

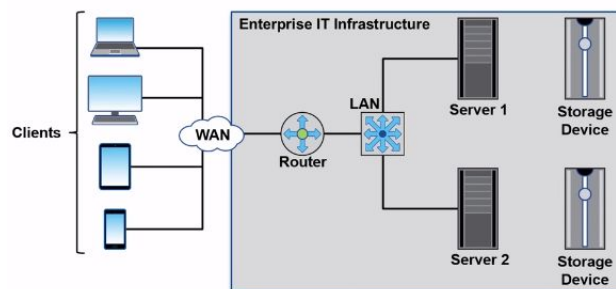
- 1.1. **Fuente.** Storage101x Introduction to Data Storage and Management Technologies
 - 1.1.1. [Enterprise IT Infrastructure Components](#)
 - 1.1.2. [Server-Storage Connectivity: Direct Attached Storage](#)
 - 1.1.3. [Server-Storage Connectivity: Networked Storage](#)
 - 1.1.4. [Data Center](#)

IT infrastructure. An enterprise IT infrastructure has three key components – **compute, network, and storage.** Compute refers to the computer systems or servers on which applications are hosted. These applications could be business applications, database applications, email applications, or mobile apps.

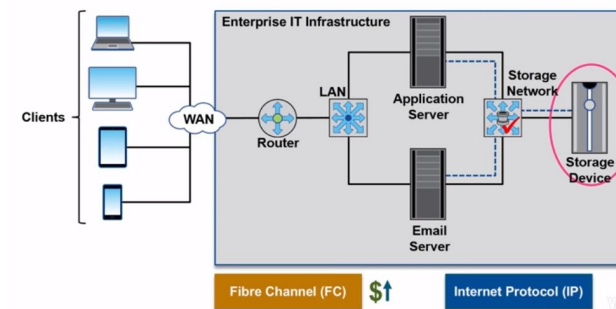
Network refers to the network components, such as switches, routers, and cables. Servers connect to an IP-based network, which allows them to communicate with one another. Clients also connect to a network, which allows them to communicate with the servers and use these applications. Storage refers to the persistent storage devices on which an enterprise stores application data. When storage is directly connected to a server, it is called as direct attached storage or **DAS**

...A main disadvantage of DAS is that it is inconvenient to transfer data over a network. Other servers in the environment can only access a storage device by going through the server to which the storage device is connected.

Components of Enterprise IT Environment



Server-Storage Connectivity: Networked Storage



One other challenge with DAS is that of storage utilization. For example, the storage device connected to a file server may be more heavily utilized than the storage device that is connected to an email server. There may come a time when the file server may start running short of space. In this case, more drives would have to be added to the DAS of the file server, even though there may be unused capacity on the DAS of the email server. Another challenge with the DAS environment is that of data loss. Each server is a single point of failure. If the server goes down, and the backup of the data of the DAS of that server has not been taken, then a huge amount of

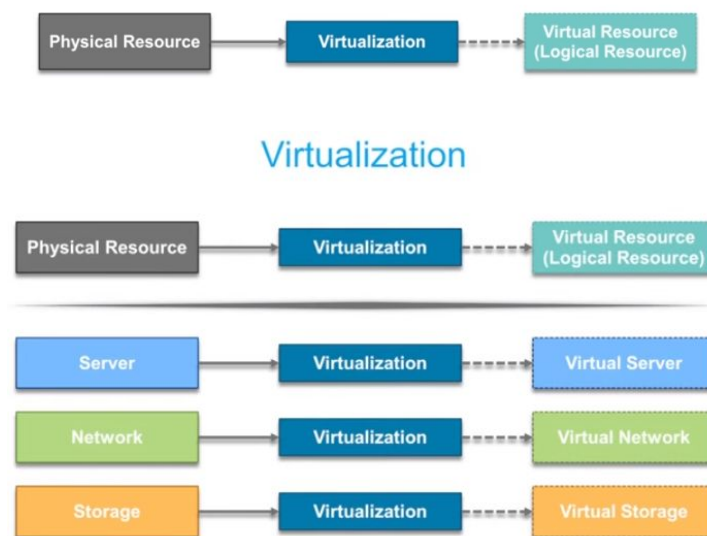
data may be lost. Taking backups of each DAS device in the environment involves a lot of time and effort.

In **networked storage**, a storage device is separated from a server. A server connects to a storage device over a network called as a storage network. In networked storage, a single storage device can be shared between multiple servers. These servers connect to the same storage device over a storage network.

The network may be a dedicated storage area network that uses the FC protocol. Such a network requires special network adapters, switches, and cables, and may be quite expensive to build. On the other hand, the network may be an IP-based network, such as an enterprise Ethernet, which is less expensive to deploy. The data is stored centrally on shared storage and so it is easier to manage and backup.

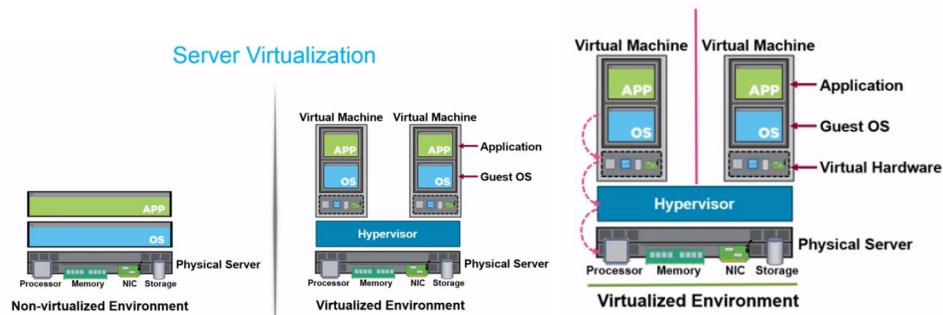
- 1.2. **Fuente.** Storage101x Introduction to Data Storage and Management Technologies
 - 1.2.1. [Virtualization Overview](#)
 - 1.2.2. [Server Virtualization](#)
 - 1.2.3. [Network Virtualization](#)
 - 1.2.4. [Storage Virtualization](#)

Virtualization is the process of using software to abstract a physical device, and create a logical or non-physical version of that physical device. This logical resource or a virtual resource is accessed and used as if it were physical resource. In the context of this course, we will primarily focus on server virtualization, network virtualization, and storage virtualization.

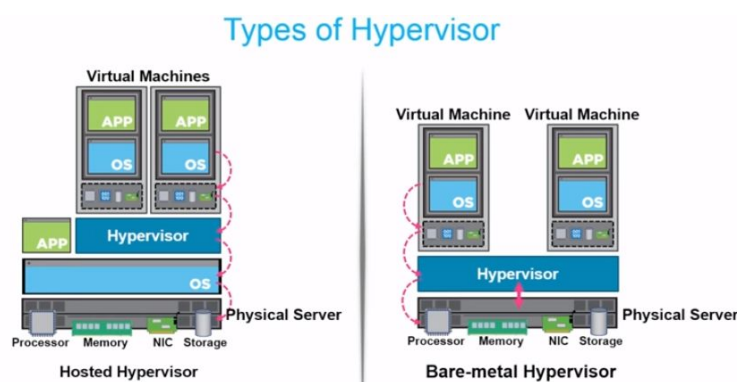


A non-virtualized environment may have considerable capital expenditure as well as operational expenditure. So **what is server virtualization**? It is a virtualization technology that enables multiple operating systems to be installed and run simultaneously on the same physical server. Server virtualization enables the creation of virtual servers on the same physical server. This virtual server is also called a virtual machine or a VM. Each VM can have a different operating system and applications installed on it. Multiple virtual machines can be running at the same time

on the same physical server. In this way it is virtually possible to run multiple operating systems simultaneously on the same physical server....Server virtualization is achieved through a software called hypervisor. The hypervisor is installed on the server hardware and enables a user to create virtual machines.



