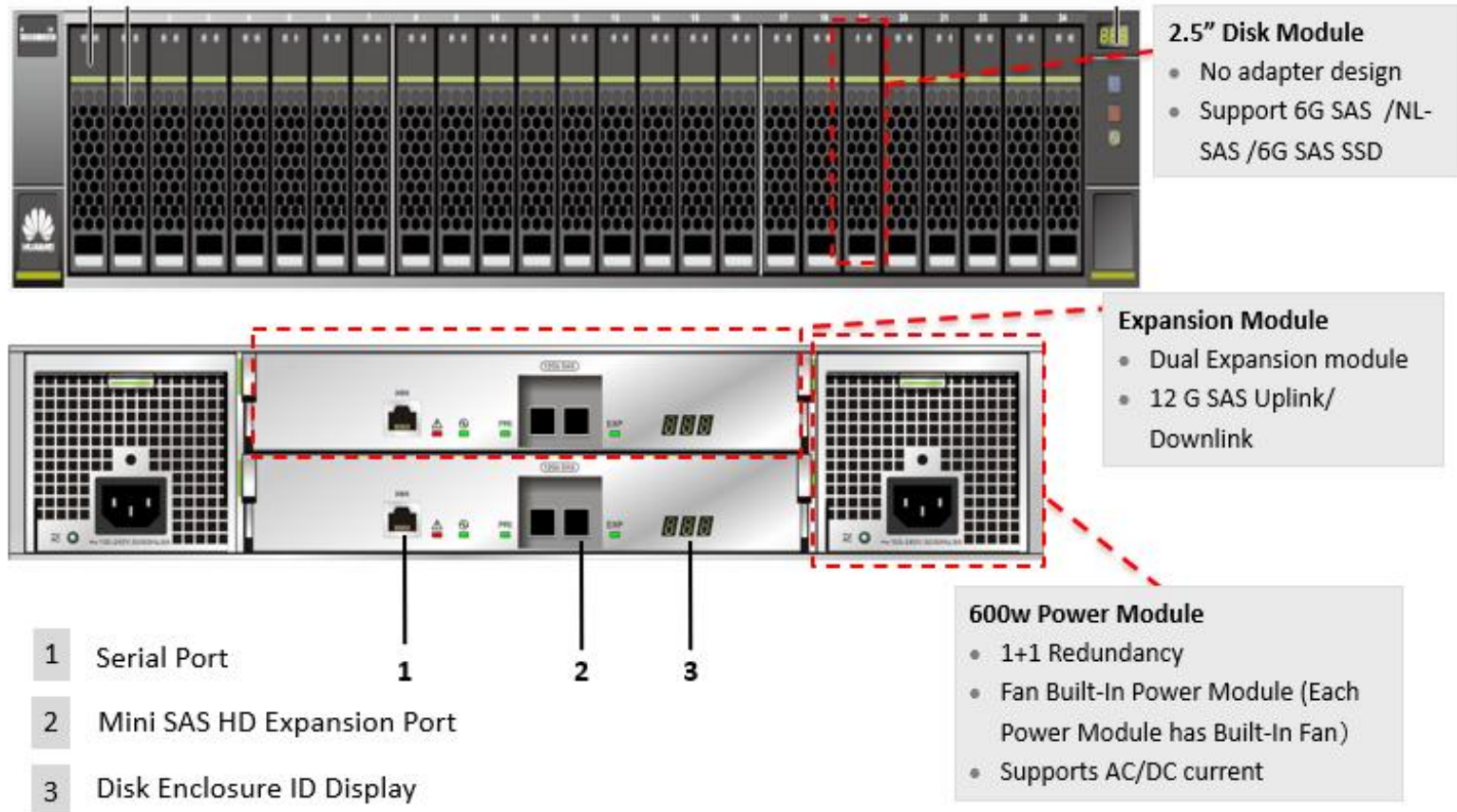
Power Module

- The AC power module supplies power to the controller enclosure, allowing the enclosure to operate normally at maximum power.



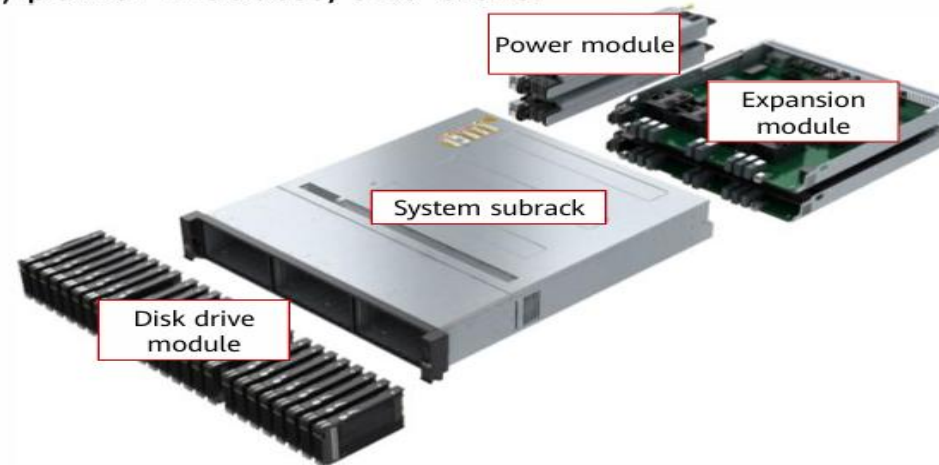
Note: Huawei OceanStor Dorado V6 is used as the example.

