Module 1 / Unit 2

Storage Devices

Objectives

On completion of this unit, you will be able to:

- Describe the features and performance characteristics of server-class hard disks and solid state drives.
- Describe features of the SCSI interface and know how to configure devices correctly.
- Describe features of the SATA and SAS interfaces and know how to configure devices correctly.

Mass Storage Drives

Like a desktop PC, a server will be configured with one or more mass storage devices to store the operating system and applications software. Disk space may also be provided on a file server as a shared resource for network users.

Unlike most desktop PCs however, server disks will typically be faster for higher performance and add resilience features such as RAID - the ability for multiple hard disks to work together to increase access speed to files or provide redundancy in case one of the drives fails.

All new servers use disk interfaces based on either Serial Attached SCSI (SAS) or Serial Advanced Technology Attachment (SATA). Some older servers may be configured with parallel SCSI interfaces.

Drives are available in several sizes. Most hard drives are 3.5" or 2.5" width (the width refers to the size of the drive bay rather than the disk unit). 2.5" now dominates the market, with 3.5" drives gradually being phased out. Internal optical and tape drive units often use the larger 5.25" form factor. There is also a distinction between half height and low profile units.

Hard Disk Drives (HDD)

Most hard disks are of the mechanical type (HDD). Data is stored on a number of metal or glass platters coated with a magnetic substance. The top and bottom of each platter is accessed by its own read/write head, moved by an actuator mechanism.

Delivery Tips

Students may not be familiar with SCSI as it is no longer on the A+ syllabus so be prepared to spend some time explaining its configuration issues.

Timings

Theory & Review Questions - 45 minutes

Discuss some of the issues in provisioning disks for servers:

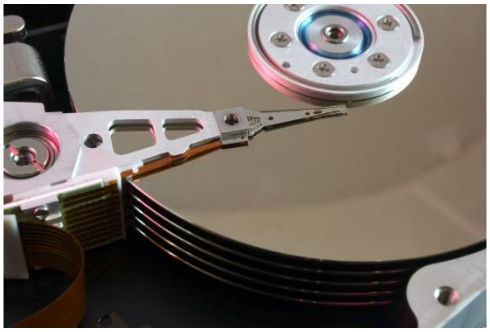
** Server role and its effect on disk usage*

** Price per gigabyte*

** Price per I/O*

** Price versus performance requirements*

** Tiered storage*



Stack of platters on spindle

The heads do not actually touch the surface of the platters. The platters are mounted on a spindle and spun at high speed and the heads "float" above them at a distance of less than a millionth of an inch. The disk unit is kept sealed to maintain a constant air pressure (important for keeping the drive heads at the correct distance from the platters) and to prevent the entry of dust.

Each side of each platter is divided into circular **tracks** and each track contains a number of **sectors**, each with a capacity of 512 bytes. The collection of tracks in the same place on each platter is called a **cylinder**. This low-level formatting is also referred to as the drive *geometry*.

*As hard drive sizes have increased, some disk models now use **Advanced Format**, with 4 kilobyte (4K) sector sizes. If supported by the OS and PC firmware, these can be used in native mode; if not, the drive controller will usually present the disk in 512 emulated (512e) mode.*

The performance of a hard disk is a measure of how fast it can read and write data. There are a number of factors that determine overall hard disk performance. One factor is the speed at which the disks can spin (measured in **Revolutions Per Minute [rpm]**). The higher the rpm, the faster the drive is. High performance drives are rated at 15,000 or 10,000 rpm; average performance is 7200 or 5400 rpm.

