Database Access Methods

File Organization

Outline

- Heap file
- Sequential file
- Indexed sequential file
- Index file
- Hashed file

Heap file

- Variable-length records ~ file is not homogeneous
- Data not sorted in any way, a record placed always at the end of the file
- Usually used only along with another supporting structure

- Insert : O(1)
- Find : O(N / b)

Unsorted sequential file

- Fixed size record
- Data not sorted in any way
- Suitable when data are collected without any relationship to other data

Block	Name	Department	…
0	Galvin Janice	Purchasing	
	Walters Rob	Marketing	
	Brown Kevin	Marketing	
1	Walters Rob	Developlment	
	Duffy Terri	Research	
	Brown Kevin	PR	
2	Duffy Terri	Developlment	
	Walters David	Production	
	Brown Kevin	Purchasing	
3	Matthew Gigi	Purchasing	
	Walters Rob	PR	
	…	…	

- Insert : O(1)
- Find : O(N / b)

Sorted sequential file

- Fixed size record
- Records sorted in the file on the primary search key

Fetch

- Sequential scan
  - O (N / b)
- Binary search
  - O (log (N / b))

Block	Name	Department	…
0	Brown Kevin	PR	
	Brown Kevin	Purchasing	
1	Brown Kevin	Marketing	
	Duffy Terri	Developlment	
2	Duffy Terri	Research	
	Galvin Janice	Purchasing	
3	Matthew Gigi	Purchasing	
	Walters David	Production	
4	Walters Rob	Marketing	
	Walters Rob	Developlment	
5	Walters Rob	PR	
	…	…	

Sorted sequential file - modification

Insert

- Inserting new record into the structure would be costly since all the following records would have to be shifted
- Auxiliary file/blocks called overflow file/bucket needs to be established where the new records are inserted
- The file is periodically reorganized

Update

- Simple if the update does not include the primary search key

Delete

- Deleted records are not directly removed since reorganization would have to take place
- A bit designating deleted records is set
- Deleted records are removed during periodical reorganization

Index

- An index is an auxiliary structure for a data file that consists of a specifically arranged structure containing key-pointer pairs
- Storage of the index
  - Main memory
  - Secondary memory
    Accessing index must also be taken into account when computing the find/fetch time

Indexed sequential file

- Fixed size record
- Structure
  - Primary file/area
  - Index/secondary file/area
  - Overflow file/area
- Data can be accessed either using the index or sequentially

Index file
2. (top) level

Block	Name	
8	Brown	6
	Walters Rob	7

Index file
1. (base) level

Block	Name	
6	Brown	0
	Clinard	1
	Duffy	2
	Leavy	3
	Peagler	4
7	Walters Rob	5
	…	

Primary file

Block	Name	…
0	Brown Kevin	
	Berkman Doloris	
1	Clinard Stephnie	
	Coolidge Emily	
2	Duffy Terri	
	Galvin Janice	
3	Leavy Shirleen	
	Matthew Gigi	
4	Peagler David	
	Shackelford Elsie	
5	Walters Rob	

Indexed sequential file - fetch

Searching for a specific value (query key)

- Check the top level of the index and identify a key-value pair with the highest value lower than the query key
- Fetch the block referenced by the value
- Repeat the previous steps with lower index levels until a primary file block is reached
- Search the primary file block for the specified key

Searching for a range of values

- Search for the lower bound key of the interval
- Sequentially scan the blocks of the primary file until record corresponding to the upper bound key is found

Indexed sequential file - fetch

Index file
2. (top) level

Block	Name	
8	Brown	6
	Walters Rob	7

Index file
1. (base) level

Block	Name	
6	Brown	0
	Clinard	1
	Duffy	2
	Leavy	3
	Peagler	4
7	Walters Rob	5
	…	

Primary file

Block	Name	…
0	Brown Kevin	
	Berkman Doloris	
1	Clinard Stephnie	
	Coolidge Emily	
2	Duffy Terri	
	Galvin Janice	
3	Leavy Shirleen	
	Matthew Gigi	
4	Peagler David	
	Shackelford Elsie	
5	Walters Rob	

Indexed sequential file - insert

- Index structure stays static when inserting data
- New records need to be stored in reserved areas (pockets) within the primary file
- Overflown data are inserted into a new block (created dynamically) pointed to by the overflown block
- Buckets can be chained and therefore theoretically the ISF does not need to be rebuilt
- Pointers to the overflow area
  - with each record
  - with each block

Indexed sequential file

Pros

- Fast access using primary search key
- Shares pros of the sequential file

Cons

- Fast access only when using primary search key
- Problems with primary file when updating

Indexed file organization

- Allow search the file according to different attributes without the need to sequentially scan the whole file
- The primary file stays unsorted or is sorted according to one key only (primary index)
- For each query key an index file can be built
  → one primary data file, multiple index files
- Basically corresponds to a standard DB table where we have one table and multiple indexes built over it (possibly of different types).

