COMP283-Lecture 3

Applied Database Management

Introduction	
DB Design	Continued…
	Disk Sizing
	Disk Types & Controllers
	DB Capacity

COMP283-Lecture 3

DB Storage: Linear Growth

- Disk space requirements increases by a constant amount for a given period:
  $F = C + (R * T)$

  F: Future disk requirement
  C: Current disk requirement
  R: Rate of growth
  T: Time period
- e.g. F = 500GB + (200GB/year * 5 years) = 1500GB

COMP283-Lecture 3

DB Storage: Compound Growth

- Disk space requirements increases by a constant fraction for a given period:

  $$F = C * (1 + R)^T$$

  F: Future disk requirement
  C: Current disk requirement
  R: Rate of growth
  T: Time period

- e.g. $F = 500GB * (1 + 25\% \text{ per year})^{5 years} = 1526GB$

COMP283-Lecture 3

DB Storage: Geometric Growth

- Disk space requirements increases by a regularly changing fraction for a given period:

  $$F = C + \frac{(Inc * (1 - R^{(1+T)}))}{(1-R)}$$

  F: Future disk requirement
  C: Current disk requirement
  Inc: Initial Increment
  R: Rate of change of growth
  T: Time period

- e.g. $F = 500GB + \frac{(100 * (1 - (1.25 \text{ per year})^{(5years+1)}))}{(1 - 1.25 \text{ per year})} = 1626GB$

COMP283-Lecture 3

DB Storage: Other Storage Considerations

- Server should have sufficient physical disks or volumes to allow separation of:
- Host OS (& DBMS programs and config files)
- Data
- Transaction Logs
- TempDB
- Backups.

Comp283-Lecture 3

DB Storage: Disk Types and Controllers

• IDE/PATA	Integrated Drive Electronics/ Parallel AT Attachment.
• SCSI	Small Computer System Interface.
• SATA	Serial Advanced Technology Attachment.
• SAS	Serial Attached SCSI.
• iSCSI	Internet Small Computer Systems Interface
• RAID	Redundant Array of Inexpensive Disks.

SATA 3.x most likely type of disk these days

RAID most likely to be used for important data

DB Storage: HD vs SSD vs SSHD

- HD - spinning rigid platters covered with magnetic material and read by a magnetic head on a moving actuator arm.
  - Now fairly cheap £67-£90 / 3TB
- SSD - Solid State Disks
  - no moving parts (therefore fast!)
  - Typically use NAND Flash memory
  - Expensive ~ £120 / 500GB
- SSHD - Solid State Hard Drive aka Hybrid Drive.
  - £60 / 1TB. Priced between HD and SSD.
  - Part SSD, part (mostly) HD
  - OS can optimise material stored on SSD part to speed up whole system.

Comp283-Lecture 3

DB Storage: HD vs SSD vs SSHD

- Time taken to access data on a Hard Disk:
  - Seek time is a measure of how long it takes the head assembly to travel to the track of the disk that contains data.
  - Rotational latency is incurred because the desired disk sector may not be directly under the head when data transfer is requested.
  - These two delays are on the order of milliseconds each.

Comp283-Lecture 3

DB Storage: HD vs SSD vs SSHD

- Average seek time ranges from under 4 ms for high-end server drives to 15 ms for mobile drives
- Most common mobile drives are about 12 ms
- Most common desktop type typically are around 9 ms.
- The first HDD had an average seek time of about 600 ms !!
- Faster seek rates typically require more energy usage to quickly move the heads across the platter, causing louder noises from the pivot bearing and greater device vibrations as the heads are rapidly accelerated during the start of the seek motion and decelerated at the end of the seek motion.
- Quiet operation reduces movement speed and acceleration rates, but at a cost of reduced seek performance.

Comp283-Lecture 3

DB Storage: HD vs SSD vs SSHD

- Disk Latency for Hard Disks
  - Delay for the rotation of the disk to bring the required disk sector under the read-write mechanism.
  - Depends on rotational speed of a disk, measured in revolutions per minute (rpm).
  - Based on the statistical relation that the average latency in milliseconds for a drive is one-half the rotational period.

$$\text{latency} = \frac{1}{\frac{R \text{ rotations}}{\text{minute}}} \times \frac{60 \text{ seconds}}{1 \text{ minute}} \times \frac{1000 \text{ ms}}{1 \text{ second}} \times \frac{1 \text{ rotation}}{2 \text{ half-rotations}} = \frac{30000}{R} \frac{\text{ms}}{\text{half-rotation}}$$

