

! This quiz has been regraded; your new score reflects 2 questions that were affected.

# Final Exam

**Due** May 4 at 7:10pm      **Points** 110      **Questions** 15

**Available** May 4 at 5pm - May 4 at 7:15pm about 2 hours      **Time Limit** None

This quiz is no longer available as the course has been concluded.

## Attempt History

	<b>Attempt</b>	<b>Time</b>	<b>Score</b>	<b>Regraded</b>
LATEST	<a href="#">Attempt 1</a>	132 minutes	98 out of 110	93 out of 110

Score for this quiz: **93** out of 110

Submitted May 4 at 7:12pm

This attempt took 132 minutes.

### Question 1

12 / 12 pts

Consider the following relation that describes a directed graph where the edges are colored: **Edges (start, end, color)**.

A **triangle** is a pattern of three directed edges  $a \rightarrow b$ ,  $b \rightarrow c$ ,  $c \rightarrow a$ . A triangle is **monochromatic** if all edges have the same color. Write a SQL query that counts the **non-monochromatic triangles** (i.e., the ones that have at least two edges with a different color).

Be careful: the triangles  $a \rightarrow b \rightarrow c \rightarrow a$ ,  $b \rightarrow c \rightarrow a \rightarrow b$  and  $c \rightarrow a \rightarrow b \rightarrow c$  are the same triangle and should be counted only once!

Your Answer:

```
SELECT Count(*) / 3
```

```
FROM Edges e1, Edges e2, Edges e3
```

WHERE e1.end = e2.start

AND e2.end = e3.start

AND e3.end = e1.start

AND (e1.color != e2.color

OR e2.color != e3.color

OR e3.color != e1.color);

## Question 2

6 / 10 pts

Consider a relational schema with one relation R(A, B) and the following query in Relational Algebra:

$$q = \pi_A(\sigma_{B=1}(R) \bowtie_{A=A'} \rho_{A \rightarrow A', B \rightarrow B'}(R))$$

Which of the following queries are equivalent to q? Choose all the correct options.

Correct Answer

$\pi_A(\sigma_{B=1}(R))$

$\sigma_{B=1}(\pi_A(R) \bowtie_{A=A'} \rho_{A \rightarrow A', B \rightarrow B'}(R))$

Correct!

$\pi_A(\sigma_{B=1}(R \bowtie_{A=A'} \rho_{A \rightarrow A', B \rightarrow B'}(R)))$

$\pi_A(R) - \pi_A(\sigma_{B \neq 1}(R))$

Correct Answer

$\pi_A(\sigma_{B=1}(R)) \cap \pi_A(R)$

## Question 3

5 / 5 pts

Consider the relation R(A, B, C, D, E, F) with functional dependencies:

A, B  $\rightarrow$  C

C  $\rightarrow$  D, E

C, F  $\rightarrow$  B

Which of the following are **keys** in R?

A, B

A, B, F

A, B, C

A, C, F

A, B, E, F

#### Question 4

8 / 8 pts

Explain your answer to the above question.

Your Answer:

A, B, F  $\rightarrow$  A, B, C, D, E, F and none of its subset implies all the attributes of the relation. So it is a valid key.

A, C, F  $\rightarrow$  A, B, C, D, E, F and none of its subset implies all the attributes of the relation. So it is a valid key.

A, B or A, B, C do not imply all the attributes of the relation.

A, B, E, F has a subset (which is A, B, F) that implies all the attributes of the relation. So it is a superkey but not a key.

**Question 5****10 / 10 pts**

We are given a relation R(A, B) with 100 pages, and a relation S(C, D) with 200 pages. In relation R, attribute A is the primary key and takes values 1, 2, 3 . . . . Each record in R is 40 bytes long, while each record in S is 10 bytes long. The size of a page is 1, 000 bytes.

How many pages do we need to store the output of the following SQL query?

```
SELECT *
FROM R, S
WHERE R.A = 1
```

**Correct!****Correct Answers**

1000

1,000

**Question 6****8 / 8 pts**

Consider a B+ tree index with order  $d = 5$ , fill factor  $F = 1$  and height  $h = 3$ . What is the total number of leaf pages in the tree?

