B561 Final Exam

C.M. Wyss A. Kaushik G. Fletcher

December 15, 2003

Instructions: You have 2 hours to complete this exam. This exam comprises 3 questions (4 pages total). This exam is worth a total of 100 points. Point distributions are marked on the questions. Ensure you attempt all questions and allot your time accordingly.

This test concerns our pirate armada. Assume the following tables exist in the database:

SHIPS(SID, Name, ArmClass, Specs)

PIRATES(PID, Pseudonym, SkillLevel, Bio)

MISSIONS(SID, PID, Date, Name, Booty, Successful)

The fields in these relations have the following sizes:

Relation	Field	Size (in bytes)
SHIPS	SID	4
	Name	128
	ArmClass	1
	Specs	1024
PIRATES	PID	4
	Pseudonym	128
	SkillLevel	1
	Bio	8096
MISSIONS	SID	4
	PID	4
	Date	4
	Name	128
	Booty	4
	Successful	1

Table 1: Field sizes (in bytes).

In addition, |SHIPS| = 150 tuples, |PIRATES| = 2500 tuples, |MISSIONS| = 10^7 tuples. The relations are stored on a disk which has a block size of 1 kilobyte (1024 bytes). The average access time to retrive/store a block to disk is 12 milliseconds.

The field SkillLevel in relation PIRATES can take one of three values: novice, expert, or ruthless. The field Successful in relation MISSIONS is a Boolean field. A ships armament class is represented by a number from 0 through 9. Assume that values in the relations are distributed uniformly over these ranges (where appropriate). *In addition, assume that the database contains mission data for exactly one year (365 days).*

Keep in mind the following guidelines when solving the problems below:

- It is easiest to assume a sequential layout of fields in records and records in blocks. However, for fields more than 8 bytes large, assume an 8-byte pointer is stored to the location of that field in memory (outide of the file storage space). Retrieving these fields thus takes an additional block I/O. You may neglect header/tombstone information.
- Make reasonable assumptions where necessary and state these explicitly.
- Show your work for every question.
- Your final cost calculations should be time measures (e.g. seconds, hours, etc.).
- Ignore all main memory calculations, and base your costs solely on disk I/O. You should not include time to output the final result.

1. (**30 points**) Consider the following query: Find triples (m_1, m_2, s) such that s had an armament class of greater than 4 and participated in two missions m_1 and m_2 on the same date. We can phrase this in SQL as follows:

```
SELECT M1.Name, M2.Name, S.Name
FROM Missions M1, Mission M2, Ships S
WHERE S.ArmClass > 4 AND
      M1.SID = S.SID AND
      M2.SID = S.SID AND
      M1.Date = M2.Date;
```

   For the questions that follow, assume we have 100 MB of main memory available to each process.

   (a) Translate the query into an RA expression and give the tree form for your expression. Your translation should be naieve (i.e. do not perform any optimization yet). How long will this plan take to run?

   (b) Assume we have B+-tree indexes on fields SID of SHIPS and ⟨SID, PID, Date⟩ of MISSIONS. Will these indices make a significant difference in the running time of your naieve plan?

   (c) Suppose we have an additional secondary index on field Date of MISSIONS. What is the running time of the (heuristically) best plan for this query? Exhibit this plan in tree form along with your analysis.

2. (**30 Points**) Tables 2 and 3 (end of test, below) show two schedules that involve three transactions T_1, T_2 and T_3. Answer the following questions for these schedules.

   (a) Is schedule A conflict serizable? If yes, give an equivalent serial schedule. If no, explain.

   (b) Is schedule B conflict serizable? If yes, give an equivalent serial schedule. If no, explain.

   (c) Is either schedule recoverable? Explain.

3. (**40 Points**) For this question, you are to write an essay of no more than 500 words on the following topic. Ensure you carefully consider your answer. Answers that are "mostly point form" are acceptable as long as they appropriately address the issues in a coherent manner. You may wish to write an outline first. Any scrap paper you use should be turned in with the exam.

   > The "serializable" properties of schedules are supposed to assist in guaranteeing Consistency and Isolation of transactions during concurrent executions. The "recoverable"-related properties of schedules are supposed to assist in guaranteeing Atomicity and Durability of transactions during concurrent exections. How well do these properties achieve their goals? Discuss in detail.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T_1						R(X)	W(X)				R(Y)	W(Y)			Cmt	
T_2	R(Z)	R(Y)	W(Y)							R(X)			W(X)	Cmt		
T_3				R(Y)	R(Z)			W(Y)	W(Z)							Cmt

Table 2: Schedule A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T_1			R(X)	W(X)				R(Y)	W(Y)					Cmt		
T_2							R(Z)			R(Y)	W(Y)	R(X)	W(X)			Cmt
T_3	R(X)	R(Z)			W(Y)	W(Z)									Cmt	

Table 3: Schedule B