Homework 1

MSAN 697, Diane Woodbridge

Due : Dec 4th (Midnight (PST))

Description

Given input file, “walmart_search_san_francisco.json” (Merchandise information containing “San Francisco”), complete the python scripts to answer Q1-Q8 (hw3_main.py, json_key_vaue.py, cassandra_function.py). *Do not add additional libraries or change the function definition.

Submit the .zip file- the name of your zipped file should be **MSAN697_HW3_LastName_Firstname.zip** on Canvas.

You can collaborate up to 3 people together (should be new collaborators).

Q1. (10 %)

- Complete “create_keyspace()” (cassandra_function.py) to create a keyspace with 'class': 'SimpleStrategy', 'replication_factor': '1'. (No need to define durable_writes).

Q2. (20 %)

- Complete “create_table()” function in “cassandra_function.py” to create a table with given field names and types.
- Create a table called “items” with “itemId” (INT), “name” (VARCHAR), “shortDescription” (TEXT), “customerRating” (DOUBLE), “numReviews” (INT) and “qty”(INT) where “itemId” is a primary key. Insert all the items from the json file and the last 2 digits of itemId as qty.

Q3. (10 %)

- Complete “select_data()” function in “cassandra_function.py” ” to find data from a table satisfying given constraints.

Q4. (15 %)

- Try to query 'MLB Women's San Francisco Giants Short Sleeve Top' on items table.
- Does select 'MLB Women's San Francisco Giants Short Sleeve Top' work in the items table table work without ALLOW FILTERING? Leave the answer (yes/no) in the code.

Q5. (10 %)

- Complete “create_materialized_view()” function in “cassandra_function.py” to create a materialized view with given field names from a base table, constraints, and primary keys.
- Create a materialized view called “materialized_items_view”, with all the columns from items, a constraint of “name IS NOT NULL” and name and itemId being primary keys.

Q6. (10%)

- Query 'MLB Women''s San Francisco Giants Short Sleeve Top' on materialized_items_view and print all the outputs.

Q7. (10%)

- Try to update data in the materialized view. (You can use update_data() or do it on the cassandra shell.)

```
cqlsh:ecommerce> UPDATE materialized_items_view SET numReviews = 10 WHERE name = 'NFL Men''s San Francisco 49Ers C Hyde 28 Player Tee' and itemId = 52507967;
```

- Can you update data in materialized views? Leave the answer (yes/no) in the code.

Q8. (15%)

- Try to update data in the base table. (You can use update_data() or do it on the cassandra shell.)

```
cqlsh:ecommerce> UPDATE items SET numReviews = 10 WHERE itemId = 52507967;
cqlsh:ecommerce> SELECT * FROM items WHERE itemId = 52507967;

 itemid   | customerrating | name                                                | numreviews | qty | shortdescription
----------+----------------+-----------------------------------------------------+------------+-----+------------------
 52507967 |           null | NFL Men's San Francisco 49Ers C Hyde 28 Player Tee |         10 |  67 |             NULL

(1 rows)
```

- When you update data in a base(original) table, is the content also updated in the corresponding materialized view?

```
cqlsh:ecommerce> SELECT * FROM materialized_items_view WHERE name = 'NFL Men''s San Francisco 49Ers C Hyde 28 Player Tee' ;
```

OUTPUT EXAMPLE:

```
ML-ITS-603436:Answer_Code dwoodbridge$ python hw3_main.py
Does select 'MLB Women''s San Francisco Giants Short Sleeve Top' work in the items table table work without ALLOW FILTERING? - Yes/No
Row(name=u"MLB Women's San Francisco Giants Short Sleeve Top", itemid=50732860, customerrating=None, numreviews=None, qty=60, shortdescription=u'NI
Can you update data in materialized views? - yes/no
When you update data in a base(original) table, is the content also updated in the corresponding materialized view? - yes/no
```