

IT Infrastructure Architecture

Infrastructure Building Blocks
and Concepts

Storage

Tapes

- When storing large amounts of data, tape is the most inexpensive option
- Tapes are suitable for archiving
 - Tape manufacturers guarantee a long life expectancy
 - DLT, SDLT, and LTO Ultrium cartridges are guaranteed to be readable after 30 years on the shelf

Tapes

- Disadvantages:
 - Tapes are fragile
 - Manual handling can lead to mechanical defects:
 - Tapes dropping on the floor
 - Bumping
 - Bad insertions of tapes in tape drives
 - Tape cartridges contain mechanical parts
 - Manually changed tapes get damaged easily
 - Frequent rewinding causes stress to the tape substrate
 - Leads to lower reliability of data reads

Tapes

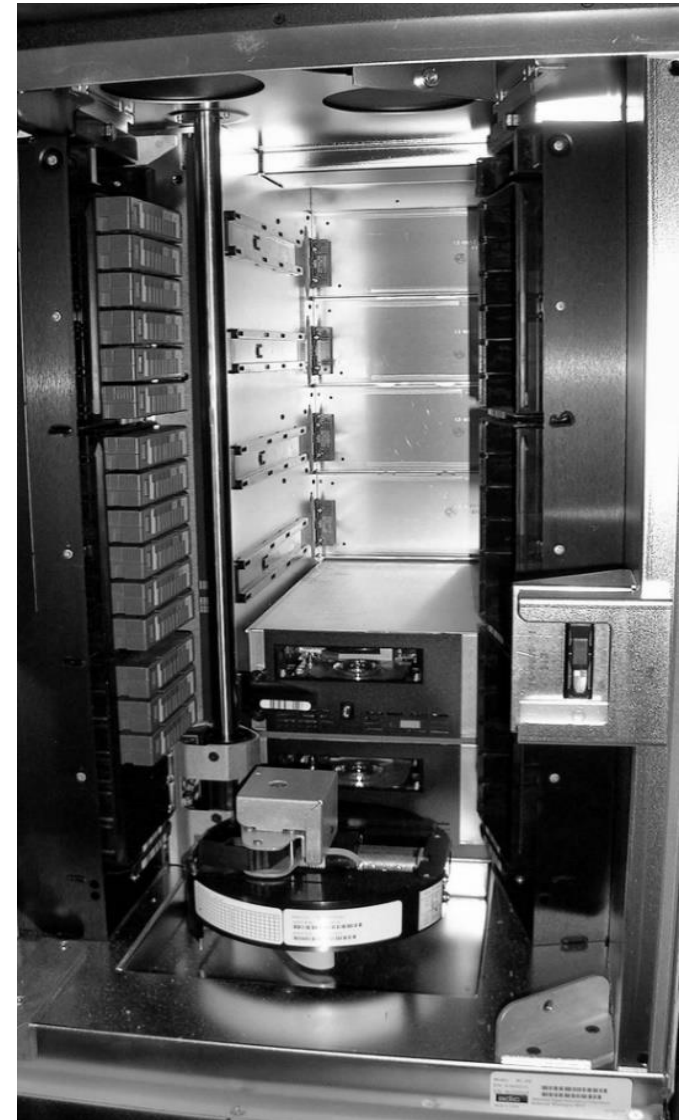
- Tapes are extremely slow
 - They only write and read data sequentially
 - When a particular piece of data is required, it must be searched by reading all data on tape until the required data is found
 - Together with rewinding of the tape (needed for ejecting the tapes) handling tapes is expressed in minutes instead of in milliseconds or microseconds

Tapes

- (S)DLT and LTO are the most popular tape cartridge formats in use today
 - LTO has a market share of more than 80%
 - LTO-7 tape cartridges can store 6 TB of uncompressed data
- Tape throughput is in the 100 to 150 MB/s range
 - The tape drive interface is capable of even higher speeds
 - Most tape drives use 4 Gbit/s Fibre Channel interfaces
 - A sustained throughput of between 350 and 400 MB/s

Tape library

- Tape libraries can be used to automate tape handling
- A tape library is a storage device that contains:
 - One or more tape drives
 - A number of slots to hold tape cartridges
 - A barcode or RFID tag reader to identify tape cartridges
 - An automated method for loading tapes



Virtual tape library

- A Virtual Tape Library (VTL) uses disks for storing backups
- A VTL consists of:
 - An appliance or server
 - Software that emulates traditional tape devices and formats
- VTLs combine high performance disk based backup and restore with well-known backup applications, standards, processes, and policies
- Most of the current VTL solutions use NL-SAS or SATA disk arrays because of their relatively low cost
- They provide multiple virtual tape drives for handling multiple tapes in parallel

