Disk Utilization with DAS

- If a server requires 100GB of storage for its OS and apps when deployed, and is expected to grow to 300GB of storage capacity, it's typical to install 500GB of disk space
- You want to leave some overhead for unexpected growth, and it might not be possible to add capacity later without bringing the server offline
- You'll often get utilization of around 30% when using Direct Attached Storage

FLACKBOX
www.flackbox.com

Disk Utilization with Centralised Storage

- Storage systems provide a centralized 'pool' of shared storage
- Devices and applications can be allocated storage as required and on enterprise class storage the size can be easily changed on the fly, non-disruptively
- Centralised storage can provide disk utilization closer to 80%

FLACKBOX
www.flackbox.com

Disk Utilization – Thin Provisioning

- Thin provisioning allows you to make it appear to the servers that they have more storage than you have actually paid for
- What you buy with centralized storage: 50 x 200GB = 10TB total
- The servers each believe they have 500GB space
- It looks like there is 25TB storage, but you only paid for 10TB

FLACKBOX
www.flackbox.com

Disk Utilization – Thin Provisioning

- The 10TB shared space is used first-come first-served
- You can add the additional space as and when needed, transparently to the servers

FLACKBOX
www.flackbox.com

Disk Utilization – Deduplication and Compression

- Deduplication detects and eliminates identical blocks
- The eliminated blocks are replaced with a pointer to a single copy of the block on disk
- Compression detects and eliminates redundant data and white space in files
- Huge space savings are possible, dependent on the amount of duplicate blocks and compressible files you have

FLACKBOX
www.flackbox.com

Disk Utilization – Centralised Storage

- Pooled storage and thin provisioning moves from a 'Just In Case' to a 'Just In Time' model of purchasing storage space
- Deduplication and compression provide additional space savings
- This provides cost savings on hardware, rack space, power and cooling, and the savings are multiplied as storage costs tends to come down over time

FLACKBOX
www.flackbox.com

Performance and Capacity

- Accessing data across a network rather than in DAS can add latency, but…
- Data can be striped across many disks in an enterprise class storage system
- Storage vendors are also at the cutting edge of new storage technologies

FLACKBOX
www.flackbox.com

Resiliency

- SAN and NAS storage systems are always built to have very high degrees of resiliency because they will almost always be mission critical systems for the enterprise
- If any single component fails there is a redundant component to take its place

FLACKBOX
www.flackbox.com

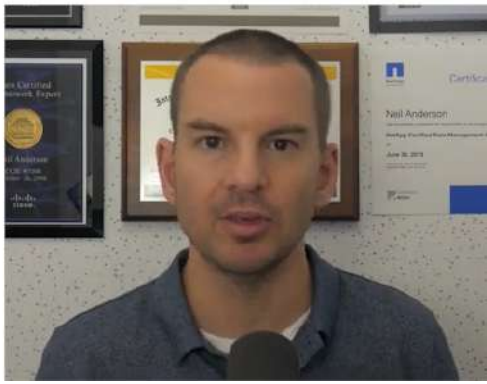
Benefits - Centralized Management

- It's much easier to manage all your storage from a centralised location, rather than separately across all your servers

FLACKBOX
www.flackbox.com

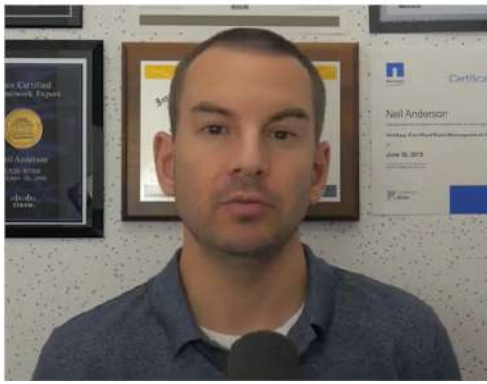
Diskless Servers

- SAN protocols allow clients to boot up from a logical disk on the remote storage
- The clients don't need to contain any disks
- This is a very popular option with blade servers



Storage Tiering

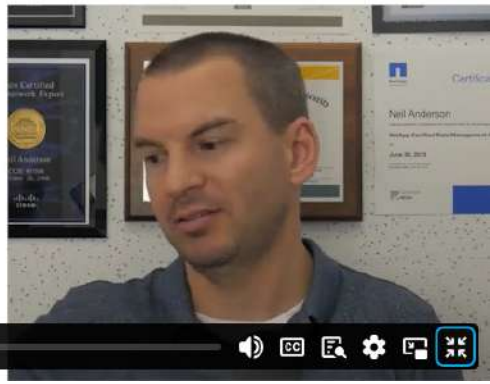
- The storage system can have media with differing attributes, such as high performance SSD drives, and high capacity (but lower performance) SATA drives
- Frequently accessed ‘hot’ data can be kept on the SSDs, and ‘cold’ data automatically archived onto the SATA drives



FLACKBOX
www.flackbox.com

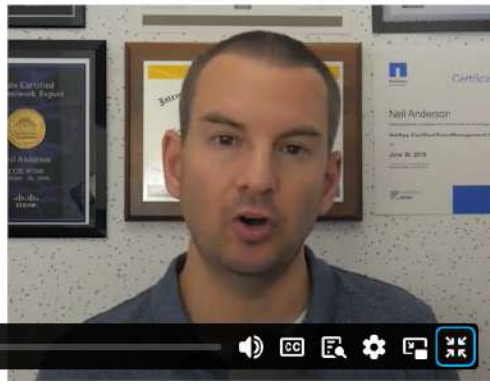
Centralized Backup

- Managing backups is very inconvenient and time consuming when you have 50 different tape drives on 50 different servers
- Centralized storage provides a centralized backup solution
- Storage systems can also backup to remote disk (rather than tape) which reduces backup windows and doesn't require loading/unloading of physical media



Snapshots

- Snapshots are a point in time copy of the file system which can be used as a convenient short term backup
- They consist of pointers to the original blocks on disk rather than being a new copy of the data, so they initially take up no space and occur nearly instantaneously



Snapshots

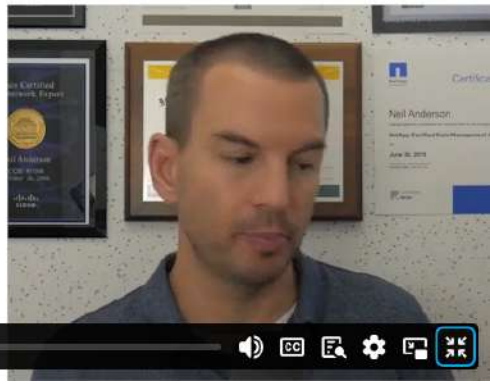
- If data gets corrupted or somebody accidentally deletes a file then you can very quickly recover from a snapshot
- Snapshots do not replace a long term offsite backup solution, but they're great for short term very quick and convenient backups and restores

FLACKBOX
www.flackbox.com

15:09

Disaster Recovery

- You can replicate data to a disaster recovery site
- You can also load balance incoming client requests for read-only data between the different sites. (You can't do this for writable data as you need to maintain one consistent copy of the data)



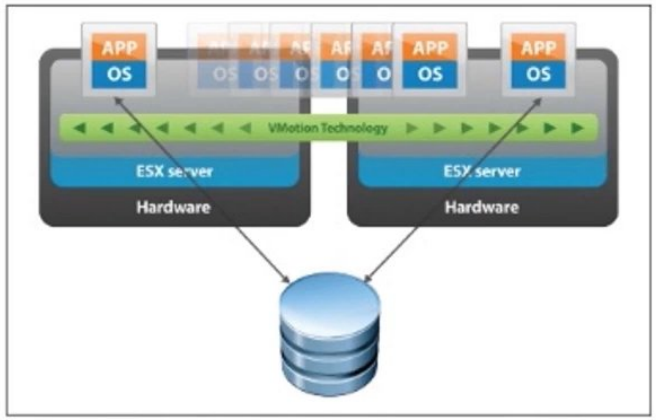
Virtualization Support - vMotion

- Software such as VMware and Hyper-V allows you to run multiple virtual servers on the same underlying physical hardware server.
- You can have a Linux web server, Exchange mail server and SQL database server for example all running on the same physical box, transparently to each of those virtual servers.

FLACKBOX
www.flackbox.com

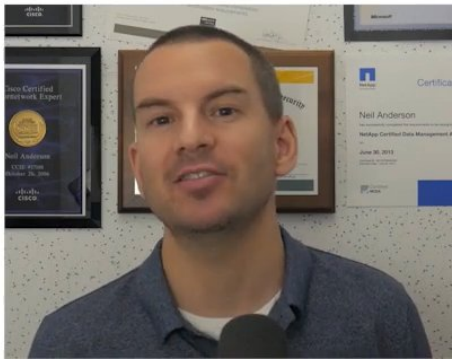
Virtualization Support - vMotion

- The killer feature of virtualisation software is the ability to move virtual servers between physical servers on the fly while they are still running.
- The virtual servers can keep on running with no outages even if the underlying physical server fails or is taken down for maintenance.
- External storage is a requirement for this feature.



HDD and SSD

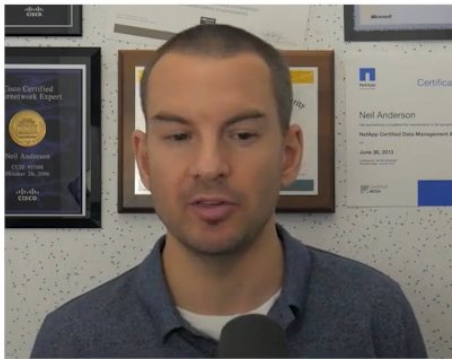
- HDD – Hard Disk Drive.
  - Has a rotating platter.
- SSD – Solid State Drive.
  - Newer, faster technology with no moving parts. Also known as ‘Flash’ in enterprise storage environments.



FLACKBOX
www.flackbox.com

SSD, SAS and SATA Drives

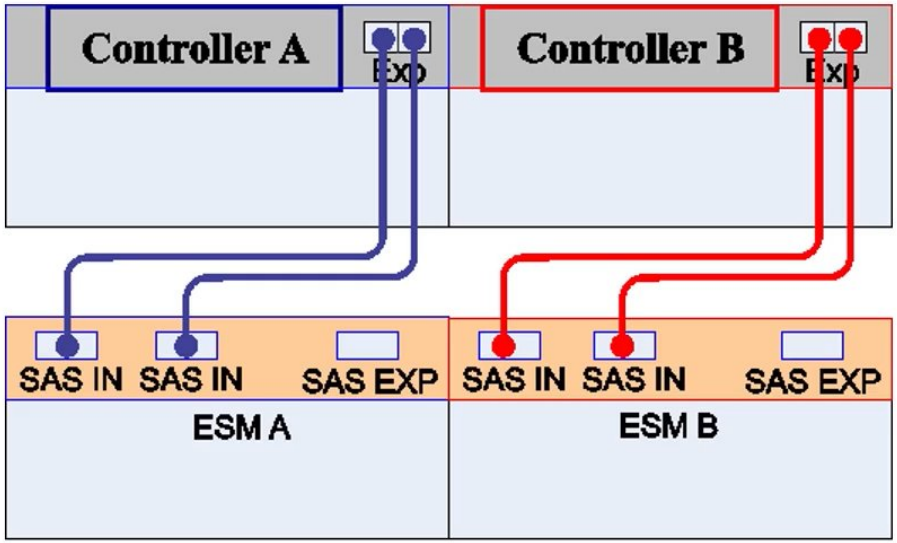
- Your PC at home most likely has a SATA (Serial Advanced Technology Attachment) SSD or HDD drive
- Servers and storage systems typically support SAS (Serial Attached SCSI) drives which provide higher performance than SATA in server environments



FLACKBOX
www.flackbox.com

SAS Architecture

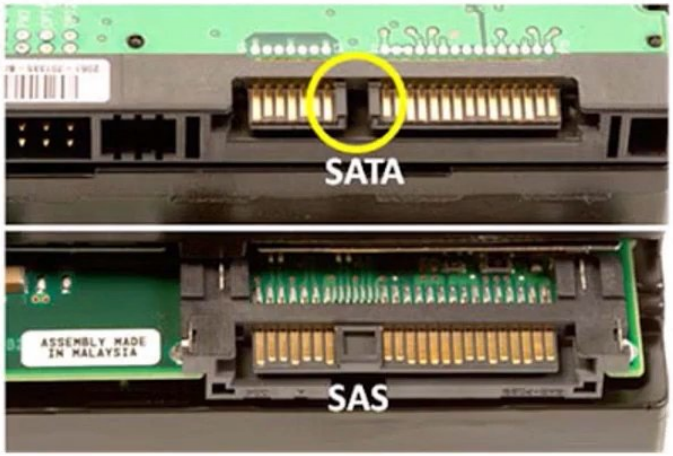
- SAS is an architecture – SAS disks fit in SAS disk shelves which are connected to storage controllers with SAS cables (or in SAS drive bays in the chassis)



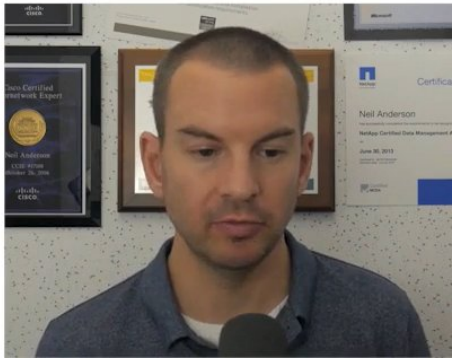
FLACKBOX
www.flackbox.com

SSD, SAS and SATA Drives

- SAS is backwards compatible with SATA
- A SAS drive bay can take SAS SSD, SAS HDD, SATA SSD or SATA HDD drives



FLACKBOX
www.flackbox.com



SSD, SAS and SATA Drives

- SSD drives provide the highest performance but have the highest cost per GB. They are used where high performance is required.
- SATA HDD drives are available in large sizes and have the lowest cost per GB, but the lowest performance. They are used to provide high capacity where performance is not important.
- SAS HDD drives provide a balance between performance, capacity and cost.

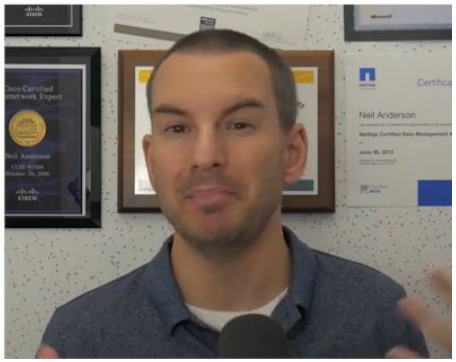
FLACKBOX
www.flackbox.com

Legacy Drive Types - SCSI and FC-AL

- SCSI (Small Computer System Interface) and FC-AL (Fibre Channel Arbitrated Loop) drives have their own connectors and are not compatible with other drive types
- They are a legacy technology

SCM Storage Class Memory

- SCM storage is an emerging technology with physical media which fits into memory slots
- SCM provides NVRAM (Non Volatile RAM) rather than DRAM (Dynamic RAM) – it is persistent storage and survives power outages
- Intel's 3D XPoint is an early SCM technology



FLACKBOX
www.flackbox.com

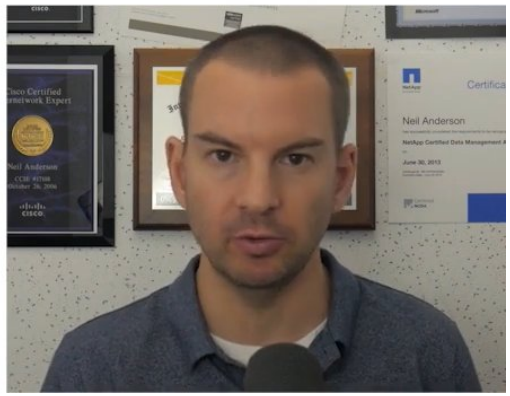
‘All Flash’ and ‘Hybrid Flash’ Storage

- Recent storage development has focused on Flash rather than HDD media.
- As SSDs come down in price HDD drives are becoming less attractive.
- ‘All Flash’ storage systems are now popular.
- ‘Hybrid Flash’ storage systems combine Flash storage for performance with HDD drives for capacity.

FLACKBOX
www.flackbox.com

RAID

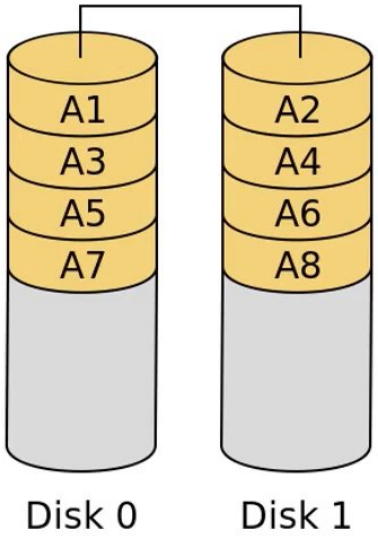
- RAID stands for Redundant Array of Inexpensive Disks or Redundant Array of Independent Disks
- Multiple physical disks are combined into a single logical unit
- This provides redundancy, improved performance, or both when compared to a single disk
- Different RAID levels provide different levels of redundancy or performance
- RAID can be managed in software by the operating system or in hardware by a RAID controller

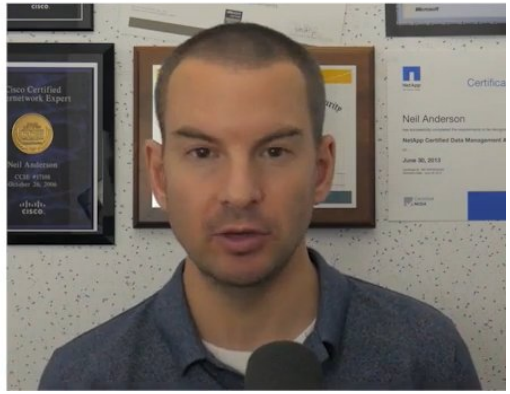


FLACKBOX
www.flackbox.com

RAID 0

- RAID 0 is also known as a striped set
- Data is split evenly across all the disks in the set
- This provides better performance, but not redundancy
- If any disk in the set fails, all the data is lost

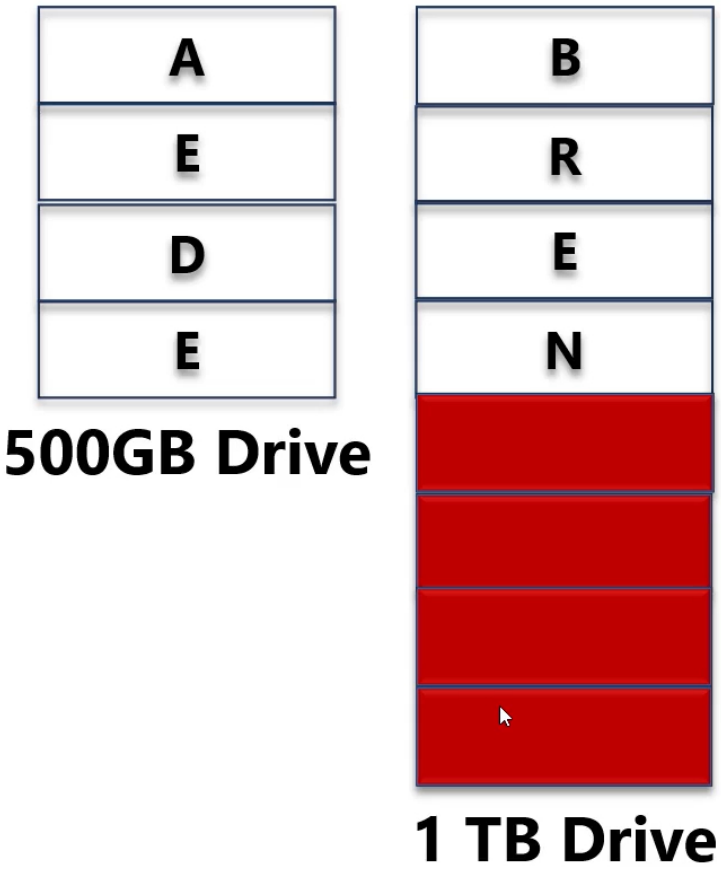




FLACKBOX
www.flackbox.com

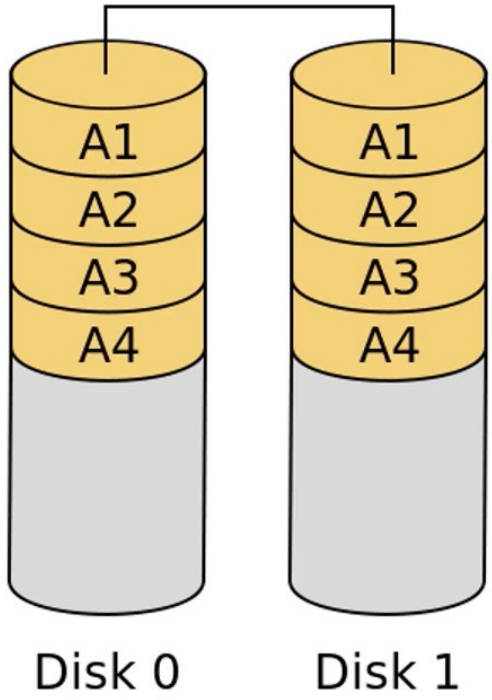
Drive Types, Speeds and Sizes

- You can mix different drive types, RPM speeds and sizes in the same storage system, but don't put them in the same RAID group
- If you use different capacity size disks, the usable size is limited to the size of the smallest disk



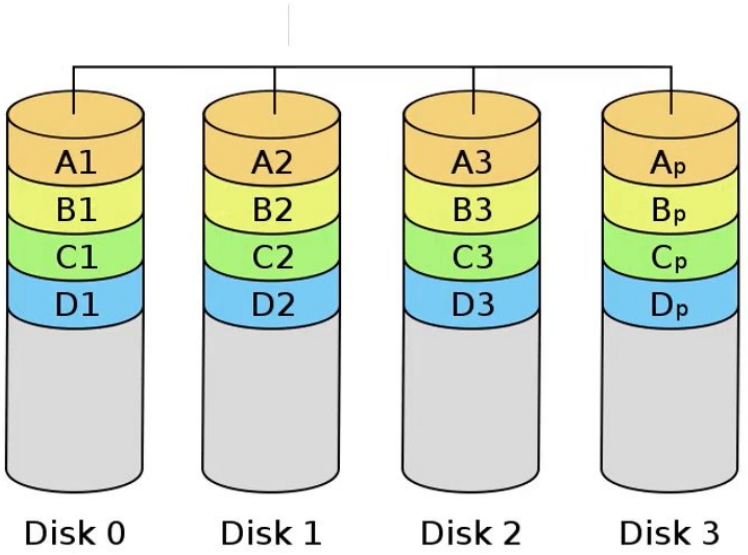
RAID 1

- RAID 1 is also known as a mirrored set
- A copy of the data is written to both disks in the set
- This provides redundancy. If one disks fails you still have a working copy of the data
- Write performance is not improved as a copy of the same data is written to both disks
- Read performance is improved as reads can be serviced by either disk

