

Storage Technologies

IT403

Prepared by
Dr. Ahmed El-Douh

Lecture 6



Storage Technologies

Lecture 6

IT403

Storage Reliability Technologies
Common Advanced Storage Technologies

Prepared by
Dr. Ahmed El-Douh

Objectives

Upon completion of this module, you will be able to:

- ▣ Describe the concepts, principles and types of RAID technologies.
- ▣ Describe the principles of RAID 2.0+ technologies.
- ▣ Understand the Multipathing technologies of hosts.
- ▣ Understand the technologies implemented for ensuring the reliability of hard disk.

Outlines

- Traditional RAID Technologies.
- RAID 2.0+ Technologies.
- Multipathing Technologies.
- Disk Reliability Technologies.

Chapter
06

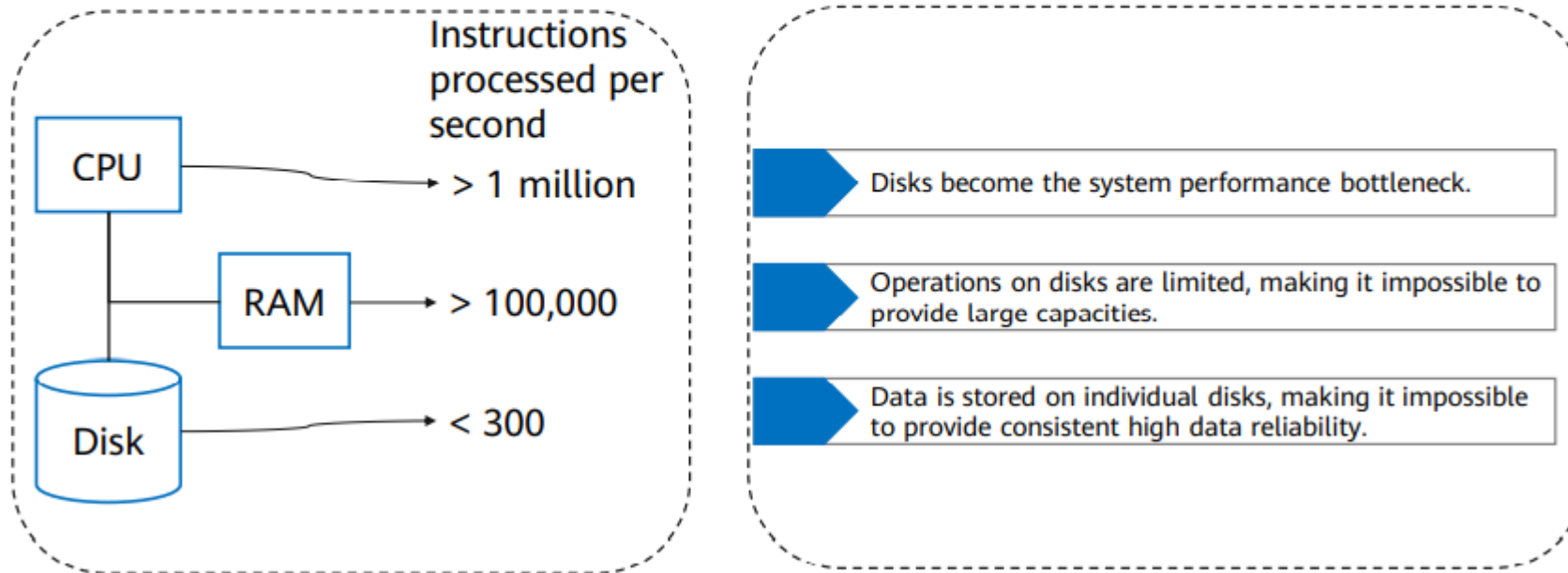
Traditional RAID Technologies



Traditional RAID Technologies

Background

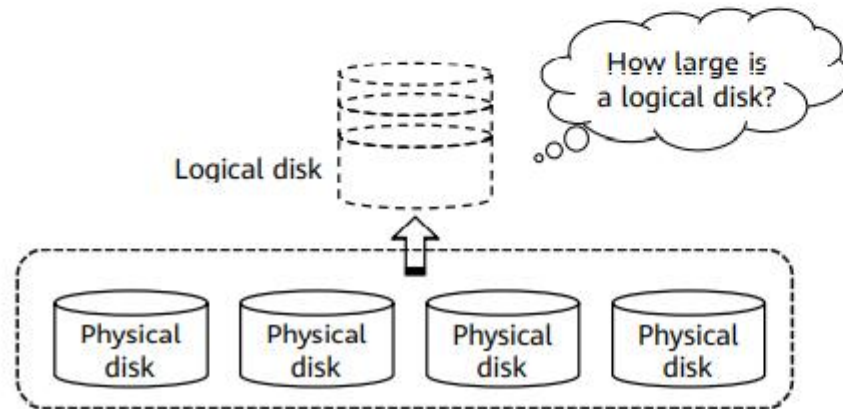
- Problems in traditional computer systems must be addressed.



Traditional RAID Technologies

What Is RAID?

- Redundant Array of Independent Disks (RAID) combines multiple physical disks into one logical disk in different ways, for the purposes of read/write performance and data security improvement.
- Implementations: hardware RAID and software RAID.



The Basic Concepts and Implementation of RAID

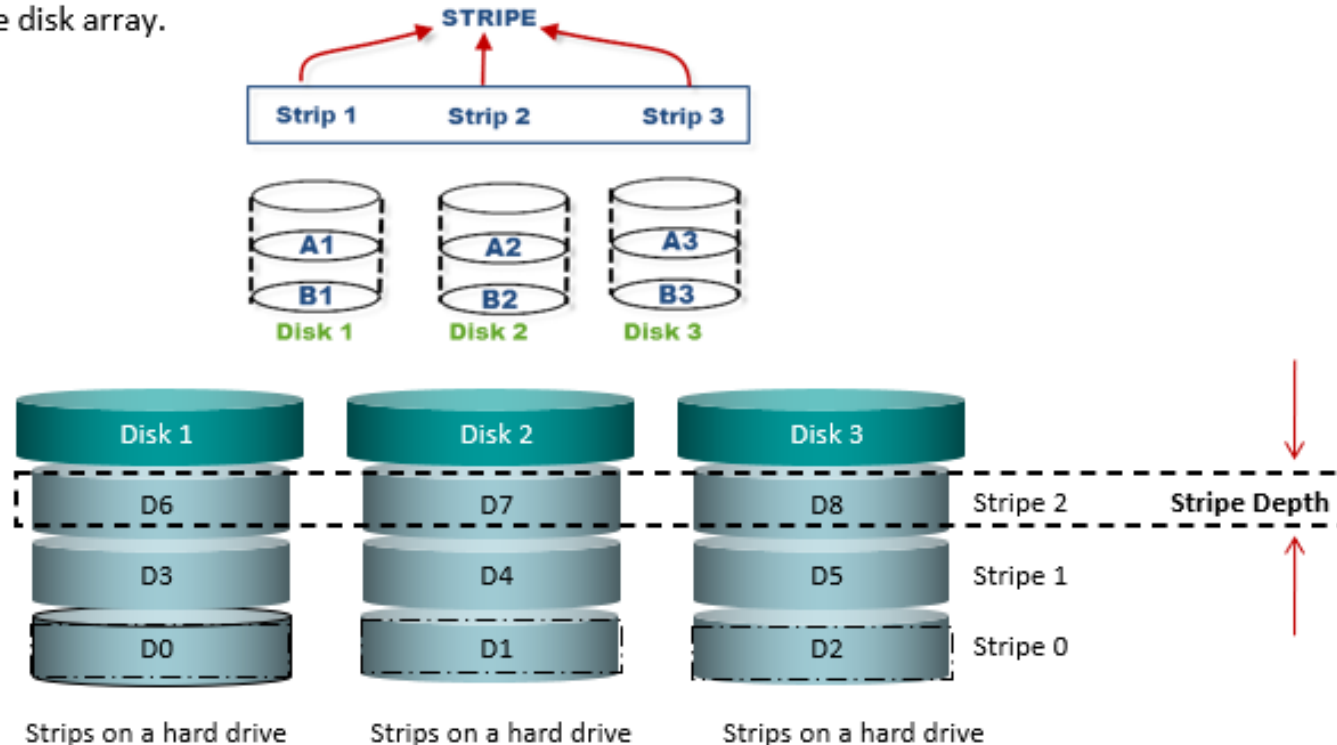
RAID: Redundant Array of Independent Disks



- Implementation Method:
 - Hardware RAID
 - Software RAID

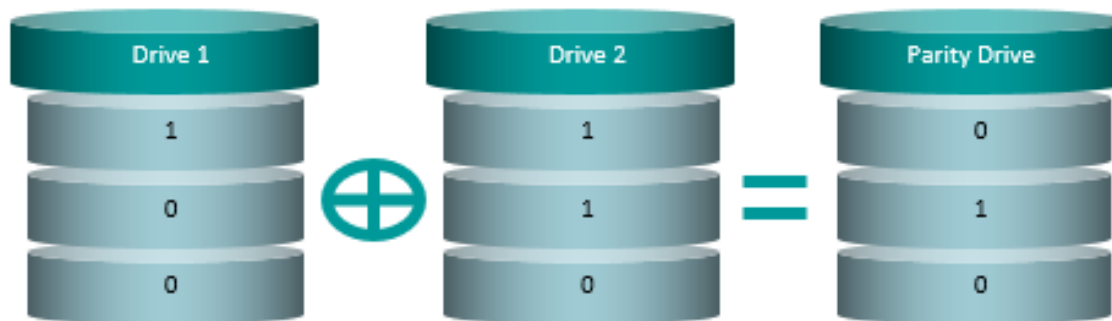
The Forms of RAID Data Organization

- Strip: A contiguously addressable block in the hard drive. It is the smallest unit that can be written with data in a hard drive and it is the element that forms a stripe.
- Stripe: Collection of strips that has the same "location" (or same ID) aligned in multiple hard drives within the disk array.



RAID Data Protection Methods

- Method 1: Save a copy of the data on another redundant hard drive.
- Method 2: Parity Check Calculation (XOR).
 - XOR computation is widely used in digital and computing science.
 - XOR computation--If both is equal then it is false, if both is different then it is true.
 - $0 \oplus 0 = 0$; $0 \oplus 1 = 1$; $1 \oplus 0 = 1$; $1 \oplus 1 = 0$;

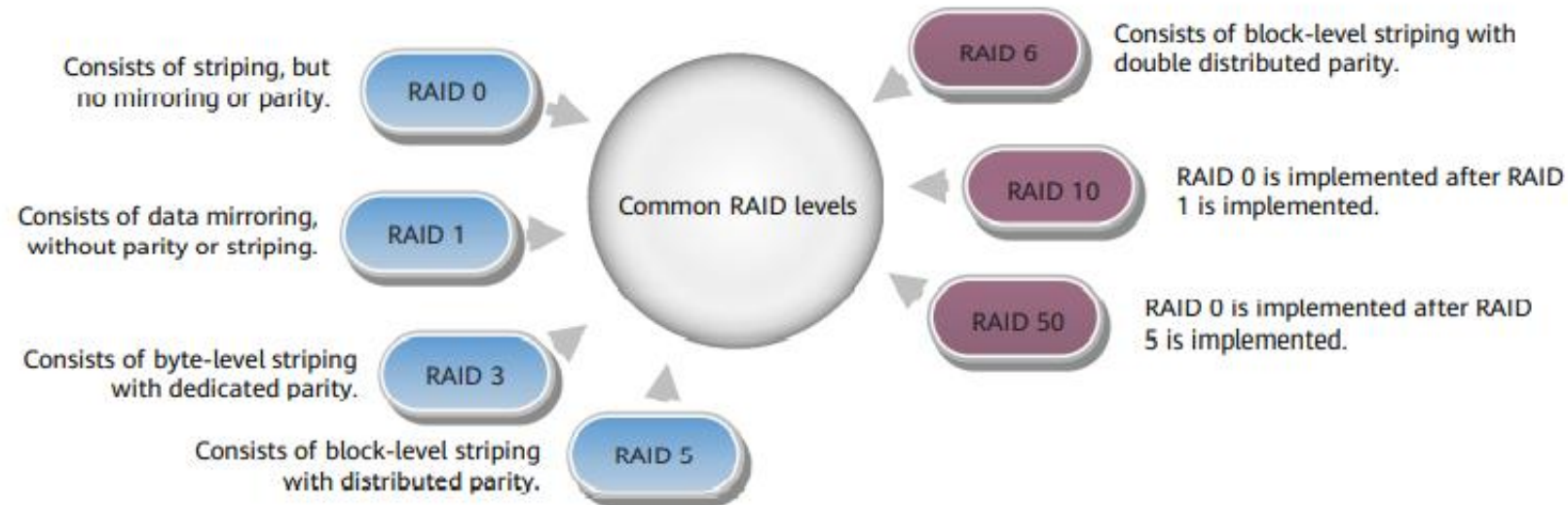


XOR Parity Check Creates A Redundant Backup of Data.

