

While inserting a new record the key of the new record is sorted along with the existing records and then inserted. For instance when we try to insert the data record DR4, the sorted insert is depicted in Figure 5



Figure 5 Sort file insert

12.4.3 Advantages and disadvantages of sequential files

In this section we discuss the main advantages and disadvantages of the sequential files.

Advantages

The main advantages of sequential files are as follows:

1. Simple and easy to implement
2. Provides best usage of the storage space.
3. Can be used for data archival needs.
4. The underlying storage systems provide low-cost storage option.
5. Can be stored in inexpensive device like tape device.

Disadvantages

The main disadvantages of the sequential files are as follows:

1. The update operation is costlier and time consuming.
2. The access speed is slow due to the sequential access.
3. It is not possible to directly access a given data record key
4. All records should have uniform structure.
5. The insert time of sorted file is high.
6. Overall processing is slow.

12.4.4 Use cases for sequential files

We generally use sequential files for long-term archival of data and for batch processing operations. We also use sequential access for storing large amount of data that tolerates slow retrieval time.

Check Your Progress – 1

1. A _____ in a file encapsulates the data of a single entity
2. Each data record is uniquely identified by a _____.
3. In _____ files, the data records are inserted in the order of their arrival.
4. For optimizing the performance, we get all the updates into a _____ and later merge it into main file.
5. The two types of sequential files are _____ and _____.
6. Define a data record for a typical university's student enrolment system.

12.5 DIRECT FILE ORGANIZATION

In the direct file organization, we directly access the file by its key. To enable this direct access, we need to map the key to the address where the data record is stored. Direct access is also known as random access. The file records are stored in direct access storage device (DASD) like hard disk, CD, magnetic disk. We randomly place the record throughout the hard disk.

We have depicted the direct file organization in Figure 6

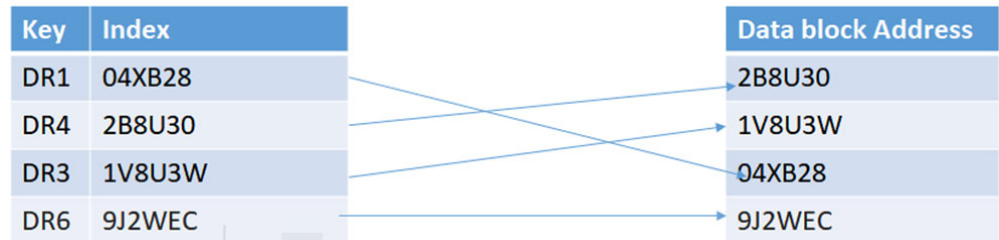


Figure 6 Direct File Organization

We use hash function to convert the key into the address. Naturally direct file access is faster compared to sequential files.

In direct file organization, the records are stored at a known address as depicted in Figure 6.

12.5.1 Hash Function and Collision Resolution

The most popular hash function is the modulo based function that computes the remainder value as the hash value based on the key. The hash value is then used to store the data record.

When there are huge set of keys, multiple keys will end up with same hash value leading to collision. In such cases, we use the collision resolution technique. The separate chaining technique chains the values to the same slot and open addressing technique probes for next available slot for placing the value.

12.5.2 Advantages and disadvantages of Direct file organization

We shall look at the main advantages and disadvantages of the direct file organization.

Advantages

Given below are the main advantages of direct file organization:

1. As we can directly access the data, the retrieval is fast that helps in transaction management systems such as relational database systems.
2. The insert, update and delete operations are faster.
3. The direct file organization is efficient in storage and retrieval of large data.
4. Key-based search is faster.
5. Can be used for real-time transactions that needs optimal performance.
6. Concurrent processing is possible

Disadvantages

Given below are the main disadvantages of direct file organization:

1. The required storage technology is costlier.
2. The usage of storage space is sub optimal.
3. Insert, delete and update needs update of the index table.

12.6 INDEXED SEQUENTIAL FILE ORGANIZATION

Indexed sequential file organization supports direct access of keys and also sequential access by keys.

We use the indexed sequential file organization for hierarchical data organization. For instance, lets consider a use case for retrieving the data record for the states. We use the indexed/direct file for getting the country. The country data record points to the sequential file storing the records of its constituent states. Indexed sequential files can be stored only on random access device such as magnetic disk.

To implement the indexed sequential file, we decouple the index and data files. The index file is structured hierarchically in tree structure whereas the data file stores the information sequentially.

We use two types of indexes static index and dynamic index.

We have depicted the depicted the indexed sequential file in Figure 7.

