

Fall 2007 CS186 Discussion Section: Week 4, 09/17 - 09/21

Your Friendly TAs

September 27, 2007

1 File Organizations

1. What is a clustered index? For which of the 3 data entry alternatives can we have a clustered index?

A clustered index is one that is organized so that ordering of the data records within the corresponding file are the same as – or close to – the ordering of data entries in the index. Alternative 1 by definition is clustered. An index that uses Alternative 2 or 3 *can* be a clustered only if the data records are sorted on the search key field. Usually though, this is not the case.

2. You are about to create an index on a relation. Discuss some considerations that guide your choices of the following.

- (a) The choice of primary index
- (b) Clustered vs unclustered indexes
- (c) Hash vs tree indexes
- (d) The use of a sorted file vs a tree based index
- (e) Choice of search key for the index.

3. Choose one of the basic file organizations (heap, sorted, or hash) that is best for a large file, for each of the following scenerios:

- (a) Search for records based on a range of field values.
- (b) Perform inserts and scans where the order of records does not matter.
- (c) Search for a record based on a particular field value.

4. Fill in the I/O costs for the operations listed in the table. Assume that the relation R takes up $p(R)$ blocks of disk space and that it contains $t(R)$ tuples, the equality and range searches are performed on column $R.A$ which contains $v(R.A)$ unique values, and that the tree indices, again on column $R.A$ have height h and l leaf blocks. Calculate the costs in terms of I/Os – not time, as it is done in your books.

File Type	Scan	$R.A = c$	$R.A > c$	Insert	Delete
Heap					
Sorted					
Clustered Tree					
Unclustered Tree					

2 Tree Indices

1. Consider the B+ tree index of order $d = 2$ shown in figure below:

- (a) Show the tree that would result from inserting a data entry with key 9 into this tree.
- (b) Show the B+ tree that would result from inserting a data entry with key 3 into the original tree. How many page reads and page writes does the insertion require?

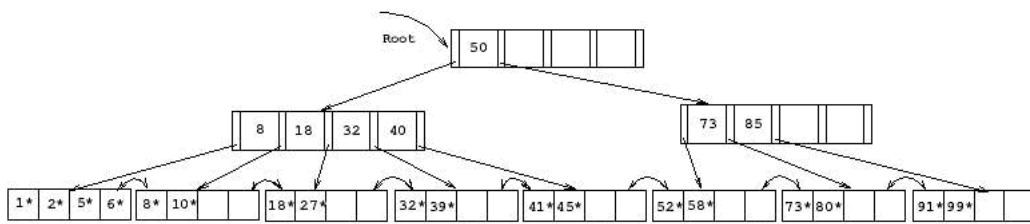


Figure 1: B+ tree