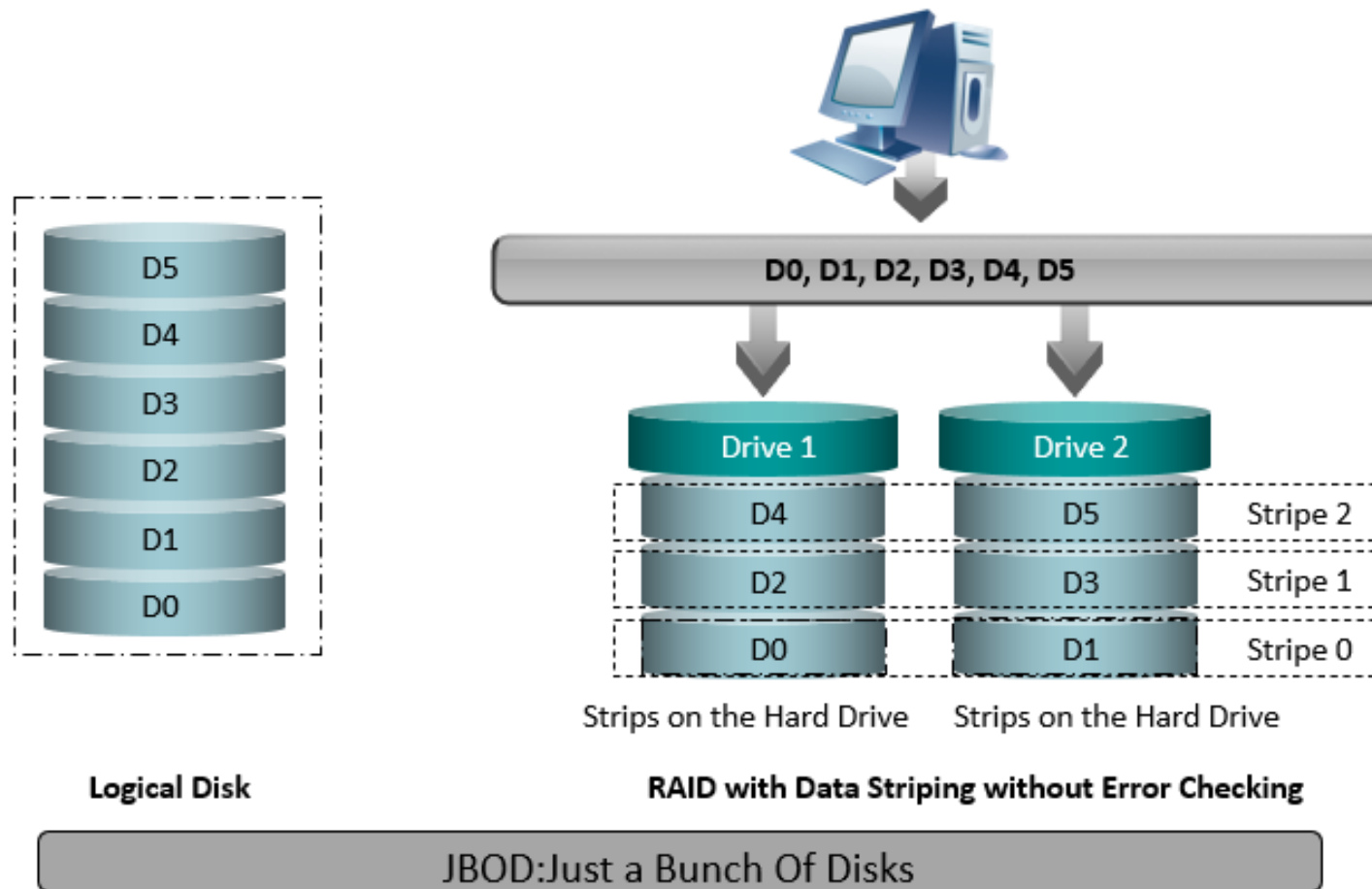
RAID Technologies

Common RAID Levels and Classification Criteria

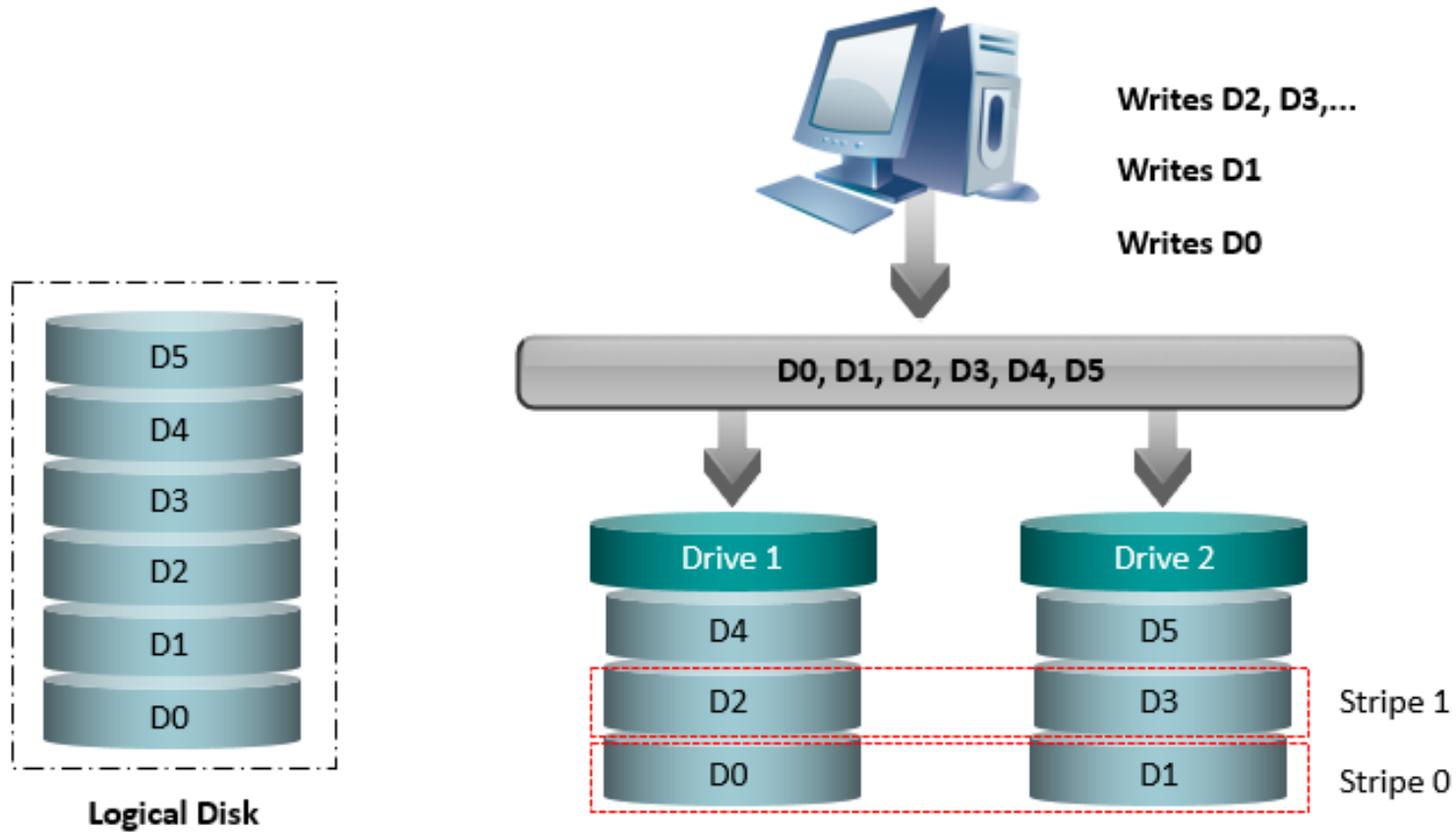
- RAID levels use different combinations of data organization forms and data protection techniques.



