Review of Last Lecture

- How to improve database access via efficient data structure
  - Unordered file
  - Ordered file
  - Hashing
  - B+ Tree
  - Indexing

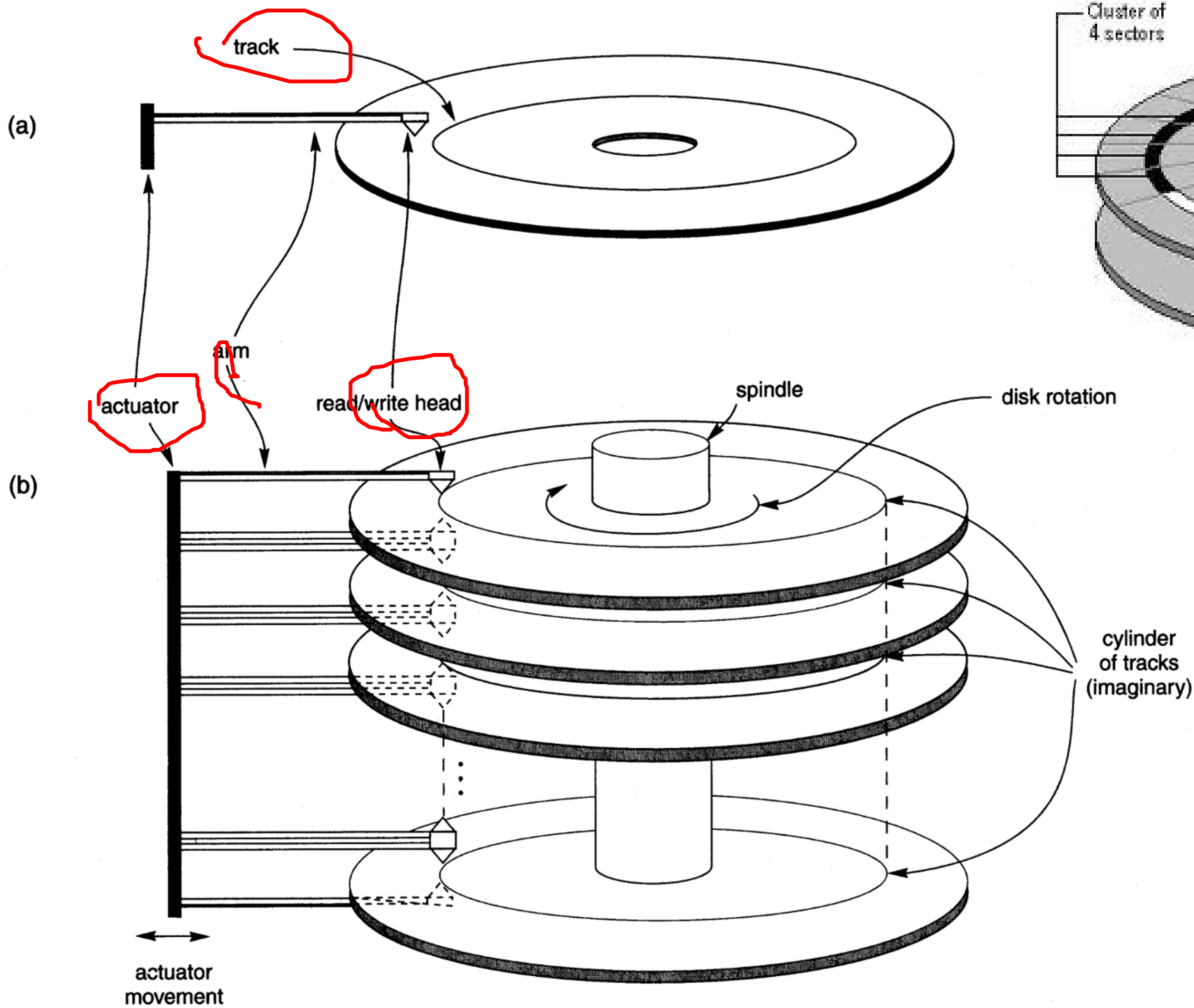
Today's Plan

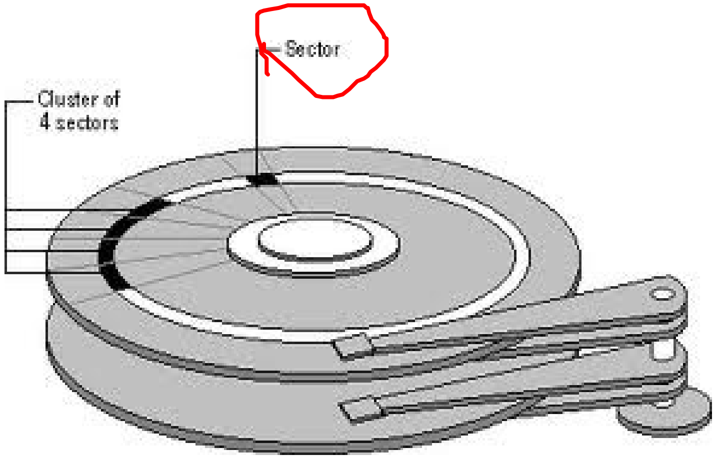
- RAID technologies
  - Striping
  - Mirroring
  - Parity
- RAID 0, 1, 5
  - Access time
  - Space efficiency
  - Fault tolerance

How to access data faster?

