

Data Storage Technology

ผู้ช่วยศาสตราจารย์ ดร.สมเกียรติ โกศลสมบัติ

สาขาวิชาวิทยาศาสตร์และนวัตกรรมข้อมูล

วิทยาลัยสหวิทยาการ มหาวิทยาลัยธรรมศาสตร์

Contents

- Data Types
 - File Types
 - RAID
 - SAN
 - NAS
 - DAS
 - References
- 
- A series of four blue brush strokes of varying lengths and orientations, starting from the left margin and extending towards the right, positioned below the 'References' list item.

Data Types

Structured



Data is well organized.

Organized by the means of relational databases.

Matured transaction, multiple concurrency techniques.

Tuples, rows and tables.

Schema dependent and less flexible.

Query performance is the highest, structured query can be performed allowing complex joins.

Semi-structured



Data is organized to some extent.

Partially organized, e.g. by XML/RDF.

Transaction is adapted from DBMS, but data concurrency can pose problems.

Tuples or graphs are possible.

Data is more flexible than structured.

Queries over anonymous nodes are possible.

Unstructured



Data is fully non organized.

Based on character and binary data.

Difficult, but achievable transaction management and data concurrency.

Versioning usually on whole data or chunks.

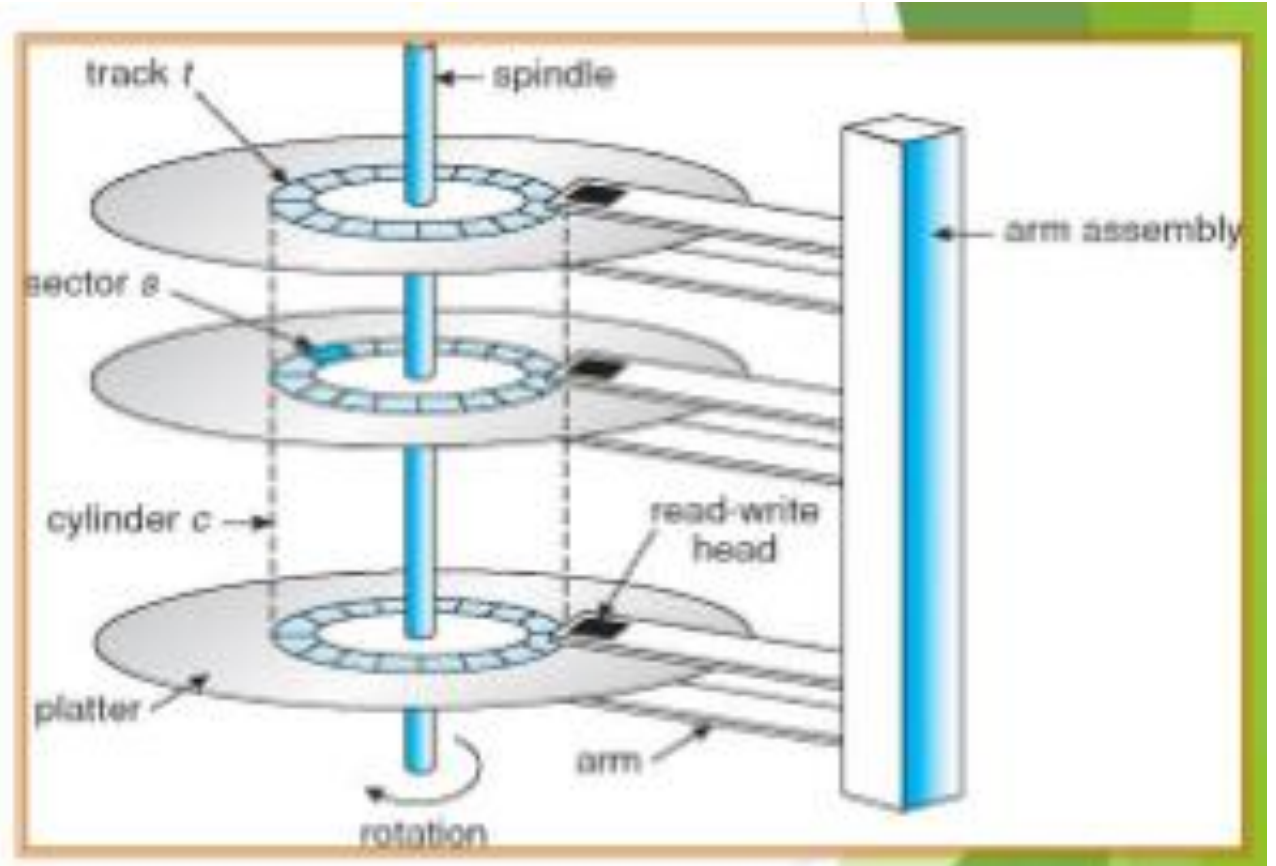
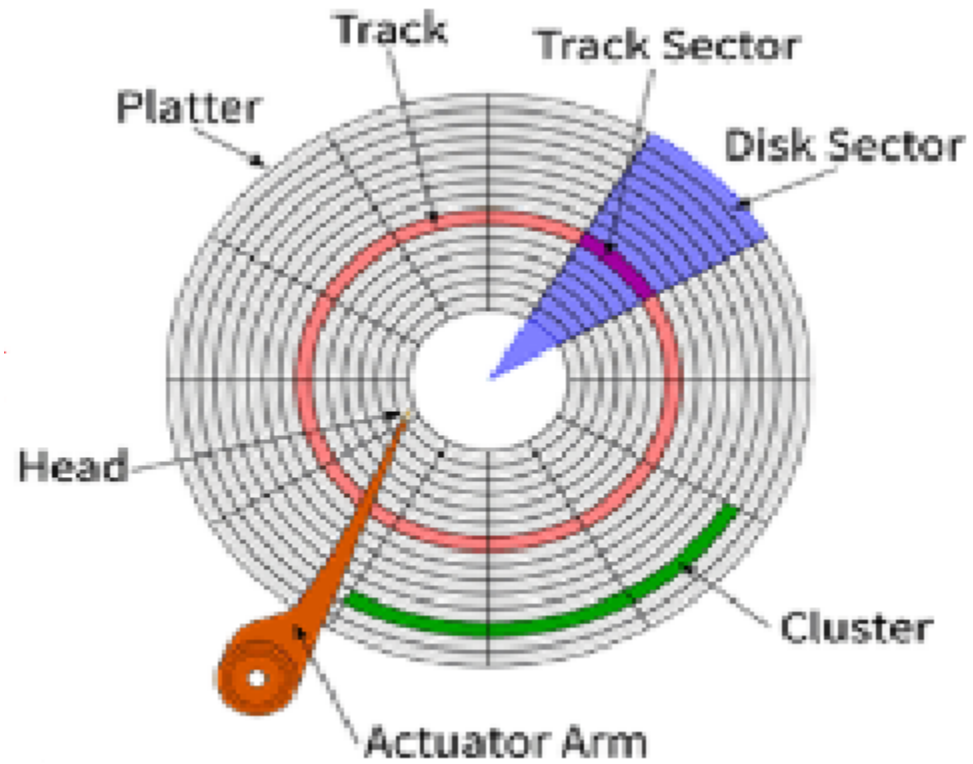
The most flexible.

