

PROJECT SPECIFICATION

Data Modeling with Cassandra

ETL Pipeline Processing

| CRITERIA | MEETS SPECIFICATIONS |
|---|---|
| Student completes the ETL pipeline procedures. | Student creates <code>event_data_new.csv</code> file. |
| Student uses the correct datatype for each Cassandra <code>CREATE</code> statement. | Student uses the appropriate datatype within the <code>CREATE</code> statement. |

Data Modeling

| CRITERIA | MEETS SPECIFICATIONS |
|---|---|
| Student creates correct data models for the queries they need to run. | Student creates the correct Apache Cassandra tables for each of the three queries. The <code>CREATE TABLE</code> statement should include the appropriate table. |
| Student can set up the data model correctly to generate the exact responses posed in the questions. | 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 <code>ALLOW FILTERING</code> to generate the results. |
| Student models the data by using appropriate table names. | 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. |
| Student has given careful thought to how the data is modeled in the table and the sequence and order in which data is partitioned, inserted and retrieved from the table. | The sequence in which columns appear should reflect how the data is partitioned and the order of the data within the partitions. |

PRIMARY KEYS

| CRITERIA | MEETS SPECIFICATIONS |
|---|---|
| The PRIMARY key for each table should uniquely identify each row in each of the tables. | The combination of the PARTITION KEY alone or with the addition of CLUSTERING COLUMNS should be used appropriately to uniquely identify each row. |

Presentation

| CRITERIA | MEETS SPECIFICATIONS |
|---|---|
| Student provides responses to the questions. | The notebooks should include a description of the query the data is modeled after. |
| Students notebook code should be clean and modular. | 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. |

Suggestions to Make Your Project Stand Out!

- You can add description of your PRIMARY KEY and how you arrived at the decision to use each for the query
 - Use Panda dataframes to add columns to your query output
-