AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTERS AND SYSTEMS ENGINEERING

Fall 2022	Course Code: *CSE346s*	*Time allowed: 2 Hrs.*

Advanced Database Systems

The Exam Consists of FIVE Questions in TWO Pages *Maximum Marks: 60 Marks* *1/2*

For the following COLLEGE database schema, answer the following questions:

D: Department (deptid, deptname)
S: Student (sid, sname, address, status, deptid)
P: Professor (pid, pname, address, deptid)
C: Course (crscode, title, description)
TR: Transcript (sid, crscode, grade)
T: Teaching (pid, crscode)

Question no. 1 (14 marks)

i. Write a SQL formula to find courses code and their grade for student "Ahmed". (3 marks)
ii. Write down Relational Algebra syntax for problem in (i). (3 marks)
iii. Tables information: Student 20 blocks; Transcript 30 blocks, Course 5 blocks, Transcript has sid as an index with h=3, consider blocking factor = 10 and $SF_{\bowtie} = 0.1$;
From this information calculate the cost for query in (i). (6 marks)
iv. If we want a minimum cost for this query, what you suggest (make any suitable assumptions). (2 marks)

Question no. 2 (10 marks)

i. Write a SQL formula to find all courses titles taught by Professor "Mohamed". (3 marks)
ii. Write down Relational Algebra syntax for problem in (i). (3 marks)
iii. Apply Dynamic algorithm to perform query optimization for query in (i). (4 marks)

Question no. 3 (8 marks)

Use the ODL language to represent the tables Student and Professor. Make sure that both student and professor are modeled as persons. Also illustrate the class of each.

Question no. 4 (14 marks)

We need a transcript with information of a certain student and courses taken. Do this task by using:

A) XML schema. (7 marks)
B) MONGODB. (7 marks)

Question no. 5 (14 marks)

A) We have a database file with size equal 800 blocks, and we want to sort it using external merge sort. Let the number of buffers in the memory equal 5. (4 marks)

i) How many initial runs does the DBMS need to perform to sort the file? 160
ii) What is the total I/O cost to sort the file? 2400
iii) What is the total number of block transfers for this algorithm? 8000
iv) What are the constraints for this algorithm to work? $B(R) \le M^2$

B) Choose either (TRUE) or (FALSE) and correct the FALSE one: (10 marks)

1. For any predicate p: $\sigma_p(R \cup S) \equiv \sigma_p(R) \cup \sigma_p(S)$.
2. Consider two relations: R(A,B) and S(A,B) with relational algebra: $\pi_{R.A}((R \cup S) - S)$; this formula is equivalent to $\pi_{R.A} R - \pi_{R.A}(R \cap S)$. F
3. The FOR clause of XQuery is used to declare variables and bind each variable to its range. T
4. CDATA is a term used about text data that will be parsed by the XML parser.

AIN SHAMS UNIVERSITY, FACULTY OF ENGINEERING
Department of Computer and Systems Engineering

Fall 2022	Course Code: *CSE 346s*	*Time Allowed: 2 Hrs.*
	Advanced Database Systems	
The Exam Consists of **Six** Questions in **Two** Pages.		**2/2**

5. NoSQL databases are built to allow the insertion of data without a predefined schema. T
6. To query data from MongoDB collection in a formatted way we use the command: >db.collection_name.find( ). F
7. ODBMS can store complex data types on the Web. T
8. In OODB, A list is an ordered collection of elements of the same type. T
9. Consider the database schema: R(A, B, C), S(B, E), T(A, D). The relational algebra query: $\sigma_{A=1 \text{ and } D>2}((R \bowtie S) \bowtie T)$ is equivalent to $(((\sigma_{A=1 \text{ and } D>2}(T)) \bowtie S) \bowtie R)$. F
10. MongoDB is a key-value database that provides high performance, high availability, and easy scalability. F

Consider the following data: Cost Estimation

SELECTION:

Linear search: Cost estimate = b_r block transfers;

If selection on a unique attribute: Cost estimate = $(b_r/2)$ block transfers

Binary search: $\lceil \log_2(b_r) \rceil + \lceil S/blf \rceil - 1$

Index scan: Cost = $(h_i + 1)$

Primary index on nonkey, equality: Cost = $h_i + n$

Equality on search-key of secondary index: Cost = $(h_i + 1)$ if search key is a candidate key;

Cost = $(h_i + n)$ if search key is not a candidate key; n for nonkey = no of records/ Distinct values

OR n= bl1/2 + r/2; as bl1 is the second index level

JOIN $R \infty S$

Nested-Loop Join: cost = $r_r * b_s + b_r$

Single loop index loop: $b_o + (r_o * (h_{i\text{-}in} + 1))$ as b_o ;r_o for outer relation; $h_{i\text{-}in}$ index of inner relation

Sort-merge join: cost= $b_s + b_r$

Intermediate Relation Sizes:

Selection: size(R) = card(R) × length(R); $card(\sigma_F(R)) = SF_\sigma(F) \times card(R)$

$SF_\sigma(A=value) = 1/card(\Pi_A(R))$

$SF_\sigma(A>value) = (\max(A) - value)/(\max(A) - \min(A))$

$SF_\sigma(A<value) = (value - \min(A))/(\max(A) - \min(A))$

$SF_\sigma(p(A_i) \wedge p(A_j)) = SF_\sigma(p(A_i)) * SF_\sigma(p(A_j))$

$SF_\sigma(p(A_i) \vee p(A_j)) = SF_\sigma(p(A_i)) + SF_\sigma(p(A_j)) - (SF_\sigma(p(A_i)) * SF_\sigma(p(A_j)))$

Projection: $Card(\Pi_A(R)) = card(R)$

Join: Special case: A is a key of R and B is a foreign key of S: $card(R \bowtie_{A=B} S) = card(S)$

More general: $card(R \bowtie S) = SF_\bowtie * card(R) * card(S)$

For external sort algorithm: number of initial run $n_R = \lceil (b_r/M) \rceil$;

total number of block transfers= $2\, b_r (\lceil \log_{dm} n_g \rceil + 1)$

$dm = M - 1$

===

END of Exam, Good Luck

Examination Committee *Exam. Date : 26th of Jan., 2023*

Prof. Hoda Korashy Mohamed