COMP90050 Advanced Database Systems: Tutorial
Winter term, 2023 (Week 1)

Exercises

Part 1: Walk through of the Group Project description

Part 2:

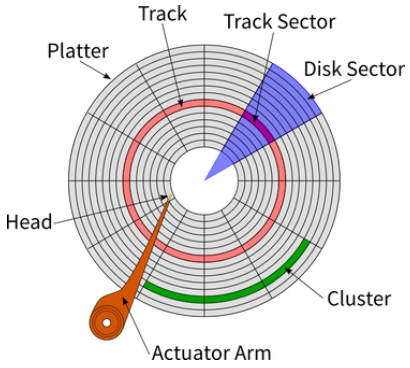
1. In a hard disk drive (HDD), the average seek time is 12 ms, rotation delay is 4 ms, and transfer rate is 4MB/sec. For simplicity, we assume in this question only 1MB equals 1000KB.

a) What is seek time delay?

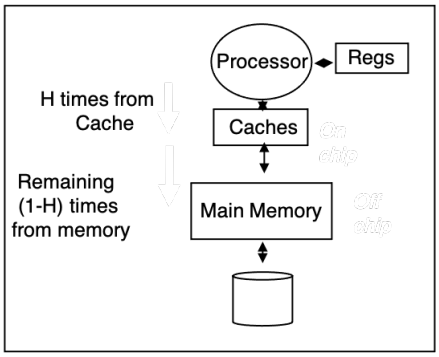
b) What is rotation delay?

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 transfer rate 4MB/sec? Is an SSD faster than an HDD for the same amount of data transfer? Why?



2. There are two different machines where machine A has a smaller cache with on average 50% cache hit ratio (H) and the other machine (machine B) has a much larger cache with on 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 overall faster effective memory access time?



3. **More details on NoSQL databases - different types of NoSQL databases:**

a. Key-value storage: A key-value database stores data as a collection of key-value pairs where a key serves as a unique identifier. All accesses to the database are done via the keys. Both keys and values can be complex.

b. Document storage: Flexible for storing different kinds of documents, where they may not all have the same sections. XML, JSON, etc. are subclasses of document-oriented databases.

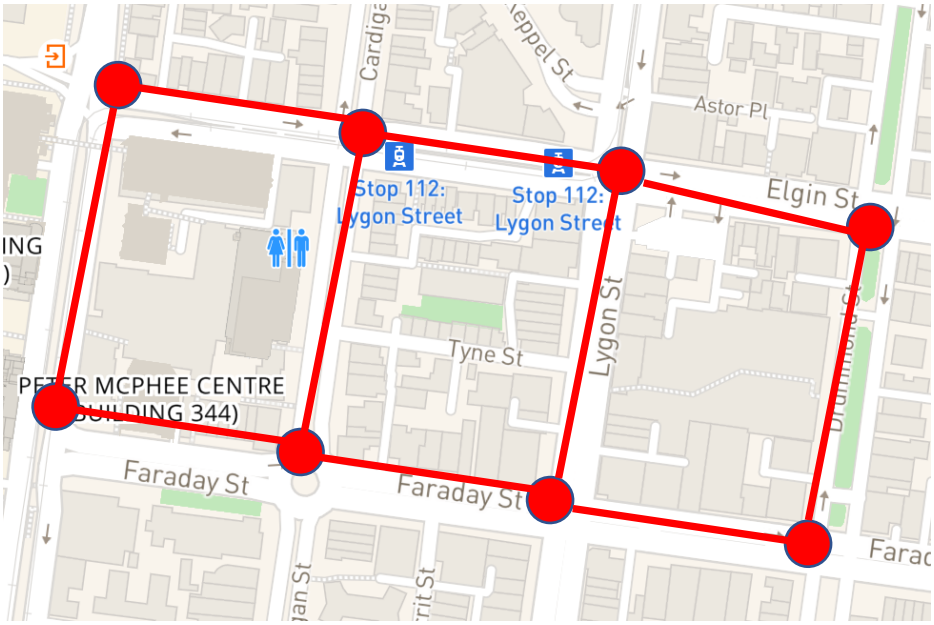
```
<contact>
  <firstname>Bob</firstname>
  <lastname>Smith</lastname>
  <phone type="Cell">(123) 555-0178</phone>
  <phone type="Work">(890) 555-0133</phone>
  <address>
    <type>Home</type>
    <street1>123 Back St.</street1>
    <city>Boys</city>
    <state>AR</state>
    <zip>32225</zip>
    <country>US</country>
  </address>
</contact>
```

Image source: Wikipedia

c. Graph storage: Graphs capture connectivity between entities. Searching and traversing by relations are very fast in such structures.