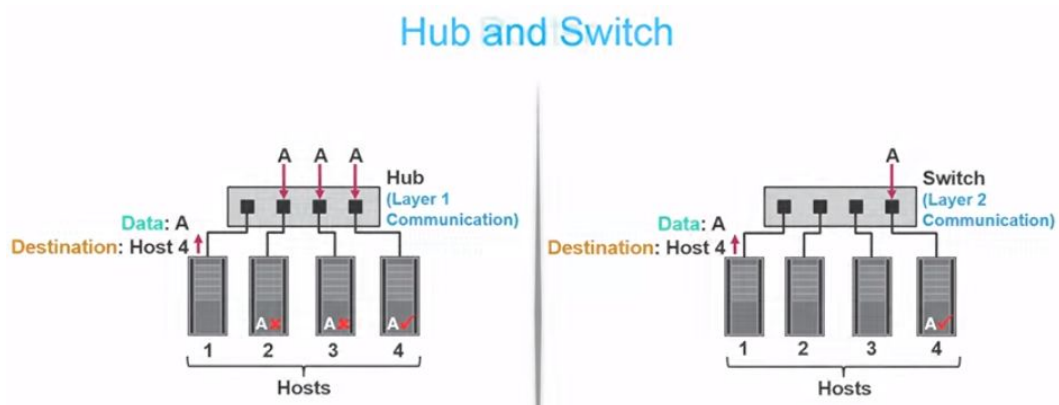
You can install an operating system and applications on a virtual machine. The operating system of a VM is called a guest OS. Each VM has virtual hardware, which appears as physical hardware to the guest OS. The IO requests sent to the virtual hardware from the guest OS are mapped by the hypervisor to the physical hardware in a transparent manner. In this way, all the VMs on the server share the same physical hardware. However the hypervisor ensures that each VM is securely isolated from the other VMs...There are two types of hypervisors – a hosted hypervisor and a bare metal hypervisor. Let’s take a look at a graphic to understand how these two work. In the **hosted hypervisor** environment, an OS is installed on the server hardware, and the hypervisor is installed on top of this OS. There may be other applications running simultaneously on that OS. The hosted hypervisor does not have direct access to the server’s hardware. All I/O requests must go through the OS on which it is installed. A **bare metal hypervisor** is installed directly on server hardware like an operating system. A bare metal hypervisor has direct access to the server’s hardware. The bare metal hypervisor can directly execute the instructions of each guest OS on the server hardware. As a result, **a bare metal hypervisor is more efficient and provides better performance** than a hosted hypervisor. Therefore, bare metal hypervisors are used for server virtualization in enterprise environments.



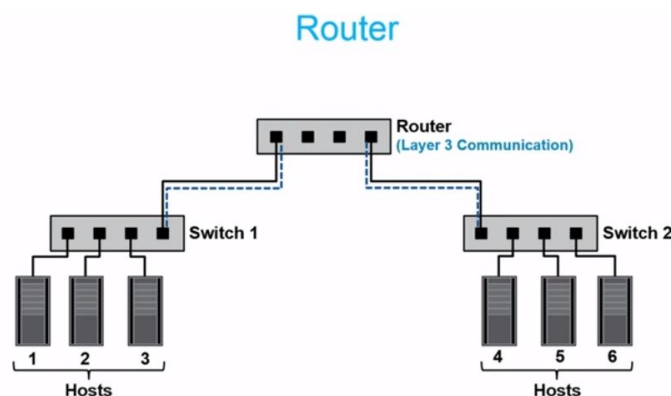
Let’s take look at some of the benefits of **server virtualization**. Server virtualization enables server consolidation and increases the utilization of individual servers. This lowers the capital expenditure that an enterprise has to incur, and also provides better return on investment. Server virtualization also reduces the floor space and power requirements, which in turn reduces the operational expenditure. Server virtualization also provides an agile and flexible environment because the time required to provision a VM is much lower than the time required to provision a

physical machine. Server virtualization also improves availability because VMs can be moved from one server to another without any significant downtime.

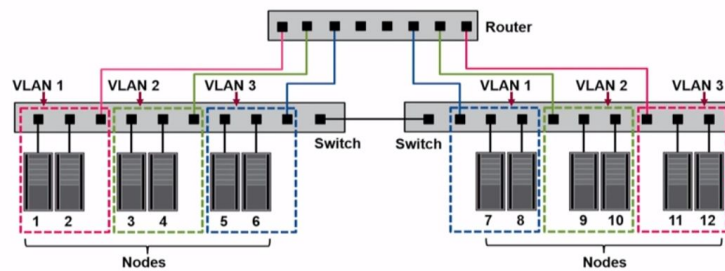
We will cover **network virtualization**. In the same way as server virtualization abstracts physical server hardware, network virtualization abstracts physical network components like switches and routers. This enables the creation of logical or virtualized network components. Network virtualization software is either built into the switch software, or it may be. We will take a look at two types of network virtualization: **virtual LANs or VLANs**, and **VM networks**. So let's take a look at a graphic to understand what is a virtual LAN or a VLAN, and understand how it works. Before we see what a VLAN is, let's see how the basic network components such as hubs and switches work. So suppose there are four hosts that are connected to a hub, and Host 1 wants to send data to Host 4. That data is broadcast to all the hosts that are connected to the same hub. This is known as a *broadcast network*. Host 4, which is the intended recipient of data will be the only host that will retain that data. A switch on the other hand, forwards the data only to the intended host. So in this graphic, we have four hosts connected to the same switch, and Host 1 wants to send data to Host 4. The switch only forwards the data to Host 4 without broadcasting it to the other hosts attached to that switch. **These hosts connected to the switch forms a local area network or LAN.**



A router is used for internetworking. So if you have two physical switches to which hosts are connected, the hosts that are connected to the first switch can communicate to the hosts that are connected to the second switch by connecting the two switches to a single router.