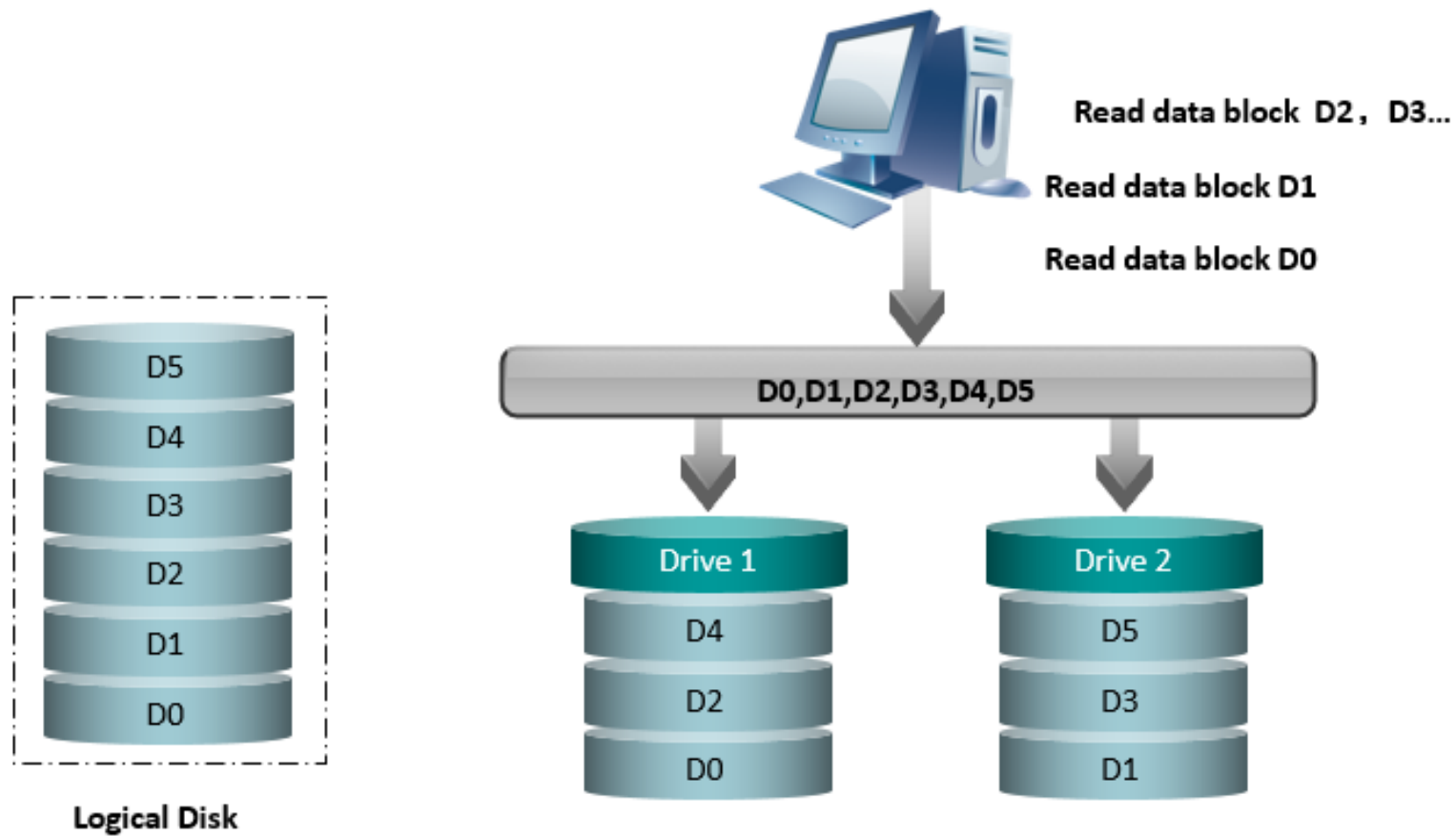
The Working Principles of RAID 0



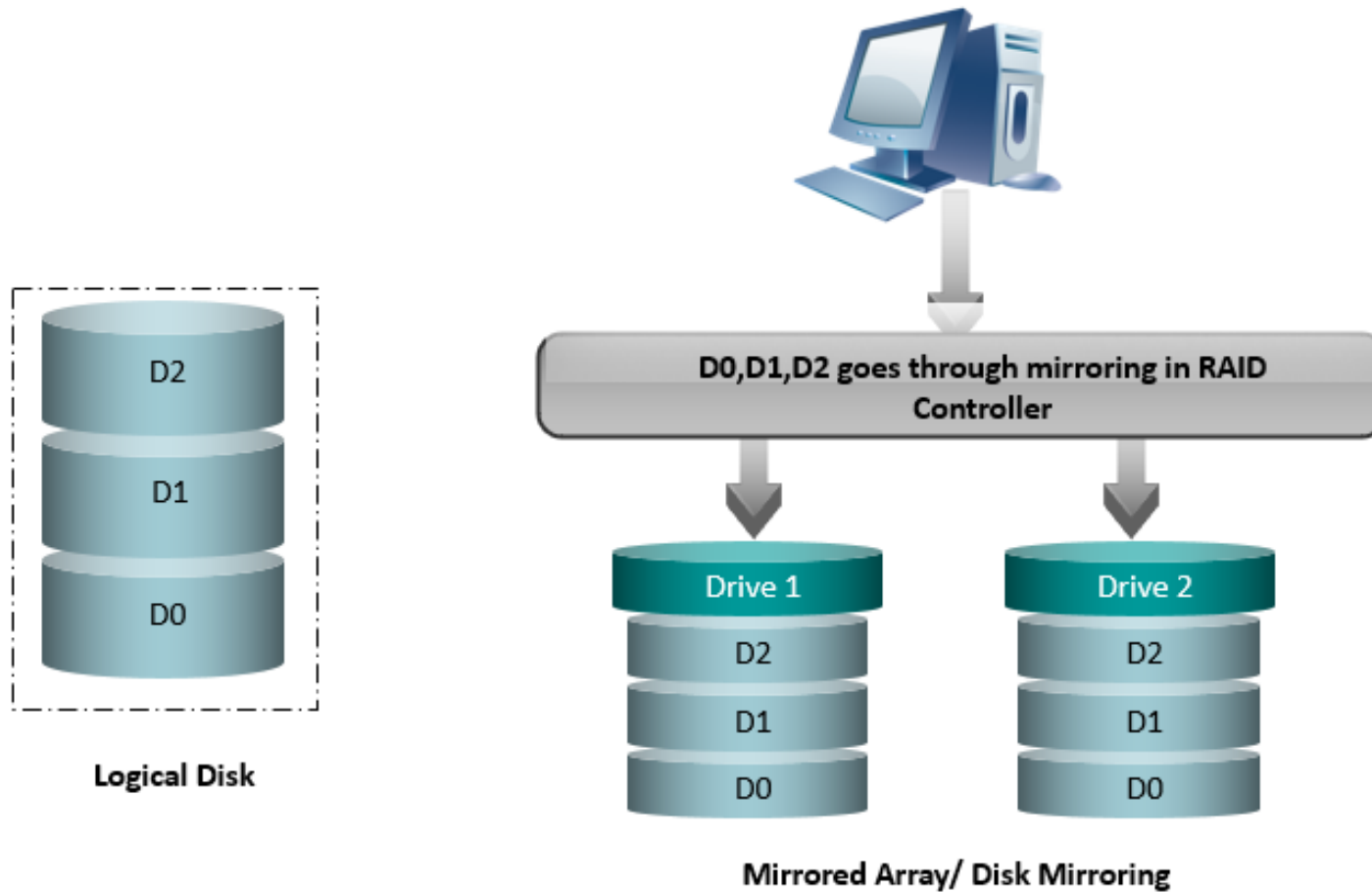
Data Writing of RAID 0



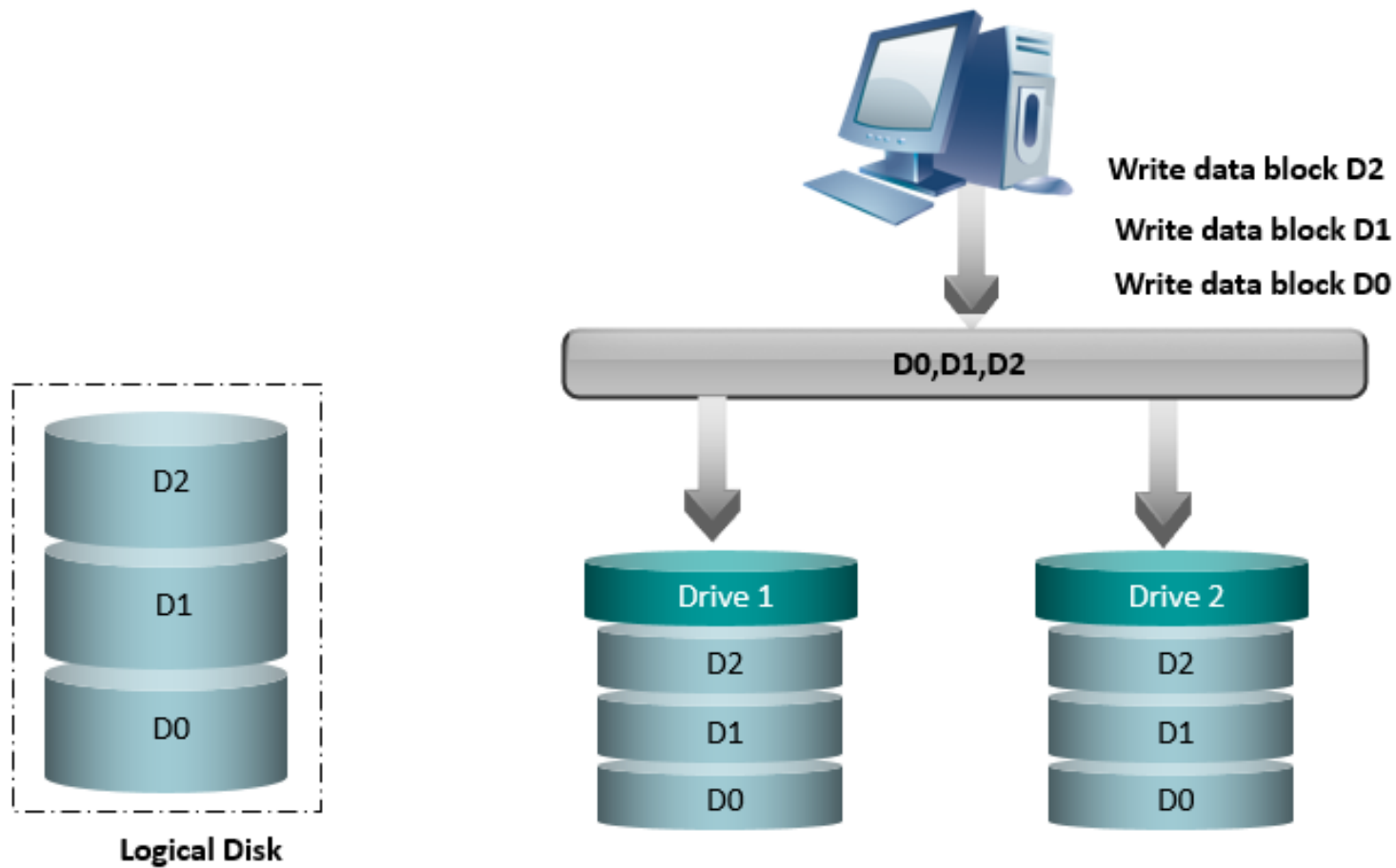
Data Reading of RAID 0



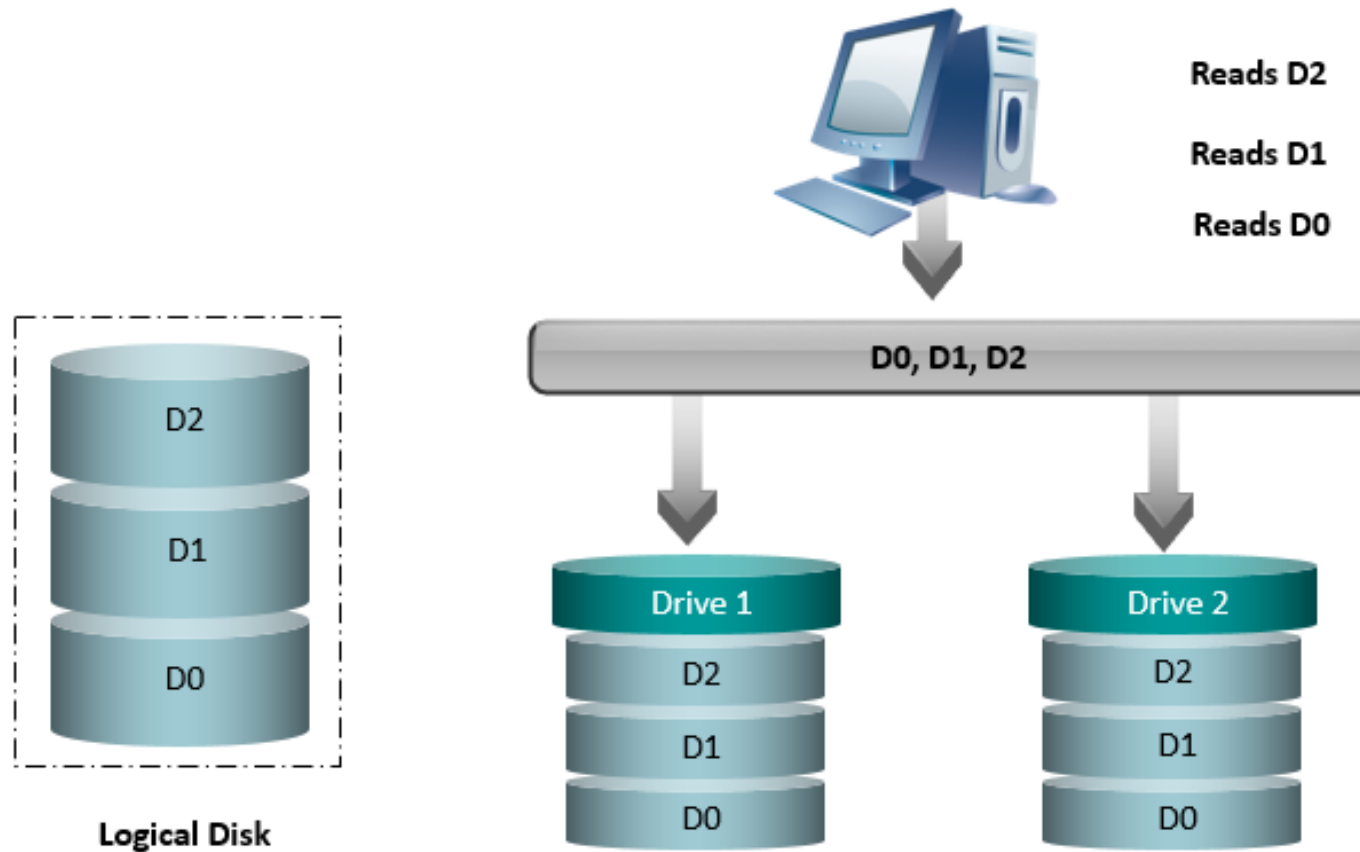
The Working Principles of RAID 1



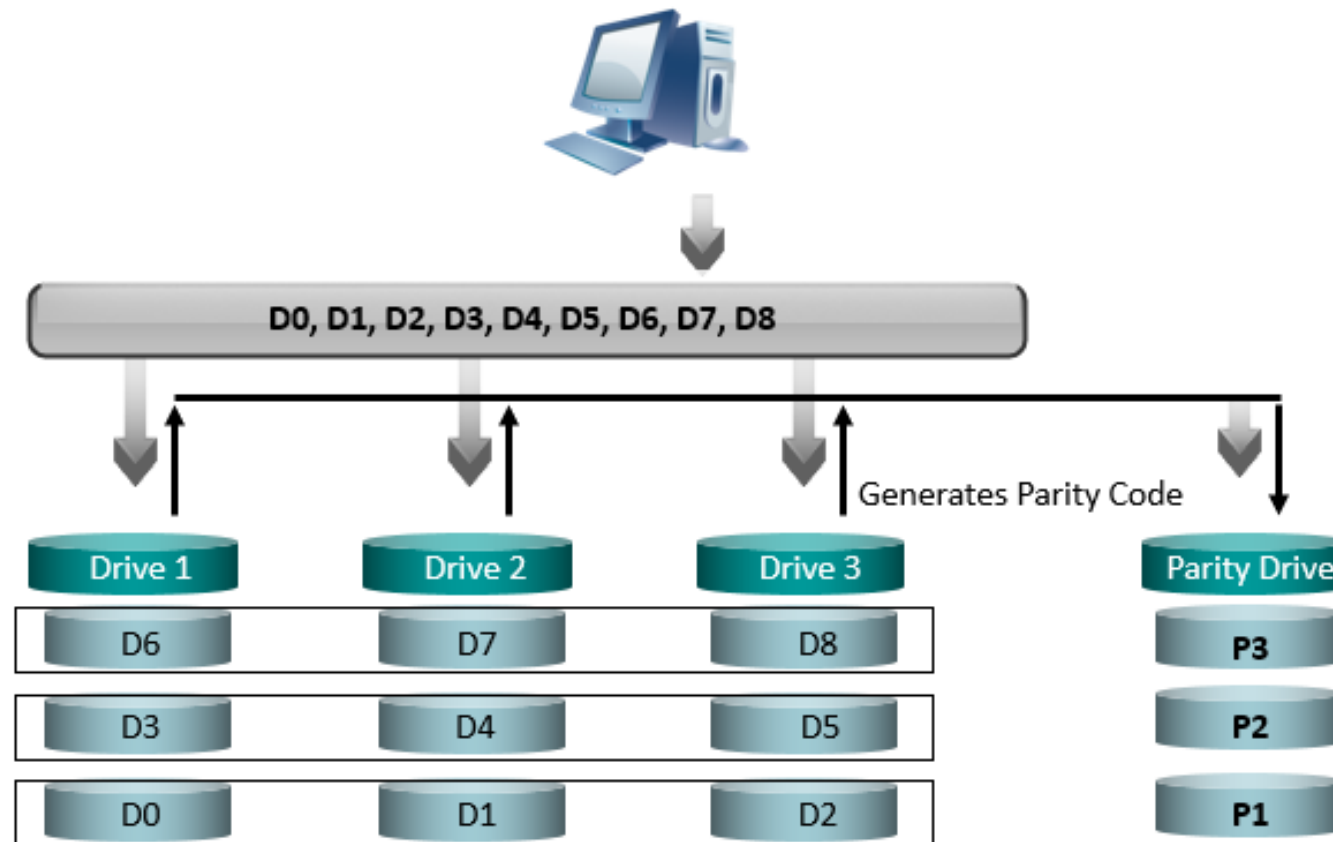
Data Writing of RAID 1



Data Reading of RAID 1

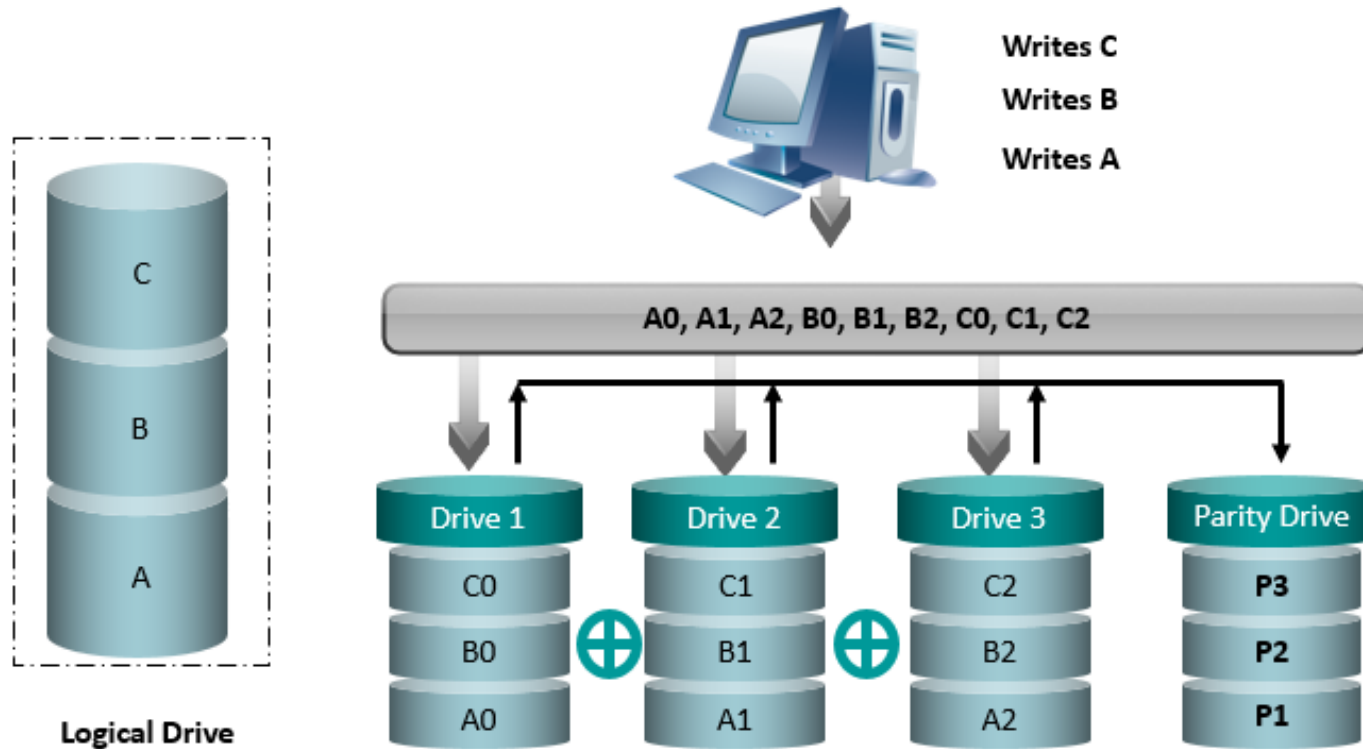


The Working Principles of RAID 3



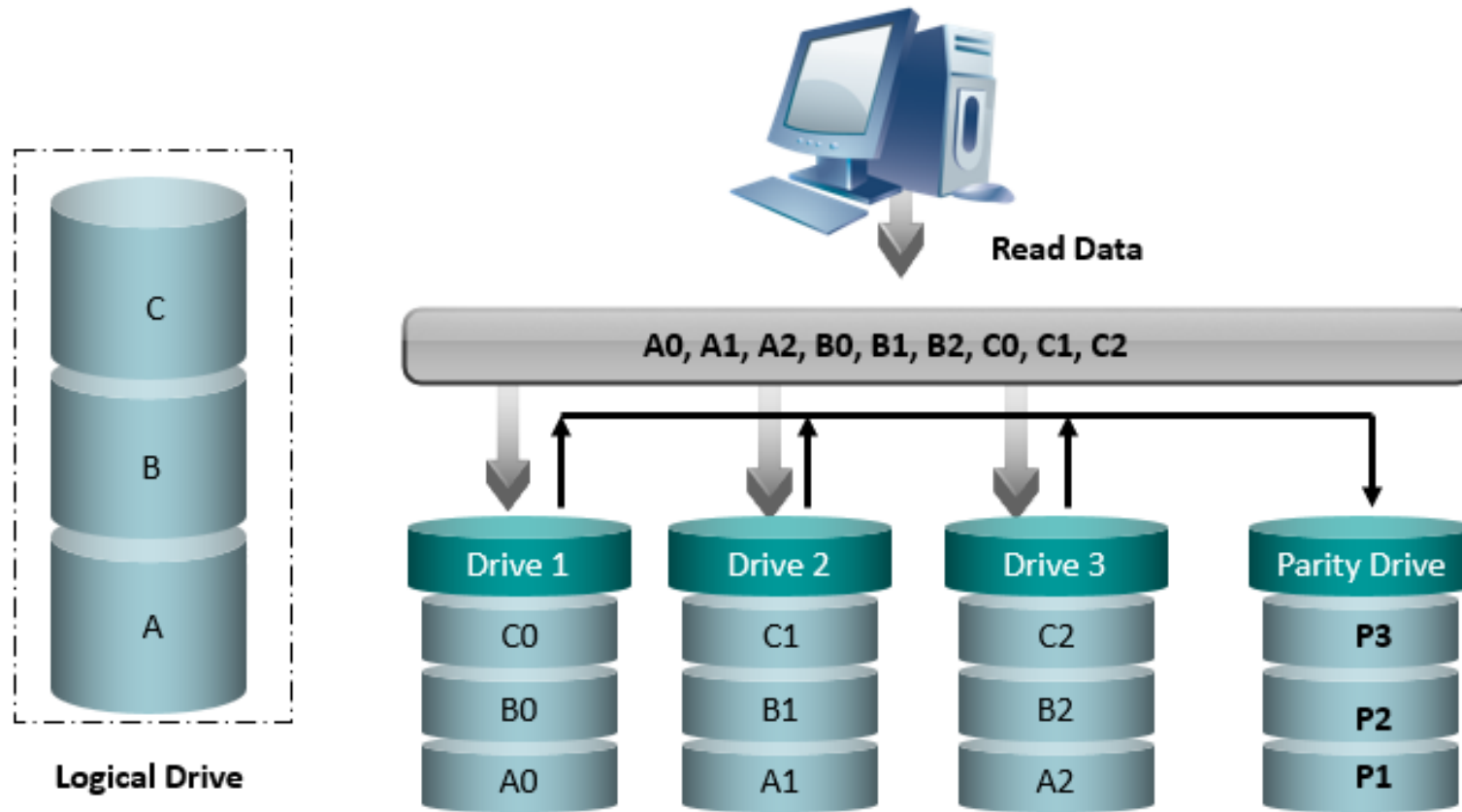
RAID with Striping and Dedicated Parity Disk

