Non-Graded Assignment Solution

1. Create books_details table

create TABLE books_details (ISBN String, title String, author String, publication_year String, publisher String, imageURLS String,imageURLM String,imageURLL String) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;

2. Load data into this table from file "BX-Books.tsv"

LOAD DATA LOCAL INPATH '/home/data/datasets/programming_in_hive/BX-Books.tsv' into TABLE books_details;

3. Create books_ratings table

create TABLE books_ratings (serialNo String,ISBN String,rating Int) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;

4. Load data into this table from file "BX-Books-Ratings.tsv"

LOAD DATA LOCAL INPATH '/home/data/datasets/programming_in_hive/BX-Book-Ratings.tsv' into TABLE books_ratings;

5. Create user_details table

create TABLE user_details (serialNo String, address String, areacode String) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;

6. Load data into this table from file "BX-Userss.tsv"

LOAD DATA LOCAL INPATH '/home/data/datasets/programming_in_hive/BX-Users.tsv' into TABLE user_details;

7. Find out the number of books published each year. (Hint: Use BX-Boooks.tsv)

select publication_year, count(*) as count from books_details group by publication_year;

8. List down the year and number of books published in each year in the ascending order of the number of books

select publication_year, count(*) as count from books_details group by publication_year order by count;

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.
select b.serialNo, b.address, a.ISBN, a.rating from books_ratings a JOIN user_details b ON a.serialNo = b.serialNo where a.rating==5;

Q2:

1. Create a table Stock_data in HIVE for storing Yahoo's stock data
create table Stock_data (exchange String,stock_symbol String,stock_date String,stock_price_open double, stock_price_high double, stock_price_low double, stock_price_close double, stock_volume double, stock_price_adj_close double) row format delimited fields terminated by ",";

2. Load CSV file yahoo_stock_AA_32_mini.csv into Stock_data table
Load data local inpath '/home/<username>/yahoo_stock_AA_32_mini.csv ' into table Stock_data;

3. Write a HIVEQL to find 52 week high and low price and its respective date

a. Query to find Max stock price and date on which it happened select MAX(stock_price_close) from Stock_data group by stock_symbol; select stock_date from Stock_data where stock_price_close=39.12;

b. Query to find Min stock price and date on which it happened   select MIN(stock_price_close) from Stock_data group by stock_symbol; select stock_date from Stock_data where stock_price_close=30.69;