Schema on-read so query performance is the lowest.

File Types

- Binary Files
 - ไฟล์ที่เก็บข้อมูลในรูปแบบเลขฐานสอง (0,1) โดยตรง แทนการใช้ตัวอักษร ASCII
 - ไม่สามารถอ่านได้โดยตรง ต้องใช้โปรแกรมเฉพาะในการเปิด
 - ไฟล์ .exe, รูปภาพ, เพลง, VDO, ฐานข้อมูล
- Text Files
 - ไฟล์ที่เก็บข้อมูลในรูปแบบตัวอักษร ASCII หรือ Unicode
 - สามารถอ่านได้ แก้ไขได้
 - ไฟล์ .txt, .csv, .html, .xml, .json
- Media Files
 - ไฟล์ดิจิทัลที่เก็บข้อมูล Multimedia
 - รูปภาพ, เสียง, VDO

File Storage



File Storage

- รูปแบบการจัดเก็บไฟล์เป็นแบบลำดับชั้น (Hierarchical) ได้แก่ File and Folder (Directory)
- Terabytes
- Simplicity
- File sharing
- High latency
- Response Time, IOPS and Throughput ต่ำ
- Scalability จำกัด

Block Storage

- รูปแบบการจัดเก็บข้อมูลลงบน Volume ที่ชี้ไปยัง Block ขนาดเล็ก ๆ
- Database and Virtualization
- Terabytes
- Low latency
- Response Time, IOPS and Throughput สูง
- Scalability ดี

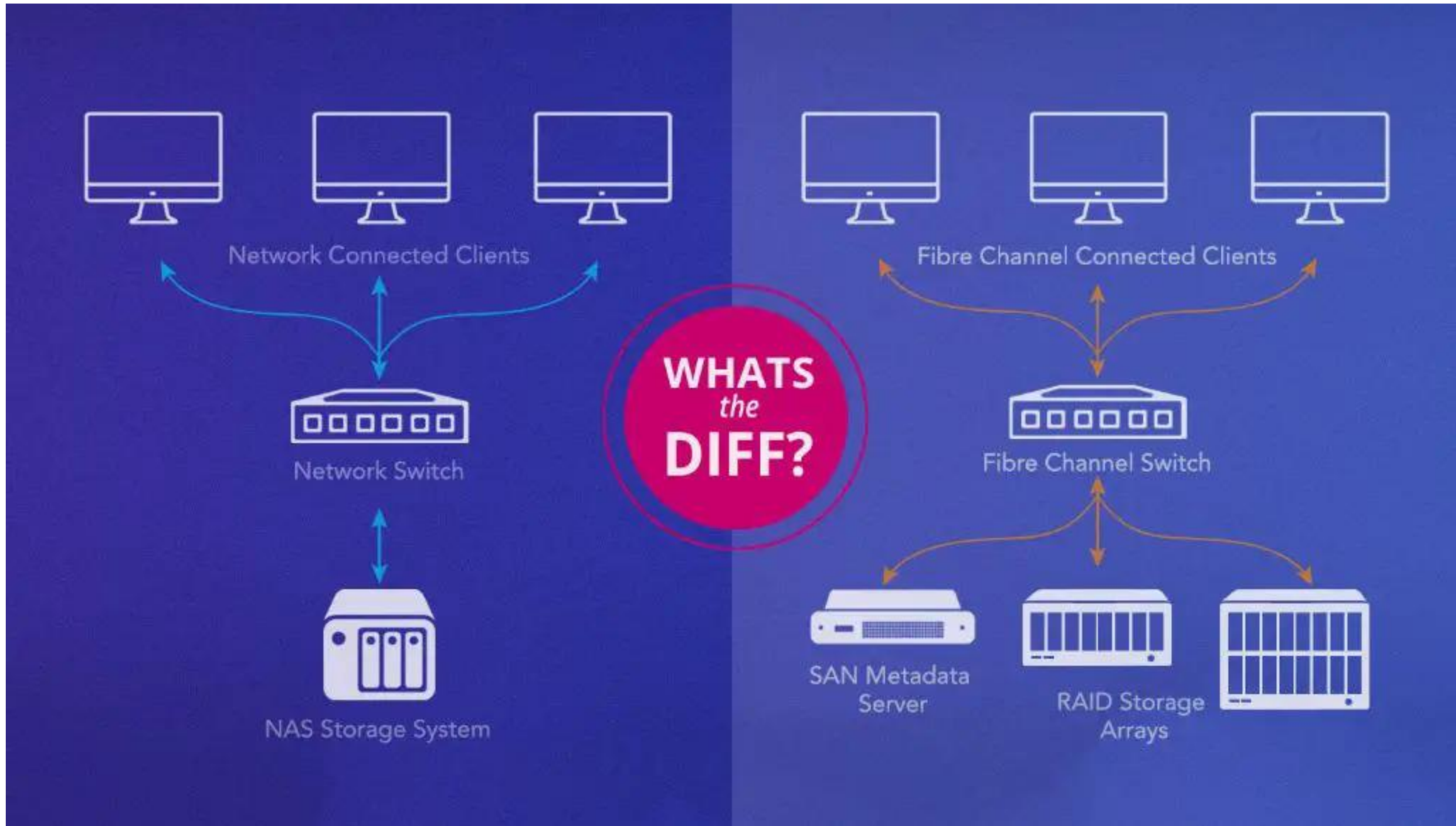
```
Filesystem      1M-blocks  Used  Available  Use% Mounted on
tmpfs            392         2       391        1% /run
/dev/mapper/ubuntu--vg-ubuntu--lv 14004    6986    6285       53% /
tmpfs           1956         0     1956        0% /dev/shm
tmpfs            5          0         5         0% /run/lock
/dev/sda2        1946       242    1586       14% /boot
tmpfs            392         1       392        1% /run/user/1000
ubuntu@ubuntu:~$
```

