0/14 Questions Answered

Vitamin 11

STUDENT NAME

Search students by name or email... ▾

Q1 Parallel Query Processing

7 Points

For this vitamin's questions, assume that data is streamed from the network to disk (i.e. pass 0 does not happen "on the fly" as the machine reads in data) Therefore, you should include the cost of reading the data from disk, once it has been streamed there from the network.

Q1.1 Inter vs Intra

2 Points

Partitioning a relation over multiple machines to perform a parallel hash join.

○ Inter-query parallelism

○ Intra-query parallelism

Splitting up a set of 10,000 queries so that each of our 100 machines only runs 100 queries.

○ Inter-query parallelism

○ Intra-query parallelism

Save Answer

Q1.2 Sorting

1 Point

Given 8 machines with 3 pages of buffer each, how many passes do you need to sort 96 pages of data. A pass is defined as reading all of the data. Assume the data is uniformly partitioned across the machines.

Enter your answer here

Save Answer

Q1.3 Hashing

1 Point

Given 8 machines with 3 pages of buffer each, how many passes do you need to hash 96 pages of data? Assume the data is uniformly partitioned across the machines, and that we have perfect hash functions that uniformly partition data. A pass is defined as reading all of the data.

Enter your answer here

Save Answer

Q1.4 Range Partition 1

1 Point

If relation R and relation S are both 32 pages and range partitioned (with uniform ranges) over 2 machines with 4 buffer pages each, what is the disk I/O cost per machine for performing a sort-merge join? (Assume that we are performing an unoptimized sort- merge join, and that data is streamed to disk after partitioning.)

Hint: Don't forget the write cost as data is coming off of the network. However, do not include the initial read cost for sending data through the network.

Enter your answer here

Save Answer

Q1.5 Round Robin Partition

1 Point

If relation R and relation S are both 32 pages and round robin partitioned over 2 machines with 4 buffer pages each, what is the network cost (number of bytes sent over the network by any machine) for performing a sort-merge join in the worst case? Assume each page is 4KB.

Provide only a numerical answer in terms of KB

Enter your answer here

Save Answer

Q1.6 Other Joins

1 Point

If we have 10 machines, relation R is 2 pages and entirely resides on Machine 1, and relation S is 20,000 pages, round-robin partitioned over all 10 machines, which of the following joins will minimize network cost? (Assume that our network only supports "unicast" messages, where a machine can only send a message to one other machine at a time).

○ Parallel Sort-Merge Join

○ Parallel Hash Join without an asymmetric shuffle

○ Parallel Hash Join with an asymmetric shuffle

○ Broadcast Join

Save Answer

Q2 Class Registration

10 Points

We have 3 machines: m1, m2, and m3. The data for the `Students` table is range partitioned on the `age` column such that m1 has 45 pages of data, m2 has 10 pages, and m3 has 50 pages. The ranges for each machine are: m1 has values 1-50, m2 has values 51-100, and m3 has values 100-150. The data is not stored in sorted files on any machine. For the following questions, how many I/Os across all machines will it take to execute each query? Assume that each machine has 1000 pages of memory.

Q2.1 Query 1

1 Point

Query 1: `SELECT MAX(age) FROM Students;`

Enter your answer here

Save Answer

Q2.2 Query 2

1 Point

Query 2: `SELECT COUNT(DISTINCT age) FROM Students;`

Enter your answer here

Save Answer

Q2.3 Query 1 vs Query 2

3 Points

Which queries, if any, would run faster if we round-robin partitioned the data instead of the range partitioning scheme described in the problem? Assume 1 I/O takes 1 ms.

□ Query 1

□ Query 2

Save Answer

Q2.4 Broadcast Teachers

1 Point

We want to join the `Teachers` table with the `Classes` table using the join condition: S.cid = C.id. The `Teachers` table has 90,000 pages and the `Classes` table has 90 pages. We have three machines: m1, m2, and m3. Currently the `Teachers` table is round robin partitioned across the 3 machines and the `Classes` table is all on m1. Each page is 1KB.

How many KB of data from the `Teachers` table will be sent across the network in a broadcast join?

Enter your answer here

Save Answer

Q2.5 Broadcast Classes

1 Point

How many KB of data from the `Classes` table will be sent across the network in a broadcast join?

Enter your answer here

Save Answer

Q2.6 Parallel Hash Classes

1 Point

How many KB of data from the `Classes` table will be sent across the network in a parallel hash join in the average case? Assume that we have perfect hash functions that divide up the data evenly.

Enter your answer here

Save Answer

Q2.7 Parallel Hash Teachers

1 Point

How many KB of data from the `Teachers` table will be sent across the network in a parallel hash join in the average case? Again assume that the data is distributed uniformly and that we have hash functions that divide up the data completely evenly.

Enter your answer here

Save Answer

Q2.8 Broadcast vs Hash

1 Point

If disk access cost and CPU cost are negligible compared to the network cost, which join should we pick to optimize for performance?

○ Broadcast Join

○ Hash Join

Save Answer

Save All Answers

Submit & View Submission >