Chapter 6: Physical Database Design and Performance

Slides:

Denormalization

- Transforming ***normalized*** relations into ***unnormalized*** physical record specifications
- Benefits:
  - Can improve performance (speed) by reducing number of table lookups (i.e. *reduce number of necessary join queries*)
- Costs (due to data duplication)
  - Wasted storage space
  - Data integrity/consistency threats
- Common denormalization opportunities
  - One-to-one relationship (Fig. 6-3)
  - Many-to-many relationship with attributes (Fig. 6-4)
  - Reference data (1:N relationship where 1-side has data not used in any other relationship) (Fig. 6-5)

Chapter 6 16

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Denormalization opportunities

- 1:1 relationship
- M:M associative entity with non-key attributes
- reference data

Chapter 6: Physical Database Design and Performance

Slides:

Partitioning

- Horizontal Partitioning: Distributing the rows of a table into several separate files
  - Useful for situations where different users need access to different rows
  - Three types: Key Range Partitioning, Hash Partitioning, or Composite Partitioning
- Vertical Partitioning: Distributing the columns of a table into several separate relations
  - Useful for situations where different users need access to different columns
  - The primary key must be repeated in each file
- Combinations of Horizontal and Vertical

Partitions often correspond with User Schemas (user views)

Chapter 6 20

Chapter 6: Physical Database Design and Performance

Slides:

Partitioning (cont.)

- Advantages of Partitioning:
  - Efficiency: Records used together are grouped together
  - Local optimization: Each partition can be optimized for performance
  - Security, recovery
  - Load balancing: Partitions stored on different disks, reduces contention
  - Take advantage of parallel processing capability
- Disadvantages of Partitioning:
  - Inconsistent access speed: Slow retrievals across partitions
  - Complexity: Non-transparent partitioning
  - Extra space or update time: Duplicate data; access from multiple partitions

Chapter 6: Physical Database Design and Performance

Slides:

Data Replication

- Purposely storing the same data in multiple locations of the database
- Improves performance by allowing multiple users to access the same data at the same time with minimum contention
- Sacrifices data integrity due to data duplication
- Best for data that is not updated often

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Physical Files

- **Physical File: A file as stored on disk**
- **Constructs to link two pieces of data:**
  - **Sequential storage**
  - **Pointers**
- **File Organization: How the files are arranged on the disk (more on this later)**
- **Access Method: How the data can be retrieved based on the file organization**
  - **Relative - data accessed as an offset from the most recently referenced point in secondary memory**
  - **Direct - data accessed as a result of a calculation to generate the beginning address of a record**

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

File Organizations

- A technique for physically arranging the records of a file on secondary storage devices.
- Goals in selecting: (trade-offs exist, of course)
  - Fast data retrieval
  - High throughput for input and maintenance
  - Efficient use of storage space
  - Protection from failures or data loss
  - Minimal need for reorganization
  - Accommodation for growth
  - Security from unauthorized use

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

File Organizations

- Sequential
- Indexed
  - Indexed Sequential
  - Indexed Nonsequential
- Hashed (also called *Direct*)
- See Table 6-3 for comparison

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Sequential File Organization

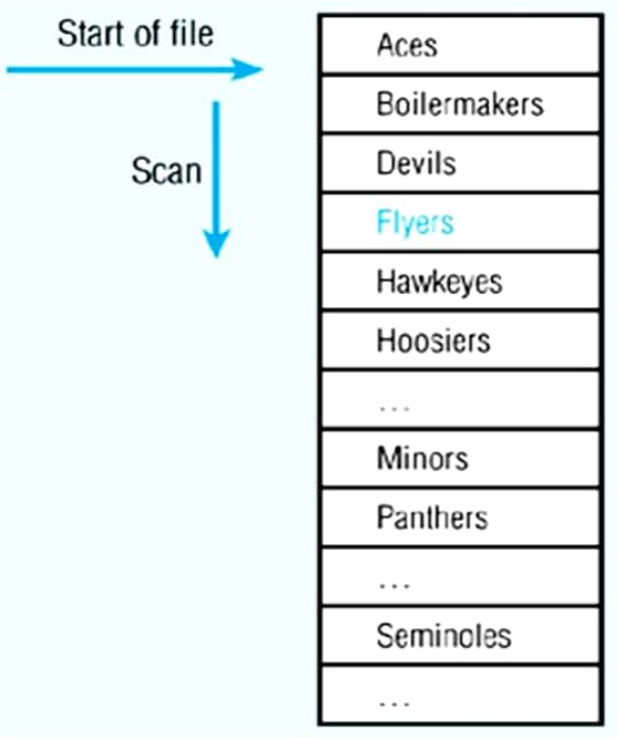
- Records of the file are stored in sequence by the primary key field values

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Comparisons of file organizations:

(a) Sequential



Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Sequential Retrieval

- Consider a file of 10,000 records each occupying 1 page
- Queries that require processing all records will require 10,000 accesses
- e.g., Find all items of type 'E'
- Many disk accesses are wasted if few records meet the condition
- However, very effective if most or all records will be accessed (e.g., payroll)