Object Storage

- รูปแบบการจัดเก็บข้อมูลแบบไม่มีลำดับชั้น ไม่มีโครงสร้าง
- ใช้ Tag ของข้อมูลด้วย Metadata ในการกำหนด
- Petabytes
- High latency มากกว่าประเภทอื่น ๆ
- Response Time, IOPS and Throughput กลาง แต่เน้นเรื่อง Flexibility สูง
- Scalability ดี

Cloud Storage ??

- การจัดเก็บข้อมูล / ฐานข้อมูล
- ประเภทของ Application
- การบริการ / การบริหาร
- งบประมาณ

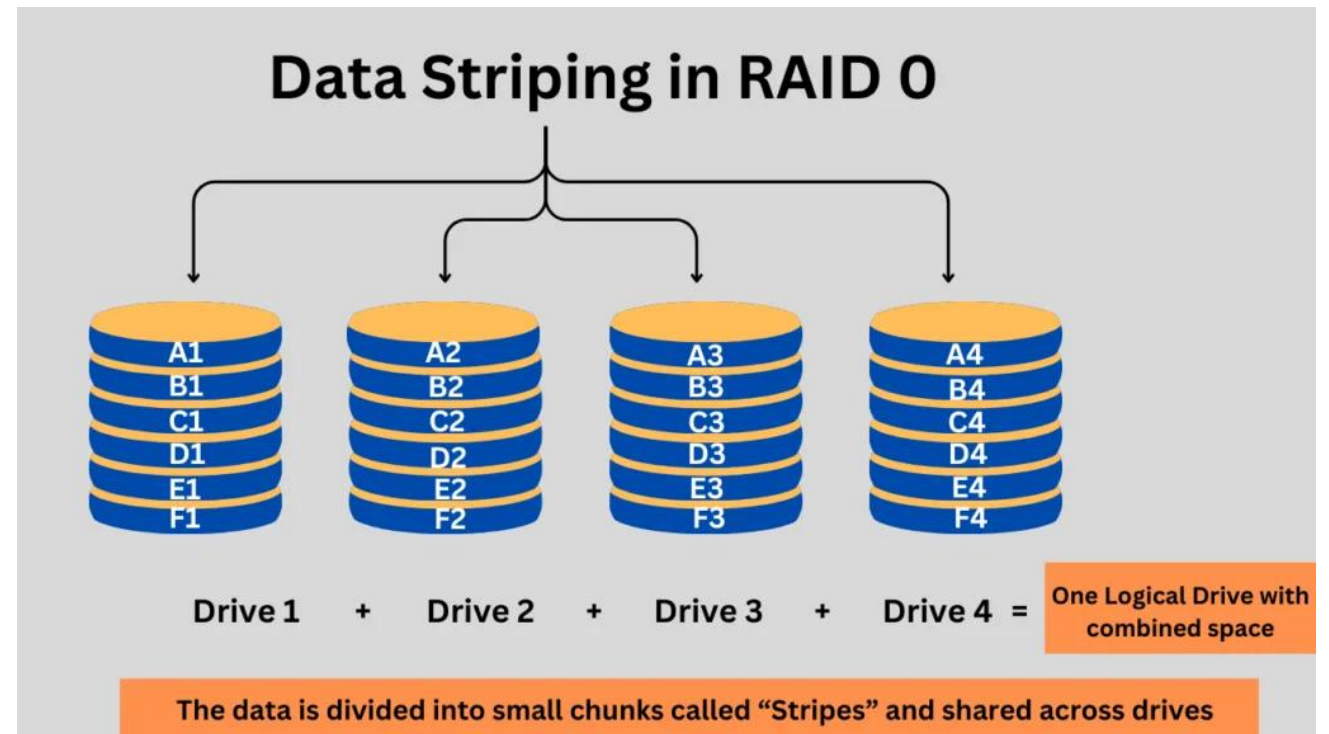
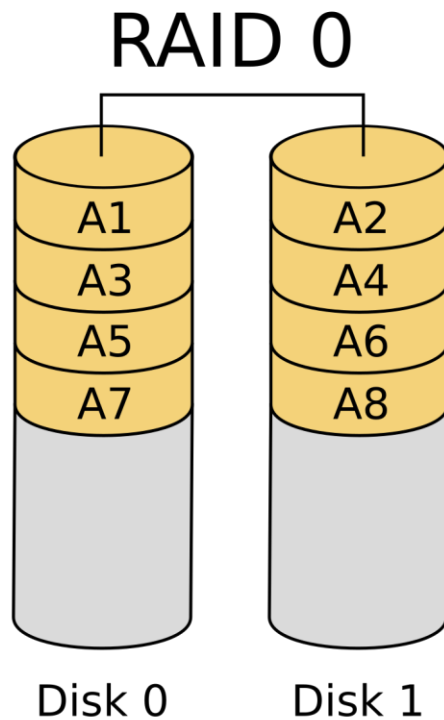


