I B.Tech.

Computer Science & Business Systems

CSE209: Data Structures & Algorithms

Unit – IV: File Organization

By
Dr. S. KAMAKSHI, AP-III / CSE
School of Computing

Agenda

- **Definition of Terms**
- **File Organization**
  - Sequential File Organization
  - Indexed Sequential File Organization
  - Direct / Radom Access File Organization
- **Comparison of File Organizations**

Definition of Terms

- Field:- A field is the smallest unit of data. Eg. in student information file: roll number, age, name of student etc.

- Record: - The collection of fields related to are combined into a record. Eg. An employee has a record with his name, designation, basic pay, allowances, deductions etc. as its fields.

  - A record may have a unique key to identify a record e.g. employee number.

- File: - It is a collection of similar records. The records will have the same fields but different values in each record. The size of a file is limited by the size of memory available.

File Organization

- File organization refers primarily to the logical arrangement of data in a file

- Different methods of organizing files
  - Sequential
  - Indexed-sequential
  - Direct access or Random access

Sequential file organization

- Records are stored in key sequence
- Adding/deleting a record requires making new file
- Records in these files can only be read or written sequentially
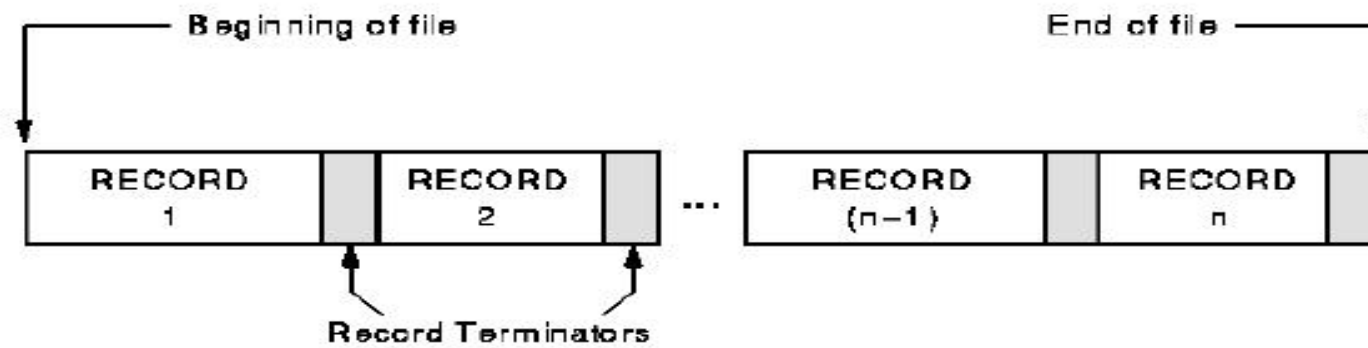
- Records are also in sequence within each block
- To access a record, previous records within the block are scanned
- Thus sequential record design is best suited for applications that process records one record after another
- Records can be read one after another without a search delay
- Records can be added only at the end of the file

ADVANTAGES &DISADVANTAGES

- Advantages:
  - Simple file design
  - Very efficient when most of the records must be processed e.g. Payroll
  - Very efficient if the data has a natural order
  - Can be stored on inexpensive devices like magnetic tape.

- Disadvantages:
  - Entire file must be processed even if a single record is to be searched
  - Overall processing is slow

Indexed Sequential Organization

- Each record of a file has a key field which uniquely identifies that record.
- An index consists of keys and addresses.
- An indexed sequential file is a sequential file (i.e. sorted into order of a key field) which has an index.
- A full index to a file is one in which there is an entry for every record.
- When a record is inserted or deleted in a file the data can be added at any location in the data file. Each index must also be updated to reflect the change. For a simple sequential index this may mean rewriting the index for each insertion.
- Index files:
  - Full Index files – One entry for each record in the main file
  - Partial Index files – One entry for each group of records. Within the group the records are in order

Example – Full Indexing

Index File

Name	Address
Ahalya	1
Arun	4
Balaji	10
Elavarasi	8
Karthick	3
Kiran	2
Manoj	5
Raghavan	6
Ravi	9
Roopa	7

Main File

Address	Name	Department	Year	CGPA
1	Ahalya	ICT	I	8.7642
2	Kiran	ICT	I	7.5648
3	Karthick	CSE	I	7.7932
4	Arun	CSE	I	6.8921
5	Manoj	CSE	I	7.4212
6	Raghavan	IT	I	9.1487
7	Roopa	IT	I	8.2890
8	Elavarasi	CSBS	I	7.8956
9	Ravi	CSBS	I	6.8212
10	Balaji	CSBS	I	9.3241

Example – Partial Indexing

Partial Index File

Department	Address
CSBS	8
CSE	3
ICT	1
IT	6

Main File

Address	Name	Department	Year	CGPA
1	Ahalya	ICT	I	8.7642
2	Kiran	ICT	I	7.5648
3	Karthick	CSE	I	7.7932
4	Arun	CSE	I	6.8921
5	Manoj	CSE	I	7.4212
6	Raghavan	IT	I	9.1487
7	Roopa	IT	I	8.2890
8	Elavarasi	CSBS	I	7.8956
9	Ravi	CSBS	I	6.8212
10	Balaji	CSBS	I	9.3241

ADVANTAGES & DISADVANTAGES

- Advantages
  - Provides flexibility for users who need both type of accesses with the same file.
  - Faster than sequential.

- Disadvantages
  - Extra storage space for the index is required

Direct/Random File Organization

- Records are read directly from or written on to the file
- The records are stored at known address
- Address is calculated by applying a mathematical function (Hashing) to the key field
- A random file would have to be stored on a direct access backing storage medium e.g. magnetic disc, CD, DVD
- Example : Any information retrieval system
  - Train timetable system

ADVANTAGES & DISADVANTAGES

- Advantages
  - Any record can be directly accessed
  - Speed of record processing is very fast
  - Concurrent processing is possible
  - Records need not be sorted
- Disadvantages
  - More complex than sequential
  - Does not fully use memory locations
  - More security and backup problems
  - Expensive hardware and software are required
  - System design is complex and costly
  - File updation is more difficult as compared to sequential files.

Comparison of File Organizations

	Sequential	Direct Access	Indexed Sequential
Type of Access	Batch	Online	Batch or Online
Data Organization	Sequentially by key value	No particular order	Sequentially and By index
Flexibility in Handling Inquires	Low	High	Very High
Speed of Retrieval	Slow	Very fast	Fast
Example	Payroll Processing	Airline Reservation	Customer Ordering