Controllers

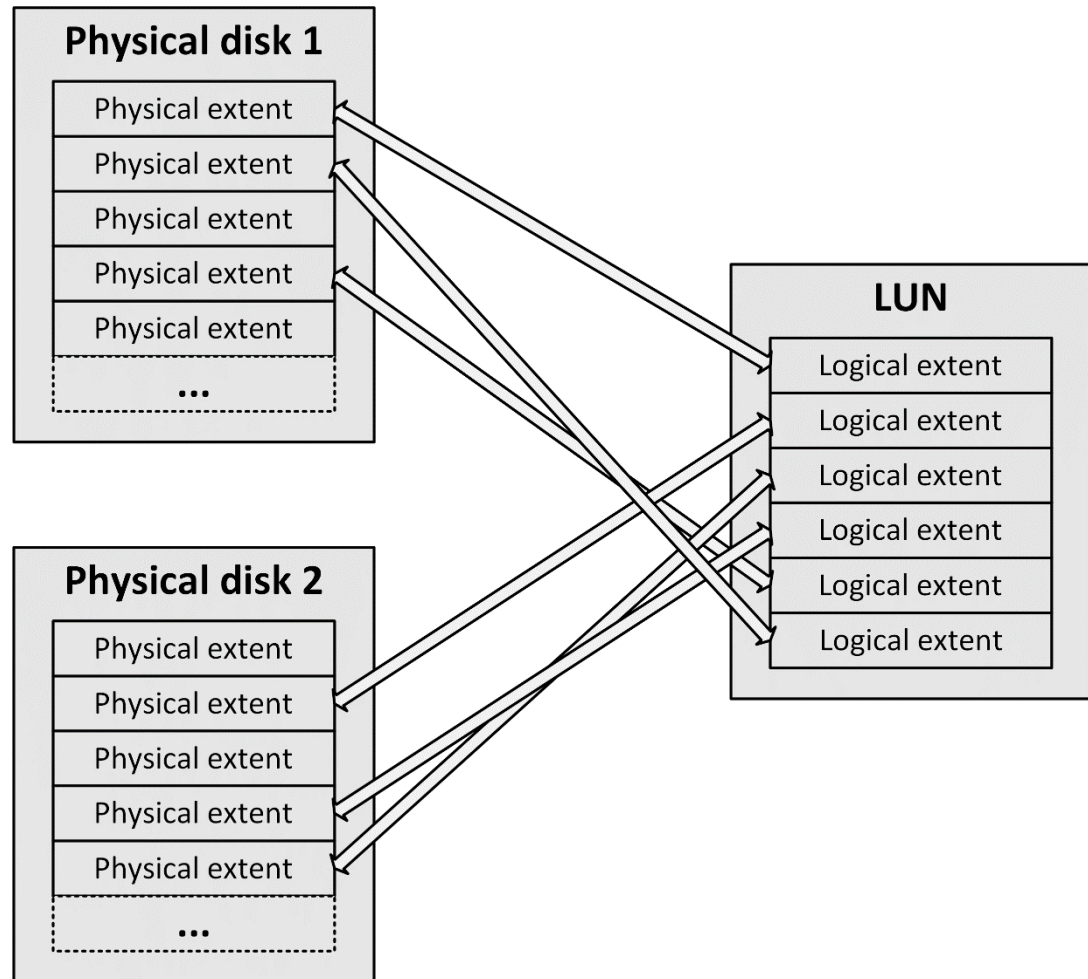
- Controllers connect disks and/or tapes to a server, in one of two ways:
 - Implemented as a PCI expansion boards in the server
 - Part of a NAS or SAN deployment, where they connect all available disks and tapes to redundant Fibre Channel, iSCSI, or FCoE connections

Controllers

- A controller can implement:
 - High performance
 - High availability
 - Virtualized storage
 - Cloning
 - Data deduplication
 - Thin provisioning

Controllers

- The controller splits up all disks in small pieces called physical extents
- From these physical extents, new virtual disks (Logical Unit Numbers - LUNs) are composed and presented to the operating system



RAID

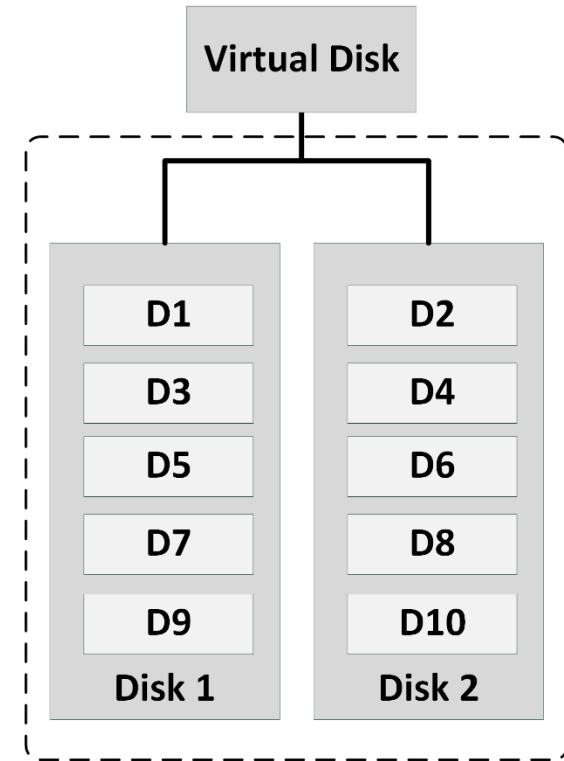
- Redundant Array of Independent Disks (RAID) solutions provide:
 - High availability of data
 - Improvements of performance
- RAID uses multiple redundant disks
- RAID can be implemented:
 - In the disk controller's hardware
 - As software running in a server's operating system