RAID

- Redundant Array of Independent Disks - RAID
- RAID is a technology that combines multiple disk drives to create a logical unit.
- Data is distributed across the drives for redundancy and performance.
- RAID protects against disk failures and provides improved read and write speeds.
- RAID 0 – Striping
- RAID 1 – Mirroring
- RAID 5 – Striping with Parity
- RAID 6 – Double Parity
- RAID 1+0 – Mirroring + Striping

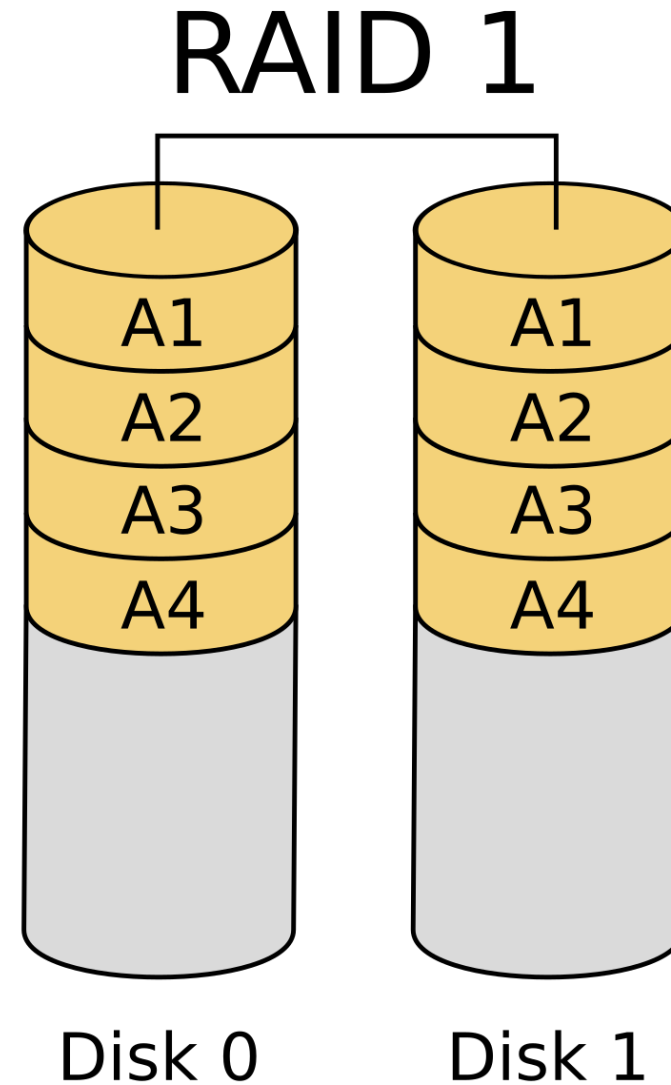
RAID 0 - Striping

- Data evenly across two or more disks without parity information redundancy or fault tolerance.

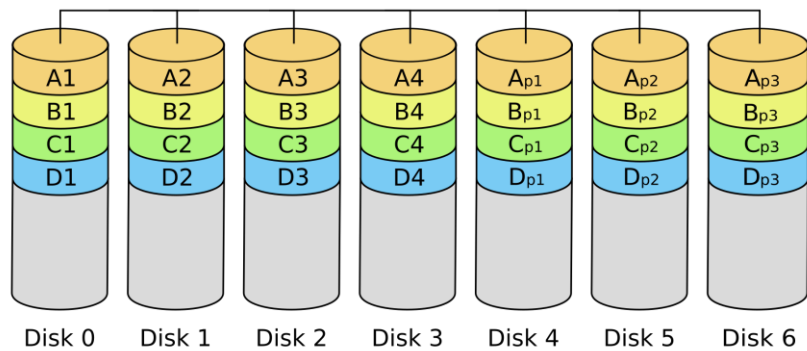


RAID 1 - Mirroring

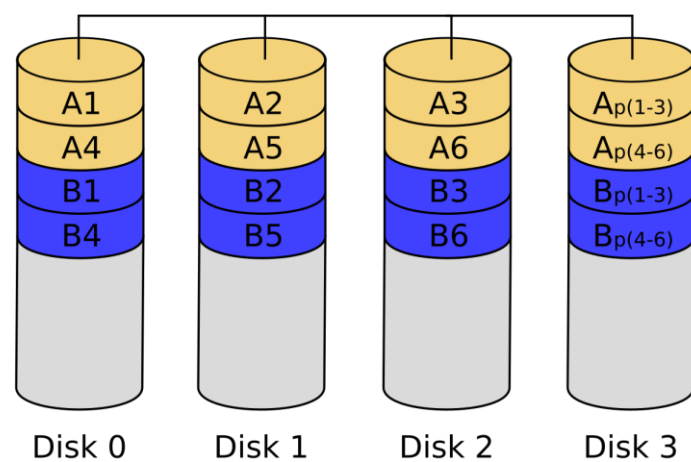
- The replication of data across two or more disks.