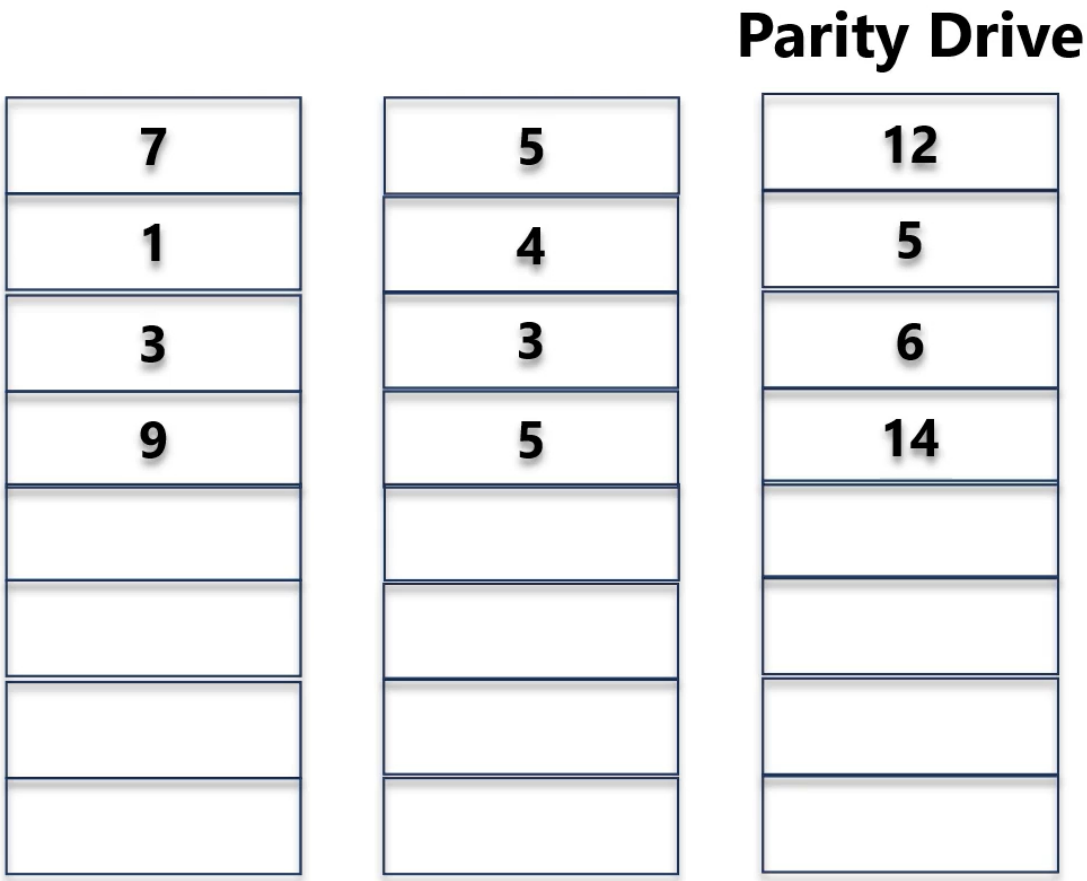


RAID 4

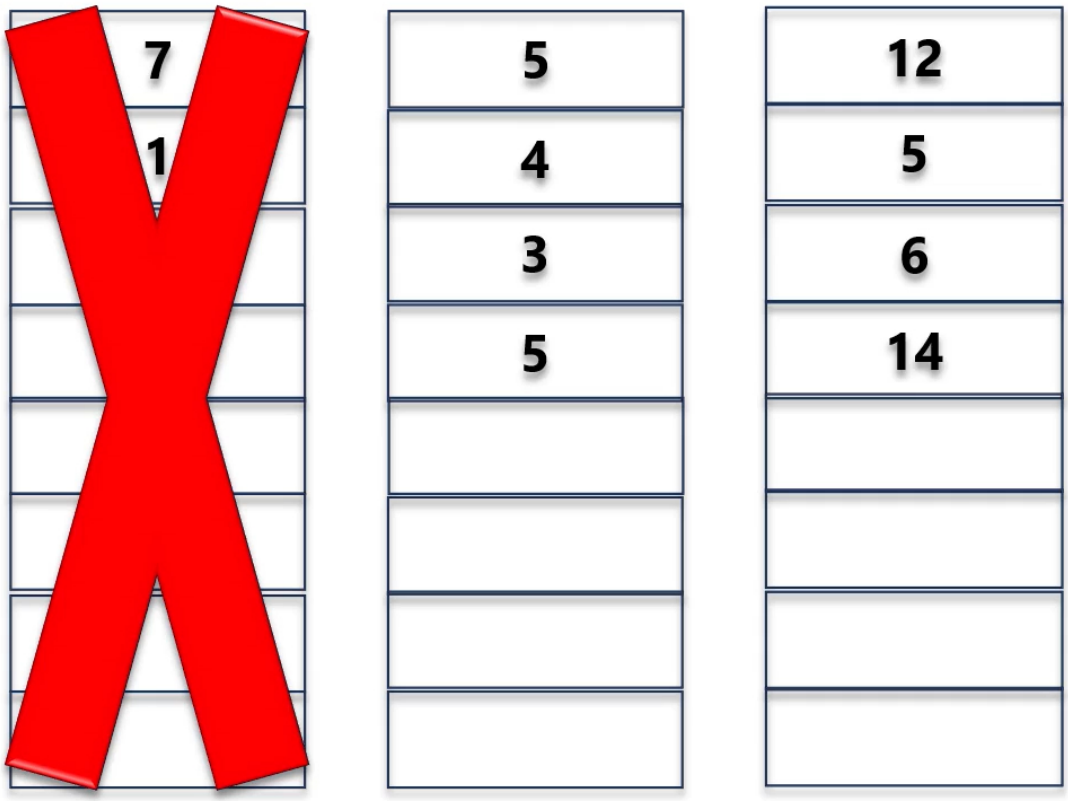
- RAID 4 uses block level striping with a dedicated parity disk
- Data can be recreated from parity if any one of the disks in the set fails
- Read performance is improved as multiple disks concurrently service reads
- Write performance is not improved as all parity data is written to the same single disk



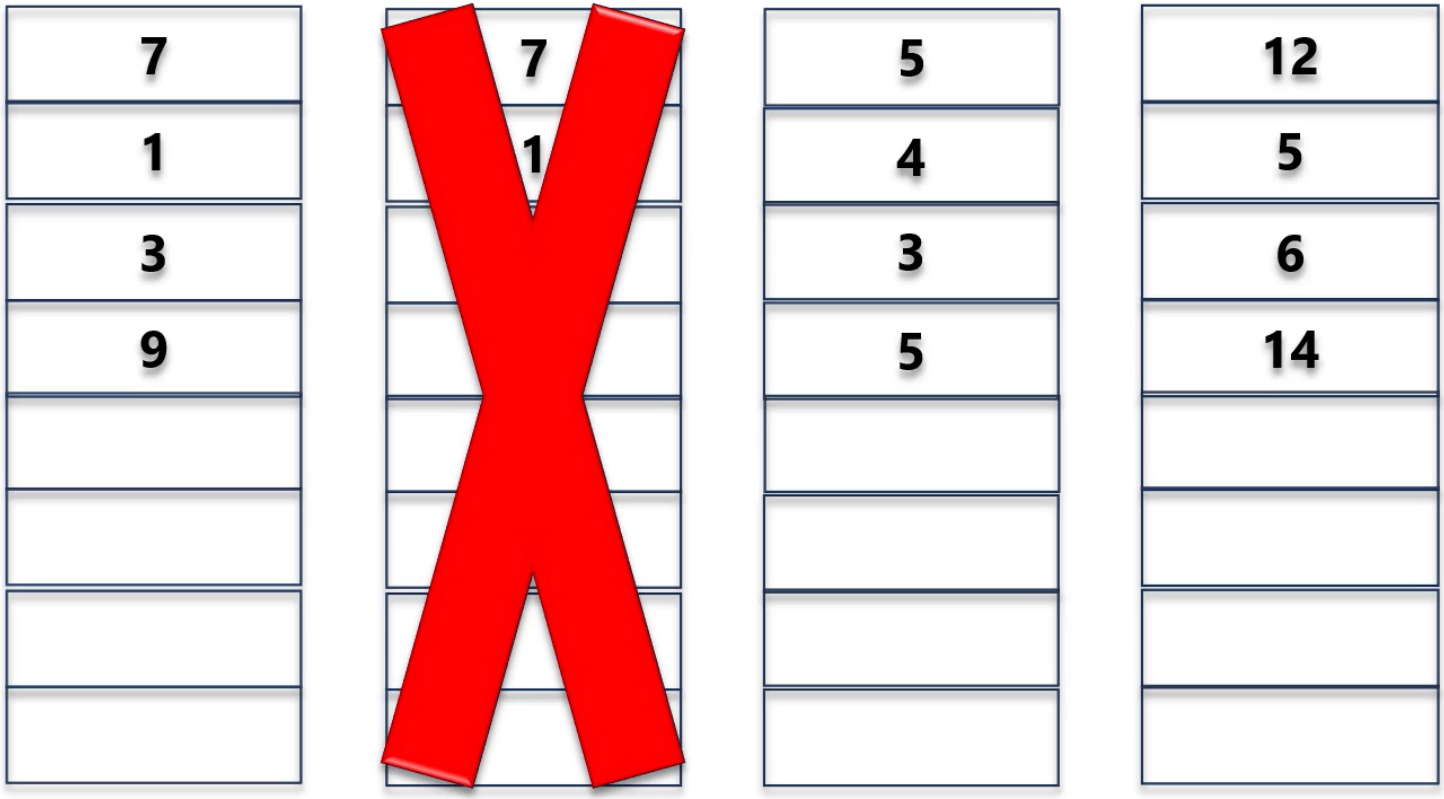
RAID 4



RAID 4



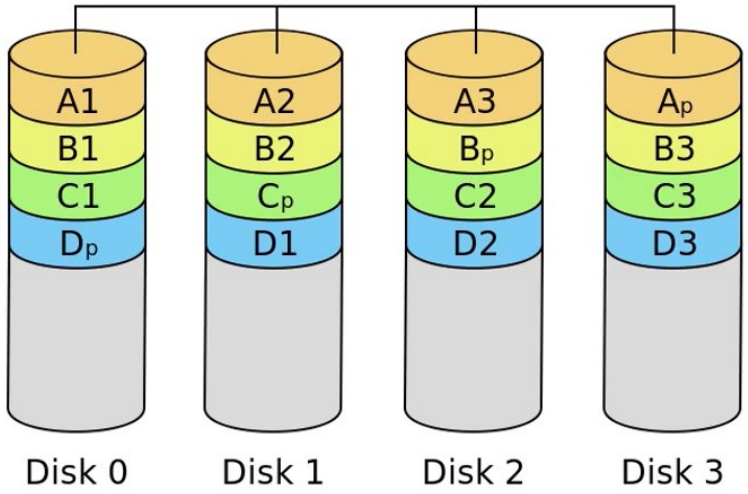
RAID 4



Performance will be degraded until the failed drive is replaced and rebuilt

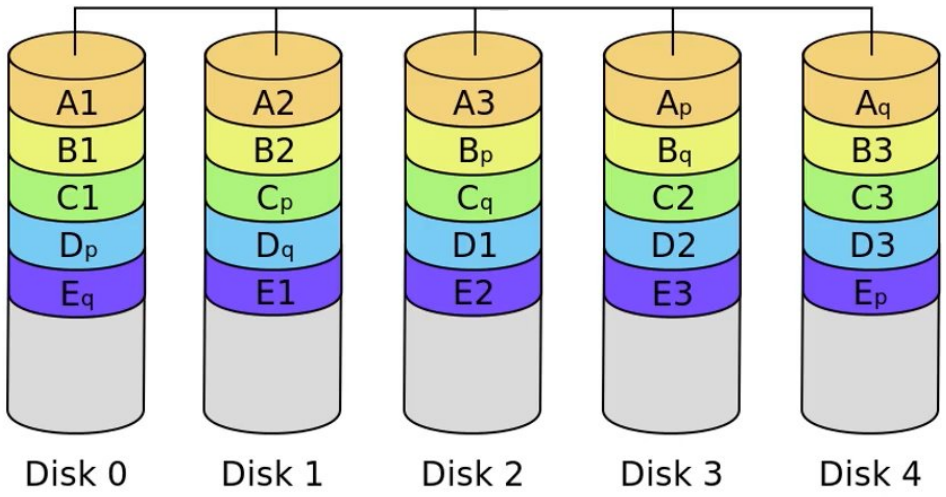
RAID 5

- RAID 5 uses block level striping with distributed parity
- Data can be recreated from parity if any one of the disks in the set fails
- Read performance is improved as multiple disks concurrently service reads
- Write performance is also improved as parity data is spread throughout the set



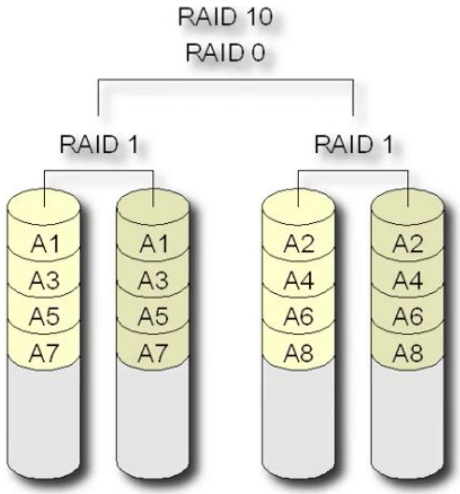
RAID 6

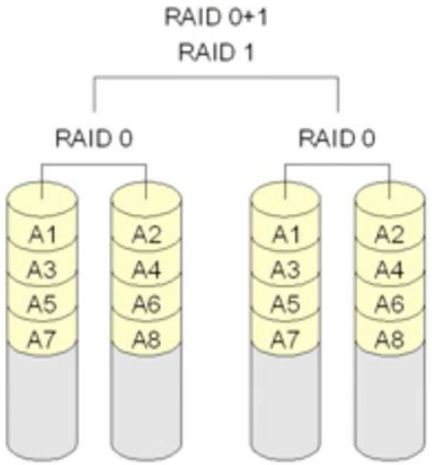
- RAID 6 uses block level striping with two parity blocks distributed throughout the set
- Data can be recreated from parity if two of the disks in the set fail
- Read and write performance is improved



Hybrid RAID

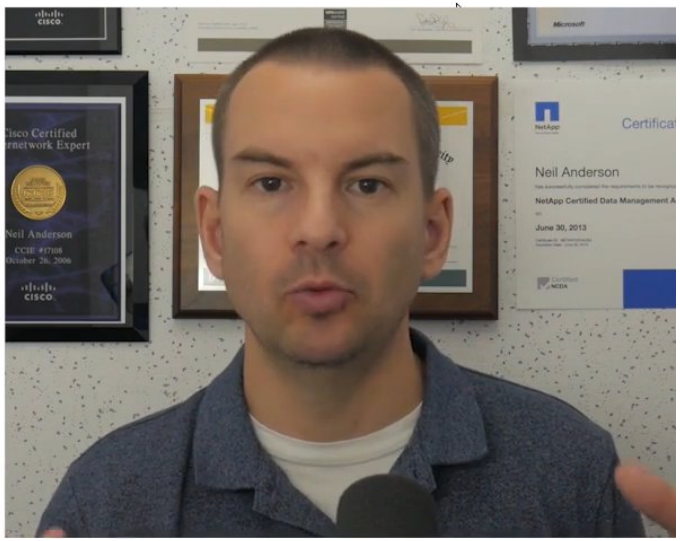
- RAID levels can be nested into a hybrid RAID set
- In RAID 10 (1+0), multiple RAID 1 mirrored sets are nested into a RAID 0 striped set
- In RAID 0+1, multiple RAID 0 sets are nested into a RAID 1 mirror
- In RAID 50 (5+0), multiple RAID 5 striped sets with parity are nested into a RAID 0 striped set





NAS Storage

- Network Attached Storage (NAS) provides file level access to storage resources.
- The management of the file system resides with the remote storage system.

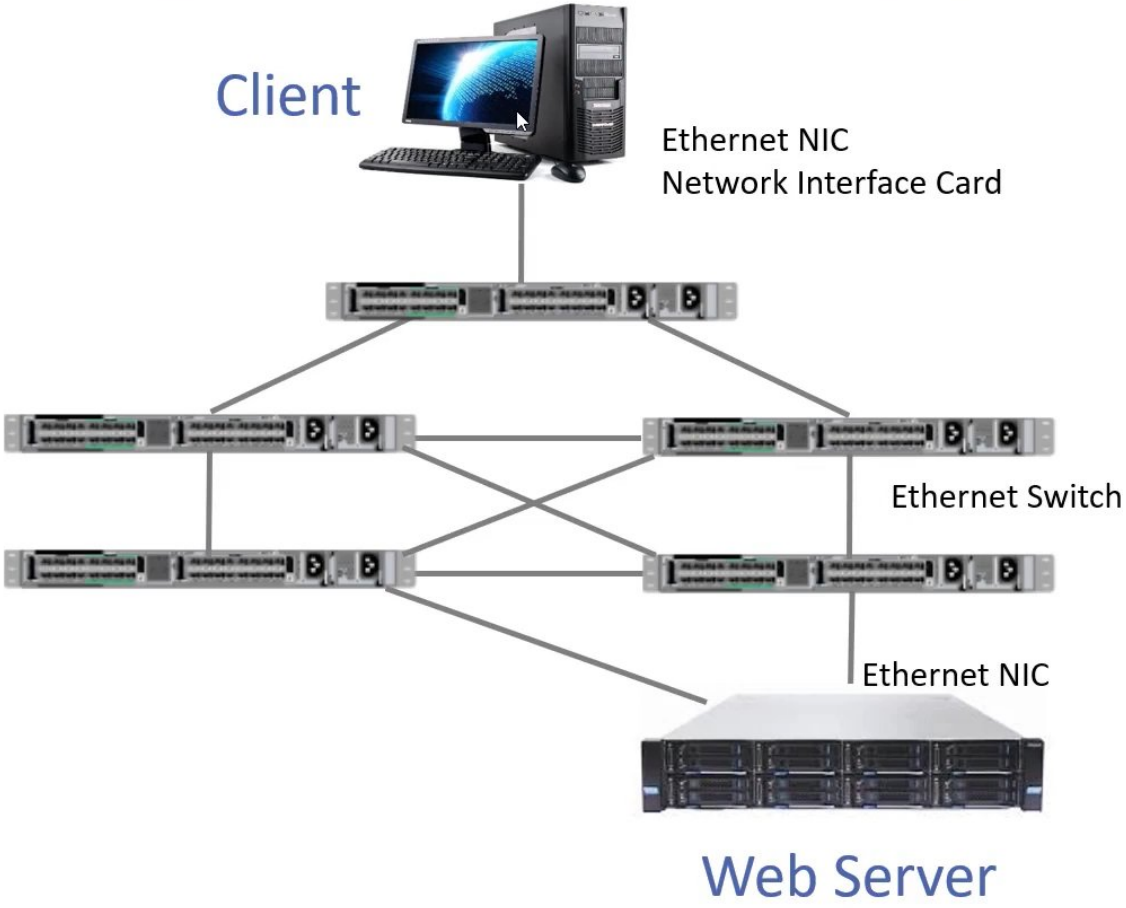


NAS Storage

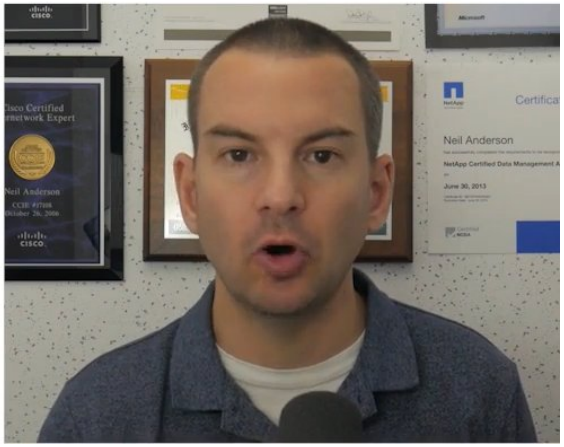
- NAS typically uses the existing data network infrastructure.

FLACKBOX
www.flackbox.com

NAS Networks

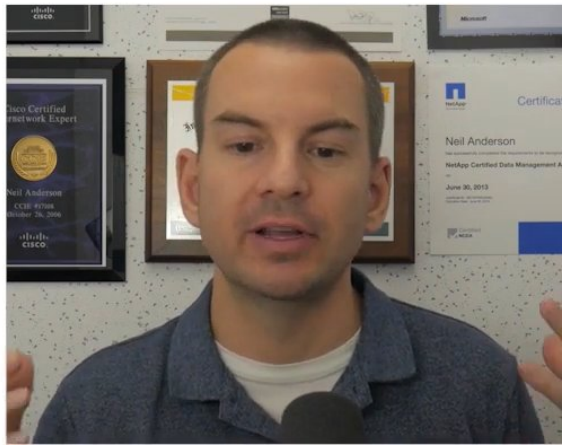


FLACKBOX
www.flackbox.com



SAN Storage

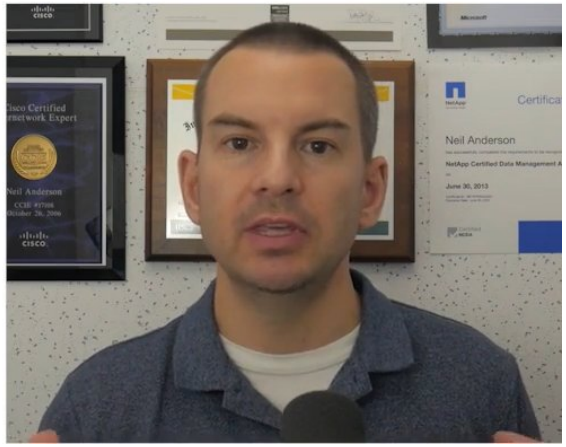
- Storage Area Networks (SAN) provide block level access to storage resources.
- The client manages its allocated space on the storage system the same as if it was a local hard drive.
- The management of the file system resides with the client.



FLACKBOX
www.flackbox.com

SAN Storage

- SAN has traditionally had a separate, dedicated storage network infrastructure.

