Name: Ryan Lockman
SID: 101430670

Midterm II
Spring 2014
CSC3287 Database Systems Concepts

74/100

Instructions:

Work alone to complete the questions. This test is an open book/open notes test. Cheating will be dealt with according to department policy.

Put your name on each page to be sure that I know whom to credit for the work.

The test, in total, is worth 100 points.

I WILL NOT ANSWER QUESTIONS DURING TESTS because the nature of this work requires you to make assumptions. If you have questions during the test, state your assumption on your paper and follow it through in your answer. Partial credit will be given for stating and following your assumptions even if they are not what I was thinking when the test was written. You may add things to your model if necessary but you must also state your assumptions that go with your addition.

Some questions have multiple answers.

Write your expressions and draw your models in the manner that you are most comfortable with. But be sure to use the formal notation taught in the book or in class and be as complete as possible with your answers.

If you finish early, check for completeness in your drawings and expressions. When you are convinced that you are done, deliver your test in one of two ways:

Email to debra.parcheta@ucdenver.edu

Or

Hand it to me at class on 4/1/2014 at 3:30 PM

Verify that two Exhibits are attached:

Museum Database EER Diagram and
Airline Database System ER Diagram

Name: Ryan Lockman
SID: 101430670

1. (5 Points) Name an attribute from the Museum Database System that would be appropriate for building a primary index. Tell why you chose that attribute.

I would choose the Name primary key attribute of the ARTIST entity. The reason for my choice is that if the records need to be ordered, ordering them alphabetically by the ARTIST name would be the best solution. +

---

2. (5 Points) Name an attribute from the Museum Database System that would be appropriate for building a secondary (non-key) index. Tell why you chose that attribute.

I would choose the Title attribute of the ART_OBJECT entity. The reason I chose this attribute was the fact that it would make a good non-key means to look up a record besides a sequential primary index. The Title could be hashed to find the specific record and duplicate Titles, if any, would still be resolved quickly because of the type of attribute it is. A Title is usually unique. +

---

3. (5 Points) Name an attribute from the Museum Database System that would be appropriate for building a clustering index. Tell why you chose that attribute.

For a clustering index you could choose the Year attribute of the ART_OBJECT entity. The reason for my choice is about the same as my previous answer. Though their would be duplicates, it would still provide good means for looking up art. Their is a lot of art produced, but for a given year it would be minimal. -2

-2

Name: Ryan Lockman
SID: 101430670

4. (15 points) Assume that a B+ Tree has been built using consecutive integers as the values in each node, starting with the integer 1 and ending with the integer 122, where p = 6, n at the leaves is p-1, n for an internal node = p+2, and the tree is 70 to 80% full at the leaves. Calculate the efficient minimum number of block accesses in the resulting B+ tree to retrieve the set of records with values. {6, 18, 62, 100, 101, 106}. Assume that the actual retrieval takes one block access and that, when the record is retrieved, the search for the next record continues from the last visited node. State any other assumptions you make.

-3 - 4 See solution

75% of 122 is 91.5 nodes.

91.5/ 6 is 15.25 block accesses.

For my assumption I was a little confused on the question. I do figure that for a B+Tree that the Big O Notation time for retrieving a set of records would be of O(log n) operations. So that would also mean time goes up linearly while n goes up exponentially.

---

5. (10 points) Refer to the Museum Database exhibit.
Use formal relational mapping notation to describe 5 different constraints on the Museum Database. (Any 5 you like.)

1. An ART_OBJECT is made by N ARTISTs. Means that a piece of art can be made by at least one ARTIST, but can be worked on by N number of ARTISTs.
2. An ARTIST can make (0 to n) pieces of ART, meaning that the ARTIST doesn't have to make any art and can be in the database, or he/she can have n number of ART_OBJECTs in the database.
3. BARROWED_ART BELONGS_TO 1 COLLECTION, meaning that it can't be from multiple COLLECTIONs. This makes sense because a piece of art is just that, a piece of art. Half of it did not come from one COLLECTION and vise versa. Also if it was among the art in one COLLECTION it would be impossible for it to be at another location in another COLLECTION.
4. A COLLECTION can have (1 to n) pieces of art. This makes sense because a COLLECTION must have at least 1 item, otherwise it wouldn't be a COLLECTION.

-5

Name: Ryan Lockman
SID: 101430670

It can also have up to n items in the COLLECTION because theirs no limit on what was collected for the COLLECTION.