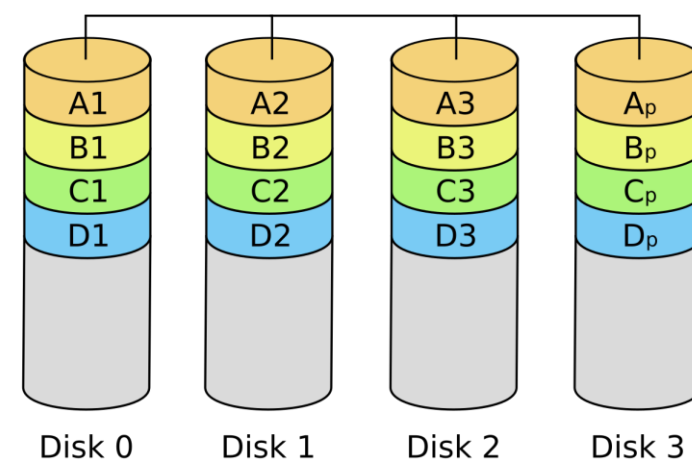
RAID 2



RAID 3

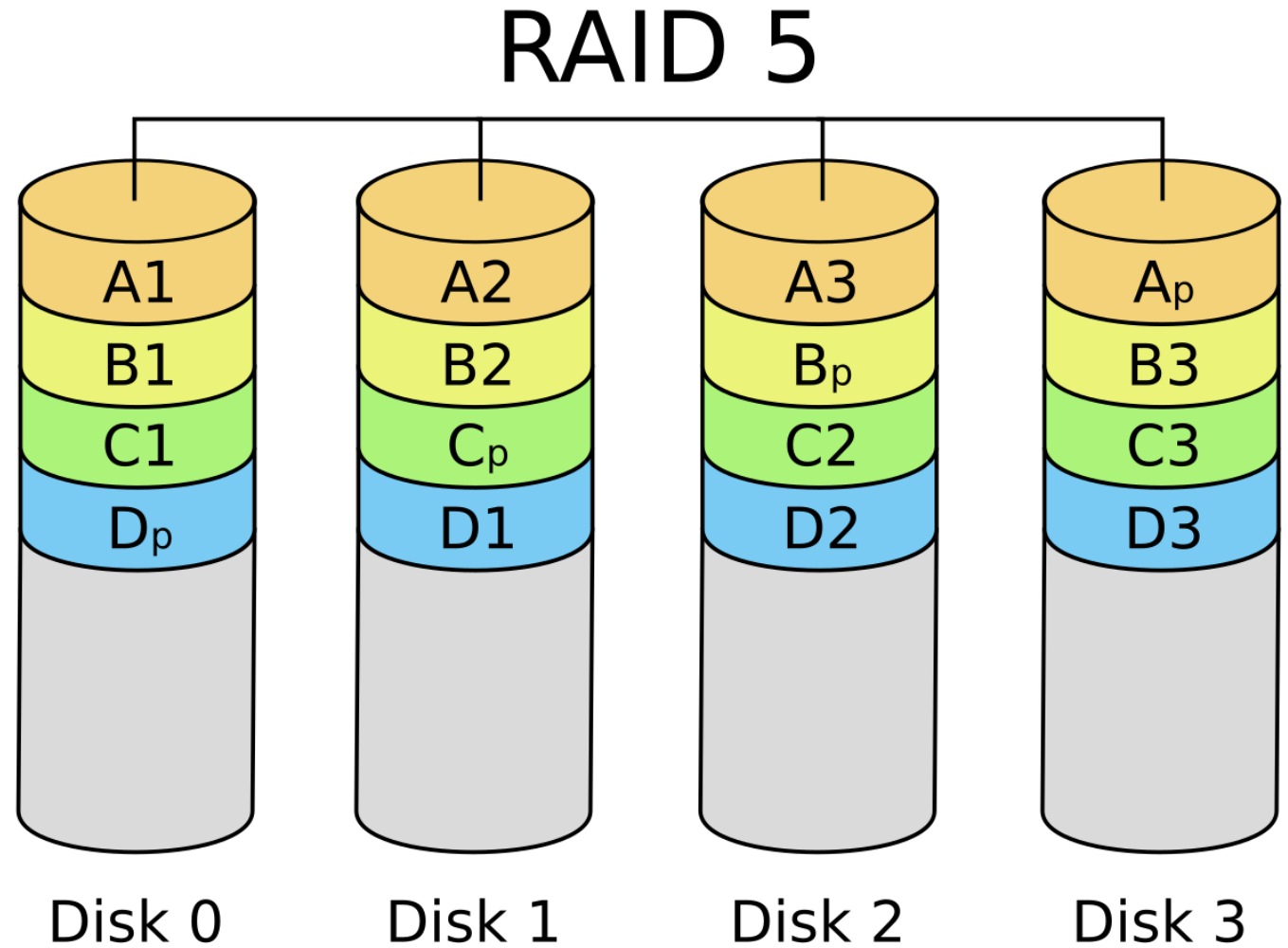


RAID 4



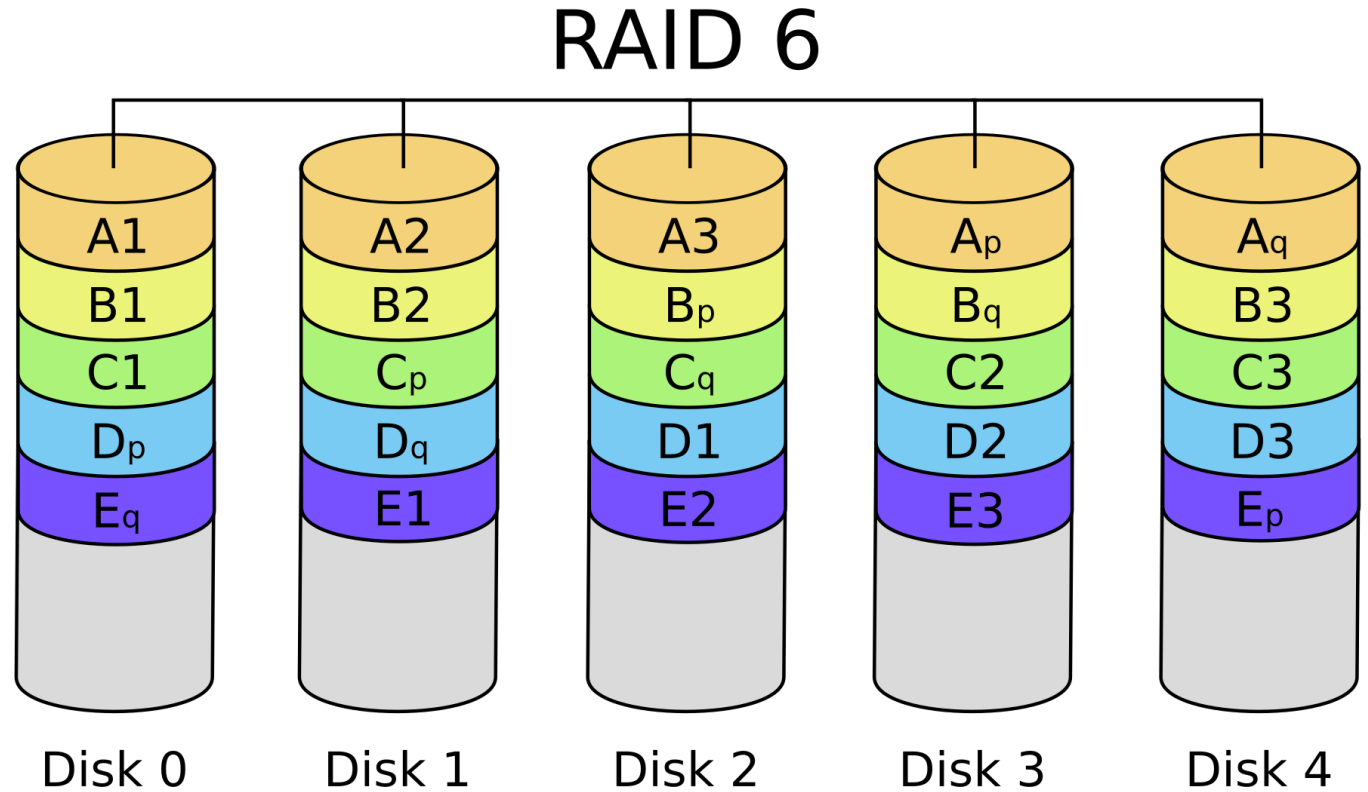