2U 2.5" Disk Enclosure



Disk Enclosure

- The disk enclosure uses a modular design and consists of a system subrack, expansion modules, power modules, and disks.



- The system subrack** integrates a backplane to provide signal and power connectivity among modules.
- The expansion module** provides expansion ports to connect to a controller enclosure or another disk enclosure for data transmission.
- The power module** supplies power to the disk enclosure, allowing the enclosure to operate normally at maximum power.
- Disks** provide storage space for the storage system to save service data, system data, and cache data.

Expansion Module



Expansion module of a 2 U SAS disk enclosure
Connects to a controller enclosure through expansion ports



Expansion module of a 2 U smart NVMe disk enclosure
Connects to a controller enclosure through expansion ports

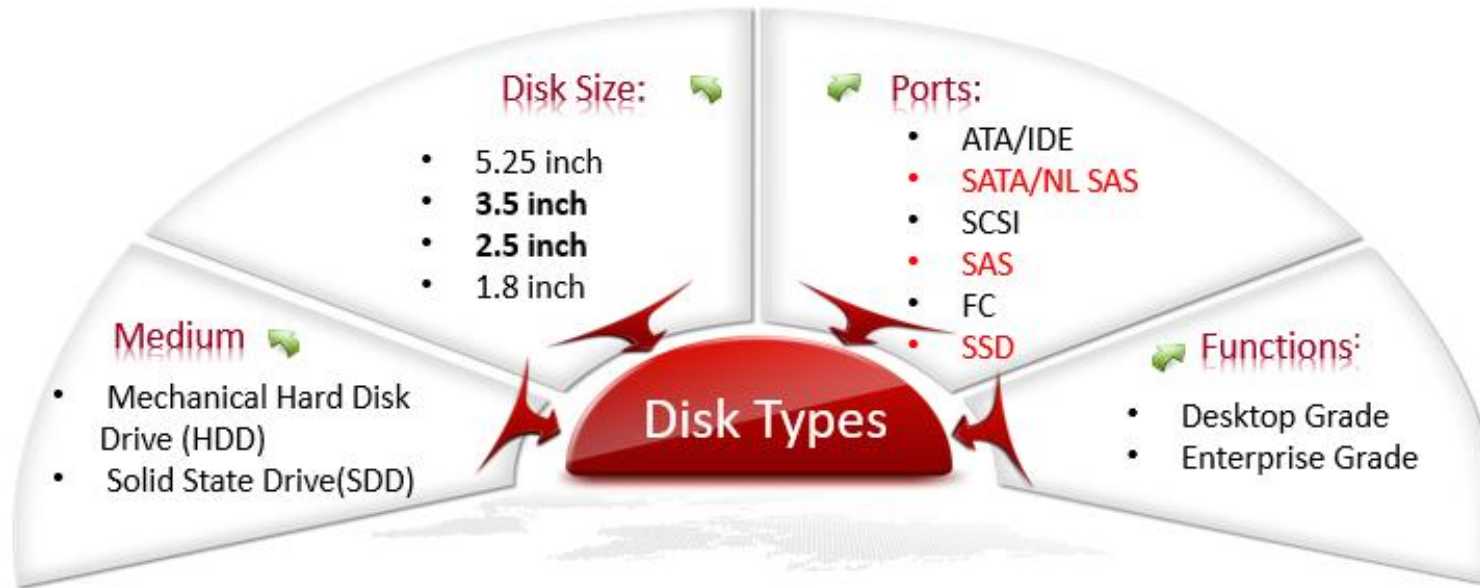
The **expansion module provides** expansion ports to connect to a **controller enclosure** or another **disk enclosure** for data transmission.

Chapter
03

Supported Hard Disk Types



Disk Types



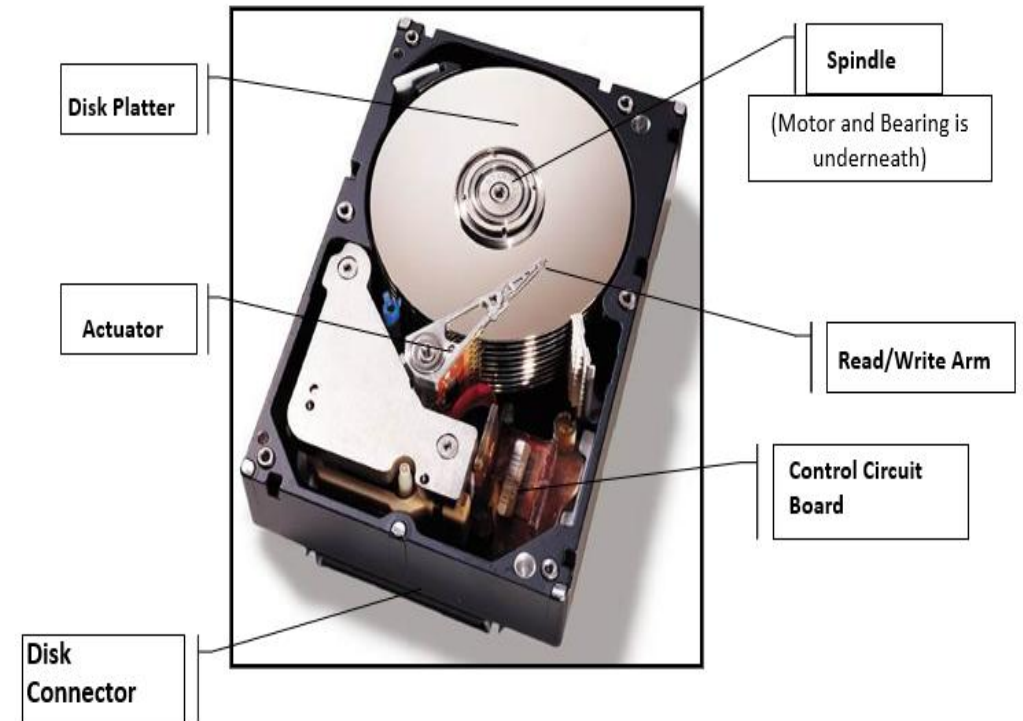
- **Desktop grade** hard disks are mainly for personal and home use and mostly used in desktop PC and laptops.
- **Enterprise grade** hard disks are mainly for enterprise level usage and mostly used in servers, storage arrays, and workstations.

