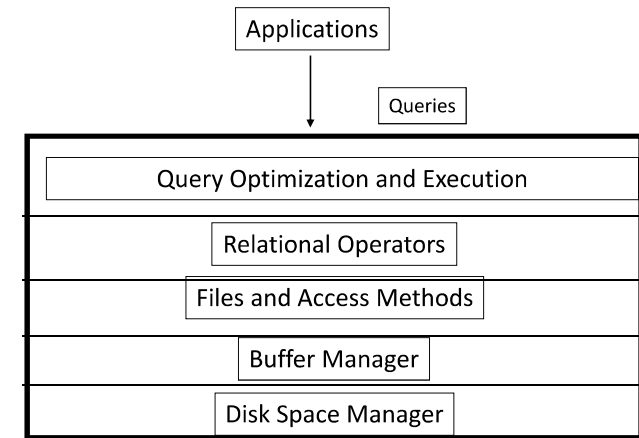


# Review: Architecture of a DBMS



## HW6: Sort-Merge Join

Instructors: Winston Hsu, Hao-Hua Chu  
Fall 2010

1

2

```

proc smjoin(R, S, R_i = S_j)
  if R not sorted on attribute i, sort it;
  if S not sorted on attribute j, sort it;

  Tr = first tuple in R;           // ranges over R
  Ts = first tuple in S;           // ranges over S
  Gs = first tuple in S;           // start of current S-partition

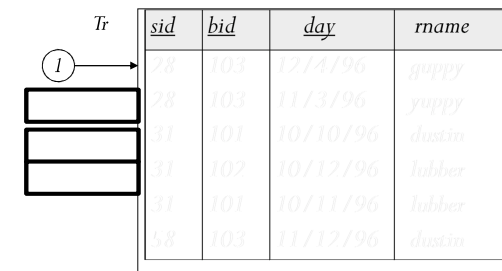
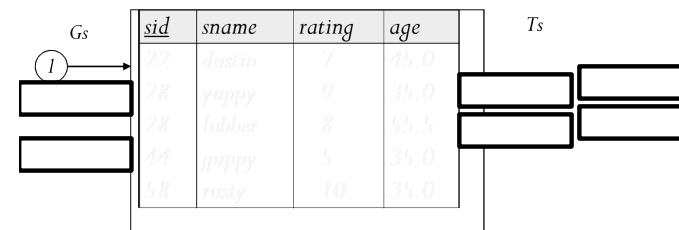
  while Tr ≠ eof and Gs ≠ eof do {
    while Tr_i < Gs_j do
      Tr = next tuple in R after Tr; // continue scan of R
    while Tr_i > Gs_j do
      Gs = next tuple in S after Gs   // continue scan of S

    Ts = Gs; // Needed in case Tr_i ≠ Gs_j
    while Tr_i == Gs_j do { // process current R partition
      Ts = Gs; // reset S partition scan
      while Ts_j == Tr_i do { // process current R tuple
        add (Tr, Ts) to result; // output joined tuples
        Ts = next tuple in S after Ts; // advance S partition scan
      }
      Tr = next tuple in R after Tr; // advance scan of R
    } // done with current R partition

    Gs = Ts; // initialize search for next S partition
  }
  
```

## Sort-Merge Join

## Scan-Merge



3

4

## Some APIs

- Sort constructor(... outfile ...)
  - Outfile : Name of sorted heapfile should be distinct
- HeapFile constructor
- HeapFile->openScan()
  - Followed by scan->getNext() and get the firstRecord
- tupleCmp() : compare tuples
- Scan->position()
  - Position the scan cursor to the record with the given rid.

5

## Other

- Output format:  
if 

recA
------

 = 

recA
------

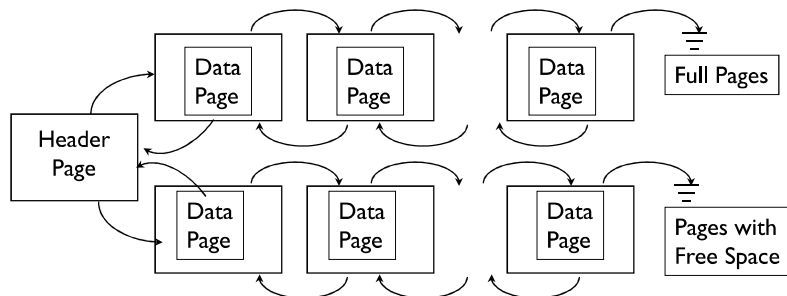
 ,then write 

recA
------

recA
------
- If there are any questions, you can post them on [ptt.cc](http://ptt.cc) / **CSIE\_DBMS**, or mail to: **217ta@csie.ntu.edu.tw**
- **Deadline: 23:59 Jan 12 (Wednesday), 2011**

6

## Heap File (Doubly Linked Lists)



- The header page id and Heap file name must be stored someplace.
- Each page contains 2 `pointers' plus data.
- The problem is that inserting a variable size record requires walking through free space list to find a page with enough space.

7

## Handin

- Due: Jan 12 (Wed) 23:59
- Email to: [217ta@csie.ntu.edu.tw](mailto:217ta@csie.ntu.edu.tw)
- 繳交格式  
hw6\_b95902080\_b95902091\_v1.zip
- 解壓縮後  
... -- hw6\_b95902080\_b95902091\_v1 + ---- ./sortMerge.C  
|  
+-----./report.pdf

8