Data Modeling with Cassandra

REVIEW

HISTORY

Meets Specifications

Congratulations on completing the project! 👏 You should be very proud of your accomplishments in building a Data Model with Cassandra. You'll find feedback and some tips to help you continue to improve on your project.
If you have any query please use knowledge board, we are happy to help!! If you find the review helpful please drop a thumbs up thanks!!
Good Luck in your Data Engineering Journey 😄

ETL Pipeline Processing

Student creates `event_data_new.csv` file.

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

- You can refer to this link in case you want to check out other data types also.

Data Modeling

Student creates the correct Apache Cassandra tables for each of the three queries. The `CREATE TABLE`

statement should include the appropriate table.

- Check out this link for some basic rules for Cassandra Data modeling.

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.

- Good job with the select queries, You are writing relevant column names instead of using `SELECT *`. Good job doing that as it is a best practice to write the names of the columns.
  Extra Reading:
- The reason why we are not using ALLOW FILTERING is that it is a very inefficient method. Check this link to read in detail about when allow filtering can be used.

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.

- Good job choosing relevant names for all the tables

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

Some DataStax documentation you can refer to:
https://docs.datastax.com/en/dse/5.1/cql/cql/cql_using/whereClustering.html

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.

- Brilliant job here! Keys for all tables are correct.
- You can refer to this link to read about performance tuning in Cassandra
- You can check this link to read about different types of keys in cassandra.

Presentation

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

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.

- TODO comments are removed correctly.

DOWNLOAD PROJECT

RETURN TO PATH