Mechanical Hard Disk Drive (HDD)



Components of a Hard Disk

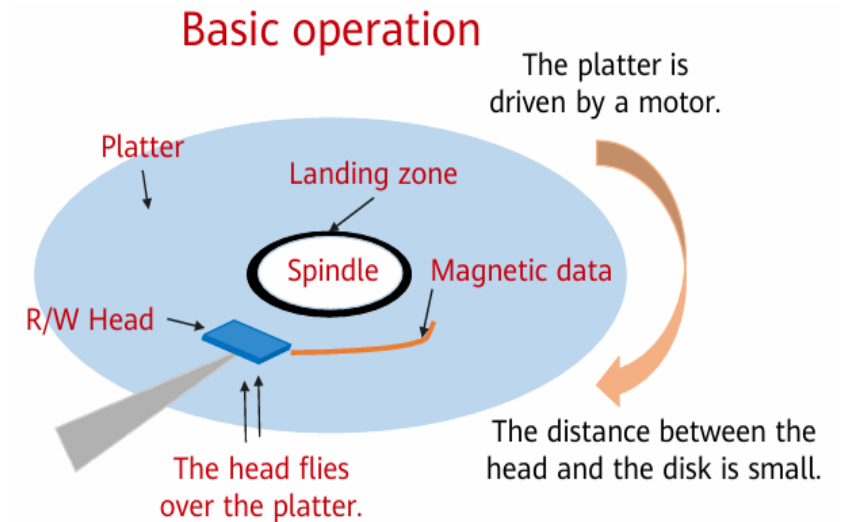
- ❑ **The Read/Write Arm** : Handles the data reading and writing
- ❑ **Platter**: Preserves or saves the written data.
- ❑ **Actuator**: Moves the Read/Write arm to the specified locations on the platter.
- ❑ **Spindle**: Rotates the disk platter and moves the disc's specified location to under the Read/Write arm.
- ❑ **Control Circuit Board**: Controls the disk speed, movement of the actuator and send commands to the disk head.
- ❑ **Disk Connector**: The interface that connects the hard disk to the rest of the system.



HDD Working Principles

❑ Working principles:

- At the beginning, the **read/write head** is in the landing zone near the **spindle** of the **platters**.
- The spindle connects to all **platters** and a motor. The **spindle** motor rotates at a constant speed to drive the **platters**.
- When the spindle rotates, there is a small gap between the head and the platter, which is called the flying height of the head.
- The **head** is attached to the end of the **actuator arm**, which drives the head to the specified position above the platter where data needs to be written or read.
- The **head** reads and writes data in binary format on the **platter** surface. The read data is stored in the **flash chip** of the disk and then transmitted to the program.



Data Organization on a Disk

- **Platter Surface:**

Each disk platter has two usable surfaces for data storage. Surfaces are numbered starting from 0, and each one corresponds to a read/write head.

- **Track:**

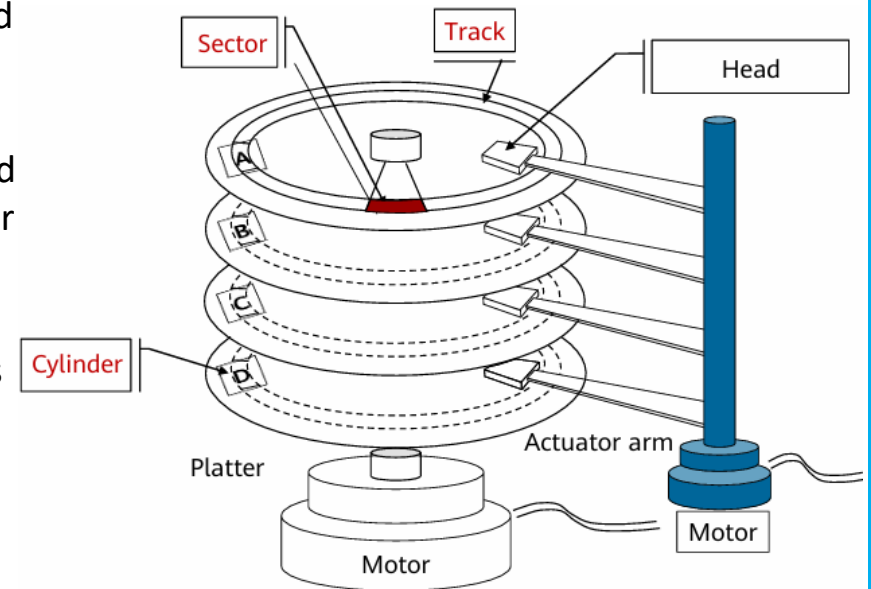
Tracks are concentric circles on a platter where data is stored. They are numbered from the outer edge inward. Each surface typically contains 300 to 1024 tracks, or more in modern disks.