Index

Primary index

- Index over the attribute based on which the records in the primary file are sorted
- If the value of the primary attribute is modified the file needs to be reorganized → should be relatively invariable
- Well-suited for range queries
- There does not have to be a primary index in the IF

Secondary index

- There can have multiple secondary indexes
- Range queries for long ranges can be very expensive (an extreme example is a sequential scan based on a secondary index which can lead to an extremely deteriorated performance)

Index

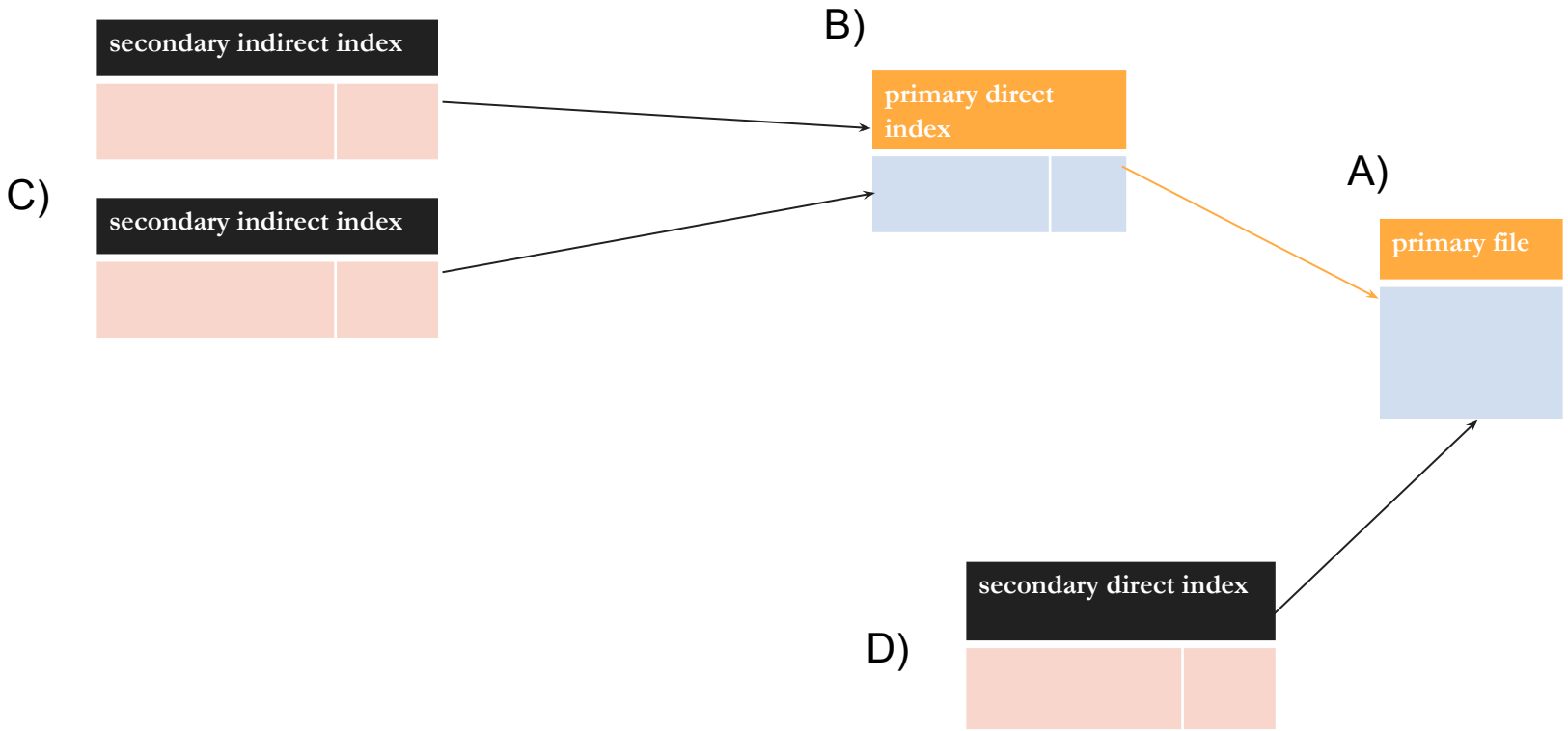
Direct index

- Index is bound directly to the record
- Primary file reorganization leads to modification the indexing structures

