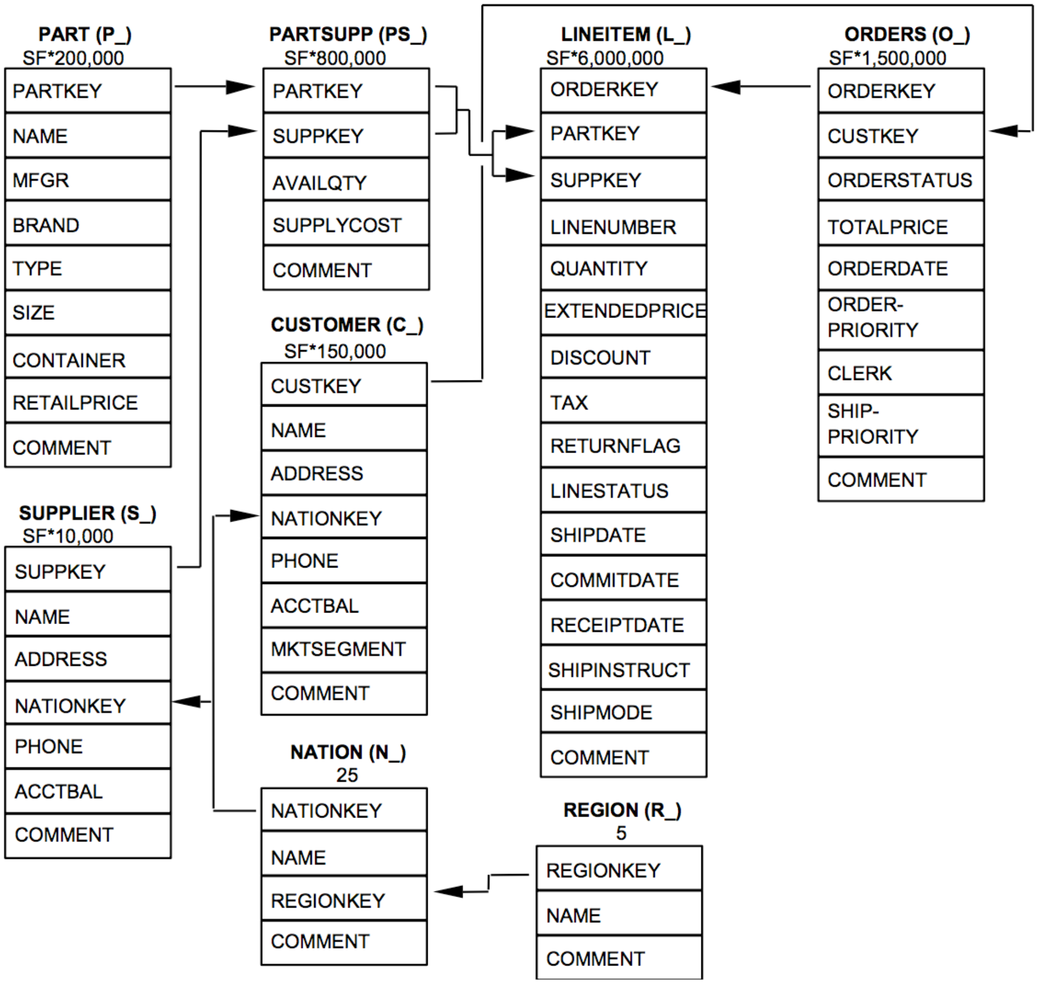
Programming Assignment 3 - 100 points

Due Date: September 19, 2025, 11:59 pm

Objective:

The goal of this assignment is to continue practicing writing SQL queries. Reading and writing SQL queries is critical component of interviews. Read and understand the following ER diagram. It shows the scenario of customers placing orders and those orders have lineitems (parts) and each of these lineitems are provided by suppliers in certain region and nation.

Figure 2: The TPC-H Schema



Legend:

- The parentheses following each table name contain the prefix of the column names for that table;
- The arrows point in the direction of the one-to-many relationships between tables;
- The number/formula below each table name represents the cardinality (number of rows) of the table. Some are factored by SF, the Scale Factor, to obtain the chosen database size. The cardinality for the LINEITEM table is approximate (see Clause 4.2.5).

Write SQL queries to get the following answers:

1. Create the tpc_h schema and create the tables. Then load the data into those tables. The database script is provided as part of the assignment. [10 points]
   The script file is named: tpch_database.sql
   There is a zip file named: tpc_h_data.zip. This zip file contains the data files for each of the tables. These files are | delimited.

2. Basic SELECTs and Filters
   a) List all parts with a retail price above 100. Return partkey, part name and retail price. [5 points]

3. Basic Joins
   a) Get the part name and supplier name for all part-supplier combinations. Return partname, suppliername. [10 points]
   b) Get all the parts where their supplycost is less than their retail price. Return the part_name, part_type, retailprice and supplycost. [10 points]

4. Aggregations and grouping
   a) Total number of orders per customer. Return customerkey, customername, numorders. [10 points]
   b) Which nation has the maximum number of customers. Return only 1 record/row. Return the name of the nation, numCustomers. [10 points]

5. Multi-table Joins
   a) Total revenue per nation order by the total revenue in descending. Return nation name, total_revenue. [15 points]
   b) Top 5 parts by number of suppliers order by numsuppliers in descreasing order followed by the partname. Return partname, numsuppliers. [15 points]
   c) Total order value from customers in each region. Return regionname, total_order_value order by total_order_value in descending. [15 points]

Tasks:

1. Submit a one single .sql file that contains all the 8 SQL queries in order of the questions. The queries should have a comment above them indicating which question that query is for. E.g. -- 4.a.
   Comment in SQL is prefixed by two dashes and a space.

*** There is no partial grading of the queries. If the query returns the correct answer, you get all the points allocated for that query, else you get 0 points for that query. Your Query will be tested against the same exact data from the tpc_h_data.zip file.***