Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)

Course-2: Analysing Big Data with SQL

Week-6: Core

Question-1: The questions in this quiz intentionally use tables that are not in the VM. You should be able to answer the questions without running any queries.
Answer-1: I acknowledge that I do not need to run any queries for the following questions. I will not be able to run them because the tables do not exist on the VM.

Question-2: Which of these queries produces the same result set as the following query?
SELECT * FROM table1
UNION
SELECT * FROM table2;
Answer-2: SELECT * FROM table1 UNION DISTINCT SELECT * FROM table2

Question-3: Choose the best query to run in Impala to return the distinct union of the columns zip_plus_4 (type STRING, has values like '94306-0001') in the california_emp table and zip (type INT, has values like 94105) in the california_offices table.
Answer-3: SELECT zip_plus4 AS zipcode FROM california_emp UNION DISTINCT SELECT CAST (zip AS STRING) AS zipcode
FROM california_offices;

Question-4: The zip column (type INT) in the california_offices table has values from 90001 to 95899. The zip column (also type INT) in the oregon_offices table has values from 97030 to 97440. Which value is guaranteed to be in the top row of the result set when you run the following query with Impala?
SELECT zip FROM california_offices
UNION ALL
SELECT zip FROM oregon_offices
ORDER BY country DESC;
Answer-4: No particular value is guaranteed in the top row

Question-5: The california_offices table has 65 rows, and the oregon_offices table has 5 rows. How many rows does the following query return when you run it with Impala? SELECT zip FROM california_offices
UNION ALL
SELECT zip FROM oregon_offices
LIMIT 2;
Answer-5: 67

Question-6: The california_offices and california_emp tables each have a column named office_id. All other columns have unique names between the two tables. Which of the following are valid join queries that Impala would run successfully on the VM, if these tables existed on the VM? Check all that apply.
Answer-6: SELECT name, california_emp.office_id, city, salary
FROM california_emp JOIN california_offices
ON california_emp.office_id = california_offices.office_id;
SELECT name, e.office_id AS office_id, city, salary FROM california_offices AS o
JOIN california_emp AS e ON o.office_id = e.office_id;

Question-7: Which of the following are valid join queries for Impala? Check all that apply.
Answer-7: SELECT o.office_id as office, AVG(salary) AS avg_salary
FROM california_emp e
JOIN california_offices o ON o.office_id = e.office_id GROUP BY office
ORDER BY avg_salary;

SELECT o.office_id AS office, COUNT(*) AS number_of_employees FROM california_emp e
JOIN california_offices o ON o.office_id = e.office_id
GROUP BY office;

SELECT name, o.office_id AS office, city
FROM california_emp e
JOIN california_offices o ON o.office_id = e.office_id ORDER BY office DESC, city DESC;

Question-8: The california_emp table includes one row with name='Sandy Tilbrook', with office_id='CA086'. There is no row in california_offices with office_id='CA086'. However, there is a office_id='CA070' in california_offices with city='Redding', but no rows in california_emp have office_id='CA070'. (There are no other rows with city='Redding'.) Choose the response that best describes how these rows will be included in the result set of this query:
SELECT name, city, salary
FROM california_emp e
INNER JOIN california_offices o ON e.office_id = o.office_id;
Answer-8: No row with name='Sandy Tilbrook' will be included, and no row with city='Redding' will be included

Question-9: Which FROM clauses could you use to return data about all the employees in california_emp, even the remote workers who are not assigned to an office (office_id=NULL) or those erroneously assigned to a non-existent office? Select all that apply.
Answer-9: FROM california_offices o RIGHT OUTER JOIN california_emp e ON e.office_id=o.office_id
FROM california_emp e LEFT OUTER JOIN california_offices o ON e.office_id=o.office_id

Question-10: Which of the following queries returns only the employees whose office IDs do not match any office IDs found in the offices table?
Answer-10: SELECT empl_id, name
FROM california_emp e
LEFT OUTER JOIN california_offices o ON e.office_id = o.office_id WHERE o.office_id IS NULL;