DB Storage: HD vs SSD vs SSHD

- Disk Latency for Hard Disks

Rotational speed [rpm]	Average latency [ms]
15,000	2
10,000	3
7,200	4.16
5,400	5.55
4,800	6.25

Comp283-Lecture 3

DB Storage: HD vs SSD vs SSHD

- Random Access Times for SSD drives are typically under 0.1 ms.
  - Data can be retrieved directly from various locations of the flash memory, access time is usually not a big performance bottleneck.
- Random Access Times for HDD drives ranges from 2.9 (high end server drive) to 12 ms (laptop HDD)
  - Need to move the heads and wait for the data to rotate under the read/write head.
- SSD much faster but have other issues e.g.:
  - Usually much slower write than read speeds (unless ££!!)
  - Individual blocks wear out after few writes - drive controller has to manage use of all blocks to even out the wear.

Comp283-Lecture 3

DB Storage: Q. What is this?

Comp283-Lecture 3

DB Storage: Q. What is this?

Comp283-Lecture 3

DB Storage: HD vs SSD vs SSHD

- “Traditional” Hard Disks are getting bigger and cheaper
- But how reliable are they?
  - Under warranty disk replacements reported between 0.5% - 13.5% !!
  - All disks will eventually fail
  - Even SSD!
- If your data is important it should be backed up.
  - Can be stored on a redundant storage system that will survive a drive failure.

http://www.enterprisestorageforum.com/storage-hardware/ssd-vs.-hdd-performance-and-reliability-1.html

Comp283-Lecture 3

DB Storage: HD vs SSD vs SSHD

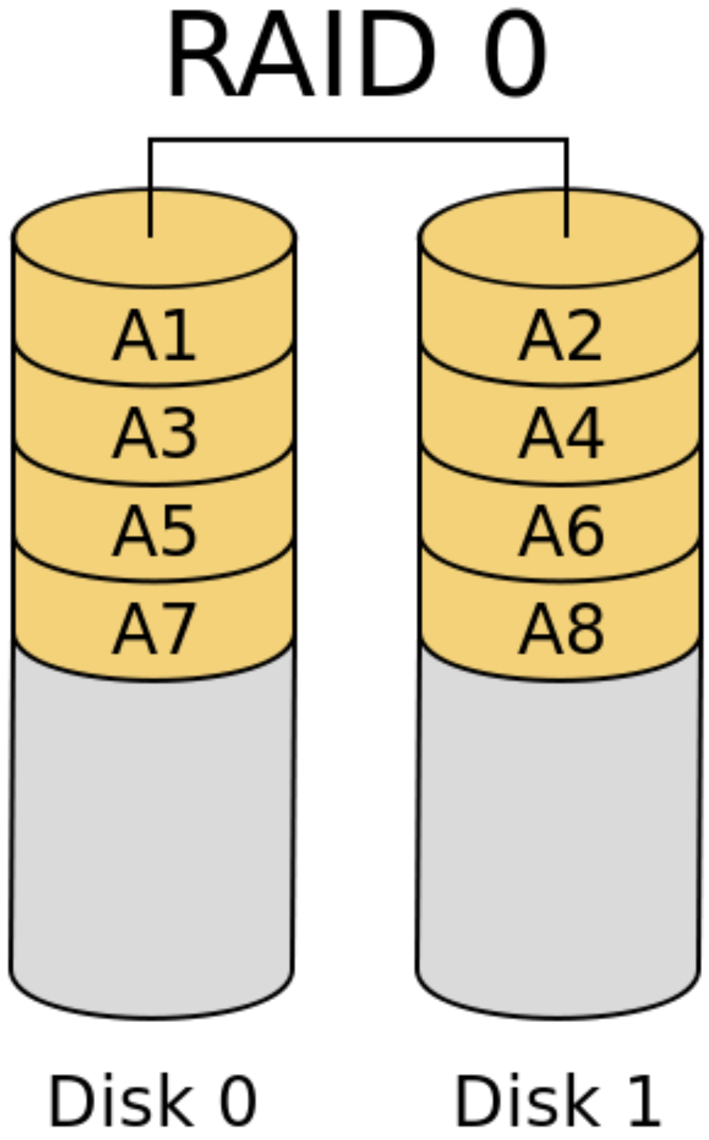
- Consumer magnetic disk error rate is 10^14 bits or an error every 12.5TB.
- Enterprise magnetic disk error rate is 10^15 bits or an error every 125TB.
- Consumer SSD error rates are 10^16 bits or an error every 1.25PB.
- Enterprise SSD error rates are 10^17 bits or an error every 12.5PB.
- Hardened SSD error rates are 10^18 bits or an error every 125PB.

