云计算与虚拟化技术

第7讲：Storage Devices

阮晓龙

13938213680 / rxl@hactcm.edu.cn

http://cloud.xg.hactcm.edu.cn
http://www.51xueweb.cn

河南中医药大学信息管理与信息系统教研室
信息技术学院网络与信息系统科研工作室

2019.4

讨论提纲

- Storage Design
  - Importance of Storage Design
  - Examining Shared Storage Fundamentals
  - Explaining RAID
  - Understanding vSAN
- Storage Array
  - Fibre Channel
  - Fibre Channel over Ethernet（FCoE）
  - iSCSI
  - Network File System（NFS）
- Case （FC-SAN / iSCSI / NFS）

1.Storage Design

1.1Importance of Storage Design

- Storage design has always been important, but it becomes more so for virtualized infrastructure, for mission-critical applications, and for offerings based on Infrastructure as a Service (IaaS).
- You can probably imagine why this is the case:
  - Advanced Capabilities
    - Many of vSphere's advanced features depend on shared storage; vSphere High Availability (HA), vSphere Distributed Resource Scheduler (DRS), vSphere Fault Tolerance (FT), and some parts of VMware Site Recovery Manager all have critical dependencies on shared storage.

1.Storage Design

- You can probably imagine why this is the case:
  - Performance
    - People understand the benefits that virtualization brings—consolidation, higher utilization, more flexibility, and higher efficiency.
    - But often, people have initial questions about how vSphere can deliver performance for individual applications when it is inherently consolidated and oversubscribed.
    - Likewise, the overall performance of the VMs and the entire vSphere cluster both depend on shared storage, which can also be highly consolidated and oversubscribed.

1.Storage Design

- You can probably imagine why this is the case:
  - Availability
    - The overall availability of your virtualized infrastructure, and by extension, the VMs running on that infrastructure, depend on the shared storage infrastructure.
    - Designing high availability into this infrastructure element is paramount.
    - If the storage is not available, vSphere HA will not be able to recover, and the VMs will be affected.

1.Storage Design

- Although design choices at the server layer can make the vSphere environment relatively more or less optimal, design choices for shared resources such as networking and storage can sometimes make the difference between virtualization success and failure.
  - This is especially true for storage because of its critical role.
  - Storage design choices remain important regardless of whether you are using storage area networks (SANs), which present shared storage as disks or logical units (LUNs); network attached storage (NAS), which presents shared storage as remotely accessed file systems; or a converged storage infrastructure using local server disks such as vSAN.
  - You can create a shared storage design that lowers the cost and increases the efficiency, performance, availability, and flexibility of your vSphere environment.

7

1.Storage Design

1.1Importance of Storage Design

You have to be fully aware of

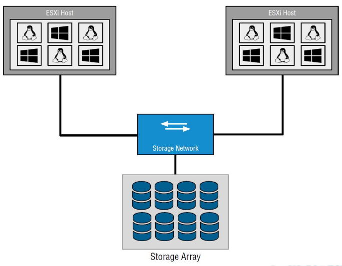
The Importance of Properly Designed Storage

8

1.Storage Design

1.2Examining Shared Storage Fundamentals

Figure 6.1
When ESXi hosts are connected to that same shared storage, they share its capabilities.



9

1.Storage Design

1.2Examining Shared Storage Fundamentals

- Several basics of storage:
  - Local storage versus shared storage
  - Common storage array architectures
  - RAID technologies
  - Midrange and enterprise storage array design
  - Protocol choices

1.Storage Design

- An ESXi host can have one or more storage options actively configured, including the following:
  - Local SAS/SATA/SCSI storage
  - Fibre Channel
  - Fibre Channel over Ethernet (FCoE)
  - iSCSI using software and hardware initiators
  - NAS (specifically, NFS)
  - InfiniBand

1.Storage Design

- Traditionally, local storage has been used in a limited fashion with vSphere.
  - because so many of vSphere's advanced features—such as vSphere HA, vSphere DRS, and vSphere FT—required shared external storage.
  - With vSphere Auto Deploy and the ability to deploy ESXi images directly to RAM at boot time, coupled with Host Profiles to automate the configuration of the ESXi Host, in some environments local storage isn't necessary.

