THE UNIVERSITY OF
MELBOURNE

COMP90050
Advanced Database Systems

Winter Semester, 2023

Lecturer: Farhana Choudhury (PhD)

Live lecture – week 1

Logistic Information

Winter Semester, 2023 - Dual delivery

Lectures Each Week: A combination of pre-recorded and live sessions

- Monday lectures will be pre-recorded and uploaded in Canvas
- Tuesday 11:00 - 2:00 will be conducted on-campus + broadcasted live via Zoom + will be available in lecture capture

Logistic Information Contd

Tutorials: one 3-hour tutorial per week

- **Starts from week 1**
- Some tutorials are conducted on-campus, and some are online as zoom sessions, links are on Canvas now
- Every student has to enrol to a tutorial officially
- First tutorial will cover details on the group project expectations and some other topics from week 1’s lectures

Logistic Information Contd

References:

1. Transaction Processing, Jim Gray and Andreas Reuter, Morgan Kaufmann, 1992
2. Database system concepts, Abraham Silberschatz, Henry F. Korth, S. Sudarshan, 2011

We will not read this book cover to cover, slides are the main source of contents

Both books are available online from UniMelb library

Logistic Information Contd

All the key information are available on Canvas:

- You need to check it a few times a week for being up to date on discussions and announcements
- It has also other basic info such as contact info for us
- Find project specs and other assessments
- Where submissions are made
- Where you can find lecture capture and other materials

Assessments

- Online quizzes – 5 quizzes worth 2% each (2 in Week 2, 2 in Week 3, and 1 in Week 4)

  -Will be open on Monday 11am, closes on Tuesday 11am

  -20 minutes strict time to complete

- Project – survey report and oral presentation on a database research topic (40%) as a **group of 4** students (presentation 15% + Report 25%)

- Final examination (50%)

Project

Step 1: Form a group of **4 students** by 2 July, 2023, 11:59pm

Step 2: Pick a topic of your interest from a list of candidate topics provided, by 2 July, 2023, 11:59pm

Canvas has details about the projects already: **Read** and **Start** with the steps above.

Presentation: During Week 4 (schedule will be uploaded)

Report submission due data: 24 July 2023, 11:59pm

Project

- All members of a group should contribute to the project. If there is significant difference among group members, we reserve the right to differentiate marks.
- Note that, there is an individual reflection component, where the members of a group will receive individual marks.

Who we are and Contact Info/Mode

Lecturer and subject coordinator

Dr. Farhana Choudhury

farhana.Choudhury@unimelb.edu.au

Head tutor

Tawfiq Islam (**tawfiqul.islam@unimelb.edu.au**)

Other tutors

Ahmad Asgharian Rezaei (a.asghariyan.rezayi@gmail.com)

Lakmal Muthugama tmuthugama@student.unimelb.edu.au

Daniel Gong <d.gong@unimelb.edu.au>

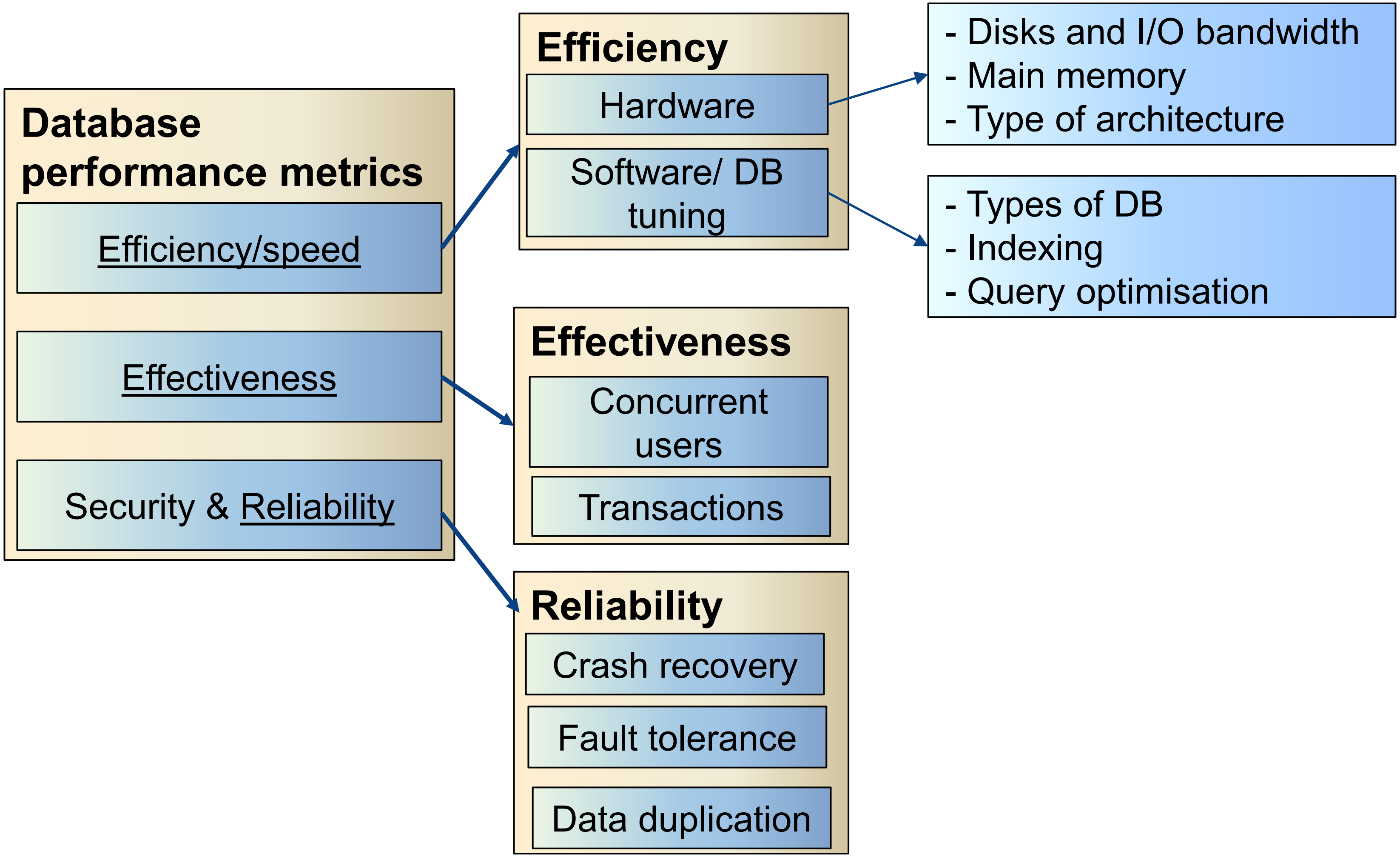
David Alexander Tedjopurnomo <davidtedjopurnomo@gmail.com>

BUT Ed DISCUSSION and ANNOUNCEMENTS first!

