Database Access Methods

File Organization Theory

Motivation

- How can we organize data records?
- Not now, let’s talk some theory first ...

Outline

- Terminology

Terminology

Data record

- Representation of an application object
- Set of fields
- Attribute is field with domain (type).
- Key

File

- Named collection of records

Database

- Collection of related data (named files) in secondary memory

Data record

Logical

- Attribute set

Physical

- Physical representation of a logical record of size R (bytes) on the medium
- Contains additional metadata (such as record delimiters, etc.)
- Records are stored in blocks of size B (bytes)

Data record

Fixed length

- File header contains number of records and length of each field
- Record can be accessed using the record number

Variable length

- Variable length of the attributes
- When similar objects relate to a nonuniform set of data
- Optional attributes
- Attributes holding records
- Workaround is to set maximum length for a given field

Record blocking

Blocking factor

- Number of record in block
- $b = \lfloor B/R \rfloor$
  - B = block size
  - R = record size

Basic division based on blocking

- non-blocked records
  1 record fits 1 block
- blocked records
  N records fit 1 block
- overflown records
  1 record fits N blocks

Record blocking

Fixed blocking

- Fixed-length records
- Possible internal fragmentation

Variable-length spanned blocking

- Variable-length records
- Record can span multiple blocks
- Hard to implement
- Need more time to read records in 2 blocks

Variable-length unspanned blocking

- Variable-length records
- No spanning
- Possible high internal fragmentation

Files

- Stored in secondary memory
- Record identified using file keys (sets of attributes)
- One file key is primary key - should be artificial

Homogeneous

- Store fixed size records of same type

Non-Homogenous

- Either variable size of different type

Files - Operations

Modification

- Insert
- Update
- Delete

Querying

- Find
  - Finds a record within the file
- Fetch
  - Gets a record into main memory

Formation/Termination

- Create/Remove

Maintenance

- Reorganize/Rebuild

Querying files

One-dimensional queries

- cars with age > 35
- cars with color = ‘red’

Multi-dimensional queries

- Total match
  - age = 12 & color = ‘red’
- Partial match
  - age = 12
- Total interval match
  - 12 < age < 25 & color in (‘red’, blue’)
- Partial interval match
  - 12 < age < 25

File organization

- How to organize a set of records in a file and how to access them
- The description of the **logical memory structure** together with **algorithms** for handling that structure
- Can contain multiple files
- Optimal choice of organization depends on the usage

File organization - Levels

Logical schema

- Algorithms
- Logical blocks (pages)
  - Logical blocks structure
  - Logical blocks relations
  - Logical blocks manipulation
- Logical files
  - Primary file
  - Auxiliary files

Physical schema

- Mapping between logical schema and physical pages
- One logical file can span multiple physical files and the other way around

Implementation schema

- Implementation of the physical files shielded from the logical level by OS