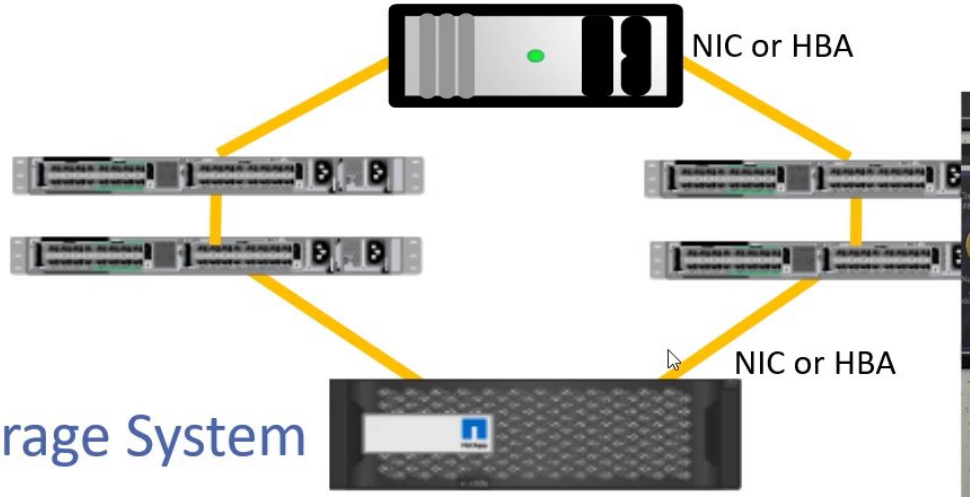


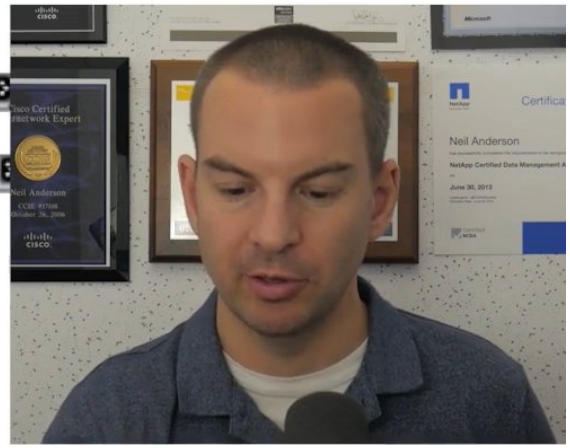
FLACKBOX
www.flackbox.com

SAN Networks

FLACKBOX
www.flackbox.com

- Server to Storage Network – Copper or Fibre

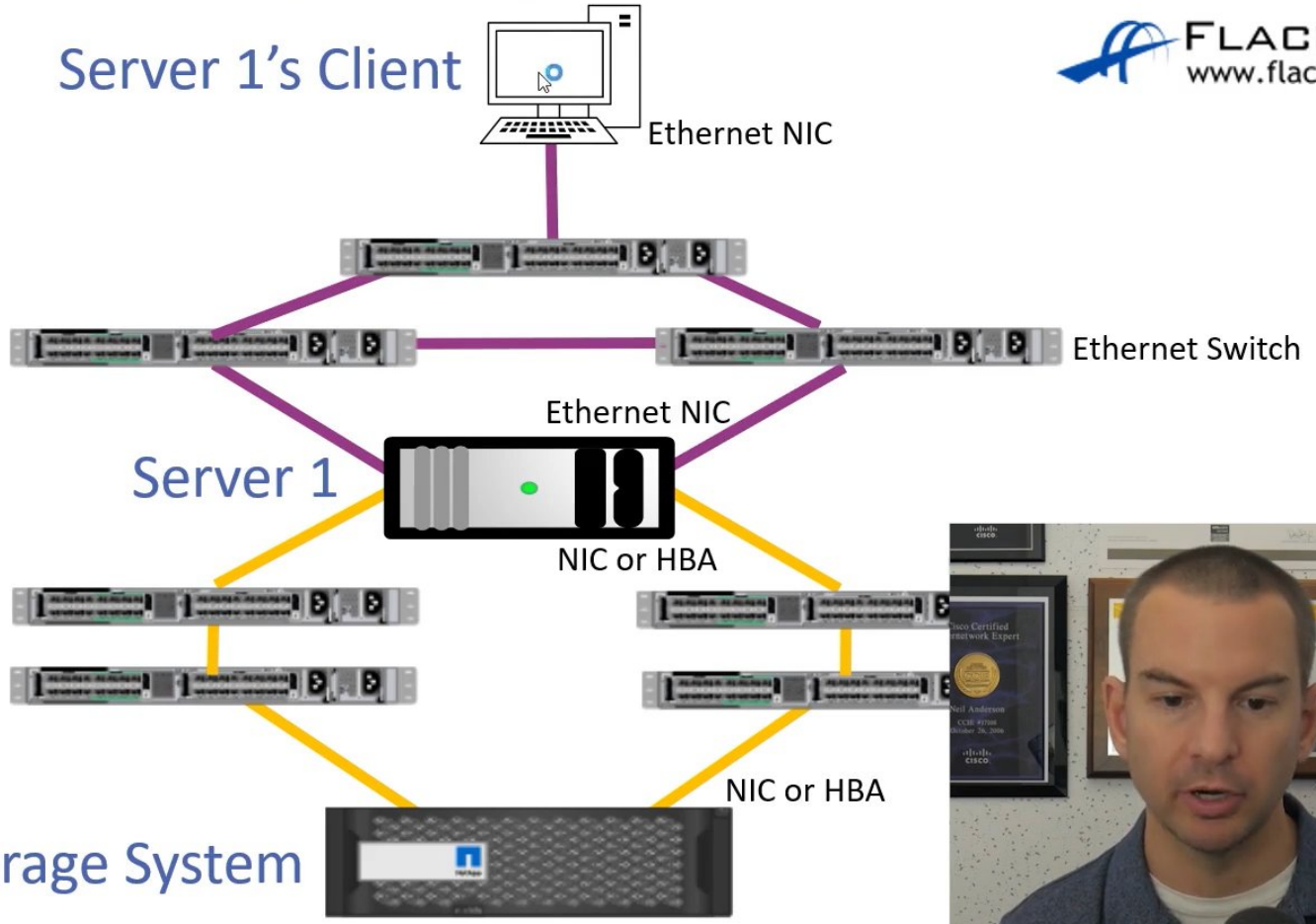




FLACKBOX
www.flackbox.com

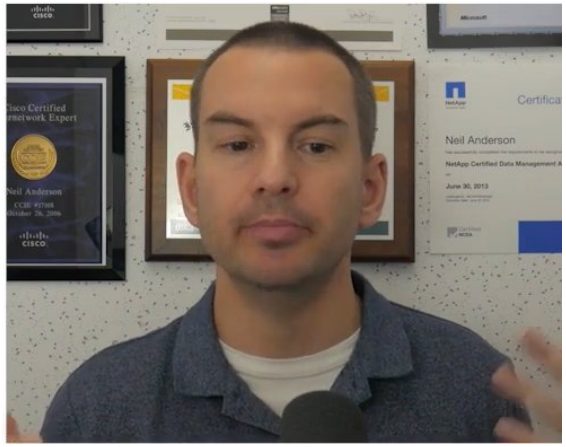
SAN Networks

- Client to Server Ethernet Data Network
- Server to Storage Network – Copper or Fibre



Boot from SAN

- As SAN provides block level access which appears the same as a locally installed drive to the client, it is possible to Boot from SAN
- Where Boot from SAN is deployed it is possible to have diskless servers
- Boot from SAN is not possible with NAS



FLACKBOX
www.flackbox.com

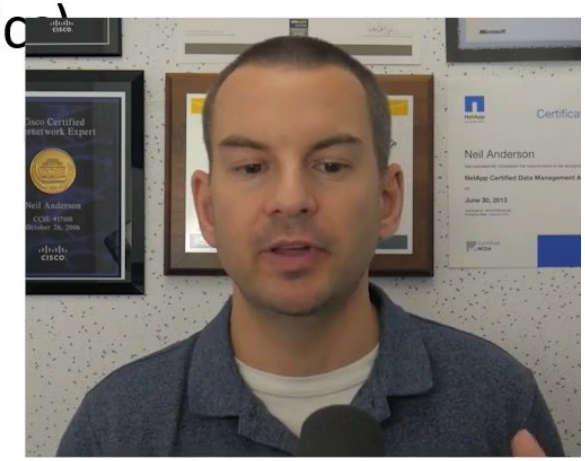
Storage Protocols Summary

- NAS Protocols
  - CIFS / SMB (Common Internet File System / Server Message Block)
  - NFS (Network File System)

FLACKBOX
www.flackbox.com

Storage Protocols Summary

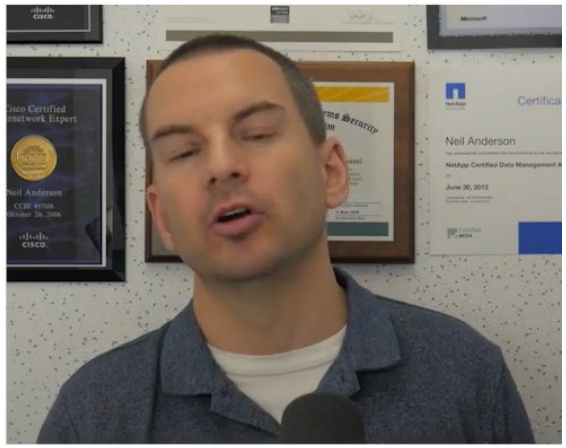
- SAN Protocols
  - FC (Fibre Channel)
  - iSCSI (Internet Small Computers Interface)
  - FCoE (Fibre Channel over Ethernet)
  - NVMeOF (Non Volatile Memory Express Over Fabrics)



FLACKBOX
www.flackbox.com

Unified Storage

- Early storage systems typically supported either SAN or NAS protocols.
- Many modern storage systems support both SAN and NAS.



FLACKBOX
www.flackbox.com

NAS Protocols – SMB / CIFS

- Server Message Block (SMB) was developed by IBM
- CIFS (Common Internet File System) is a version of SMB developed by Microsoft
- The two terms are often used interchangeably, however newer versions of SMB are used in today's operating systems

NAS Protocols – SMB / CIFS

- CIFS was developed by Microsoft for Windows based clients
- Samba allows UNIX based clients to access CIFS shares
- Servers ‘share’ and clients ‘use’ or ‘map’ the share

FLACKBOX
www.flackbox.com

NAS Protocols - N

- NFS is the Network File System
- NFS was developed by Sun Microsystems for Unix based clients
- It is commonly used for VMFS Datastores in VMware (which can also use SAN)
- Windows based clients can access NFS exports
- Servers ‘export’ and clients ‘mount’ the export

SAN Terminology

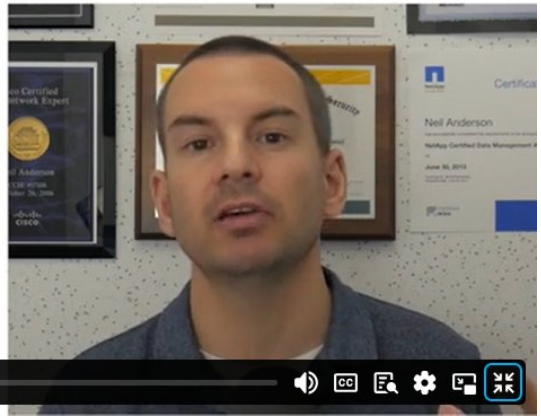
- A **LUN** (Logical Unit Number) represents a disk that will be presented to a host
- LUNs are specific to SAN (not NAS) protocols

- The client is known as the **Initiator**
- The storage system is known as the **Target**

FLACKBOX
www.flackbox.com

SAN Protocols – Fibre Channel (FC)

- Fibre Channel is the original SAN protocol, and is still very popular
- It uses dedicated adapters, cables and switches and is different than Ethernet at all layers of the OSI stack, including the physical level
- FCP is used to send SCSI commands over the Fibre Channel network



SAN Protocols – Fibre Channel (FC)

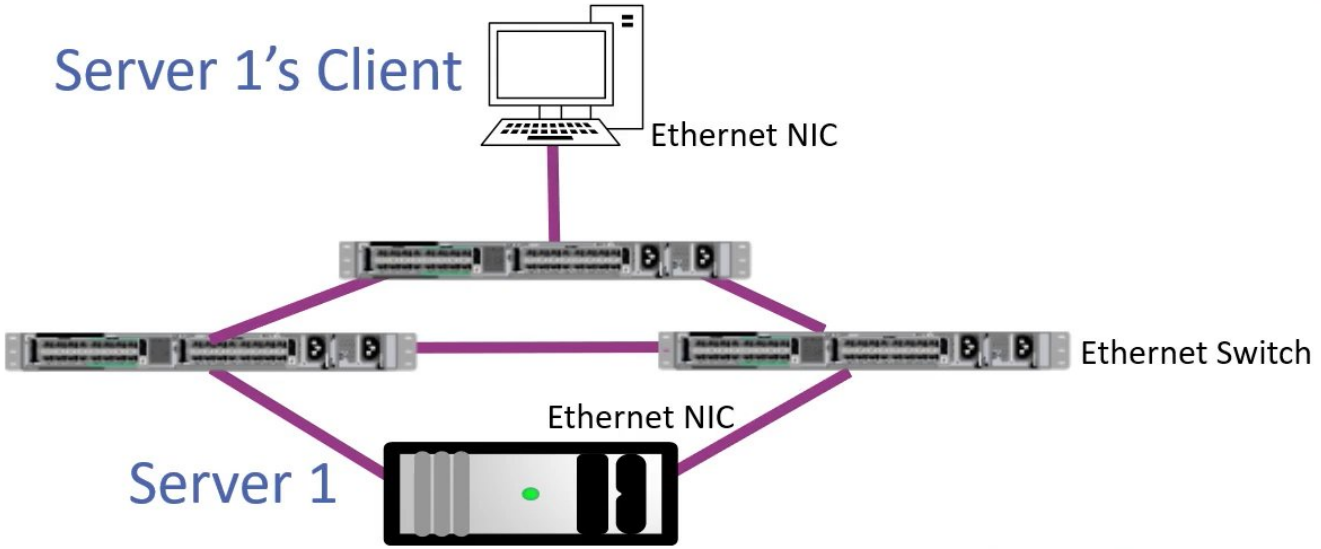
- Fibre channel is a very stable and reliable protocol
- It is lossless, unlike TCP and UDP
- It supports bandwidths of 1, 2, 4, 8, 16, 32, and 128 Gbps

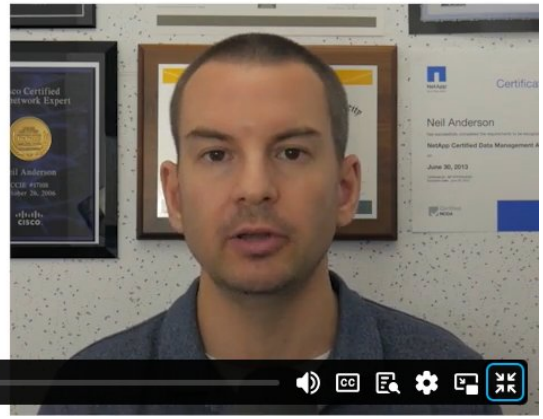
FLACKBOX
www.flackbox.com

03:26

Fibre Channel Networks

- Client to Server Ethernet Data Network



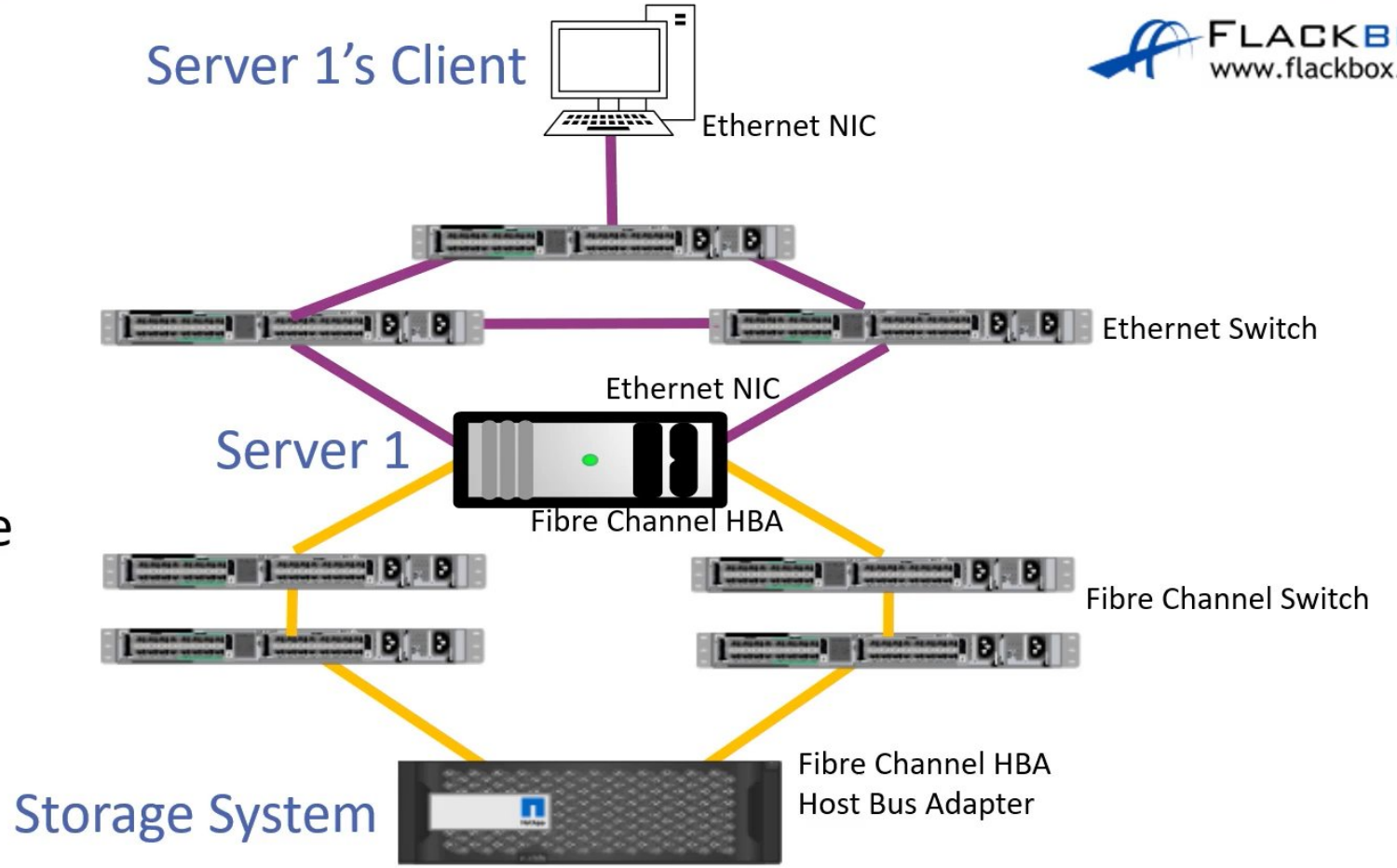


FLACKBOX
www.flackbox.com

04:20

Fibre Channel Networks

- Client to Server Ethernet Data Network
- Server to Storage Fibre Channel Network



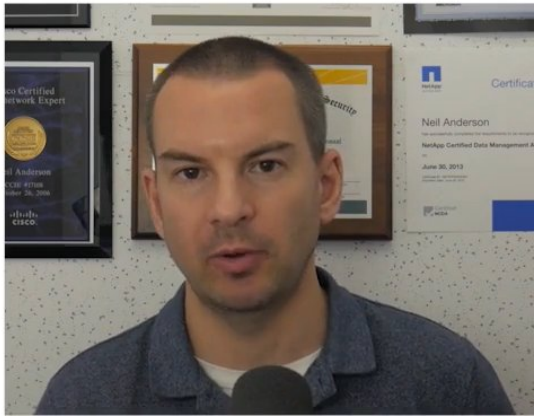
World Wide Names

- FCP uses WWN World Wide Names for it's addressing
- WWNs are 8 byte addresses made up of 16 hexadecimal characters
- Their format is: **15:00:00:f0:8c:08:95:de**

FLACKBOX
www.flackbox.com

World Wide Node Names

- The WWNN World Wide Node Name is assigned to a node in the storage network
- The same WWNN can identify multiple network interfaces of a single network node



FLACKBOX
www.flackbox.com

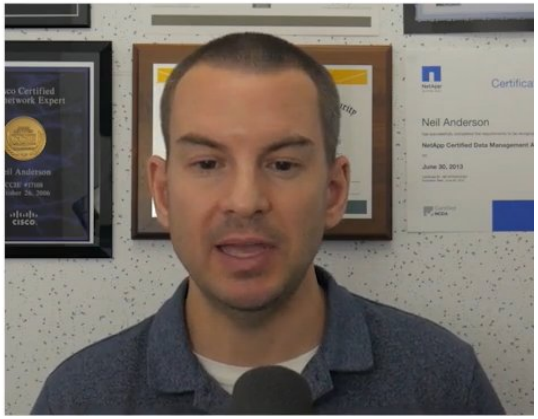
World Wide Port Names

- A different WWPN World Wide Port Name is assigned to every individual port on a node
- A multi-port HBA will have different WWPN's on each port
- WWPN's are the equivalent of MAC addresses in Ethernet
- The WWPN is burned in by the manufacturer and guaranteed to be globally unique

FLACKBOX
www.flackbox.com

World Wide Port Names

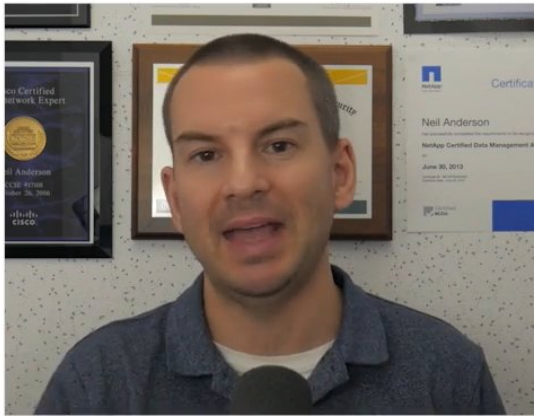
- WWPNs are assigned to HBAs on both the clients and the storage system
- We are primarily concerned with WWPNs, not WWNNs, when configuring Fibre Channel networks



FLACKBOX
www.flackbox.com

Aliases

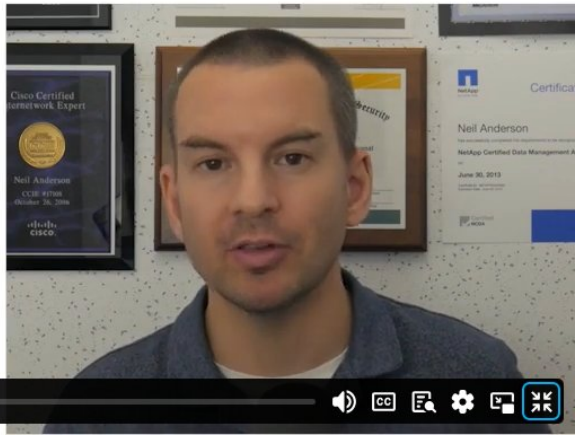
- Aliases can be configured to make configuration and troubleshooting easier
- For example we could create an alias named EXCHANGE-SERVER for the Exchange Server with WWPN 15:00:00:f0:8c:08:95:de
- Aliases can be configured on both the fibre channel switches and the storage system



