Course code : **CSE2007**
Course title : **Database Management System**
Module : **6**
Topic : **2**

Operation on Files

Objectives

This session will give the knowledge about

- Operations on Files
- Files of Unordered Records (Heap Files)
- Files of Ordered Records (Sorted Files)

Operations on Files

A file organization refers to the organization of the data of a file into records, blocks, and access structures; this includes the way records and blocks are placed on the storage medium and interlinked.

An access method, on the other hand, provides a group of operations that can be applied to a file.

Retrieval operations

- No change to file data

Update operations

- File change by insertion, deletion, or modification

Records selected based on selection condition

Operations on Files

Examples of operations for accessing file records

- Open - Prepares the file for reading or writing.
- Reset - Sets the file pointer of an open file to the beginning of the file.
- Find (or Locate) - Searches for the first record that satisfies a search condition.
- Read - Copies the current record from the buffer to a program variable in the user program
- FindNext - Searches for the next record in the file that satisfies the search condition.

Operations on Files

- Delete - Deletes the current record and (eventually) updates the file on disk to reflect the deletion.
- Modify - Modifies some field values for the current record and (eventually) updates the file on disk to reflect the modification.
- Insert - Inserts a new record in the file by locating the block, transferring that block into a main memory buffer, writing the record into the buffer, and (eventually) writing the buffer to disk to reflect the insertion.
- Close - Completes the file access by releasing the buffers and performing any other needed cleanup operations.

Operations on Files

The preceding (except for Open and Close) are called record-at-a-time operations because each operation applies to a single record.

It is possible to streamline the operations Find, FindNext, and Read into a single operation, Scan, whose description is as follows:

Scan - If the file has just been opened or reset, Scan returns the first record; otherwise it returns the next record.

If a condition is specified with the operation, the returned record is the first or next record satisfying the condition.

Operations on Files

Additional set-at-a-time higher-level operations may be applied to a file. Examples of these are as follows:

- FindAll - Locates all the records in the file that satisfy a search condition.
- Find (or Locate) n - Searches for the first record that satisfies a search condition and then continues to locate the next n − 1 records satisfying the same condition.
- FindOrdered - Retrieves all the records in the file in some specified order.
- Reorganize - Starts the reorganization process. As we shall see, some file organizations require periodic reorganization.

Files of Unordered Records (Heap Files)

Heap or pile file is the simplest and most basic type of organization where the records are placed in the file in the insertion order, so new records are inserted at the end of the file.

Inserting a new record is very efficient

Searching for a record requires linear search

Files of Unordered Records (Heap Files)

Deletion techniques

Rewrite the block - To delete a record,

- a program must first find its block,
- copy the block into a buffer,
- delete the record from the buffer, and
- finally rewrite the block back to the disk.

Files of Unordered Records (Heap Files)

Use deletion marker - Another technique used for record deletion:

- Have an extra byte or bit, called a deletion marker, stored with each record.
- A record is deleted by setting the deletion marker to a certain value.
- A different value for the marker indicates a valid (not deleted) record.

Files of Ordered Records (Sorted Files)

Ordered (sequential) file

- Records sorted by ordering field
  - Called ordering key if ordering field is a key field

Advantages

- Reading records in order of ordering key value is extremely efficient
- Finding next record
- Binary search technique

Files of Ordered Records (Sorted Files)

Average access times for a file of b blocks under basic file organizations

Type of Organization	Access/Search Method	Average Blocks to Access a Specific Record
Heap (unordered)	Sequential scan (linear search)	$b/2$
Ordered	Sequential scan	$b/2$
Ordered	Binary search	$\log_2 b$

Files of Ordered Records (Sorted Files)

Dis advantages

- Inserting and deleting records are expensive operations for an ordered file because the records must remain physically ordered.
- One option for making insertion more efficient is to keep some unused space in each block for new records.
- Another frequently used method is to create a temporary unordered file called an overflow or transaction file. With this technique, the actual ordered file is called the main or master file. New records are inserted at the end of the overflow file rather than in their correct position in the main file.

Summary

This session will give the knowledge about

- Operations on Files
- Files of Unordered Records (Heap Files)
- Files of Ordered Records (Sorted Files)