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

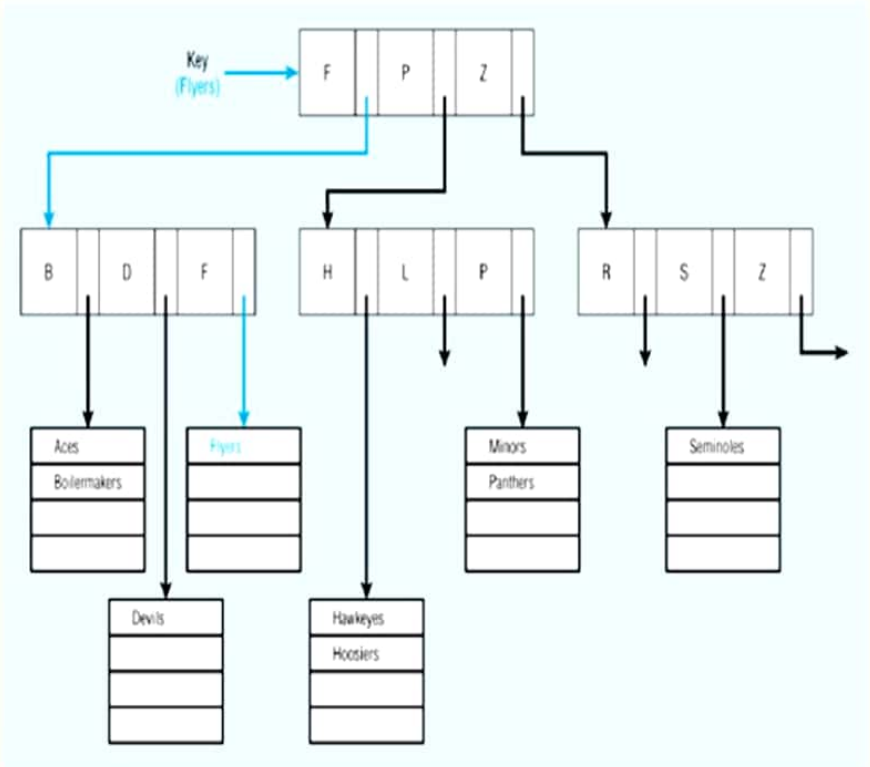
Indexed File Organization

- Index concept is like index in a book
- Indexed-sequential file organization: The records are stored sequentially by primary key values and there is an index built on the primary key field (and possibly indexes built on other fields, also)

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

(b) Indexed



Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Hashed File Organization

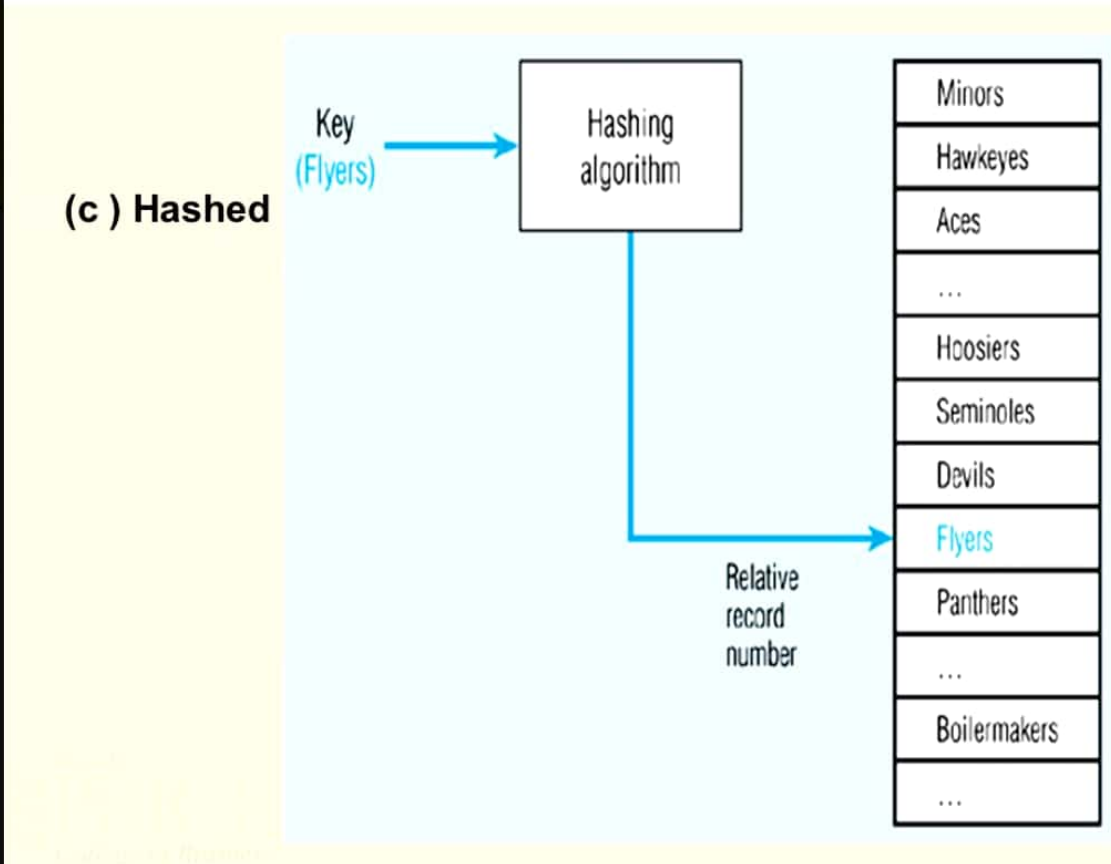
- Hashing Algorithm: Converts a primary key value into a record address
- Division-remainder method is common hashing algorithm

(more to come on this)

Chapter 6 Physical Database Design.

Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:



Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Hashing

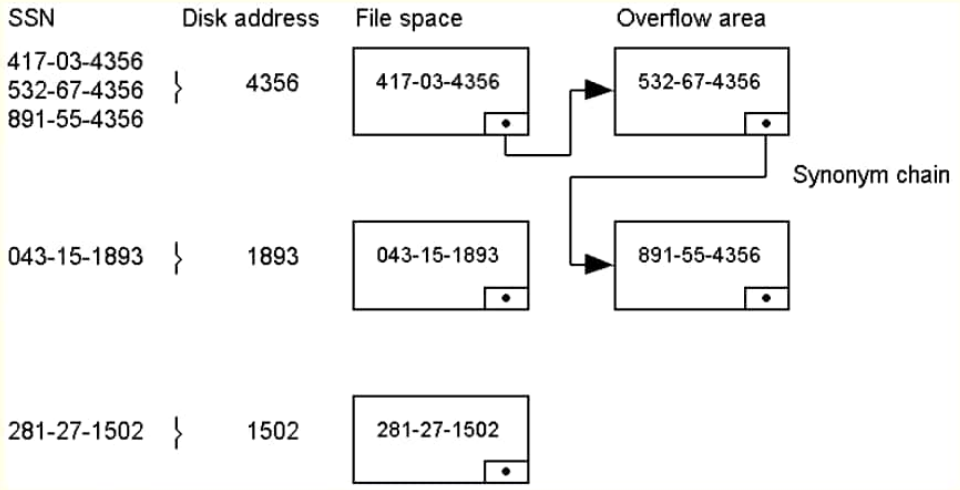
- A technique for reducing disk accesses for direct access
- Avoids an index
- Number of accesses per record can be close to one
- The hash field is converted to a hash address by a hash function

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Hashing

- **hash address = remainder after dividing SSN by 10000**



Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Shortcomings of Hashing

- Different hash fields may convert to the same hash address
  - these are called Synonyms
  - store the colliding record in an overflow area
- Long synonym chains degrade performance
- There can be only one hash field per record
- The file can no longer be processed sequentially

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Indexing

- An index is a table file that is used to determine the location of rows in another file that satisfy some condition

ITEMTYPE INDEX	
ITEMTYPE	
C	
C	
C	
C	
E	
E	
F	
N	
N	
N	

ITEM			
ITEMNO	ITEMNAME	ITEMTYPE	ITEMCOLOR
1	Pocket knife - Nile	E	Brown
2	Pocket knife - Thames	E	Brown
3	Compass	N	-
4	Geo positioning system	N	-
5	Map measure	N	-
6	Hat - polar explorer	C	Red
7	Hat - polar explorer	C	White
8	Boots - snakeproof	C	Green
9	Boots - snakeproof	C	Black
10	Safari chair	F	Khaki

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Maintaining an Index

- Adding a record requires at least two disk accesses:
  - Update the file
  - Update the index
- Trade-off:
  - Faster queries
  - Slower maintenance (additions, deletions, and updates of records)
  - Thus, more *static* databases benefit more overall

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Querying with an Index

- Read the index into memory
- Search the index to find records meeting the condition
- Access only those records containing required data

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Rules of Thumb for Using Indexes

1. Indexes are most useful on larger tables
2. Index the primary key of each table (may be automatic, as in Access)
3. Indexes are useful on search fields (WHERE)
4. Indexes are also useful on fields used for sorting (ORDER BY) and categorizing (GROUP BY)
5. Most useful to index on a field when there are many different values for that field

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Rules of Thumb for Using Indexes

6. Find out the limits placed on indexing by your DBMS (Access allows 32 indexes per table, and no index may contain more than 10 fields)
7. Depending on the DBMS, null values may not be referenced from an index (thus, rows with a null value in the field that is indexed may not be found by a search using the index)

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

Rules for Adding Derived Columns

- Use when aggregate values are regularly retrieved.
- Use when aggregate values are costly to calculate.
- Permit updating only of source data.
- Create triggers to cascade changes from source data.

Chapter 6 Physical Database Design. Physical Design u The purpose of the physical design process is to translate the logical description of the data into.

Slides:

One Other Rule of Thumb for Increasing Performance

- Consider contriving a shorter field or selecting another candidate key to substitute for a long, multi-field primary key (and all associated foreign keys)