Data Writing of RAID 3

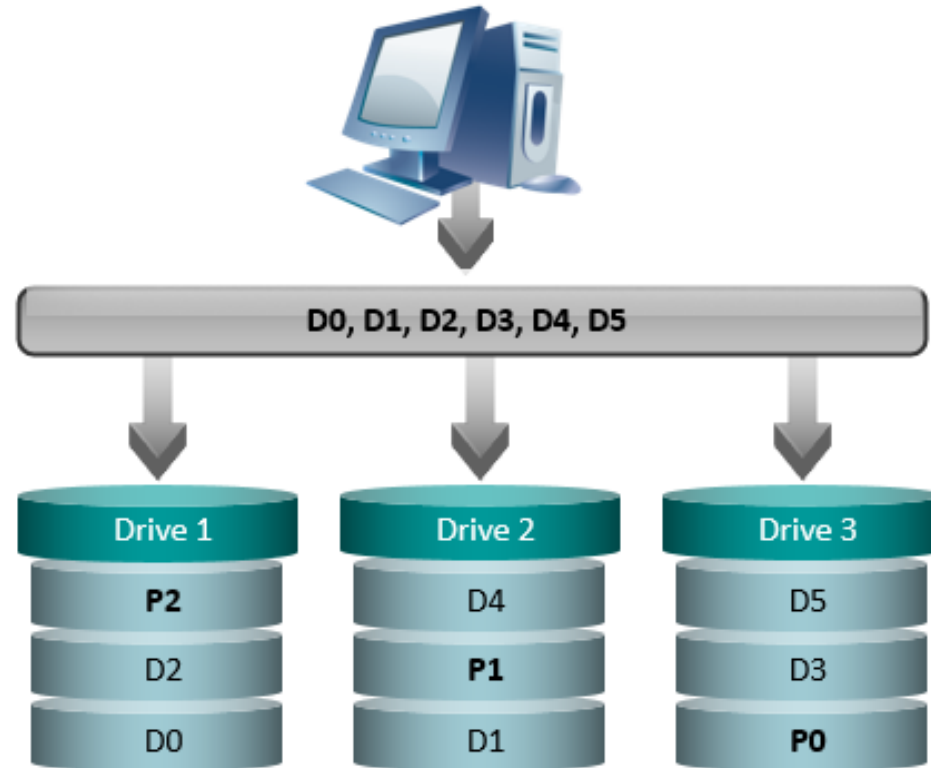


Note: When new data to be written is significantly small and only need to be written to 1 or 2 disks, it will cause RAID 3 "Write Penalty".

Data Reading of RAID 3

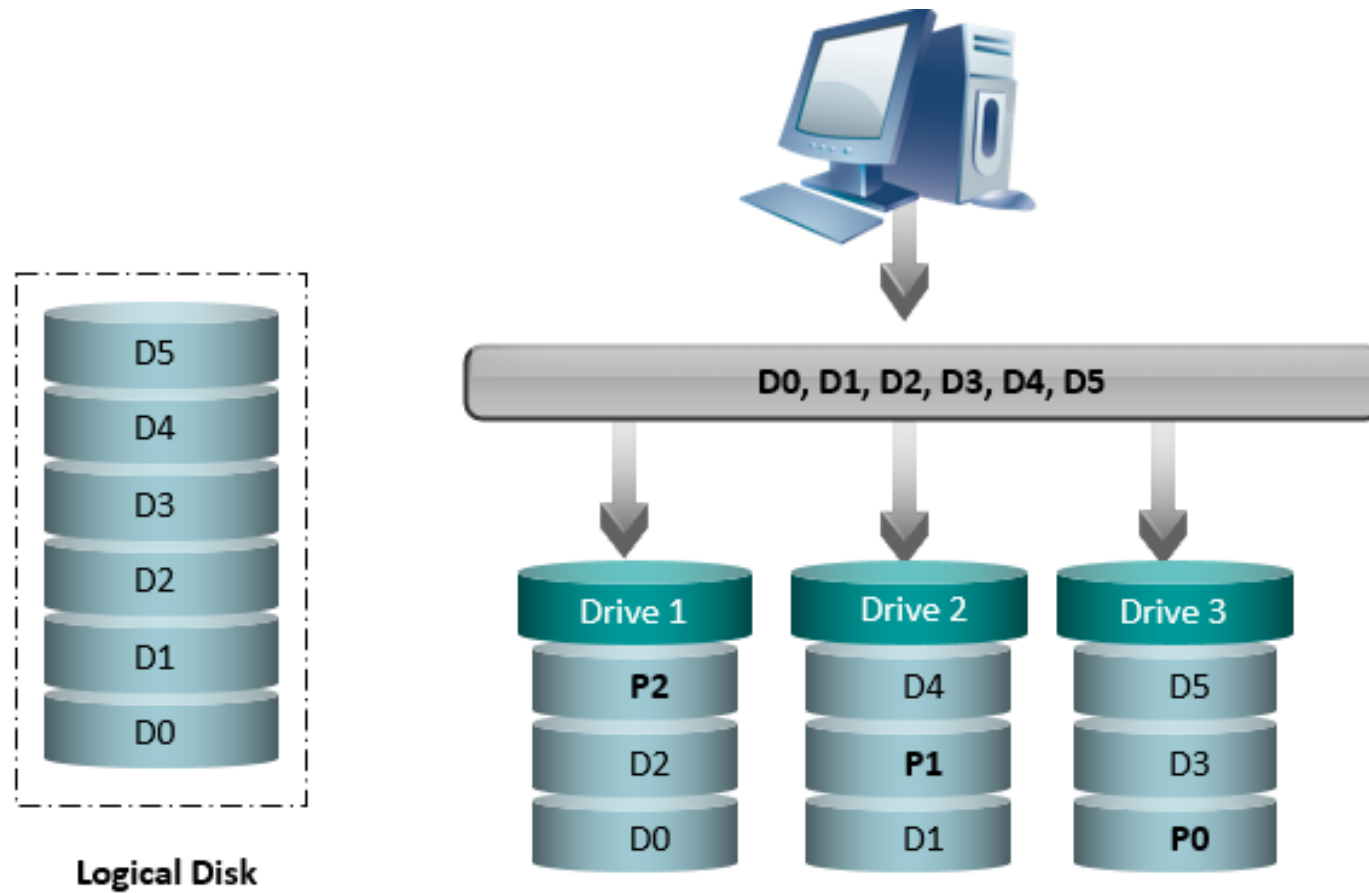


Working Principles of RAID 5

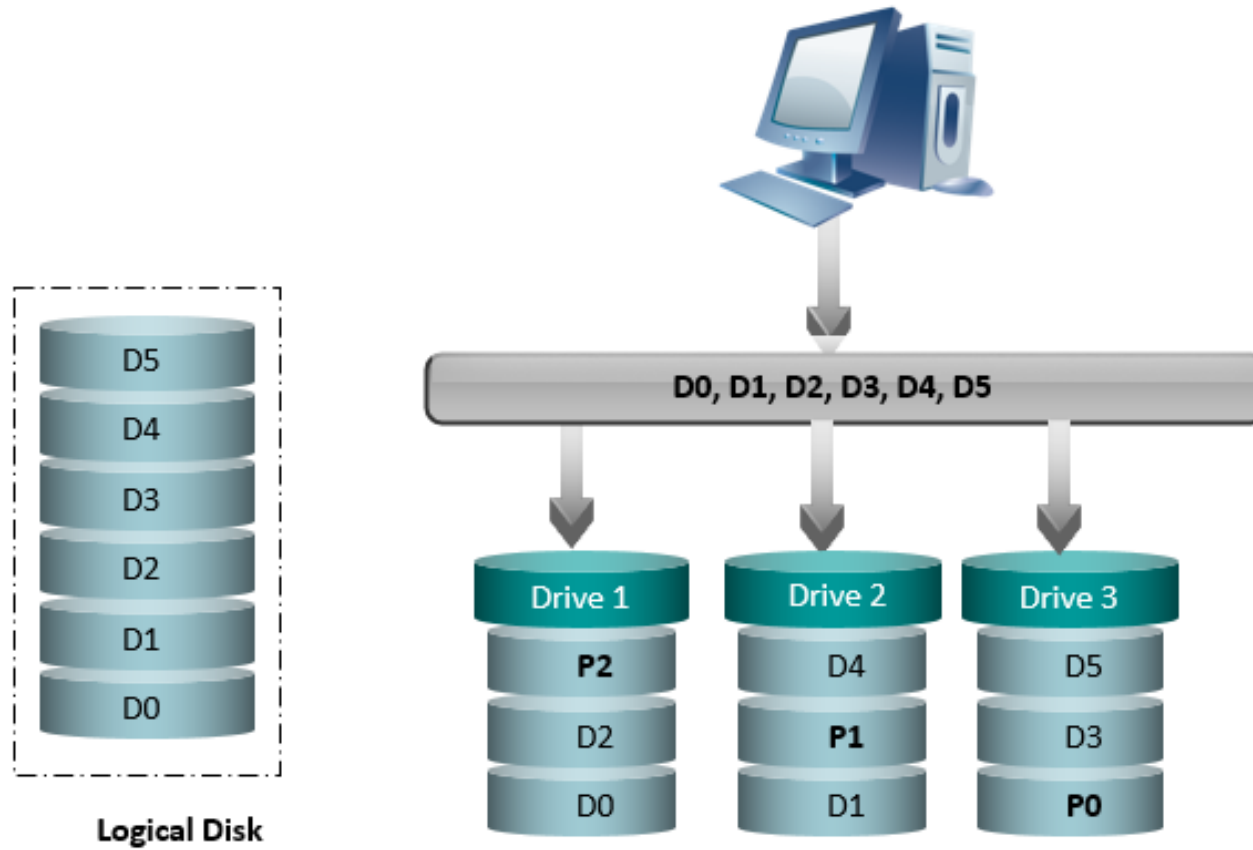


RAID with Distributed Parity

Data Writing of RAID 5



Data Reading of RAID 5

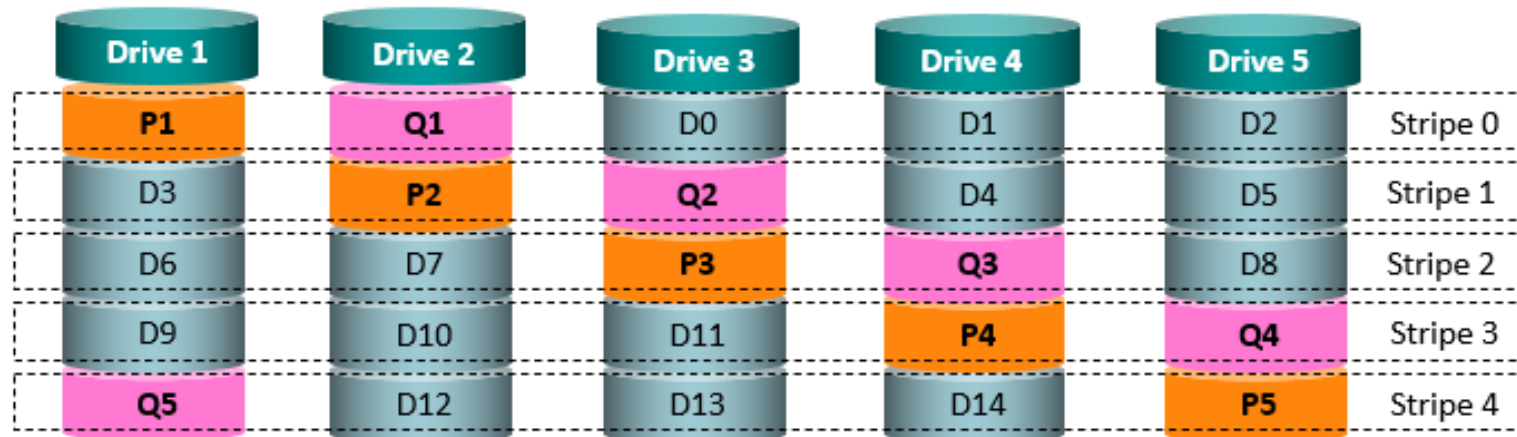


Overview of RAID 6

- RAID 6
 - Has 2 types of parity options for this RAID type.
 - Requires at least $N+2$ ($N>2$) number of disk to form this array, commonly used in high data reliability, high availability application scenarios.
- The common RAID 6 technologies are:
 - RAID6 P + Q
 - RAID6 DP (Double Parity)

Working Principles of RAID 6 P+Q

- RAID 6 P+Q calculates 2 different sets of parity data P and Q, when data is lost, the data can be recovered based on P and Q. The parity information of P and Q is calculated based on the formula below:
 - $P = D0 \oplus D1 \oplus D2...$
 - $Q = (\alpha \oplus D0) \oplus (\beta \oplus D1) \oplus (\gamma \oplus D2)...$



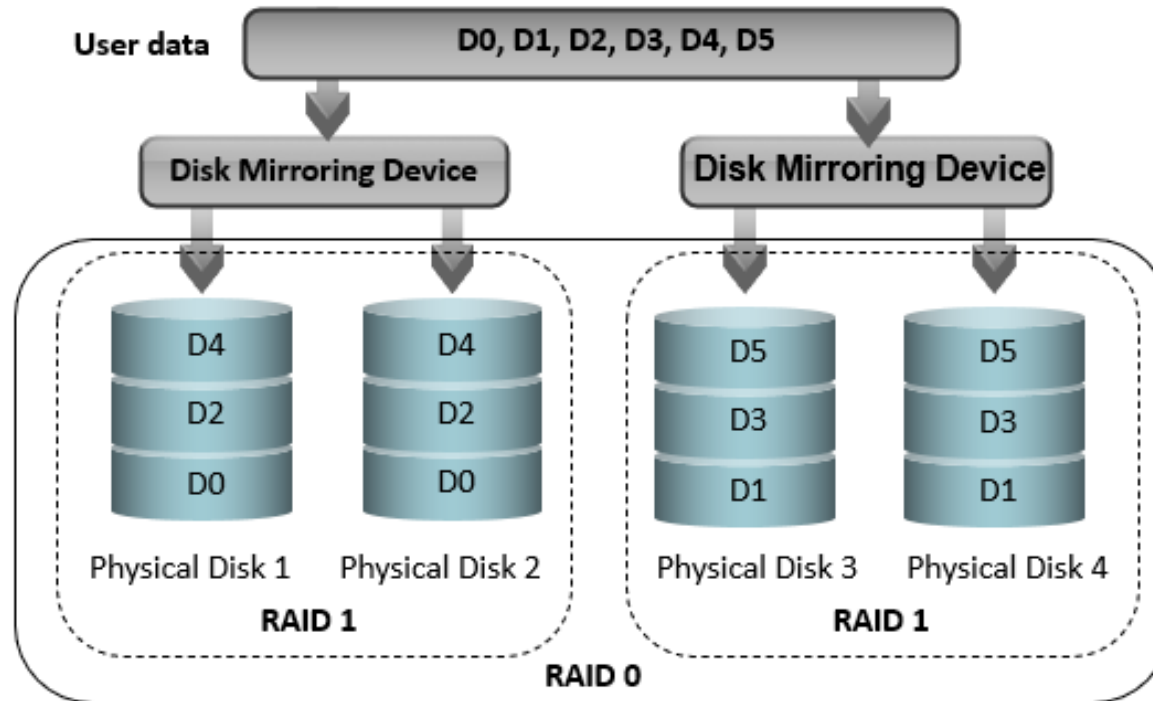
Working Principles of RAID 6 DP

- DP=Double Parity, which is an addition of a diagonal parity disk used to store parity information, which is added on top of the basis of XOR parity disks used by RAID 4.
- P0-P3 within the horizontal parity disk are the horizontal data parity information for each disks(Drive1-4).
 - Example: $P0 = D0 \text{ XOR } D1 \text{ XOR } D2 \text{ XOR } D3$
- DP0-DP3 within the diagonal parity disk are the diagonal parity information for data in Drive1-4 and horizontal parity drive.
 - Example: $DP0 = D0 \text{ XOR } D5 \text{ XOR } D10 \text{ XOR } D15$