1.Storage Design

- No Local Storage? No Problem!
  - What if you don't have local storage? (Perhaps you have a diskless blade system, for example.)
  - There are many options for diskless systems, including booting from Fibre Channel/iSCSI SAN and network-based boot methods like vSphere Auto Deploy.
  - There is also the option of using USB or SD Card boot, a technique that we've employed on numerous occasions.
  - Both Auto Deploy and USB boot give you some flexibility in quickly reprovisioning hardware or deploying updated versions of vSphere.

1.Storage Design

- Shared storage is the basis for most vSphere environments because it supports the VMs themselves and because it is a requirement for many of vSphere's features.
  - Shared external storage in SAN configurations (which encompasses Fibre Channel, FCoE, and iSCSI) and NAS (NFS) is always highly consolidated.
  - This makes it efficient.
    - Similar to the benefits of physical-tovirtual consolidation with regard to CPU and memory, SAN/NAS or vSAN can take the direct attached storage in physical servers that are 10% utilized and consolidate them to eighty 80% utilization.

1.Storage Design

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - External Connectivity
    - The external (physical) connectivity between the traditional storage array and the hosts (in this case, the ESXi hosts) is generally Fibre Channel or Ethernet, though InfiniBand and other rare protocols exist. The characteristics of this connectivity define the maximum bandwidth (given no other constraints, and there usually are other constraints) of the communication between the ESXi host and the shared storage array.
    - External connectivity is typically referred to as front-end or FE connectivity and most often tied to a fabric for distributed sharing and scalability purposes.

1.Storage Design

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - Storage Processors
    - Different vendors have different names for storage processors, which are considered the brains of the array.
    - They handle the I/O and run the array software. In most modern arrays, the storage processors are not purpose-built application-specific integrated circuits (ASICs) but instead are general-purpose x86 CPUs.
    - Some arrays use PowerPC, some use specific ASICs, and some use custom ASICs for specific purposes.
    - But in general, if you cracked open an array, you would most likely find an Intel or AMD CPU.

1.Storage Design

1.2Examining Shared Storage Fundamentals

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - Array Software
    - Although hardware specifications are important and can define the scaling limits of the array, just as important are the functional capabilities the array software provides.
    - The capabilities of modern storage arrays are vast, similar in scope to vSphere itself, and vary wildly among vendors.

1.Storage Design

1.2Examining Shared Storage Fundamentals

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - Cache Memory
    - Every array differs as to how cache memory is implemented, but all have some degree of nonvolatile memory used for various caching functions, delivering lower latency and higher IOPS throughput by buffering I/O using write caches and storing commonly read data to deliver a faster response time using read caches.
    - Nonvolatility (meaning the ability to survive a power loss) is critical for write caches because the data is not yet committed to disk, but it's not critical for read caches. Cached performance is often used when describing shared storage array performance maximums (in IOPS, MBps, or latency) in specification sheets.

1.Storage Design

1.2Examining Shared Storage Fundamentals

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - Cache Memory
    - These results generally do not reflect real-world scenarios. In most real-world scenarios, performance tends to be dominated by the disk performance (the type and number of disks) and is helped by write caches in most cases, but only marginally by read caches (with the exception of large relational database management systems, which depend heavily on read-ahead cache algorithms).
    - One vSphere use case helped by read caches is a situation where many boot images are stored only once (through the use of vSphere or storage array technology), but this is also a small subset of the overall VM I/O pattern.

1.Storage Design

1.2Examining Shared Storage Fundamentals

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - Disks
    - Arrays differ as to which type of disks (often called spindles) they support and how many they can scale to support.
    - Drive capabilities are defined by a number of attributes.
      - First, drives are often separated by the drive interface they use: Fibre Channel, serial-attached SCSI (SAS), and serial ATA (SATA). With the exception of enterprise flash drives (EFDs), drives are typically described by their rotational speed, noted in revolutions per minute (RPM). Fibre Channel drives typically come in 15K RPM and 10K RPM variants, SATA drives are usually found in 5400 RPM and 7200 RPM variants, and SAS drives are usually 15K RPM or 10K RPM variants.

