

ASSIGNMENT-5

A.

1. Sequential access method
2. access code
3. Cluster
4. Hashed Tree
5. Data Warehouse
6. Data Mining
7. Data Redundancy
8. Metadata
9. Bt tree allows rapid random access as well as sequential access.
10. File Management System

B.

1. Define heap file organization
It is the simplest and most basic type of organisation. It works with datablocks in heap file organisation, the records are inserted at the file's end. When the records are inserted, it doesn't require the sorting and ordering of records.

When the data block is full, the new record is stored in some other block. This new block need not to be the very next block but it can select any data block in the memory to store new records. This heap file is also known as an unordered file.

2. Define ~~Sequential~~ heap file organisation.

~~It is the simplest and most basic type of organisation. It works with datablock in~~

A sequential file contains records organised by the order in which they were entered. The order of the records is fixed.

Records in sequential file can be read or written only sequentially. After you place a record into a sequential file, you cannot shorten, lengthen or delete the record.

3. Define hash file organisation.

~~Hash file organisation~~ uses the computation of hash function on some fields of the records.

The hash function's output determines the location of disk blocks where records are to be placed. When a record has to be

retrieved using hash key columns then the address is generated and the whole record

is retrieved using that address in the same way. When a new record has to be inserted then the address is generated using the hash key and record is directly inserted. The same is done for delete and update.

4. What is a file system.

File system is the way in which files are named and logically placed for storage and retrieval. Without this, stored information wouldn't be isolated into individual files and would be difficult to identify and retrieve.

C

1. Explain the B+ tree indexes of multiple keys with a suitable eg.

A B+ tree is an advanced self-balancing tree. It uses a multilevel indexing system. Leaf node in the B+ tree represent actual data references. It keeps all of the leaf nodes at the same height. A link list is used to connect the leaf nodes in B+ tree so it can allow both the random and sequential access.

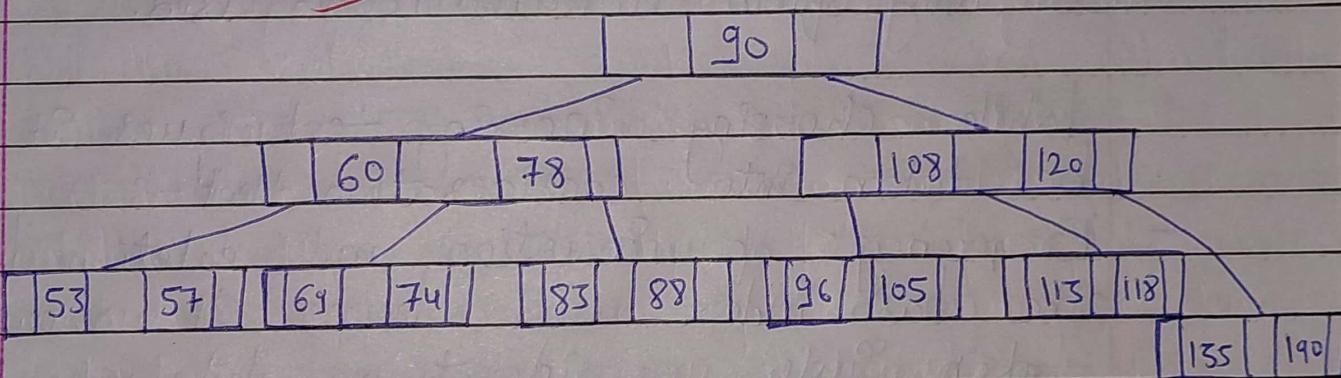
Both keys and records can be placed in the interval and leaf nodes of the B tree.

In a B+ tree records or data can only be kept in leaf nodes, whereas key values can only be placed on the interval nodes.

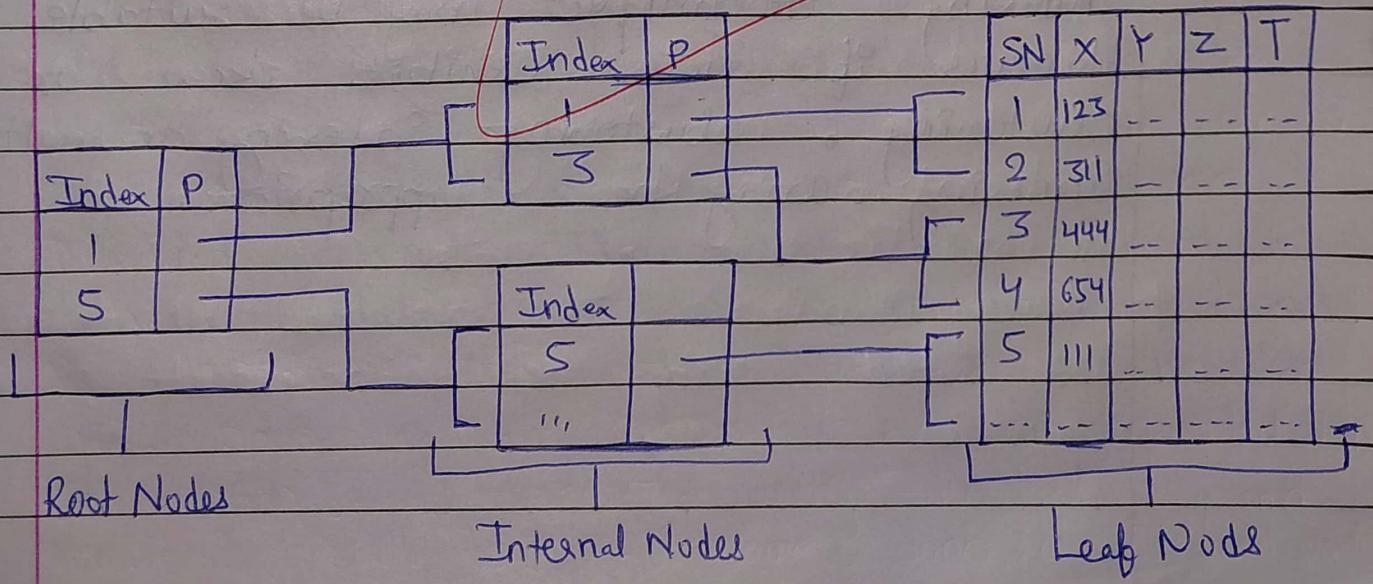
To make search queries more efficient, the leaf nodes of the B+ tree in data structure are connected together in the form of singly linked list.