THE UNIVERSITY OF MELBOURNE

Discussion on the topics of the pre-recorded lecture and additional contents

Core Concepts of Database management system



Database performance metrics

- Efficiency/speed
- Effectiveness
- Security & Reliability

Efficiency

- Hardware
  - Disks and I/O bandwidth
  - Main memory
  - Type of architecture
- Software/ DB tuning
  - Types of DB
  - Indexing
  - Query optimisation

Effectiveness

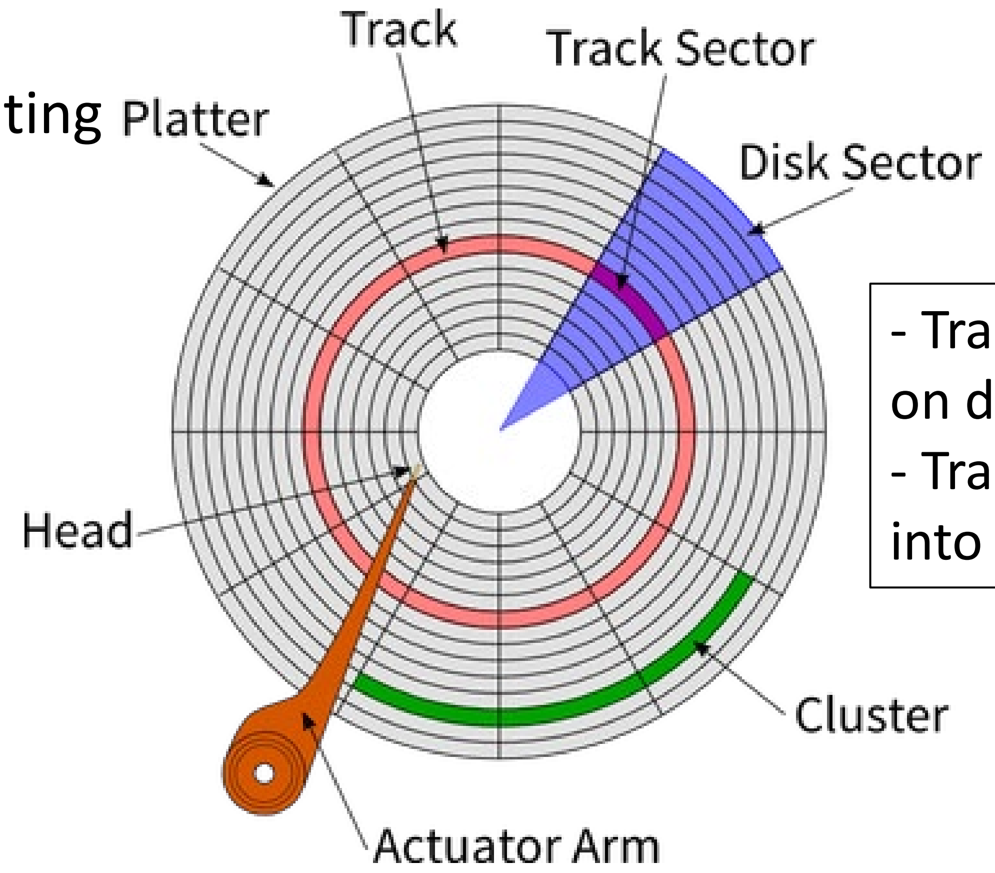
- Concurrent users
- Transactions

Reliability

- Crash recovery
- Fault tolerance
- Data duplication

Basic Hardware of a classical disk

Rapidly rotating



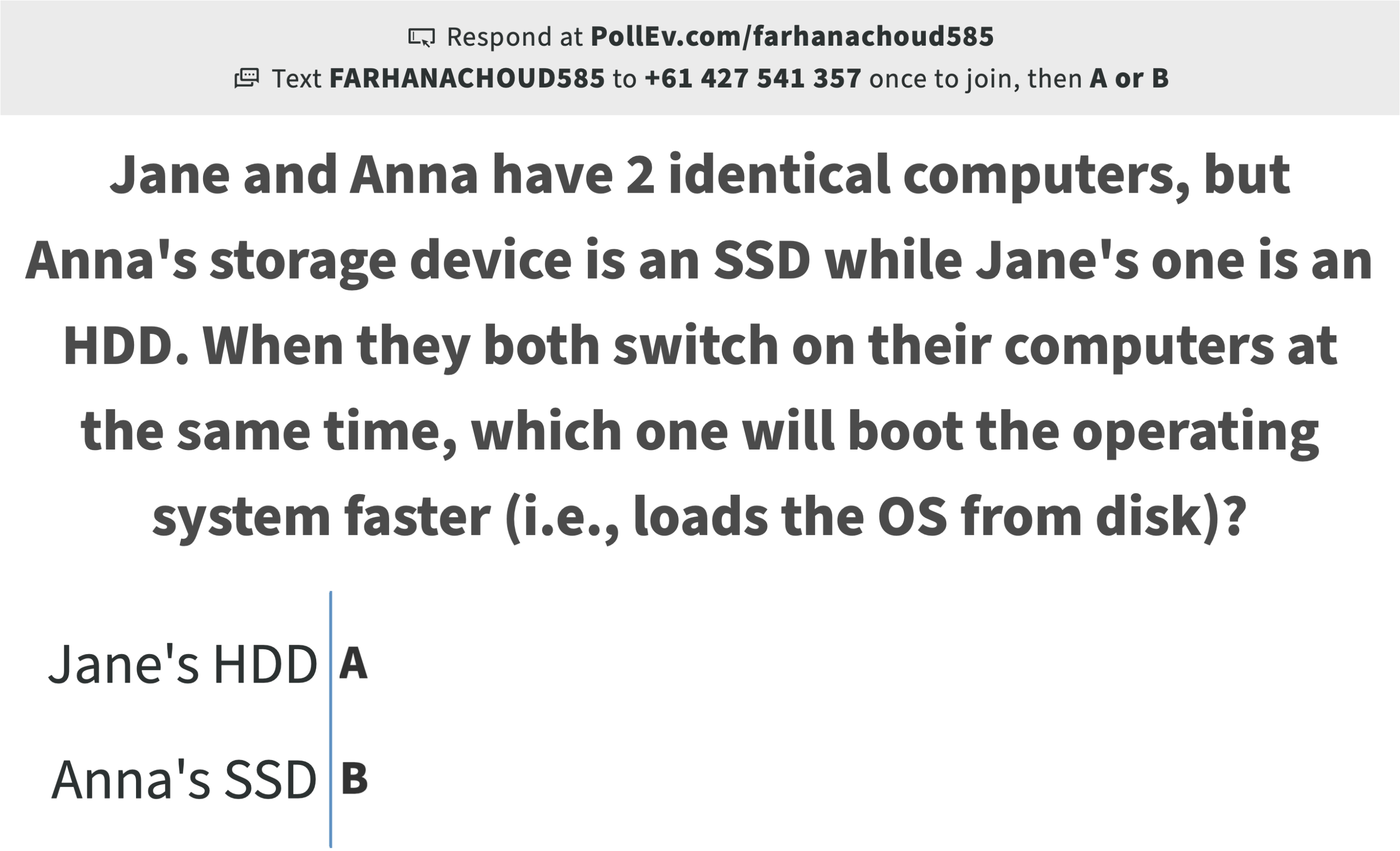
- Tracks: circular path on disk surface
- Tracks are subdivided into disk sectors

with magnetic head, which reads
and writes data to the platter surfaces

SSD (Solid-State Drive/Solid-State Disk)

- **No moving parts** like Hard Disk Drive (HDD)
- No seek/rotational latency
- No start-up times like HDD

