Discussion Handout (2/5/2003)

Storing Data: Disk and Files

1. Underline the basic file organization (heap, sorted, or hash) that is best for a large file where the most frequent operations are as follows:
   a) *Heap/Sorted/Hash:* Search for records based on a range of field values.
   b) *Heap/Sorted/Hash*: Perform inserts and scans where the order of records does not matter.
   c) *Heap/Sorted/Hash:* Search for a record based on a particular field value.
2. Consider a relation stored as a randomly ordered file for which the only index is an unclustered index on a field called **sal**. If you want to retrieve all records with **sal>20**, is using the index always the best alternative? Explain.
3. Why are clustered indexes faster than unclustered indexes ? Is there anything that can be done to improve the performance of unclustered indexes ?
4. In the organization of a data page, what is a slot directory? What is it used for? In the organization of a data record, what is a field offset? What is it used for?
5. Consider the page format for variable-length records that uses a slot directory.
   One approach to managing the slot directory is to use a maximum size (i.e. a maximum number of slots) and allocate the directory array when the page is created.
   a) Discuss the pros and cons of this approach with respect to the approach discussed in the text.
   b) Suggest a modification to this page format that would allow us to sort records (according to value in some field) without moving records and without changing the record ids.
6. Consider the following relation describing employees in a company:
   Employee (eno: integer, name: char(20), phone: integer, age: integer, salary: real).
   Assume that this being stored in a heap file whose page size is 4KB where 20 bytes is used for some “magic” purpose.
   a) What is the most number of tuples that can be stored per slotted page ?
   b) Compute the number of records that you can store per page for each of the following cases
      i. Fixed length record format, PACKED page format
      ii. Fixed length record format, UNPACKED, BITMAP page format
      iii. Fixed length record format, Slotted page format
      iv. Variable length record format, Slotted page format