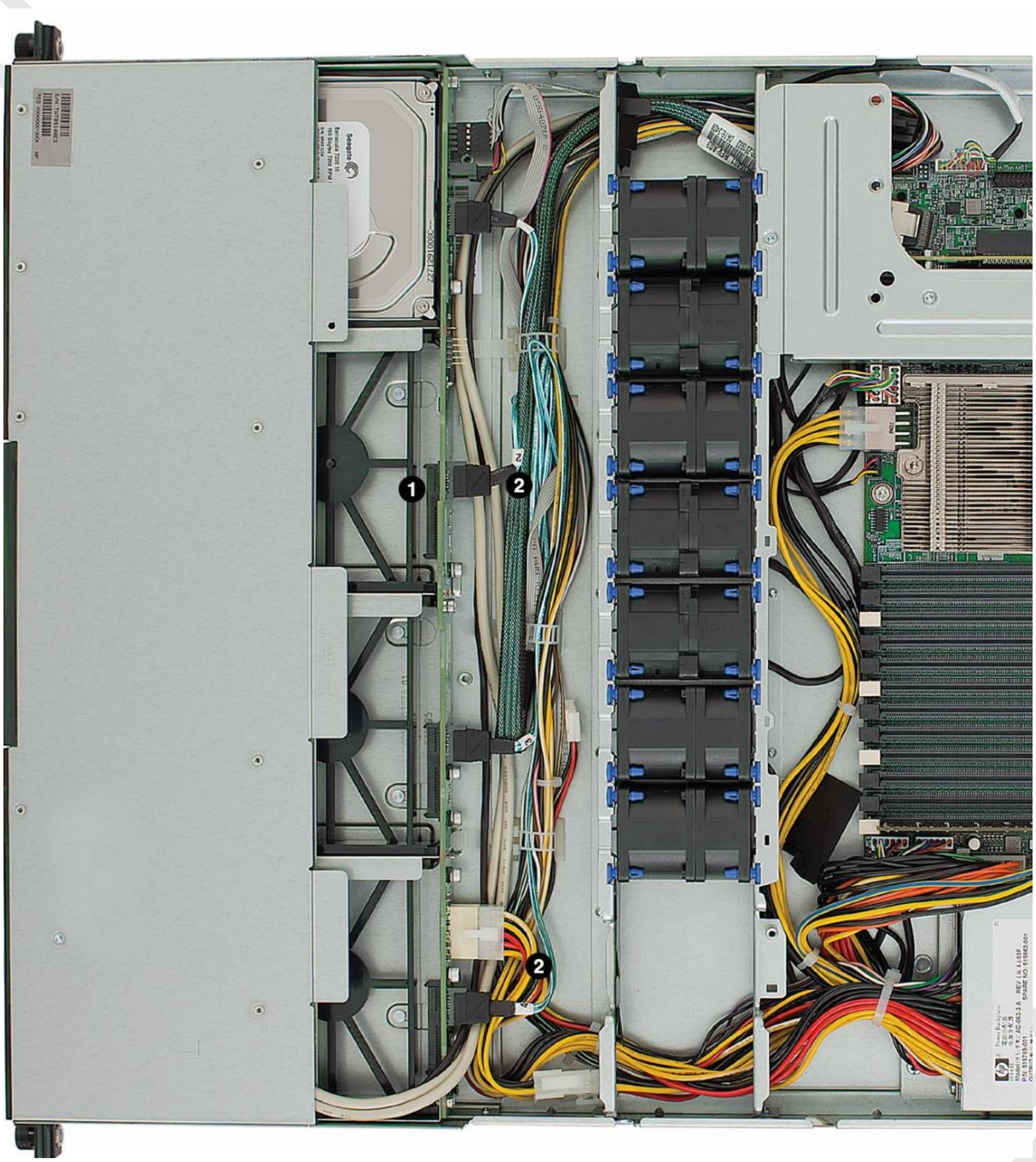
RPM is one factor determining **access time** (measured in milliseconds), which is the delay that occurs as the read/write head locates a particular track position (**seek time**) and sector location (**rotational latency**) on the drive. A high performance drive will have an access time below 3 ms; a typical drive might have an access time of around 6 ms.

The **internal transfer rate** (or data or disk transfer rate) of a drive is a measure of how fast read/write operations are performed on the disk platters. A 15K drive should support an internal transfer rate of up to about 180 MBps while 7.2K drives will be around 110 MBps. The **external transfer rate** (often simply described as the transfer rate) measures how fast data can be transferred to the CPU across the bus. Cache memory can help to sustain better transfer rates. A high performance disk may feature an 8 MB or better cache.

The last, but by no means least, characteristic of a hard drive is its capacity. Advances in hard disk technology have enabled disks of up to 8 terabytes (8000 GB) to be produced but often smaller capacities are used because they offer better reliability and redundancy (as part of an array).

Enclosures and Backplanes

A drive is housed in an enclosure or caddy. The drive may then be connected to the host bus adapter via a cable but on enterprise-class servers it is more usual for the enclosures to be connected via a **backplane**. Rather than using cabled connectors, the drives plug (or "mate") into a combined data and power port on the enclosure. This means that drives can be easily added and removed from the front of the case without having to open the chassis.



Enclosure and backplane on an HP server - 1) The drive mates with the port on the backplane card; 2) Data and power cables on the other side of the backplane card connect to the drive port and PSU

The drives are secured and released from the server using a latch. Many server drives are hot-swappable, meaning that they can be added or removed without powering down the server.

The SCSI Interface

Parallel SCSI is obviously diminishing in importance but students still need to know all the configuration issues.

The **Small Computer System Interface (SCSI)** has been in use as an expansion bus interface since the 1980s. There have been many revisions to the standard. Originally a parallel interface, SCSI is now used for serial connections for disk drives (Serial Attached SCSI) and peripheral devices (Firewire or IEEE 1394).

More information about SCSI can be located on the T10 Committee website (www.t10.org) and from the SCSI Trade Association (www.scsita.org).

Parallel SCSI Standards

The original iterations of the SCSI standards defined a parallel bus, originally 8 bits wide but later updated to 16 bits (wide SCSI). As a parallel interface, SCSI is more-or-less obsolete but the main iterations are listed below:

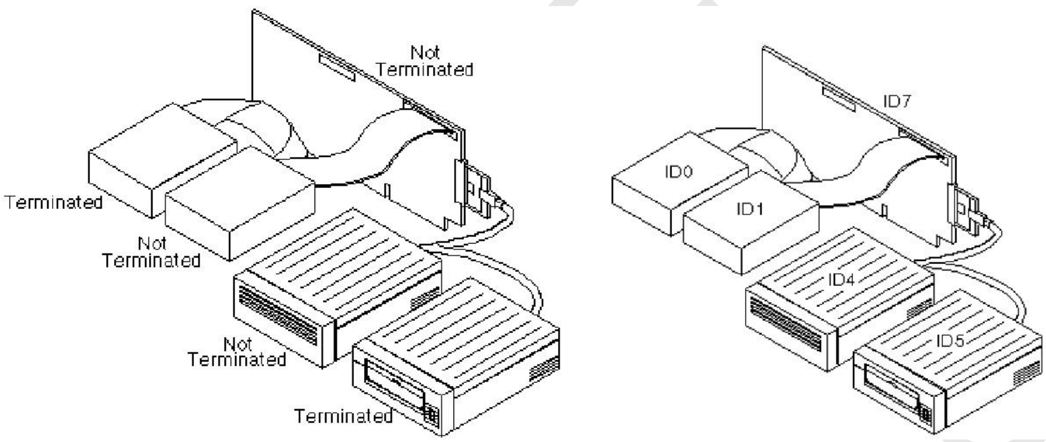
Interface	Devices (Excluding Host Adapter)	Rate	Max Cable Length (m)			Connector
			SE	LVD	HVD	
SCSI-1	7	5 MBps	6	-	25	50-pin
Fast SCSI	7	10 MBps	3	-	25	50-pin
Fast-Wide SCSI	15	20 MBps	3	-	25	68-pin
Ultra SCSI	7	20 MBps	1.5	-	25	50-pin
Wide Ultra SCSI	15	40 MBps	-	-	25	68-pin
Ultra2 SCSI	7	40 MBps	-	12	25	50-pin
Wide Ultra2 SCSI	15	80 MBps	-	12	25	68-pin / 80-pin
Ultra3 SCSI (Ultra160 SCSI)	15	160 MBps	-	12	-	68-pin / 80-pin
Ultra320 SCSI	15	320 MBps	-	12	-	68-pin / 80-pin
Ultra640 SCSI	15	640 MBps	-	12	-	68-pin / 80-pin

