UDACITY

‹ Return to "Data Engineering Nanodegree" in the classroom

DISCUSS ON STUDENT HUB ›

Data Modeling with Cassandra

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Dear student,

Congratulations on passing this project! 💯 👏🏻 I believe you've learned a lot through this challenging project.

Wish you the best in your future Data endeavors!

ETL Pipeline Processing

✓

Student creates `event_data_new.csv` file.

Good job! I see the event_data_new.csv file, which indicates you followed the ETL pipeline to create the csv file.

✓

Student uses the appropriate datatype within the `CREATE` statement.

Nice work! You are using the correct datatype, such as int for the itemInsession and the ids. You are also using the correct string datatype for the relevant columns as all of your INSERT statements are processing the data correctly.

Data Modeling

✓

Student creates the correct Apache Cassandra tables for each of the three queries. The `CREATE TABLE` statement should include the appropriate table.

Great job! You followed the one table per query rule of Apache Cassandra. You are not replicating the same table for all three queries, which defies that rule.

✓

Student demonstrates good understanding of data modeling by generating correct SELECT statements to generate the result being asked for in the question.

The SELECT statement should NOT use `ALLOW FILTERING` to generate the results.

Suggestion

The rubric requires that "Student demonstrates good understanding of data modeling by generating correct SELECT statements to generate the result being asked for in the question."

Here, instead of `select *`, you should just select the information that is asked for.

✓

Student should use table names that reflect the query and the result it will generate. Table names should include alphanumeric characters and underscores, and table names must start with a letter.

Nice! the table names reflect the query and the result it generates.

✓

The sequence in which columns appear should reflect how the data is partitioned and the order of the data within the partitions.

Excellent work!

This is probably one of the most important learning that I want you to walk away with from this lesson. Apache Cassandra is a partition row store, which means the partition key determines which node a particular row is stored on. In case of composite partition key, partitions are distributed across the nodes of the cluster and how they are chunked for write purposes. Any clustering column(s) would determine the order in which the data is sorted within the partition.

PRIMARY KEYS

✓

The combination of the PARTITION KEY alone or with the addition of CLUSTERING COLUMNS should be used appropriately to uniquely identify each row.

Excellent work at understanding and then implementing the COMPOSITE PRIMARY KEY to uniquely identify each row.

Presentation

✓

The notebooks should include a description of the query the data is modeled after.

Good work including headers to denote descriptions of the query! It was clear how each query was being addressed.

✓

Code should be organized well into the different queries. Any in-line comments that were clearly part of the project instructions should be removed so the notebook provides a professional look.

Good Job!
The Notebook code is well formatted! 👍🏽

DOWNLOAD PROJECT

RETURN TO PATH