FLACKBOX
www.flackbox.com

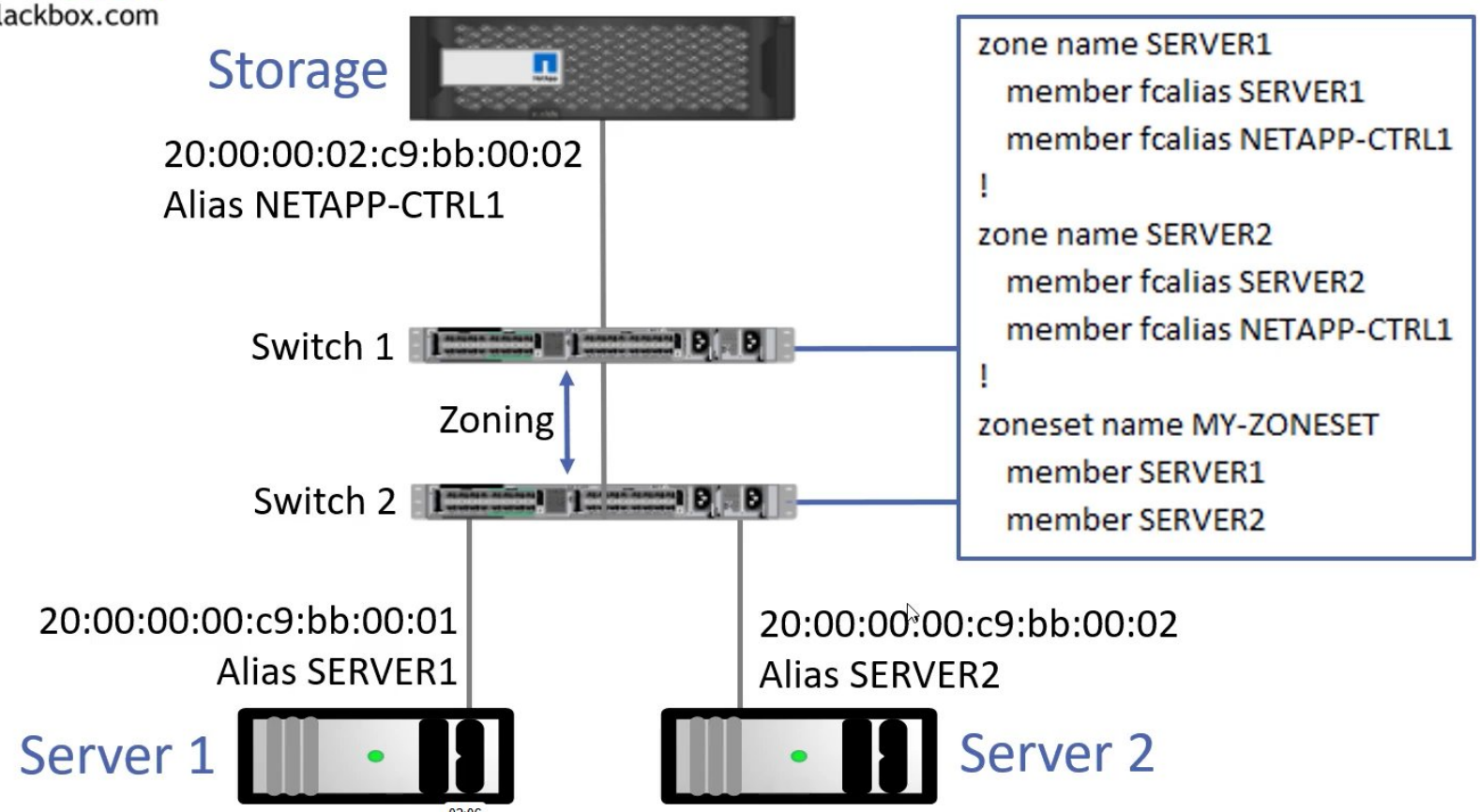
Zoning

- For security, zoning will be configured on the switches to control which hosts are allowed to communicate with each other
- Initiators (servers) will be allowed to communicate with targets (the storage system), but initiators will not be allowed to communicate with each other over fibre channel



Zoning

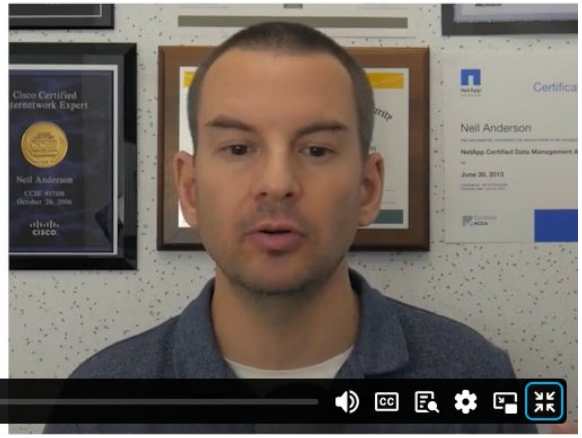
FLACKBOX
www.flackbox.com



02:06

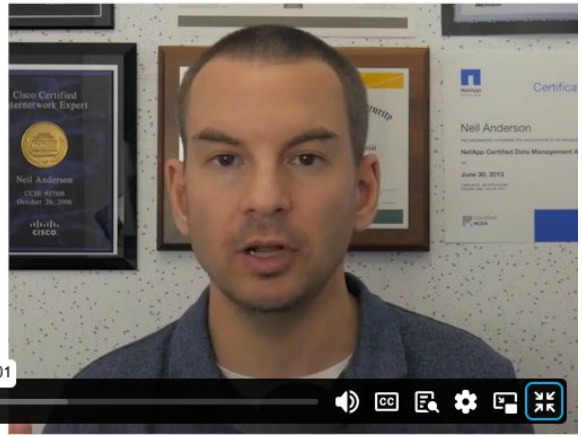
LUN Masking

- It is critical that the right LUN is presented to the right host
- If the wrong host is able to connect to a LUN then it is liable to corrupt it
- Zoning prevents unauthorized hosts from reaching the storage system, but it will not prevent a host from accessing the wrong LUN once it gets there



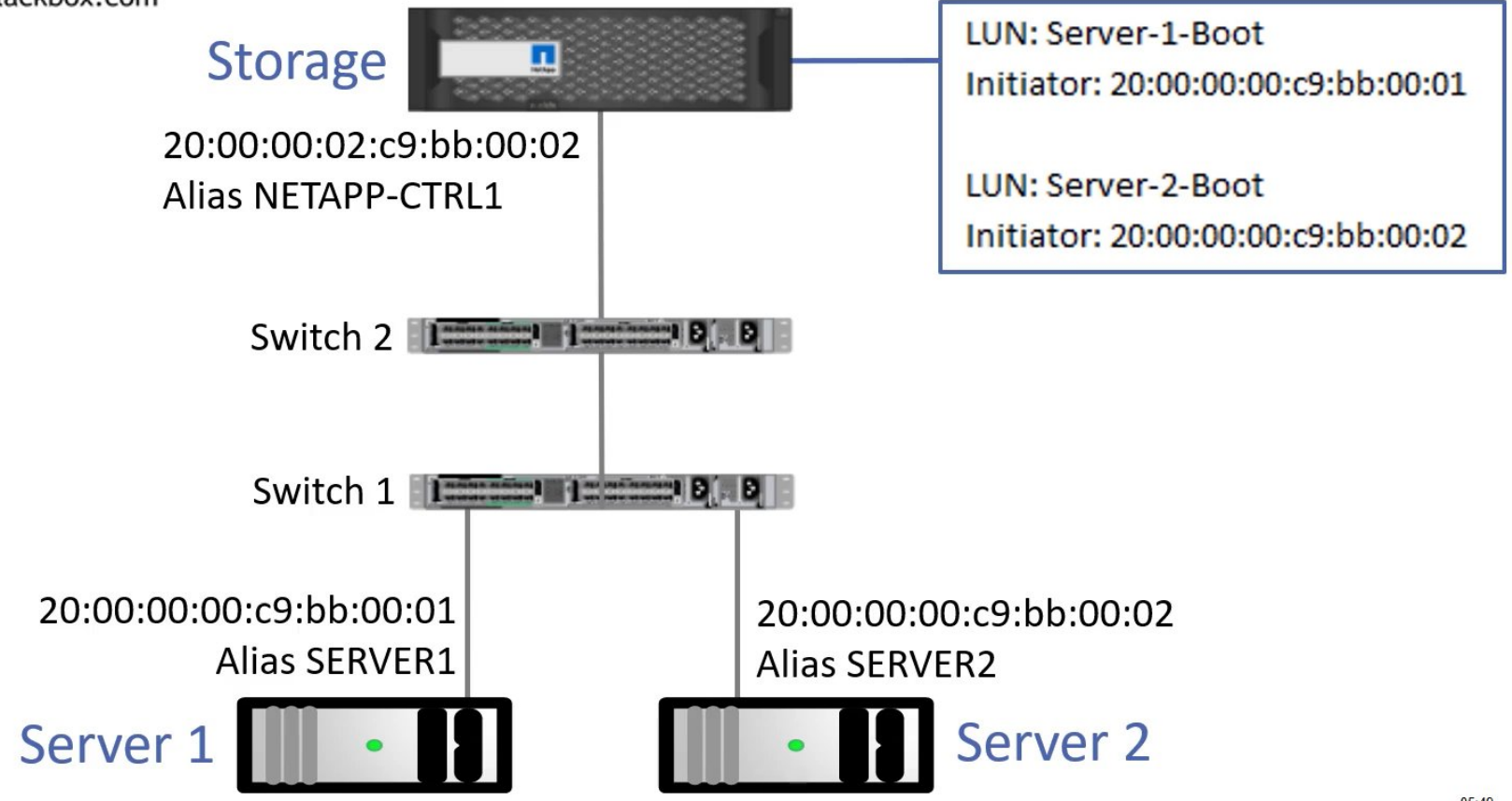
LUN Masking

- LUN Masking is configured on the storage system to lock a LUN down to the host or hosts who are authorized to access it
- To secure your storage you need to configure zoning on your switches **and** LUN masking on your storage system



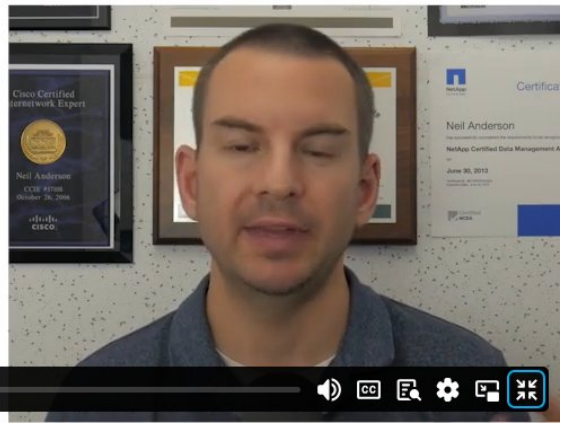
LUN Masking

FLACKBOX
www.flackbox.com



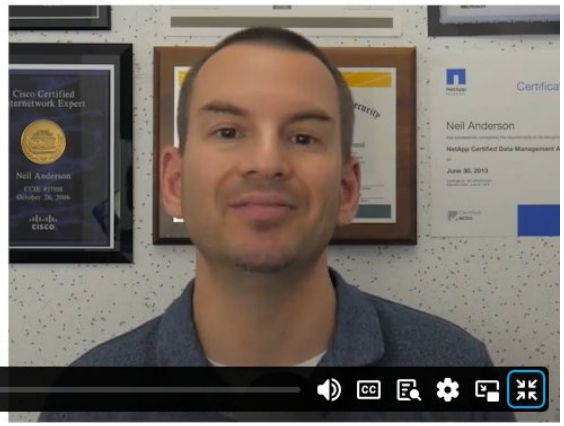
Switch Domain IDs

- Each switch in the fibre channel network will be assigned a unique Domain ID
- One switch in the network will be automatically assigned as the Principle Switch. It assigns Domain IDs to the other switches
- Each switch learns about the other switches in the network and how to route to them based on their Domain ID



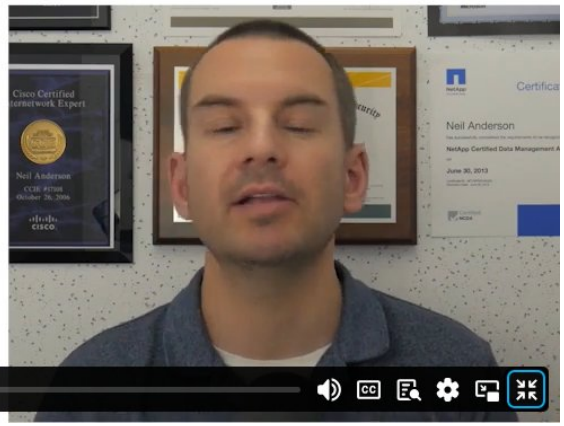
FLOGI Fabric Login

- When a server's or storage system's HBA powers on it will send a FLOGI Fabric Login request to its locally attached fibre channel switch
- The switch will then assign it a 24 bit FCID Fibre Channel ID address
- The FCID assigned to hosts is made up of the switch's Domain ID and the switch port the host is plugged into



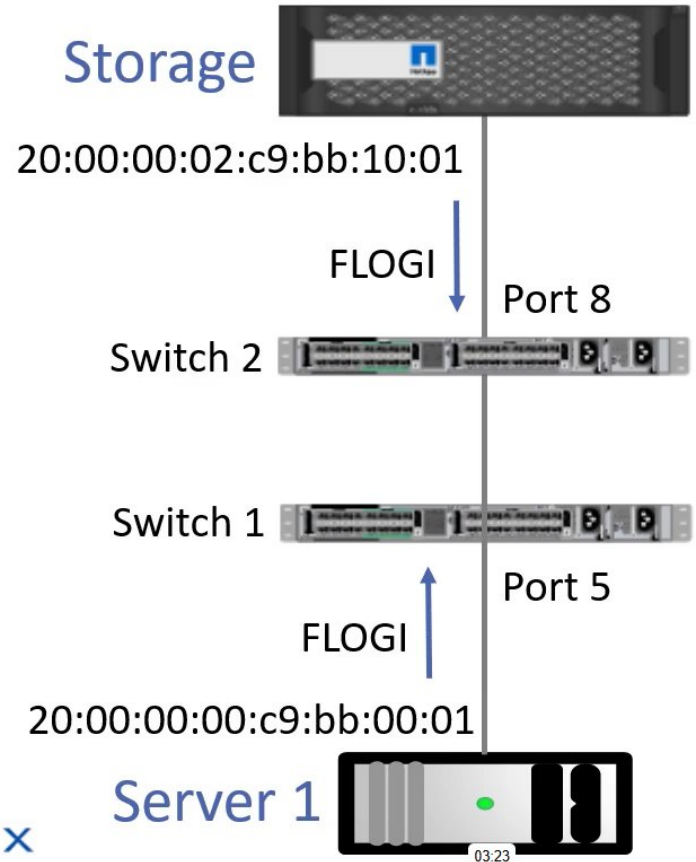
FLOGI Fabric Login

- The FCID is similar to an IP address. It is used by fibre channel switches to route traffic between servers and their storage
- Switches maintain a table of FCID to WWPN address mappings and which port the host is located on



02:29

The Fabric Login

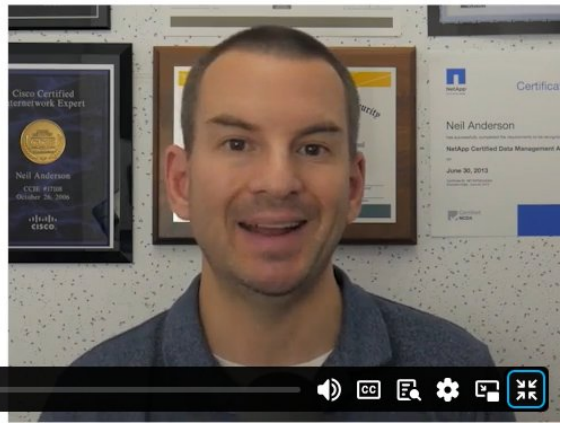



```
FC-Switch-2# show flogi database
--------------------------------------------------------------------------------
INTERFACE        FCID       PORT NAME               NODE NAME
--------------------------------------------------------------------------------
fc1/8            0xcf2000  20:00:00:02:c9:bb:10:01 20:00:00:02:c9:bb:11:01
                    [NETAPP-CTRL1]
```

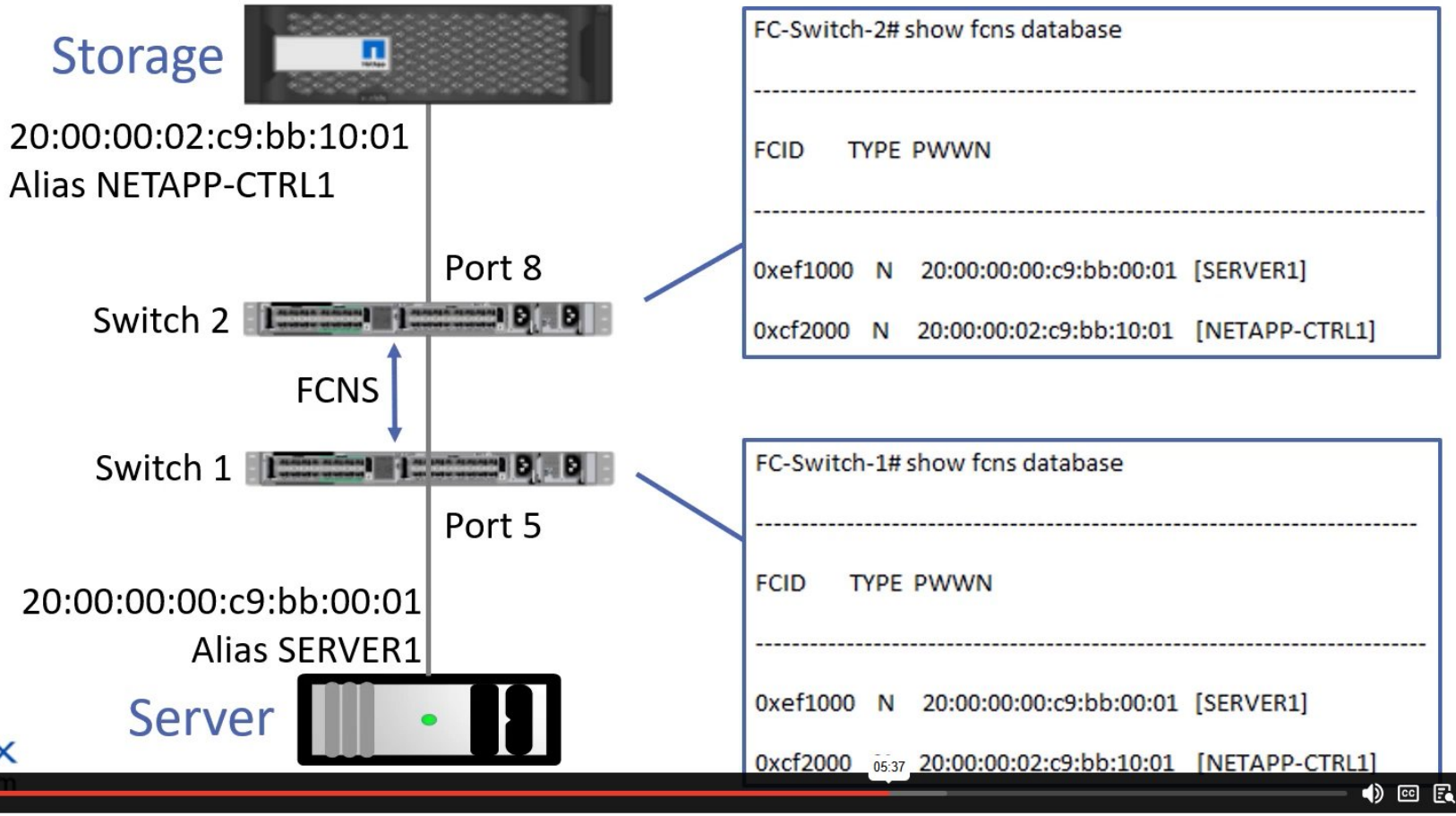
```
FC-Switch-1# show flogi database
--------------------------------------------------------------------------------
INTERFACE        fcid       PORT NAME               NODE NAME
--------------------------------------------------------------------------------
fc1/5            0xef1000  20:00:00:00:c9:bb:00:01 20:00:00:00:c9:bb:01:01
                    [SERVER1]
```

FCNS Fibre Channel Name Service

- Fibre channel switches share the FLOGI database information with each other using the FCNS Fibre Channel Name Service
- Each switch in the network learns where each WWPN is and how to route traffic there

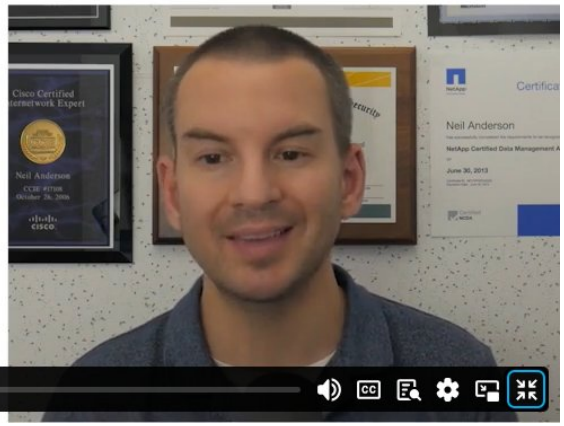


FCNS Fibre Channel Name Service



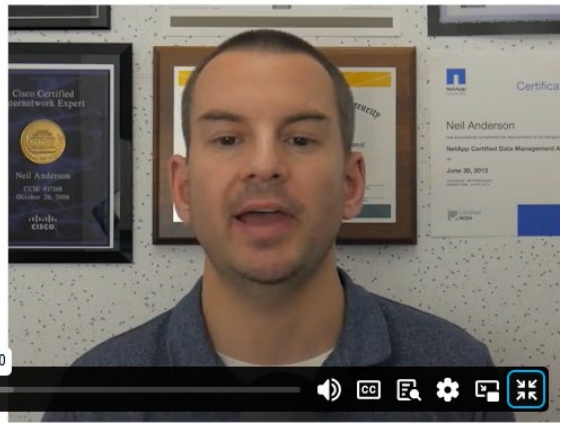
PLOGI Port Login

- After the FLOGI Fabric Login process is complete, the initiator will send the PLOGI Port Login
- Based on the zoning configuration on the switch, the host will learn its available target WWPNs