Pollev.com/farhanachoud585



Respond at **PollEv.com/farhanachoud585**

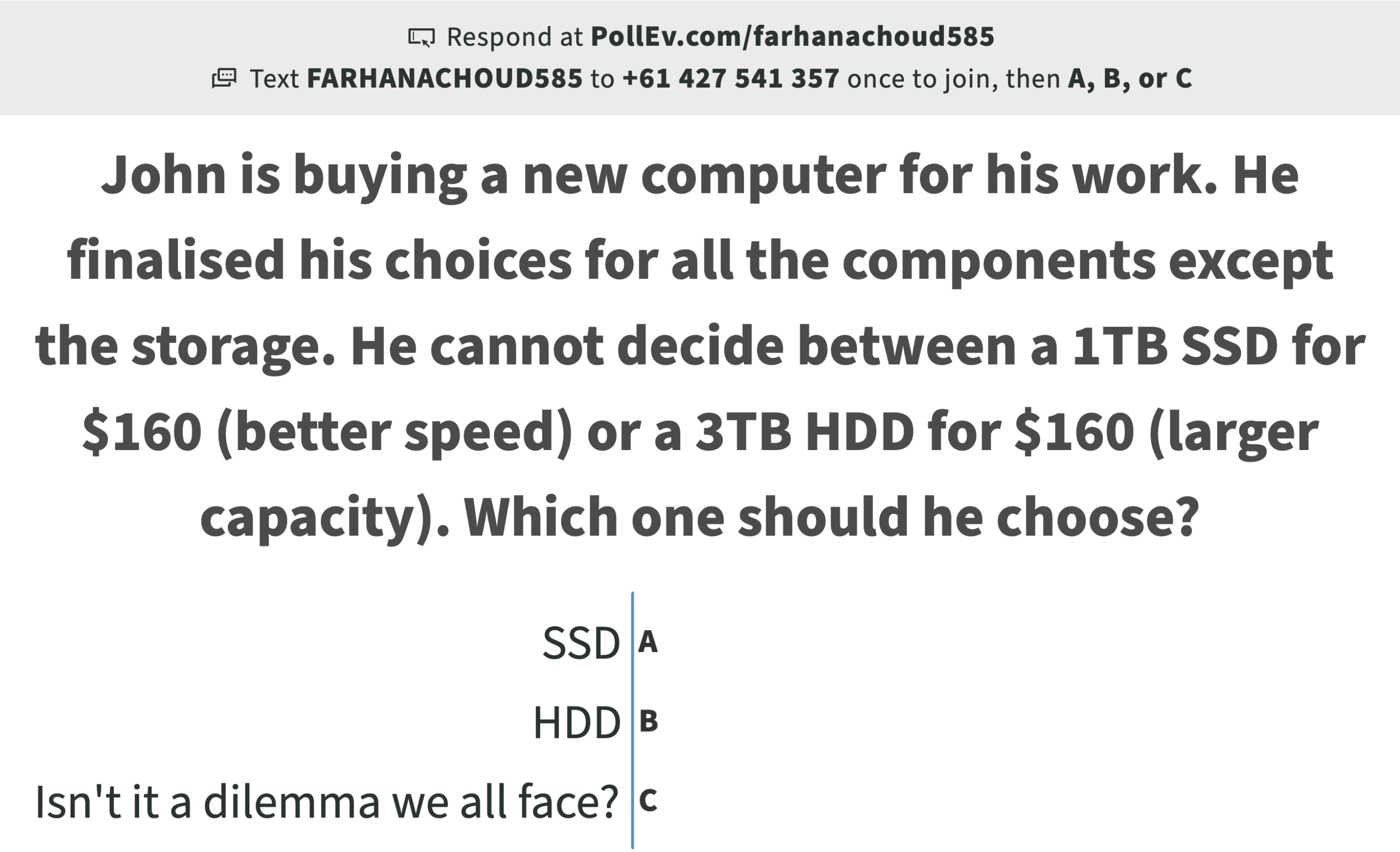
Text **FARHANACHOUD585** to **+61 427 541 357** once to join, then **A or B**

Jane and Anna have 2 identical computers, but Anna's storage device is an SSD while Jane's one is an HDD. When they both switch on their computers at the same time, which one will boot the operating system faster (i.e., loads the OS from disk)?

Jane's HDD **A**

Anna's SSD **B**

Pollev.com/farhanachoud585



Respond at **PollEv.com/farhanachoud585**

Text **FARHANACHOUD585** to **+61 427 541 357** once to join, then **A, B, or C**

John is buying a new computer for his work. He finalised his choices for all the components except the storage. He cannot decide between a 1TB SSD for $160 (better speed) or a 3TB HDD for $160 (larger capacity). Which one should he choose?

- SSD **A**
- HDD **B**
- Isn't it a dilemma we all face? **C**

What if some more contexts are given?

Disk access

Disk access time for HDD

$$Disk\ \ access\ \ time = seek\ \ time + rotational\ \ time + \frac{transferlength}{bandwidth}$$

What is the Disk access time for a transfer size of 8KB, when average seek time is 12 ms, rotation delay 4 ms, transfer rate 4MB/sec?

