CS 2550 – Principles of Database Systems (Spring 2017)
Dept. of Computer Science, University of Pittsburgh

Assignment #3: Physical Organization & Relational Algebra

Release: Feb 9, 2017 Due: 8:00 PM, February 17, 2017.

Goal

Study storage organization, practice different indexing and hashing structures as well as review relational algebra that is the basis for query optimization.

Description

Answer the following questions [for a total of 100 points]

Physical Organization

1. [10 points] Assume a length field for variable-length attributes with a fixed size of 1 byte. How can you store longer than 255 bytes in such a variable-length attribute? Assume that you would like to store strings of 1024 bytes.

2. [15 points] Suppose that we are using linear hashing on a file that contains records with the following search-key values:

   2, 3, 5, 7, 11, 17, 9, 13, 4, 19, 20, 29, 31, 25, 23

   Suppose the hash function is $h_0(x) = x \ \ mod \ \ 4$ and buckets can hold three records, and there is one overflow bucket to be used to minimize the number of splits.

   (a) Show the linear hash structure for this file (showing BSplit, BLast pointers) after each split.

   (b) Show how the structure changes as the result of deleting 7, 23, 31 (showing BSplit, BLast pointers).

3. [10 points] Show the extensible hashing structure that result from inserting the keys of Problem 2 using the least significant bits. Suppose that buckets can hold three records.

4. [15 points] Construct a B^+-tree for the following set of key values:

   2, 3, 5, 7, 8, 17, 15, 19, 11, 20, 29, 31, 25, 23

   Assume that the tree is initially empty and values are added in the order listed above. Construct B^+-tree for the cases where the degree is four.

   Show the B^+-tree in the following **four** stages: after inserting 15, 20, after inserting all the keys, and after deleting both 17 and 19. Number each node incrementally as they are being created.

5. [15 points] Consider a relation R with r tuples for which you need to build an index structure for its primary key. Assume that the size of the primary key is $V = 9$ bytes, the block size is $B = 512$ bytes, a data block pointer is $P_D = 7$ bytes and an index block pointer is $P_I = 6$ bytes.

   What is the main reason that you will select a B+-tree instead of a B-tree? Make your case by showing the number of tuples r held by each tree of level/depth 4, assuming that each node in each tree is 71% full.

6. [15 points] Suppose we have a database system that uses only the fixed length records to store tuples. In the case of HDD, traditionally the page organization scheme is the slotted page. As Solid State Disk (SSD) drive technology matures, we would like to use SSD to store our tuples. If the slotted page scheme is not appropriate to continue using with the SSDs, suggest one appropriate page organization scheme for SSDs.

Relational Algebra Review:

7. [10 points] Using the schema from homework 2 (helpdesk), answer the following queries using relational algebra.
   (a) List all the pairs of machine ids that have at least one ticket with the same category id, with each pair being listed only once (e.g., (3,2) is the same as (2,3)). A pair with the same ids is invalid (e.g., (2,2)).
   (b) List the average number of tickets for all the machines that are in the same building.
8. [10 points] Using the schema from homework 2 (helpdesk), and assuming that the relations TICKETS and ASSIGNMENT have 6 and 12 tuples, respectively, find the arity and cardinality of the following relations: (For those whose accurate values can not be determined, give the min and max values)
   (a) $\Pi_{machine\_name} TICKETS$
   (b) $TICKETS * ASSIGNMENT$

What to submit

You are required to submit **exactly one** PDF file under your **pitt_user_name** (e.g, pitt01.pdf) with your name and pitt user name at the top. *No hand-written and scanned answers will be accepted and graded.*

How to submit your assignment

1. Submit your assignments through the Web-based submission interface (go to the class web page `http://db.cs.pitt.edu/courses/cs2550/current.term` and click the Submit button). **It is your responsibility to make sure the assignment was properly submitted.**
2. You must submit your assignment before the due date (**11:59pm, Friday, Feb. 17, 2017**).

Academic Honesty

The work in this assignment is to be done *independently*. Discussions with other students on the assignment should be limited to understanding the statement of the problem. Cheating in any way, including giving your work to someone else will result in an F for the course and a report to the appropriate University authority.