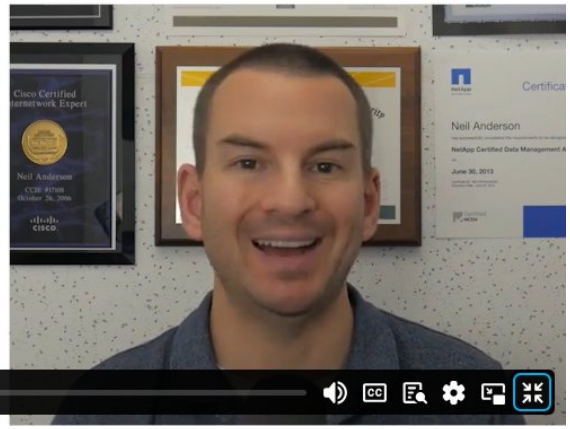
PRLI Process Login

- Finally the initiator host will send a PRLI Process Login to its target storage
- The storage system will grant access to the host based on its configured LUN masking



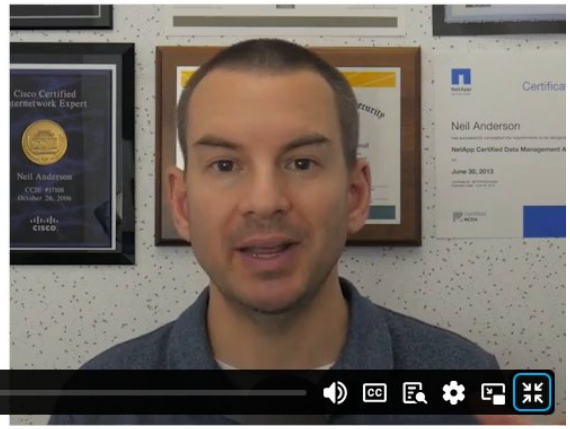
Redundant SAN Fabrics

- Server's access to their storage will invariably be mission critical for the enterprise, so there should be no single points of failure
- Redundant Fibre Channel networks will be put in place, known as Fabric A and Fabric B (or SAN A and SAN B)
- Each server and storage system host will be connected to both fabrics with redundant HBA ports



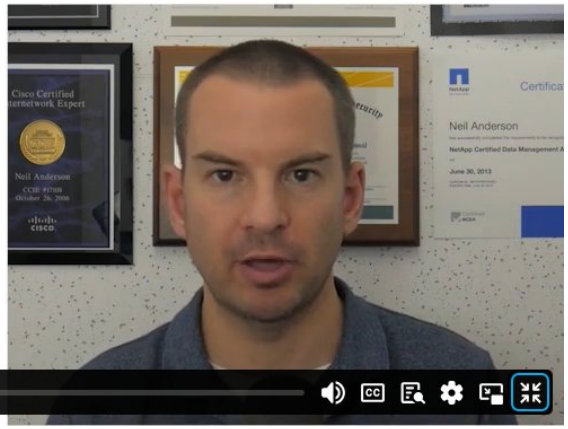
Redundant SAN Fabrics

- Fibre Channel switches distribute shared information to each other such as Domain IDs, the FCNS database, and zoning
- If an error in Fabric A was able to propagate to Fabric B this would bring down both fabrics and drop the servers connection to their storage

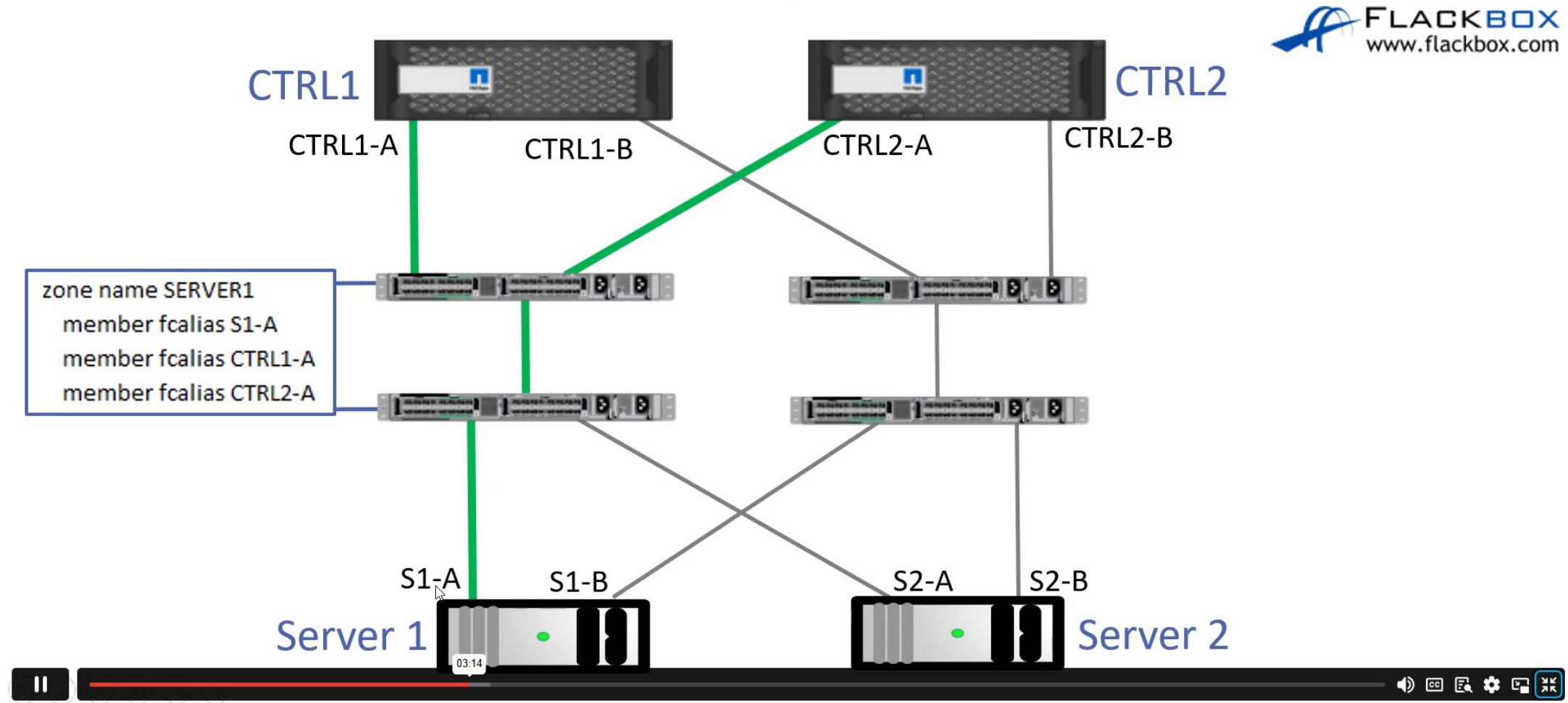


Redundant SAN Fabrics

- For this reason switches in different sides of the fabric are NOT cross connected to each other – both sides of the fabric are kept physically separate
- End hosts are connected to both fabrics, but the switches are not

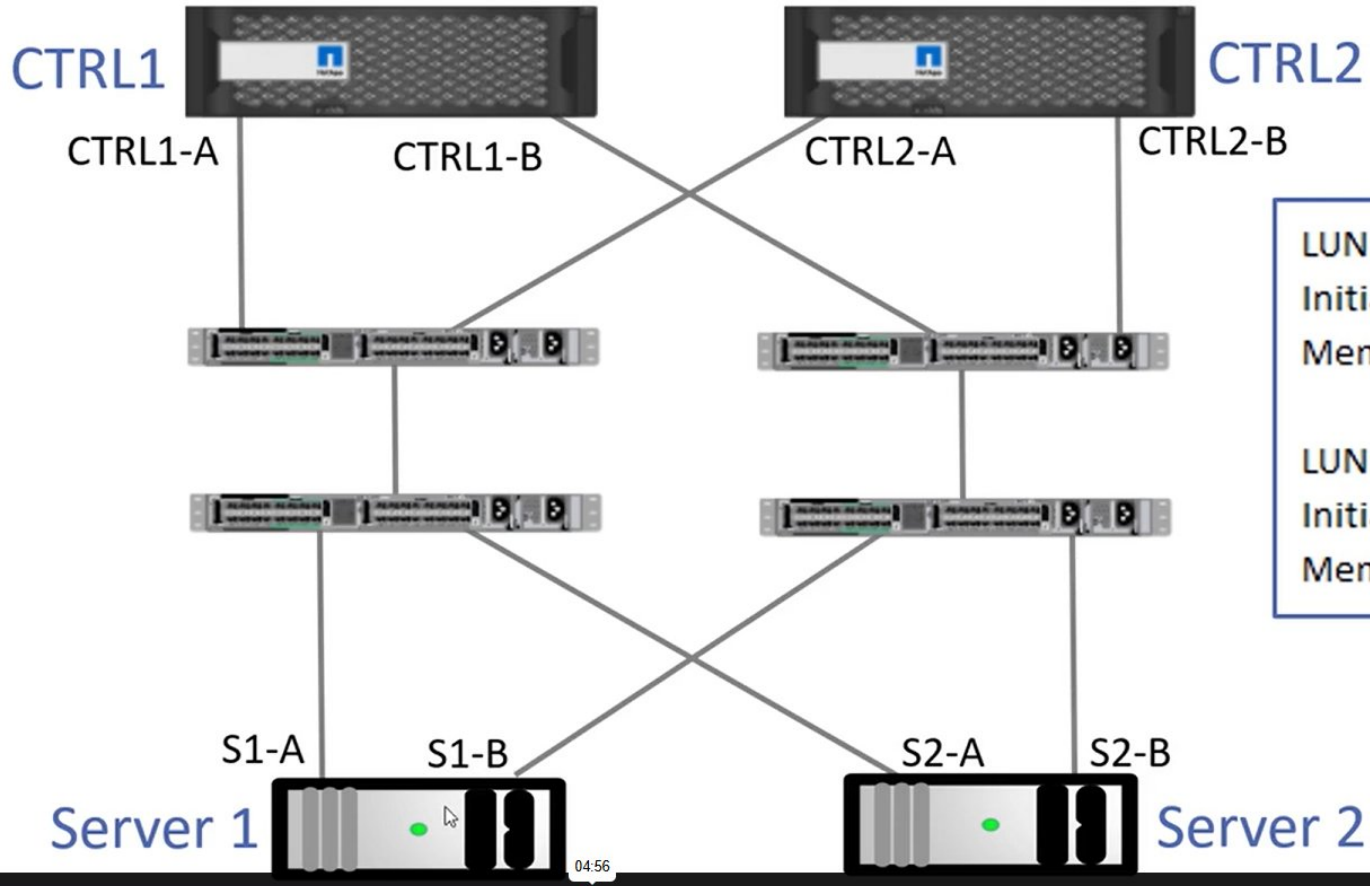


Zoning on Fabric A Switches



LUN Masking on Storage System

FLACKBOX
www.flackbox.com

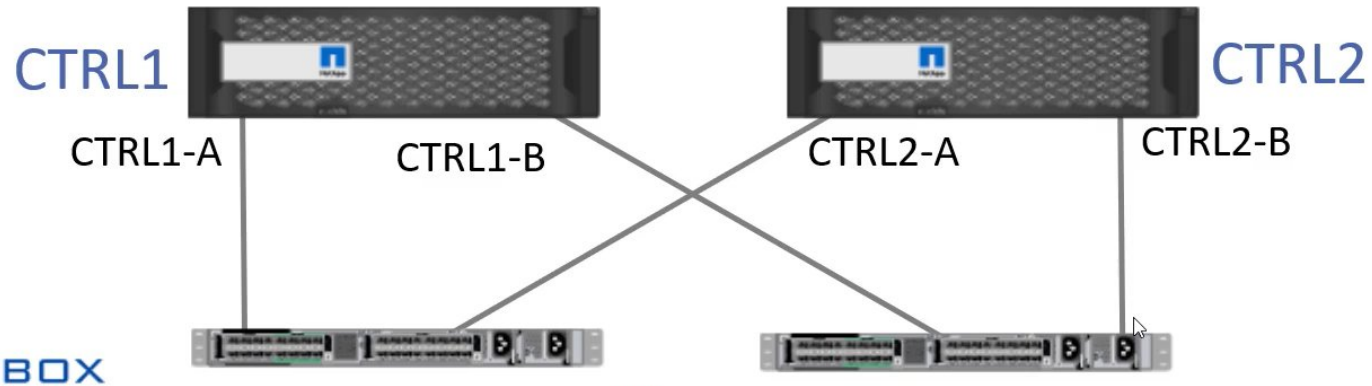


LUN: Server-1-Boot
Initiator Group: Server-1
Members: S1-A and S1-B

LUN: Server-2-Boot
Initiator Group: Server-2
Members: S2-A and S2-B

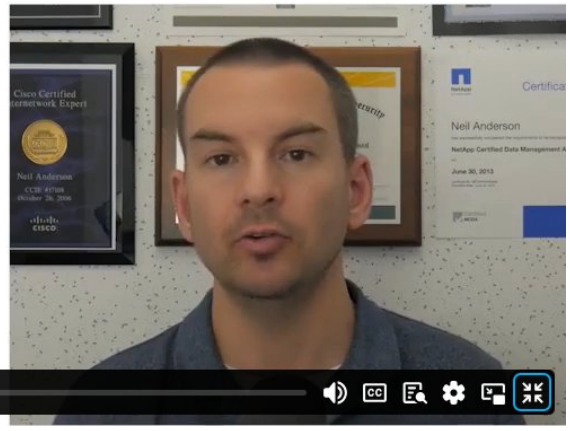
Target Portal Groups

- All of the ports on the storage system which initiators can access their storage through are added to a Target Portal Group
- In our example, ports CTRL1-A, CTRL1-B, CTRL2-A and CTRL2-B are added to a Target Portal Group



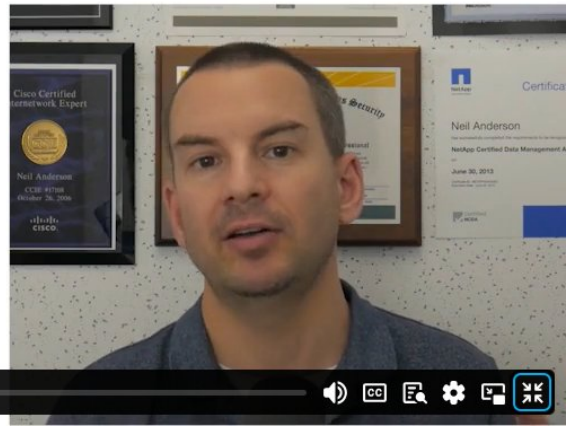
ALUA Asymmetric Logical Unit Assignment

- ALUA Asymmetric Logical Unit Assignment is used by the storage system to tell the client which are the preferred paths for it to use
- Direct paths to nodes owning the LUN are marked as optimized paths
- Other paths are marked as non-optimized paths



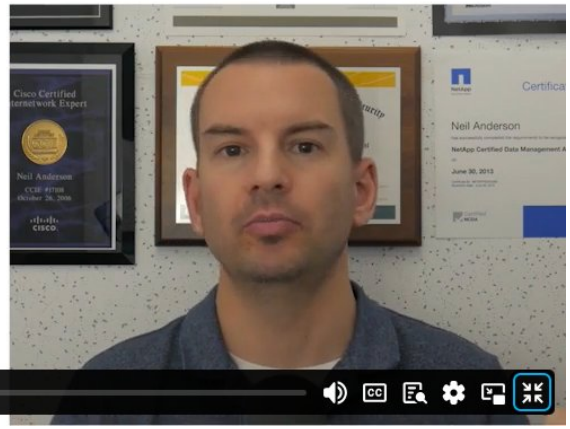
Multipathing

- During the Process Login, initiators will detect ports available to connect to their storage on in the Target Portal Group, and ALUA will notify which are the preferred paths
- Multipathing software on the initiator will choose which path or paths to take to the storage



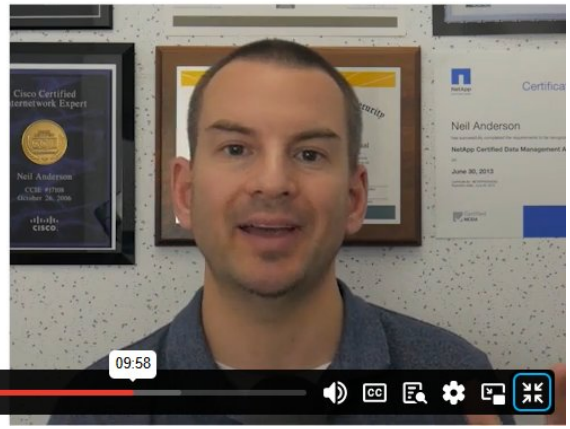
Multipathing

- All popular operating systems have multipathing software which supports active/active or active/standby paths
- The client will automatically fail over to an alternate path if the one it is using fails



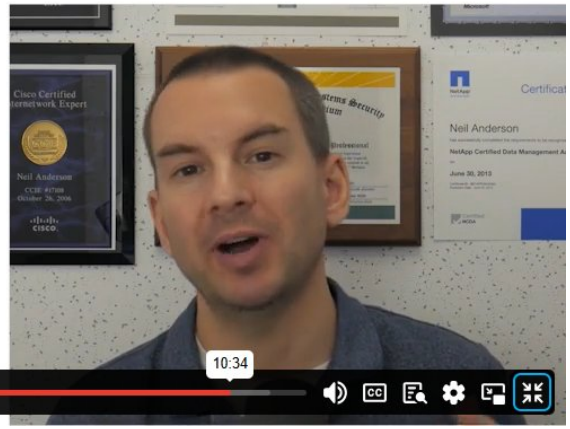
Multipathing

- Client connectivity to SAN storage is fundamentally different to how Ethernet networking works
- In Ethernet networking, all the routing and switching decisions are handled by network infrastructure devices
- In SAN storage, multipathing intelligence is enabled on the client end hosts



Fibre Channel Multipathing

- In Fibre Channel, the initiator will automatically detect the available paths to its storage through the FLOGI, PLOGI and PRLI process
- Multipathing software on the initiator will then choose which path or paths to use

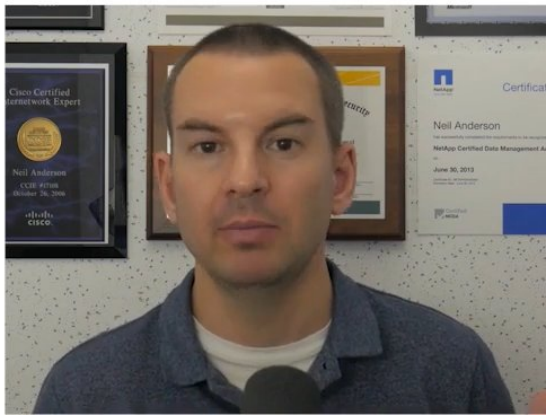


SAN Protocols – Fibre Channel over Ethernet (FCoE)

- Fibre Channel over Ethernet (FCoE) became possible with the advent of 10Gbps Ethernet, which has enough bandwidth to support both data and storage traffic on the same physical network infrastructure

SAN Protocols – Fibre Channel over Ethernet (FCoE)

- **FCoE uses the Fibre Channel Protocol, encapsulated in an Ethernet header to run over Ethernet Networks**
- QoS is used to guarantee the required bandwidth to the storage traffic
- It retains the reliability and performance of Fibre Channel

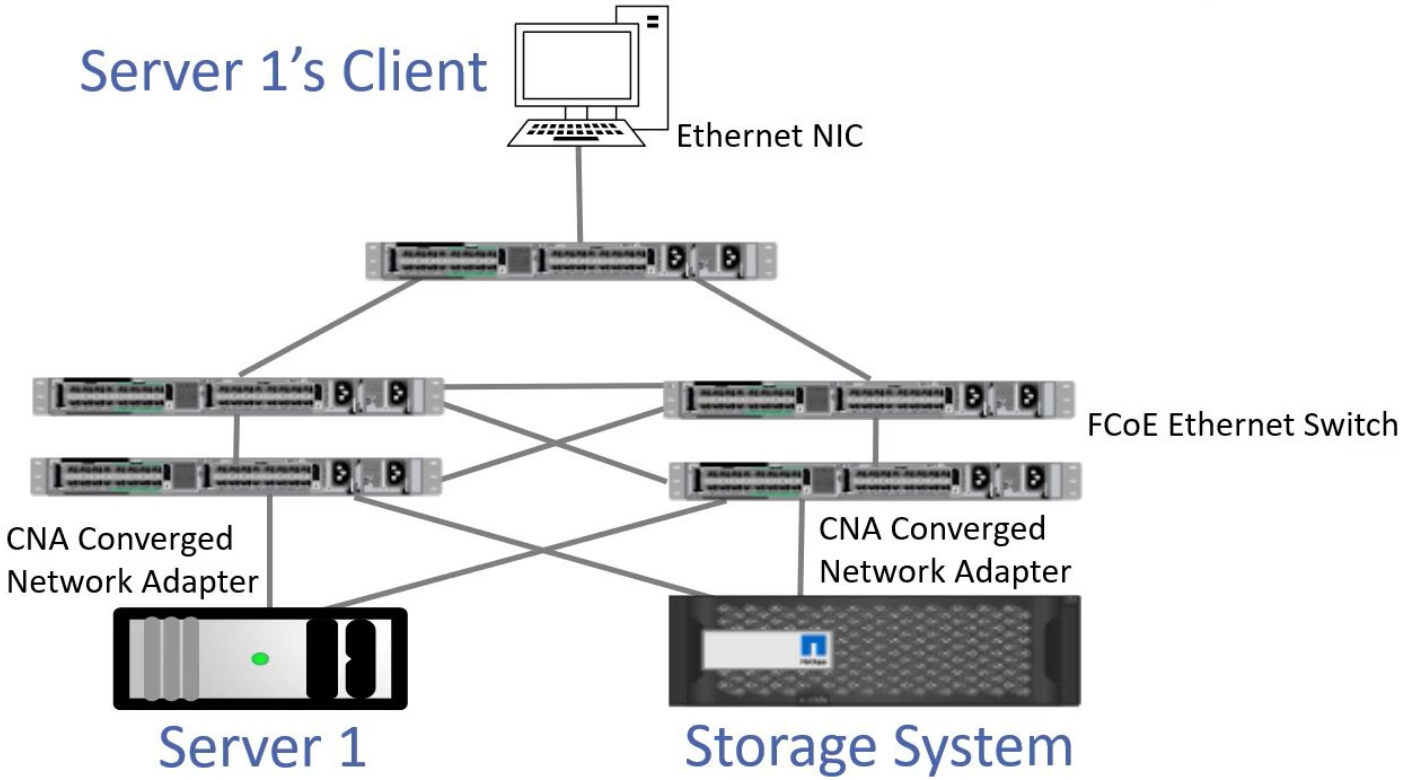


FLACKBOX
www.flackbox.com

FCoE Networks

FLACKBOX
www.flackbox.com

Server to Storage
and
Server to Client
uses shared 10Gbps
Ethernet Network

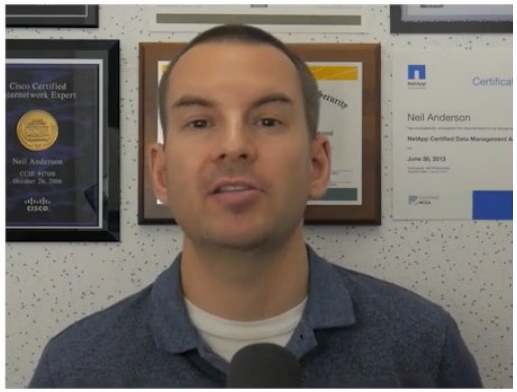


FCoE Operation

- Fibre Channel over Ethernet (FCoE) works exactly the same way as native Fibre Channel
- FCP works the same as when it's travelling over a native Fibre Channel network, just encapsulated in an Ethernet header so it can travel over an Ethernet network
- We still have WWPNs on the initiator and target, and use the FLOGI, PLOGI and PRLI process

FCoE Operation

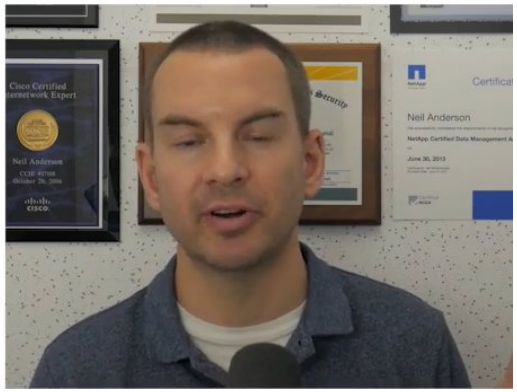
- In FCoE both the storage and the data traffic is going out the same shared physical interface
- The storage traffic uses FCP so it requires a WWPN
- The Ethernet data traffic requires a MAC address
- The way that Ethernet data traffic and FCP storage traffic works is totally different so how can we support them both on the same physical interface?