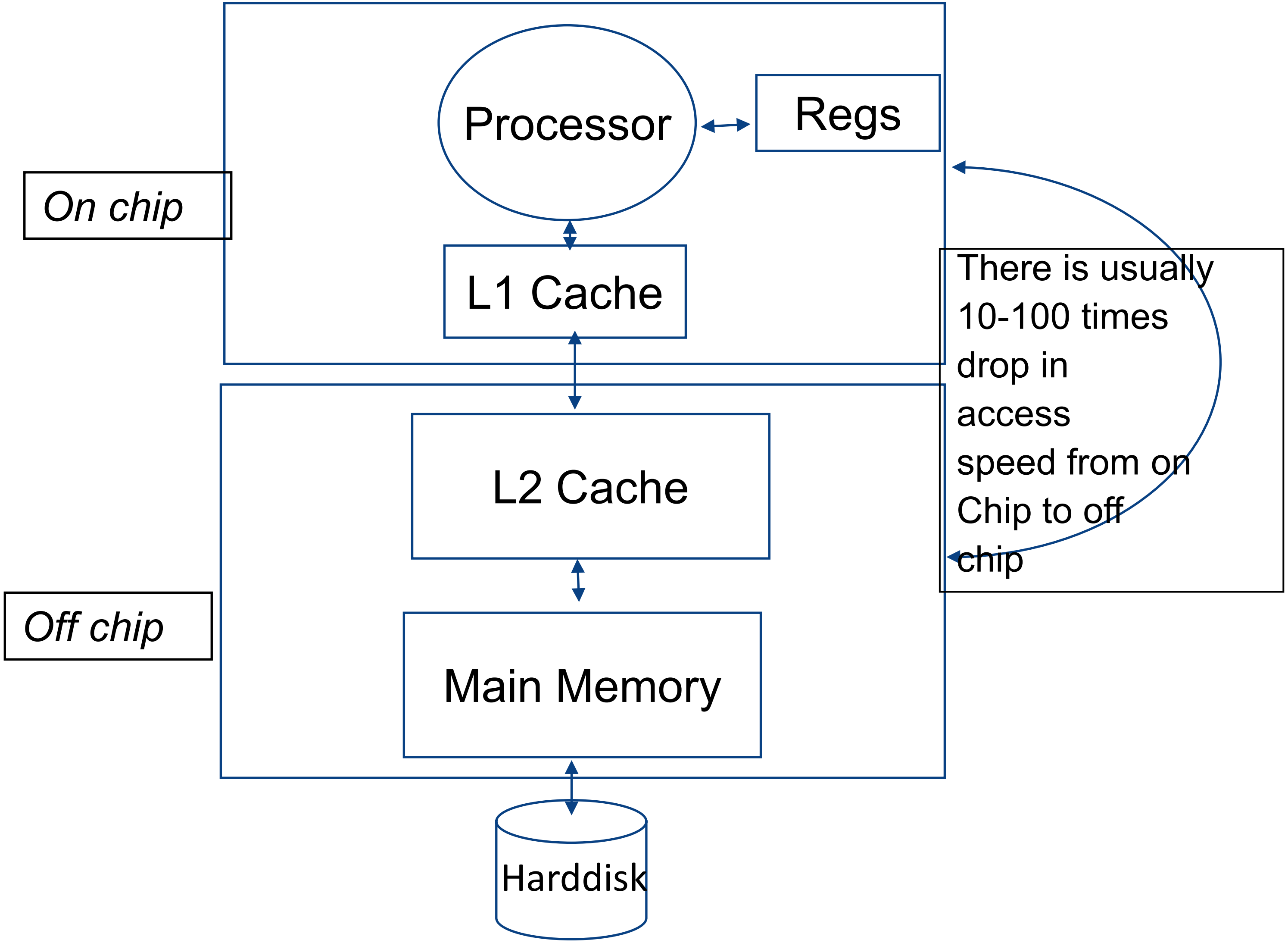
Solve related problems in tutorials

Disk access time for SSD

$$Disk\ access\ time = \frac{transferlength}{bandwidth}$$

So where do we store data: The Memory Hierarchy

Basic structure



There is usually 10-100 times drop in access speed from on Chip to off chip

Memory hierarchy

$$Hit\ ratio = \frac{references\ satisfied\ by\ cache}{total\ references}$$

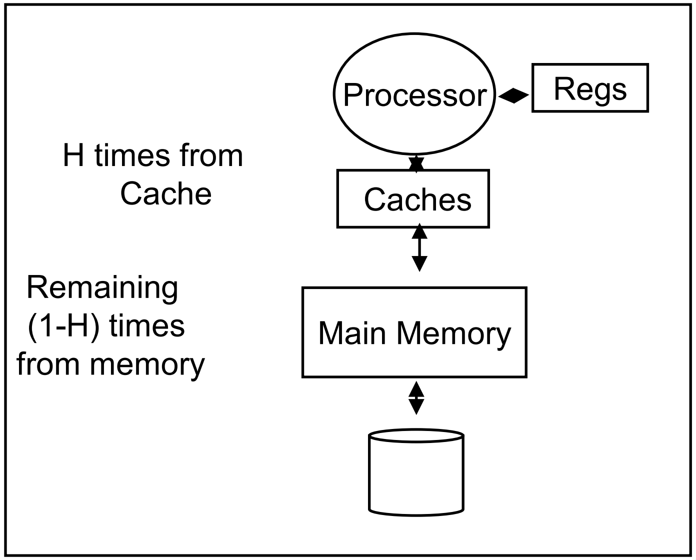
Effective memory access time,

EA = H*C+(1-H)*M

where H = hit ratio,

C = cache access time;

M = memory access time



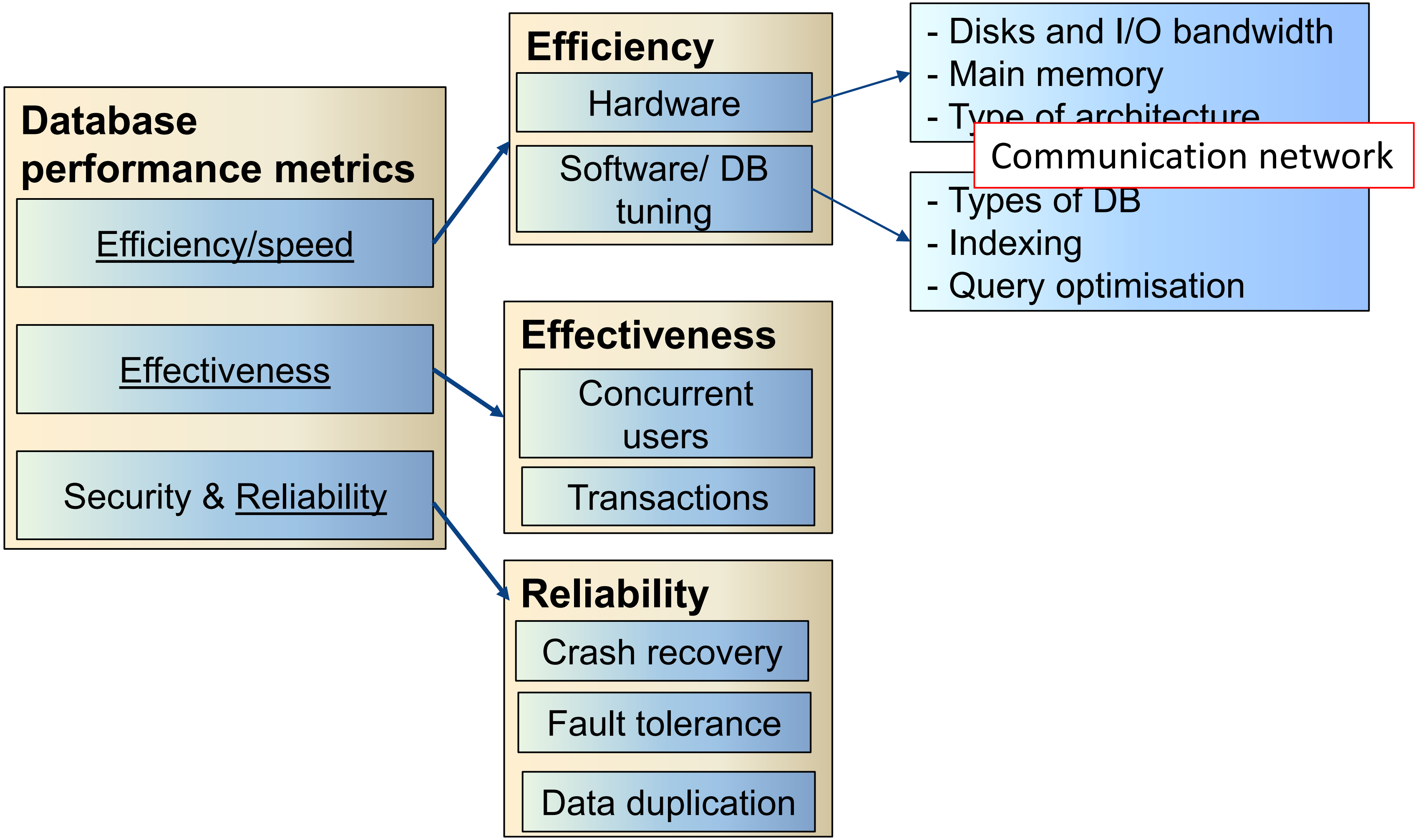
Hit ratio	Effective access time as multiple of C, M = 100 C
50.00%	50.5
90.00%	10.9
99.90%	1.1

Solve related problems in tutorials

What does this all mean?

- **Harddisk was the foundation stone for many DBMS design choices**
- What if CPU problems dominate?
- Another recent player is networking:
  - More and more databases are distributed
  - The network hardware speeds are at the speed of light already
  - Can this be the next determining front for DBMS design choices?
  - Some of these are already at play in various new systems!

Core Concepts of Database management system



Database performance metrics

- Efficiency/speed
- Effectiveness
- Security & Reliability

Efficiency

- Hardware
  - Disks and I/O bandwidth
  - Main memory
  - Type of architecture
  - Communication network
- Software/ DB tuning
  - Types of DB
  - Indexing
  - Query optimisation

Effectiveness

- Concurrent users
- Transactions

Reliability

- Crash recovery
- Fault tolerance
- Data duplication

Communication Costs

Increasingly, another item to model the cost of is data transfer:

transmit time = (distance/c) + (message_bits/bandwidth)

c = *speed of light* (200 million meters/sec) with fibre optics

*This means we can **no longer reduce latency on contemporary hardware further** and increasingly the motto is that the **message length should be large to achieve better utilization**.*

Can you relate the same idea for reading from HDD?

Types of Database Systems

How the data are stored –

Simple file

- As a plain text file. Each line holds one record, with fields separated by delimiters (e.g., commas or tabs)

RDBS

- As a collection of tables (relations) consisting of rows and column. A primary key is used to uniquely identify each row.