The links can be material or immaterial:

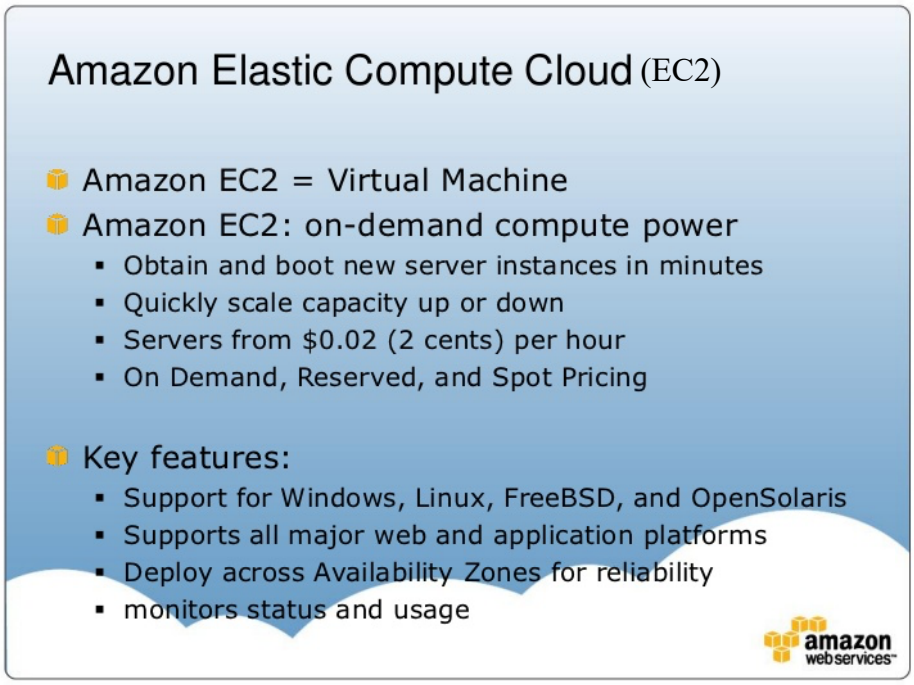
- Links between two streets are junctions;
- Links between people as their Facebook connections (non-material links)
- a graph is a structure amounting to a set of objects (called vertices) where some pairs of the objects are connected/related in some sense. A connection is called an edge.



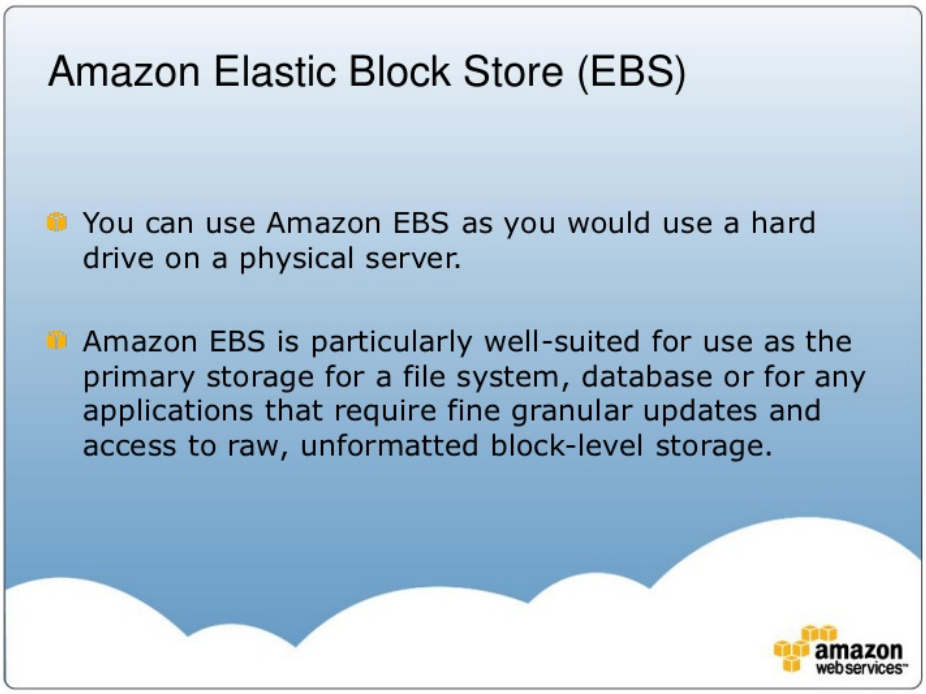
d. Discuss example applications of different type of NoSQL databases.