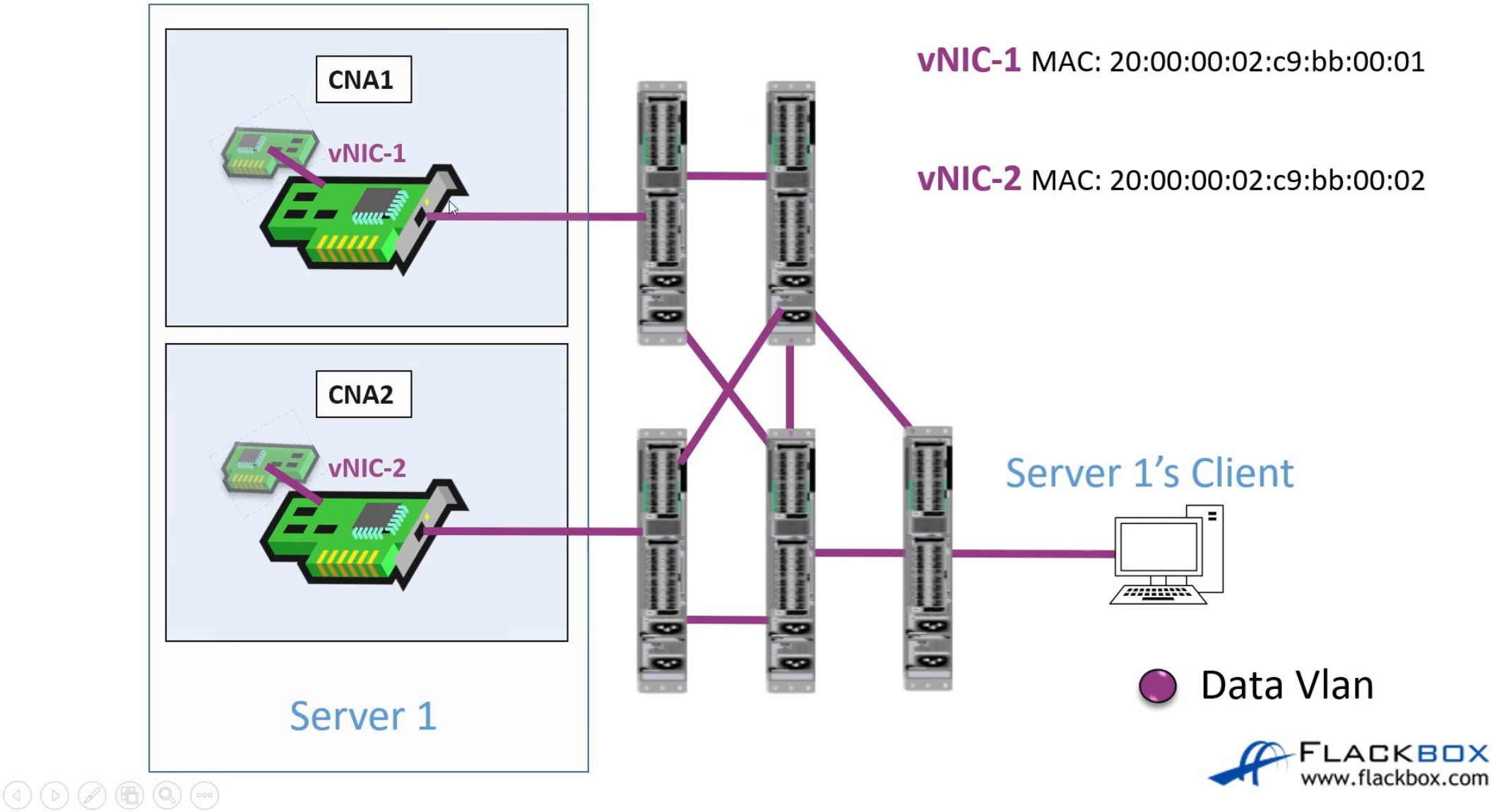


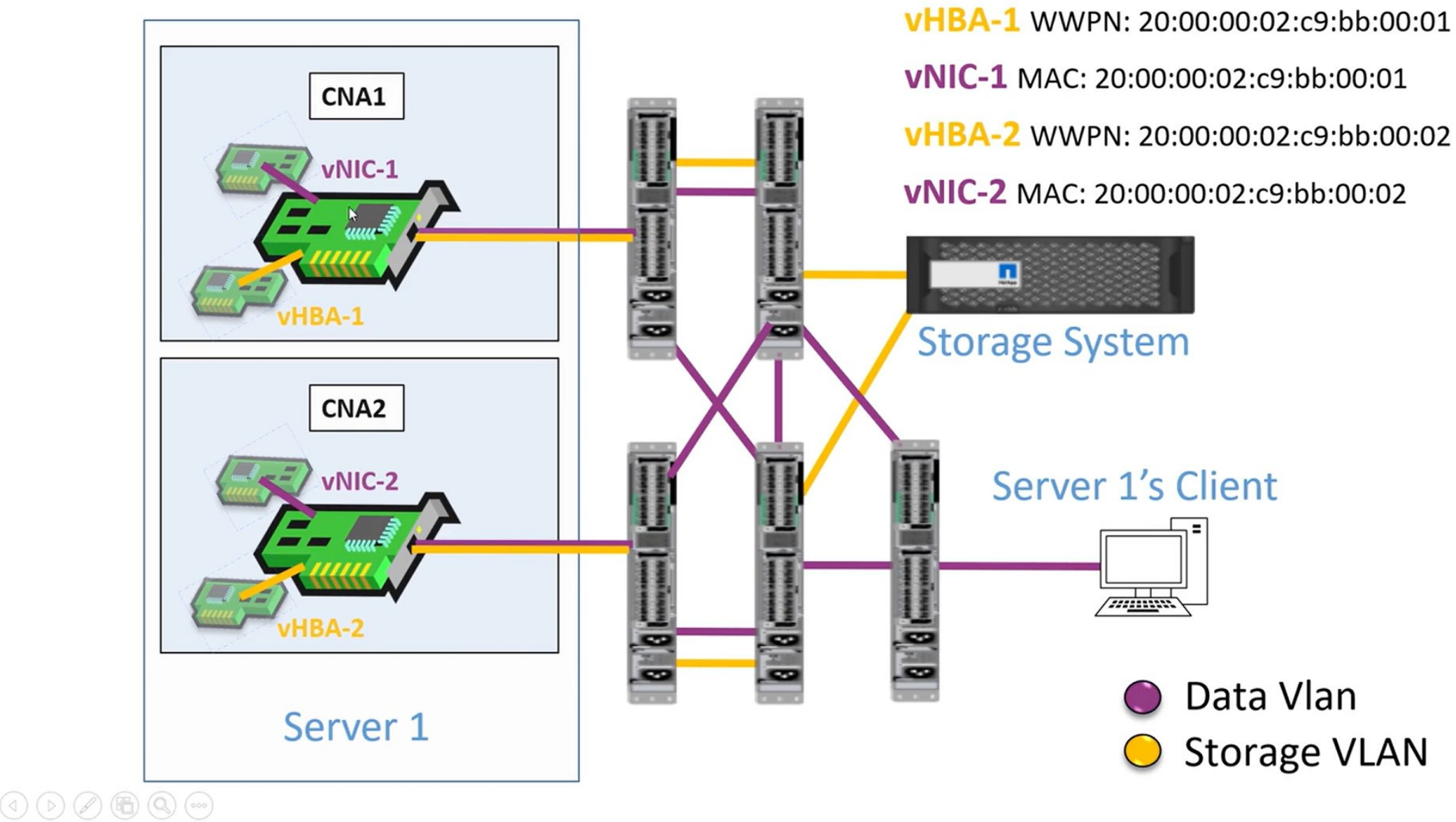
FCoE Operation

- Answer - We virtualize the physical interface into two virtual interfaces:
- A virtual NIC with a MAC address for the Ethernet data traffic
- A virtual HBA with a WWPN for the storage traffic

- The storage and data traffic are split into two different VLANs




CNA1
vNIC-1
CNA2
vNIC-2
Server 1
vNIC-1 MAC: 20:00:00:02:c9:bb:00:01
vNIC-2 MAC: 20:00:00:02:c9:bb:00:02
Server 1's Client
Data Vlan
FLACKBOX
www.flackbox.com


vHBA-1 WWPN: 20:00:00:02:c9:bb:00:01
vNIC-1 MAC: 20:00:00:02:c9:bb:00:01
vHBA-2 WWPN: 20:00:00:02:c9:bb:00:02
vNIC-2 MAC: 20:00:00:02:c9:bb:00:02
CNA1
vNIC-1
vHBA-1
CNA2
vNIC-2
vHBA-2
Server 1
Storage System
Server 1's Client
Data Vlan
Storage VLAN

Lossless FCoE

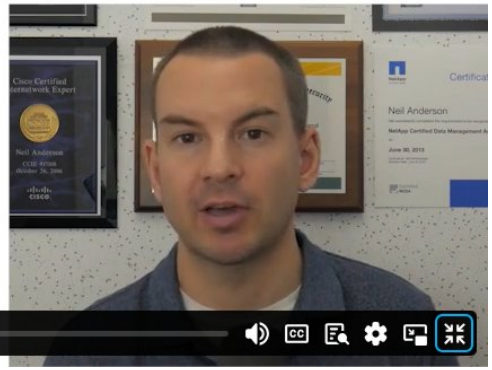
- Fibre Channel is a lossless protocol, it ensures that no frames are lost in transit between the initiator and target
- Ethernet is not lossless. TCP uses acknowledgements from the receiver back to the sender to check traffic reaches it's destination. If an acknowledgement is not received, the packet will be resent.
- FCoE uses FCP which assumes a lossless network, so we need a way to ensure our storage packets are not lost while traversing the Ethernet network

Lossless FCoE

- The PFC Priority Flow Control FCoE extension for Ethernet is used to ensure lossless delivery
- PFC works on a hop by hop basis
- Each NIC and switch in the path between initiator and target must be FCoE capable
- FCoE capable NICs are known as CNA Converged Network Adapters

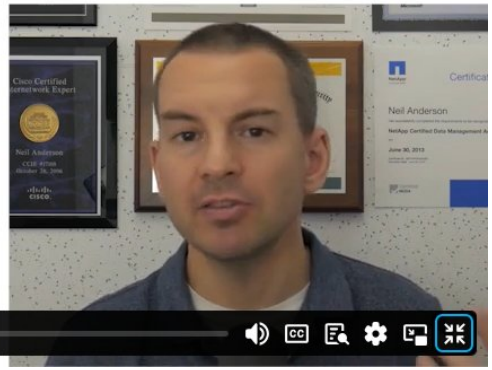
SAN Protocols - iSCSI

- iSCSI is the Internet Small Computer System Interface protocol
- It runs over Ethernet networks and was originally viewed as a less expensive alternative to Fibre Channel
- It has higher packet overhead than and has traditionally been seen as having lower reliability and performance than Fibre Channel
- It is a very mature and popular SAN technology



SAN Protocols - iSCSI

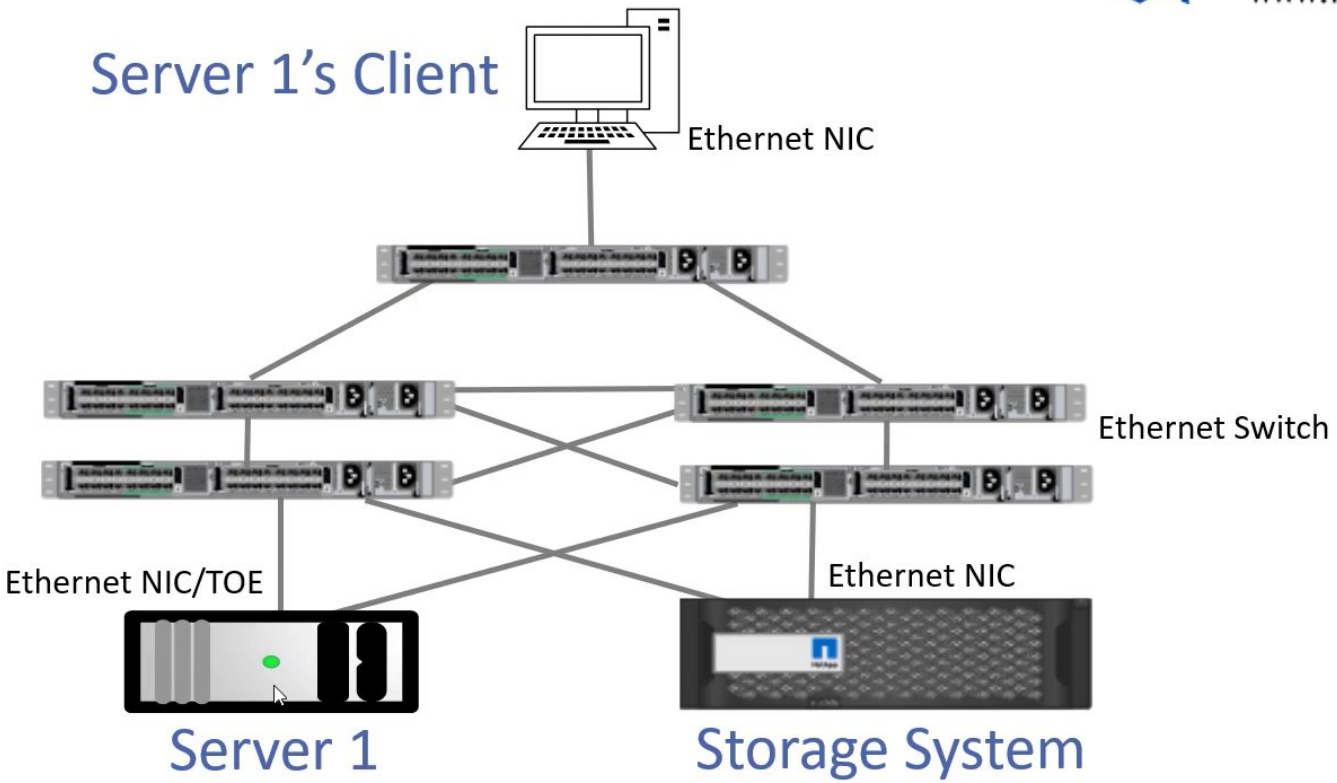
- As it runs over Ethernet, it can share the data network or have its own dedicated network infrastructure
- TOE (TCP Offload Engine) cards are specialist adapters which can be used to offload the storage TCP/IP processing from a server's CPU. They are sometimes called iSCSI HBAs.



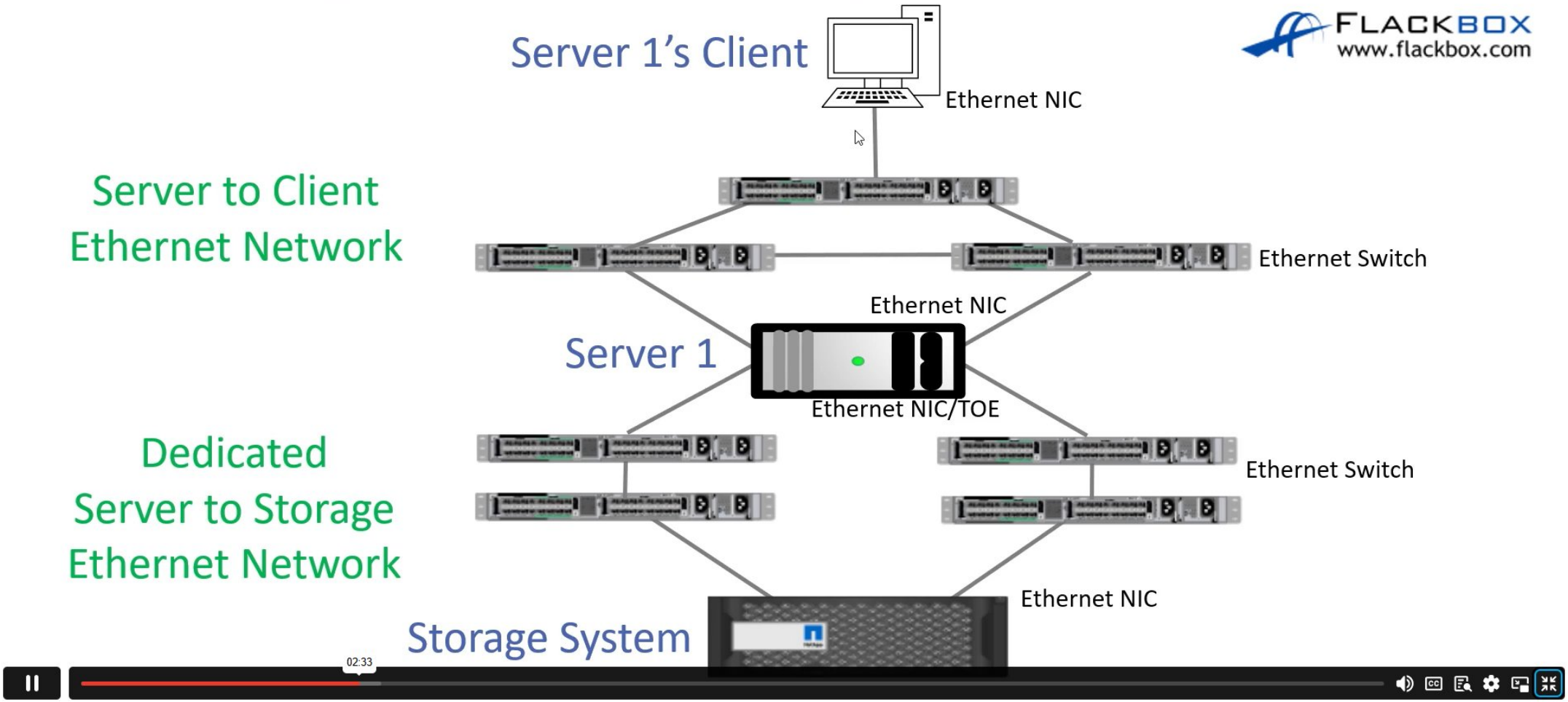
iSCSI Networks – Shared Network

FLACKBOX
www.flackbox.com

Server to Storage
and
Server to Client
uses shared
Ethernet Network



iSCSI Networks – Dedicated Storage Network

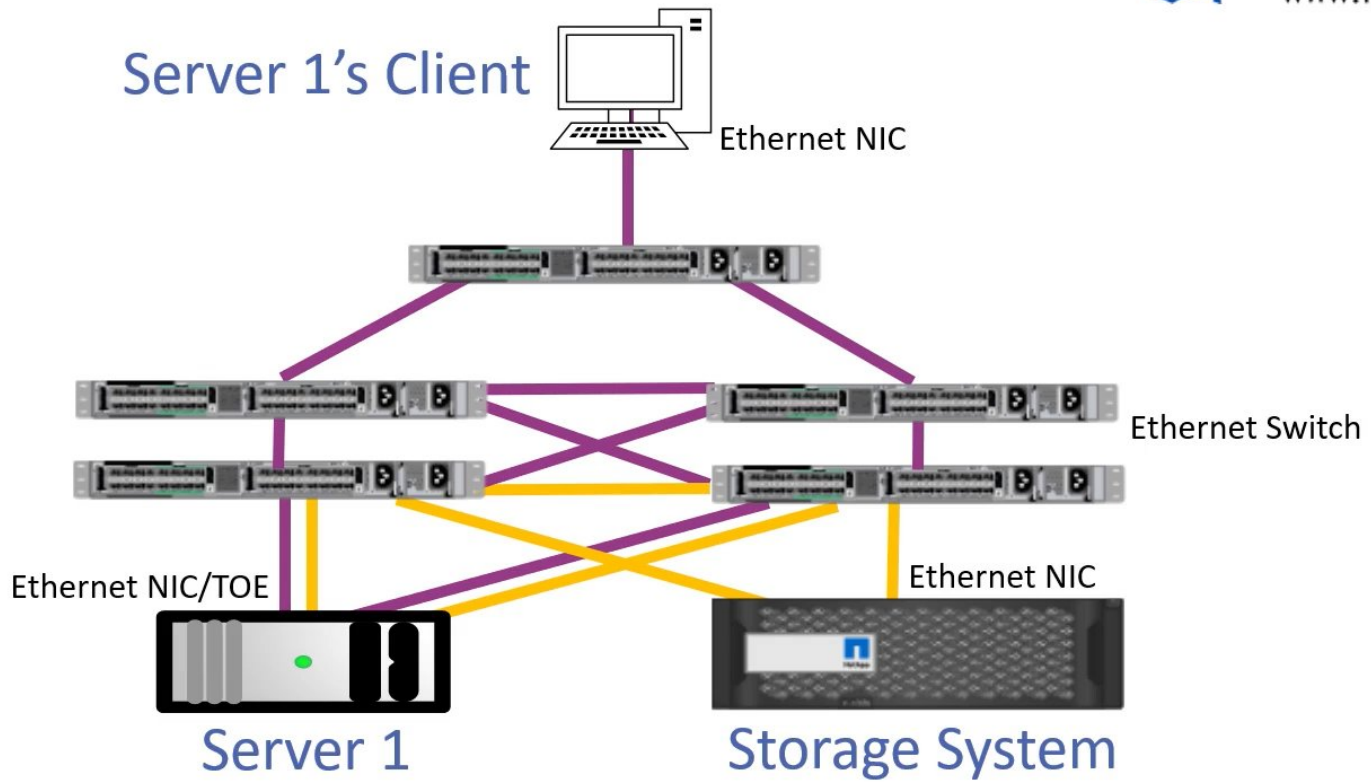


iSCSI Networks – Dedicated VLAN Network

FLACKBOX
www.flackbox.com

Server to Storage
and
Server to Client
uses shared
switches, but traffic
is split into
Separate VLANs