1. Come up with a better way to answer the query
   - Query optimization, already discussed in Chapter 03
2. Come up with a better way to logically organize the data
   - Data structure of database and indexing
3. Come up with a better way to physically store the data on hard drive
   - Storage technology and RAID

How Disk is Accessed





https://blogs.umass.edu/Techbytes/2017/04/04/hard-drives-how-do-they-work/

Cost of Accessing Disk

- The time to Reading/Writing a disk is referred to as Input/output (I/O) cost
- Broken down into 3 steps
  1. Seek time: time it takes to move the arm to the proper track
     - 4 – 10 milliseconds (ms), often specified by the manufacturer.
  2. Rotation time: time it takes to rotate the disk to put the right sector underneath the read/write head.
     - Typical RPM is 7200, translated into 4 ms.
  3. Transfer time: time to move data between memory and disk
     - Defined by its interface (S-ATA, USB etc.). Typically couple hundred Mbps to couple Gbps
     - Compared to seek time and rotation time, transfer time is negligible.

A Mid-end HD (FYI)

WD Black 4TB Performance Desktop Hard Disk Drive - 7200 RPM SATA 6 Gb/s 64MB Cache 3.5 Inch - WD4001FAEX

by Western Digital

1,553 customer reviews | 78 answered questions

Available from these sellers.

Capacity: **4 TB**

1 TB $79.00	2 TB $148.89	3 TB from 1 seller
4 TB **used from $149.99**		

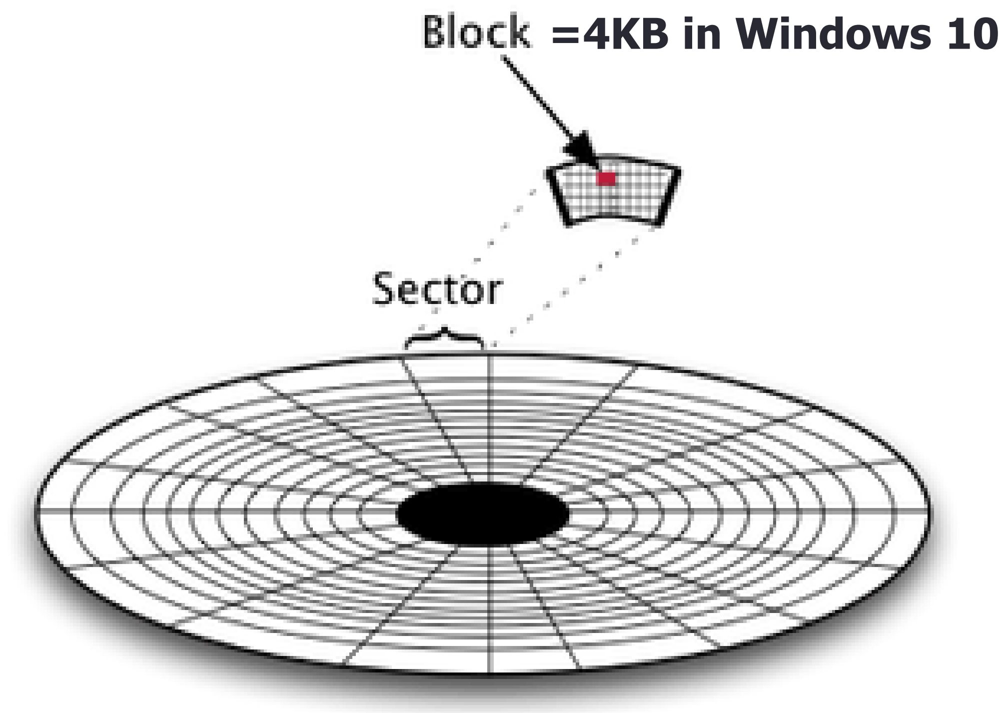
- WD Black high performance 3.5-Inch SATA hard drives combine 7200 RPM spin speed, 64 MB cache, and SATA 6 Gb/s interface for the ultimate in power computing.
- High performance electronics architecture features dual processors and bigger, faster caches for maximum read and write speeds.
- StableTrac The motor shaft is secured at both ends to reduce system-induced vibration and stabilize platters for accurate tracking, during read and write operations.
- 4 TB capacity holds up to 800,000 digital photos, 1,000,000 MP3 files, and 300 hours of HD video.
- 5 year limited warranty.