- **Cylinder:**

A cylinder consists of tracks with the same number on all platter surfaces. Data is read/written sequentially from head 0 downwards. The switching between cylinders is called **seek**.

- **Sector:**

Each track is divided into sectors, which are the smallest addressable units of storage. Most sectors store 512 bytes, though some disks use larger 4 KB sectors.



Disk Capacity and Cache

- Disk capacity
 - $\text{Disk capacity} = \text{Number of cylinders} \times \text{Number of heads} \times \text{Number of sectors} \times 512 \text{ bytes}$.
The unit is MB or GB. The disk capacity is determined by the **capacity of a single platter** and the **number of platters**.
- Cache
 - Because the processing speed of a CPU is much faster than that of a disk, the CPU must wait until the disk completes a read/write operation before issuing a new command. To solve this problem, a cache is added to the disk to improve the read/write speed.

Factors Relevant to Disk Performance



Rotation speed

Primary factor that determines the throughput in the case of sequential I/Os

Seek speed

Primary factor that affects the random I/O performance

Single platter capacity

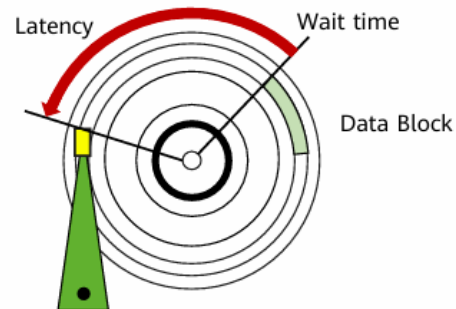
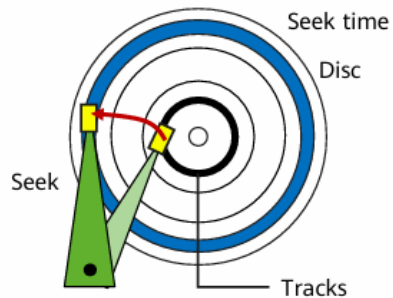
Indirect factor for disk performance

Port speed

The least important factor for disk performance

Average Access Time

- The average access time is determined by:
 - Average seek time
 - Average latency time



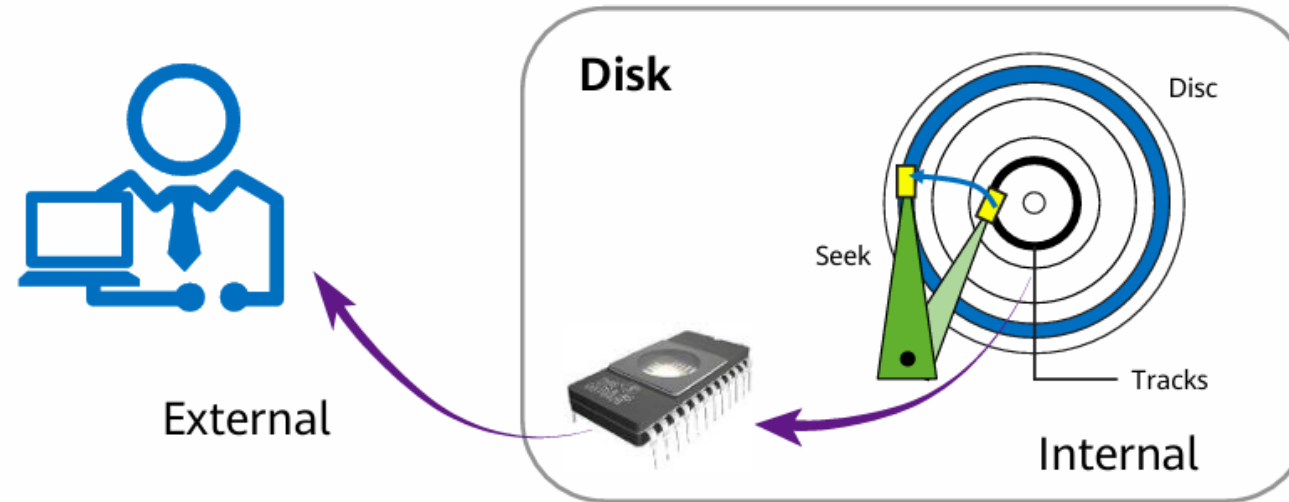
▣ Average Seek Time:

- The time it takes for the read/write head to move from its current position to a target track.
 - It's a key factor in the disk's internal data transfer rate—the shorter, the better.

▣ Average Latency Time:

- The waiting time for the desired sector to rotate under the head after reaching the correct track.
 - Typically, it's half the time of a full disk rotation. So, higher rotation speed means lower latency.

Data Transfer Rate



Data Transfer Rate

- It refers to how fast data is transferred to and from the hard disk, whether during reading or writing.
 - The unit is usually megabytes per second (MB/s), and it is divided into two types:

Internal Transfer Rate → Reflects the hard disk's actual performance in reading/writing data from the disk itself.

External Transfer Rate → Indicates the speed at which data is delivered from the hard disk to the computer.