- Data Vlan
- Storage VLAN



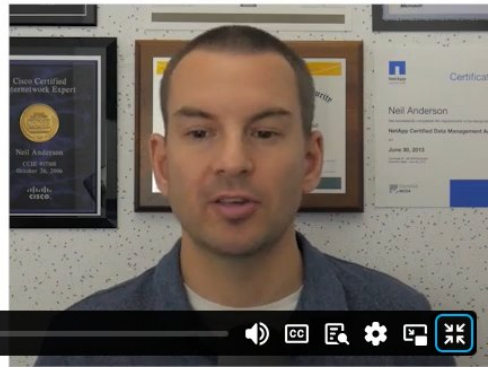
iSCSI Addressing

- Fibre Channel uses World Wide Names to identify initiators and targets
- iSCSI uses IQN iSCSI Qualified Names (or less commonly EUI Extended Unique Identifier) for its addressing
- The IQN can be up to 255 characters long and has the following format:

  iqn.yyyy-mm.naming-authority:unique name
- For example iqn.1991- 05.com.microsoft:testHost

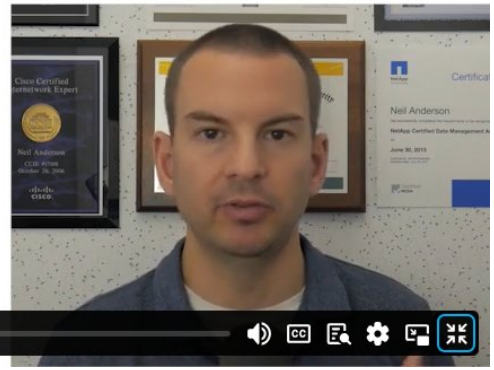
iSCSI Addressing

- The IQN is assigned to the host as a whole, similar to the WWNN in Fibre Channel
- iSCSI runs over Ethernet so individual ports are addressed by IP address



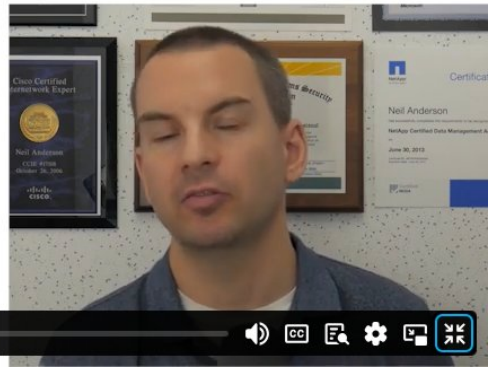
iSCSI Addressing

- iSCSI does not support Fibre Channels FLOGI/PLOGI/PLRI process so an administrator must explicitly point the initiator at its target by specifying one of the IP addresses in the Target Portal Group
- It will then discover the target's IQN and the other ports in the TPG



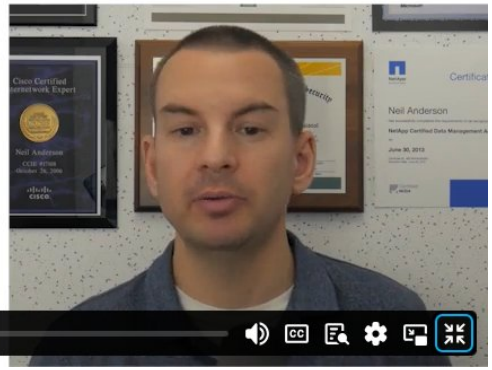
iSCSI Addressing

- Multipathing software on the initiator can then choose which path or paths to take
- Although it runs over Ethernet, iSCSI is still a SAN protocol with multipathing intelligence on the initiator



iSCSI Security

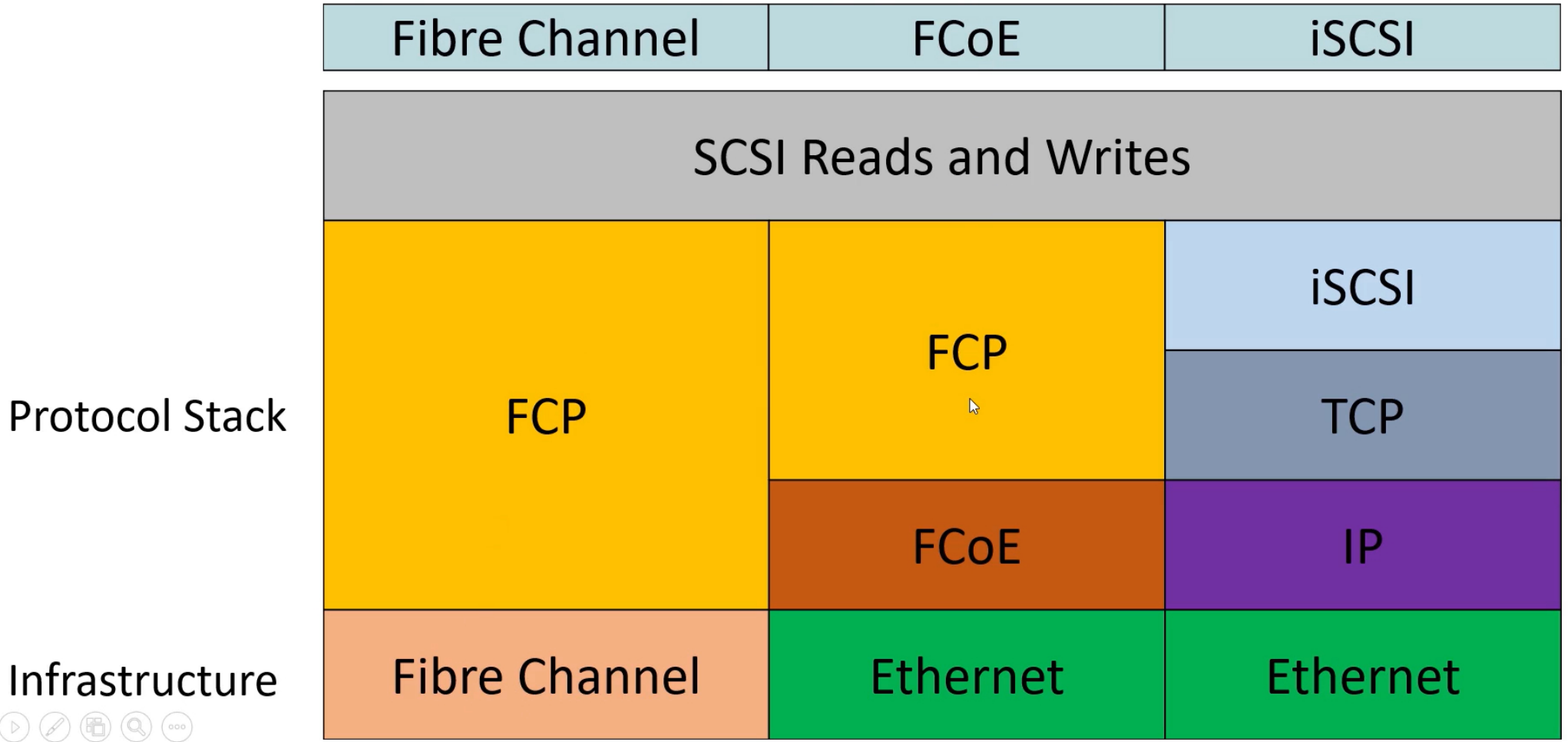
- LUN Masking is configured in the same way as in Fibre Channel, using the IQN rather than WWPN on the storage system to identify the client
- Zoning is not supported in iSCSI
- Password based authentication is typically configured on the initiator and target to guard against spoofing attacks
- End to end IPSec encryption can also be enabled to enhance security



SAN Adapter Summary

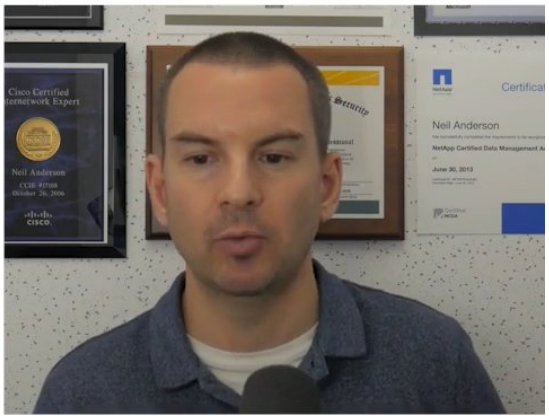
- NIC: Network Interface Adapter. Traditional Ethernet network card, it is used for NAS protocols and iSCSI.
- TOE: TCP Offload Engine. Used to offload the TCP/IP processing from a server's CPU, it can enhance performance for NAS protocols and iSCSI.
- HBA: Host Bus Adapter. Fibre Channel equivalent of a NIC.
- iSCSI HBA: Ethernet TOE card optimized for iSCSI.
- CNA: Converged Network Adapter. 10Gb Ethernet card which supports FCoE.
- UTA: Universal Target Adapter. NetApp proprietary card which supports FCoE or Fibre Channel.

SAN Protocol Stack Comparison



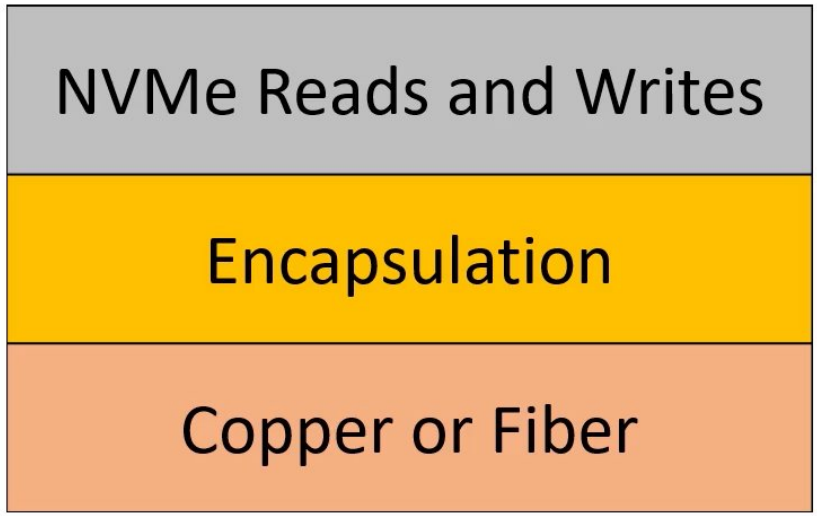
NVMeOF (NVMe Over Fabrics)

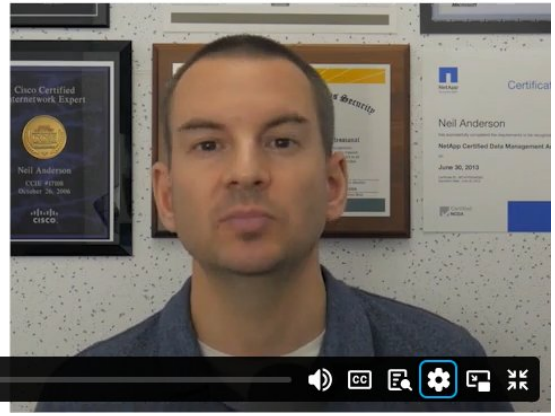
- NVMeOF (NVMe Over Fabrics) allows the NVMe command set to be used for block access over a network, providing lower latency over the wire than traditional block access protocols.



FLACKBOX
www.flackbox.com

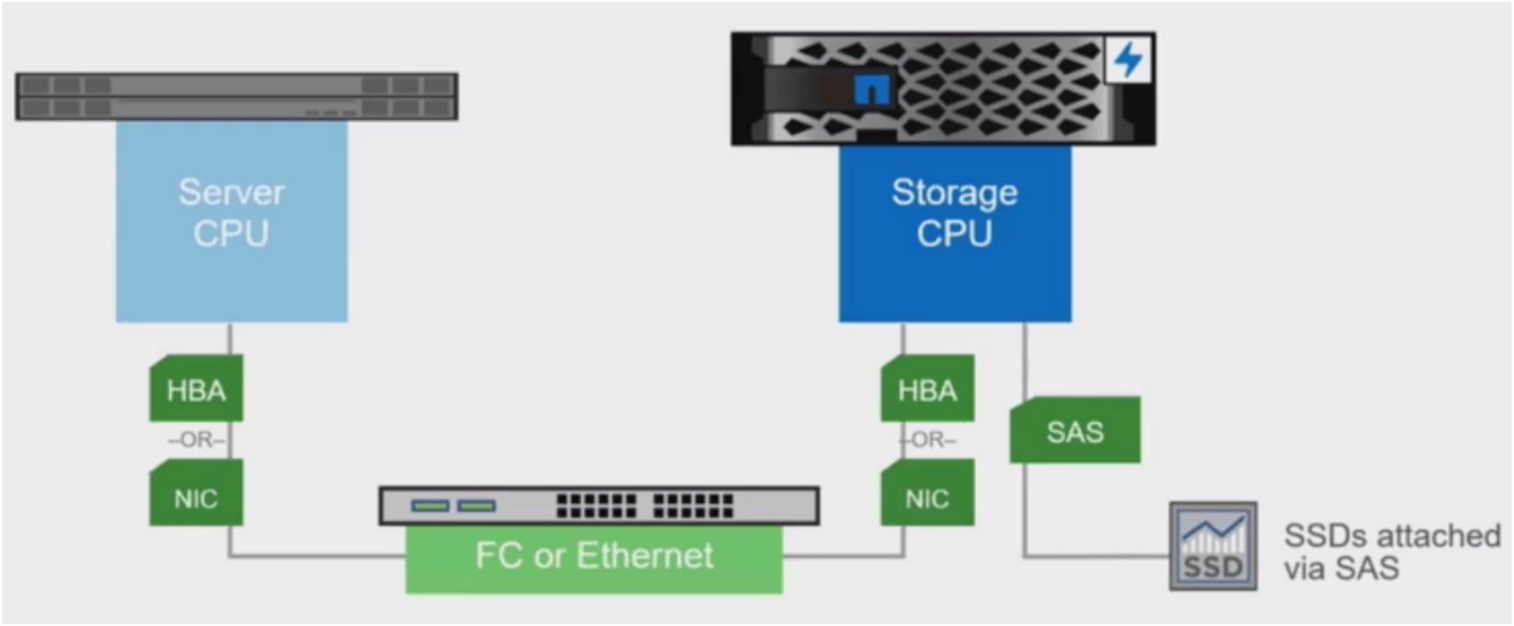
NVMeOF Protocol Stack

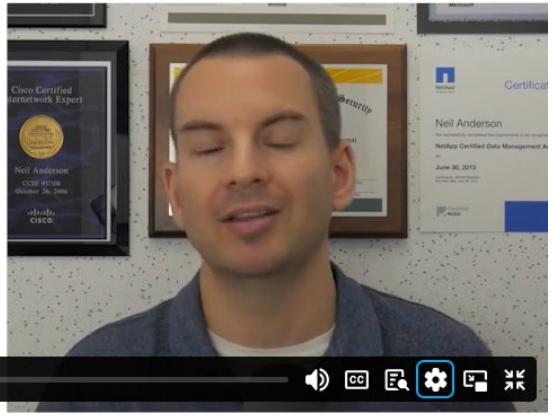




02:45

NVMeOF Front End, SAS Back End

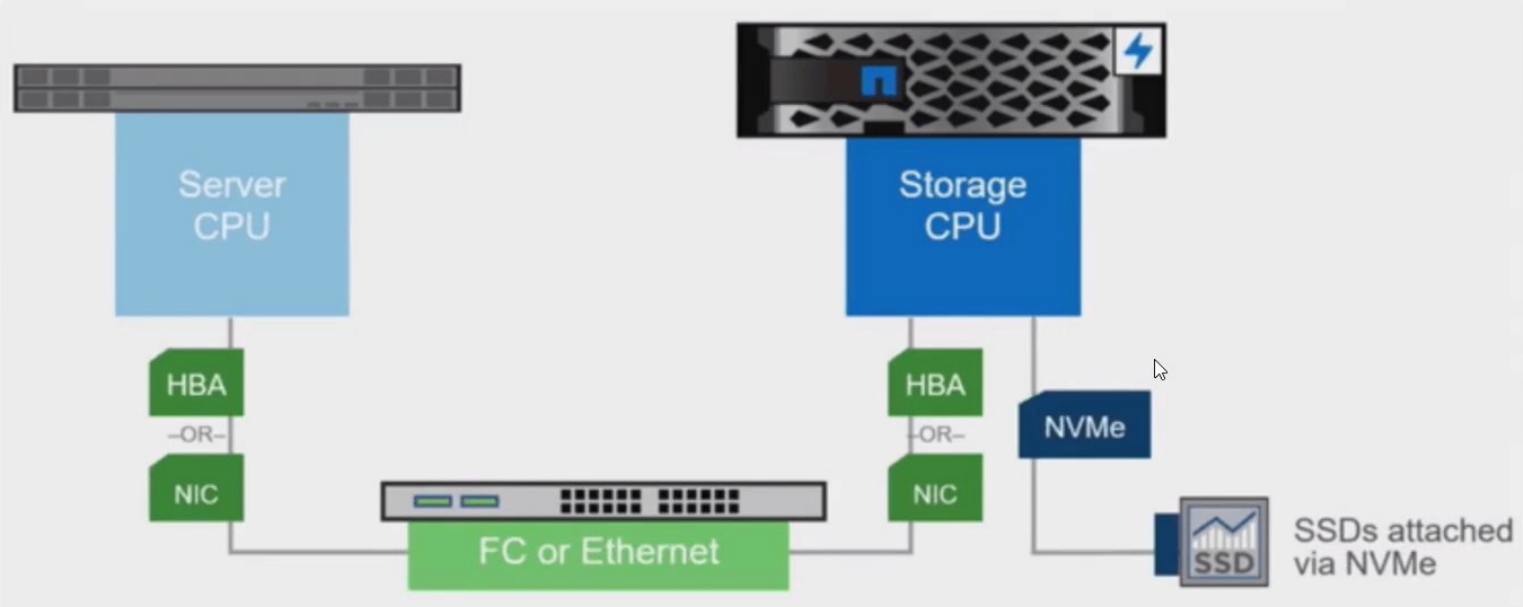


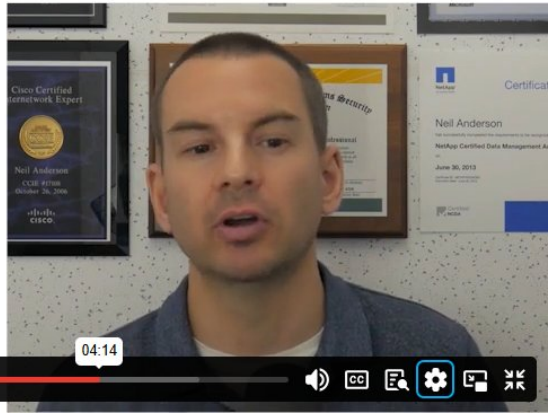


FLACKBOX
www.flackbox.com

03:05

End-to-end NVMeOF





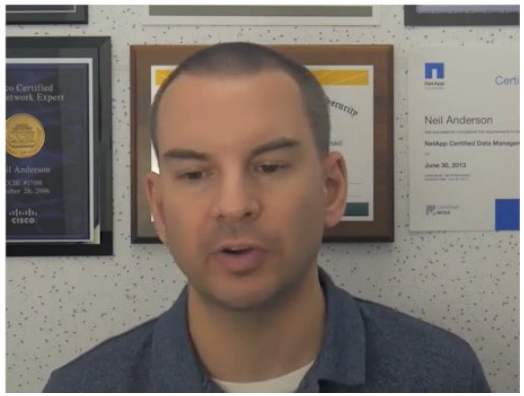
FLACKBOX
www.flackbox.com

Block Storage

- Block storage stores and manages data in blocks which are accessed via low-level storage protocols using SCSI and NVMe commands.
- SAN protocols provide block access over a network. The experience for the user and applications is similar to accessing a local disk.

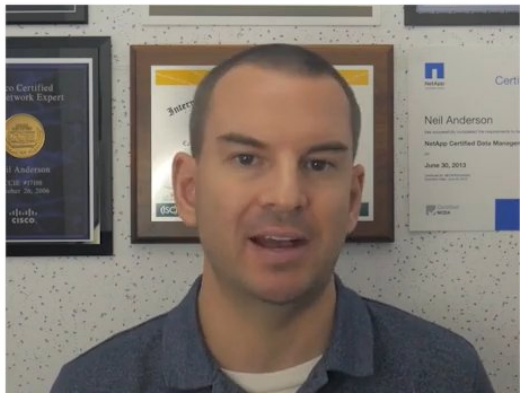
Block Storage

- The direct access to the data reduces overhead by minimizing abstraction layers.
- Higher level tasks such as multi-user access, sharing, locking and security are usually handled by the operating system.



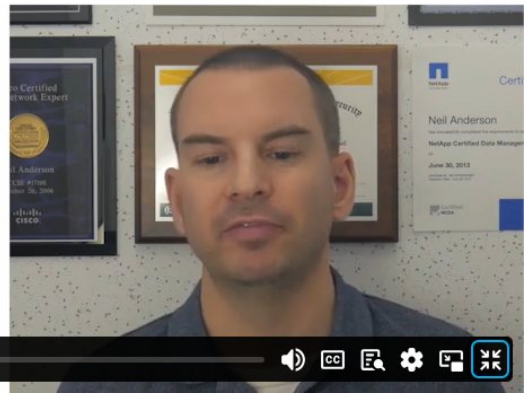
Block Storage Metadata

- There is no storage-side metadata associated with the block, only its address.
- The block is simply a chunk of data that has no description, no association and no owner.



Block Storage Use Cases

- Block storage is considered the best solution for performance sensitive, transactional, and database oriented applications.
- It is mostly used for primary storage and with the client and storage system both located in the same physical location. Adding distance between the application and storage harms performance.