**Correct!****Correct Answers**

1331

1,331

## Question 7

Original Score: 10 / 10 pts Regraded Score: 5 / 10 pts

⚠ This question has been regraded.

Consider the following extendible hash index. What is the smallest number of entries you can insert so that the global directory doubles in size?

[exhash-2.pdf](#)

[Minimize File Preview](#)



//

---

 1

---

 2

---

 3**Correct Answer****You Answered**

4**Question 8****0 / 5 pts**

We are given two relations: R with 20,000 pages and S with 10,000 pages. We are performing a key-foreign key join between R and S, wherein S has the foreign key attribute. Suppose that R is already sorted on the join attribute.

**What is the I/O cost of the Sort Merge Join algorithm** that uses replacement sort to create the initial runs? Do not count the cost of writing the join result to disk. Assume that the size of the buffer is 100 pages.

**DU Answered**

70000

**Correct Answers**

50000

50,000

**Question 9****7 / 10 pts**

Explain your answer to the above question.

Your Answer:

Since R is already sorted on the join attribute, we only need to sort S on the join attribute. Since we are using the replacement sort to create the initial runs, we will have  $10000 / (2 * 100) = 50$  runs after pass 0. Pass 0 uses  $2 * 10000 = 20000$  I/Os. 50 runs fit in the buffer, so we only need one pass to sort these runs, which is again 20000 I/Os. Once we have two relations sorted, given that it is a foreign key join, we only need  $M(R)$

+ M(S) = 20000 + 10000 = 30000 I/Os to complete SMJ algorithm. Add these I/Os together and it costs 70000 I/Os in total.

## Question 10

15 / 15 pts

Suppose we have a schema with two relations:

- R(A,B) has 2,000,000 tuples.
- S(C,D,E) has 3,000,000 tuples.

Consider the following SQL query:

```
SELECT COUNT(*)
FROM R, S
WHERE R.B = S.C AND R.E = 10 ;
```

Assume that:

- Each page holds 1,000 tuples
- The selectivity of the predicate (R.E = 10) is 1%
- The buffer has size 100 pages.

**What is the I/O cost of the best query plan?** Describe your answer in detail by presenting the annotated query plan (Here you just need to describe the sequence and implementation of each operator, and where you use materialization or pipelining).

Your Answer:

The best annotated query plan:

Firstly do a selection (R.E = 10) on relation R with a hash index or a B+ tree index. The intermediate result can be pipelined. Secondly do a join of R and S on R.B = S.C with BNLJ algorithm and count the number of tuples during BNLJ.

This plan will cost (3020 + OUT) I/Os.

Since the selectivity is 1%, there will be 20000 tuples after the selection, which will fit in  $20000 / 1000 = 20$  pages. By searching in a hash index or

B+ tree index, we will only need 20 I/Os to read these pages. These 20 pages will be pipelined. During BNLJ, since 20 pages fit in the buffer and are pipelined to the join operator, we only need 3000 I/Os to read the pages from S. We will be able to count the number of tuples during the process of BNLJ. In total, it is  $20 + 3000 + \text{OUT} = (3020 + \text{OUT})$  I/Os.

**Question 11**

4 / 4 pts

If the conflict graph of an interleaved schedule has no edges, then any serial schedule is equivalent to the original schedule.

Correct!

 True False**Question 12**

4 / 4 pts

If a transaction reads a data object after it has been written by an uncommitted transaction, then isolation is always violated.

 True False

Correct!

**Question 13**

3 / 3 pts

Consider the following SQL query:

```
SELECT *
FROM R
WHERE R.A = 1 AND R.B > 10 ;
```

A B+ tree index on (B,A) matches the selection predicate in the above query.

Correct!

True

False

### Question 14

3 / 3 pts

A bitslice index can be used to answer range queries efficiently.

Correct!

True

False

### Question 15 Original Score: 3 / 3 pts Regraded Score: 3 / 3 pts

ⓘ This question has been regraded.

Replacement sort can never create a sorted run with less than B-1 pages, where B is the size of the buffer.

DU Answered

True

orrect Answer

False

Quiz Score: **93** out of 110