CS522 Advanced Database Systems
Data Storage

Chengyu Sun
California State University, Los Angeles

The Megatron 3000 Approach

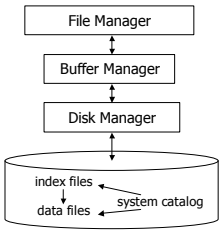
- Text file
- One tuple per line
- Fields are separated by #

e.g. /usr/db/students

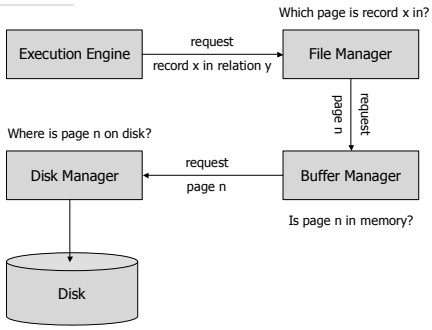
```
Smith#123#CS
Jones#522#EE
...
```

The Real DBMS Approach

- File Manager
- Buffer Manager
- Disk Manager



Record Request



Why So Many "Managers"?

- Isn't the OS supposed to handle file I/O??
- Don't we have *both* OS and disk buffering already??

Disk Manager

- Abstracts storage as a sequence of pages
  - allocate, de-allocate
  - read, write
- Keeping track of free disk blocks
  - Linked list
  - Bitmap directory

Buffer Manager

- Buffer pool
  - pages in the buffer pool are called frames
- Replacement policy
- Additional information *per frame*
  - pin-count
  - dirty

How Buffer Manager Works

- If page in buffer pool, increment pin-count of the frame; otherwise
  - Choose a frame for replacement
  - Write out the frame if it's "dirty"
  - Read in frame
  - Increment pin-count of the frame
- Return frame

A Few Finer Points

- When do we increment/decrement pin-count??
- What if all frames are pinned??
- Concurrency and recovery??

Replacement Polices

- Optimal??
- FIFO
- Random
- Least Recent Used (LRU)
- Most Recent Used (MRU)
  - Sequential flooding
- Clock

LRU Replacement Implementation

- A queue of pointers to frames with 0 pin count
  - Add a pointer to a frame to the queue when the pin count of the queue reaches 0

Clock Replacement Implementation

- Current frame
- Per frame
  - pin-count
  - referenced

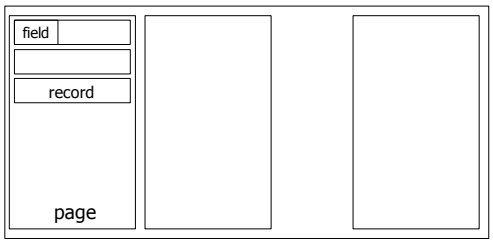
More Replacement Strategies

- Buffer pool partition
  - bind a partition to a database, or relation, or index
  - apply different replacement strategies
- Considering more parameters
  - page type
- Higher level controls
  - Love/hate
  - Buffer reservation

Project 1 and Accounts

- CS Server accounts
- Turnin Server accounts
- Minibase discussion

Data Organization



"File"

Heap File

- Unsorted
  - why??
- Operations
  - *create* and *delete* files
  - iterate through all records (*scan*)
  - *insert* a record
  - *get*, *update*, and *delete* a record with a given id

Fields for Common SQL Types

- char(n), varchar(n), bit(n)
- int, float, decimal(n,m)
- date, time
- text, BLOB

Fixed-Length Records

- Fixed number of fields
- Fixed length of each field

System Catalog

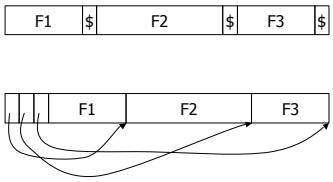
- For each relation
  - name, file name, file structure
  - attribute name and type
  - index name
  - constraints
- Index and view information
- Accounts and privileges
- Statistics