RAID 5 – Striping with Parity

- Consists of block-level striping with distributed parity

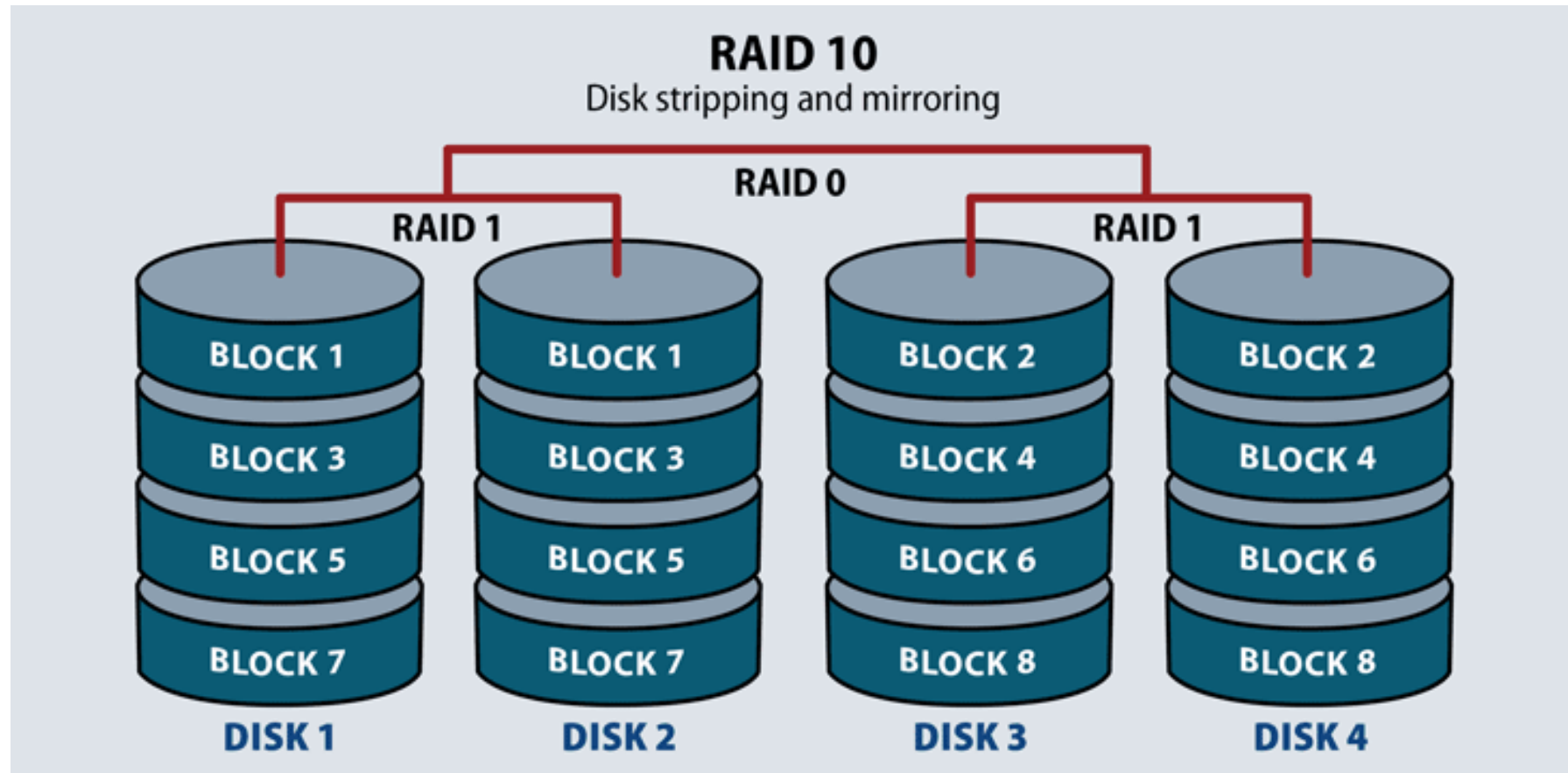


