

File organisation and Access

- File organisation refers to the way data is stored in a file.

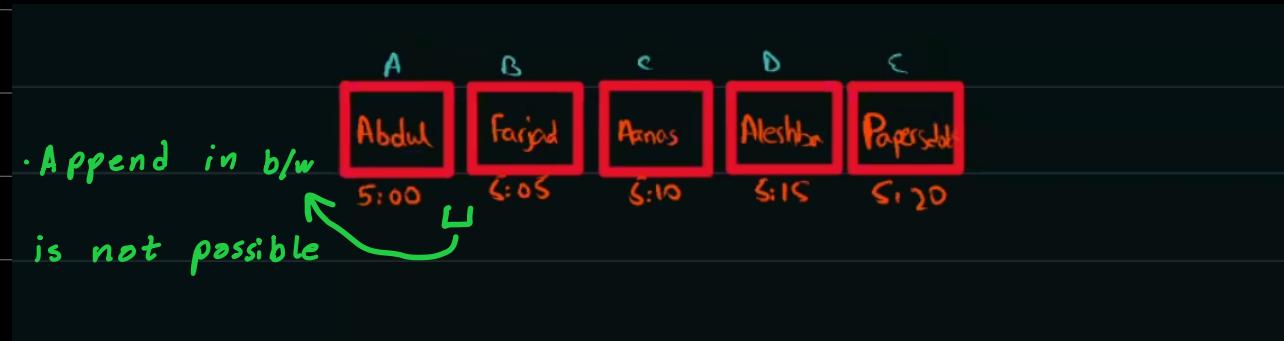
Organisation

① Serial

② Sequential

③ Random

Serial



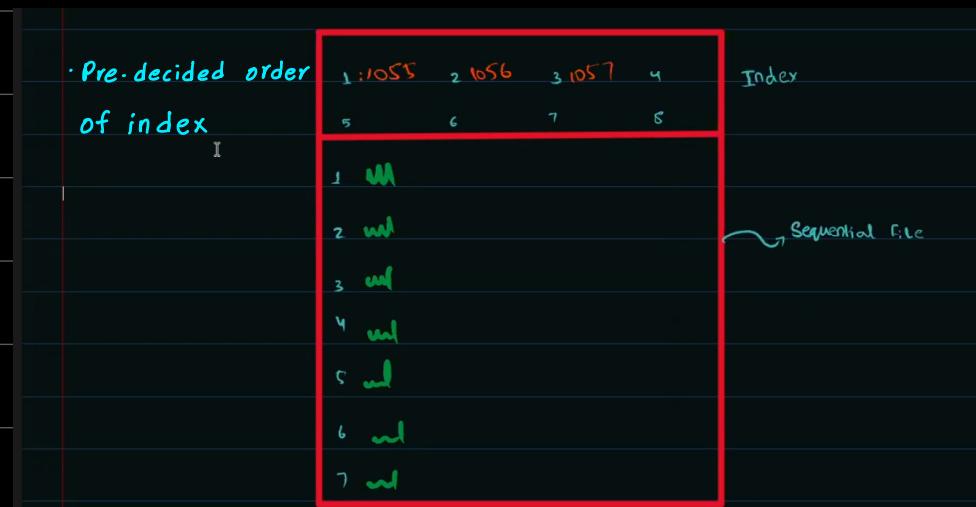
- Data is stored in chronological order (arranged in order of time)
- Easy to add/append data to end of file
- Allows the data to be read in order it is taken

- No key fields need to be used. (key field is a unique number which is used to access data directly)

Q- When will we use serial file organisation method?

- Chronological order matters
- Easier to append data at the end of a file.
- Re-organisation and re-sorting is not required
- Small file, so easy to search.

Sequential

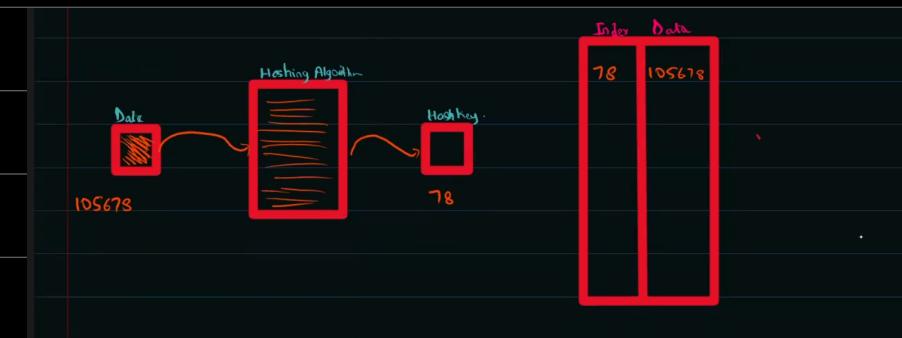


Q- Where will we use sequential file organisation method?

- Unique field (so we could use it for indexing)
- Sorted in a predecided order

Random

Hashing Algorithm: It is basically a mathematical function to find a hash key to access data



Hashing Algorithm

① Folding

② Truncation

③ Modular Div

Folding

* Not in syllabus, just added for understanding

- All Hashing algorithms
are applied on data



Truncation

- Take alternate numbers

0 5 4 1 9 7 2 2
0 4 9 2
09
Hash key.