1.Storage Design

1.2Examining Shared Storage Fundamentals

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - Disks
    - Arrays differ as to which type of disks (often called spindles) they support and how many they can scale to support.
    - Drive capabilities are defined by a number of attributes.
      - Second, EFDs, which are now mainstream, are solid state and have no moving parts; therefore, rotational speed does not apply (nor does the name spindle). The type and number of disks are very important. Coupled with how they are configured, this determines how a storage object (either a LUN for a block device or a file system for a NAS device) performs. Shared storage vendors generally use disks from the same disk vendors, so this is an area of commonality across shared storage vendors.

1.Storage Design

1.2Examining Shared Storage Fundamentals

- The elements that make up a shared storage array consist of external connectivity, storage processors, array software, cache memory, disks, and bandwidth:
  - Disks
    - The following list is a quick reference on what to expect under a random read/write workload from a given disk drive:
      - 7,200 RPM SATA: 80 IOPS
      - 10K RPM SATA/SAS/Fibre Channel: 120 IOPS
      - 15K RPM SAS/Fibre Channel: 180 IOPS
      - Commercial solid-state drives (SSD) based on Multi-Level Cell (MLC) technology: 1,000–100,000s IOPS
      - Enterprise flash drives (EFD) based on Single-Level Cell (SLC) technology and much deeper, very high-speed memory buffers: 6,000–100,000s IOPS

1.Storage Design

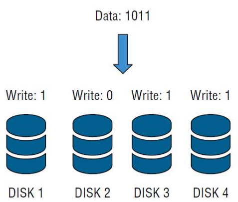
- RAID is used to increase data availability and to scale performance beyond that of a single drive. Every array implements various RAID schemes.
  - RAID schemes address this by leveraging multiple disks together and using copies of data to support I/O until the drive can be replaced, and the RAID protection can be rebuilt.
  - Each RAID configuration tends to have different performance characteristics and different capacity overhead impact.
  - We recommend that you view RAID choices as a significant factor in your design.

1.Storage Design

- RAID 0
  - This RAID level offers no redundancy and no protection against drive failure.
  - In fact, it has a higher aggregate risk than a single disk because any single disk failing affects the whole RAID group.
  - Data is spread across all the disks in the RAID group, which is often called a stripe.
  - This level delivers fast performance, but it is the only RAID type that is usually not appropriate for any production vSphere use because of the availability profile.

1.Storage Design

Figure 6.2
In a RAID 0 configuration, the data is striped across all the disks in the RAID set, providing very good performance but very poor availability.

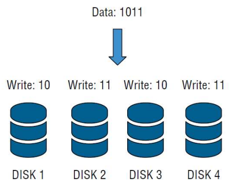


1.Storage Design

- RAID 1, 1+0, 0+1
  - These mirrored RAID levels offer high degrees of protection but at the cost of 50% loss of usable capacity. This is versus the raw aggregate capacity of the sum of the capacity of the drives.
  - RAID 1 simply writes every I/O to two (or more) drives and can balance reads across all drives (because there are multiple copies). This can be coupled with RAID 0 to form RAID 1+0 (or RAID 10), which mirrors a stripe set, or to form RAID 0+1, which stripes data across pairs of mirrors.
  - This has the benefit of being able to withstand multiple drives failing, but only if the drives fail on different elements of a stripe on different mirrors, thus making RAID 1+0 more fault tolerant than RAID 0+1. The other benefit of a mirrored RAID configuration is that, in the case of a failed drive, rebuild times can be very rapid, which shortens periods of exposure.

1.Storage Design

Figure 6.3
This RAID 10 2+2 configuration provides good performance and good availability, but at the cost of 50% of the usable capacity.



1.Storage Design

- Parity RAID (RAID 5, RAID 6)
  - These RAID levels use a mathematical calculation (an XOR parity calculation) to represent the data across several drives. This tends to be a good compromise between the availability of RAID 1 and the capacity efficiency of RAID 0.
  - RAID 5 calculates the parity across the drives in the set and writes the parity to another drive. This parity block calculation with RAID 5 is rotated among the disks in the RAID 5 set.

