INDEX STRUCTURE, QUERY PROCESSING

UNIT-4

Reference : Database Concepts, Korth, Silbertz
Fundamentals of Database System, Elmasri, Navathe

File Organization

A file organization is a technique for physically arranging the records of a file on secondary storage devices. When choosing the file organization, we should consider the following factors:

- ✓ ***Speed of data retrieval***
- ✓ ***Speed of processing data***
- ✓ ***Speed of update operations***
- ✓ ***Storage space usage efficiency***
- ✓ ***Failure and data loss protection***
- ✓ ***Reorganization needs(frequency)***
- ✓ ***Scalability (capability to grow, as more records are added to the file)***
- ✓ ***Security***

Sequential File Organization

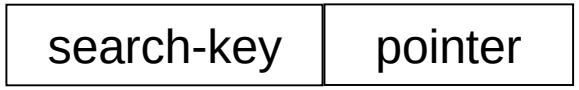
In a Sequential File Organization the records in the file are stored in a sequence according to the Primary key value. To locate a particular record, a program must scan the file from the beginning until the desired record is located.

In an Indexed File Organization, *the records are stored either sequentially or non-sequentially and an index is created that allows the applications to locate the individual records using the index.*

In the Indexed file organization, if the records are stored sequentially based on the primary key value then the file organization ***is called an indexed sequential organization.***

Basic Concepts

- Indexing mechanisms used to speed up access to desired data.
  - E.g., author catalog in library
- **Search Key** - attribute to set of attributes used to look up records in a file.
- An **index file** consists of records (called **index entries**) of the form



- Index files are typically much smaller than the original file
- Two basic kinds of indices:
  - **Ordered indices:** search keys are stored in sorted order
  - **Hash indices:** search keys are distributed uniformly across “buckets” using a “hash function”.

Index Evaluation Metrics

- Access types supported efficiently. E.g.,
  - records with a specified value in the attribute
  - or records with an attribute value falling in a specified range of values
- Access time
- Insertion time
- Deletion time
- Space overhead