Object oriented

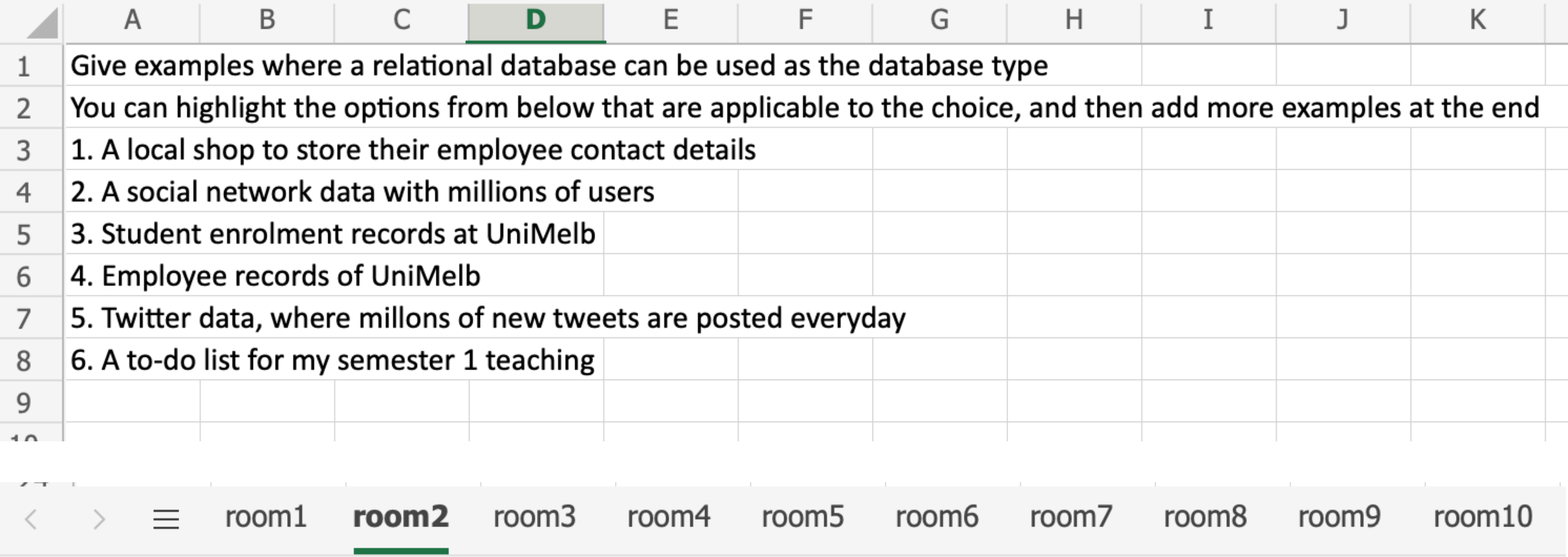
- Data stored in the form of 'objects' directly (like OOP)

No-SQL

- Non relational – database modelled other than the tabular relations. Covers a wide range of database types.

Time for breakout rooms!

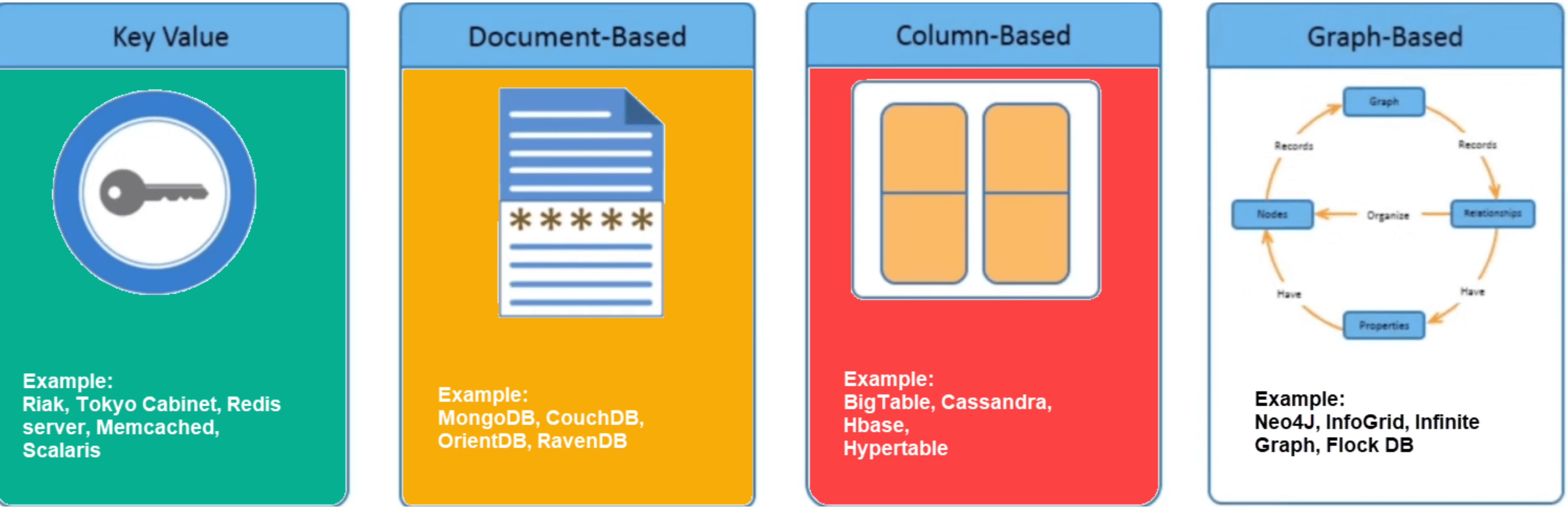
tinyurl.com/yvxt9zbh



A B C D E F G H I J K
room1 room2 room3 room4 room5 room6 room7 room8 room9 room10

	A
1	Give examples where a relational database can be used as the database type
2	You can highlight the options from below that are applicable to the choice, and then add more examples at the end
3	1. A local shop to store their employee contact details
4	2. A social network data with millions of users
5	3. Student enrolment records at UniMelb
6	4. Employee records of UniMelb
7	5. Twitter data, where millons of new tweets are posted everyday
8	6. A to-do list for my semester 1 teaching
9	

Types of NoSQL databases



Key Value

Example:
Riak, Tokyo Cabinet, Redis server, Memcached, Scalaris

Document-Based

Example:
MongoDB, CouchDB, OrientDB, RavenDB

Column-Based

Example:
BigTable, Cassandra, Hbase, Hypertable

Graph-Based

Example:
Neo4J, InfoGrid, Infinite Graph, Flock DB

Image source: https://jameskle.com/writes/no-sql

Examples and discussions in tutorials

Database Architectures

Centralized	Data stored in one location
Distributed	Data distributed across several nodes, can be in different locations
WWW	Stored all over the world, several owners of the data
Grid	Like distributed, but each node manages own resource; system doesn’t act as a single unit.
P2P	Like grid, but nodes can join and leave network at will (unlike Grid)
Cloud	Generalization of grid, but resources are accessed on-demand

Each architecture itself is quite a broad topic –
we will only look at what they are and a couple of important properties

Database Architectures

Cloud computing

Updated URL: https://youtu.be/mxT233EdY5c

More on Storage

- Storage today does not come as one disk
- They come as a system and increasingly complex
- Storage systems can determine the performance and also **fault tolerance** of a database
- In Database Management Systems (DBMSs), rarely data is stored in one location as well
- Storage is now much larger, involves multiple disks, and accessed over a network and at many sites

