Comparing File Organizations

- Heap files (random order; insert at eof)
- Sorted files, sorted on <*age, sal*>
- Clustered B+ tree file, Alternative (1), search key <*age, sal*>
- Heap file with unclustered B + tree index on search key <*age, sal*>
- Heap file with unclustered hash index on search key <*age, sal*>

Cost Model for Our Analysis

We ignore CPU costs, for simplicity:

- **B:** The number of data pages
- **R:** Number of records per page
- **D:** (Average) time to read or write disk page
- Measuring number of page I/O's ignores gains of pre-fetching a sequence of pages; thus, even I/O cost is only approximated.
- Average-case analysis; based on several simplistic assumptions.

☛ *Good enough to show the overall trends!*

Operations to Compare

- Scan: Fetch all records from disk
- Equality search
- Range selection
- Insert a record
- Delete a record

Assumptions in Our Analysis

- Heap Files:
  - Equality selection on key; exactly one match.
- Sorted Files:
  - Files compacted after deletions.
- Indexes:
  - Alt (2), (3): data entry size = 10% size of record
  - Hash: No overflow chains.
    - 80% page occupancy => File size = 1.25 data size
  - B+Tree:
    - 67% occupancy (typical): implies file size = 1.5 data size
    - Balanced with fanout F (133 typical) at each non-level

Assumptions (contd.)

- Scans:
  - Leaf levels of a tree-index are chained.
  - Index data-entries plus actual file scanned for unclustered indexes.
- Range searches:
  - We use tree indexes to restrict the set of data records fetched, but ignore hash indexes.

Cost of Operations

	Scan	Equality	Range	Insert	Delete
Heap File	BD	.5BD	BD	2D	Search + D
Sorted File	BD	$D\log_2 B$	$D(\log_2 B$ + #matching pages)	Search + BD	Search + BD
Clustered Tree Index	1.5BD	$D\log_F 1.5B$	$D(\log_F 1.5B$ + #matching pages)	Search + D	Search + D
Unclustered Tree Index	BD(R+.15)	$D(1+\log_{F.15B})$	$D(\log_{F.15B}$ + #matching **recs**)	Search + 3D	Search + 3D
Unclustered Hash Index	BD(R+.125)	2D	BD	4D	4D

☛ *Several assumptions underlie these (rough) estimates!*