Modular Division

- Data = 42
- MOD = 5
- hash key = 42 MOD 5

$$5 \overline{)42} \quad \begin{array}{r} 8 \\ -40 \\ \hline 2 \end{array}$$

2 → hash key

Lookup file: File used to find data

4 A bank has 95000 customers. Each customer has a unique ID. When a customer uses an Automated Teller Machine (ATM) to obtain cash, their current balance is checked. The balance is stored in a file which has the following fields:

- the customer ID (6-digit number in the range 100000 to 689999)
$$\frac{100000}{100000} \overline{)689999} \quad \begin{array}{r} 6 \\ -60 \\ \hline 89 \\ -80 \\ \hline 99 \\ -90 \\ \hline 9 \end{array}$$

The file can store a maximum of 100000 records.

(a) Give a reason why a random organisation would be appropriate for this file.
spread of access increases.
used as a lookup file. [1]

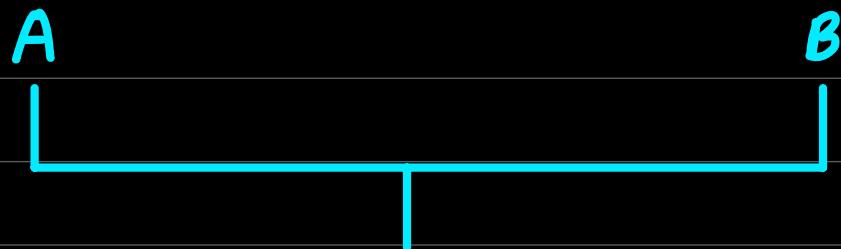
(b) An algorithm for inserting a new record in this file uses the following hash function:
 $\text{RecordKey} \leftarrow \text{CustomerID} \text{ MOD } 100000$
where RecordKey is the record position in the file.

(i) Complete the table to show the values generated by the hash function for the given customer IDs.

CustomerID	RecordKey
802139	2139
700004	4
689998	89998
102139	2139

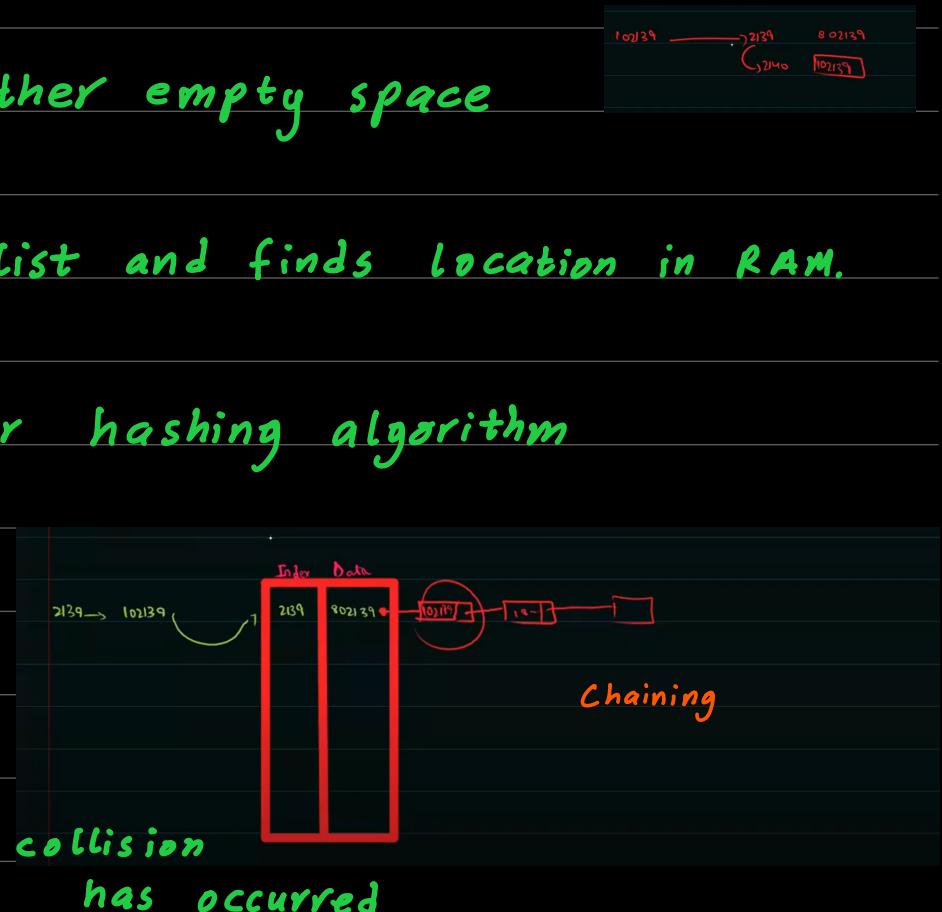
(ii) State the range of possible values for RecordKey.
Minimum value of RecordKey: 0
Maximum value of RecordKey: 99999 [2]

Collisions in Hashing Algorithm



Same Hash key

- We have three methods to deal with hashing collision
- Bucketing (closed hash): Find another empty space
- Chaining (open hash): Uses linked list and finds location in RAM.
- Rehashing (closed hash): Use another hashing algorithm
- Data is stored in form of linked list which is stored in RAM.
- Pointers are used to link linked lists to the memory location where hashing collision has occurred



Q- Describe Random file organisation

- Record location is calculated
- Using a hashing algorithm on a key field
- If a record can not be stored
- Then subsequent location is searched known as closed hash
- or an overflow area is searched known as open hash

Sequential and Direct Access

- Sequential: Particular sequence
- Direct: By using key, you can access data directly.

Q- Explain how sequential access is used?

- Earliest reading is accessed first
- and each successive reading is read
- until the final reading is read.