Optimize Disk Access

- Data is transferred between disk and main memory in blocks
  - Block: a contiguous sequence of sectors from a single track
  - Block sizes range from 512 bytes to several kilobytes
    - Smaller blocks: more transfers from disk
    - Larger blocks: more space wasted due to partially filled blocks
    - Typical block sizes today range from 4 to 16 kilobytes
- Minimize the cost of disk I/O
  - Read/use consecutive blocks on disk: eliminate seek/rotational latency for all blocks beyond the first one
  - Read/use all of the records within one block (get multiple (useful) records from one physical I/O)

Block: Example

- I want to store "0123456789" on HD (10 bytes)

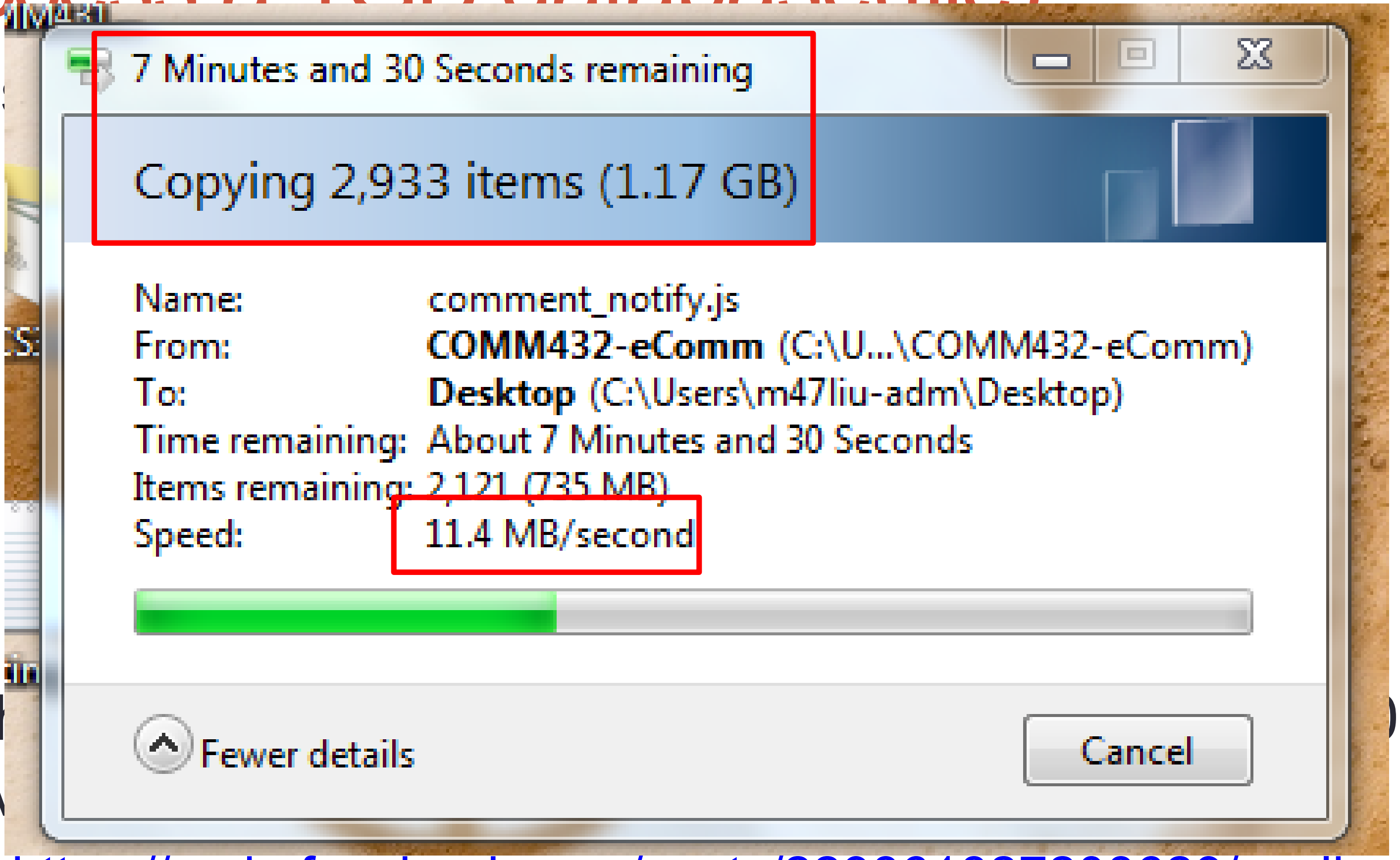


- I want to retrieve my SID "0123456789" from the HD

Image Source: https://jmetz.com/2018/01/storage-hdd-and-ssd-partitioning-and-i-o-operations/

Example: How much time does it take to access a 1GB database(file)?