5. An EXHIBITION can have N number of ART_OBECTs that are SHOWN_IN. This means that their can be any number of ART_OBJECTs that are currently on display at the exhibit.

6. (5 points) Referring to the Museum Database exhibit, write the relational algebra expression that produces a list of the COLLECTIONs where the COLLECTION.Type was recorded as "Painting."

-Symbol below is sigma symbol. Yuk It's in MSWord Symbol Table -2

$O\sim_{type=Painting}(COLLECTION)$

---

7. (5 points) Refer to the Airline Database system. Name the entity(ies), if any, that would have a fixed record size. Tell why you would know that.

I would say that SEAT, FLIGHT_LEG, LEG_INSTANCE, and AIRPLANE. The reason for this is because SEAT, has only a Seat_no which would be fixed size, FLIGHT_LEG, has only a fixed size Leg_no, LEG_INSTANCE, has a fixed Date and could have fixed digits(003, 123, etc...) for No_of_avail_seats, and AIRPLANE, because the Airplane_id is of fixed length and Total_no_of_seats could be of fixed length(123, 347, 099, etc...). +

---

8. (5 points) Refer to the Museum Database exhibit. Specify the size of a record in the PERMANENT COLLECTION.

The size of the PERMANENT COLLECTION would be of fixed length. If I made the judgment on its actual underlying size I would say Cost would be a maximum of 10 digits, Date_Acquired would be a maximum of 8 digits, and On_Display would be a maximum of 5 digits. One record size would then equal 23 bytes. -2 inheritance and data types

---

9. (5 points) Continue to refer to the Museum database. Given a block size of 1024, what is bfr for the records in ART_OBJECT.

If B = 1024 and we had a record with a fixed size of 107(I calculated this from the total bytes of each Art_OBJECT attribute) then the bfr = 1024/107, which would equal 9.57 records. This would be rounded to floor $\lfloor \_ \rfloor$, and would mean that 9 records would fit. +

Name: Ryan Lockman
SID: 101430670

10. (10 points)

a. Referring to the Airline Database Exhibit, tell me, for each entity what type of record blocking technique would be used and why.

As I stated in question 7, SEAT, FLIGHT_LEG, LEG_INSTANCE, and AIRPLANE are of fixed size. Because they are fixed I would use the unspanned technique. The reason for this is that since they are fixed records and that block size is greater than the record size it would make each record start at a specific location in the block allowing for a simpler record processing. AIRPORT, AIRPLANE_TYPE, FLIGHT, and FARE are of variable length size so a spanning technique would be better suited. The reason for spanning is to not loose the unused space within a block. You would just store a pointer to the rest of the last record in the block.

+

but would R->B for these?

---

11. (5 points) Given the relational nature of the Airline database, what sort of file allocation would work best and why?

For the entire airline database I would use Linked Allocation. The reason behind this choice is that an airline database would most likely need to expand in the future, there are always things that you could add in or change. With linked allocation you are able to achieve expansion through pointers to the next block. This does make reading the file slower but the expansion can pay off in the long run.

+

---

12. (5 points) Write the generic form of an aggregation statement using Relational Algebra notation.

-2

-Symbol below is aggregate symbol.

Q ≡ ← <grouping attributes> ℱ <function list> (R)

---

13. (5 points) Write an actual aggregation statement using data from the Airline Database System. (Any aggregation that you can produce is fine.)

-2

Q ≡ ← Code ℱ AVERAGE(Amount)(FARE)

Name: Ryan Lockman
SID: 101430670

14. (15 points) Consider the two tables T1 and T2 shown in Figure 6.15 below. Show the results of the following operations on the tables. (Extra blank page provided for answers if needed.)

A. Q

A	B	C	P	Q	R
10	b	6	10	a	5
25	c	3	25	a	6

-2

B. Q

A	B	C	P	Q	R
10	b	6	15	b	8
10	b	5	15	b	8

C. Q

A	B	C	P	Q	R
10	b	6	10	a	5
NIL	NIL	NIL	15	b	8
NIL	NIL	NIL	25	a	6

-2

D. Q

A	B	C	P	Q	R
10	b	6	NIL	NIL	NIL
25	c	3	NIL	NIL	NIL
10	b	5	NIL	NIL	NIL

-2

E. Q

A	B	C	P	Q	R
10	b	6	10	a	5
25	c	3	15	b	8
10	b	5	25	a	6

-2

F. Q

A	B	C	P	Q	R
10	b	5	10	a	5

-2