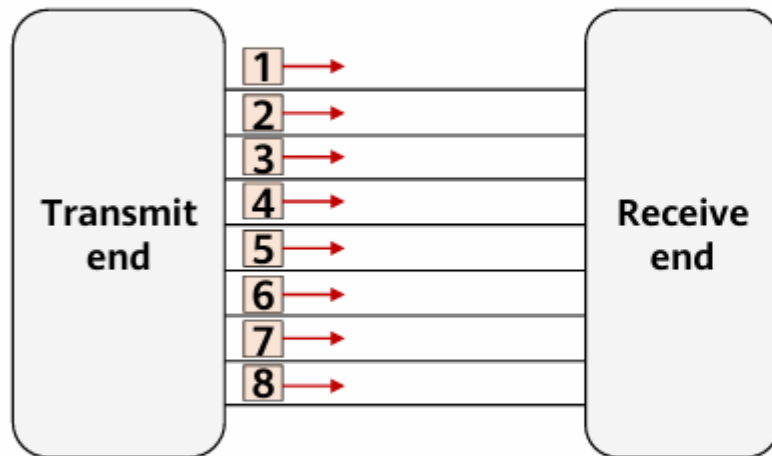
Disk IOPS and Transmission Bandwidth

- IOPS
 - Input/Output operations per second (IOPS) is a key indicator to measure disk performance.
 - IOPS is calculated by the seek time, rotation latency, and data transmission time.
- Transmission bandwidth (throughput)
 - Indicates the amount of data that is successfully transmitted in a unit time, that is, the speed at which data streams are transmitted. For example, if it takes 10s to write 10,000 files of 1 KB size, the transmission bandwidth is only 1 MB/s; if it takes 0.1s to write a 10 MB file, the transmission bandwidth is 100 MB/s.

Parallel and Serial Transmission

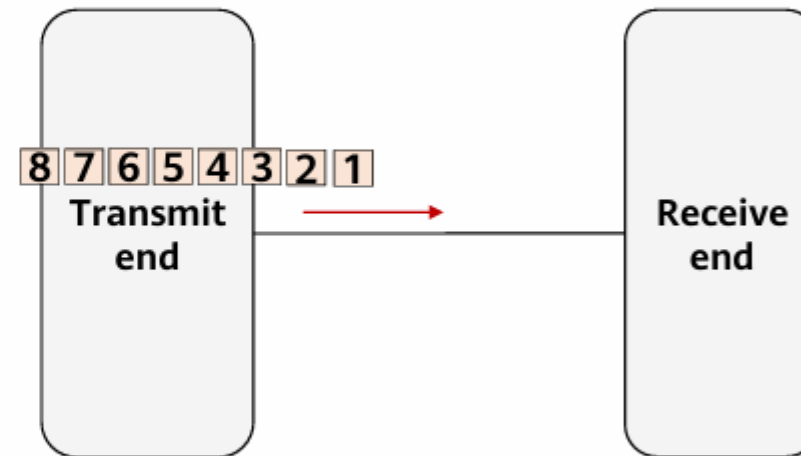
- For example, the methods for transmitting numbers 1 to 8 are as follows:

Parallel transmission



Multiple lines are connected between two ends, and one number is transmitted on each line.

Serial transmission



Only one line is connected between two ends. Eight numbers are sent on this line in sequence. The receive end has all numbers after eight transmissions.

