JIGSAW ACADEMY
THE ONLINE SCHOOL OF ANALYTICS

Non-Graded Assignment
T5.2 Programming in Hive

Q1: Data sets BX-Books.tsv, BX-Book-Ratings.tsv and Bx-Users.tsv are present in location /home/data/datasets/programming_in_hive in Lab. All are tab-delimited files i.e. field separator is tab character. Using the given datasets, perform the queries mentioned below:

Format details of BX-Books.tsv:

Field Name	Sample Record
ISBN	195153448
Title	Classical Mythology
Author	Mark P. O. Morford
Year of Publication	2002
Publisher	Oxford University Press
Image URL-S	http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg
Image URL-M	http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg
Image URL-L	http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg

Format details of BX-Books-Ratings.tsv

Serial No	276725
ISBN	034545104X
Rating	0

Format details of BX-Users.tsv

Serial No	1
Address	nyc,newyork,usa
Area Code	0

Perform the following tasks using HIVE Scripts:

1. Create books_details table
2. Load data into this table from file “BX-Books.tsv”
3. Create books_ratings table

4. Load data into this table from file “BX-Books-Ratings.tsv”
5. Create user_details table
6. Load data into this table from file “BX-Userss.tsv”
7. Find out the number of books published each year. (Hint: Use BX-Boooks.tsv)
8. List down the year and number of books published in each year in the ascending order of the number of books
9. Write a query to join book_ratings and user_details tables on SerialNo field to display SerialNo and address of all users who have purchased books of rating 5.

Paste the HIVE Script for all the steps and the output for steps 7 to 9.

Q2:

Financial Analysis with Hive:

Leading financial service organizations are using big data analytics for both compliance and regulation, as well as building the business by getting advanced insights into customers, markets and operations.

We need to find out 52 week high and low price and the date on which it happened for each of the stock given in the data. You need to load this data into Hive table and then use SQL query for completing this assignment.

Find attached yahoo_stock_AA_32_mini.csv file in LMS for the assignment. The format of the data is:

exchange,stock_symbol,date,open,high,low,close,volume,adj close

Expected steps:

1. Create a table Stock_data in HIVE for storing Yahoo’s stock data
2. Load CSV file yahoo_stock_AA_32_mini.csv into Stock_data table
3. Write a HIVEQL to find 52 week high and low price and its respective date

Paste the code for each step and the output of the Query.