RAID

- RAID can be implemented in several configurations, called RAID levels
- In practice, five RAID levels are implemented most often:
 - **RAID 0** - Striping
 - **RAID 1** - Mirroring
 - **RAID 10** - Striping and Mirroring
 - **RAID 5** - Striping with distributed parity
 - **RAID 6** - Striping with distributed double parity

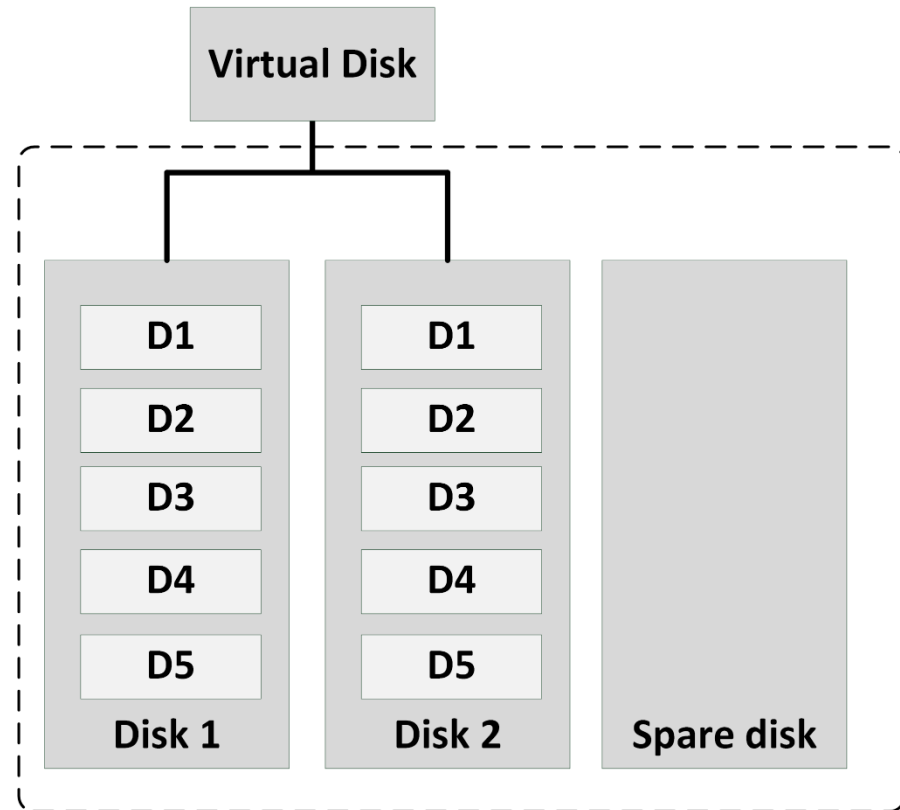
RAID 0 - Striping

- RAID 0 is also known as striping
- Provides an easy and cheap way to increase performance
- Uses multiple disks, each with a part of the data on it
- RAID 0 actually lowers availability
 - If one of the disks in a RAID 0 set fails, all data is lost
- Only acceptable if losing all data on the RAID set is no problem (for instance for temporary data)