- U…



- Th… 0TB ev…
  - https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/

Parallelizing Disk Access using RAID Technology

- Secondary storage technology must take steps to keep up in performance and reliability with processor technology.

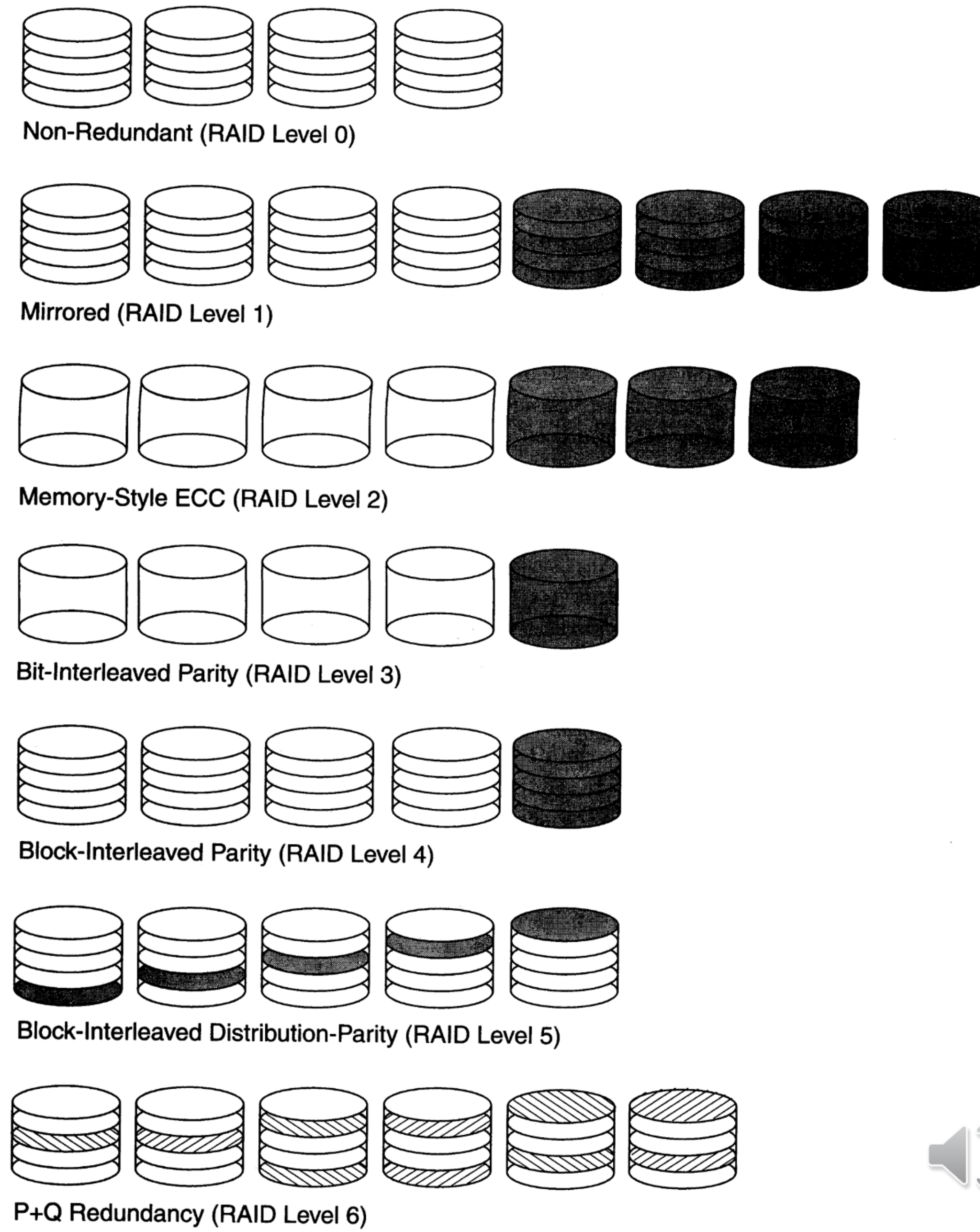
- A major advance in secondary storage technology is represented by the development of RAID, which originally stood for **Redundant Arrays of Inexpensive Disks**.

- The main goal of RAID is to even out the widely different rates of performance improvement of disks against those in memory and microprocessors. And to provide improved reliability as well.