Types of Indexing

*** Single –Level Indexing and *Multi- Level Indexing**

1. **Primary index**

2. **Clustered Index**

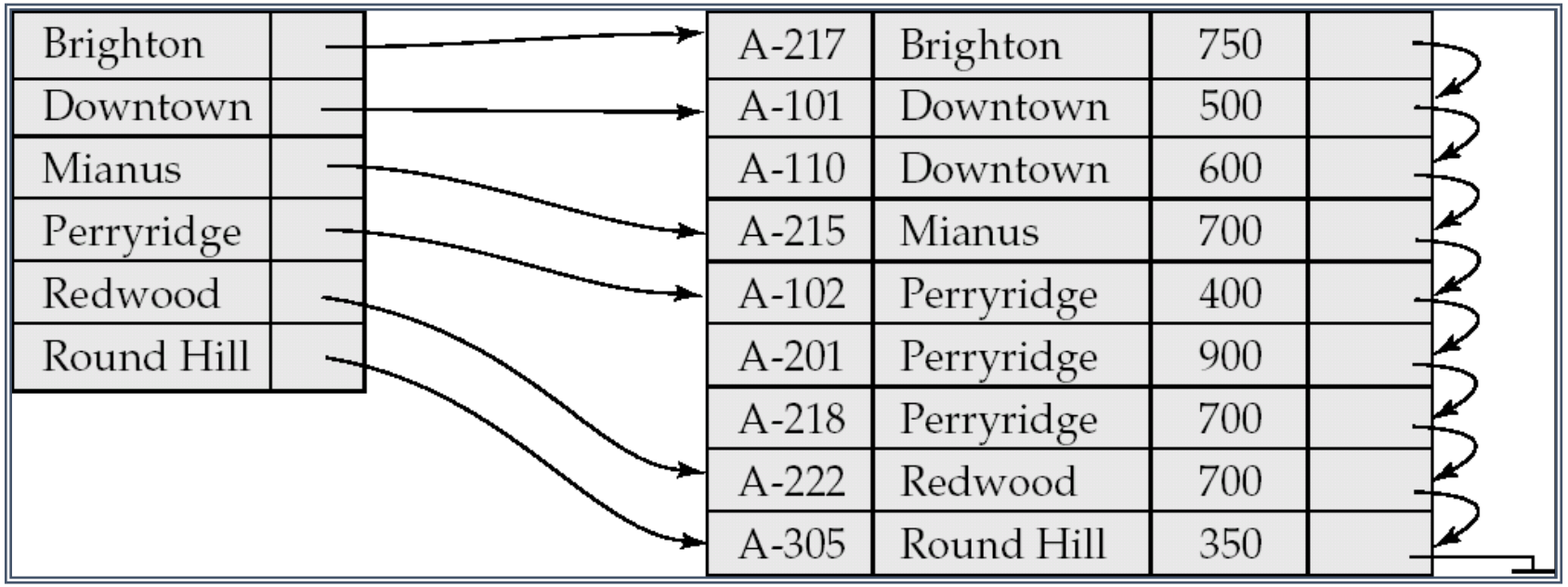
3. **Secondary index**

Types of Indexing contd…

- In an **ordered index,** index entries are ***stored***, ***sorted*** on the search key value. E.g., author catalog in library.
- **Primary index:** in a sequentially ordered file, the index whose search key specifies the sequential order of the file.
  - Also called **clustering index**
  - The search key of a primary index is usually but not necessarily the primary key.
- **Secondary index**: an index whose search key specifies an order different from the sequential order of the file. Also called non-clustering index**.**
- **Index-sequential file:** ordered sequential file with a primary index**.**

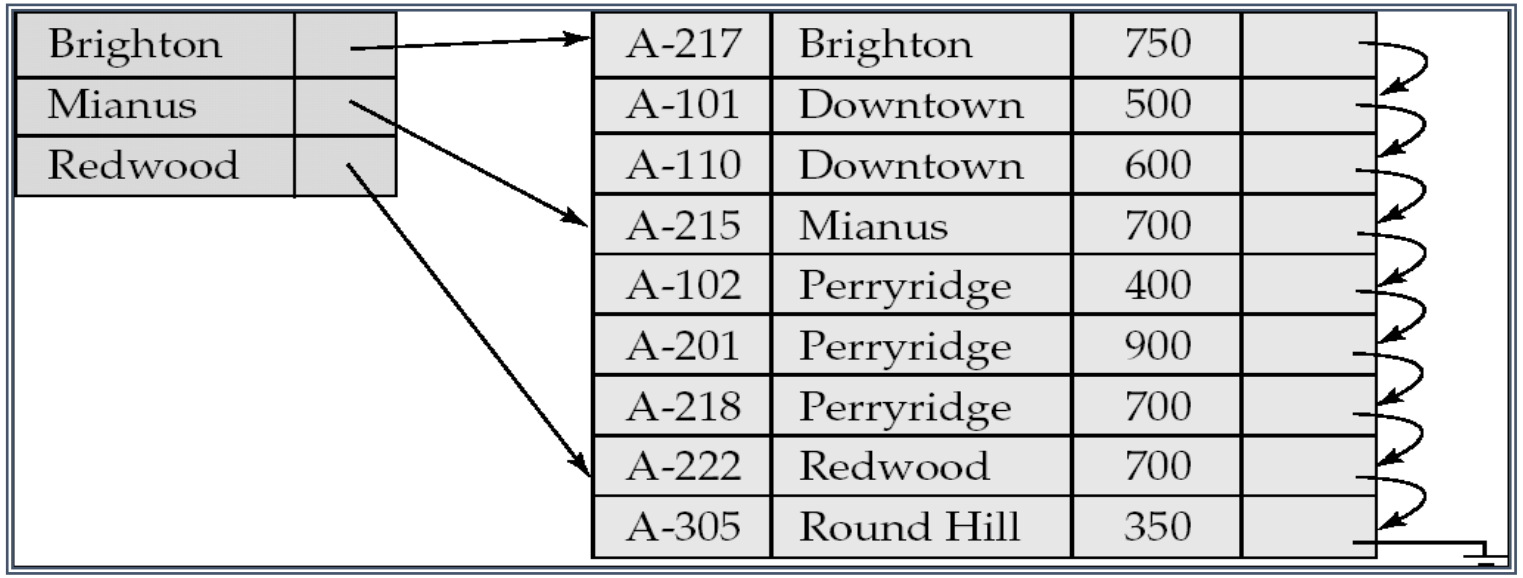
Dense Index Files

- Dense index — Index record appears for every search-key value in the file.



Sparse Index Files

- **Sparse Index:** contains index records for only some search-key values.
  - Applicable when records are sequentially ordered on search-key
- To locate a record with search-key value *K* we:
  - Find index record with largest search-key value < *K*
  - Search file sequentially starting at the record to which the index record points



Sparse Index Files (Cont.)

- **Compared to dense indices:**
  - Less space and less maintenance overhead for insertions and deletions.
  - Generally slower than dense index for locating records.

- **Good tradeoff**: sparse index with an index entry for every block in file, corresponding to least search-key value in the block.