The number of devices listed in the table excludes *the host adapter. So for example, narrow SCSI supports 8 devices including the host adapter and wide SCSI supports 16.*

Parallel SCSI Configuration Issues

Given the number of different versions of the standard, SCSI configuration is relatively complex.

- **Host adapter** - the SCSI host adapter must be installed and recognized by the system for devices to be attached and detected. A third-party driver might have to be installed for the host adapter to be recognized.
- **Bus width** - SCSI originally supported 8 devices (the host adapter counts as a device). Wide SCSI supports up to 16 devices.
- **Signaling** - SCSI specifies three signaling methods. Most buses and devices now use LVD (Low Voltage Differential). SE (Single Ended) devices can be added to a LVD bus, but it reduces the performance of the whole bus. H(igh)VD is incompatible with the other two and must not be mixed.
- **Termination** - a SCSI bus must be terminated at both ends, usually by enabling termination on the first and last devices in the chain. Termination may either be enabled internally on the device by setting a switch or by physically connecting a terminator pack to a device or the host adapter. There are passive and active terminators. Passive terminators are generally used with older devices (pre-Ultra SCSI). When installing a terminator pack, the terminator must match the signaling type (SE, LVD, HVD, or SE/LVD). Termination is also made more complex if there is a mix of narrow (8-bit) and wide (16-bit) devices on the bus.



Correct termination and device identification of a SCSI chain

- **ID** - each SCSI device must be allocated a unique ID, from 0 to 7 (or 15 for wide SCSI). IDs may be allocated automatically or by setting a jumper or click-wheel on the device itself. The order of SCSI ID priorities (from highest to lowest) is 7 through to 0 then 15 through to 8.

The host adapter is usually set to 7 or 15. A bootable hard disk is usually allocated ID 0.

Note that the host adapter counts as one device.

- **Logical Unit Number (LUN)** - some parallel SCSI devices can perform more than one function (an auto-loading tape drive for instance). In this case, each function must be allocated a **Logical Unit Number (LUN)** from 0 to 7 or 0 to 15. This is normally assigned by the manufacturer.

Longer (64-bit) LUNs are also used in Storage Area Networks (SAN) based on Fiber Channel or iSCSI. See *Unit 3.2* *for more information about SANs.*

Parallel SCSI Connectors

A SCSI port is denoted by the following symbol: . There are numerous SCSI connectors. The most common are:

- IDC50 - a 50-pin internal connector used with early SCSI devices (SCSI-1).
- CN50 - a 50-pin Centronics-style connector used for external connections in early SCSI devices.
- HD68 - 68-pin connectors used for internal and external ports. 68-pin adapters support Wide SCSI.

Internal (left) and external male HD connectors

- Single Connector Attachment (SCA) - an 80-pin connector that incorporates both a power connector and configuration wires, allowing for hot swappable drives.

Serial Attached SCSI (SAS)

Make sure students understand that SAS delivers better performance than SCSI or SATA.

Serial Attached SCSI is the next generation of SCSI interface. It uses a serial interface with full-duplex communication over 2-pair wiring (much like PCI Express) but retains support for the SCSI command set.

SAS components can be rated at 3 Gbps (SAS-1), 6 Gbps (SAS-2), or 12 Gbps (SAS-3).

A significant feature of SAS is support for *thousands* of devices (up to 16,384), using an Ethernet switch-like device called an **expander**. Each device is identified by a unique, manufacturer-coded ID, so there is no manual configuration to be performed. Also, SAS does not require termination, removing another complex configuration issue.

