

Exercise 7

1.) Big Data In The Cloud:

- a.) Wikipedia says that big data is data that is simply too large to be dealt with by traditional data processing software, while IBM says that big data does not have distinct searchable fields, meaning it has less to do with the size directly, and more to do with how the data is structured. The reasoning behind this is that with new technology constantly being brought into play, the data gathered by these new machines is getting exponentially bigger and harder to manage. Wikipedia had more characteristics to define big data that IBM did not include; Scalability, Exhaustive, Relational. IBM took a bigger emphasis on explaining that the solution to 'Big Data' is machine learning. Humans cannot process the amount of data we collect, and the number of petabytes of data we create each day is growing exponentially, so it's smart to think that data analytics is going to become more of an AI job over the next several years.
- b.) I think this is one of the single biggest (technological) challenges that we will find ourselves facing in the next several decades. As we rely more and more on computers, AI, virtual reality, digitized currency, blockchain, etc., we will create so much more data than we can reasonably store and use. I think that without the use of machine learning for data analysis and compression, we will either have too much data to store, or we will have too much data to try to send around the world at one time.
- c.) NoSQL is designed with the idea of processing large amounts of data in mind. According to Wikipedia, scalability is a strong point for NoSQL, while some programming languages, even low level ones like C++, can have issues when scaling for large amounts of data.

2.) Ubiquitous B-Tree:

- a.) Unlike binary trees, B-trees are self balancing, all leaves are at the same depth meaning that there are no branches that are longer than others. With a Binary tree, as data is added randomly, you run the risk of having 'long' paths and 'short' paths. The Longest path in a B-Tree of n keys is about $\log_D n$ nodes, where D is the order of the tree. A find operation may traverse N nodes in an unbalanced tree of N records, but a B-tree that is balanced will only require the search of $\log_D n$ nodes. Because of the self balancing ability, a B-tree will always have a lower height.
- b.) In a B+ tree the keys are all stored in the leaf nodes. This means that when searching a tree, it does not matter what values are encountered as long as the path leads to the correct leaf. When doing sequential processing of a file, no node will be accessed more than once, so there only needs to be space for one node in main memory. They also allow for dynamic allocation and deletion of memory. This means that B+ Trees are better suited to applications that use both random and sequential processing.

- c.) The InnoDB index is a B+ tree because it follows the pattern of **storing keys in the leaf** nodes along with **data**. Each node also can store more than one key, which speeds up indexing time. The leaf nodes are linked by 1 or 2 pointers, making the performance faster when doing sequential insertion or indexing.

3.) Lore: A database Management System

- a.) With OOP languages like C++ you have the ability to create **objects**, which are groups of attributes with methods that can be used on them to manipulate or view data within that object. In LORE, you can create an Object Exchange Model. An Object Exchange Model shows **interactions between objects**.
- b.) Websites and Mobile Platforms are a good example of this because there is still a pattern to the data, but there is a lot of flexibility within the structure and usage of the data. Relational tables are going to be better suited for areas where the data structure is more rigid, while still having an emphasis on structure and pattern, such as accounting software or perhaps banking transaction databases.
- c.) A path search in LORE requires more complex queries, but it is where the search is only used to extract a specified condition. LORE is more complex than SQL and it has some functionalities not present in SQL.