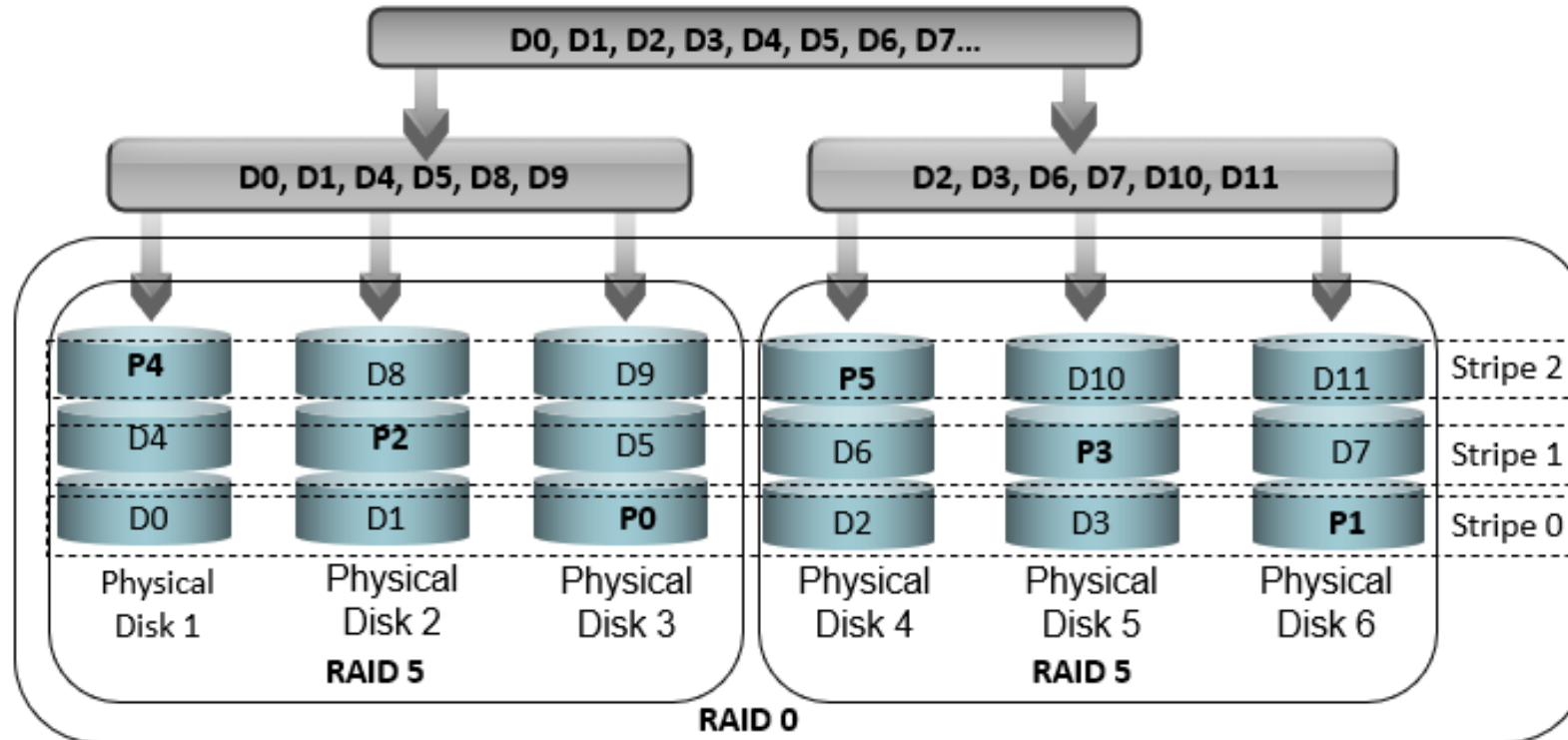
Hybrid RAID - RAID 10

- RAID 10 is a RAID type that combines mirroring and striping by implementing RAID 1 mirroring and then RAID 0. RAID 10 is one of the widely used RAID types.



Hybrid RAID - RAID 50

- RAID 50 is the RAID type that combines RAID 5 and RAID 0. The first level is RAID 5 and the second level is RAID 0.