SAS also goes some way to uniting the SATA and SCSI standards. It provides both hardware support (the same connectors and cable) and software support (through the SATA Tunneling Protocol) for SATA drives. This offers the opportunity to mix low-cost SATA drives with high-cost, high-performance SAS drives in an integrated storage solution.

Serial Attached SCSI Installation

SAS devices feature a combined 7-pin data port and 15-pin power port (though some drives may also have a legacy 4-pin Molex power port too). In fact, on most SAS devices there are two data connectors (for redundancy). This is referred to as "dual-port". Obviously the backplane or cable must also support a redundant connection for this to work.

SAS connections are typically either single-lane (a simple adapter to device connection) or multi-lane (a single port on the adapter is connected to four devices).

Ports can either be "straight-through" or flush / surface mounted with right-angle connectors. Flush-mounted fittings are used in blade servers and other systems with restricted space. SAS backplane connectors are designed to "blind-mate", which means that a connection is made reliably when a drive caddy is inserted into a backplane. The design of the pins also reduces the chance of damage through ESD or a power spike (the full power pins connect after the other pins).

Hot-pluggable drives are not screwed into the chassis but slot into a drive cage. The cables for the devices connect to the drive cage (backplane) rather than the drive units.

Hot-pluggable drives on HP Proliant server

When removing a drive, you will probably need to use a utility or the OS to stop the device. This completes any cached write operations and prevents the NOS from trying to write data to the device while it is being removed. You also need to power down the device, either using software or a switch on the drive bay. Indicators on the drive bay should show when a drive is safe to remove. The drives are physically released and inserted using a lever or latch mechanism.

Unused drive slots should be filled with blanks to maintain the correct airflow and cooling within the chassis.

Serial Attached SCSI Connectors

A number of connectors are associated with the use of SAS host adapters and drives:

- SFF 8087 - internal mini connector for both adapter card ("host") and drive or backplane ports ("target").
- SFF 8088 - external mini connector for both adapter card and drive enclosures.
- SFF 8484 ("Multilane") - legacy 32-pin internal or external HD connector (host) supporting four 7-pin lanes (target). The remaining pins are used for "sideband" signals (LEDs and status monitoring).
- SFF 8470 ("Infiniband") - legacy jackscrew connector for both internal and external use.
- SFF 8482 - internal connector compatible with both SAS and SATA drives. This type of connector would be used principally to attach SATA devices (such as DVD drives) to an SAS bus.

The Serial ATA Interface

Serial ATA (SATA) was developed to address the limitations of the now obsolete parallel ATA or IDE interface. SATA would be used on low-end server hardware as a cheaper option than SAS.

4 SATA motherboard ports in front of an IDE port on an Intel motherboard

As the name suggests, SATA transfers data in *serial* format. This allows for thinner, longer, more flexible cables (up to 1m [39"]) with smaller, 7-pin data connectors. Each port supports a single device.

SATA cable for HP workstations

The first commercially available SATA standard supports speeds of up to 1.5 Gbps. This standard was quickly augmented by SATA revision 2 (3 Gbps) and then SATA revision 3 (6 Gbps). Another key advantage of SATA over PATA is that SATA is a **hot swappable** interface. This means that a compatible drive can be connected or disconnected while the system is running.

Other additions in SATA revision 2 include the use of port multipliers, which allow up to 15 drives to be connected to a single SATA adapter, and **Native Command Queuing (NCQ)**, which enables the drive to analyze read/write operations and perform them in the most efficient manner, depending on the location of data on the disk

SATA revision 3 adds some extensions to NCQ to support isochronous data transfer (prioritizing real time data such as video to ensure smooth playback). SATA revisions 3.1 and 3.2 add better support for Solid State Drives (SSD).

More information on SATA standards can be obtained from www.sata-io.org.

Hot Swapping

One of the major advantages of SATA over PATA is the support for hot swapping and consequently better compatibility with RAID configurations. Serial ATA 15-pin power connectors have been redesigned to provide support for both hot plugging and a 3.3V power supply in addition to the usual 5V and 12V.

