UDACITY

Data Engineering Capstone

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Hey,

Congratulation, the project is very detailed and nicely presented. The project is meeting all expectation and passes all the rubrics.

The detailed write up is an absolute delight to read, the presentation is very clear. Hallmark of any great project is the ability of the project to be able to reproduce by anyone else, and the write up like yours definitely helps.

Tip: If you are interested in knowing more about data modeling in the real world, please read this series of posts and this experience sharing

Further Readings

- If you are interested in knowing more about the heavy-read system design, I would suggest that you read this tutorial about Netflix system.
- I would suggest that you read this experience sharing about the lesson learned when implementing data pipeline

Write Up

The write up includes an outline of the steps taken in the project.
The purpose of the final data model is made explicit.

The write up describes a logical approach to this project under the following scenarios:

- The data was increased by 100x.
- The pipelines would be run on a daily basis by 7 am every day.
- The database needed to be accessed by 100+ people.

The choice of tools, technologies, and data model are justified well.

Execution

All coding scripts have an intuitive, easy-to-follow structure with code separated into logical functions. Naming for variables and functions follows the PEP8 style guidelines. The code should run without errors.

The project includes at least two data quality checks.

- The ETL processes result in the data model outlined in the write-up.
- A data dictionary for the final data model is included.
- The data model is appropriate for the identified purpose.

I liked the way the write up is presented, I can see there is very detailed step by step ETL process, the data dictionary is very comprehensive and more than serves the purpose

The sample queries provided helps to validate the model and confirms that the data model is able to answer the questions that it was created to do so

The project includes:

- At least 2 data sources
- More than 1 million lines of data.
- At least two data sources/formats (csv, api, json)

Your project meets all the requirements.👏

⤓ DOWNLOAD PROJECT

Rate this review

START