More on Storage Systems

We will discuss the following types of storage systems

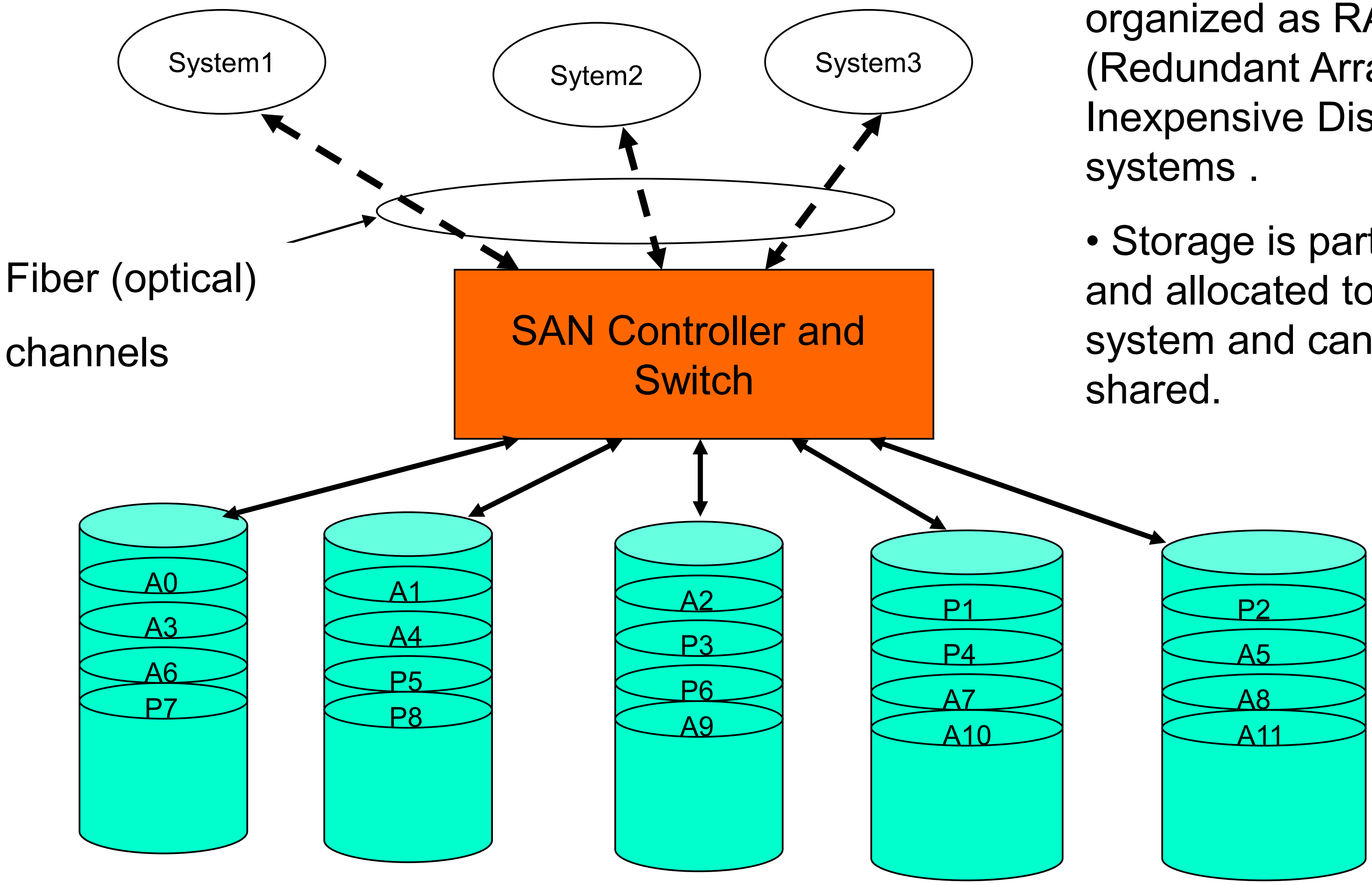
- RAID : Redundant Array of Independent Disks – different ways to combine multiple disks as a unit

  (Presented in pre-recorded lecture!)

- Storage Area Networks

Storage Area Networks (SAN)

A dedicated network of storage devices



- Storage can be organized as RAID (Redundant Array of Inexpensive Disks) systems .
- Storage is partitioned and allocated to each system and can also be shared.

More on SANs

- They are used for shared-disk file systems
- **Automated backup** functionality
- It was the fundamental storage for data center type systems with mainframes for decades
- Different versions evolved over time to allow for more data, but fundamentals are the same even today
- But in short, **failure probability of one disk is different than 100s of disks -** which requires design choices

Fault Tolerance

The property that enables a system to continue operating properly in the event of the failure of some of its components.

We have covered

- Statistics crash course
- Lifecycle of a system
- Different fault tolerance techniques

# P(A and B) = P(A)*P(B) assuming A and B are independent events.

# P(A or B) = P(A) + P(B) - P(A and B)

= P(A) + P(B) - P(A)*P(B) [Assuming A and B are independent]

≈ P(A) + P(B) [if P(A) and P(B) are very small]

Some activities on statistics!

PollEv.com/farhanachoud585

Respond at **PollEv.com/farhanachoud585**

Text **FARHANACHOUD585** to **+61 427 541 357** once to join, then text your message

Fault tolerance by RAID

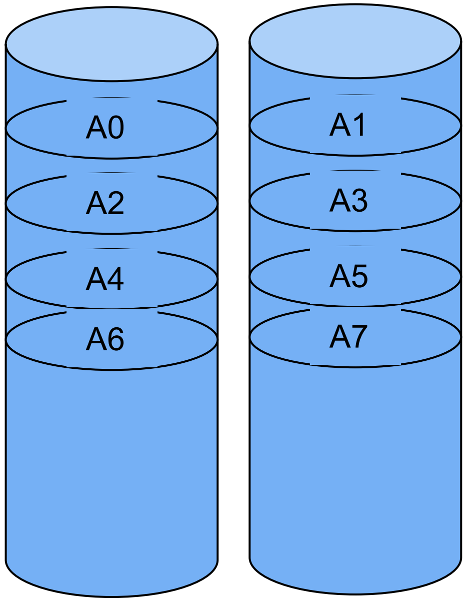
Redundant Array of Independent Disks – different ways to combine multiple disks as a unit for fault tolerance or performance improvement, or both of a database system

Choosing the suitable RAID level

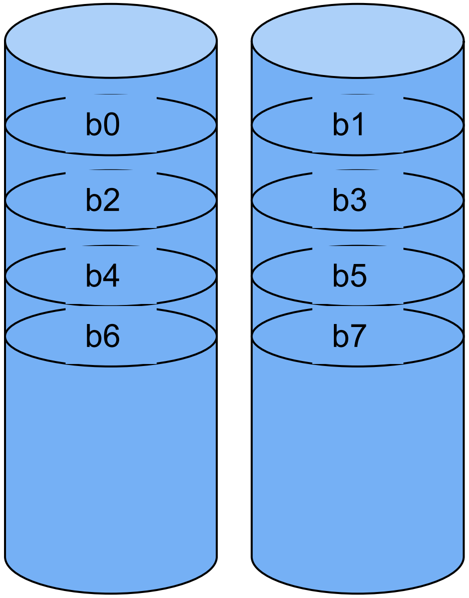
The factors to consider:

- Reliability
- Performance
- Storage utilization
- Price/number of disks

Example summary: RAID 0 and RAID 2



Block level striping



Bit level striping (rare)

Storage utilization?

What is the minimum number of disks needed?