Different Level of RAID

- Different RAID organizations were defined based on different combinations of the two factors of granularity of data interleaving (striping) and pattern used to compute redundant information.
  - Raid level 0 has no redundant data and hence has the best write performance.
  - Raid level 1 uses mirrored disks.
  - Raid level 2 uses memory-style redundancy by using Hamming codes, which contain parity bits for distinct overlapping subsets of components. Level 2 includes both error detection and correction.
  - Raid level 3 uses a single parity disk relying on the disk controller to figure out which disk has failed.
  - Raid Levels 4 and 5 use block-level data striping, with level 5 distributing data and parity information across all disks.
  - Raid level 6 applies the so-called P + Q redundancy scheme using Reed-Soloman codes to protect against up to two disk failures by using just two redundant disks.
- We only focus on 0, 1 and 5 as they are the most common

RAID: Diagram



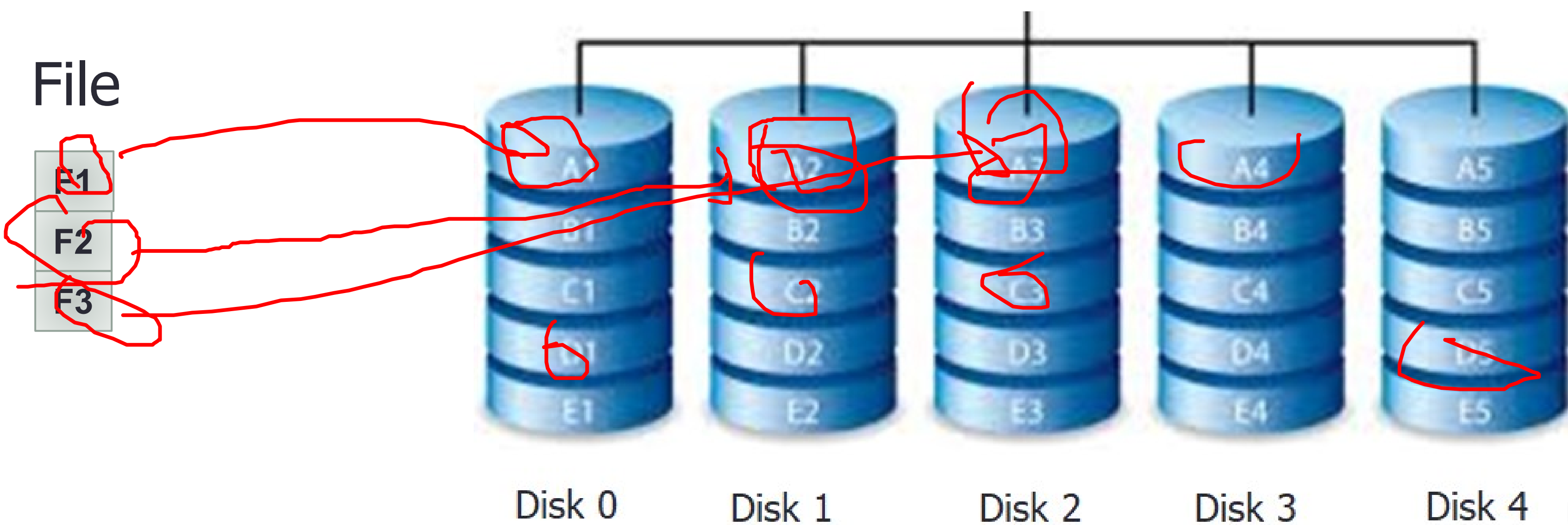
Key RAID Concepts

- Striping
- Mirroring
- Parity
- Performance matrix
  - Read/write speed
  - Space efficiency
  - Fault tolerance

- Good information source: https://en.wikipedia.org/wiki/Standard_RAID_levels

Data Striping

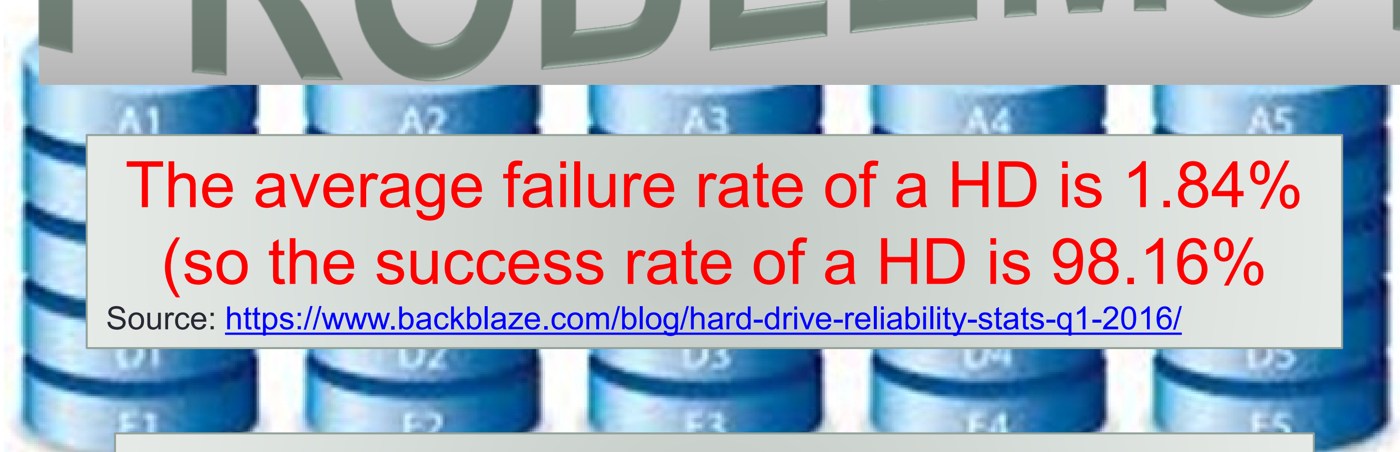
- Data striping intentionally distributes data blocks of the same file over multiple disks to speed up access performance.
    - It utilizes parallelism to improve disk performance as each disk can move and rotate independently. So each disk can access a data block simultaneously.