RAID 6 – Double Parity

- Extends RAID 5 by adding a second parity block

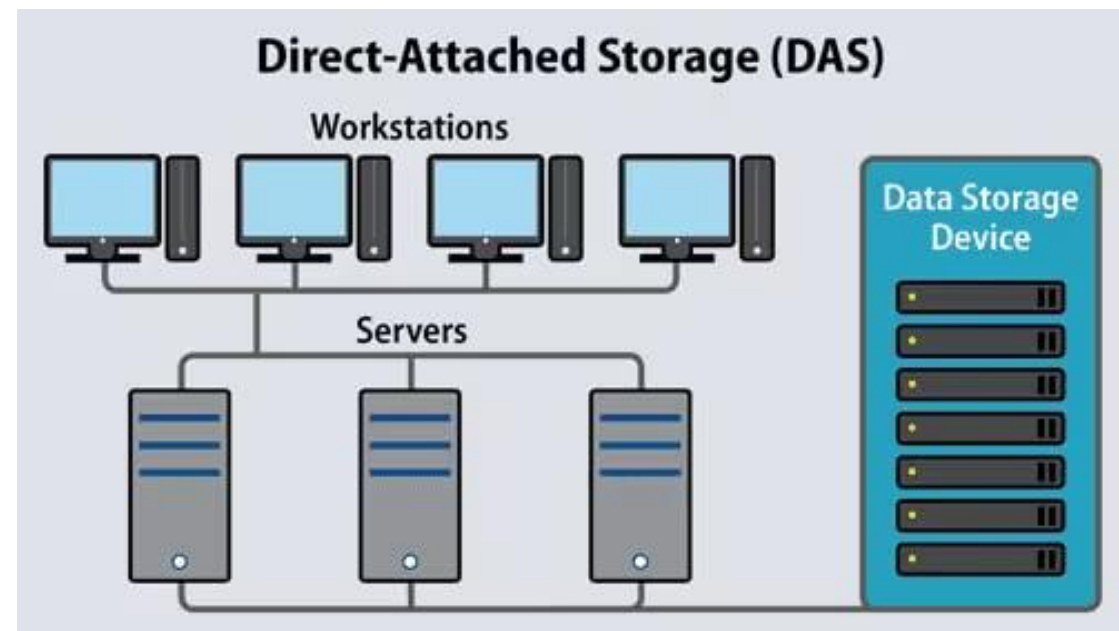
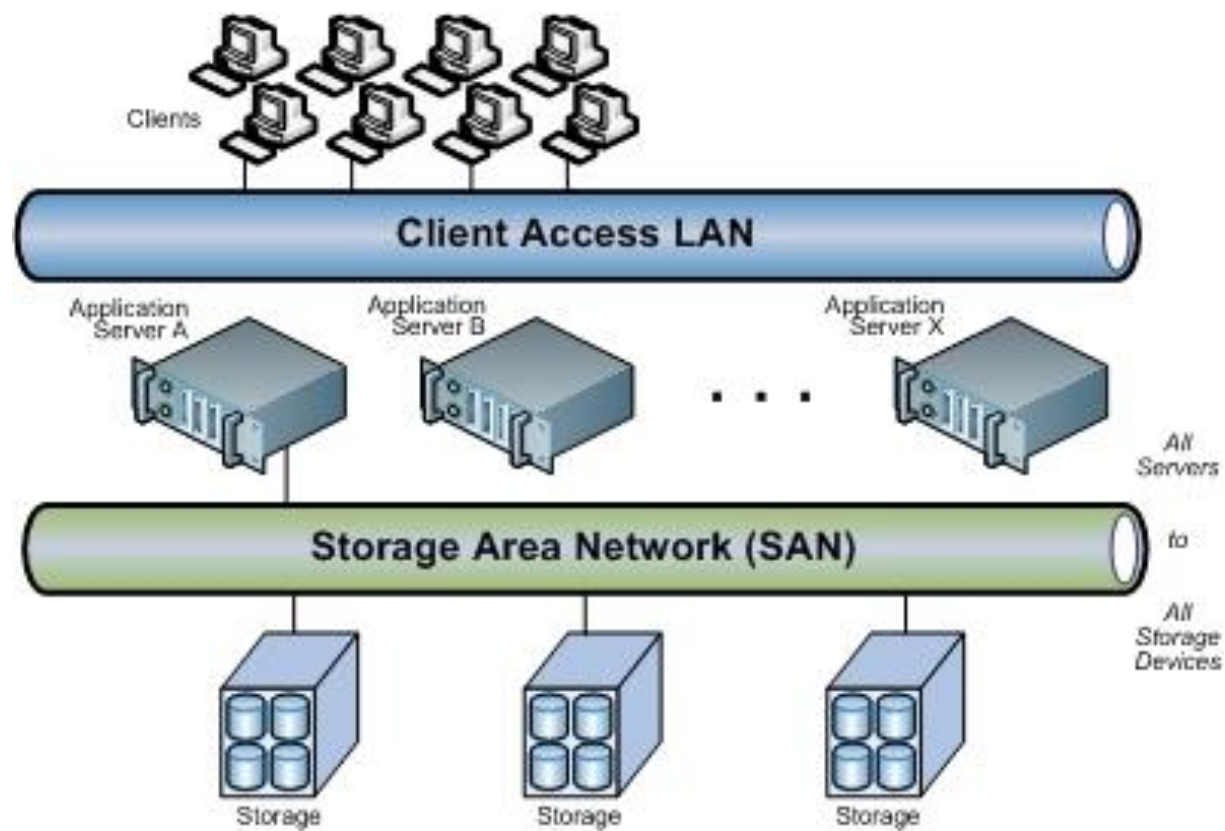