Provides balanced I/O of disk drives

Provides higher throughput (~doubles)

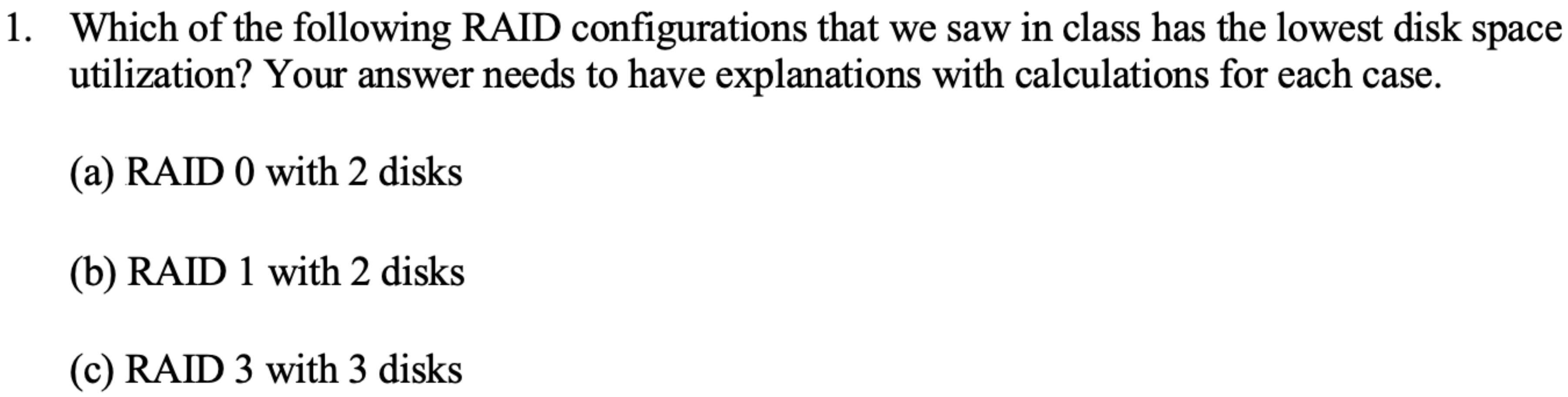
Any disk failure will be catastrophic

MTTF reduces by a factor of 2

Higher throughput at the cost of increased vulnerability to failures

Calculation of MTTF values – in tutorials

RAID – tutorial exercise



1. Which of the following RAID configurations that we saw in class has the lowest disk space utilization? Your answer needs to have explanations with calculations for each case.

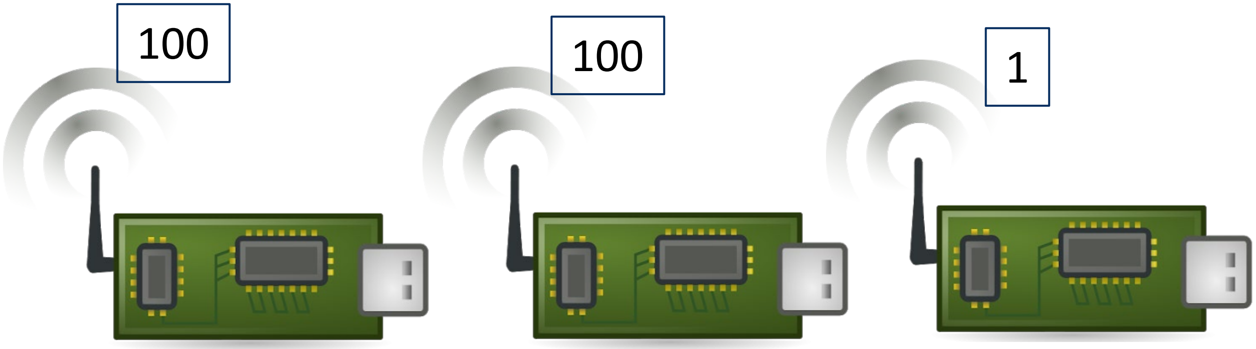
   (a) RAID 0 with 2 disks

   (b) RAID 1 with 2 disks

   (c) RAID 3 with 3 disks

Fault Tolerance by voting

Use more than one module, voting for higher reliability

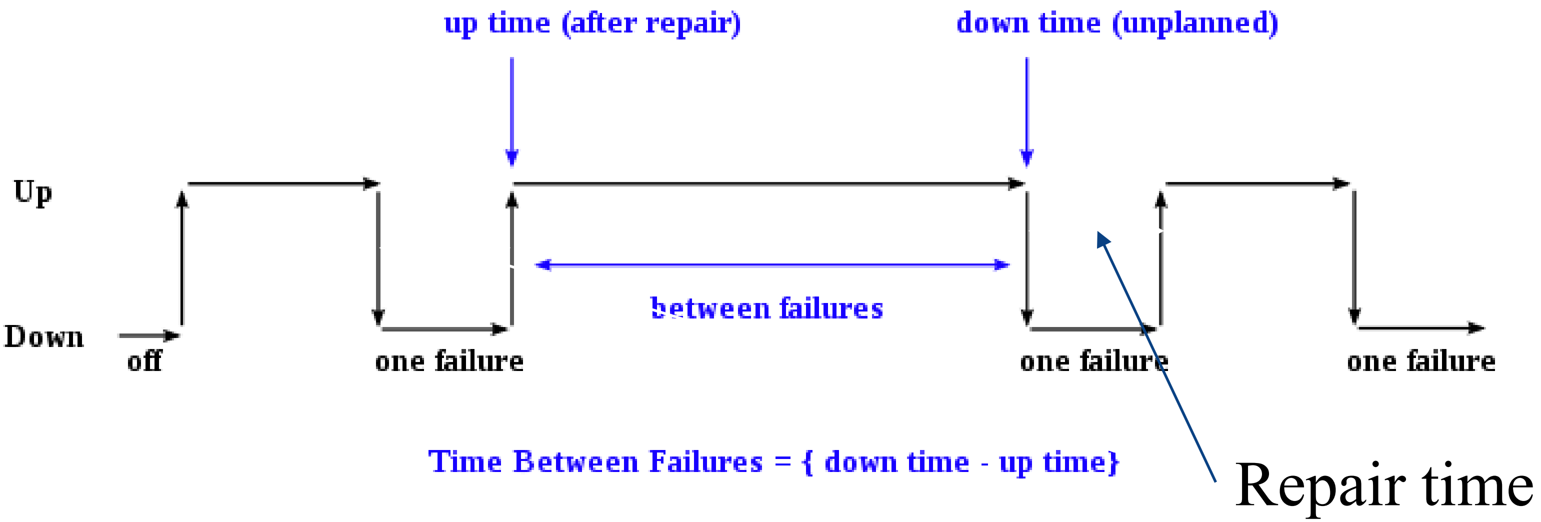


Failvote - Stops if there are no majority agreement

Failfast (voting)- Similar to failvote except the system senses which modules are available and uses the majority of the available modules.

Supermodule – A system with multiple modules that use voting when multiple modules are working/available, but still work even when only one is available