RAID 0

- Only striping is used
- Performance

PROBLEMS?



The average failure rate of a HD is 1.84%
(so the success rate of a HD is 98.16%

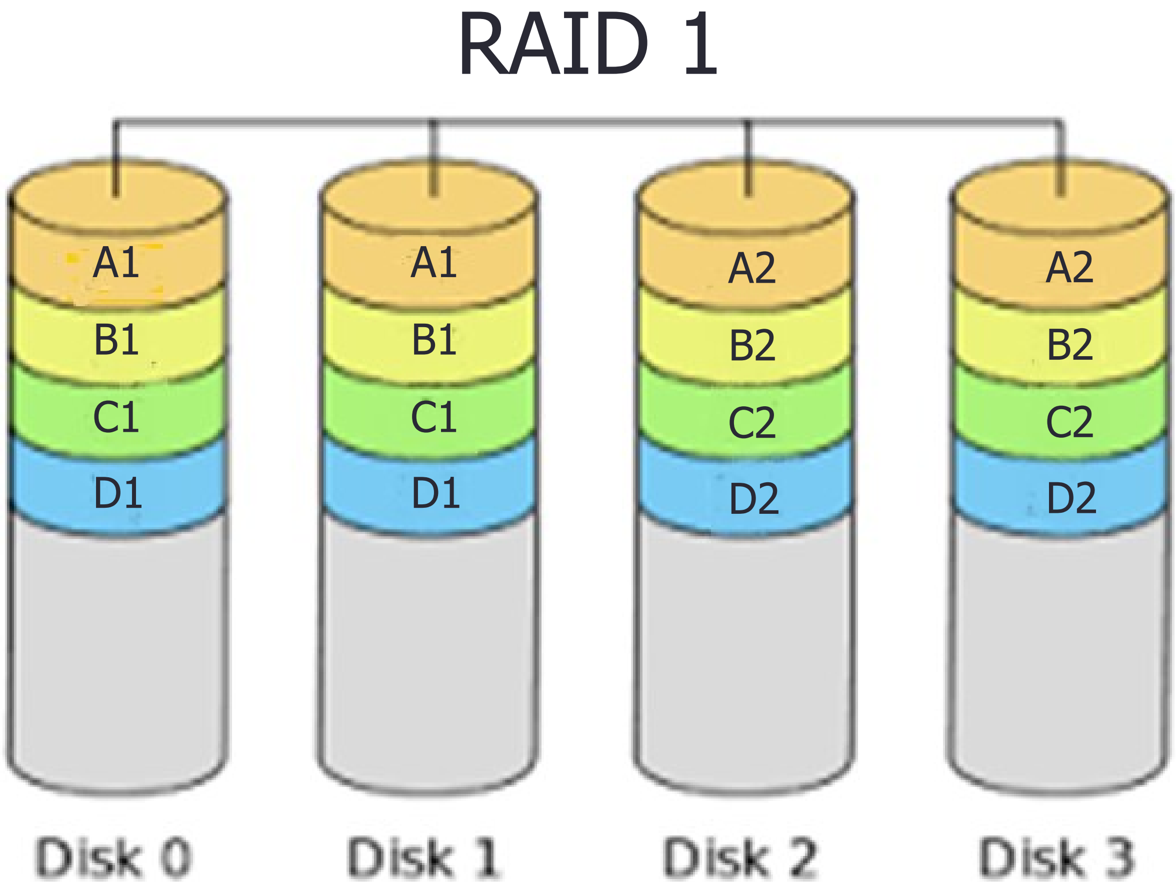
Source: https://www.backblaze.com/blog/hard-drive-reliability-stats-q1-2016/

Dis 4

- With 5 drives, speed improve 5 times.
- But chance of data loss due to HD failure?