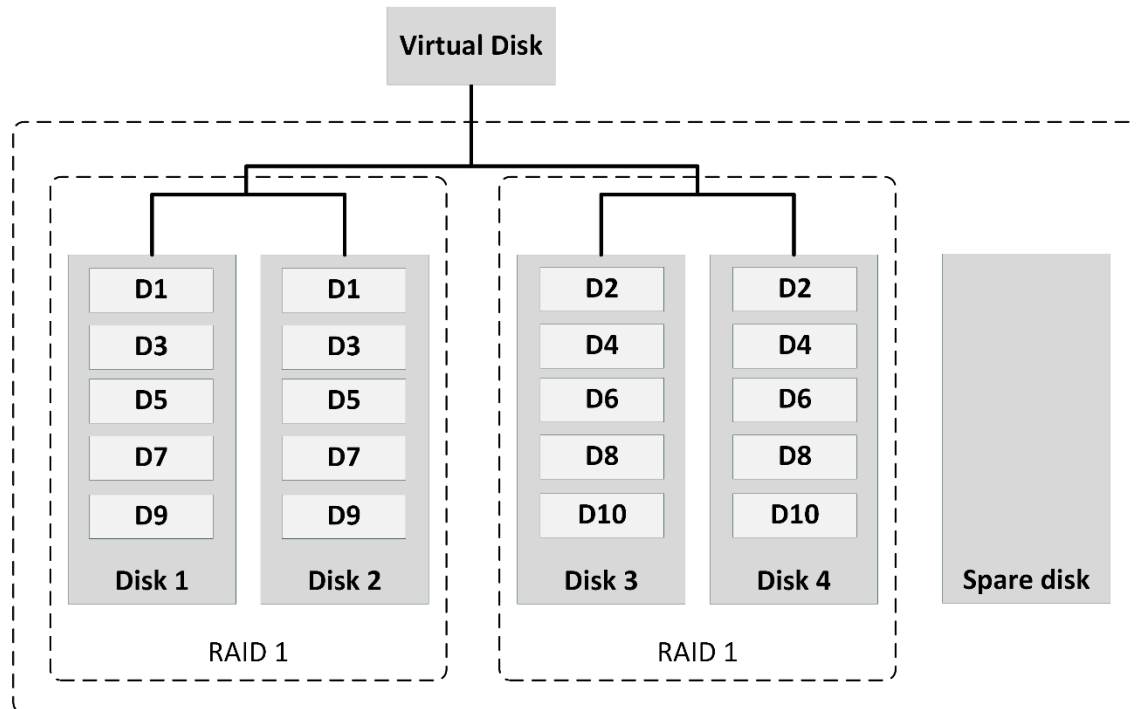
RAID 1 - Mirroring

- RAID 1 is also known as mirroring
- A high availability solution that uses two disks that contain the same data
- If one disk fails, data is not lost as it is still available on the mirror disk
- The most reliable RAID level
- High price
 - 50% of the disks are used for redundancy only
- A spare physical disk can be configured to automatically take over the task of a failed disk



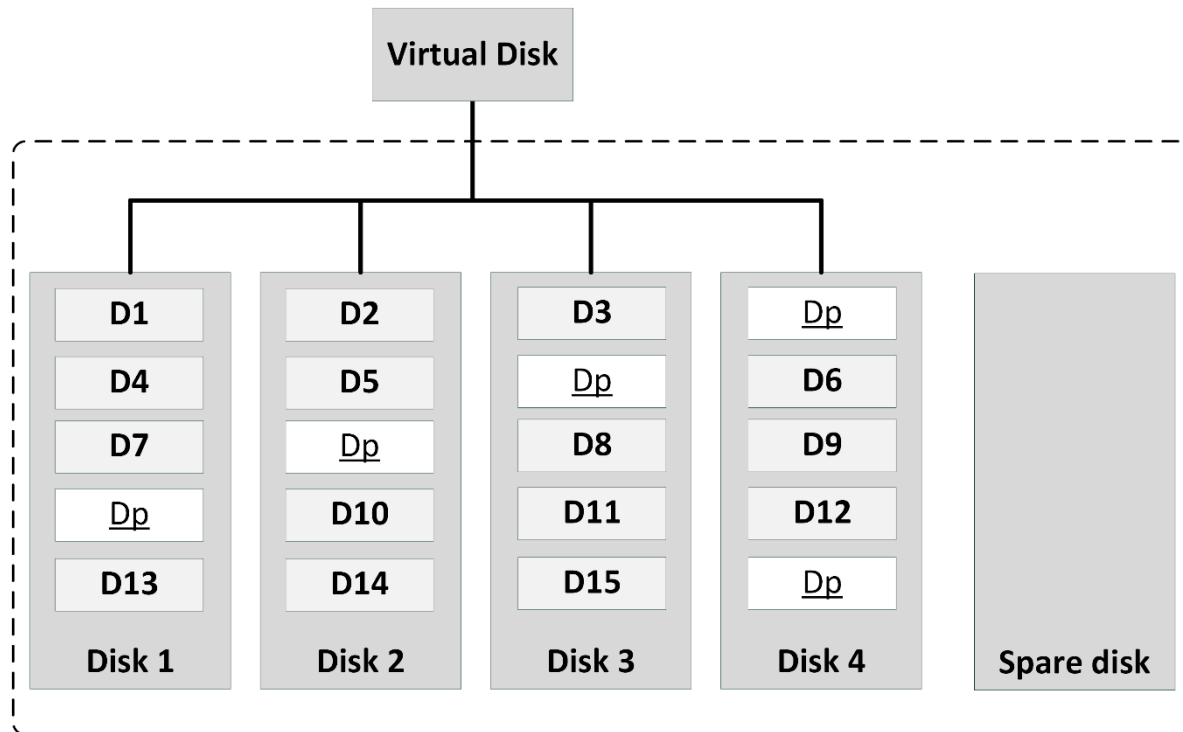
RAID 10 - Striping and mirroring

- RAID 10 uses a combination of striping and mirroring
- Provides high performance and availability
- Only 50% of the available disk space is used