Putting this into rather brutal context, consider the data sheet for the 8TB Archive Drive from Seagate. This has an error rate of 10^14 bits. That is one URE every 12.5TB. That means Seagate will not guarantee that you can fully read the entire drive twice before encountering a URE.

Comp283-Lecture 3

DB Storage: RAID Storage

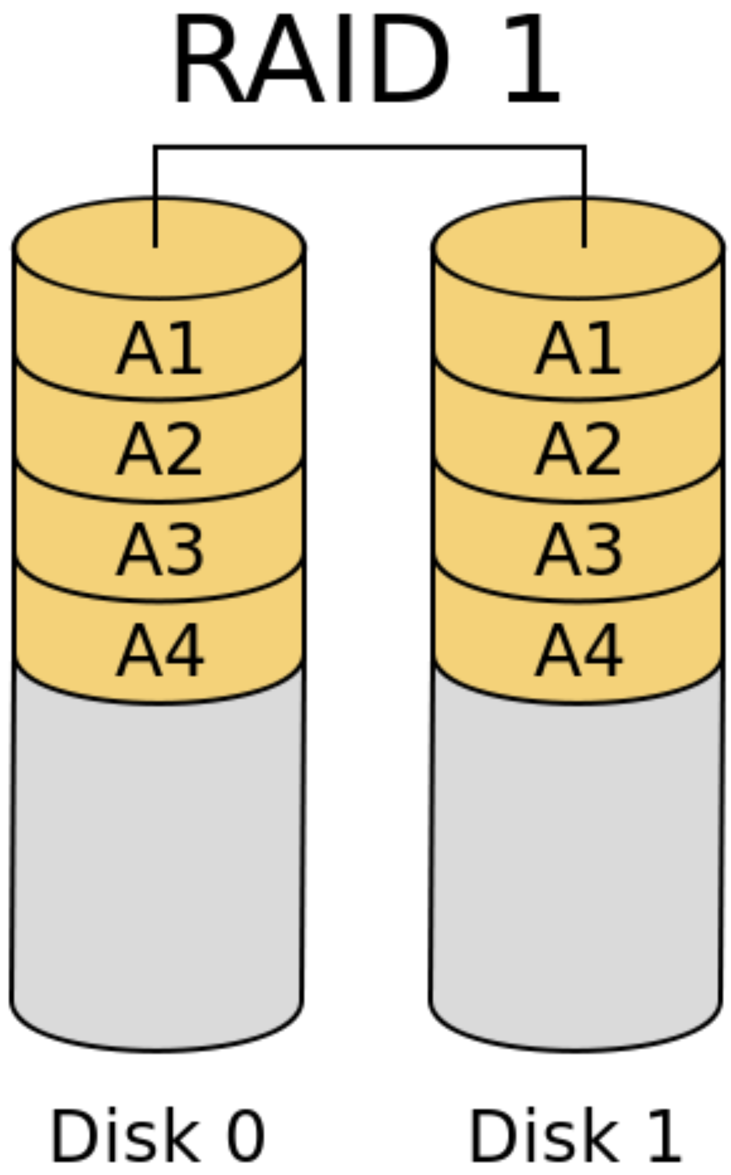
- RAID 0 (aka a stripe set or striped volume)
- Splits ("stripes") data evenly across two or more disks, without parity information, redundancy, or fault tolerance.
- The failure of one drive will cause the entire array to fail and result in total data loss.
- Why use RAID0?



Comp283-Lecture 3

DB Storage: RAID Storage

- RAID 1 consists of an exact copy (or mirror) of a set of data on two or more disks
- a classic RAID 1 mirrored pair contains two disks.
- This configuration offers no parity, striping, or spanning of disk space across multiple disks, since the data is mirrored on all disks belonging to the array
- The array can only be as big as the smallest member disk.
- Why do this?