B+ trees are used to store vast amounts of data that are too large to fit in the main memory. The interval nodes of the B+ tree (the keys to access records) are securely stored in memory, whereas leaf nodes are placed in the secondary memory due to the restricted amount of main memory.

~~Ex of B+tree of order 3 :-~~



In B+ tree the index of indices is formed in multilevel indexing as :-



2. What are various issues one should consider while choosing file organisation and indexing techniques?

When choosing file organisation we should consider following factors.

- i. Fast Retrieval
- ii. High amount of work for processing data
- iii. Efficient use of storage space
- iv. Minimizing need for reorganization
- v. Protection from failures and data loss
- vi. Accommodating Growth
- vii. Security from unauthorized used.

While choosing indexing techniques it should be taken into consideration that -:

- Frequency of insertion and deletion determines the hash index and hence it should be taken into consideration while choosing an indexing technique.
- Data distribution within the database, if the data is evenly distributed, a uniform indexing technique may be suitable. However if data exhibits skew or clustering, clustering indexing or multi-level indexing may be appropriate.

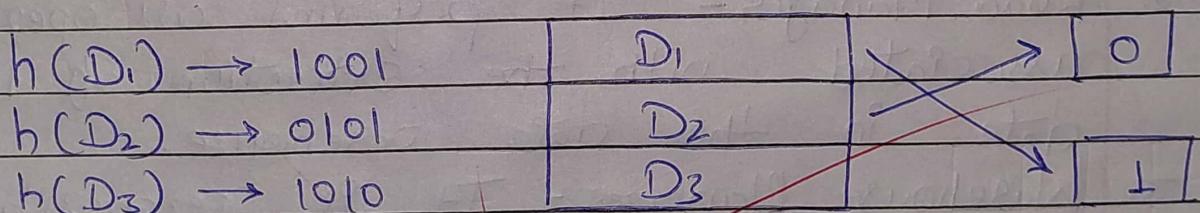
(D)

1. How does Dynamic Hashing manage file expansion? What are the problems with static hashing?

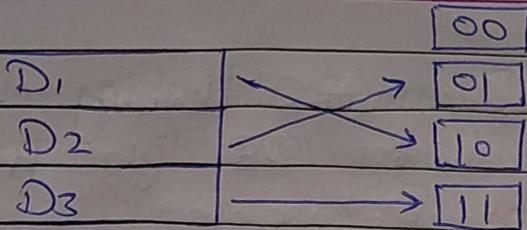
The drawbacks of static hashing is that it does not expand or shrink dynamically as the size of the database grows or shrinks.

In Dynamic Hashing, data buckets grows or shrinks (added or removed dynamically) as the records increase or decrease and hence is also called extended hashing. In dynamic hashing, the hash function is made to produce large no. of values.

For eg. - there are 3 data records D_1, D_2 and D_3 . The hash function generates 3 addresses - 1001, 0101 and 1010 respectively. This method of storing considers only part of this address - especially only first one bit to store the data. So it tries to load three of them at address 0 and 1.



But the problem is that no bucket address remains for D_3 . The bucket has grown dynamically to accommodate D_3 . So it changes the address to have 2 bits rather than 1 bit and it update the existing data to have 2 bit address. Then it tries to accommodate D_3 .



So,

- In dynamic hashing the performance doesn't decrease as the data grows in the system. It simply increase the size of memory to accommodate the data.
 - As the memory is well utilized (it grows and shrink with data) there will not be any unused memory lying.
2. Describe the various components of data warehouse and explain different data model used to store data with example.
- Components of datawarehouse includes :-
 1. Operational Source - It is a data source consisting operational and external data.
 2. Load Manager - It performs all operations associated with the extraction of loading data in the data warehouse.
 3. Warehouse Manager - It is responsible for warehouse management processes like analysis, aggregation, backup and collection of data, denormalization of data.
 4. Query Manager - It performs all tasks associated with the management of user queries.

5. Detailed Data - It is used to store all the detailed data in database schema. Detailed data is loaded into data warehouse to complement the data collected.
6. Summarized Data - It is part of data warehouse that stores predefine aggregations generated by warehouse manager.
7. Archive and Backup data - The detailed data in the database schema. Data is relocated to storage archive such as magnetic tapes or optical disks.
8. Metadata - It is used for extraction and loading process, warehouse management process and query management process.
9. End-User Access Tool - It consist of analysis, reporting and mining. By using these tool, user can link with warehouse.

- Different data models used to store data are:
1. Relational data model - It designs the data in the form of row and columns within the table. Thus this model uses tables for representing data and in between relationships.
 2. Entity-relationship model - It is the logical representation of data as objects and relationships among them. Here objects are known as entities and relationship is among is association among these entities.

3. Object Based Data Model - It is an extension of ER model with notions of functions, encapsulations and object identity as well. This model supports a rich type system that includes structured and collection types.
4. Semistructured Data Model - It allows the data specifications at places where the individual data items of the same type may have diff. attribute sets. The XML is widely used for representing the semi-structured data.

Final Exam