VLAN Example

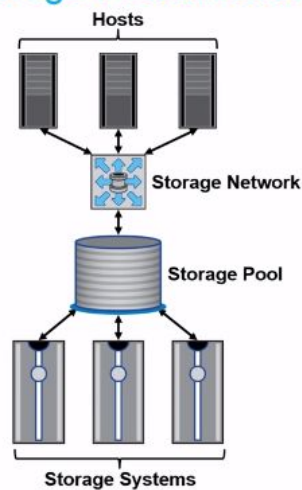


Network virtualization is used to divide a local area network into smaller virtual networks called **VLANs or virtual LANs**. A VLAN is a logical network that is created on a switch by grouping switch ports. A VLAN enables communication between a group of nodes with a common set of functional requirements. In the figure, we have six nodes that are connected to a switch. We can arrange these ports into groups of three, and each group represents a VLAN. So we have VLAN 1, VLAN 2, and VLAN 3. Each VLAN is identified by a unique 12 bit VLAN ID. A node can be made a member of a VLAN by assigning it that particular VLAN ID. Each VLAN behaves like a LAN and is managed independently. Two nodes that belong to the same VLAN can communicate with each other without the routing of frames. So for example, if Node 1 wants to send data to Node 2, that transmission happens within the switch itself. However, two nodes that belong to different VLANs can communicate to each other only through the router even if they are on the same switch. For example, if Node 2 wants to communicate with Node 3 that data passes through the router. This is because Node 2 and Node 3 are on different VLANs.

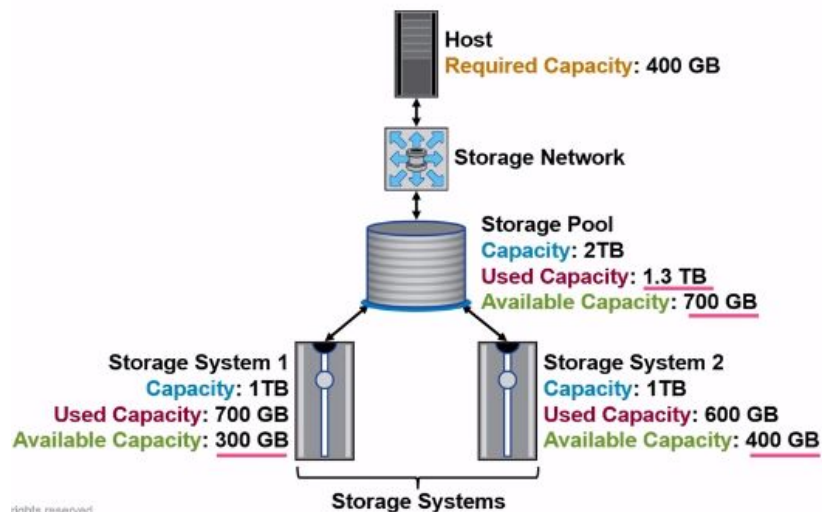
Storage virtualization refers to logically combining the storage capacity from multiple storage devices into a single storage pool. An enterprise may have multiple storage devices and there may be free storage space across these devices. Storage virtualization aggregates this free space from across the devices to create a common storage pool. Users can request and use capacity from the pool without knowing the underlying storage configuration.

Storage systems virtualization can store data from multiple applications. S.V. software can be implemented on the storage system or on separate hosts. In this graphic, we have two storage devices each of 1 terabyte capacity.

Storage Virtualization



Storage Virtualization Example



Using storage virtualization the individual capacities of these two storage devices can be aggregated to create a pool of 2 terabyte capacity. In the first storage device there is 300 GB of free storage space. In the second storage device there is 400 GB of free space. So in the pool there is 1.3 terabyte of used capacity, and 700 GB of free capacity. If a user needs about 400 GB of storage space, storage virtualization helps to allocate 200 GB from each device. The user is not aware that the capacity comes from a pool that is created from multiple storage devices.

Storage virtualization improves storage capacity utilization across multiple storage devices. It simplifies storage management because virtualized storage can be managed from a single administrative console. A storage administrator can also see utilization trends and growth patterns more clearly, and can make better upgrade or capacity planning decisions. Storage virtualization also increases flexibility. When storage space is decoupled from a physical storage device, it is simple to migrate and copy across devices and even geographic locations.