28

1.Storage Design

1.3Explaining RAID

- Parity RAID (RAID 5, RAID 6)
  - Parity RAID schemes can deliver very good performance, but there is always some degree of write penalty. For a full-stripe write, the only penalty is the parity calculation and the parity write, but in a partial-stripe write, the old block contents must be read, a new parity calculation must be made, and all the blocks must be updated. However, generally modern arrays have various methods to minimize this effect. Read performance, on the other hand, is generally excellent because a larger number of drives can be read from than with mirrored RAID schemes. RAID 5 nomenclature refers to the number of drives in the RAID group.

29

1.Storage Design

1.3Explaining RAID

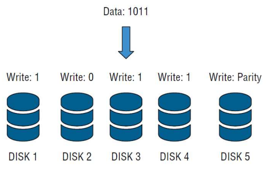
- Parity RAID (RAID 5, RAID 6)
  - RAID 5 can be coupled with stripes, so RAID 50 is a pair of RAID 5 sets with data striped across them. When a drive fails in a RAID 5 set, I/O can be fulfilled using the remaining drives and the parity drive, and when the failed drive is replaced, the data can be reconstructed using the remaining data and parity.

30

1.Storage Design

1.3Explaining RAID

Figure 6.4
A RAID 5 4+1 configuration offers a balance between performance and efficiency.

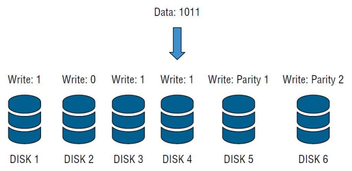


31

1.Storage Design

1.3Explaining RAID

Figure 6.5
A RAID 6 4+2 configuration offers protection against double drive failures.



32

1.Storage Design

1.4Understanding vSAN

- vSphere 5.5 introduced a brand-new storage feature, Virtual SAN, or simply vSAN.
  - At a high level, vSAN pools the locally attached storage from members of a vSAN–enabled cluster and presents the aggregated pool back to all hosts within the cluster. This could be considered an "array" of sorts because just like a normal SAN, it has multiple disks presented to multiple hosts.
  - vSAN does not require any additional software installations. It is built directly into ESXi itself.
  - Managed from vCenter Server, vSAN is compatible with all the other cluster features that vSphere offers, such as vMotion, HA, and DRS, even use Storage vMotion to migrate VMs on or off a vSAN datastore.

河南中医药大学 / 阮晓龙 / 13938213680 / http://cloud.xg.hactcm.edu.cn / http://www.51xueweb.cn

33

1.Storage Design

1.4Understanding vSAN

- vSAN uses the disks directly attached to the ESXi hosts and is simple to set up, but there are a few specific requirements.
- Listed here is what you'll need to get VSAN up and running:
  - ESXi 5.5 or newer hosts
  - vCenter Server 5.5 or newer
  - One or more SSDs per host
  - One or more HDDs per host for hybrid mode
  - Storage controllers must be on the vSAN HCL
  - Minimum of three hosts per vSAN cluster
  - Maximum of 64 hosts per vSAN cluster
  - 1 Gbps network between hosts (10 Gbps highly recommended and required for allflash vSAN)

河南中医药大学 / 阮晓龙 / 13938213680 / http://cloud.xg.hactcm.edu.cn / http://www.51xueweb.cn

1.Storage Design

1.4Understanding vSAN

- There are two types of vSAN configurations.
  - The first is an all flash-based configuration that provides vSAN clusters with the highest performance available for both data in cache as well as data at rest. The original configuration that was introduced with vSphere 5.5 is a "hybrid" approach. It uses both flash-based and magnetic hard disks.
  - vSAN requires at least one flash-based device in each host. Hybrid vSAN uses the flash tier as a read and write cache just as some external SANs do. When blocks are written to the underlying datastore, they are written to the flash tier first, and later the data can be relocated to the HDDs (capacity tier) if it's not considered to be frequently accessed.
  - vSAN's read/write cache ratio is 70% read, 30% write.

