

Assignment (Video- 16 to 18): Database System Impl. (COP6726)

Name: Vikas Chaubey, UFID: 3511 5826, Email: vikas.chaubey@ufl.edu

- 1) **Sort based Algorithm:** Sorting means putting the elements in a given sequence in a certain order which could be increasing or decreasing. The orientation of order is decided based on a fixed rule which is followed through the sequence to sort it. most of the times frequently used orders are numerical order and lexicographical order. There are various types of sorting algorithm available which can sort the data but the important factor to consider a sorting algorithm is its efficiency. For a particular use case one sorting algorithm might prove to be more efficient compared to others. in order to make an algorithm efficient which utilizes sorting algorithm as a component, it is important that sorting algorithm used inside is also efficient. Sorting is also often useful for canonicalizing data and for producing human-readable output. Sorting algorithm must follow the two conditions, the first one is that the output should be in non-decreasing order and the output should be the permutation of original sequence, i.e. the output should contain the same elements as the original input. For optimum efficiency, the input data in fast memory should be stored in a data structure which allows random access rather than one that allows only sequential access.
- 2) **In memory sorting:** when the large amount of unordered data needs to be sorted then external sorting is used. when the amount of data which needs sorting is huge then in that case it is not possible to import data on the main memory and use traditional sorting techniques to sort it. It is because main memory has limited storage capacity. In that case to do sorting a different technique is required. This technique is called external sorting because data is kept in the hard disk and chunks of data are moved to main memory to perform sorting operation. External sorting uses a hybrid sort merge strategy. In the sorting phase, chunks of data small enough to fit in main memory are read, sorted, and written out to a temporary file. In the merge phase, the sorted sub-files are combined into a single larger file. for example external merge sort algorithm , first divides the whole file data into small chunks which are small enough so that they can fit on the main memory , these chunks are called runs , these runs are sorted individually and they when all runs are sorted these are merged together to form bigger runs , this is done until the whole file is sorted. External sorting algorithms generally fall into two types, distribution sorting, which resembles quicksort, and external merge sort, which resembles merge sort.
- 3) **Hashing Based Algorithms:** A hashing algorithm is consisting of cryptographic hash function. This algorithm maps the data of random size into hash of a fixed size. generally, a hash function is a one-way function on which these algorithms are based that means the output of the hash function is not feasible to be reversed or inverted. however, in recent years several hashing algorithms have been cracked and compromised. A widely known hash function designed to be a cryptographic hash function, which is now so

easy to reverse — that we could only use for verifying data against unintentional corruption. an ideal cryptographic function should be able to compute the hash value for any input in fast manner. also, the output given by cryptographic function should be impossible to invert, it should only be possible by brute force but no other method, the hash created for each input should be unique, the hash function should not generate similar hashes for different inputs. Every change to a message, even the smallest one, should change the hash value. It should be completely different. It's called the avalanche effect.

- 4) **In memory hashing:** In DBMS, hashing technique is used to search the location of desired data on the disk without making use of index structure. using hashing techniques databases index the data and data retrieval is also done, as it is faster to search the specific data items using the shorter hashed key instead of using its original value. A hashing algorithm is a complex mathematical calculation that takes an input and return a value called a hash value or hash. When used for memory addressing the hash value generated is the memory location of where the record is stored. A database contains millions of records hence locating a single record in a table can be time consuming, when the record is searched using linear search. using binary trees this process could be made faster but binary tree requires are to be sorted in order to do search efficiently. In case of practical applications, the data base data is not sorted hence searching records is a slow process. That's where hashing can prove to be efficient
- 5) **Join:** A database join query performs a join operation in relation algebra which combines columns from one or more tables in a relational database. It creates a set that can be saved as a table or used as it is. There various types of joins such as symmetric vs asymmetric.
- 6) **Union:** the union operation I used to combine the result of two or more select statements in database. It combines the result sets of two separate clauses in the database.
- 7) **Nested Loop Algorithm:** A simple nested loop algorithm reads rows from the first table in a loop one at a time, passing each row to a nested loop that processes the next table in the join. This process is repeated as many times as there remain tables to be joined.