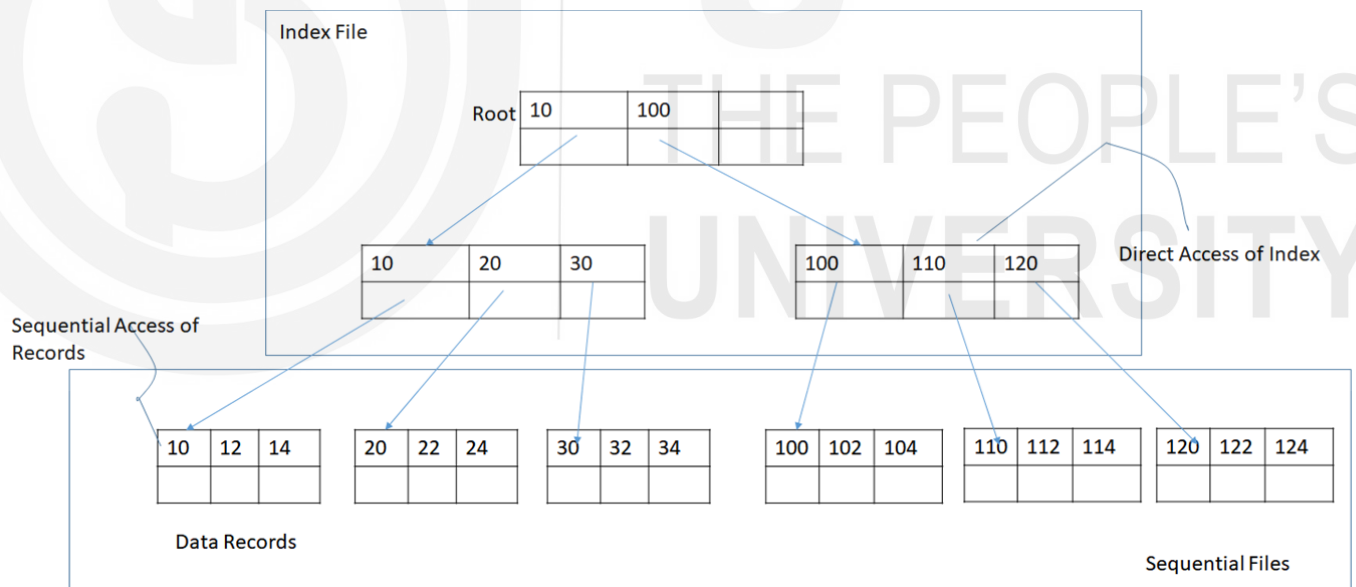


Figure 7 Indexed sequential file

As depicted in Figure 7, the indexed sequential file is depicted as a two-level hierarchy. The first level holds the index to each of the three records in the data file. The second level holds the sequential set of records.

The first level is accessed directly where we get the keys from the index file. In the second level we store the data record sequentially. As the data records are stored sequentially, the key in the first level just points to the first record in the second level from where it can be accessed sequentially.

When we have to insert the record, we update the data file based on the sequence and update the index accordingly. Static index and dynamic index are two main types of indexes. In static indexes the update to the records in the data file does not change the index structure whereas in the dynamic index, the updates to the data file changes the index structure.

12.6.1 Advantages and disadvantages of Indexed sequential file organization

We shall look at the main advantages and disadvantages of the indexed sequential file organization

Advantages

Given below are the main advantages of indexed sequential file organization:

- As the records are directly accessed, the access speed is high
- Record insert is very fast.

Disadvantages

Given below are the main disadvantages of indexed sequential file organization:

- Usage of storage space is sub optimal
- The implementation is costly due to the required of costly hardware
- Extra storage for index file is required.

Check Your Progress – 2

1. In the direct file organization, we directly access the file by its ____
2. The function that converts the key to address is called ____
3. The first level in the indexed sequential access is done ____ and second level is accessed ____
4. The data records in direct file organization are stored in ____ storage device
5. ____ file organization for managing hierarchical data

12.7 SUMMARY

In this unit we mainly discussed about file organization and various types of file organization. A file is a collection of data records. The key operations of the file are File creation, file update, file deletion, file merging, file searching. File organization defines the way the data record is stored and retrieved from the files. We write the data records in a specific sequence in sequential files. There are mainly two types of sequential files - Pile Files and sort files. In Pile files, the data records are inserted in the order of their arrival. In case of sort files, the data records are sorted based on the key and are inserted into the file. In the direct file organization, we directly access the file by its key. Indexed sequential file organization supports direct access of keys and also sequential access by keys.

12.8 SOLUTIONS/ANSWERS

Check Your Progress – 1

1. Data record.

2. Key
3. Pile files
4. Transaction log
5. Pile files and sort files
6. A sample student enrollment structure is as follows:
Student Number, Student Name, Date of Enrollment, Address, Programme

Check Your Progress – 2

1. Key
2. Hash function
3. Directly and sequentially
4. Direct access
5. Indexed sequential

12.9 FURTHER READINGS

References

- https://en.wikipedia.org/wiki/File_system
- https://en.wikipedia.org/wiki/Sequential_access
- https://en.wikipedia.org/wiki/Indexed_file
- <https://en.wikipedia.org/wiki/ISAM>