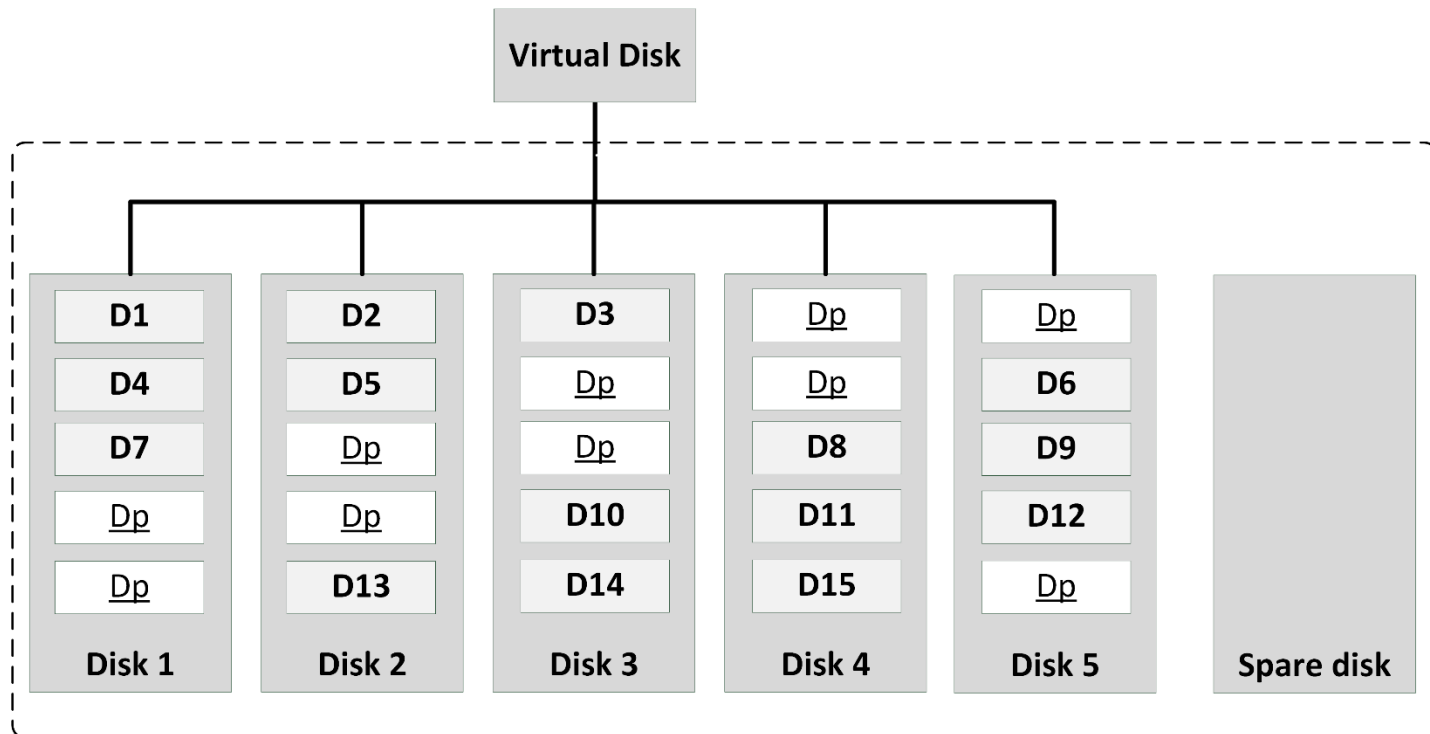
RAID 5 - Striping with distributed parity

- Data is written in disk blocks on all disks
- A parity block of the written disk blocks is stored as well
- This parity block is used to automatically reconstruct data in a RAID 5 set (using a spare disk) in case of a disk failure



RAID 6 - Striping with distributed double parity

- RAID 6 protects against double disk failures by using two distributed parity blocks instead of one
- Important in case a second disk fails during reconstruction of the first failing disk



