Ramakrishna Mission Vivekananda Educational and Research Institute
PO Belur Math, Howrah, West Bengal 711 202
School of Mathematical Sciences
Department of Computer Science

MSc CS : Batch 2021-23, Semester II and Batch 2020-22 Semester IV, Final Exam
ETBDC
Champak Kumar Dutta

*S*tudent Name (in block letters): Date: 17 June 2022
*S*tudent Roll No: *M*ax Marks: 100
*S*ignature: Time: 4hrs

Answers must be properly justified to deserve full credits.

1. (25 points) Topic Apache Spark Streaming:

   Input Data: [”timestamp”: ”2018-09-25 10:05:00”, ”symbol”: ”MSFT”, ”priceData”: ”high”: ”114.6100”, ”close”: ”114.5400”, ”open”: ”114.5200”, ”low”: ”114.4900”, ”volume”: ”53615”, ”timestamp”: ”2018-09-25 10:05:00”, ”symbol”: ”ADBE”, ”priceData”: ”high”: ”265.9100”, ”close”: ”265.9100”, ”open”: ”265.5820”, ”low”: ”265.5820”, ”volume”: ”8283”, ”timestamp”: ”2018-09-25 10:07:00”, ”symbol”: ”GOOGL”, ”priceData”: ”high”: ”1183.1600”, ”close”: ”1183.1600”, ”open”: ”1183.1600”, ”low”: ”1183.1600”, ”volume”: ”2335”, ”timestamp”: ”2018-09-25 10:08:00”, ”symbol”: ”FB”, ”priceData”: ”high”: ”162.2700”, ”close”: ”162.0700”, ”open”: ”161.8300”, ”low”: ”161.8000”, ”volume”: ”239757”, ”timestamp”: ”2018-09-25 10:09:00”, ”symbol”: ”FB”, ”priceData”: ”high”: ”165.2700”, ”close”: ”165.0700”, ”open”: ”165.8300”, ”low”: ”165.8000”, ”volume”: ”239760”, ”timestamp”: ”2018-09-25 10:10:00”, ”symbol”: ”FB”, ”priceData”: ”high”: ”167.2700”, ”close”: ”167.0700”, ”open”: ”167.8300”, ”low”: ”167.8000”, ”volume”: ”239766”]

   (a) (20 points) Calculate the following indicator value of the stock in a 2-minute sliding window for the last 6 minutes.
   Indicator reference: C = The most recent closing price L3 = The lowest price traded of the 3 previous trading sessions H3 = The highest price traded during the same 3-day period

   Note: Data, format of the data and output are not critical for evaluation and you may use any data with the same attributes in any format. But unit test output (i.e. logs) are to be produced in exam hall for all entry points used and this output will be used for marking.

   (b) (5 points) Price is considered overbought when the indicator values is above 80 and oversold when below 20. Compare usability of Spark and Flink frameworks for solution to this problem.

   You may like to refer: https://timepasstechies.com/spark-textfilestream-to-find-relative-strength-index-or-rsi-of-stocks-with-sliding-window-and-reducebykeyandwindow-example/

2. (75 points) Please develop scripts/programs based on the following.

   Input Data: Jan_2020_ontime.csv (refer from classroom: Graph Algo -¿ GraphFrame)
   Input Description: Airport Arrival and Departure Details
   Environment: Spark GraphFrames and CypherQL
   Tools: No restriction

   (a) (5 points) Spark Graphframes: Find two airports having direct connection with distances greater than 500 units (miles)

   (b) (5 points) Spark Graphframes: Find all tuples of airports (a, b, c) following the conditions mentioned below:
   Airport a has a direct flight to airport b then airport b has a direct flight to airport c where airport a has no direct flight to airport c and airport c is not the same as airport a

(c) (25 points) Cypher QL: Find pair of airports having direct connection with distances greater than 500 units (e.g. miles)

(d) (40 points) Cypher QL: Find all tuples of airports (a, b, c) following the conditions mentioned below: Airport a has a direct flight to airport b, then airport b has a direct flight to airport c where airport a has no direct flight to airport c and airport c is not the same as airport a

Output is important for evaluation and should match for both the frameworks.