Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)

Course-1: Foundations of Big Data Analysis with SQL

Week-4: SQL Tools for Big Data Analysis

Question-1: You need a database system for a company with a massive physical warehouse operation. They want to keep track of their inventory, recording every shipment in and every shipment, with well-defined descriptions of each item and where they are stored. They also want to be able to analyse their operations to answer questions such as whether certain items should be stored closer together, or how often a particular item sells out. Which of the following would be the best choice for this?
Answer-1: A non-transactional operational system designed for structured data, such as Apache Kudu.

Question-2: You need a database system for a library of millions of large text documents, to help users find the documents that contain the information they need. Which of the following would be the best choice for this?
Answer-2: A search system, such as Cloudera Search or Elastic-search.

Question-3: Which of the following are features of SQL on RDBMSs that are also kept for working with big data systems? Check all that apply.
Answer-3: SELECT statements
- Seeing data as tables with column names

Question-4: Which of the following is the reason why we lose many features of SQL when moving from traditional RDBMSs to big data systems?
Answer-4: Many of the lost features require transactions, which are notoriously challenging for big data systems and not typically implemented

Question-5: Which of the following are features of SQL for working with traditional RDBMSs that we lose when moving to working with big data systems? Check all that apply.
Answer-5: Foreign key constraints
- Database triggers and stored procedures

Question-6: Which of the following are features of SQL for working with big data systems that are not typically found in SQL for traditional RDBMSs? Check all that apply.
Answer-6: Complex Data Types

Question-7: A company has a small on-premises cluster that they are rapidly outgrowing, and they are considering switching to cloud storage, or maintaining a hybrid solution. The following describes some factors going into their decision. Which are reasons that potentially support using a cloud cluster rather than an on-premises cluster? (Note that a hybrid option might still be best!)
Answer-7: The company hopes some upcoming new products will drastically increase their storage needs, though their processing needs probably will increase less dramatically
- The company's assets and budgets afford extra room for operating expenditures, but they are trying to keep capital expenditures to a minimum

Question-8: Which of the following accurately describes how the data dictionary in a traditional RDBMS being tightly coupled to the data is different from the table definitions in a big data system being loosely coupled to the files? Check all that apply.
Answer-8: The contents of the data dictionary accurately describe every table in a RDBMS, while the table definitions in a big data system describe what is expected in some files, but even those files may not match exactly.
- The data dictionary governs what is stored as data in a RDBMS, while the files in a big data system are completely ungoverned.