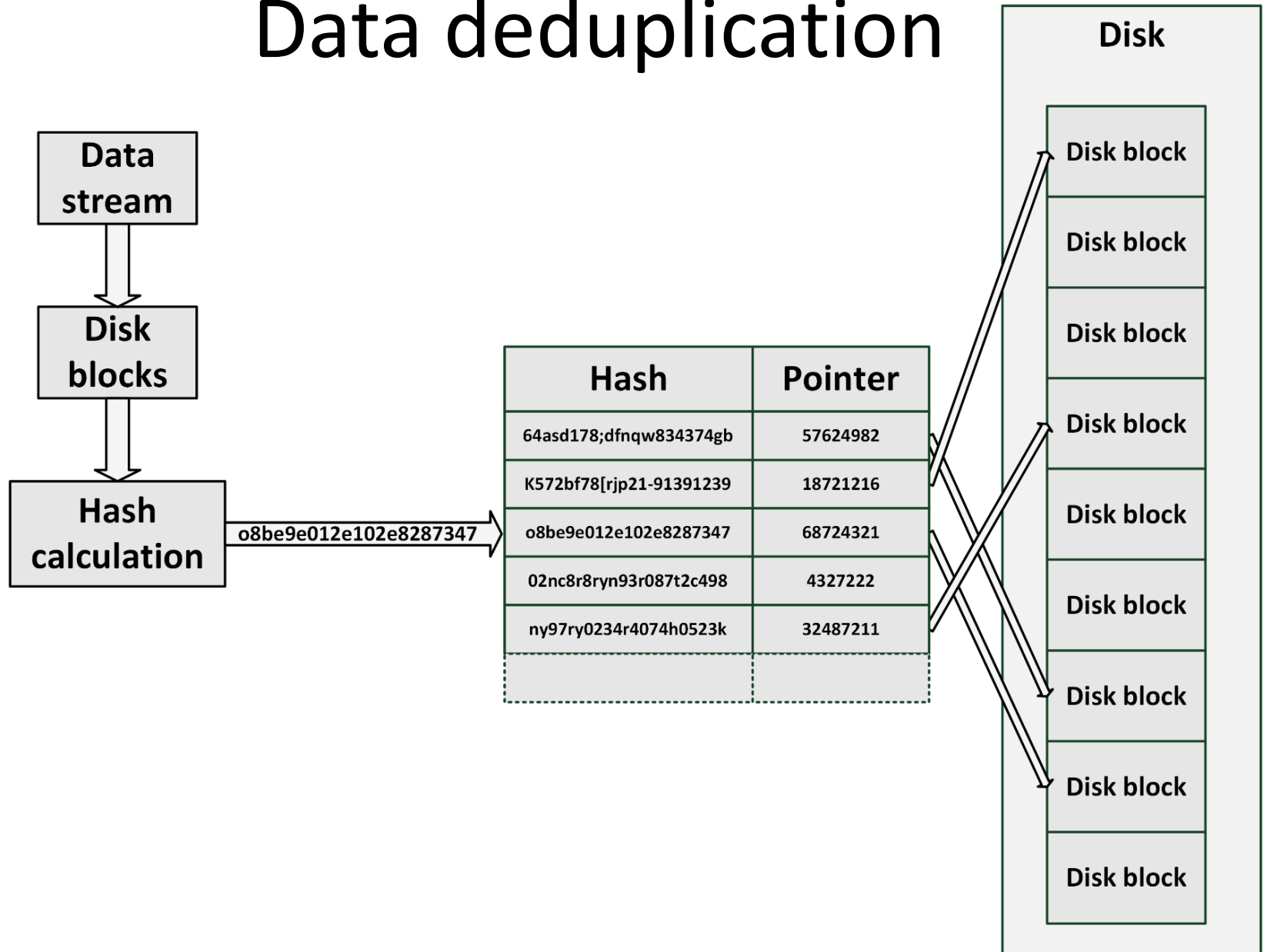
Data deduplication

- Data deduplication searches the storage system for duplicate data segments (disk blocks or files) and removes these duplicates
- Data deduplication is used in archived as well as in production data

Data deduplication

- The deduplication system keeps a table of hash tags to quickly identify duplicate disk blocks
 - The incoming data stream is segmented
 - Hash tags are calculated of those segments
 - The hashes are compared to hash tags of segments already on disk
 - If an incoming data segment is identified as a duplicate, the segment is not stored again, but a pointer to the matching segment is created for it instead

Data deduplication



Data deduplication

- Deduplication can be done inline or periodically
 - **Inline** deduplication checks for duplicate data segments before data is written to disk
 - Avoids duplicate data on disks at any time
 - Introduces a relatively large performance penalty

Data deduplication

- **Periodically:** writing data to disk first, and periodically check if duplicate data exists
 - Duplicate data is deduplicated by changing the duplicate data to a pointer to existing data on disk, and freeing disk space of the original block
 - This process can be done at times when performance needs are low
 - Duplicate data will be stored on the disks for some time

Cloning and snapshots

- With cloning and snapshotting, a copy of data is made at a specific point in time that can be used independently from the source data
- Usage:
 - Create a backup at a specific point in time, when the data is in a stable, consistent state
 - Creating test sets of data and an easy way to revert to older data without restoring data from a backup
- Cloning: the storage system creates a full copy of a disk, much like a RAID 1 mirror disk

Cloning and snapshots

- Snapshot: represents a point in time of the data on the disks
 - No writing to those disks is permitted anymore, as long as the snapshot is active
 - All writing is done on a separate disk volume in the storage system
 - The original disks still provide read-access

Thin provisioning

- Thin provisioning enables the allocation of more storage capacity to users than is physically installed
 - About 50% of allocated storage is never used
- Thin provisioning still provides the applications with the required storage
 - Storage is not really available on physical disks
 - Uses automated capacity management
 - The application's real storage need is monitored closely
 - Physical disk space is added when needed
- Typical use: Providing users with large sized home directories or email storage

Direct Attached Storage (DAS)