1.Storage Design

1.4Understanding vSAN

- vSAN doesn't use the traditional RAID concepts explained, it uses what VMware is calling RAIN, or reliable array of independent nodes.
  - vSAN uses a combination of vSphere APIs for Storage Awareness (VASA) and storage policies to ensure that VMs are located on more than one disk and/or host to achieve their performance and availability requirements. This is why VMware recommends 10 Gbps networking between ESXi hosts when using vSAN.
  - A VM virtual disk could be located on one physical host but could be running on another host's CPU and memory. The storage system is fully abstracted from the compute resources. In all likelihood the VMs virtual disk files could be located on multiple hosts in the cluster to ensure a level of redundancy.

2.Storage Array

2.1Fibre Channel

- vSphere offers several choices for shared storage protocol, including Fibre Channel, Fibre Channel over Ethernet/FCoE/, iSCSI, and Network File System (NFS), which is a form of NAS.
- SANs are most commonly associated with Fibre Channel storage because Fibre Channel was the first widely adopted protocol used with SANs. However, SAN refers to a network topology, not a connection protocol. In fact, SAN refers to the ability to create block storage access through the use of a network, and although people often use the acronym SAN to refer to a Fibre Channel SAN, you can create a SAN topology using different types of protocols, including iSCSI, FCoE, and InfiniBand.

2.Storage Array

- SANs were initially deployed to aggregate storage inside a datacenter while maintaining some of the characteristics of local or direct attached SCSI devices. A SAN is a network where storage devices (logical units, or LUNs, just as on a SCSI or SAS controller) are presented from a storage target (one or more ports on an array) to one or more initiators.
- An initiator can come in both hardware and software forms. Hardware adapters, such as host bus adapters (HBA) for Fibre Channel and iSCSI, or converged network adapters (CNA), for iSCSI and FCoE are common, though software-based initiators are available for iSCSI and FCoE as well.
- Today, Fibre Channel HBAs have roughly the same cost as high-end multiport Ethernet interfaces or local SAS controllers, and (depending on the type) the per-port cost of a Fibre Channel switch is about twice that of a high-end managed Ethernet switch.

2.Storage Array

- Fibre Channel typically uses an optical interconnect (though there are copper variants) because the Fibre Channel protocol assumes a very high-bandwidth, low-latency, and lossless physical layer. Standard Fibre Channel HBAs today support very-high-throughput, 4 Gbps, 8Gbps, 16 Gbps, and 32 Gbps connectivity in single-, dual-, or quad-port options.
- For end-to-end compatibility (in other words, from host to HBA to switch to array), every storage vendor maintains a similar compatibility matrix. From a connectivity standpoint, almost all cases use a common OM2 (orange-colored cables) multimode duplex LC/LC cable. The newer OM3 and OM4 (aqua-colored cables) are used for longer distances and are generally used for 10 Gbps Ethernet and 8/16 Gbps Fibre Channel. Different optical transceivers have different distance tolerances and using the wrong transceiver with the inappropriate cable can result in unpredictable storage networking performance.

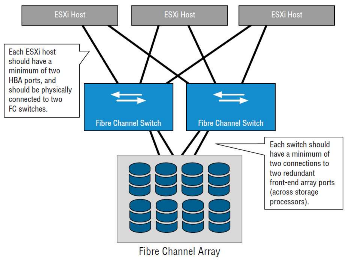
2.Storage Array

- The Fibre Channel protocol can operate in three modes: point-to-point (FC-P2P), arbitrated loop (FC-AL), and switched (FC-SW). Point-to-point and arbitrated loop are rarely used today, though they may have specific use cases. FC-AL is commonly used by some array architectures to connect their backend spindle enclosures, but the protocol is more generally used to connect to tape-based backup devices. Most modern arrays use switched fabric designs, which have higher bandwidth per disk enclosure and greater deployment flexibility.
- Best practice for block-based storage systems is to have equal and redundant systems for purposes of high availability (HA). "SAN A/B" design is common and often expected in storage environments. each ESXi host has a minimum of two HBA ports, and each is physically connected to two Fibre Channel switches. Each switch has a minimum of two connections to two redundant front-end array ports.

2.Storage Array

2.1Fibre Channel

Figure 6.8
The most common Fibre Channel configuration—a switched Fibre Channel (FC-SW) SAN. This enables the Fibre Channel LUN to be easily presented to all the hosts while creating a redundant network design.



