CMPT 354
Database Systems

Simon Fraser University
Spring 2017

Instructor: Oliver Schulte

Assignment 4: System Implementation, XML.

Instructions: Check the instructions in the syllabus. The university policy on academic dishonesty and plagiarism (cheating) will be taken very seriously in this course. *Everything submitted should be your own writing or coding.* You must not let other students copy your work. On your assignment, put down your **name**, the number of the assignment and the number of the course. Spelling and grammar count.

Group Work: Discussions of the assignment is okay, for example to understand the concepts involved. If you work in a group, put down the name of all members of your group. There should be no group submissions. Each group member should write up their own solution to show their own understanding.
For the due date please see our course management server https://courses.cs.sfu.ca .

Part I: Storage and Indexing

Consider the following instance of the Students relation, sorted by *gpa*.

sid	name	age	gpa
53831	Madayan	11	1.8
53832	Guldu	12	2.0
53688	Smith	19	3.2
53666	Smith	18	3.4
53650	Jones	19	3.8

For the purposes of this question, assume that these tuples are stored in a sorted file in the order shown. Each page can store up to three data records. So the first three tuples are on page 1, the fourth is on page 2 etc.

1. Explain what the *data entries* in each of the following indexes contain. If the order of entries is significant, say so and explain why. For definition of terms, refer to the text. You do *not* need to show the entire index (e.g. index nodes for tree indexes) just the data entries (e.g. just the leaves in a tree index).

    a) A hash index on *gpa* using Alternative (1).
    b) A clustered tree index on *gpa* using Alternative (2).
    c) An unclustered hash index on *name* using Alternative (3).

2. Consider a delete operation specified using an equality condition on a key. Assuming that no record qualifies, what is the cost for the three file organizations: heap file, sorted file, unclustered hash index? Present your analysis using the same parameters as in the lectures (you can ignore the H and C parameters from the text)

    B = number of data pages
    R = number of records per page
    D = time to read or write disk page

Grading Criteria:

Total Marks: 40 Marks
1. Question 1 (20 Marks)
2. Question 2 (20 Marks)

Part II: Query Evaluation

1. Consider the following SQL query.

   SELECT S.sname
   FROM Sailors S, Reserves R, Boats B
   WHERE S.sid = R.sid AND R.bid = B.bid AND B.colour = ‘green’
   INTERSECT
   SELECT S.sname
   FROM Sailors S, Reserves R, Boats B
   WHERE S.sid = R.sid AND R.bid = B.bid AND B.colour = ‘blue’

   Write a query evaluation plan (relational algebra tree) for evaluating this query. (You do not have to annotate the nodes with access methods, just specify the relational operators for each node.)

2. (Bonus Question) Consider the following schema with the Sailors relation:

   Sailors (sid: integer, sname: string, rating: integer, age: real)

   For each of the following indexes, list whether the index matches the given selection conditions. If there is a match, list the primary conjuncts.

   (a) A hash index on the search key <Sailors.sid, Sailors.rating>

   - $\sigma_{\text{rating = 10 AND sid=500}}$ (Sailors)
   - $\sigma_{\text{ rating > 10 AND sid=500}}$ (Sailors)

   (b) A B+-tree on the search key <Sailors.sid, Sailors.rating>

   - $\sigma_{\text{sid< 500 AND rating=10}}$ (Sailors)
   - $\sigma_{\text{sid= 500 AND rating > 10}}$ (Sailors)
   - $\sigma_{\text{sid< 500}}$ (Sailors)
   - $\sigma_{\text{rating > 10}}$ (Sailors).

3. Emp(eid:integer, salary: integer, age: real, did: integer)

   Assume that there is an unclustered index on *age* and an unclustered index on *eid*.

   a. How would you use the indexes to enforce the constraint that *eid* is a key?
   b. Give an example of an *update* that is definitely speeded up because of the available indexes. (English description of the update is sufficient). Briefly explain how the indexes cause the speedup.

c. Give an example of an *update* that is definitely slowed down because of the available indexes. (English description of the update is sufficient). Briefly explain how the indexes cause the slowdown.

Grading Criteria:

Total Marks: 40 Marks

1. Question 1 (10 Marks)
2. Question 2 (18 Marks)
3. Question 2 (10 Marks)

Part III: XPath Questions

Refer to: http://www.w3schools.com/xml/xpath_examples.asp
For the below questions, use the XML document from the above page, which contains an XML structure for a bookstore. Please test your queries to ensure that your XPath queries work. The following website seems to evaluate Xpath expressions correctly http://www.online-toolz.com/tools/xpath-editor.php . Write Xpath expressions for the following queries.

(a) Select all books published after 2003.

(b) Display the language of the third book.

(c) Display only the book titles that are categorized for cooking.

Grading Criteria:

Total Marks: 18 Marks
6 Marks for each query

Optional XPath Questions

Using the same document as in the example above.

(d) For each title in the bookstore, display its language.
(e) Display only the book titles that are categorized for cooking *or* for children.
(f) Display (only) the prices of all and only the book titles that are categorized for cooking.

What to Submit

X-Path Question: Please submit a .pdf showing a screenshot with each of your queries and the result of the query. Call this xpath.pdf .

All other questions: Write out the answers in a word processing program and submit a pdf with them. Call this main.pdf .

Please submit both .pdfs on courses.cs.sfu.ca.

A note on the relational algebra tree (query evaluation question): We'd prefer you use a drawing program, but we'll accept hand-drawn trees as long as they are easy to read after scanning. Remember that by using a drawing program, you make it easy to modify your answer as you get better and better ideas.