System Catalog as a Collection of Relations

SC_Attributes

attr_name	rel_name	type	position
attr_name	SC_Attributes	String	1
rel_name	SC_Attributes	String	2
type	SC_Attributes	String	3
postion	SC_Attrributes	Integer	4
sid	Students	Integer	1
sname	Students	String	2
gpa	Students	Real	3
cid	Courses	String	1
cname	Courses	String	2

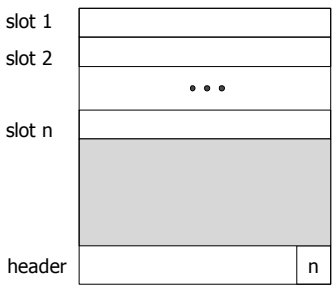
Variable-Length Records



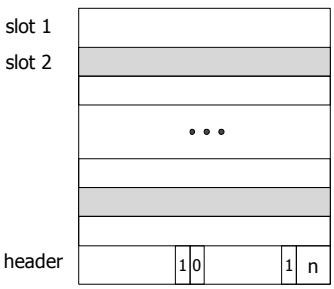
Organize Records into Pages

- Record ID (rid)
  - <page id, *slot* number>
- Concerns
  - insert, delete, update, and search records in a page

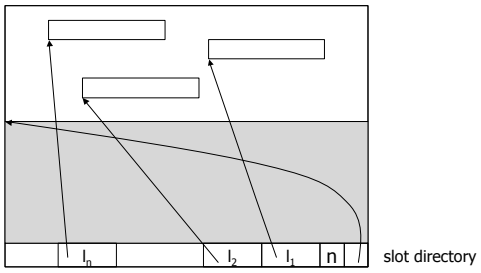
Fixed-length Records – Packed Page



Fixed-length Records – Unpacked Page



Variable-Length Records – Slotted Page



(record offset, record length)

About Slotted Page

- Delete, insert, update, search??
- Does the <page id, slot number> scheme still work?

From Pages to Files

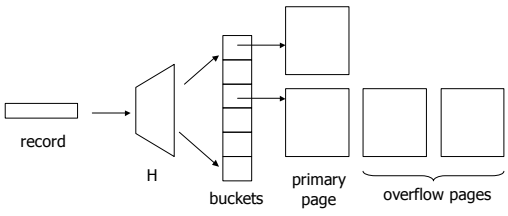
- ??

Other File Organizations – Sorted File

- Sorted file (Sequential file)
  - Key
    - multi-field key??
  - Usually *packed*
- Maintenance – *almost sorted*
  - Sliding
  - Overflow page
  - Sequential page

Other File Organizations – Hashed File

- Hashed file
  - H(key) → primary page
  - overflow page

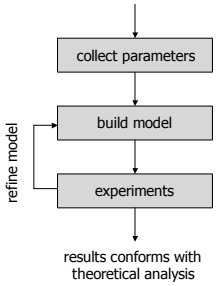


A Simple Cost Model

- The model
  - **B** data pages
  - **R** records per page
  - Avg time to read or write a disk page: **D**
  - Avg time to process a record: **C**
  - Time to compute the hash value of a record: **H**
- The method – the complexity of a DB algorithm can usually be *estimated* by its number of disk page accesses
- The problem(s)??

R&D Methodology

- Simplified model
- Extensive experimentation



Three File Organizations

- Heap file
- Sorted file
- Hashed file

Five Operations

- Scan
- Equality selection
- Range selection
- Insert
- Delete

Cost Estimations

	Heap	Sorted	Hashed
Scan	BD	BD	BD
E. Selection	0.5BD	DlogB	D
R. Selection	BD	DlogB	BD
Insert	2D	Search + BD	2D
Delete	Search+D	Search + BD	Search + D

Readings

- Stanford book: Chapter 12
- [Wisconsin book: Chapter 7, 8]
- [PAX paper]