2.Storage Array

2.1Fibre Channel

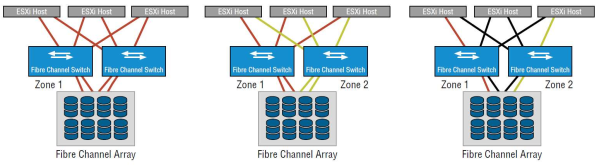


Figure 6.10
There are many ways to configure zoning. From left to right: multi-initiator/multi-target zoning, single-initiator/multi-target zoning, and single-initiator/single-target zoning.

2.Storage Array

2.2Fibre Channel over Ethernet

- Fibre Channel, as a protocol, is organized into different parts so that it is decoupled from the lower-level physical layer. the Fibre Channel standard makes provisions to run the protocol over different transportation types, including Layer 3/4 TCP/IP, Layer 2 Ethernet, pseudowire, and other transportation mechanisms. These backbone changes all fall under the purview of the FC-BB standards.
- Fibre Channel is designed to guarantee in-order delivery and, as implemented in datacenters today, requires a lossless, low-jitter, high-bandwidth physical layer connection. To ensure that the same type of performance can be achieved using Ethernet, which is traditionally a lossy medium and more forgiving of errors on the wire, additional considerations were required on the Ethernet side.

2.Storage Array

2.2Fibre Channel over Ethernet

- To address this need, the IEEE created a series of standards that enhance traffic delivery, the result of which makes a perfect combination for running lossless FCoE traffic simultaneously with lossy LAN traffic. Three key standards, all part of the Datacenter Bridging (DCB) effort, make this possible:
  - Priority Flow Control (PFC, also called Per-Priority Pause)
  - Enhanced Transmission Selection (ETS)
  - Datacenter Bridging Exchange (DCBX)

2.Storage Array

2.2Fibre Channel over Ethernet

- There is an additional standard in Ethernet called IEEE 802.1pp that allows a link between devices to be separated into eight classes of service (CoS) values, called priorities. The term priority is somewhat of a misnomer as the term does not refer to the importance of the traffic, but rather the class that the traffic belongs to. This becomes the foundation for multiprotocol traffic because it permits users to place traffic on specific priorities, each having its own specific behavioral characteristics.
- Priority Flow Control (IEEE 802.1Qbb) is the standard that creates the lossless behavior on a specific priority without affecting other traffic on other CoSs (priorities). When using a lossless no drop priority, it is possible to isolate FCoE traffic and maintain in-order delivery through the use of judicious PAUSE frames, which pause the traffic until such time that it can be delivered with the frames in order.

2.Storage Array

2.2Fibre Channel over Ethernet

- ETS and DCBX are part of the same standard document (IEEE 802.1Qaz) and refer to two specific capabilities.
  - First, ETS provides minimum bandwidth requirements for traffic groups. In the most common deployment of multiprotocol traffic, FCoE is given 50% of bandwidth and the remaining LAN traffic is given the other 50%. However, these are minimum guarantees, which means that each type of traffic is guaranteed to have at least 50% of the available bandwidth. If, on the other hand, FCoE traffic is not currently using all of its available bandwidth, the LAN traffic can use whatever additional capacity is available. But when FCoE needs its bandwidth back, it gets it, at least to the 50% setting.

2.Storage Array

- ETS and DCBX are part of the same standard document (IEEE 802.1Qaz) and refer to two specific capabilities.
  - The second part, DCBX, is simply an extension of the Link Layer Discovery Protocol (LLDP), which permits settings to be exchanged between devices. For example, when a CNA comes online, it can receive its settings (including ETS, FCoE settings, and so forth) from the switch using the DCBX protocol. Used together, these three protocols allow Fibre Channel frames to be transported in a lossless fashion, independent of lossy traffic being transported along the same wire at the same time.

2.Storage Array

2.2Fibre Channel over Ethernet

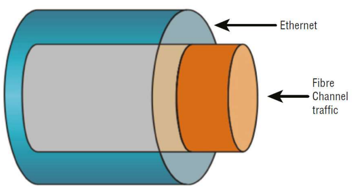
- ETS and DCBX are part of the same standard document (IEEE 802.1Qaz) and refer to two specific capabilities.
  - The second part, DCBX, is simply an extension of the Link Layer Discovery Protocol (LLDP), which permits settings to be exchanged between devices. For example, when a CNA comes online, it can receive its settings (including ETS, FCoE settings, and so forth) from the switch using the DCBX protocol. Used together, these three protocols allow Fibre Channel frames to be transported in a lossless fashion, independent of lossy traffic being transported along the same wire at the same time.