Many drives retain a 4-pin Molex port for compatibility with legacy power supplies). Molex-SATA conversion adapters are also available.

SATA power connector

SATA and SAS

As mentioned earlier, SAS includes hardware and software support for SATA devices. The reverse is not true however; SAS devices *cannot* be plugged into a SATA bus.

Solid State Drives (SSD)

Recently, flash memory based drives (**Solid State Drives [SSD]**) have been introduced to the market.

There are broadly two types of SSD: Single Layer Cell (SLC) and Multi Layer Cell (MLC). SLC stores 1 bit per cell while MLC stores 4 or more bits per cell, yielding higher capacities at lower cost. MLC can be slower however as it requires substantial error correction processing. It is also perceived as not reliable enough for enterprise server applications as it supports fewer write cycles, though improvements in the technology (enterprise-grade eMLC technology for instance) may change that perception over time.

HP SSD with SATA interface (Image © 2014 Hewlett-Packard Development Company)

The advantages of flash memory-based SSDs are that the lack of moving parts makes them quieter, more power efficient, and less prone to catastrophic failure or damage due to shock (dropping or moving a device rapidly for instance). Read times are better because seek time and consequently the effect of file fragmentation is eliminated. They are also less susceptible to data loss in the event of power failure. Most drives still feature DRAM-based write cache to improve performance. In the event of a power failure, unwritten cache would be lost. However, the DRAM cache may be backed up by a battery to cover this eventuality.

The main disadvantage is the high cost; a *64 GB* SSD costs a bit more than a *2 TB* HDD (a 1 TB SSD can cost the same as a top-end server).

SSD Interfaces

An SSD might be installed as the server's only internal drive. The SSD would normally be used to install the OS and software applications while a disk array or other shared storage would be used for data files.

An SSD might be installed to an SAS or SATA port as with a mechanical HDD. SSDs are now often provided as PCIe adapter Add In Cards (AIC) though, taking advantage of the higher bandwidths available from a native PCIe interface. Where SAS uses SCSI and SATA uses the **Advanced Host Controller Interface (AHCI)** to communicate with the bus, PCIe-based SSDs use the **Non-Volatile Memory Host Controller Interface Specification (NVMHCI)** or **NVM Express (NVMe)** for short. One of the big advantages of NVMe over SCSI and AHCI is support for much longer command queues (64,000 versus 64 in SCSI).

An SSD can also be packaged as a 2.5" drive and connect to an NVMe controller via a SFF 8639 / U.2 port.

SSD Performance

SSDs normally outperform HDDs but there are situations where they can perform worse than HDDs, when serving large (GB) files for example.

When making a detailed comparison between different types of storage technology, you need to compare performance against different types of data transfer. For example, read and write performance are not equivalent. There are also differences between sequential access (reading data from the same "block" as might happen when transferring a large file) and random access (reading data from different locations on the drive or transferring lots of small files for instance). Along with the data throughput (measured in Mbps) and latency / access time, you may need to consider the number of **Input / Output Operations per Second (IOPS)** that can be achieved by a device for different kinds of data transfer operation.

Flash chips are also susceptible to their own type of degradation over the course of many write operations, so the drive firmware and operating system must use **wear leveling** routines to prevent any single storage location from being overused and optimize the life of the device.

Review Questions / Module 1 / Unit 2 / Storage Devices

Answer these questions to test what you have learned in this unit. You can submit your answers and review the model answers on the course website.

1) True or false? LFF and SFF represent two different fixed drive technologies?

2) What two mechanical factors affect hard disk access times?

3) How would you expect hot-pluggable drives to be attached to a server?

4) What type of SCSI hard disk subsystem would you expect to be provisioned on a new server - SAS or Ultra320?

5) What are the main issues to consider when configuring a parallel SCSI bus?

6) True or false? SATA and SAS use compatible connectors.

7) What is the principal factor reducing the operational life of SSDs and what process is used to mitigate it?

8) You need to configure servers to cache files for a content delivery network. What type of storage technology is best suited to this task?