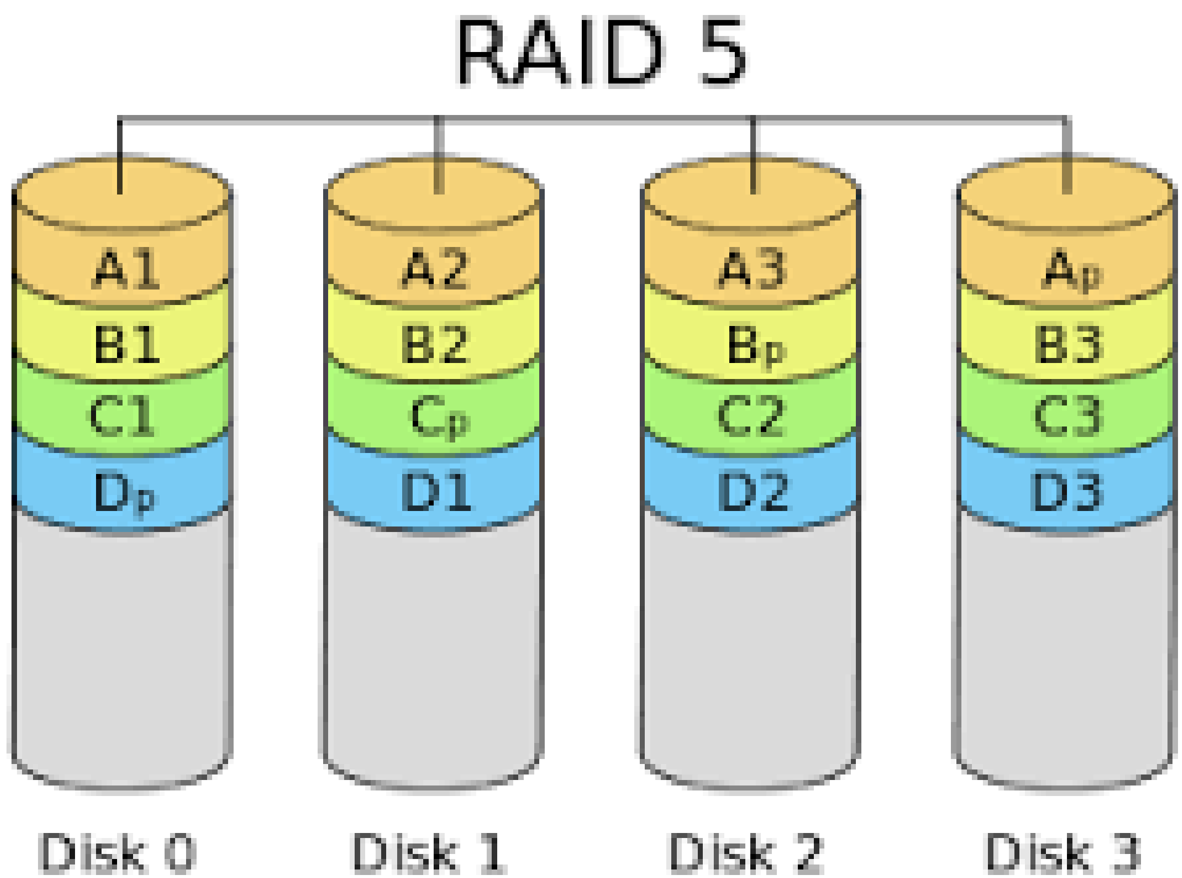
RAID 1

- Only use mirroring
- Performance



RAID 5

- Striping and parity are used



Parity

- A parity bit, or check bit, is a bit added to a string of binary code … 1-bits in the string … d.

111X101

7 bits of data	(count of 1-bits)	… bits including parity	
		even	odd
0000000 ?	0	0000000**0**	0000000**1**
1010001 ?	3	1010001**1**	1010001**0**
1101001	4	1101001**0**	1101001**1**
1111111	7	1111111**1**	1111111**0**

- Source: https://en.wikipedia.org/wiki/Parity_bit

Parity at work: Example

- 5 data blocks, each block contains 4 bit, even parity.

4 bits of data
0 0 0 0
0 1 0 1
0 1 0 1
1 0 1 0
1 1 0 1
[illegible]

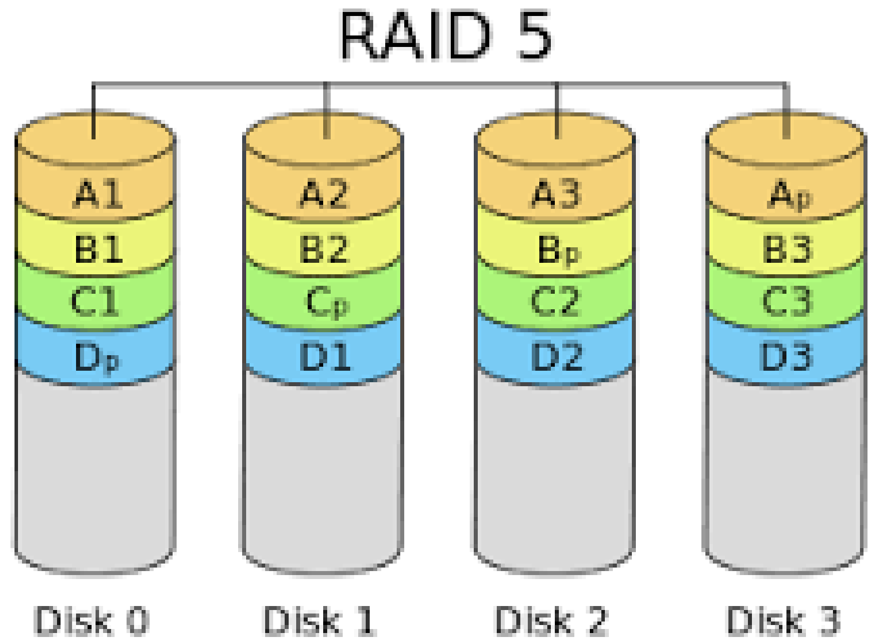
- One block corrupted

4 bits of data
A B C D
0 1 0 1
0 1 0 1
1 0 1 0
1 1 0 1
0 1 1 1

Striping + Parity

- 5 data blocks, each block contains 4 bit, even parity. Assume B2 is empty (0 0 0 0).

File A	
F1 0 0 0 0	A1
F2 0 1 0 1	A2
F3 0 1 0 1	A3
F4 1 0 1 0	B3
F5 1 1 0 1	B1



Even Parity for F1(stored in A1), F2 (stored in A2) and F3 (stored in A3)

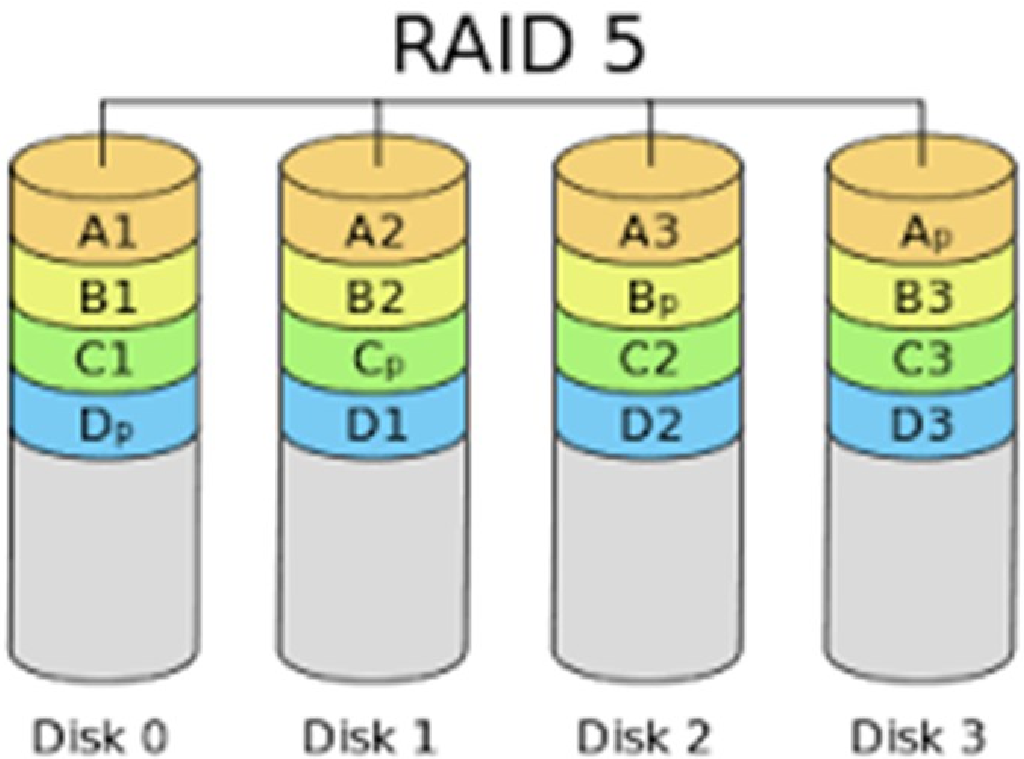
0 0 0 0	Ap

Even Parity for F4 (stored in B3), F5 (stored in B1) and B2 (0 0 0 0)

0 1 1 1	Bp

RAID 5

- Performance
- Fault Tolerance



Summary of Today's Lecture

- RAID technologies
  - Striping
  - Mirroring
  - Parity
- RAID 0, 1, 5
  - Access time
  - Space efficiency
  - Fault tolerance