





1. Cloud computing and data transformation

Question:

A cloud data analyst is working with a large dataset. They decide to transform and analyze the dataset using cloud computing resources. How will this action help the cloud data analyst?

-  They will avoid the use of enterprise-grade level tools.
-  They will avoid all resource challenges in the transformation process.
-  **They will have access to the required computational power virtually.**
-  They will have unrestricted access to their company data.





Feedback:

Cloud computing provides scalable and on-demand computational power, which is ideal for processing large datasets. It doesn't eliminate all resource challenges or bypass enterprise tools, but it does offer virtual access to powerful infrastructure.

2. Aggregation error

Question:

A cloud data analyst is working with a large dataset. They make a mistake and choose the wrong aggregation method. Which type of errors will this introduce to the data?

-  Bulk reading process errors
-  Bulk transformation errors
-  **Data transformation process errors**
-  Data entry process errors

Feedback:

Choosing the wrong aggregation method affects how data is transformed and summarized, leading to **data transformation errors** that can misrepresent insights.

3. Aggregation for storage optimization

Question:

A cloud data analyst is tasked with optimizing the storage space of a large dataset from a music streaming service. The dataset generates a data point with the time and day every time a song is played. The streaming service only wants to know how many times per day a song is played. How can data aggregation help the cloud data analyst optimize the storage space to have only one data point?

- ☐ By aggregating the total plays per song per customer
- ☐ By aggregating the total songs played in a day
- ☐ By aggregating the total plays per song by hour
- ☒ **By aggregating the total plays per song by day**

Feedback:

Aggregating by **song and day** reduces the granularity of the data, saving storage while still providing the needed insight.

4. Charting aggregated data

Question:

A cloud data analyst is working with a dataset that tracks the use of a website. As a next step, they aggregate the data and use this information to create charts. What is the cloud data analyst doing?

- ☐ Evaluating customer preference
- ☐ Analyzing network traffic
- ☐ Comparing dataset data
- ☒ **Finding trends in the data**

Feedback:

Aggregated data visualized in charts helps identify **trends**, such as user behavior over time or peak usage periods.

5. Data duplication

Question:

In data analysis, what is the term for an exact copy of a record?

- ☐ Backup data
- ☒ **Duplicated data**
- ☐ Clean data
- ☐ Aggregated data

Feedback:

An **exact copy** of a record is called a **duplicate**, which analysts often need to detect and remove to ensure data quality.

6. Identifying partial duplicates

Question:

You are working as a data analyst deduplicating data. Based on the information in the following table, which is a partial duplicate record?

Name	Last name	Age	Address	Phone
Daniel	Brown	58	198 Maple St.	545-656-7894

Options:

- ☒ **Name: Daniel, Last name: Brown, Age: 58, Address: 189 Mapel St., Phone: 545-656-7894**
- ☐ Name: Daniel, Last name: Bern, Age: 58, Address: 442 Main St., Phone: 545-67
- ☐ Name: Darlean, Last name: Brown, Age: 45, Address: 198 Maple St., Phone: 545-656-7894
- ☐ Name: Daniel, Last name: Brown, Age: 58, Address: 198 Maple St., Phone: 545-656-7894

Feedback:

The first option is a **partial duplicate** due to a typo in the address ("Mapel" vs "Maple") but matches on other fields.

7. Join types

Question:

A cloud data analyst is joining table A and table B. What type of join should the cloud data analyst use to obtain both matched and unmatched rows from one or both tables?

- ☒ **Outer join**
- ☐ Left inner join
- ☐ Inner join
- ☐ Center outer join

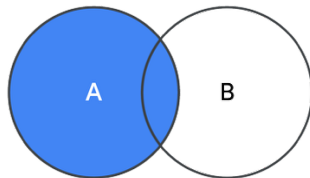
Feedback:

An **outer join** includes all matched and unmatched rows from both tables, making it ideal for comprehensive data merging.

8. Venn diagram join type

Question:

Which join method does the following Venn diagram illustrate?



- ☒ **Left outer join**
- ☐ Inner join
- ☐ Right outer join
- ☐ Center join

Feedback:

A **left outer join** returns all records from the left table (A), including matches from the right table (B) where available.

9. Purpose of an algorithm in data derivation

Question:

A cloud data analyst is tasked with a data problem. As a part of the process, they decide to use data derivation and create an algorithm. What is the purpose of the algorithm?

- ☐ Perform mathematical calculations
- ☒ **Transform the data into new data**
- ☐ Delete the duplicates from the dataset
- ☐ Split the dataset into two small datasets

Feedback:





In **data derivation**, algorithms are used to **generate new data** from existing data, often by applying rules or formulas.

10. Challenges with data derivation aging

Question:

A cloud data analyst is working with derived data. To gain the insight they need, they

will need to use data derivation aging. What is a problem they can encounter by doing this?

-  **The data may not be accurate**
-  The data may be inaccessible
-  The data may have duplicates
-  The data may be aggregated

Feedback:

Aging derived data can lead to **accuracy issues**, especially if the data becomes outdated or no longer reflects current conditions.