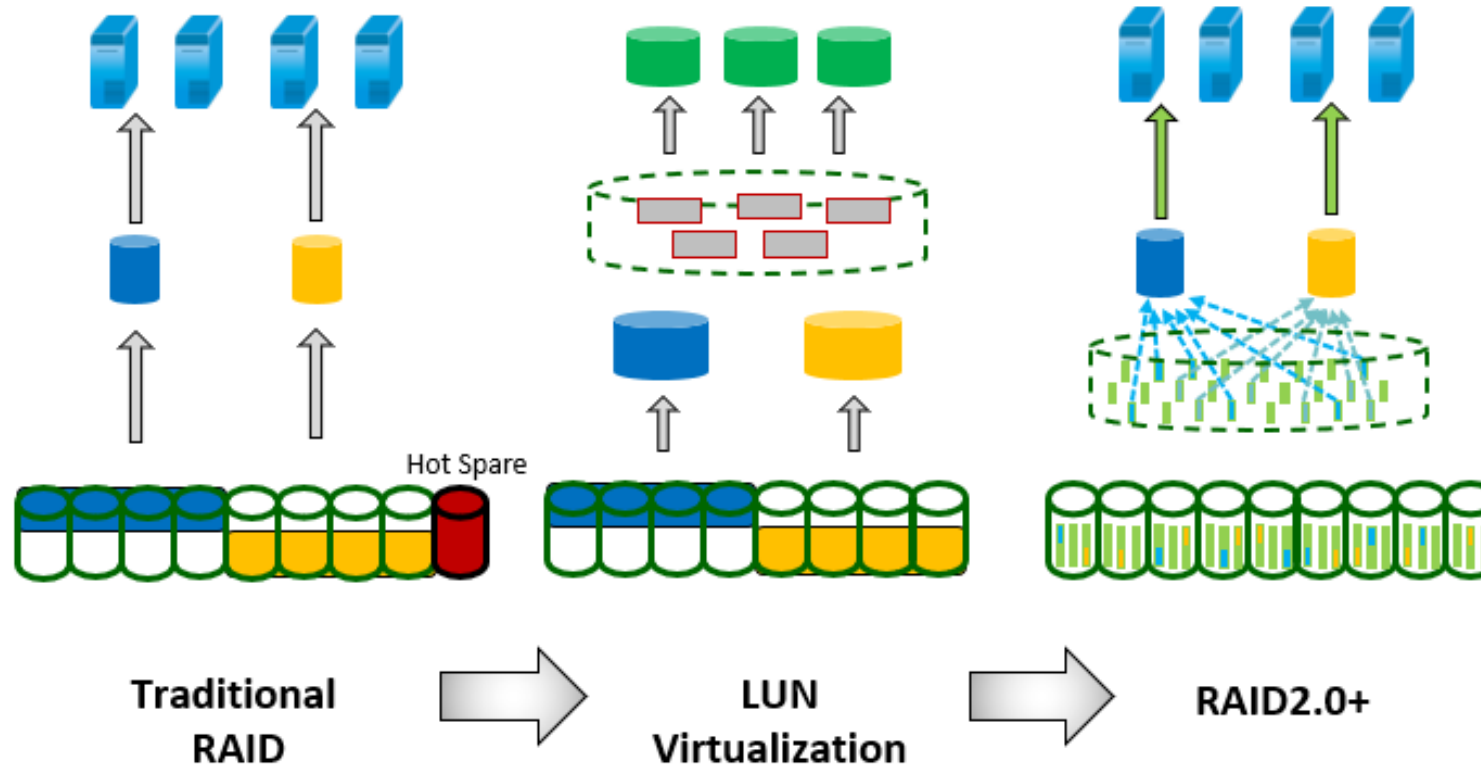


Chapter
06

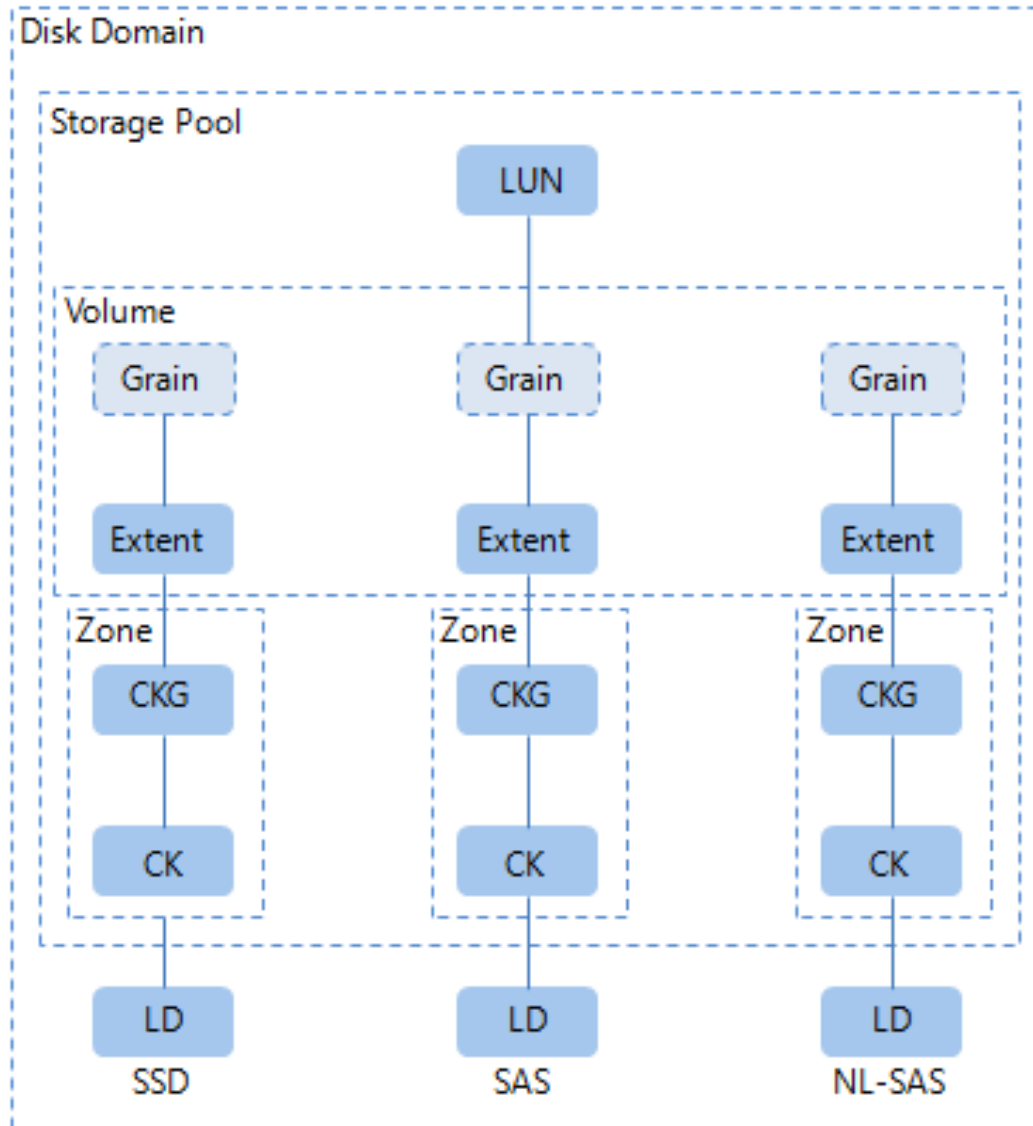
RAID 2.0+ Technologies



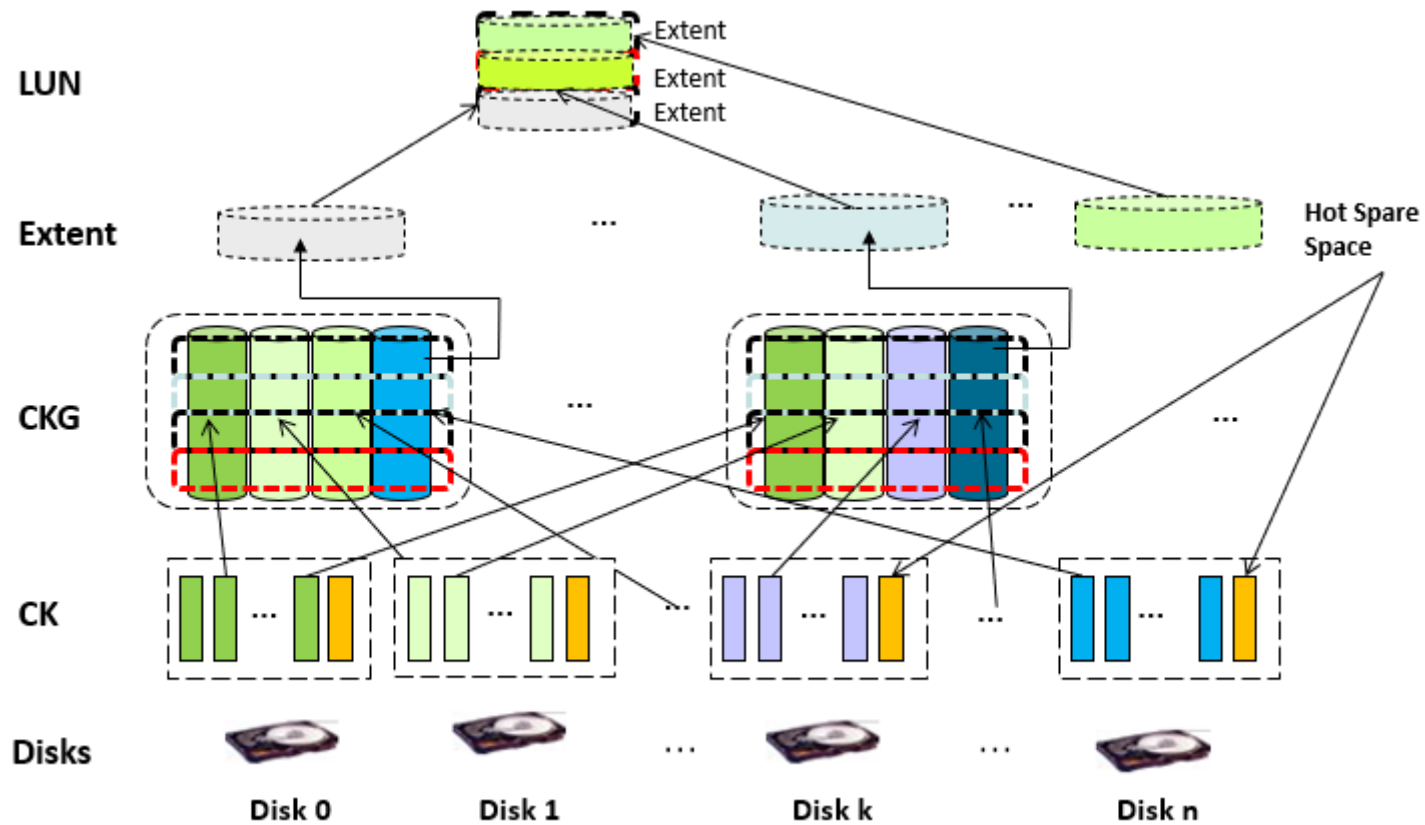
Development of RAID Technologies



Logical Objects of RAID 2.0+ Software

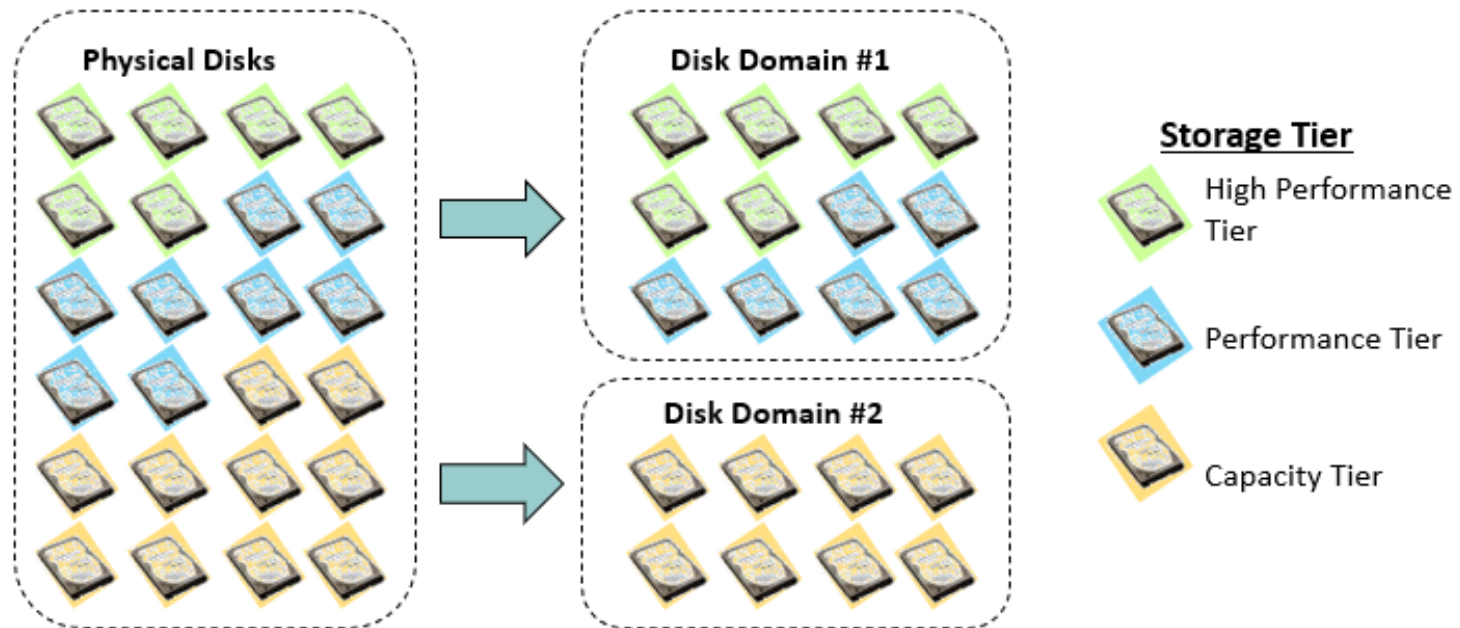


Basic Principles of RAID 2.0+



Disk Domain

- Disk Domain refers to a group of combined disk drives (can also be all the drives within the system). These drives capacity are combined with reserved hot spare space and then provided as a unified storage resource to the storage pool.



Storage Pool & Tier

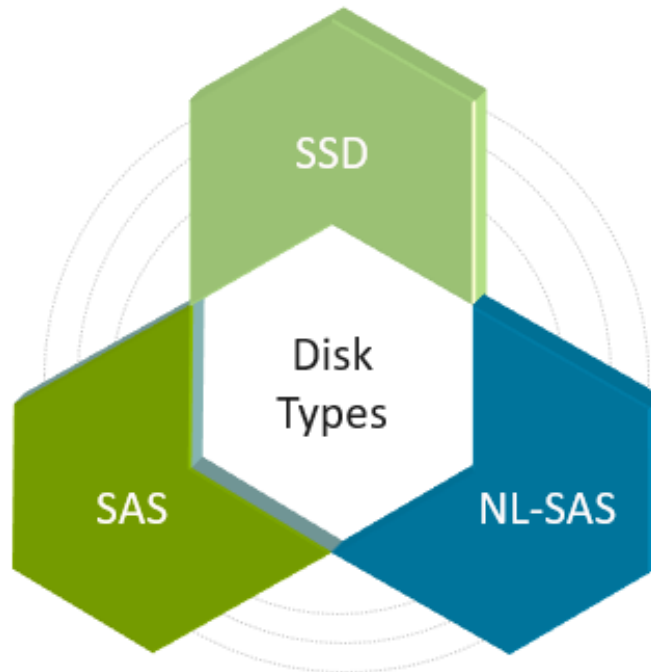
- Storage Pool is the container for storing storage resources. All the storage space used by servers are provided by the storage pool.
- Storage Tier is the integration of storage media with similar performance within the storage pool. It is used to manage the storage media with different performances and makes it easier to provide different storage space based on performance requirements.

Storage Tier	Tier Name	Supported Disk Type	Application
Tier0	High Performance Tier	SSD	Highest performance and cost, suitable for high frequency data access and data storage.
Tier1	Performance Tier	SAS	Medium performance and cost, suitable for medium frequency data access and data storage.
Tier2	Capacity Tier	NL-SAS	Low performance, low cost and high capacity. Suitable for large sized data and low frequency data access and data storage.

Disk Group (DG)

- Disk Group consist of the combination or integration of multiple similar types of hard drives within a Disk Domain.

The types of hard drives includes SSD, SAS and NL-SAS.



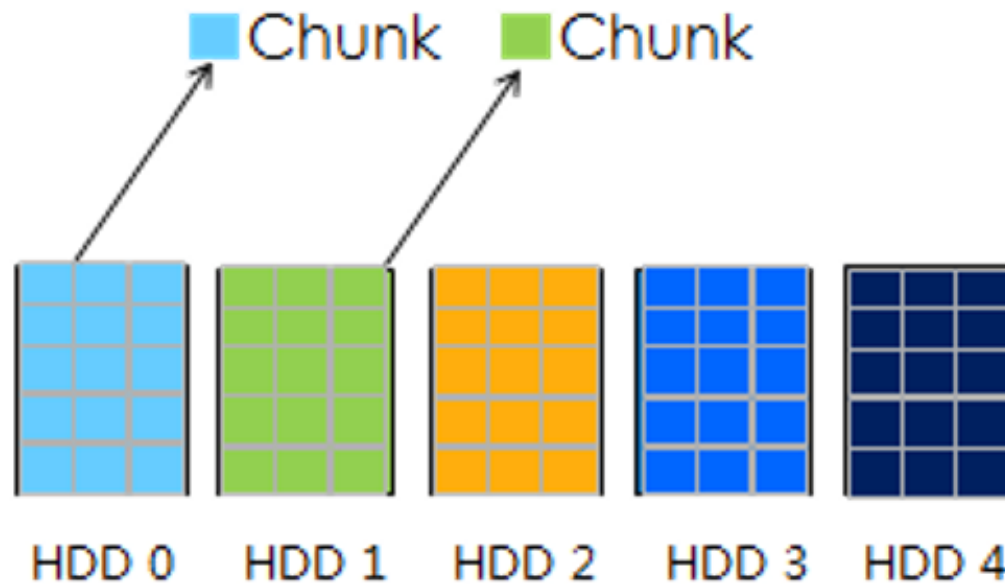
Logical Drive (LD)

- Logical Drive (LD) is the drive that is managed by the storage system which corresponds to the physical hard drive.



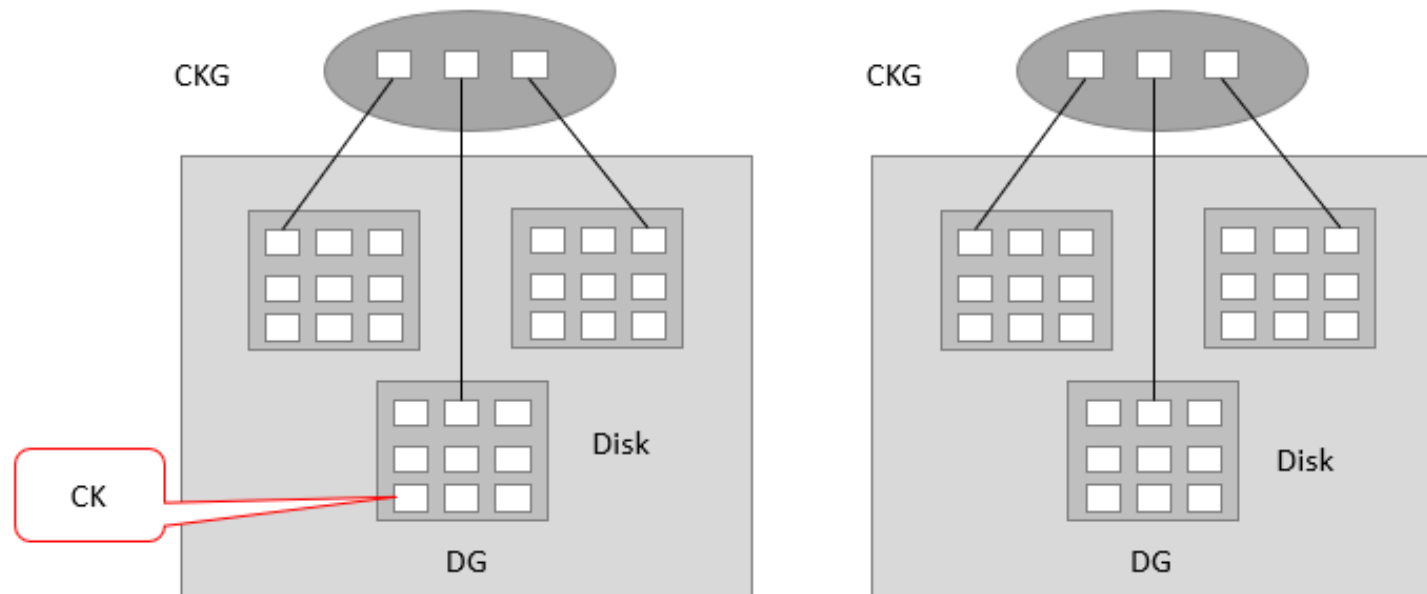
Chunk (CK)

- Chunk or abbreviated as CK, refers to the fixed size physical storage space that is divided from the hard drive capacity within the storage pool, and it is the basic unit for forming RAID.



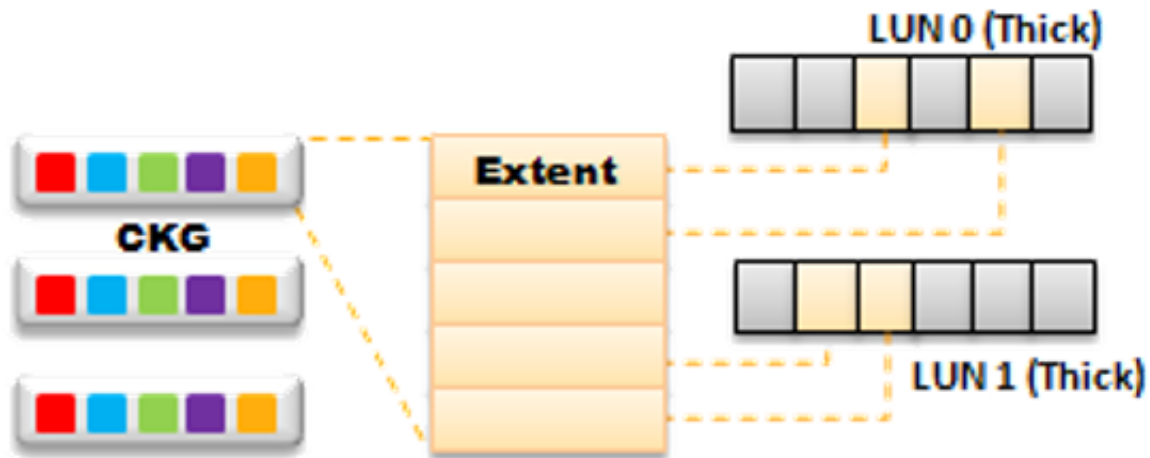
Chunk Group (CKG)

- Chunk Group (CKG), is formed from Chunks (CK) of multiple disks from the same Disk Group (DG). CK is the logical storage unit based on RAID, and it is the smallest unit for resource allocation in the Storage Pool.



Extent

- Extent is the logical storage space of fixed size that is formed using the CKG as the fundamental unit. The size of extent can be changed, and it is the smallest unit for hot data analysis and data migration (data migration granularity).



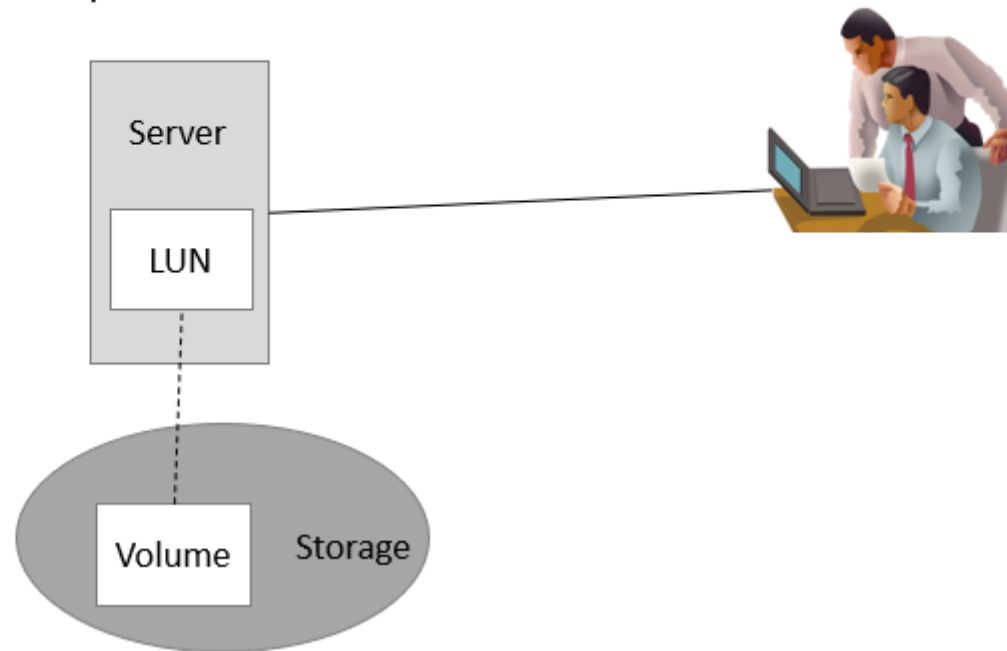
Grain

- In Thin LUN mode, the Extent is further divided into finer and smaller blocks. These blocks are called Grain. Thin LUN uses Grain for space allocation. The LBA within the Grain is continuous.



Volume & LUN

- Volume is the internal management object within a storage system.
- LUN is the storage unit that can be directly mapped to host for data read/write.
It is the external representation of Volume.

