

Cost Model for Analysis

- For simplicity, ignore CPU costs:
 - **B**: number of data pages
 - **R**: number of records per page
 - **D**: (average) time to read or write disk page
- Average-case analysis; based on several simplistic assumptions

Heap Files

- Scan: _____
- Equality select
 - Equality selection on key: _____
 - Search not on a key: _____
- Range select: _____
- Insert
 - Assume always insert at the end of file: _____
- Delete
 - Write the page back: _____
 - If the search is based on a given rid: _____
 - If the search is based on equality or range select:

Sorted Files

- Scan: _____
- Equality select: search based on sorted order
 - Match only 1 record: _____
 - Match more than 1 records: _____

- Range select: _____
- Insert: _____
- Delete
 - Read and rewrite all subsequent pages: _____
 - If the search is based on a given rid: _____
 - If the search is based on equality or range selects:

Clustered Indexed Files

- Empirical study shows that the pages usually have 67% occupancy. The # of physical pages is: _____

- Scan: _____

- Equality select: search based indexed key

 - Match only 1 record: _____

 - Match more than 1 records: _____

- Range select: _____

- Insert: _____

- Delete: _____

Unclustered Tree Index

- Assume that the size of data entry in the index is 1/10 of data records. If the index pages have 67% occupancy, the # of leaf pages is: $\sim 0.15B$. The number of data entries on a page is $6.7R$.
- Scan: _____
- Equality select: search based on indexed key
 - Match only 1 record: _____
 - Match more than 1 records: _____

Unclustered Tree Index (cont.)

- Range select: _____
 - If more than 10% of records satisfy the select condition, we are better off retrieve all the data records, sort them based on selection attribute, and retain the records satisfying the selection
- Insert: _____
- Delete: _____