Exercise 1 Solution

1. In a hard disk drive (HDD), the average seek time is 12 ms, the rotation delay is 4 ms, and the transfer rate is 4MB/sec. For simplicity, we assume in this question only 1MB equals 1000KB.

   (a) What is the seek time delay?

   (b) What is the rotation delay?

   (c) What will be the disk access time for a transfer size of 8MB? What will be the disk access time for a transfer size of 8KB?

   (d) In a solid state drive, what will be the disk access time for a transfer size of 8MB when the transfer rate is 4MB/sec? Is an SSD faster than an HDD for the same amount of data transfer (Assuming the base sequential data transfer rates are the same for the given two drives.)? Why?

Solution:

(a) The seek time delay/seek latency is the period that the head of the actuator arm moves from a position to a required track.

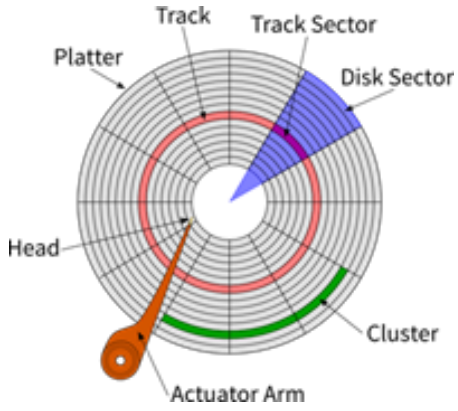

Figure 1: Components of a Hard Disk Drive (HDD).

(b) The rotation delay/rotation latency is the waiting period that the rotation of the disk brings the required track sector to the head of the actuator arm.

(c) Disk access time for 8MB:

$$seektime + rotationtime + \frac{transferlength}{Bandwidth}(ms) \tag{1}$$

$$\longrightarrow 12 + 4 + \frac{8}{4} \times 1000(ms) = 2016ms$$

Disk access time for 8kB:

$$seektime + rotationtime + \frac{transferlength}{Bandwidth}(ms) \tag{2}$$

$$\longrightarrow 12 + 4 + \frac{8}{4 \times 1000} \times 1000(ms) = 18ms$$

Comments: A comparison of the two cases highlights that sequentially reading large data pays off as seek time is buried under a lot of transfer time. For example, in the first case, seek time is only 0.6% of the total time while nearly all the time is spent on transferring data. In the second case, seek time is 66.7% of the total time while only a small fraction of the time is spent on data transfer.

(d) Disk access time of SSD:

$$\frac{transferlength}{Bandwidth} = \frac{8}{4}(s) = 2(s) \tag{3}$$

Unlike an HDD, an SSD does not have any moving parts. Hence there is no rotation delay or seek delay in an SSD. Therefore, the speed of SSDs is higher than that of HDDs given the assumptions in the question.

2. There are two different machines where machine A has a smaller cache with an average 50% cache hit ratio (H) and the other machine (machine B) has a much larger cache with an average 90% cache hit ratio. However, the memory access time of machine A is 100C and the memory access time of machine B is 400C (i.e., memory access in machine A is faster than memory access in machine B), where C is the cache access time. Which machine has an overall faster effective memory access time?

Solution:
Effective memory access time of A:

$$0.5 \times C + (1 - 0.5) \times 100C = 50.5C \tag{4}$$

Effective memory access time of B:

$$0.9 \times C + (1 - 0.9) \times 400C = 40.9C \tag{5}$$

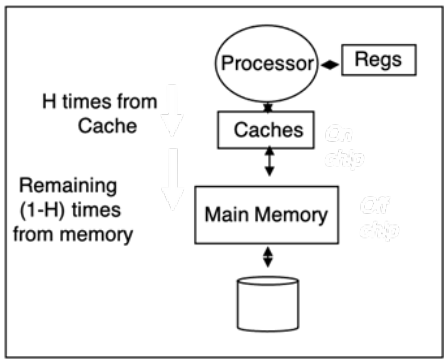

Figure 2: Hierarchical Memory Structure.

3. Supplementary details on NoSQL databases (provided in the Exercise).

4. Discuss example applications of different types of NoSQL databases.

Solution:

- *Applications for key-value databases*: Suitable if the dataset does not need a complex relational table type of structure but can be expressed with simple key-value pairs. The simple structure allows faster insertion and search and scales quickly. For example, a shopping cart database in an e-commerce site.
- *Applications for document storages*: Well suited when different kinds of documents do not always have the same structure/sections. For example – a database of news articles.
- *Applications for column-based NoSQL databases*: They are used in data analysis applications/tasks.
- *Applications for graph databases* : Well suited for storing connection data such as social network connections and spatial data.

5. Consider the different scenarios below and discuss which database architecture is the most suitable choice and why:

- FriendBook is a new startup app that will launch its operation soon. They have only one office without much budget right now, but they are expecting high growth in the scale of millions of users across the globe in a couple of years. Which of the following database architecture is the most suitable choice for this scenario?
  - **Cloud storage**
  - World wide web

  - Distributed database
  - Centralised database

  Solution:

  Comments: Cloud storage allows for data to be stored across multiple servers in data centers, making it easier to scale horizontally as the number of users grows. This type of architecture also provides better reliability and fault tolerance.

- FriendBook is a new social network site that will launch its operation soon. They have offices in many major cities in the USA. They need a database that can handle millions of users across the globe. For preserving privacy and security, they need their own data storage system, which is not shared or owned by any other company. Which of the following database architecture is the most suitable choice for this scenario?

  - Cloud storage
  - World wide web
  - **Distributed database**
  - Centralised database

Solution:

Comments: Unlike the previous scenario, if data is transferred and stored in a 3rd party storage like the cloud, the security is not in the hands of FriendBook (including encryption guarantee, data discloser agreement, etc.). Hence, having the setup of their own distributed database (as they are located across many cities with many users across the globe) is a more suitable solution.

Note: There is no universal truth or final answer on which architecture should be chosen for an application in some cases. The characteristics, advantages, and disadvantages guide us on which one is more suitable over the others and even then some decisions are borderline if the pros – cons are approaching the same level between the two choices.