Gurjus Singh

MSDS 432 Foundations of Data Engineering

July 5th, 2020

Module 2 – Reading Comprehension

1. **Define and describe the Relational Model.**

   The relational model is one which most developers probably know well with SQL. The way it works is there are relations otherwise known as tables and each table contains an entity. The table contains tuples or rows, and attributes or columns. The history of the relational model is that it was used as preferred method beginning in the 1980s. It has been around since the last 25-30 years. It is used mostly for business data such as sales, customers, inventory, employee data, and payroll.

2. **Define and describe Document Databases.**

   Is a type of NO SQL database which stores data in documents. This data can be stored in a JSON format. It is human readable, semi structured and is hierarchical. Another thing is it. Documents are usually stored as a single continuous string. The advantage of this is all of the data is in the same document which is called having the advantage of "locality". There are also disadvantages to the locality rule such as loading the data all at once, when the only part of the document is needed. This can waste time and space.

3. **Define and describe Graph Databases.**

   Graph databases are databases that have vertices and edges. These databases can be used to represent relationships or geographic locations. For example, GPS systems use graph databases to represent a route, and find the fastest route using Djikstra's algorithm. Graph Databases can also be used to represent your friends list on Facebook where one vertex represents a person and edge represents whether that person has a friend.

4. Discuss the differences between three different Query Languages covered in this chapter.

Three types of query languages discussed are Cypher, SPARQL and Datalog. The differences between each is Cypher is used for property graphs, while SPARQL is used for the triple stores model which is written as (subject, predicate, object) form. SPARQL is much older than Cypher although they are similar in syntax. Datalog meanwhile is older than both Cypher and SPARQL and is used for systems such as Datomic and Cascalog. Its model is kind of like the triple stores model. The data is written in predicate(subject, object) form.

5. Discuss the differences between Linked Lists and Arrays from the Algorithms reading (Bhargava)

Differences between Linked List and Arrays is extracting elements in Arrays involves indexing while in Linked Lists it involves traversing through the whole LinkedList in order to find what you are looking for. Arrays are much trickier when you want to add

elements, as you have to shift over your elements to add one more while Linked lists are just a collection of elements that point to the address of the next element, and the element can be stored anywhere in memory. Also running times are different for each as it takes linear time to insert and delete in an Array while it takes constant time to do that in Linked Lists. Meanwhile reading has better time complexity in Arrays than in Lists. These are the key differences between both data structures.