Chapter

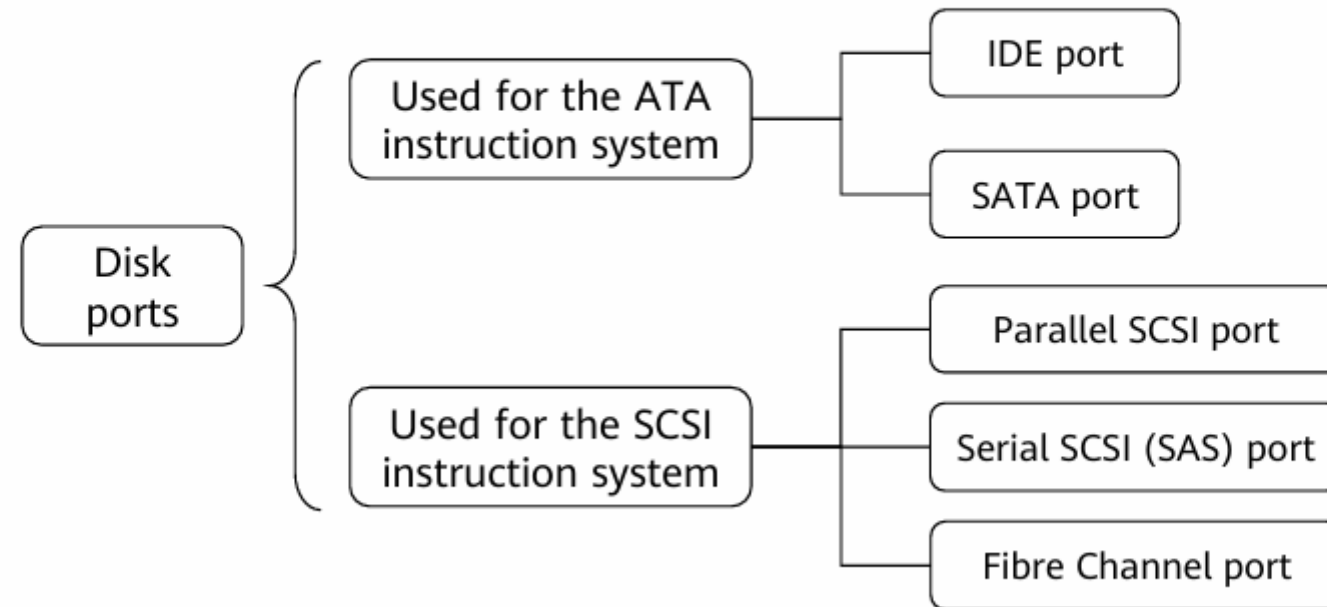
03

HDD Port Technology



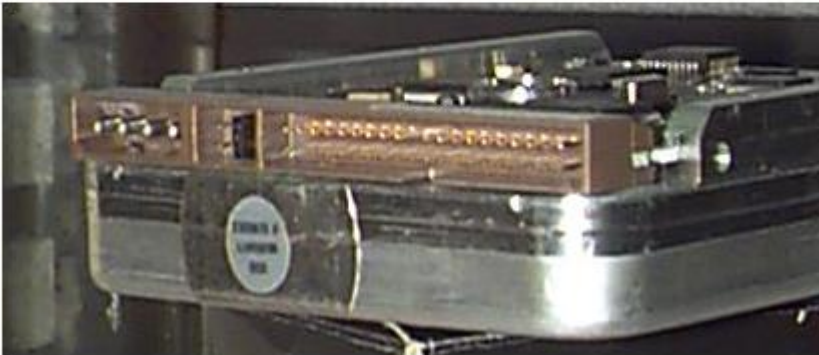
HDD Port Technology

- A disk must provide a simple port for users to access its data. Generally, disks provide the following physical ports:



IDE Disk Port

- The integrated drive electronics (IDE) port is also called the parallel ATA port.
 - ATA stands for Advanced Technology Attachment.
 - The ATA disk is also called the IDE disk.
 - The ATA port uses the parallel ATA technology.



SATA Port

- SATA is short for serial ATA.
 - SATA ports use serial transmission and provide a higher rate than IDE ports.
 - SATA uses a point-to-point architecture and supports hot swap.



SATA Version	Line Code	Transfer Rate	Throughput
1.0	8b/10b	1.5 Gbit/s	150 MB/s
2.0	8b/10b	3 Gbit/s	300 MB/s
3.0	8b/10b	6 Gbit/s	600 MB/s

SCSI Port

- SCSI is short for Small Computer System Interface.

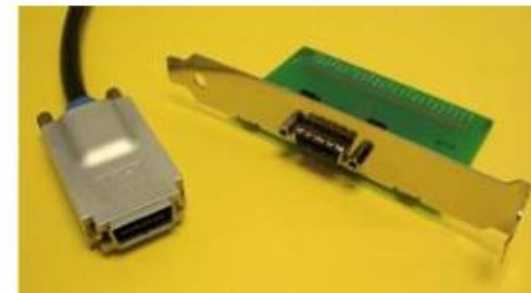


SCSI Port

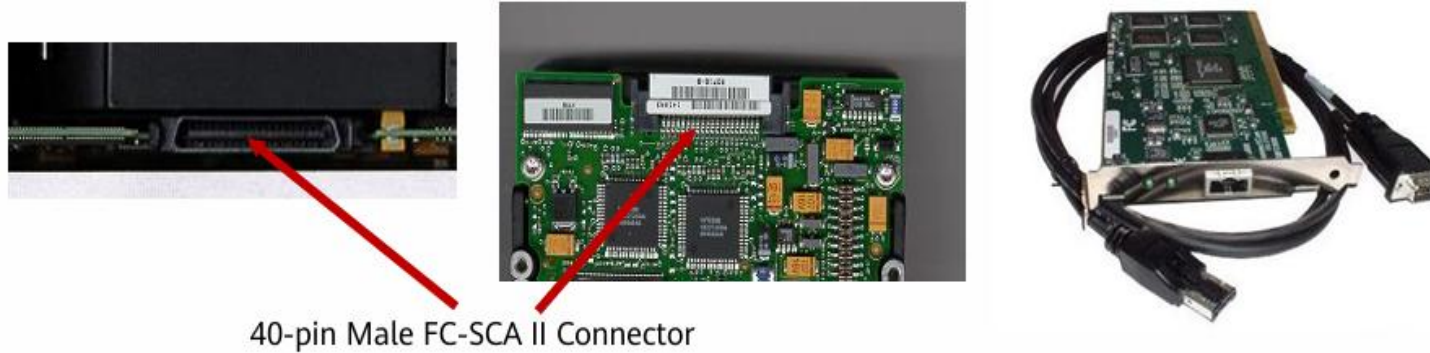
- SCSI disks were developed to replace IDE disks to provide higher rotation speed and transmission rate.
- SCSI is originally a bus-type interface and works independently of the system bus.
- **Advantages:**
 - Applicable to a wide range of devices. One SCSI controller card can connect to 15 devices simultaneously.
 - High performance (multi-task processing, low CPU usage, fast rotation speed, and high transmission rate)
 - SCSI disks can be external or built-in ones and are hot-swappable.
- **Disadvantages:**
 - High cost and complex installation and configuration.

SAS Port

- SAS is short for Serial Attached SCSI.
 - SAS is a point-to-point, full-duplex, and dual-port interface.
 - SAS is backward compatible with SATA.
 - Rate: 600 Mbit/s per channel
 - SAS features high performance, high reliability, and powerful scalability.



Fiber Channel Port



- ❑ **Fiber Channel** disks use the Fiber Channel arbitrated loop (**FC-AL**).
- FC-AL is a dual-port serial storage interface based on the SCSI protocol.
- FC-AL supports full-duplex mode.
- **Fiber Channel** provides a universal hardware transmission platform for upper-layer protocols (SCSI and IP). It is a serial data transmission interface that features high speed, high reliability, low latency, and high throughput.