A system's lifecycle



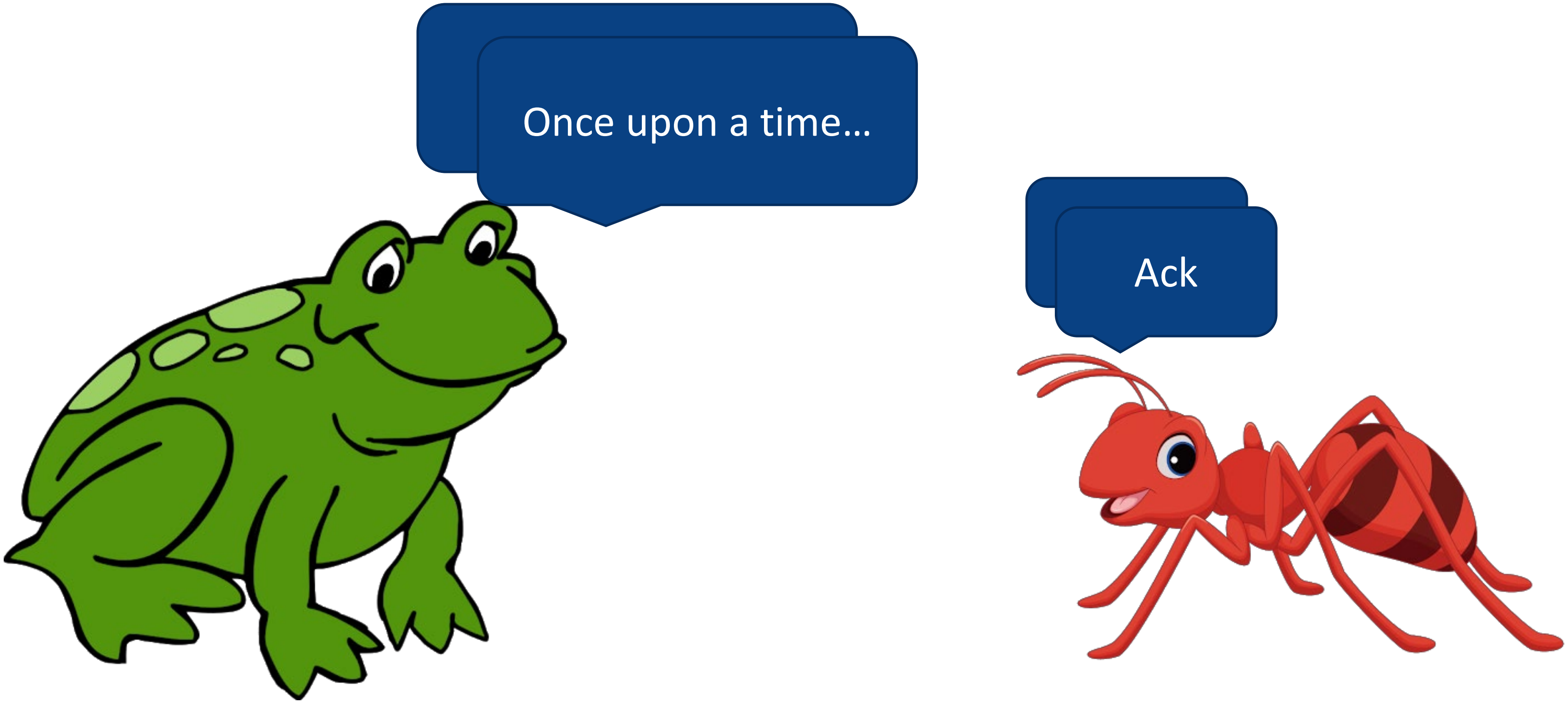
Module availability : measures the ratio of service accomplishment to elapsed time

the **time** elapsing before a failure is experienced

$$= \frac{\overbrace{\text{Mean time to failure}}}{\text{Mean time to failure} + \text{mean time to repair}}$$

Availability of a system: (i) Without repair (ii) With repair

Communication reliability



Communication reliability

If acknowledgement not received -

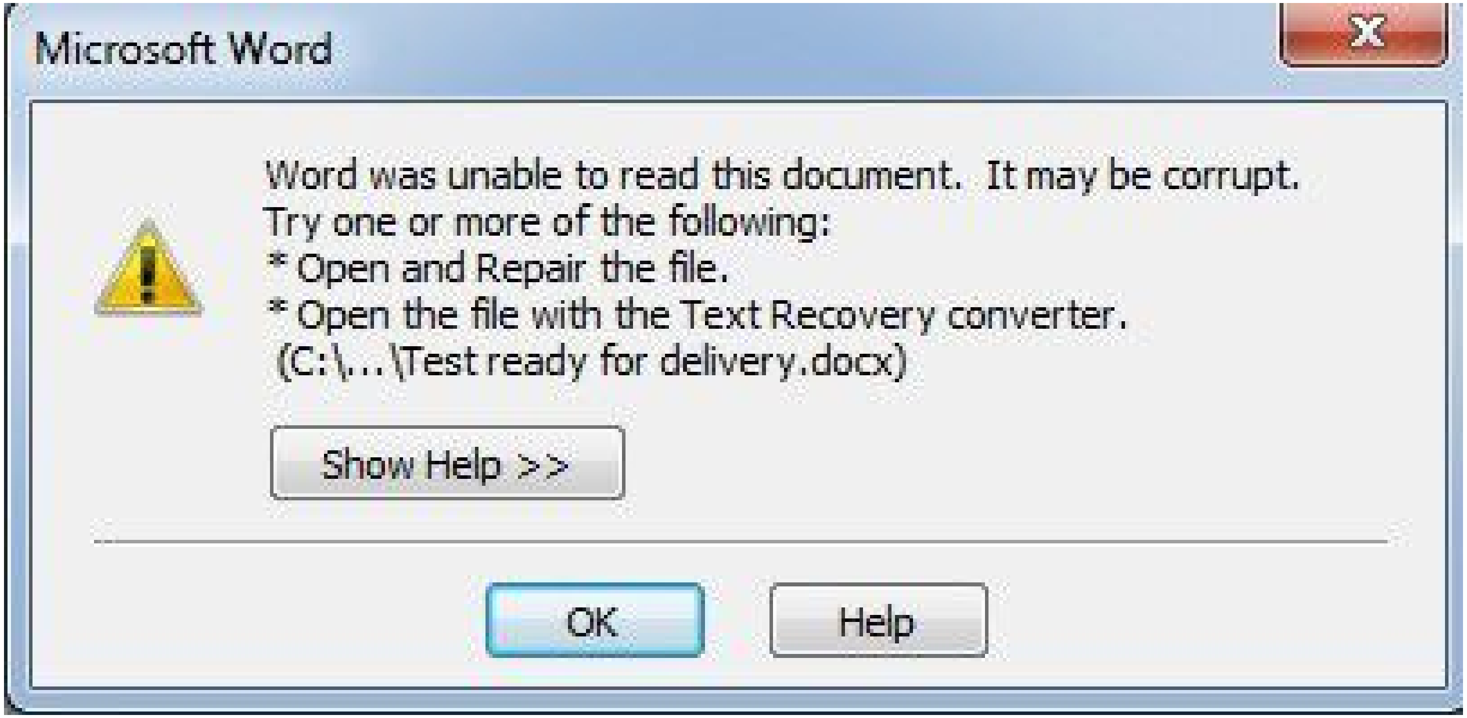
I will tell a story

I will tell a story

I will tell a story

Disk writes

Either entire block is written **correctly** on disk or the contents of the block is unchanged.



Word was unable to read this document. It may be corrupt.
Try one or more of the following:
* Open and Repair the file.
* Open the file with the Text Recovery converter.
(C:\...\Test ready for delivery.docx)

- What type of systems use duplex writes?
- Difference with RAID 1?

Cyclic Redundancy Check (CRC)

An error detection algorithm

1. A polynomial needs to be specified

2. A sequence of bitwise exclusive-or (XOR) operation needs to be performed

3. The final CRC value needs to be stored for each data block (or the data unit on which CRC is performed)

4. Data correctness can be checked with CRC -

    a. its corresponding CRC value is retrieved

    b. A sequence of bitwise XOR operation needs to be performed to find out the correctness of data

Non-examinable resource:

Reliability, disk write reliability, RAID, CRC, etc. in use in real systems –

Dell EMC Unity (storage unit by Dell)

https://www.delltechnologies.com/asset/en-us/products/storage/industry-market/h17076-dell-emc-unity-data-integrity.pdf (duplex write is the same as synchronous replication in that document).

Summary

Performance and Reliability are important

Achieving reliability requires additional hardware/algorithms

- Effect of Hardware on performance – different memory hierarchy
- Hardware reliability
- Communication reliability
- Disk write reliability