

Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)

Course-2: Analysing Big Data with SQL

Week-5: Core

Question-1: The fly.flights table has the following schema: Choose the valid SELECT statements. Check all that apply.

column	type
year	smallint
month	tinyint
day	tinyint
dep_time	smallint
sched_dep_time	smallint
dep_delay	smallint
arr_time	smallint
sched_arr_time	smallint
arr_delay	smallint
carrier	string
flight	smallint
tailnum	string
origin	string
dest	string
air_time	smallint

Answer-1: SELECT carrier, COUNT(*) FROM fly.flights GROUP BY carrier ORDER BY carrier;
SELECT * FROM fly.flights ORDER BY distance;

Question-2: Select all the statements that return the same result as SELECT * FROM flights ORDER BY carrier;

Answer-2: SELECT * FROM flights ORDER BY carrier ASC;

Question-3: Suppose you want to find the longest-distance flights in the fly.flights table for a particular carrier, and then find the flights with the shortest air time.

Write a query to return the data in fly.flights for American Airlines (carrier is AA) so that they are sorted by distance with the longest distance first, and for those that tie distances, by air_time with the shortest air time first. Execute the query in Hue using Impala. What's the shortest air time for the longest distance?

Answer-3: 411

Question-4: Write and run a SQL query to determine which airport in the fly.airports table is closest to the geographical (not magnetic) North Pole, using the following calculation for the distance in kilometers, using the latitude (lat) column: $\text{distance} = 6371 * 2 * \text{asin}(\text{least}(1, \sin(\text{radians}(90 - \text{lat}) / 2)))$

(Note: The least function chooses the minimum value among two or more scalar values — similar to the MIN function, but MIN works on values in a column.)

Which airport is closest to the geographical North Pole?

Answer-4: Wiley Post Will Rogers Memorial Airport

Question-5: Select the queries that will return exactly the same result as the query: `SELECT * FROM fly.planes ORDER BY year DESC;`

when executed by Impala. Check all that apply.

Answer-5: `SELECT * FROM fly.planes ORDER BY year DESC NULLS FIRST;`

Question-6: Select the queries that will run without error in Hive. Check all that apply.

Answer-6: `SELECT model, year FROM fly.planes ORDER BY 2019-year` `SELECT model, type FROM fly.planes ORDER BY type;`

`SELECT * FROM fly.planes ORDER BY type;`

`SELECT model, 2019 - year AS age_in_2019 FROM fly.planes ORDER BY age_in_2019;`

Question-7: Select the valid SQL queries. Check all that apply.

Answer-7: `SELECT arr_time, AVG(arr_delay) AS avg_arr_delay`

`FROM flights WHERE origin = 'LAX' GROUP BY arr_time`

`HAVING avg_arr_delay > 45`

`LIMIT 1000;`

Question-8: Which clause should you use with Impala to return rows 1001 through 1050 of a result set?

Answer-8: `LIMIT 50 OFFSET 1000`

Question-9: Select the appropriate uses for the LIMIT clause. Check all that apply.

Answer-9: Reduce the compute resources used by the SQL engine

Protect against returning an unexpectedly large number of rows Return a few rows from a table to inspect some of the values

Question-10: In what order does a SQL engine execute the clauses of a SELECT statement?

Answer-10: FROM, WHERE, GROUP BY, HAVING, SELECT, ORDER BY, LIMIT