More Multidimensional Indexes

CS157B

Chris Pollett

Mar. 7, 2005.

Outline

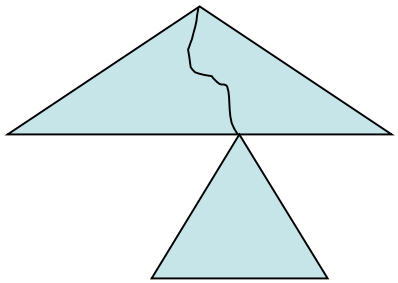
- Tree-Like Structures for Multidimensional Data
- Bitmap Indexes
- Query Execution

Different Kinds of Tree Indexes

- Multiple Key Indexes
- kd-trees
- Quad trees
- R-trees

Multiple Key Indexes

- If have a two attribute index, make a tree index on first attribute.
- The leaves of this tree are possible values for this attribute.
- The pointer followed from this value points to an index on the second attribute.

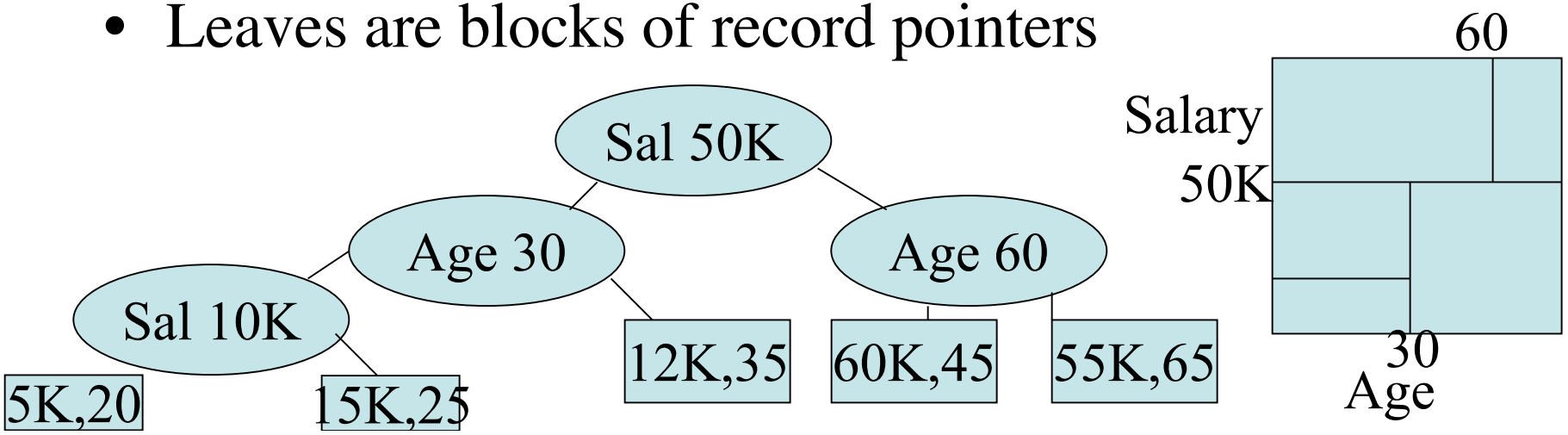


Uses of Multiple-Key Indexes

- Might have such an index on Age and Salary.
- If in a query specify both a fixed Age values (say 35) and a fixed Salary (say 20K), the results can be fast.
- However suppose wanted all Age's for a given Salary -- this would be slow if Age was the first thing indexed.
- Multiple key indexes work well for range queries and nearest neighbor queries. (Treat nearest neighbor as a range query)

kd-trees

- kd-trees are a generalization of binary search trees.
- Each node has some attribute A and some value V.
- The tree satisfies that beneath a node labeled as above all the records to the left of the node are less than V, those to the right are greater than V .
- Leaves are blocks of record pointers

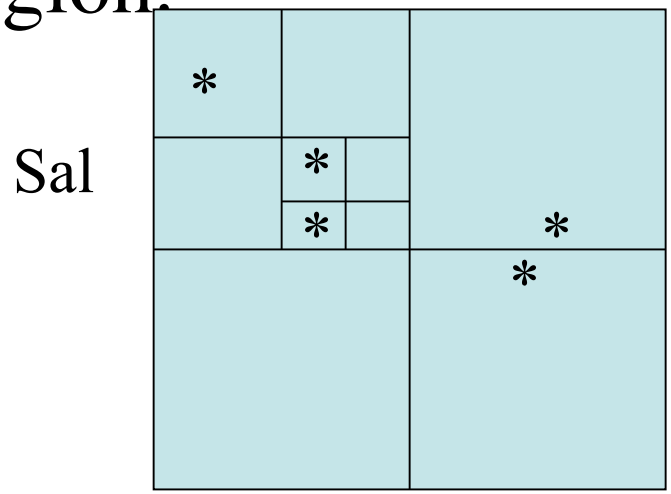


Operations on kd-trees

- Basic search is straightforward.
- Partial match search involves following both paths out of a node if do not know its value.
- Similarly, for range query, if range straddles a query have to follow both branches.
- Nearest neighbor search is treated as a special kind of range query.
- To optimize for disk rather than main memory, rather than split on one value V at a node, split n ways based on block size.

Quad-trees

- Split regions into four subregions if it takes more than one block to store the contents of a region.

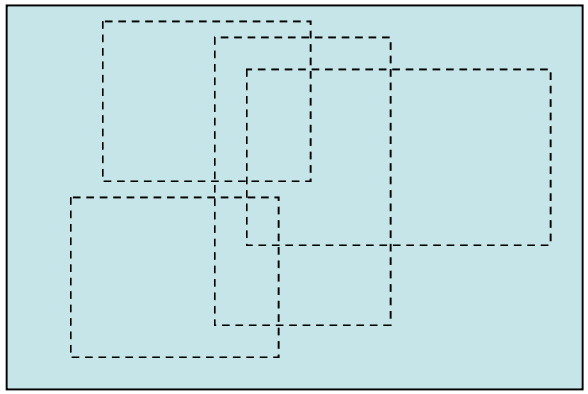


Here a * indicates at one blocks worth of data. Operations similar to kd-trees. Can be generalized if have keys on more than two attributes.

R-Trees

- Each node in an R-tree is supposed to contain all values that lie within a region.
- Regions are allowed to overlap.
- Regions can be arbitrary shapes but are usually rectangles.
- R-trees are useful for “Where am I?” queries

Example region for a node and its subregions



Bitmap Indexes

- A n-bit vector V is a string of 0 and 1's of length n.
- A bitmap index for a field F is a collection of n-bit-vectors, one for each value of F.

Value	Vector
foo	100100
bar	010010
baz	001001

A 1 indicates that in that row the value occurred.

Uses of Bitmaps

- Can be very fast at doing operations like intersection and union, since these map to ANDs and ORs
- Also, good for range queries

Compressed Bitmaps

- Frequently, have lots more 0's than 1's in a given bitmap.
- Makes sense to do things like run-length encoding to compress the bitmap.
- There are algorithm for doing each of the operations we mentioned on the compressed bitmaps.

Query Execution

- We are going to start talking about how query results are computed by the DBMS.
- An overview of the operations involved might look like: