

Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)

Course-1: Foundations of Big Data Analysis with SQL

Week-3: Big Data

Question-1: Suppose you want to store a petabyte of data, and you want to run a report that requires reading and processing 250 terabytes of that data. What is a key difference in the technology you'll use for this, versus a need to store, process one or two megabytes of data.

Answer-1: Distributed storage and processing

Question-2: The following are records in a contact list.

```
{'name';'Étienne', 'email';'etienne@example.com', 'mobile';'555-8372'}
```

```
{'name';'Brayden', 'home';'555-2202', 'work';'555-2800'}
```

```
{'name';'Diana', 'mobile';'555-6575', 'email';'dprince@example.com'}
```

Is this contact list an example of structured, semi-structured, or unstructured data?

Answer-2: Semi-structured

Question-3: Is an online credit card or bank account statement an example of structured, semi- structured, or unstructured data?

Answer-3: Structured

Question-4: You plan to gather data from various sources. Which of the following sources do you think will definitely give you structured data?

Answer-4: Tables you capture from another relational database system

- A survey in which every question is a rating from 1 to 5

Question-5: Which of the following describe a reason why RDBMSs are a poor choice for big data? Check all that apply.

Answer-5: The structured nature of RDBMSs imposes costs in terms of storage and processing, which becomes prohibitive with really big amounts of data.

- Because a large amount of unstructured data would need to be stored as a BLOB or CLOB, RDBMSs provide little to no support for working with such data.

Question-6: Look at the following data:

Now imagine that, instead of four rows, you have 4000 rows, and all are similar to the rows you see here. Which of the following questions can you answer from this data? Check all that apply.

id	name	grade_level	gpa	age
930	Olufunmilayo Ayton	11	4.00	16
667	Vincent Michaelson	10	2.53	15
907	Asa Quigg	10	3.57	
168	Kiran Patil	11	3.28	17

Answer-6: What is the number of students in the table?

- What is the number of students in each grade level?
- What are the names of all the students at the same grade level as Kiran Patil?

Question-7: Consider the following data (in this case, a list of JSON objects): {'shop':'Dicey', 'game':'Monopoly', 'qty':7, 'aisle':3, 'price':17.99} {'shop':'Dicey', 'game':'Clue', 'qty':3, 'price':9.99}

{'shop':'Board Em', 'game':'Monopoly', 'qty':11, 'aisle':2, 'price':25.00} {'shop':'Board Em', 'game':'Candy Land', 'qty':4, 'aisle':2}

{'shop':'Board Em', 'game':'Risk', 'qty':7, 'aisle':3, 'price':35.00} {'shop':'Board Em', 'game':'Stratego', 'qty':'low stock'}

Which of the following questions can you definitely answer from this data? (Hint: Take note of missing values and inconsistent data types, which would make the answers unknown or uncertain.)

Answer-7: What is the price of Risk at the Board Em shop?

What are the games that start with the letter C?

Question-8: Which of the following questions could be answered quickly and easily by treating the complete plays of Shakespeare as data, separated by title and type (tragedy, comedy, or history)? Check two answers.

Answer-8: Which of the tragedies includes, or mentions, someone named Lucilius?

- How many plays are histories?