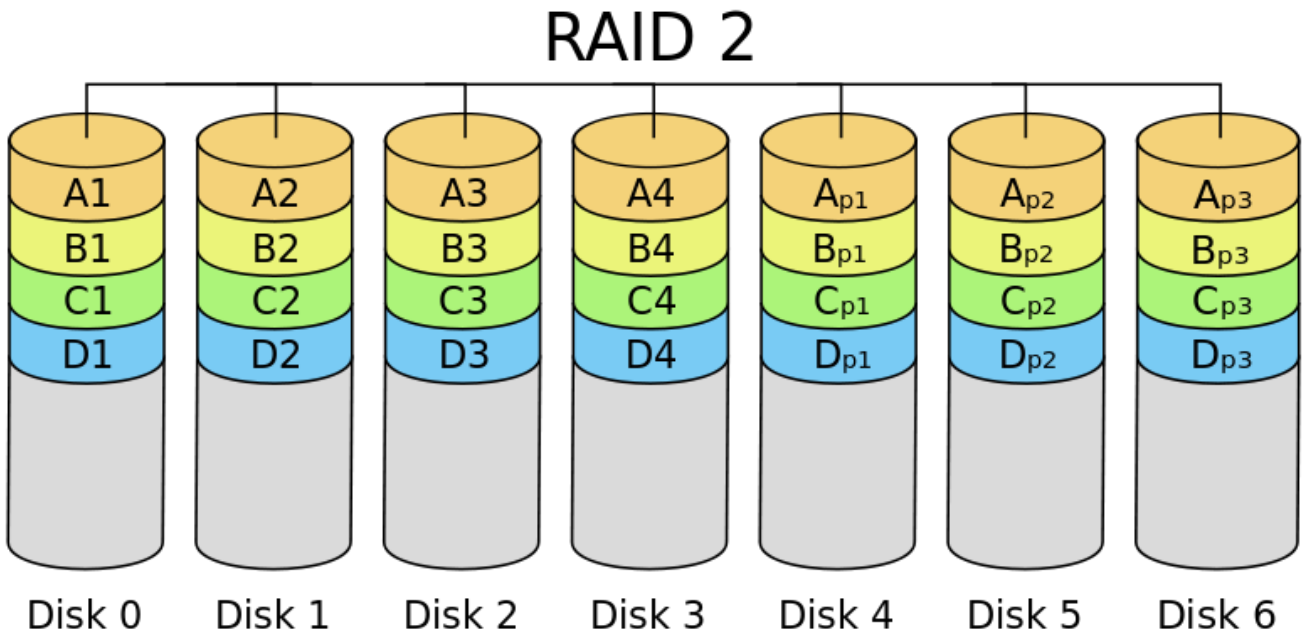


Comp283-Lecture 3

DB Storage: RAID Storage

- RAID 2 stripes data at the bit (rather than block) level, and uses a Hamming code for error correction.

- disks are synchronized by the controller to spin at the same angular orientation (they reach index at the same time).

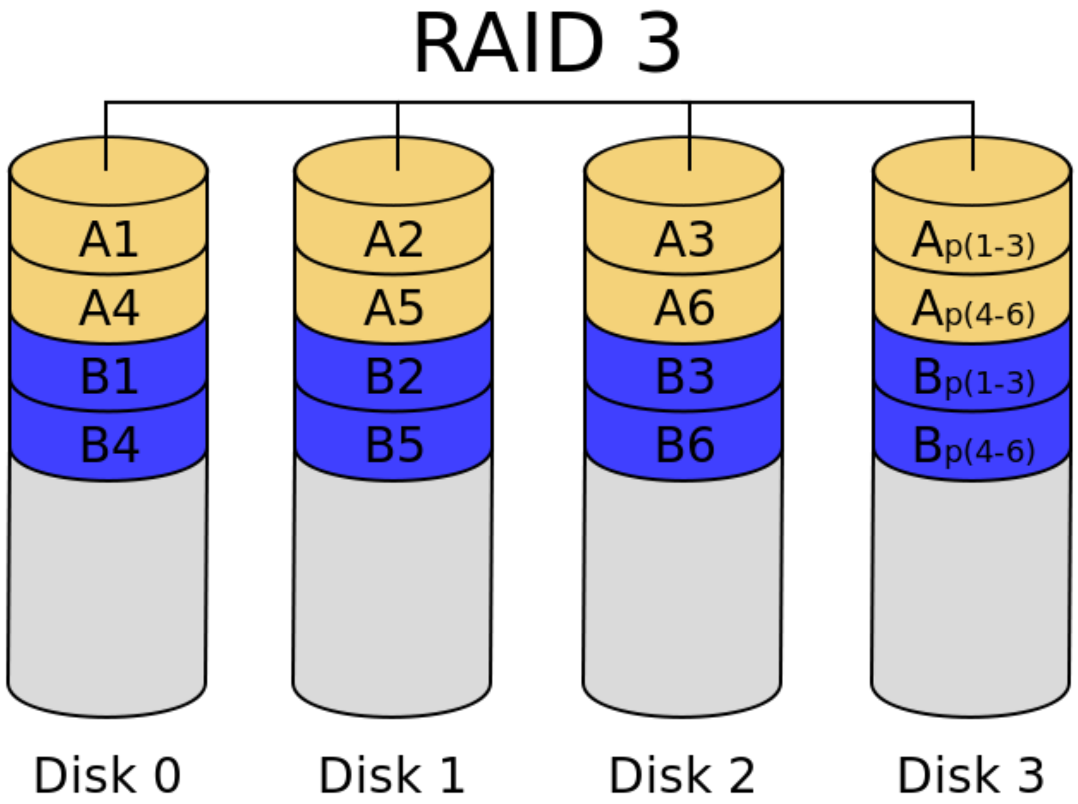


- Extremely high data transfer rates are possible.
- The array can only be as big as the smallest member disk.

