Lab7: Columnar Database

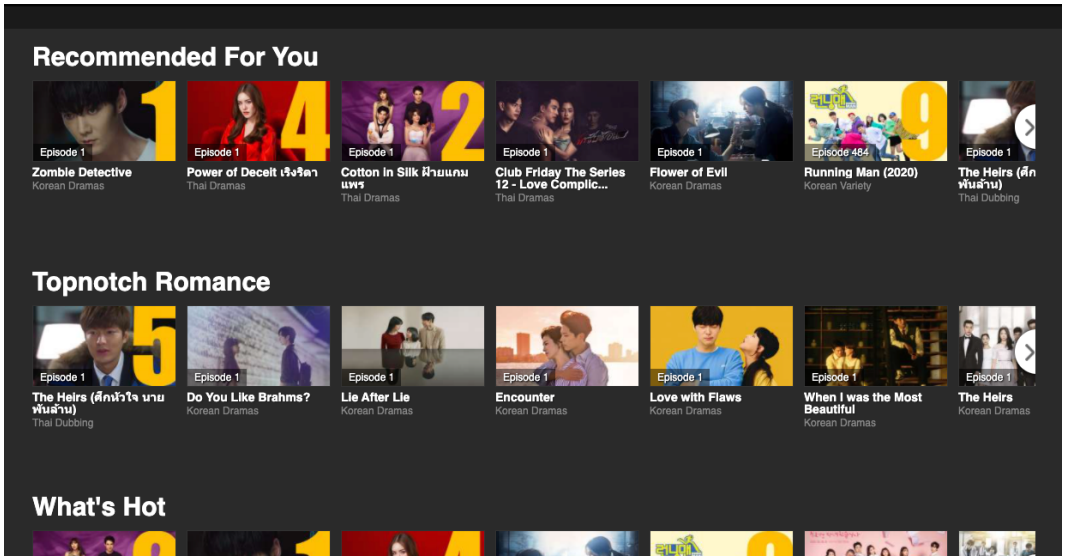
Objectives: Learn the a column-family model using Apache Cassandra about data model design, data manipulation, and Cassandra CQL collection to collect and retrieve data from columnar database.

Estimated Time : 3 hours

Lab Instruction

Case Study: *WeLoveSeries Application* is a streaming service that allows members to watch Asian TV-Series which come from many Asian countries such as Korea, Thai, China, Japan and India. This application should support high concurrency more than one million members per day around the world. Therefore, we would like to build a data model for improving the performance of this application using Apache Cassandra. At first, we have to convert data model from relational database to Apache Cassandar. Then, design data model depends on the query patterns that serves the below queries in a TV-Series and categories tables.

Q1) get a TV-Series by id.
Q2) get a category of TV-Series by id.
Q3) get all TV-Series under a particular category.
Q4) get the categories of a particular TV-series.



An Example Home Page of *WeLoveSeries* Application

Design a column-family model and write commands to collect and retrieve the data in Apache Cassandra for 14 questions using the given relational database schema.

Table1. Series Information

series_id	series_name	country	release_year
S001	Zombie Detective	Korea	2020
S002	Vagabond	Korea	2019
S003	Good Morning Call	Japan	2016
S004	Count Your Lucky Stars	China	2020

Table2. Category Information

cat_id	cat_name
C001	Mystery
C002	Comedy
C003	Crime
C004	Action
C005	Romance

Table3. Series_Category

series_id	cat_id
S001	C001
S001	C002
S002	C003
S002	C004
S003	C002
S004	C002
S004	C005

1. Create a keyspace in Apache Cassandra.

[Task 1.] Write command to create a keyspace named “lab7_keyspace”

2. Write commands to create column-family/table to serve above 4 queries.

[Task 2.] Write command to create table for Q1)
[Task 3.] Write command to create table for Q2)
[Task 4.] Write command to create table for Q3)

[Task 5.] Write command to create table for Q4)

3. Insert data to into a table for Q1-Q4, using command SELECT * FROM <table_name> to verirify the data and capture the results.

[Task 6.] Write insert command into table for Q1)
[Task 7.] Write insert command into table for Q2)
[Task 8.] Write insert command into table for Q3)
[Task 9.] Write insert command into table for Q4)

3. Query data in table Q1-Q4 and capture the results
[Task 10.] List all TV-Series data (in Q1)
[Task 11.] List a category name which has category id 'C002' (in Q2)
[Task 12.] List all TV-Series under category 'C002' (in Q3)
[Task 13.] Show the number of TV-Series of each category (in Q3)
[Task 14.] List all TV-Series which category CONTAINS 'ACTION' (using INDEX) (in Q4)

Lab Submission

Submission System: Google Classroom

Total TASKS: 14