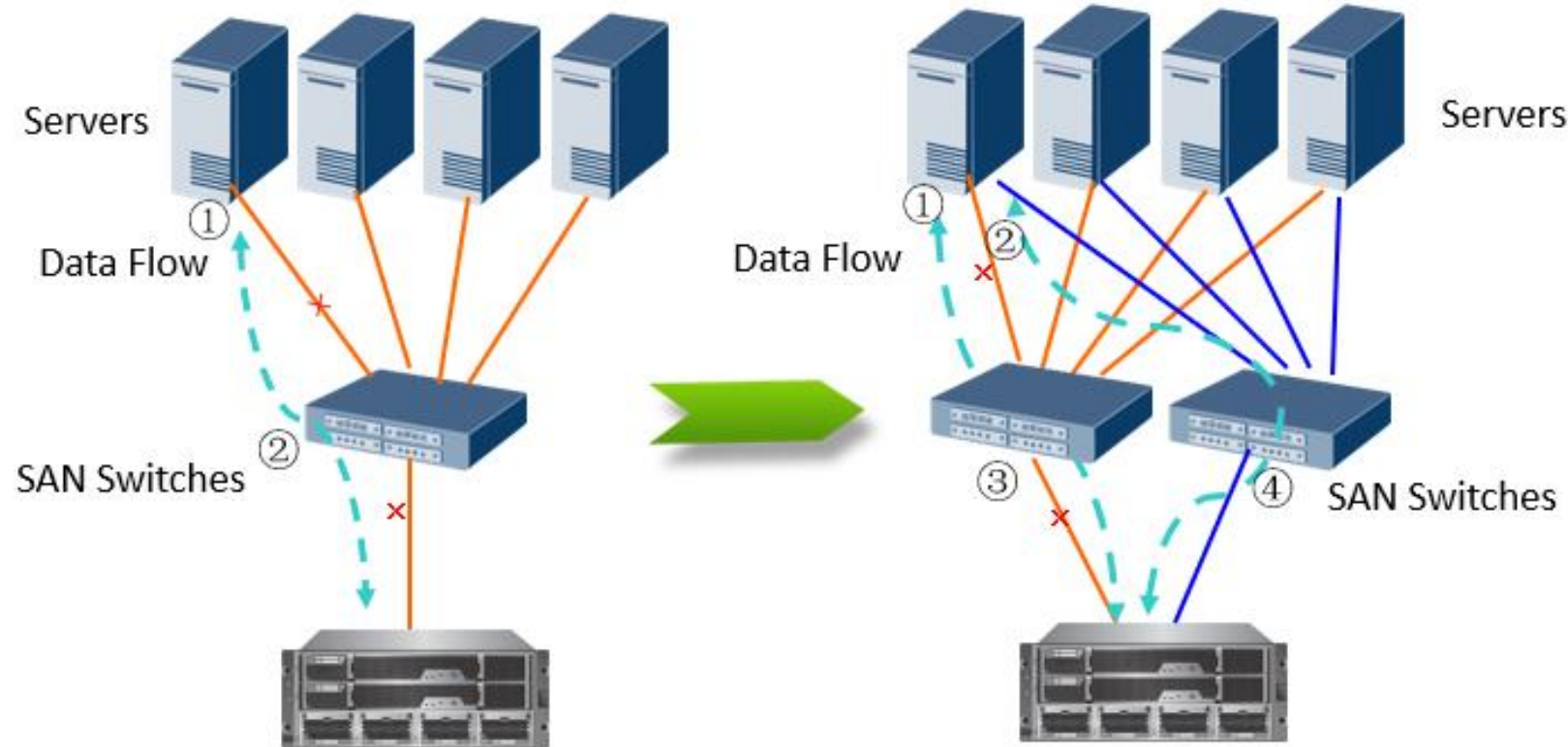


Chapter
06

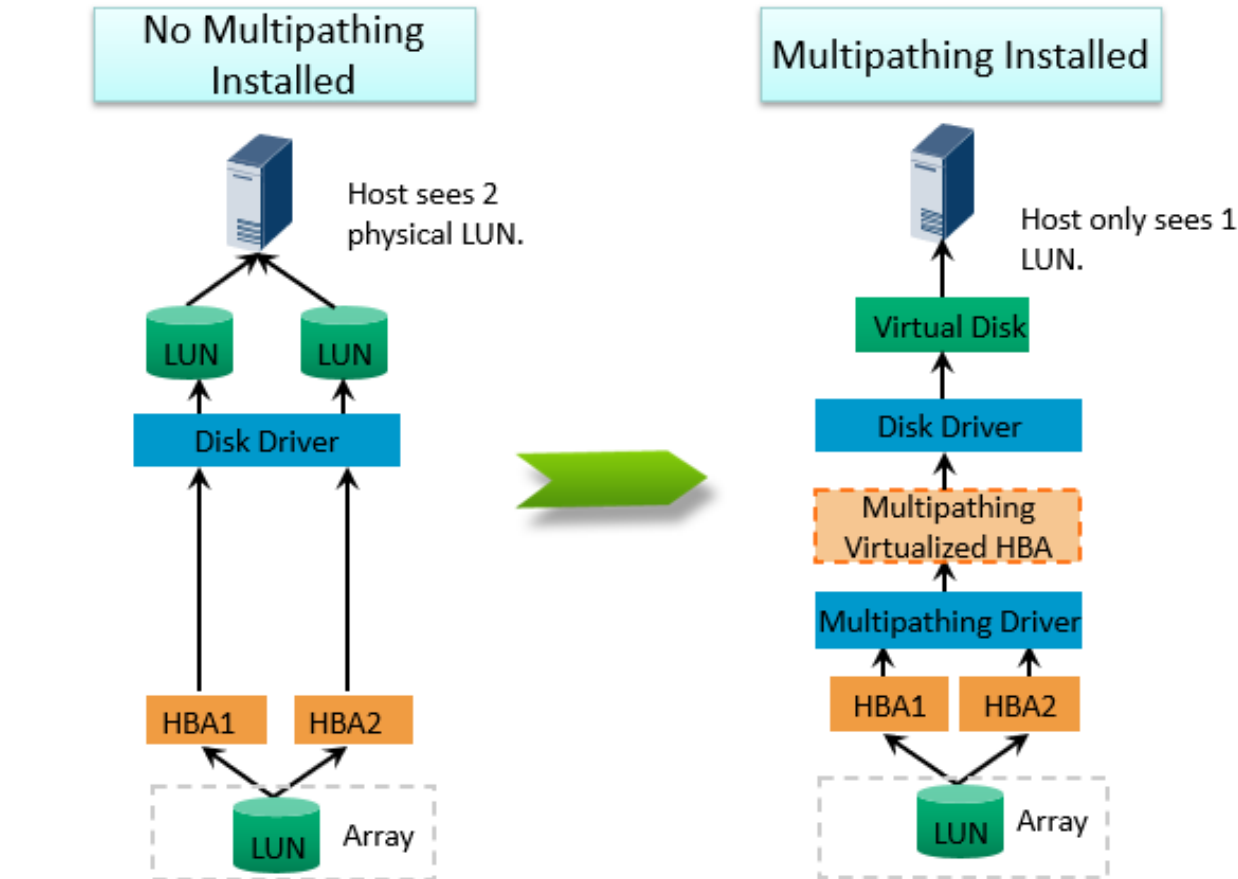
Multipathing Technologies



Multipathing Technologies

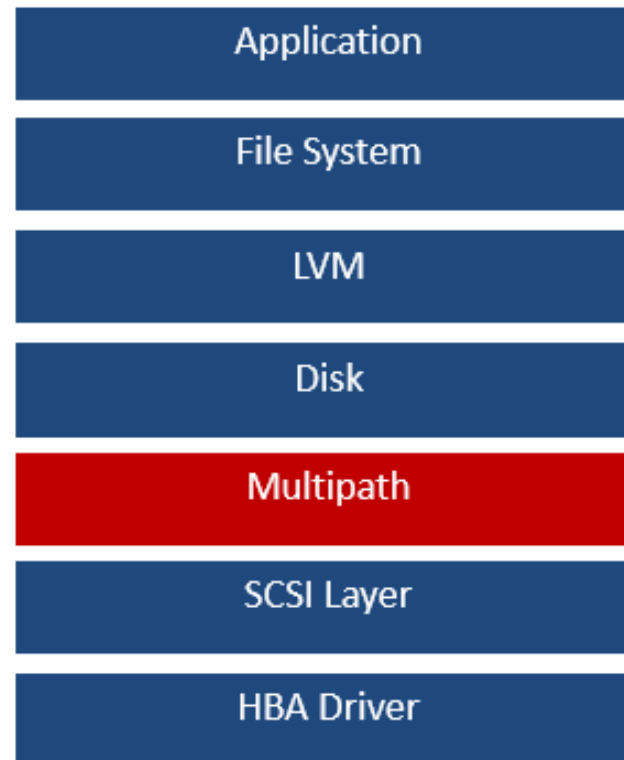


Working Principles of Multipathing Technologies

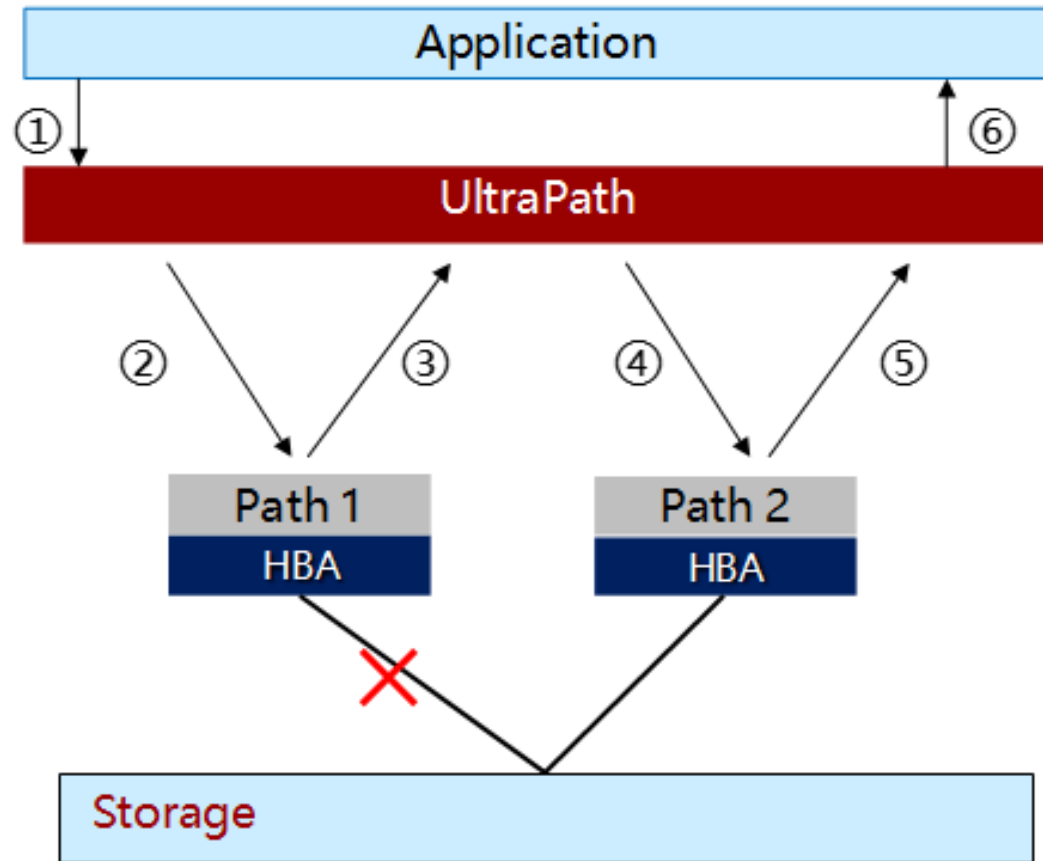


Integration Method Of Multipathing Software with the Operating System

- In Windows, Linux and Solaris platform, the layer where the UltraPath driver lies is shown on the diagram below:



UltraPath Feature - Failover



UltraPath Feature - Failback

- UltraPath will retransmit the I/O back to the normal route after detecting that the normal path has recovered from failure.
- There are two methods of path recovery:
 - For systems that supports hot plugging technologies (such as Windows), the broken path between the host and storage will cause the SCSI device to be removed, correspondingly when the link has recovered it will trigger the automatic connection of the SCSI device. In this scenario, UltraPath will detect the path recovery instantly.
 - For systems that doesn't support hot plugging technologies (such as AIX, earlier version of Linux), UltraPath detects path recovery in periodic interval cycle. This means that UltraPath will check the paths in the configured specific interval of time to see if the path is recovered or not.

Chapter
06

Disk Reliability Technologies



Disk Reliability

Components of a Hard Disk:



Mechanical Parts



Electrical Parts



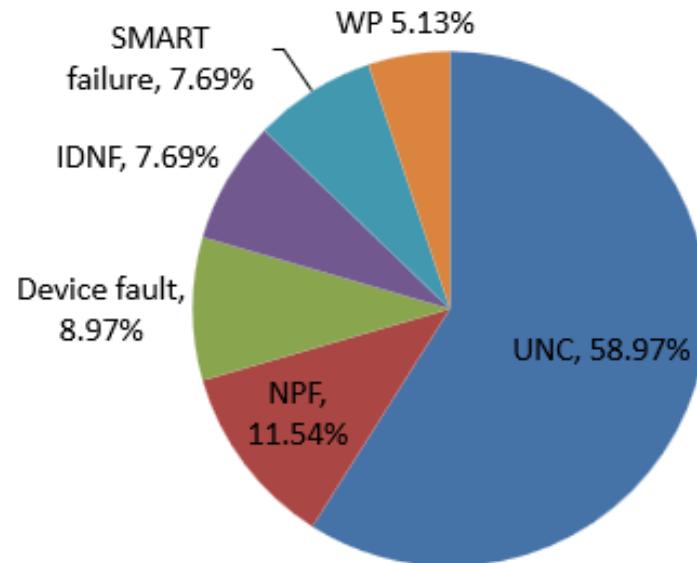
Thousands
of lines of
Firmware
codes.

Software

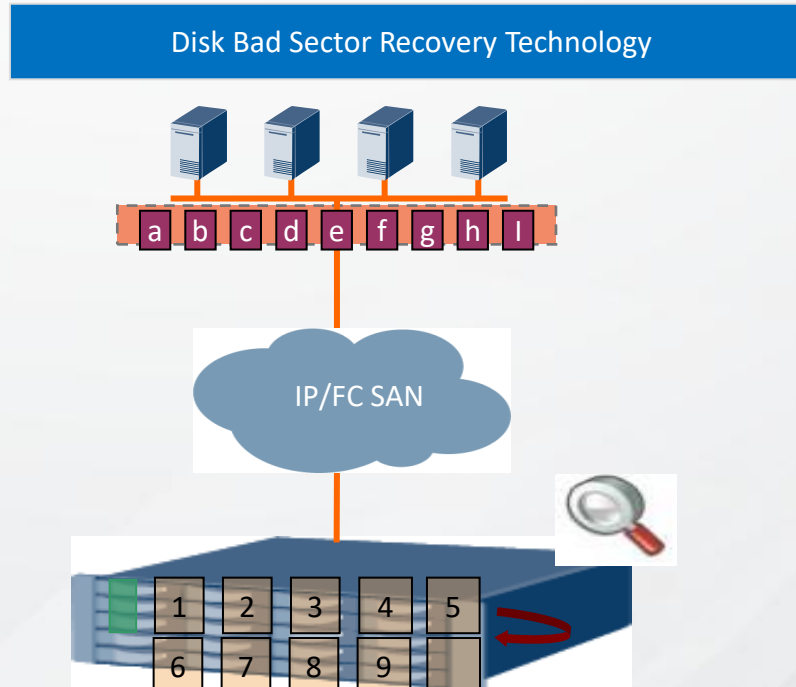


Types of Disk Failures

- TOP 1: UNC (Uncorrectable) about 60% of the disk errors are caused by bad sectors.
- TOP 2: NPF (No Problem Found) about 12% of the disk errors found no problems after the disk is returned to factory for analysis.
- TOP 3: Disks will predict failures before it happens, but the failure detection rate is just 7.69%, most of the failures are not able to be predicted in advance.