Comp283-Lecture 3

DB Storage: RAID Storage

- RAID 3 consists of byte-level striping with a dedicated parity disk.



- Generally cannot service multiple requests simultaneously - happens because any single block of data will, by definition, be spread across all members of the set and will reside in the same location.

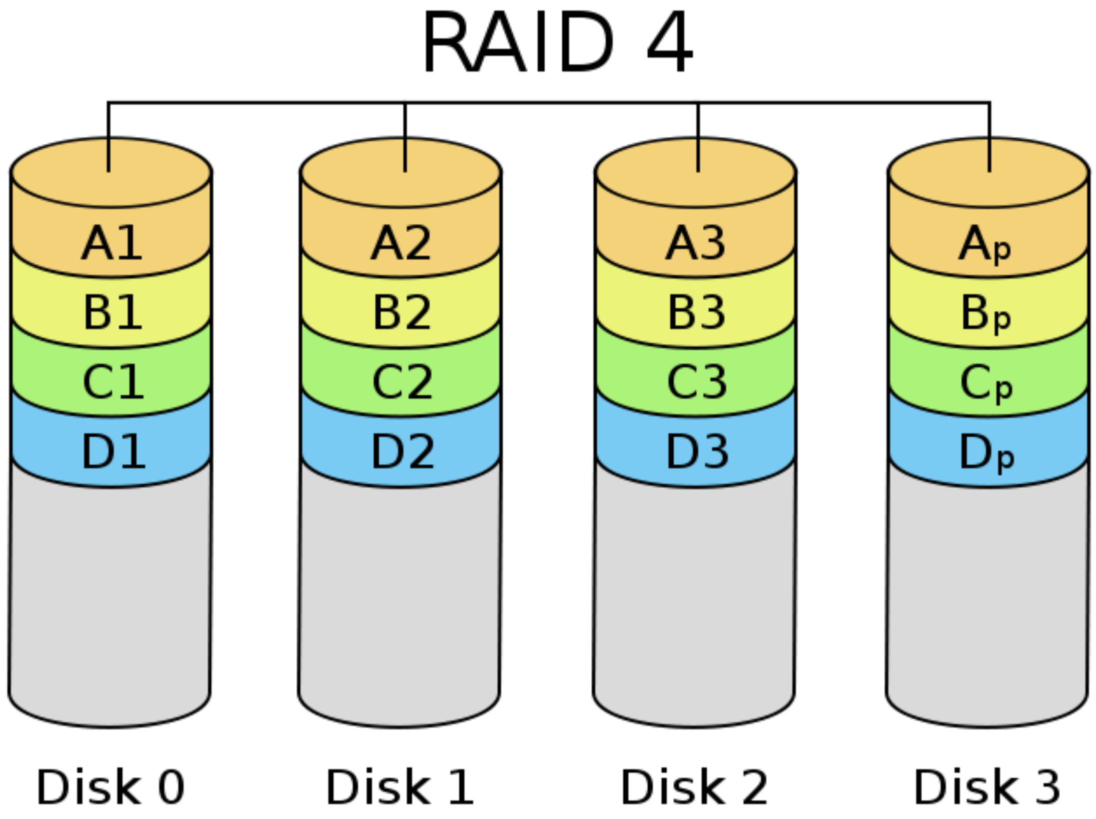
- Extremely high data transfer rates for R/W are possible.
- Suitable for e.g. uncompressed video editing.

Comp283-Lecture 3

DB Storage: RAID Storage

- RAID 4 consists of block-level striping with a dedicated parity disk.