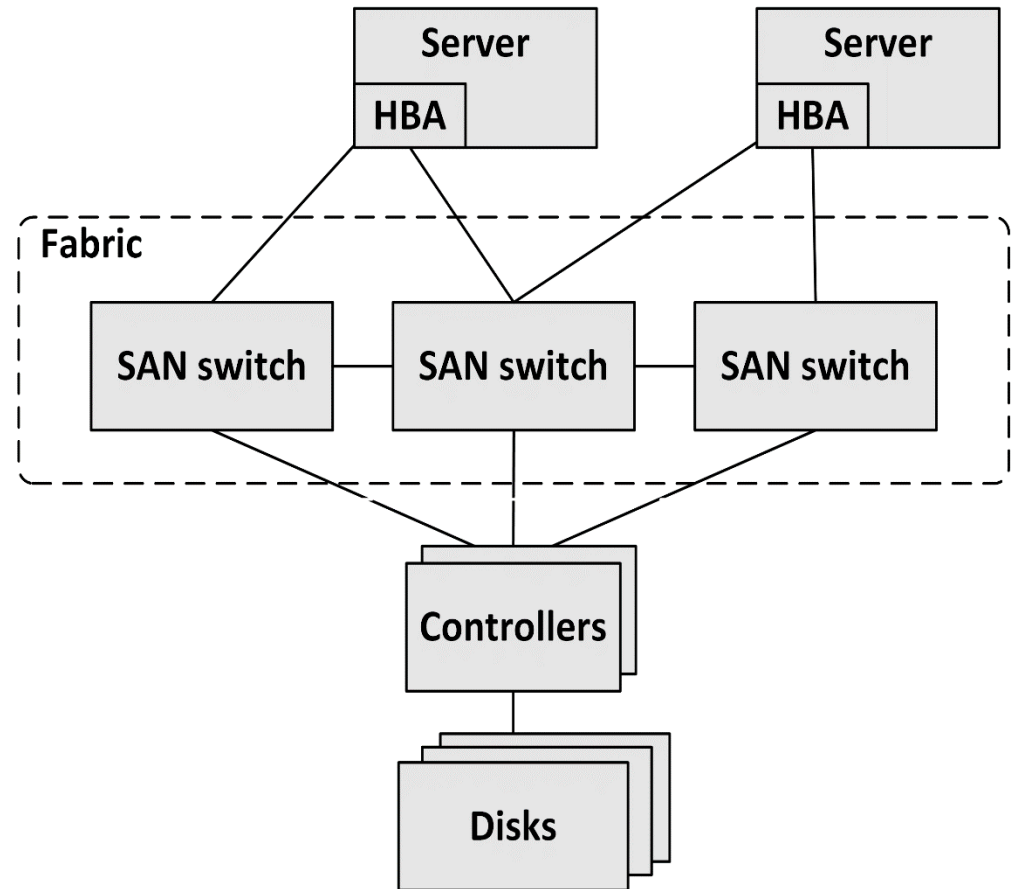
- DAS – also known as local disks – is a storage system where one or more dedicated disks connect via the SAS or SATA protocol to a built-in controller, connected to the rest of the computer using the PCI bus
- The controller provides a set of disk blocks to the computer, organized in LUNs (or partitions)
- The computer's operating system uses these disk blocks to create a file system to store files

Storage Area Network (SAN)

- A Storage Area Network (SAN) is a specialized storage network that consists of SAN switches, controllers and storage devices
- It connects a large pool of central storage to multiple servers
- A SAN physically connects servers to disk controllers using specialized networking technologies like Fibre Channel or iSCSI
- Via the SAN, disk controllers offer virtual disks to servers, also known as LUNs (Logical Unit Numbers)
- LUNs are only available to the server that has that specific LUN mounted

Storage Area Network (SAN)

- The core of the SAN is a set of SAN switches, called the Fabric
 - Comparable with a LAN's switched network segment
- Host bus adapters (HBAs) are interface cards implemented in servers
 - Comparable to NICs used in networking
 - Connected to SAN switches, usually in a redundant way



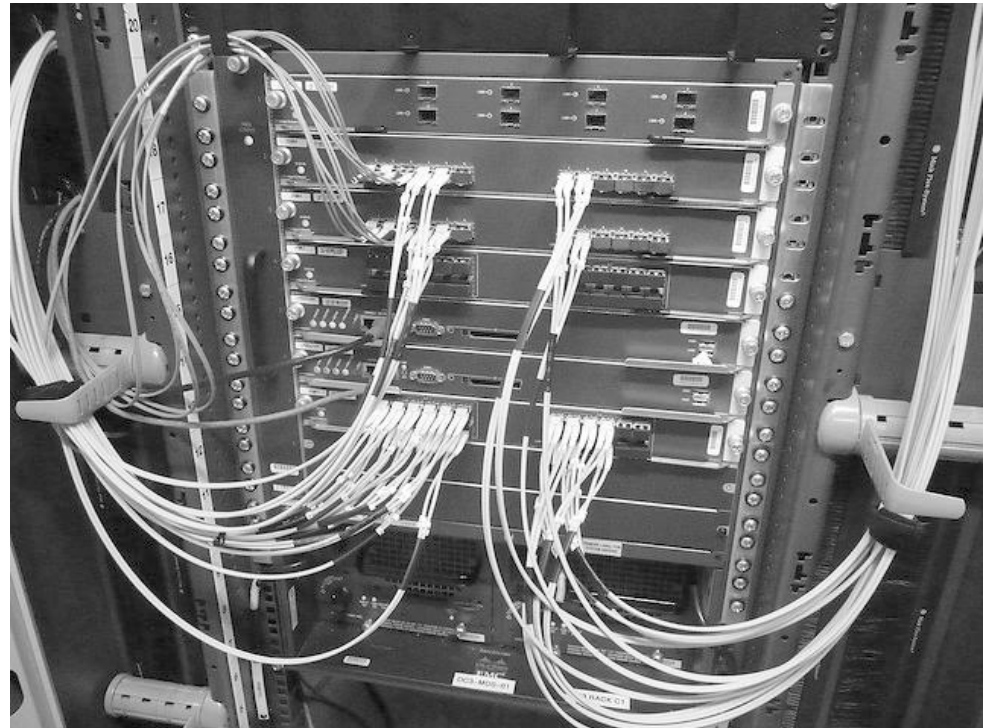
Storage Area Network (SAN)

- In SANs, a large number of disks are installed in one or more disk arrays
- The number of disks varies between dozens of disks and hundreds of disks
- A disk array can easily contain many hundreds of terabytes (TB) of data or more