Chapter
03

Solid State Drives (SSD)



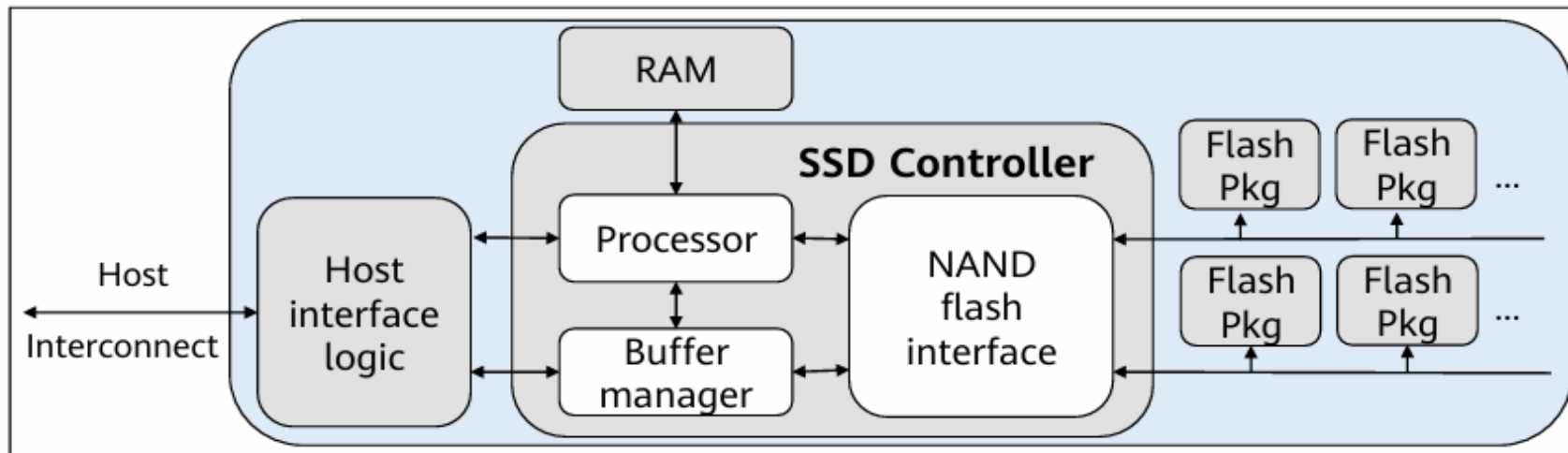
SSD Overview

- As price lowers and storage capacity increases, SSD is becoming more and more popular in the market.
- **SSD Principles:**
 - Uses flash technology to store data.
 - No mechanical structure with lower energy consumption, low heat signature and low noise.
- Limited lifetime due to SSD usage frequency.
- **3 types of SSD:**
 - SLC (Single Level Cell)
 - MLC (Multi Level Cell)
 - TLC (Triple Level Cell)



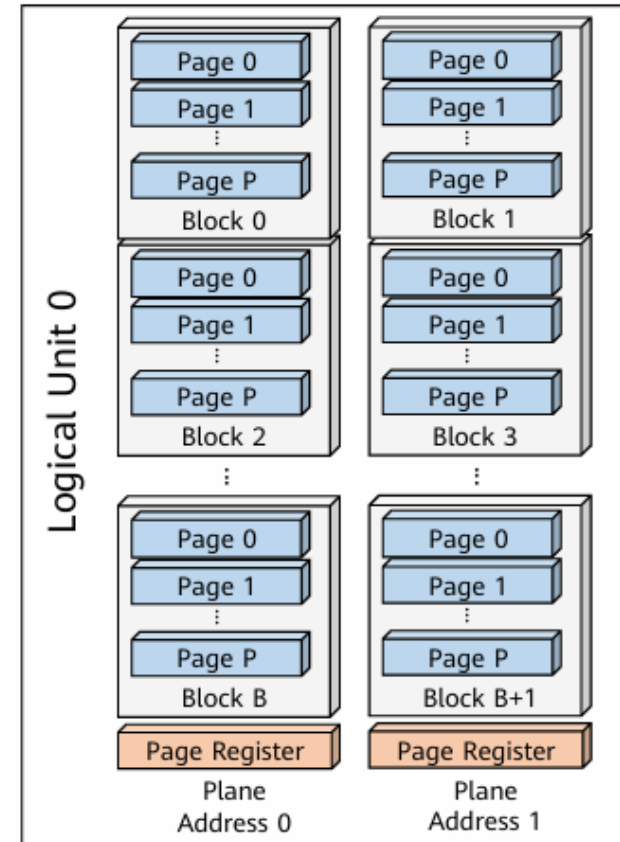
SSD Architecture

- An SSD consists of a control unit and a storage unit (mainly flash memory chips).
 - Control unit: SSD controller, host interface, and DRAM
 - Storage unit: NAND flash



