

PG-DBDA SEP 2022 BATCH KHARGHAR

MODULE: BIG DATA ANALYTICS

DATE : 10th June, 2023

MARKS : 40 MARKS

Please create a doc/txt/pdf file with 12 digits student id, which will contain the code along with the screenshots of the output or result. While taking the screenshot make sure that you are visible in all the images.

---

**Q1) Map Reduce Use NYSE DATA SET FOR THE BELOW QUESTION:**

**[ Marks: 15 ]**

**Q.) Find out the average closing price for each stock.**

Here, we have chosen the stock market dataset on which we have performed map-reduce operations. Following is the structure of the data. Kindly Find the solutions to the questions below.

**Data Structure**

- 1. Exchange Name**
- 2 Stock symbol**
- 3. Transaction date**
- 4. Opening price of the stock**
- 5. Intra day high price of the stock**
- 6. Intra day low price of the stock**
- 7. Closing price of the stock**
- 8. Total Volume of the stock on the particular day**
- 9. Adjustment Closing price of the stock**

**Field Separator – comma**

---

**Q2.) Hive:**

- **Use Airlines dataset for the Below Questions: [ Marks: 15 ]**

1. Which airports have the highest altitude ?
  2. How many routes are operated by active airlines from the United States ?
  3. Which airlines operate routes that have less than 3 stops number of stops top 10 alphabetically?
  4. How many airlines have a specific IATA code 'W9'?
  5. Find the airlines that operate routes with a specific equipment as 'AN4' and codeshare enabled.
- 

**Pyspark:**

- **Use Airlines dataset for the Below Questions: [ Marks: 10 ]**

1. What is the total revenue generated in each year?
  2. What is the average revenue per flight for each airline?
  3. What is the total number of booked seats for each quarter in a given year?
-