Disk Smart Scan & Bad Sector Recovery



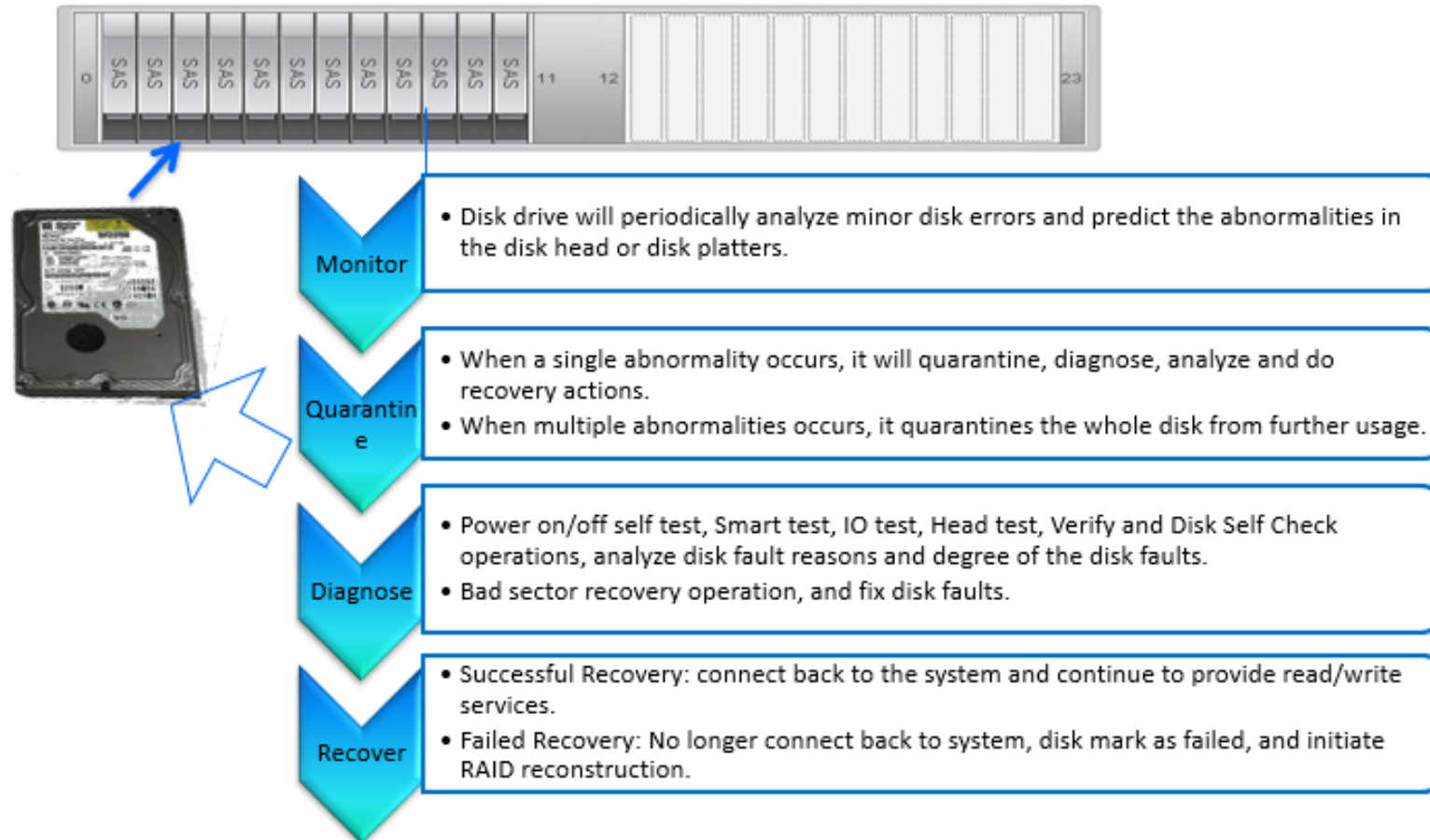
- **Technical Principle:**

- Characteristics that produces bad sectors: Temporal Locality, Spatial Locality, Data Decay.
- Storage systems periodically scans all the storage media to find any hidden errors.
- Scanning algorithm: cross scanning, dynamic rate adjustment.
- When the system finds a bad sector, it tries to restore the data using the redundancy capability of RAID, after it recovers the data, it writes it back to the disk.

- **Technical Advantage:**

- Detect hidden data safety risks and run recovery actions to minimize the overall system risks due to faulty drives.
- Increases the disk drives service life and protects the user investment.

Smart Online Disk Diagnostics System



Disk PreCopy

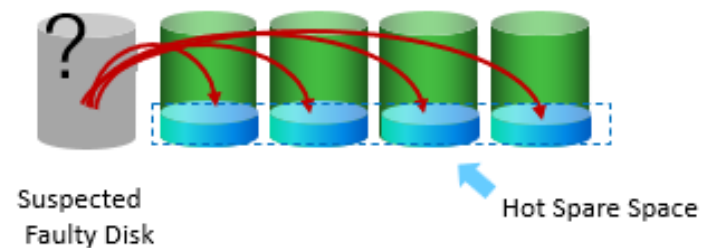
Working Principles of Disk PreCopy

- 1 In normal status, disk status are monitored in real-time.

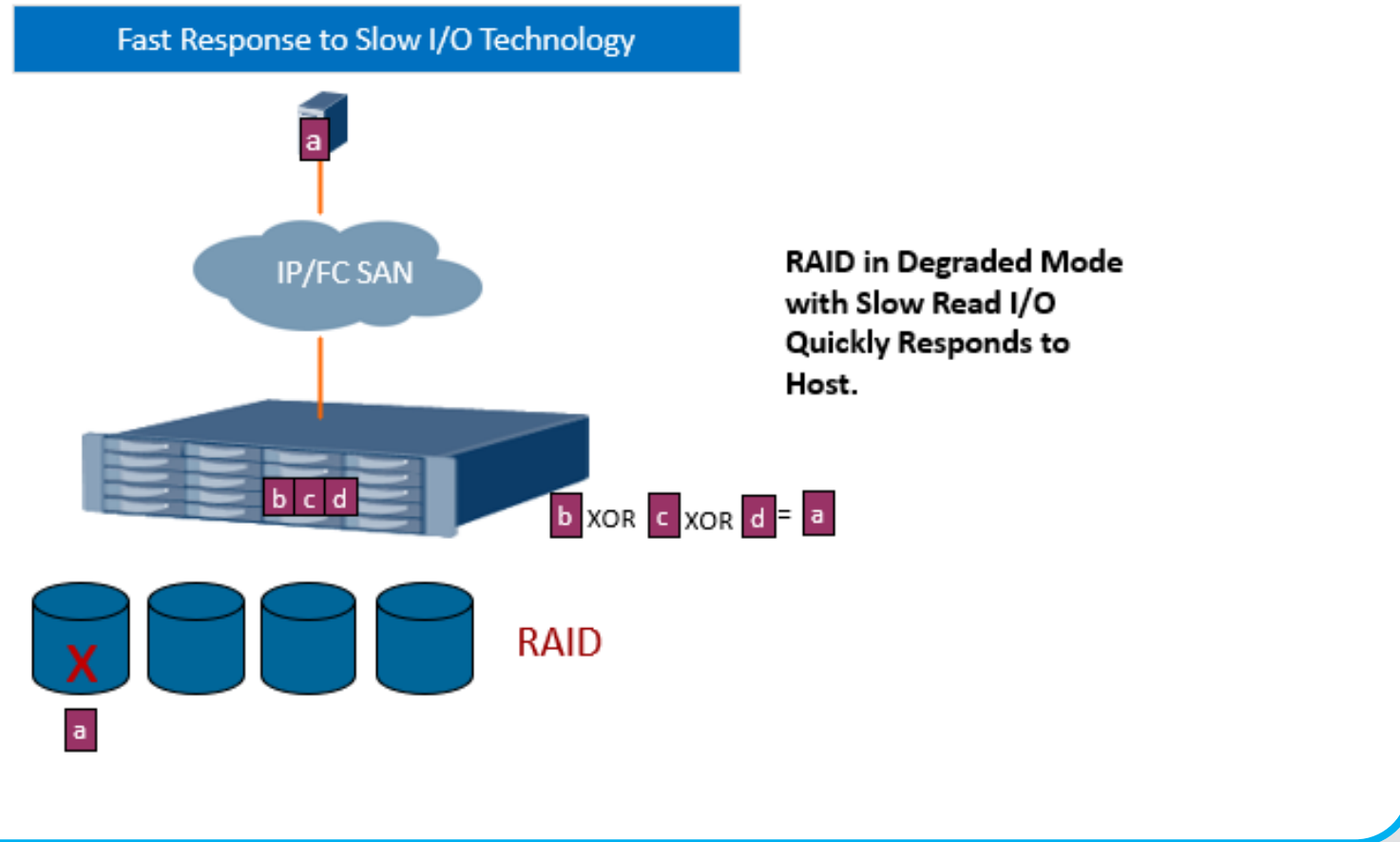


RAID PreCopy

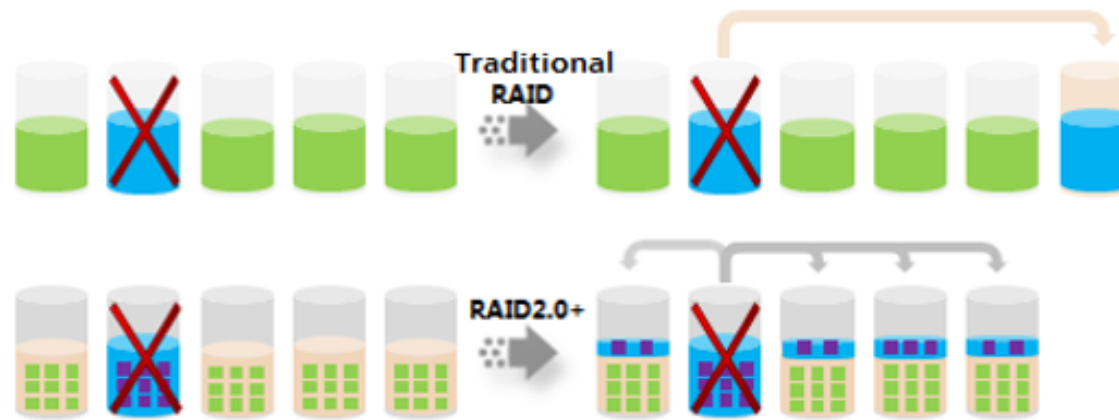
- 2 When a certain hard disk shows symptoms of failures: Move or migrate all the data within that disk to the hot spare space.



Fast Response to Slow I/O

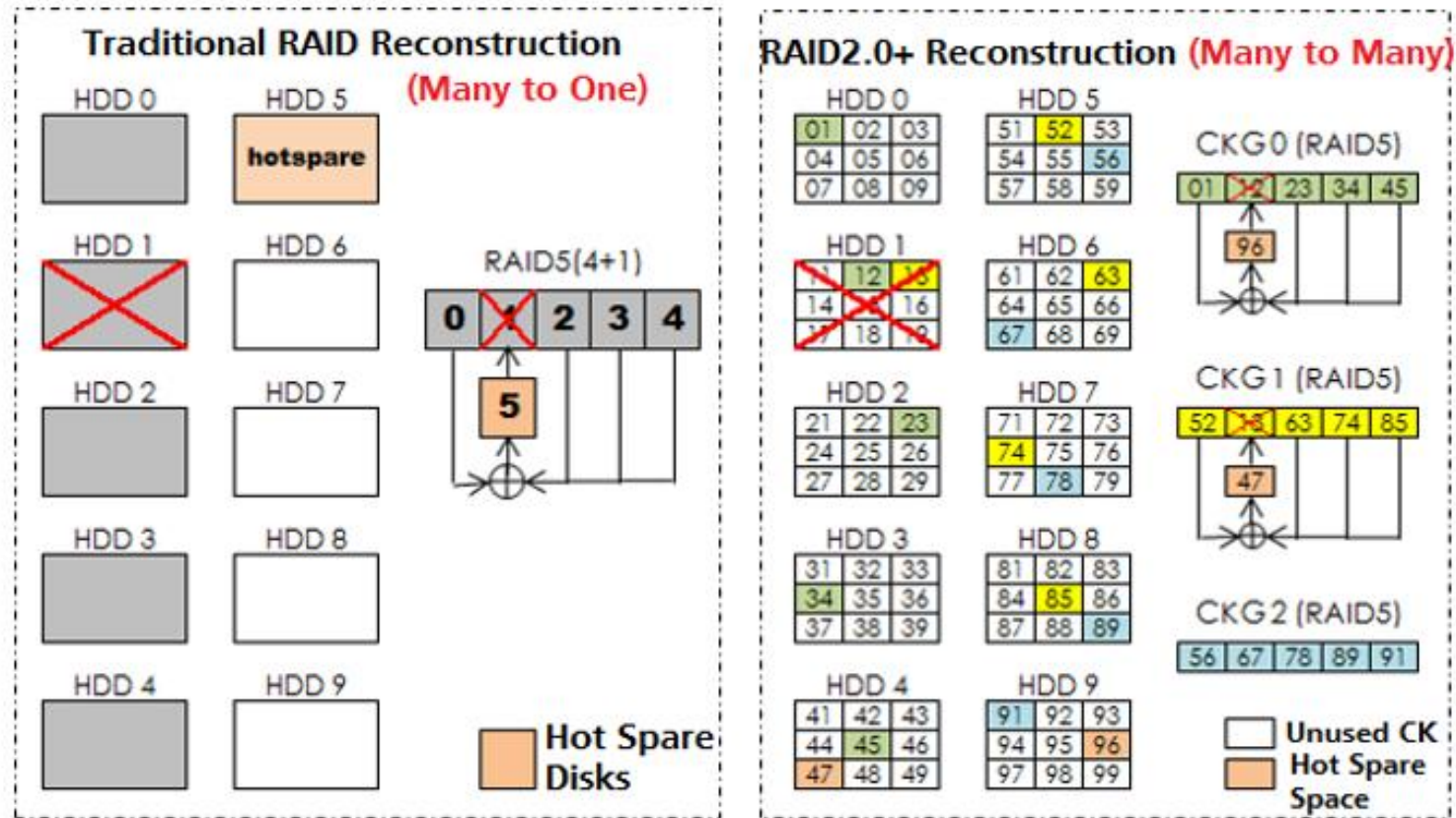


Overview of RAID Reconstruction Technologies



Traditional RAID	RAID2.0+
Need to manually configure the independent global or local hot spare disks.	Distributed hot spare space, no need for manual configuration.
Many to one reconstruction, where the reconstructed data are written in sequence to a single hot spare disk.	Many to many reconstruction, where reconstructed data is written in parallel to many disk drives.
Hot spot exists and longer reconstruction time.	Load balancing and short reconstruction time.

Technical Principles of RAID Reconstruction



- Note: The diagram above uses RAID 5 (4+1) as example, the working principles for other RAID type are similar.



Thank You