Final Exam Preparation:
Instructions:

- The Exam consists of two sections for 50 points max:
    - **Section A (20 points)** – objective and multiple choice, compare and contrast, and true/false questions
    - **Section B (30 points)** – Subjective content on topics provided below
- Attempt all sections as per instructions provided.
- Use space provided to write your answers.
- Grade Scope
- Show all your work. Partial credit will be allocated, if all intermediate steps are shown.
- Maximum time: 60 mins
- No Plagiarism

Topics:

Relational Algebra

- ☐ Given the relational algebra convert into its equivalent in SQL

SQL

- ☐ Explain the semantics of the SELECT query
- ☐ How aliasing is brought about in SQL?
- ☐ Writing queries for a given database.

SQL and Nested Queries

- ☐ What is a Sub-Query in SQL?
- ☐ What is the difference between Correlated and Uncorrelated subqueries?
- ☐ Why do we need Nesting in SELECT queries in SQL?
- ☐ Explain the use of Scoping rules in sub-queries?

Functional Dependencies

- ☐ The Armstrong's Axioms
- ☐ The Three Inference Rules of FD's and how you would prove them using the Armstrong's Axioms
- ☐ The difference between Trivial, Non-trivial, and completely non-trivial FD's
- ☐ Establishing Closure using FD's
- ☐ Establishing the cover of a set of FD's

Normal Forms

- ☐ What is normalization in DBMS?
- ☐ What are the different kinds of normal forms?
- ☐ What is the difference between a prime attribute and a non-prime attribute?
- ☐ What is BCNF?

Query Processing

- ☐ Steps of Query Execution
- ☐ Physical Query Plan operators and types
- ☐ Iterators and its stages
- ☐ One Pass algorithm

ALL LABS are included in the final.

Sample Questions:

SECTION A: Answer all questions (20 points)

1. Write the equivalent SQL expression for the following:
$$\pi_{sid}(\pi_{pid}(\sigma_{color=red\ or\ color=green}Parts) \bowtie catalog)$$

2. Using appropriate examples, explain the difference between the following pairs of terms:
   a. Prime attribute and Primary Key
   b. 2NF and 3NF
   c. Physical Query Plan and Logical Query Plan
   d. Correlated Queries and Uncorrelated Queries
   e. Pseudo Transitivity axiom and Transitivity axiom
   f. Trivial FD and Non-Trivial FD

3. Consider the schema:
   - Suppliers (sid: integer, sname: string, address: string)
   - Parts (pid: integer, pname: string, color: string)
   - Catalog (sid: integer, pid: integer, price: real)

   The keys for each relation in the schema are underlined, and the domain of each field is listed after the field name. The *Catalog* relation lists the prices charged for parts by *Suppliers*.

   Write the following queries in SQL.

   a. Using correlated subqueries, list the sids of suppliers who supply red colored parts "or" who's store are located in "Ruston City".
   b. List pairs of sids such that the supplier with the first sid charges more for some part than the supplier with the second sid.

4. State true or false:
   a. The logical query plan results in the creation of a parse tree
   b. Index scan is a physical query plan operator uses an index to read blocks of a data file.
   c. To achieve query optimization, select statements are pushed as far down the parse tree as possible to reduce the number of branches in the parse tree.
   d. An iterator in the physical query plan comprises of three steps (a) Open(), (b) GetNext(), (c) Close()
   e. To achieve query optimization, extra project statements are added near the leaves of the parse tree to reduce the size of tuples going up the tree.

SECTION B: Answer any 2 out of 3 (30 points)

1. Compute the closure for relation schema R (A, B, C, D, E), given the set of FD's: F = {A → BC, CD → E, B → D, E → A}, and List the candidate keys for R.

2. For each of the following transformations expressed in relational algebra, give an intuitive explanation of what it means and determine its validity.
   a. $\pi_{L_1}\left(\pi_{L_2}(R)\right) = \pi_{L_1}(R)$
   b. $(R \times S) \times T = R \times (S \times T)$
   c. $\sigma_{c1}(\sigma_{c2}(R)) = \sigma_{c2}(\sigma_{c1}(R))$

3. What are the factors that affect the cost of a query?
4. Build the parse trees for the following queries, and optimize them using the "rewrite rules". Given the relations: MOVIE(title, year, director, language), and Starsin(title, year, StarName, language),
   a. `SELECT starname FROM GermanMovies NATURAL JOIN Starsin WHERE Director = 'ABCDEF'`
      [Note GermanMovies is a view as shown below:
      `CREATE VIEW GermanMovies AS SELECT * FROM Movie WHERE language = 'German';`
   b. `SELECT Dept_Name FROM Dept, Manager WHERE Manager.DNO=DEPT.DNO and Manager.Salary = 50000;`
5. As discussed in class, determine the 3NF for the following relation and set of FD.

Student(SSN, SName, Saddress, HScode, HSname, HScity, GPA, priority)

FD = {HScode → HSname, HScity, GPA→ Priority, SSN → SName, Saddress, HScode, HSname, HScity, GPA, priority}.

6. Writing of Triggers, and SQL – correlated and uncorrelated nested queries.
7. Explain the steps involved during query execution?
8. Explain the strategy adopted using one-pass algorithm to execute the following binary operations, Assuming R is the bigger relation:
   a. $R \cap S$
   b. $R - S$
   c. $S - R$