河南中医药大学 / 阮晓龙 / 13938213680 / http://cloud.xg.hactcm.edu.cn / http://www.51xueweb.cn

2.Storage Array

2.2Fibre Channel over Ethernet

Figure 6.11
FCoE encapsulates Fibre Channel frames into Ethernet frames for transmission over a lossless Ethernet transport.



河南中医药大学 / 阮晓龙 / 13938213680 / http://cloud.xg.hactcm.edu.cn / http://www.51xueweb.cn

2.Storage Array

- FCoE uses whatever physical cable plant that 10 Gb Ethernet uses.
  - 10 GbE connectivity varies between optical (same cables as Fibre Channel),
  - Twinax (which is a pair of coaxial copper cables),
  - InfiniBand-like CX cables,
  - 10 Gb shielded twisted pair (UTP) use cases via the 10GBase-T standard.
- Each has its specific distance-based use cases and varying interface cost, size, and power consumption.

2.Storage Array

- Be careful not to let cost be the deciding factor in choosing appropriate physical layer connectivity. It is all too easy to mismatch transceivers and cabling simply because they “fit” together.
  - FCoE has more stringent requirements for bit error rates (BERs), the use of 10GBase-T is particularly tricky.
  - FCoE, on the other hand, requires tighter controls, and so it’s important to note that only Cat 6a (not just Cat 6) and Cat 7 are supported for FCoE traffic.
  - Because of the higher resistance of copper (compared to optical cabling) the supported distance winds up being around 30 meters.

2.Storage Array

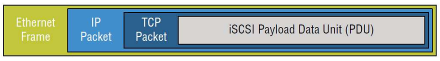
- iSCSI brings the idea of a block storage SAN to customers with no Fibre Channel infrastructure.
  - iSCSI is an Internet Engineering Task Force (IETF) standard for encapsulating SCSI control and data in TCP/IP packets, which in turn are encapsulated in Ethernet frames.
  - TCP retransmission is used to handle dropped Ethernet frames or significant transmission errors. Storage traffic can be intense relative to most LAN traffic. This makes it important that you minimize retransmits, minimize dropped frames, and ensure that you have a bet-the-business Ethernet infrastructure when using iSCSI.

52

2.Storage Array

2.3iSCSI

Figure 6.12
Using iSCSI, SCSI control and data are encapsulated in both TCP/IP and Ethernet frames.

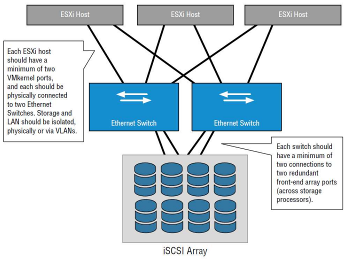


53

2.Storage Array

2.3iSCSI

Figure 6.13
Notice how the topology of an iSCSI SAN is the same as a switched Fibre Channel SAN.



54

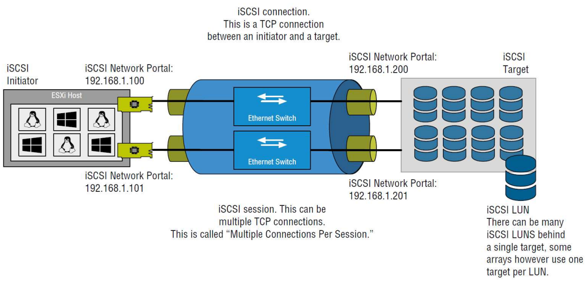
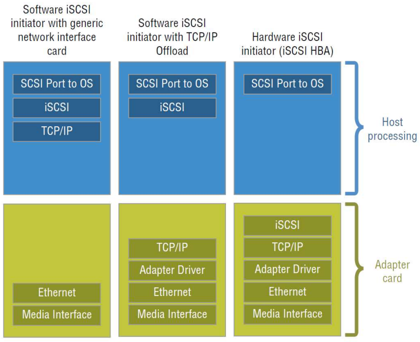


Figure 6.14
The iSCSI IETF standard has several different elements.

