Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)

Course-3: Managing Big Data in Cluster and Cloud Storage

Week-3: Core

Question-1: Which of these data types allows the greatest precision?
Answer-1: DECIMAL

Question-2: Which of the following are character string data types used by Hive and Impala?
Answer-2: STRING
- CHAR
- VARCHAR

Question-3: Which of the following are numeric data types?
Answer-3: DOUBLE
- BIGINT

Question-4: Which of these are good advice for choosing data types, assuming you want flexibility with both Hive and Impala?
Answer-4: Choose the smallest integer type that accommodates the required range
- Use the STRING type rather than CHAR or VARCHAR, for general use

Question-5: Which statements about the typeof function are true?
Answer-5: You can use it with expressions as well as column references You can use it with Impala

Question-6: You have a column of integer values that you expect to range from -10 to 127. Using Impala, what's the smallest integer data type you can use that allows you to identify which values are out of this range? The ranges of the integer data types are provided below as a reminder.

Integer Type	Range
TINYINT	-128 to 127
SMALLINT	-32,768 to 32,767
INT	-2,147,483,648 to 2,147,483,647 (approximately 2.1 billion)

Answer-6: SMALLINT

Question-7: Which are advantages of using Parquet files?
Answer-7: Columnar format (which improves performance for some access patterns)
- Good interoperability (can be used by many applications)

Question-8: Why is interoperability an important consideration when choosing a file type for storing data?
Answer-8: You might need to use different tools to access the same data

Question-9: Which are important considerations when choosing a file type for storing data?
Answer-9: Data size and query performance
- Ingest pattern
- Data lifetime