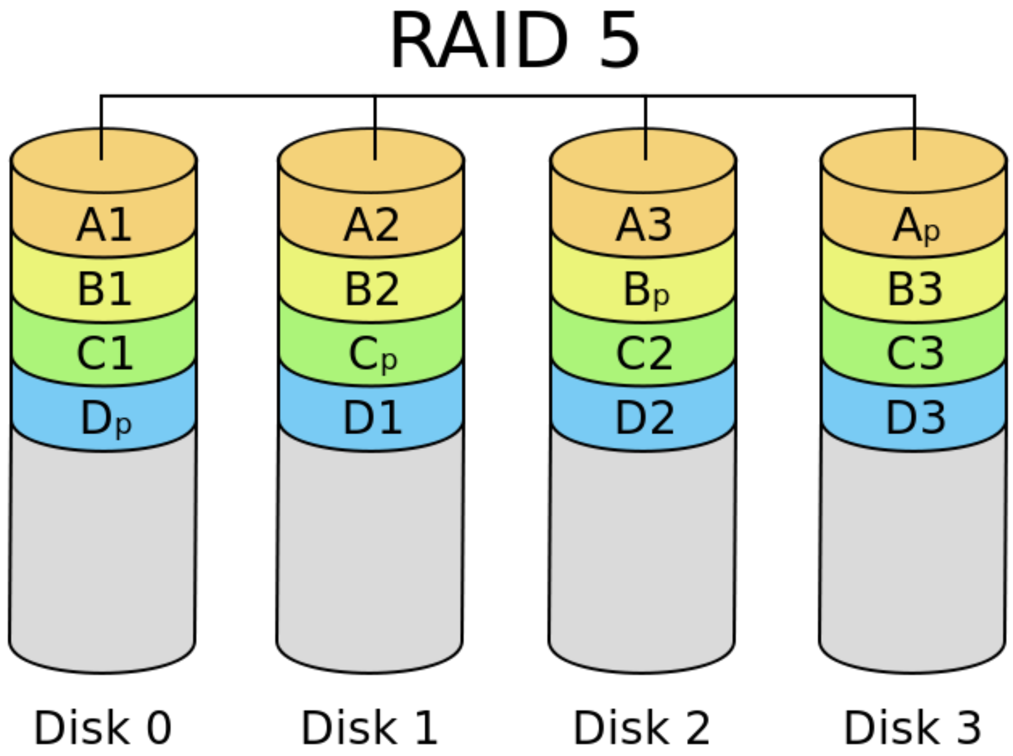
- Provides good performance of random reads, while the performance of random writes is low due to the need to write all parity data to a single disk.



DB Storage: RAID Storage

- RAID 5 consists of block-level striping with distributed parity.
- Unlike in RAID 4, parity information is distributed among the drives.

- Requires that all drives but one be present to operate.
- Upon failure of a single drive, subsequent reads can be calculated from the distributed parity such that no data is lost.