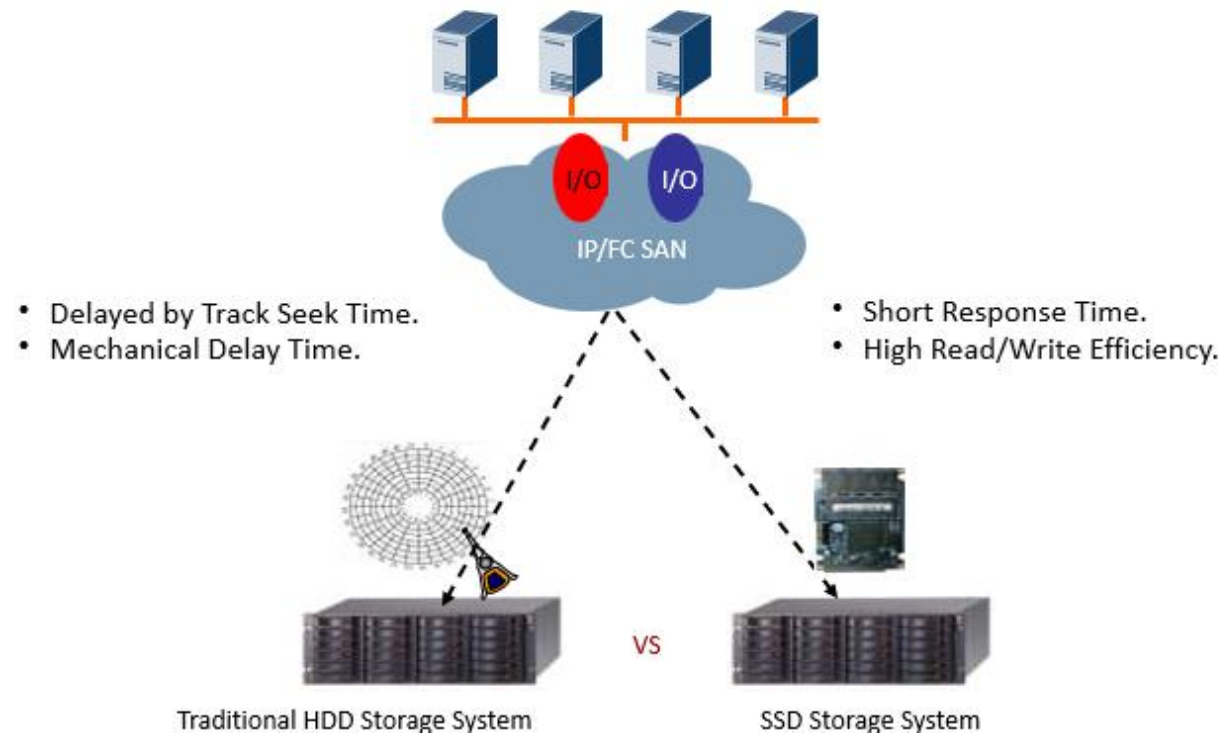
NAND Flash

- Internal storage units in NAND flash include:
 - LUNs, planes, blocks, pages, and cells
- Operations on the NAND flash include erase, program, and read.
- NAND flash is a non-volatile medium. A block must be erased before new data is written to it. A program/erase (P/E) cycle is the process of erasing a block and then writing it again.



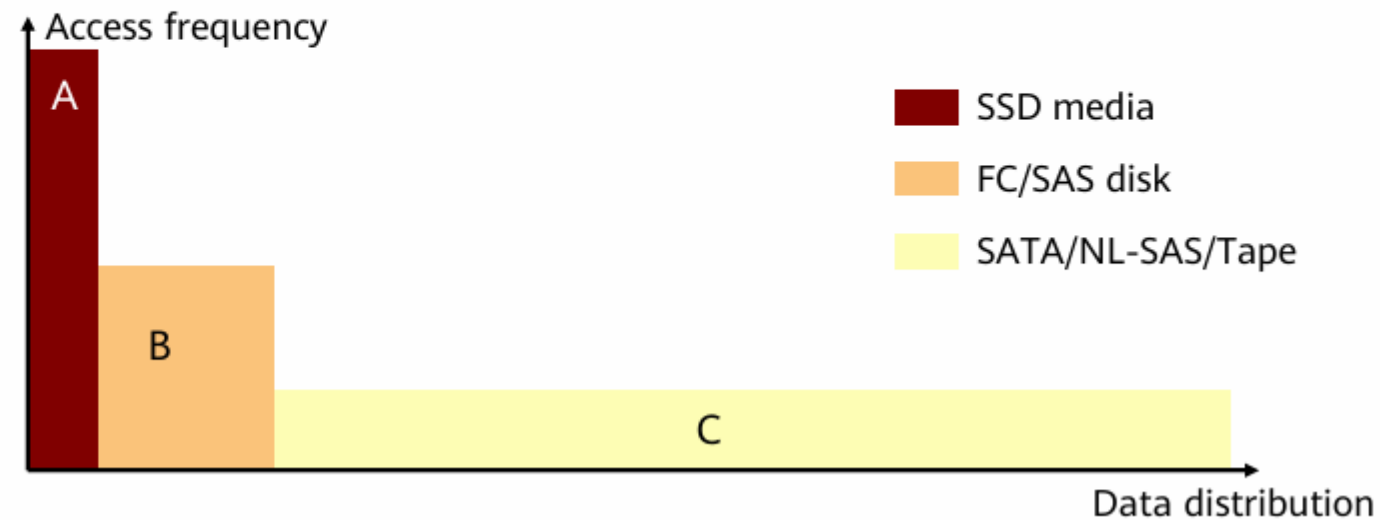
Advantages of SSD Performance

Performance Advantages of SSD Technology



Use of SSDs in Storage Systems

- Class A applications: high-concurrency applications featuring random reads and writes, such as databases
- Class B applications: large files, images, and streaming media featuring sequential reads and writes
- Class C applications: data backup or rarely used applications



Summary

- Architecture of Storage Systems → general design and main components of storage.
- Supported Hard Disk Types → different types of disks used in storage systems.
- Mechanical Hard Disk Drive (HDD) → description of traditional hard disks and how they work.
- HDD Port Technology → the interfaces used to connect HDDs with the system.
- Solid State Drives (SSD) → modern drives that are faster, more efficient, and consume less power compared to HDDs.



Thank You