SAN connectivity protocols

- The most used SAN connectivity protocols:
 - Fibre Channel
 - FCoE
 - iSCSI



Fibre Channel

- Fibre Channel (FC) is a dedicated level 2 network protocol, specially designed for transportation of storage data blocks
- Speeds: 2 Gbit/s, 4 Gbit/s, 8 Gbit/s, or 16 Gbit/s
- Runs on:
 - Twisted pair copper wire (i.e. UTP and STP)
 - Fiber optic cables
- The Fibre Channel protocol was specially developed for the transport of disk blocks
- The protocol is very reliable, with guaranteed zero data loss

Fibre Channel

- Three network topologies:
 - Point-to-Point
 - Two devices are connected directly to each other
 - Arbitrated loop
 - Also known as FC-AL
 - All devices are in a loop
 - Switched fabric
 - All devices are connected to Fibre Channel switches
 - A similar concept as in Ethernet implementations
- Most implementations today use a switched fabric

FCoE

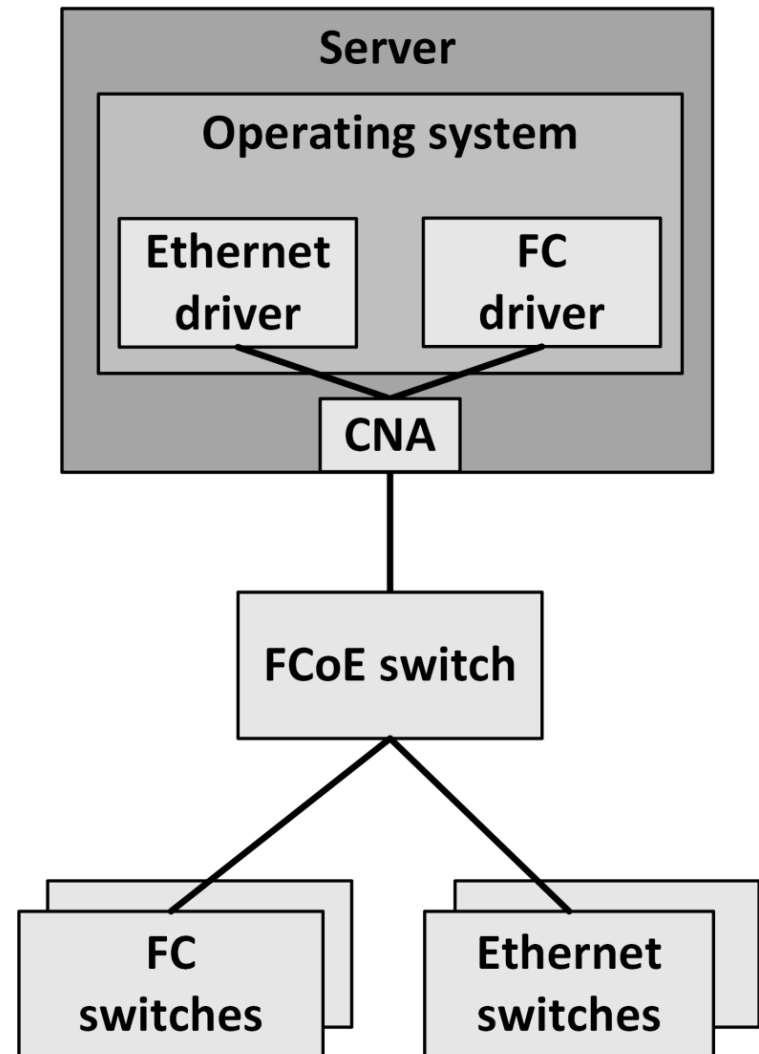
- Fibre Channel over Ethernet (FCoE) encapsulates Fibre Channel data in Ethernet packets
- Allows Fibre Channel traffic to be transported over 10 Gbit or higher Ethernet networks
- FCoE eliminates the need for separate Ethernet and Fibre Channel cabling and switching technology
- PCoE needs at least 10 Gbit Ethernet with special extensions, known as Data Center Bridging (DCB) or Converged Enhanced Ethernet (CEE)

FCoE

- Ethernet extensions:
 - Lossless Ethernet connections
 - A FCoE implementation must guarantee that no Ethernet packets are lost
 - Quality of Service (QoS)
 - Allows FCoE packets to have priority over other Ethernet packets to avoid storage performance issues
 - Large Maximum Transfer Unit (MTU) support
 - Allows Ethernet packets of 2500 bytes in size, instead of the standard 1500 bytes
 - Also known as Jumbo frames

FCoE

- FCoE needs specialized Converged Network Adapters (CNAs)
- CNAs support the Ethernet extensions
- They present themselves to the operating system as two adapters:
 - Ethernet Network Interface Controller (NIC)
 - Fibre Channel Host Bus Adapter (HBA)



iSCSI

- iSCSI allows the SCSI protocol to run over Ethernet LANs using TCP/IP
- Uses the familiar TCP/IP protocols and well known SCSI commands
- Performance is typically lower than that of Fibre Channel, due to the TCP/IP overhead
- With 10 or 40 Gbit/s Ethernet and jumbo frames, iSCSI is now rapidly conquering a big part of the SAN market