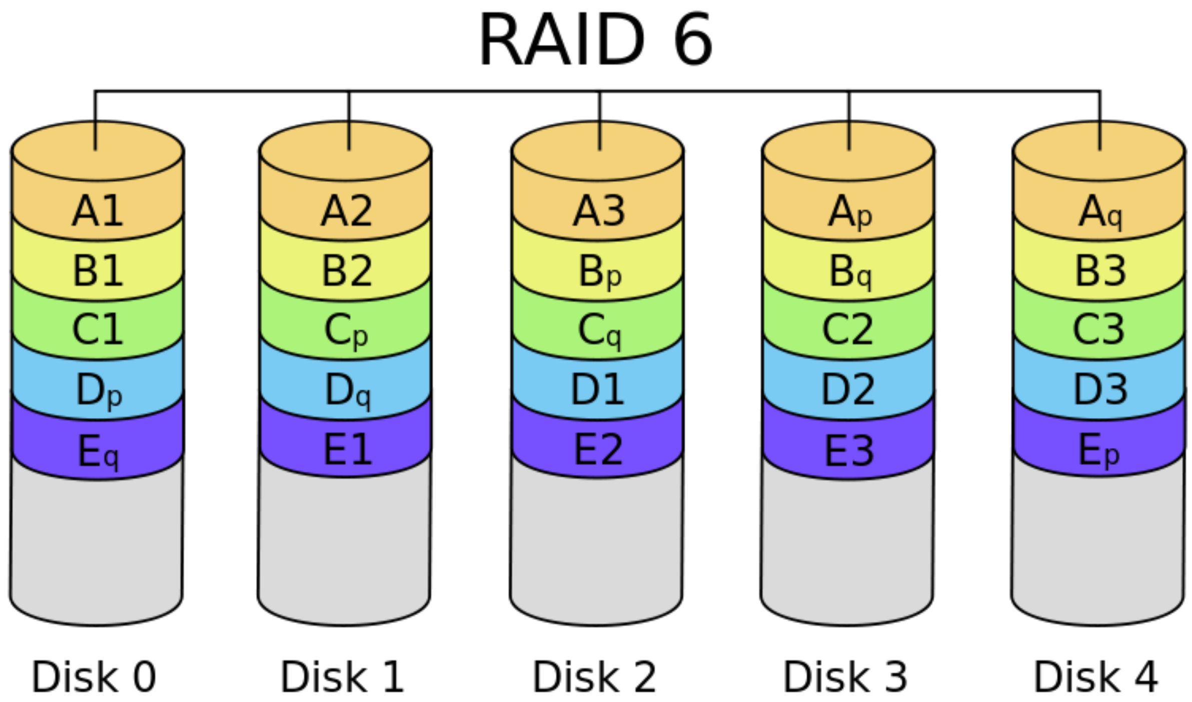


RAID 5 requires at least three disks.

Comp283-Lecture 3

DB Storage: RAID Storage

- RAID 6 extends RAID 5 by adding another parity block i.e. block-level striping with two parity blocks distributed across all member disks.
- Several methods have been used to implement RAID 6 dual parity.
- RAID 6 = RAID 5 + 1 extra disk

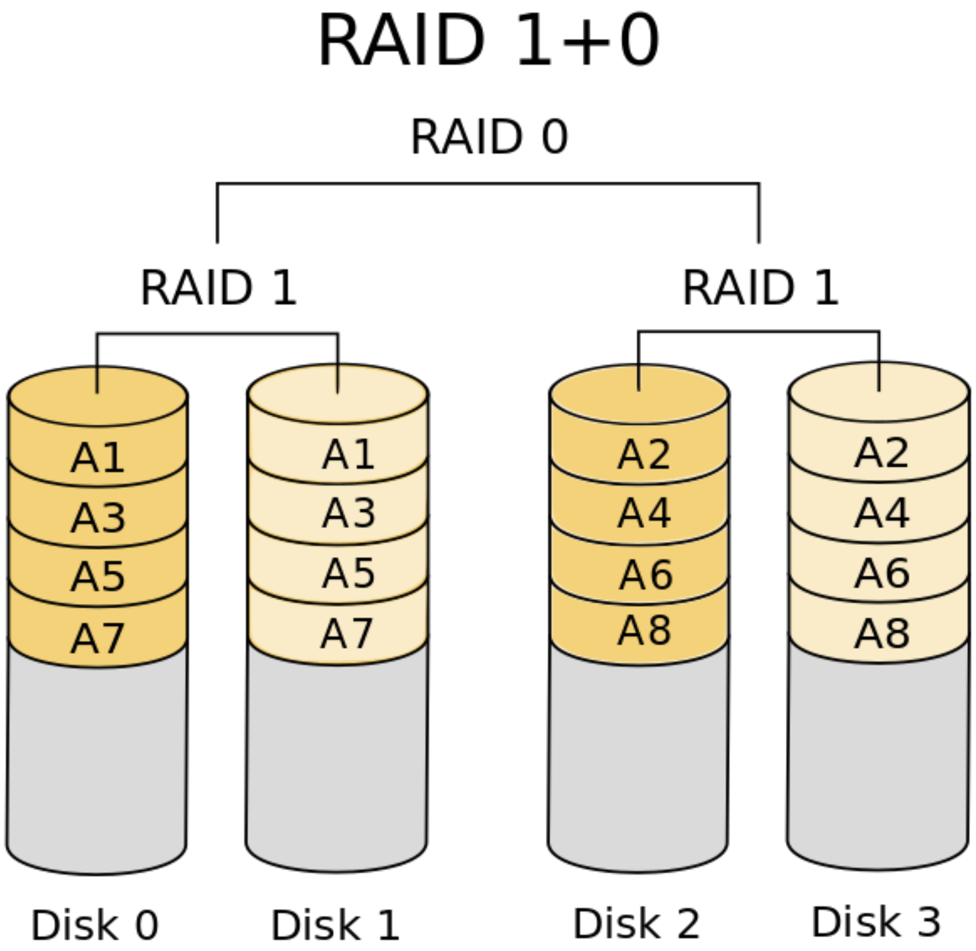


- RAID 6 does not have a performance penalty for read operations, but does have performance penalty on write operations because of overhead associated with parity calculations.

Comp283-Lecture 3

DB Storage: RAID Storage

- RAID 1+ 0 aka RAID 10 combines RAID 0 (striping) with RAID 1 (mirroring)



- Provides better throughput and latency than all other RAID levels except RAID 0.
- Preferable RAID level for I/O-intensive applications such as database, email, and web servers, as well as for any other use requiring high disk performance.

