

**0/5 Questions Answered**

## Vitamin 6

**STUDENT NAME**

### Q1 Join Costs

6 Points

Consider the join( $R, S$ ) on  $R.a = S.b$ , given the following information about the relations to be joined. The cost metric is the number of page I/Os unless otherwise noted, and the cost of writing out the result should be uniformly ignored. For join( $R, S$ ) assume  $R$  is the outer loop and  $S$  the inner loop.

- Relation  $R$  contains 10,000 tuples and has 10 tuples per page.
- Relation  $S$  contains 2000 tuples and also has 10 tuples per page.
- Attribute  $b$  of relation  $S$  is the primary key for  $S$ .
- Both relations are stored as simple heap files.
- Neither relation has any indexes built on it.
- 52 buffer pages are available.

#### Q1.1 PNLJ

2 Points

What is the cost of joining  $R$  and  $S$  using a page-oriented nested loop join?

What is the minimum number of buffer pages required for the cost of the above question to remain the same?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

Save Answer

### Q1.2 BNLJ

2 Points

What is the cost of joining R and S using a block nested loop join?

Enter your answer here

What is the minimum number of buffer pages required for the cost of the above question to remain the same?

- ☐ 1
- ☐ 2
- ☐ 50
- ☐ 51
- ☐ 52

Save Answer

### Q1.3 Sort Merge Join

2 Points

What is the cost of joining R and S using sort-merge join? (Consider the cost of read and write for sorting; this is ordinary sort-merge, not optimized sort-merge, so the sort and merging steps are fully independent).

Enter your answer here

What is the minimum number of buffer pages required for the cost of the above question to remain the same?

- ☐ 1
- ☐ 2
- ☐ 20
- ☐ 31
- ☐ 32
- ☐ 33
- ☐ 52

Save Answer

### Q2 Hash Joins

2 Points

#### Q2.1 Partition Size

**True or False.** All partitions must be no larger than B-2 pages when we begin the build and probe stage.

- ☐ True
- ☐ False

 False

## Q2.2 Hash Functions

**True or False.** No two partitioning passes may use the same hash function, but the hash functions may be reused in the probing pass.

- ☐ True
- ☐ False

Save Answer

## Q3 Grasshopper Join

2 Points

You are the organizer for the 2020 annual grasshopper exhibition. You are given two tables: Table R, which contains information about all students taking CS 186, and Table S, which contains information about all UC Berkeley students who want to attend the exhibition. You want to find out how many students in CS 186 are attending the exhibition, and whether they have enough sleep minutes to put off their project and have a great time!

### Q3.1 Perfect Grass

We will perform Grace Hash Join using tables R and S. Operate under the following assumptions:

- $[R] = 60$ ,  $[S] = 20$
- $B$  (number of buffer pages) = 6

Assuming that we use perfect hash functions that uniformly partition our data at every step of the join process, what is the I/O cost of executing Grace Hash Join?

Enter your answer here

### Q3.2 Imperfect Grass

Now, assume that the first hash function we use was imperfect. As a result, we end up with partitions after the first pass that look like this:

- Partition 1:  $[R] = 20, [S] = 4$
- Partition 2:  $[R] = 20, [S] = 7$
- Partition 3:  $[R] = 10, [S] = 2$
- Partition 4:  $[R] = 4, [S] = 5$
- Partition 5:  $[R] = 6, [S] = 2$

Assuming that we use a perfect hash function on each partition following this initial mishap of a partitioning phase, what is the total number of I/Os, from start to finish, needed to execute a Grace Hash Join between R and S?

**Remember:** we don't count the final write cost in our I/Os.

Enter your answer here

Save Answer

Save All Answers

Submit & View Submission >