Figure 6.15
Some parts of the stack are handled by the adapter card versus the ESXi host CPU in various implementations.



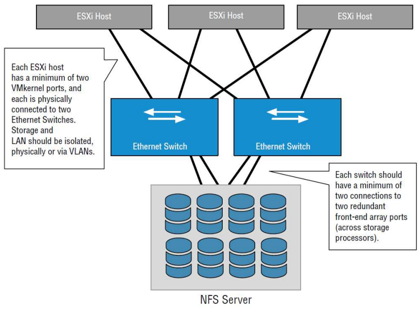
2.Storage Array

- NFS protocol is a standard originally developed by Sun Microsystems to enable remote systems to access a file system on another host as if it were locally attached. vSphere 6.0 and later implements a client compliant with both NFSv3 and NFS v4.1 using TCP.
- When NFS datastores are used by vSphere, no local file system (such as VMFS) is used. The file system is on the remote NFS server. This means that NFS datastores need to handle the same access control and file-locking requirements that vSphere delivers on block storage using the vSphere Virtual Machine File System, or VMFS. NFS servers accomplish this through NFS file locks.

2.Storage Array

- The movement of the file system from the ESXi host to the NFS server also means that you don't need to handle zoning or masking tasks. This makes an NFS datastore one of the easiest storage options to simply get up and running.
- On the other hand, it also means that all of the high availability and multipathing functionality that is normally part of a Fibre Channel, FCoE, or iSCSI storage stack is replaced by the networking stack.

Figure 6.16
The topology of an NFS configuration is similar to iSCSI from a connectivity standpoint but very different from a configuration standpoint.



2.Storage Array

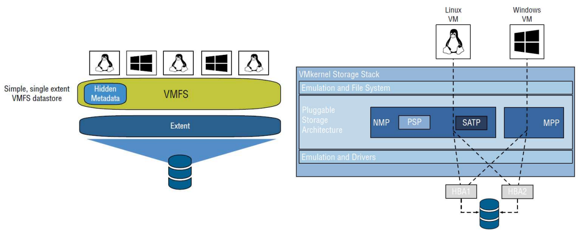
2.5Making Basic Storage Choices

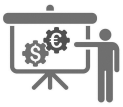
Table 6.1: Shared storage choices

Feature	Fibre Channel SAN	iSCSI SAN	NFS	vSAN
ESXi boot (boot from shared storage)	Yes	Hardware initiator or software initiator with iBFT support	No	No
VM boot	Yes	Yes	Yes	Yes
Raw device mapping	Yes	Yes	No	No
Dynamic extension	Yes	Yes	Yes	Yes
Availability and scaling model	Storage stack (PSA), ESXi LUN queues, array configuration	Storage stack (PSA), ESXi LUN queues, array configuration	Network Stage (NIC teaming and routing), network and NFS server configuration	Storage stack (local), Network Stage (NIC teaming and routing)
VMware feature support (vSphere HA, vMotion, Storage vMotion, vSphere FT)	Yes	Yes	Yes	Yes

3.Case

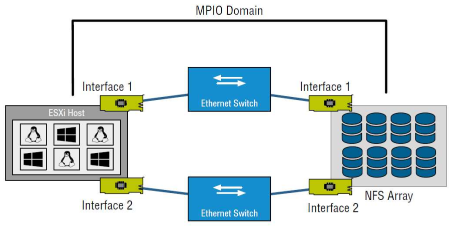
3.1FC-SAN



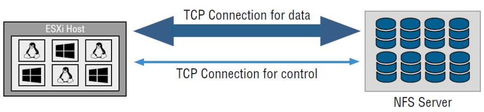


- 演示
  - 基于Windows Server 2016实现iSCSI
  - VMware ESXi访问iSCSI
  - VCSA实现iSCSI共享存储

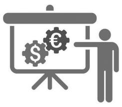
3.Case



3.Case



64



- 演示
  - 基于Windows Server 2016实现NFS
  - VMware ESXi访问NFS
  - VCSA实现NFS共享存储

65