

[Return to "Data Engineering Nanodegree" in the classroom](#)

[DISCUSS ON STUDENT HUB](#)

# Data Modeling with Cassandra

## REVIEW

## HISTORY

### Requires Changes

4 SPECIFICATIONS REQUIRE CHANGES

You've done a good job on this project!

Please make the necessary changes to complete the project.

Looking forward to reviewing your next submission!

### ETL Pipeline Processing



Student creates `event_data_new.csv` file.

Nicely done correctly creating the `event_data_new.csv` file



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

Great job using appropriate datatypes within the `CREATE` statement:

- The `artist_name` and `song_title` do have the `text` / `varchar` datatype

- The `length` does use a `float` / `double` datatype

## Data Modeling



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

Good job creating correct Apache Cassandra tables designed to be used to answer the questions in the prompt by adhering to the one table per query rule of Apache Cassandra



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.

Well done selecting specific columns from the table without having to call `SELECT *`, which creates unnecessary performance overhead



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.

The tables are named as 'musicapp\_history1', 'musicapp\_history2' and 'musicapp\_history3'. Please change the table names in such a way that it reflects the query and the results it will generate. For example, for query 2, an appropriate table name should reflect song playlist in session, such as `song_playlist_session`.



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

Nicely done following a the order of the COMPOSITE PRIMARY KEY for the `CREATE` and `INSERT` statements. The data is being inserted and retrieved in the same order as how the `COMPOSITE PRIMARY KEY` is set up.

## 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.

For Query 2, the `itemInSession` column should be specified as a clustering column instead of being part of the composite key so that the results are sorted by that column.

## Presentation



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

Please make sure to add a Markdown cell as a header for each of the queries and include descriptions of the query that 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.

Please make sure to remove any in-line comments that were part of the project instructions so the notebook presents a professional look.

Kindly remove the TO-DO statements, such as the ones below:

TO-DO: `Query 1: Give me the artist, song title and song's length ...`

TO-DO: `Add in the SELECT statement to verify the data was entered`

 RESUBMIT

 DOWNLOAD PROJECT



## Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

[Watch Video](#) (3:01)

RETURN TO PATH

Rate this review