File Storage

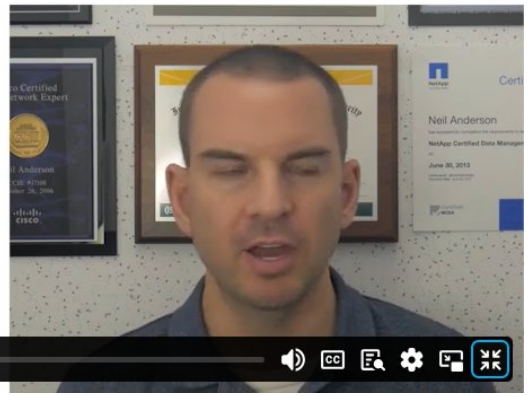
- File storage, used by NAS protocols, stores data as a file hierarchy in a file system.
- The hierarchy is similar to a physical file cabinet with folders (or ‘directories’) and subfolders.
- The user or application connects to the file system through a share (CIFS/SMB) or by mounting an export (NFS).

File Storage Metadata

- File system metadata is recorded separately from the file itself and records basic file attributes such as file name, creation date, creator, file type, most recent change and last access.
- The metadata is fixed and standardized to the file system.
- Adding custom metadata (extended attributes) requires a custom application and database.

File Storage Use Cases

- File storage is well suited to general purpose data, especially data which is edited frequently and concurrently by multiple users or applications.
- It is designed to be accessed over both the local network and remotely.



Block and File Storage Limitations

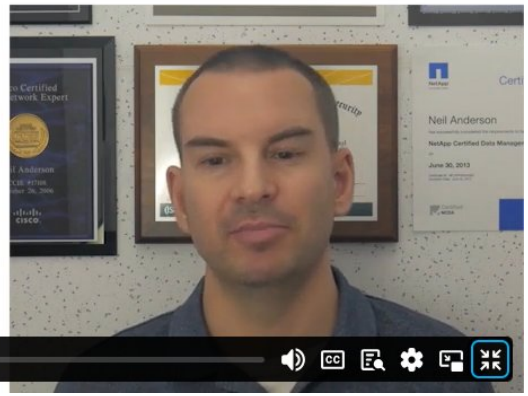
- Block and file storage systems can be scaled out by adding more disks and nodes but they are typically limited in scale to a single geographic location physically.
- NAS file index tables (inodes) have a maximum size and can affect performance if they grow too large.
- Block and file stores need to be backed up offsite for resiliency.

Object Storage

- Object storage organizes information into containers of flexible sizes, referred to as objects.
- Objects are stored and managed in a flat organisation of flexibly sized containers called buckets or containers.
- Buckets can span multiple nodes and geographic locations.

Object Storage Scalability

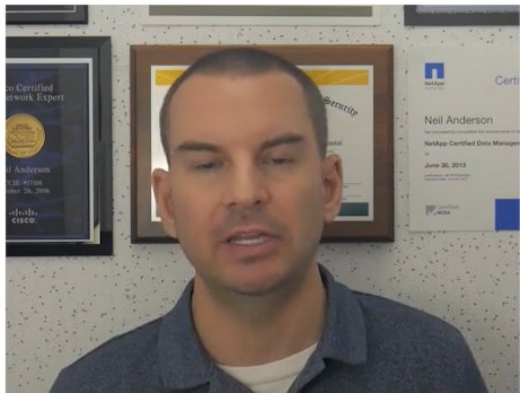
- Object based storage architectures can be scaled out and managed simply by adding additional nodes ***which can be across multiple locations***.
- It is very commonly offered by cloud providers, but on-premises and hybrid solutions (where objects can be stored both on-premises and in the cloud) are also available.



Object Storage

player.vimeo.com – To exit full screen, press Esc

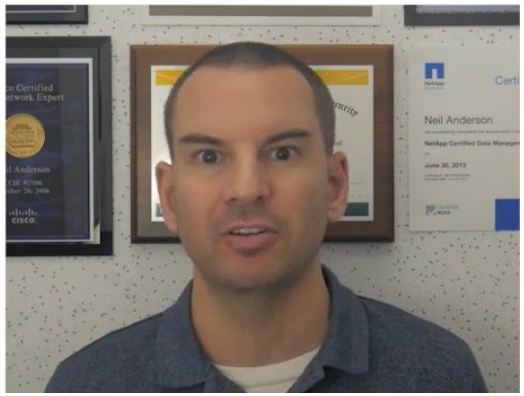
- Every object contains three things:
  - The data itself.
  - Customisable metadata.
  - A globally unique identifier.



FLACKBOX
www.flackbox.com

Object Storage Metadata

- ***Customisable*** metadata links directly with the object and contains additional descriptive properties.
- This can be used for better indexing, for example by including ‘black’ and ‘cat’ tags in a video object, or the patient’s name and injury type in an x-ray.

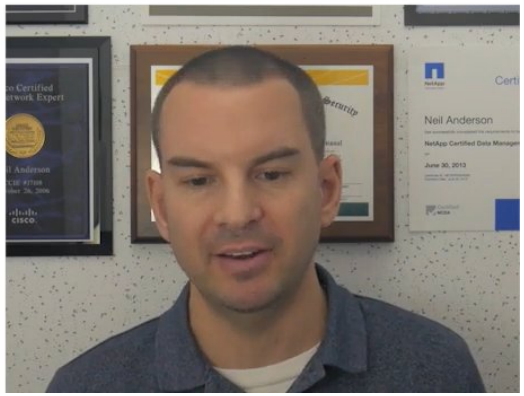


Object Storage Metadata (Cont.)

- It can also be used for better management, for example indicating replication instructions, when to move to a different tier of storage or location, and when to delete.

Globally Unique Identifier

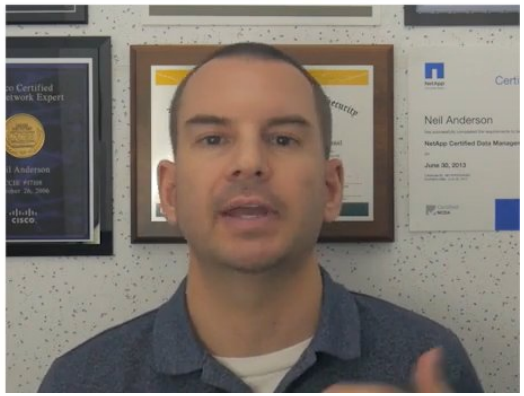
- A globally unique identifier is used instead of a file name and path as in NAS.
- This address is unique across the namespace, and is used to find the object over the distributed system without having to know the physical location of the data.
- This removes the complexity and scalability challenges of a NAS hierarchical file system based on complex file paths.



FLACKBOX
www.flackbox.com

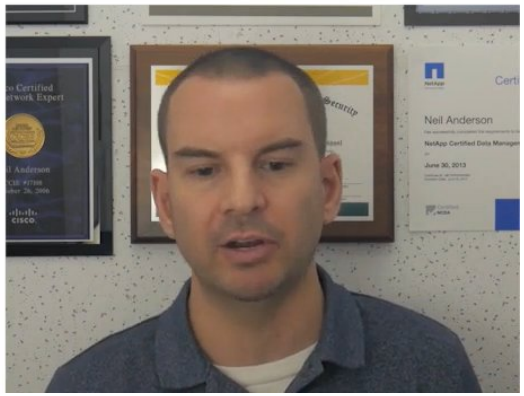
Data Protection

- Object Stores provide resiliency through replication and/or erasure coding.
- Replication is used to store objects on multiple nodes which can be geographically dispersed. This is most suitable for small files.



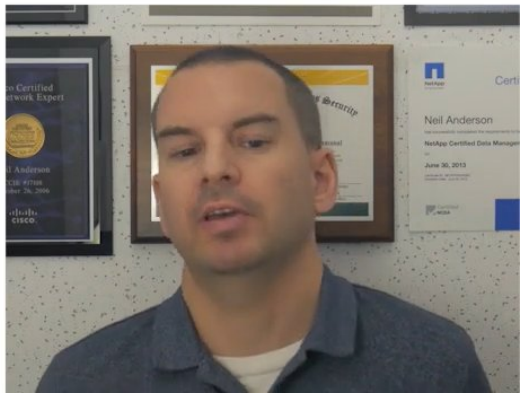
Data Protection (Cont.)

- With erasure coding the object is broken up into smaller distributed parts. Parity information allows the data to be reconstructed if there is node failure. This is suitable for larger files.
- If one or more nodes fail, the data can still be made available without the application or the end user being impacted.



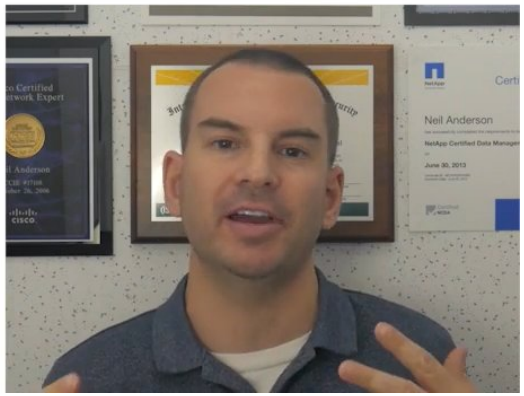
Object Storage Versioning

- Object Storage supports versioning, where the old version of the object can be automatically saved if it is changed.
- It does not typically support file locking and files can't be updated in place (this makes the simple data protection possible).
- If multiple users update the same object concurrently, the system will simply write different versions of the object.



Object Storage Versioning (Cont.)

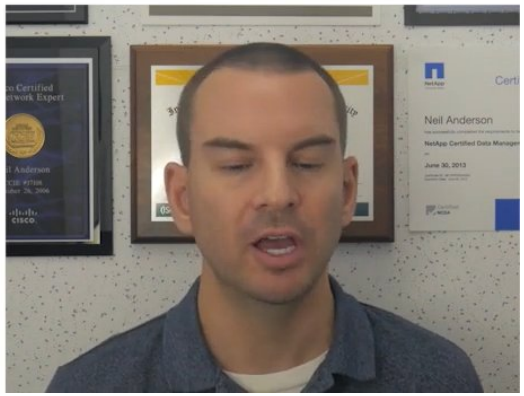
- It is designed for Puts and Gets, not for data which will have multiple edits from multiple users such as transactional databases or Word documents.
- This has traditionally made it more suitable for secondary (backup and archive) rather than primary data.



FLACKBOX
www.flackbox.com

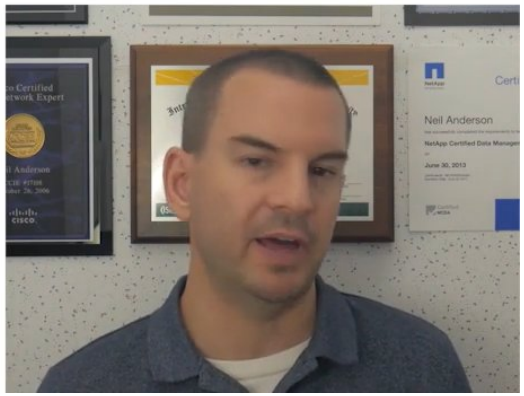
Cost

- Object storage is typically used as secondary storage where performance is not a priority.
- It is usually one of the lowest cost storage options from a cloud provider.
- On-premises Object Storage platforms can typically be bought as an appliance or software only.
- Lower cost commodity hardware can be used for the underlying infrastructure on a software only purchase.



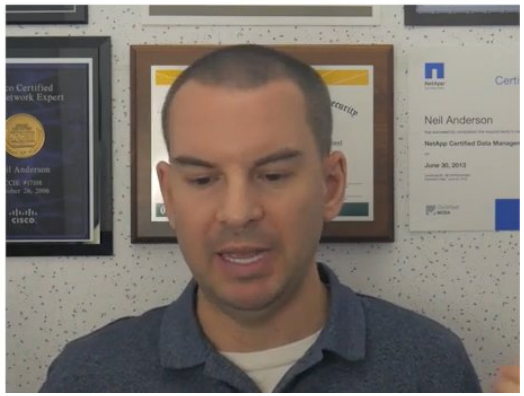
Object Storage Protocols

- Object Storage uses RESTful APIs that use HTTP requests to GET, PUT, POST and DELETE data, and HEAD to request metadata information.
- Open Standard API examples:
  - CDMI, the Cloud Data Management Interface from the Storage Networking Industry Association (SNIA)
  - Amazon Web Services Simple Storage Service (S3)
  - OpenStack Swift



Object Storage Protocols

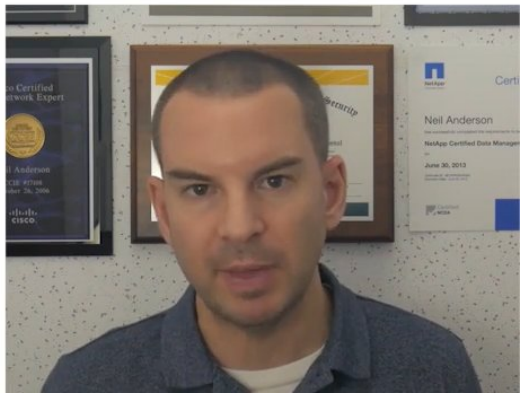
- End users can access the storage directly via a web interface or application such as Cyberduck.
- Support for access via NAS protocols is typically also included, either natively or by using a cloud gateway.
- The cloud gateway can be a physical appliance or virtual machine.



FLACKBOX
www.flackbox.com

Object Storage Benefits Summary

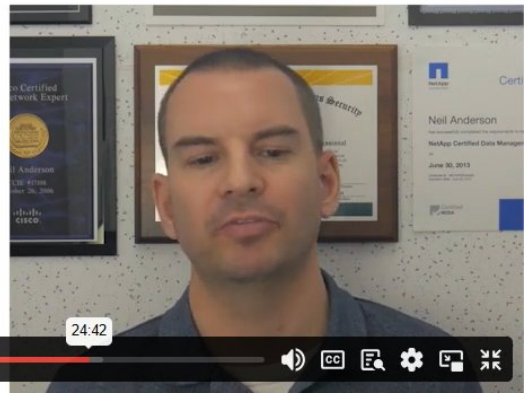
- Single namespace with almost infinite scale
- Scales across multiple physical locations
- Performance does not degrade with scale
- Customizable metadata for better indexing and management
- Supports data management functions such as replication at object-level granularity
- Typically low cost



FLACKBOX
www.flackbox.com

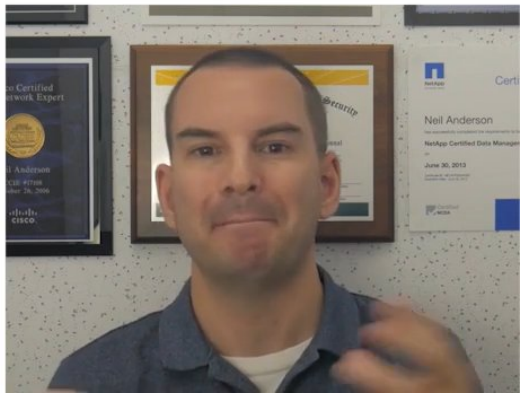
Object Storage Limitations Summary

- Generally lower performance, so not suitable for databases or other applications which require high performance.
- Does not have locking and file-sharing facilities so not suitable for data which may be accessed concurrently and changed by multiple users or applications.



Object Storage Use Cases

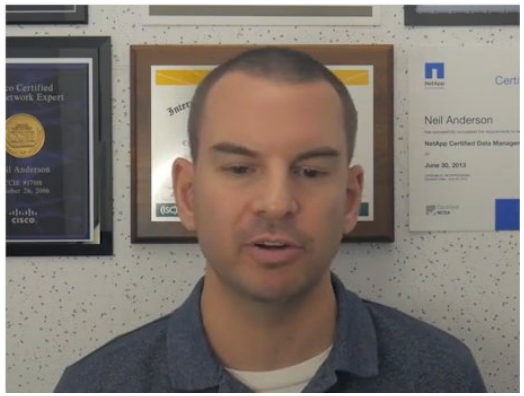
- Object storage is best suited as a massively scalable store for unstructured data that's updated infrequently.
- It can be used as an additional storage tier beyond transactional storage for inactive data, or as archival storage.
- As object storage evolves it may become more suited to primary data.
- It's well suited for file content in the cloud space, especially images and videos.



FLACKBOX
www.flackbox.com

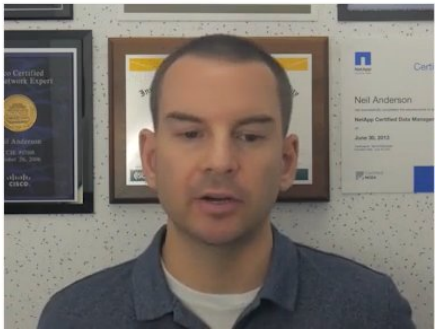
Object Storage Examples

- Cloud Providers:
  - Amazon Web Services Simple Storage Service (S3)
  - Microsoft Azure Blob Storage
- Proprietary:
  - Facebook Haystack
- On-Premises/Hybrid:
  - NetApp StorageGRID



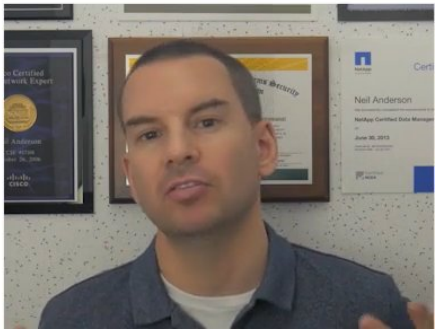
Traditional Data Center Deployments

- The main hardware building blocks of a data center are compute (the servers), networking and storage
- In a traditional deployment the organization's internal IT team will produce the data center design then procure, install and configure the equipment



Traditional Data Center Deployments

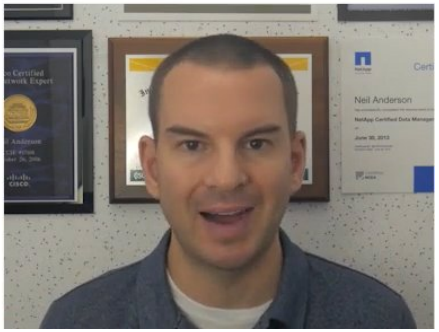
- The main hardware building blocks of a data center are compute (the servers), networking and storage
- In a traditional deployment the organization's internal IT team will produce the data center design then procure, install and configure the equipment



FLACKBOX
www.flackbox.com

Traditional Data Center Deployments (Cont.)

- The equipment will typically come from different vendors and the organization will typically have different teams involved in designing and managing the solution
- The initial deployment and later addition of additional hardware is time consuming



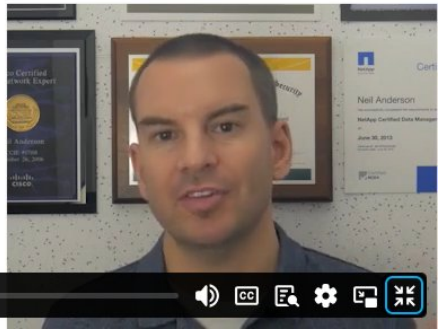
FLACKBOX
www.flackbox.com

Converged Infrastructure

- With Converged Infrastructure, different compute, storage and networking hardware vendors team together to produce validated designs
- The validated designs are best practice configurations for a specific workload
- Converged software management tools are also often included in the design

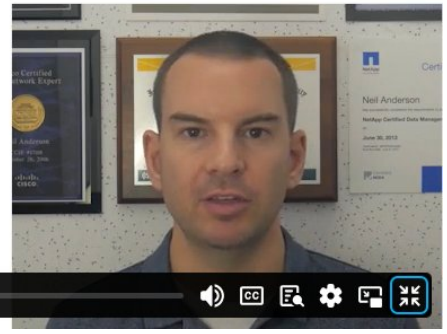
Converged Infrastructure (Cont.)

- The end result looks the same as an organisation's internal IT department might produce in a traditional solution
- Less solution architect expertise is required by the organisation



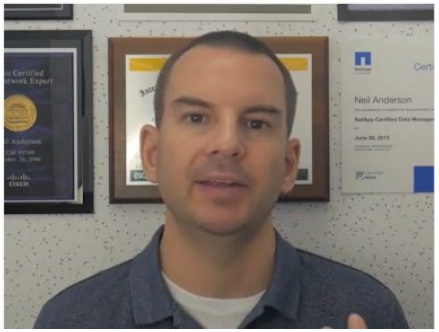
Converged Infrastructure (Cont.)

- The validated designs will be in the form of documentation which the organisation can deploy themselves, or pre-built systems which can be ordered
- The vendors may offer a single point of contact for support
- This makes it faster to purchase and deploy, and easier to run and support



Converged Infrastructure (Cont.)

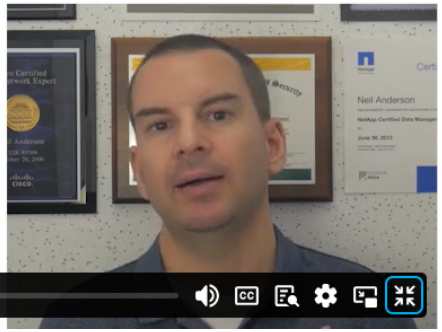
- Customisation is usually possible
- The solution can be scaled up through customisation, and scaled out by adding more racks



FLACKBOX
www.flackbox.com

Converged Infrastructure Examples

- FlexPod: Cisco and NetApp
- VxBlock: Cisco and Dell EMC
- HPE Converged: HPE and Cisco/Arista



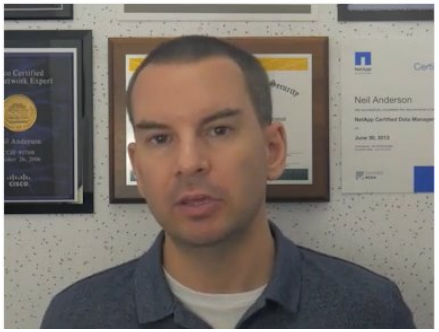
Hyper Converged Infrastructure HCI

- In HCI compute, storage and networking are software defined – they are virtualized resources pooled across all nodes and made available to all client applications.
- The integrated technologies can be managed as a single system through a common tool set.
- HCI can be sold as software that can be installed on a buyer's existing hardware or as hardware purchased specifically for the installation.

FLACKBOX
www.flackbox.com

Hyper Converged Infrastructure HCI

- In HCI hardware systems compute and storage are deployed in an all-in-one modular hardware chassis
- The solution can be scaled out by adding more chassis'
- As with Converged Infrastructure, they are fast to purchase and deploy, and easy to run and support



FLACKBOX
www.flackbox.com

Hyper Converged Infrastructure Examples

- Nutanix
- HPE Simplivity
- Cisco Hyperflex
- Dell EMC VxRail
- NetApp HCI

FLACKBOX
www.flackbox.com