Part 3

4. Review commercially available different cloud computing and database services from Amazon



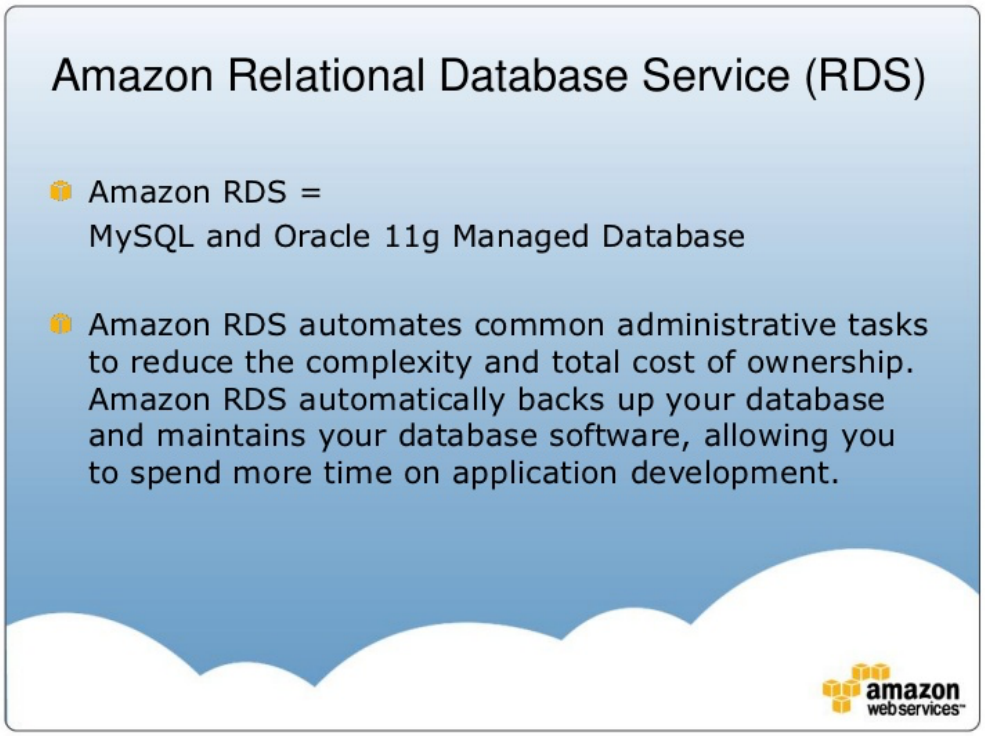
1



Amazon Simple Storage Service (S3)

- In traditional on-premise applications, this type of data would ordinarily be maintained on **SAN** or **NAS**. However, a cloud-based mechanism such as Amazon S3 is far more agile, flexible, and geo-redundant.
- Amazon S3 is a highly scalable, durable and available distributed object store designed for mission-critical and primary data storage with an easy to use web service interface.

amazon web services



5. Discuss the advantages and disadvantages of different database architectures for different application scenarios.

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

I. FriendBook is a new startup app that will launch its operation soon. They have only one office with not much budget right now, but they are expecting a 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?

a. Cloud storage
b. World wide web
c. Distributed database
d. Centralised database

II. FriendBook is a new social network site that will launch its operation soon. They have offices in many major cities of 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?

a. Cloud storage
b. World wide web
c. Distributed database
d. Centralised database