RAID 1+0 – Mirroring + Striping



SAN

- Storage Area Network
- SAN presents storage devices to a host such that the storage appears to be locally attached.
- SANs provide block-level storage rather than file-level storage.
- SANs allow multiple servers to connect to shared block-level storage devices.
- SANs are typically composed of host, switches, storage elements, and storage devices that are interconnected using a variety of technologies, topologies, and protocols.
- SAN is a dedicated high-speed network used for storage devices.
- SANs use a high-speed network, such as Fibre Channel, iSCSI, or Fibre Channel over Ethernet (FCoE), dedicated specifically for storage communication.
- This isolates storage traffic from general LAN traffic and enables much faster data access.
- SANs also allow centralized management, monitoring and backup of storage. Because SANs operate at the block level, the connected servers do not have to manage file-level operations.



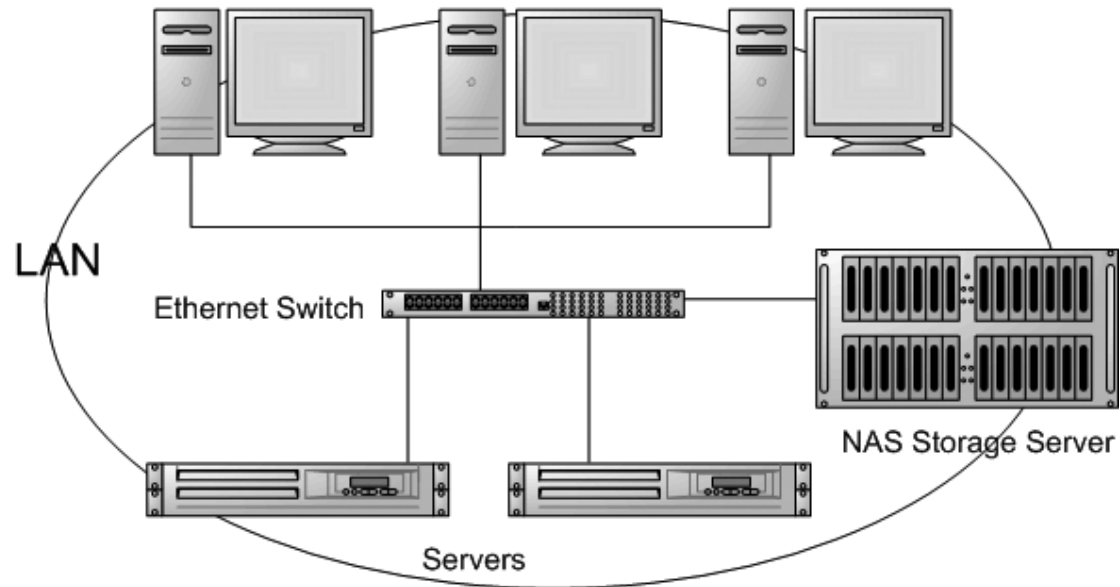
NAS

- Network attached storage
- NAS is a file-level storage device that enables access to data over a network.
- It acts as a file server but with added features like inbuilt storage drives, built-in RAID support, and management software.
- NAS is optimized for file sharing and data access over a network.
- Multiple users and client devices can connect to the NAS device via standard network protocols like SMB/CIFS, NFS, FTP or HTTP and access shared folders and files.
- Data is stored at the file level rather than block level.

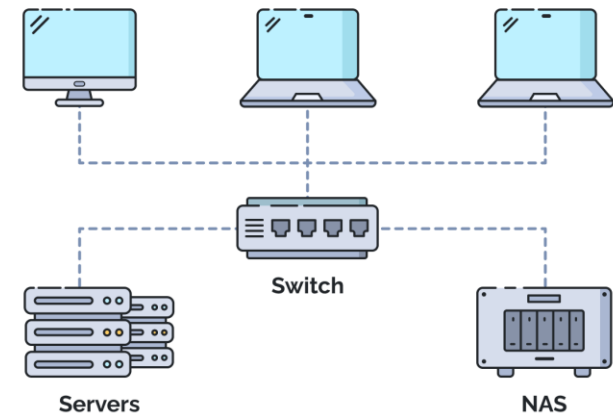
Network Attached Storage (NAS)

Network Attached Storage

Clients



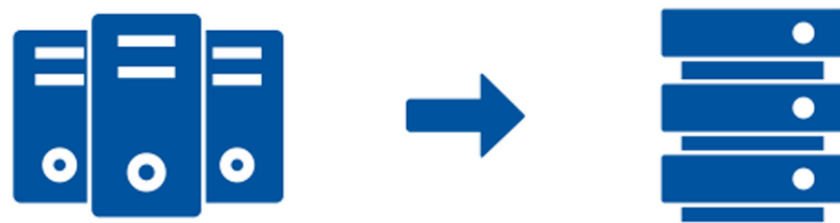
What is Network Attached Storage (NAS)



DAS



SAN



NAS



References

- <https://www.pecan.ai/blog/what-is-structured-data/>
- <https://en.wikipedia.org/wiki/RAID>
- <https://darwinsdata.com/what-is-raid-vs-san-vs-nas/>
- <https://arash-hatami.ir/content/images/2020/08/novalan-almacenamiento-de-informacion-das-san-nas.png>