Hadoop: Spark Assignment - 3

Problem Statement:

You work as a Big Data Engineer at GrapeVine Pvt. Ltd. Your company is currently working as a Data Analytics consultant for a hedge fund. Due to the size of the available dataset, the company requires you to increase computational efficiency using Apache Spark. You have been assigned certain tasks for the fulfillment of this analysis through stock market backtesting.

Dataset:

https://drive.google.com/file/d/1B9ZVyuYTV8-EFx22eU0FUrq8pWNPZbQm/view?usp=sharing

Dataset Description:

The dataset used for this assignment is 'Reliance NSE Stock Market Data.' The relevant fields that will be put to use in further analysis are as follows:

time – The timestamp of the data record (separated by 5-minute intervals)
open – The opening price of the stock
high – The highest point of the stock in the last 5-minute interval
low – The lowest point of the stock in the last 5-minute interval
close – The price of the stock at the end of the 5-minute interval

The rest of the fields or columns can be ignored.

Tasks To Be Performed:

1. Find out the average 'close' price of Reliance throughout the duration of the dataset.
2. If a Reliance stock was bought at the beginning of the trading day, '2020-04-07' (YYYY-MM-DD), at the close price of the first 5-minute window, scan the dataset to find out the point to sell the stock to maximize profits. You are required to print the specific timestamp.
3. Find out the net profit or net loss to be accumulated if one stock of Reliance is bought at the opening of every 5-minute slot and sold at the lowest possible point in that 5-minute slot.

4. Find out the net profit or net loss to be accumulated if one stock of Reliance is bought at the opening of every 5-minute slot and sold at the highest possible point in that 5-minute slot.