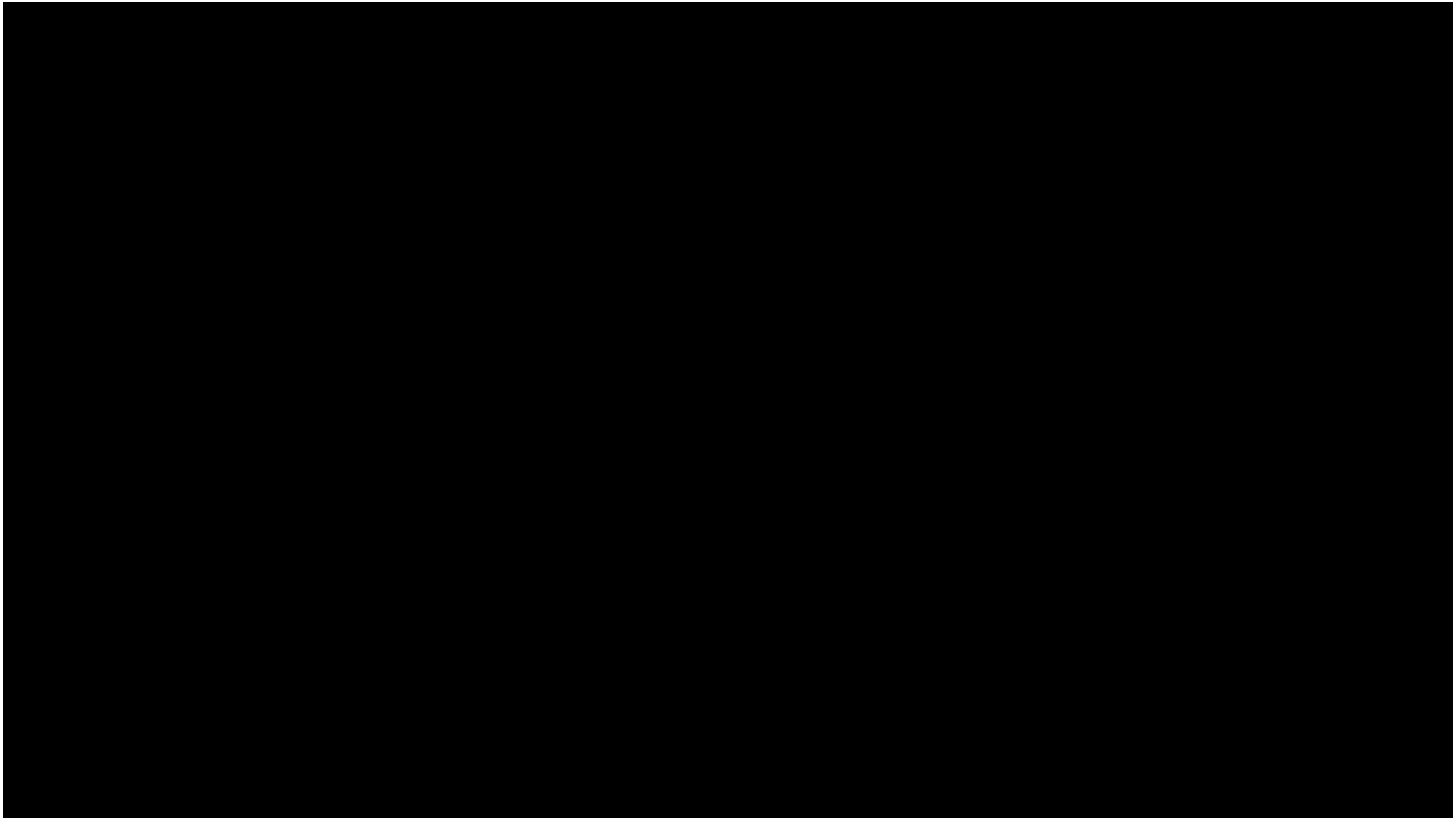
Requires a minimum of 4 drives.

Comp283-Lecture 3

DB Storage: HD vs SSD vs SSHD

- Choice of Drives for backup system / RAID depends on nature of the service requirements
  - Data Throughput / Bandwidth
  - Read only or Read / Write?
  - Cost
- Amount of storage not so much of an issue these days in most cases.

Comp283-Lecture 3 Drobo - BeyondRAID



COMP283-Lecture 3

Conclusion

- More factors to be considered for DB design
- Storage Considerations and disk types
- Identified Storage planning tasks.