Indirect index

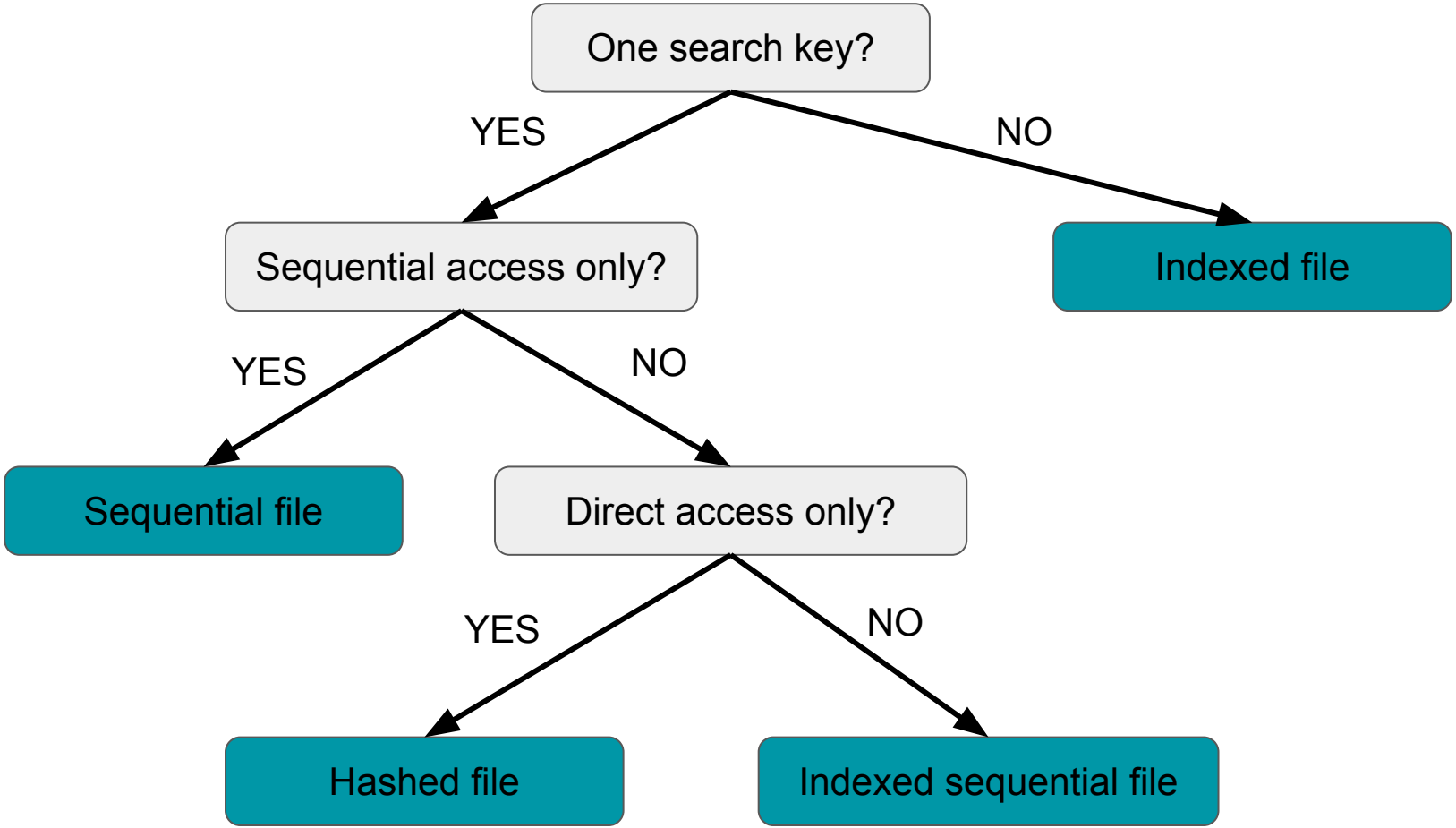
- Indirect indexes contain keys of the primary index and not pointers to the primary file
- Accessing a record needs one more accesses to the primary index
- If the primary file is reorganized, the secondary indexes stay intact

Index


secondary indirect index
B)
primary direct index
C)
secondary indirect index
A)
primary file
secondary direct index
D)

Hashed file organization

- Direct access with one unique key
- Use hash function to map records to pages/blocks addresses
- If the data can not fit into a page/block when inserting an overflow strategy is employed
- Placement within the page is not specified
- When file is being reorganized, the pages are not filled only to, e.g., 80%


One search key?
YES
NO
Sequential access only?
Indexed file
YES
NO
Sequential file
Direct access only?
YES
NO
Hashed file
Indexed sequential file