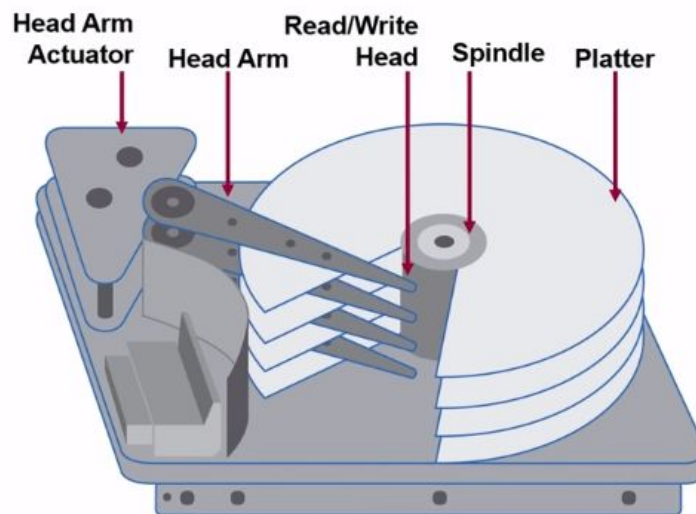
- 1.3. **Fuente.** Storage101x Introduction to Data Storage and Management Technologies
 - 1.3.1. [Types of Data Storage Devices](#)

Types of Storage Devices

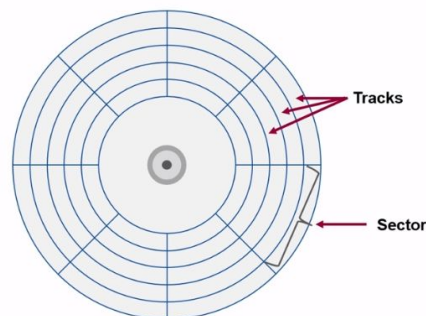
- Magnetic disk drive (Hard disk drive)
- Solid state drive (SSD)
- Optical disc drive
- Magnetic tape drive

A **magnetic disk drive** is an electromechanical storage device. It is also called the hard drive. The disk drive is composed of platters, read/write heads, spindle motor, the actuator assembly, and the drive controller.

Magnetic Disk Drive



Magnetic Disk Drive: Tracks and Sectors



The entire platter is divided into several such circular tracks which form concentric circles running all the way down to the center of the platter. Each track is then divided into blocks called sectors. Data is stored on the platter in the form of blocks, and the sector is the smallest addressable unit of a disk drive. Each sector is typically 512 bytes in size. So when you store data it is stored in the form of 512 byte sized blocks...**General Architecture is presented...**

Now let's take a look at disk drive performance. **There are various parameters on which we can measure disk drive performance.** The first parameter is the drive speed...Drive speed is measured in revolutions per minute or rpm. Some common disk drive speeds are 5,400 rpm, 7,200 rpm, 10,000 rpm, and 15,000 rpm. The second parameter is the seek time... Seek time is

measured in milliseconds. Drives that have lower seek times provide better performance. The next parameter is rotational latency... Drives with faster RPMs have lower rotational latency. The next parameter is the data transfer rate. The data transfer rate represents the number of bytes of data that can be transferred to and from the drive per unit time. This is typically represented in Megabytes per second or Mbps...

In computing, a protocol is a set of commands and rules that enables two entities to communicate...

In storage, there are various protocols that allow a server and a storage device to connect and exchange data. Usually **each protocol also provides its own physical interface specifications**. This affects the type of connectors and the cables that are used to connect the storage drive to a server. There various protocols for connecting a storage drive to a server. Some common protocols are:

Storage Interfaces

- Serial Advanced Technology Attachment (SATA)
- Small Computer System Interface (SCSI)
- Serial Attached SCSI (SAS)
- Nearline SAS (NL-SAS)
- Fibre Channel (FC)

On the storage device, these protocols are implemented on the drive controller. On the server, these protocols are either implemented on the motherboard or by using adapters that plug into the motherboard. These protocols are applicable to both DAS and networked storage environments. **The SATA interface** is commonly found in consumer desktops and laptops. In enterprises, they provide cheap, low performance, and high capacity storage. SATA drives are typically used for data backups and archiving. **The SCSI interface** is popular for enterprise storage. SCSI drives provide parallel transmission and are used for high-performance, mission-critical workloads. SAS is the serial point-to-point variant of the SCSI protocol, and it is also used in high-end computing. **Nearline SAS** or NL-SAS is a hybrid of SAS and SATA interfaces. These drives have a SAS interface and support the SAS protocol. But in the back end they use SATA drives. They cost lesser than SAS drives, and provide the benefit of the SCSI command set. At the same time they offer the large capacities of SATA drives. **The Fibre Channel or FC** protocol is also based on the SCSI protocol. FC is a widely used standard for networked storage. It provides very high throughputs, with the latest standard supporting transfer rates of up to 16 gigabits per second.