

Scalero - Data Engineer

This test is designed to assess the skills and competencies of ideal candidates for the Data Engineer position at Scalero. It will evaluate knowledge in the development of analytical products, pipelines, and presentation. Additionally, technical skills such as programming languages, queries and infrastructure will be evaluated.

It is essential that you can effectively communicate each stage of this test in your own words to evaluate your communication skills with future stakeholders.

The following link contains test data ready to be processed.

(<https://www.kaggle.com/datasets/devarajv88/target-dataset/data?select=customers.csv>),

Steps to take:

- Import the provided data into a relational database of your choice. MySQL, Postgres, etc.
- Design an efficient ETL (Extract, Transform, Load) pipeline to ingest, process, and store the dataset for further analysis. It's not necessary to implement an ETL at this stage; a proposal is enough.
- Identify missing values, inconsistencies, or duplicates in the dataset and propose strategies for cleaning.
- Add recommendations on efficient data storage and retrieval.

Prepare a presentation for the following test.

- Analyze store performance, top stores, bottom stores. Try and explain why some might be more successful than others.
- Provide sales projections for future years.
- Predict future demand for products based on historical sales.
- Add any additional insights you find relevant while analyzing the data.
- Share the presentation along with any queries written to analyze and build visualizations.