CPrE/SE 419: SOFTWARE TOOLS FOR LARGE-SCALE DATA ANALYSIS, SPRING 2022

LAB #8: STREAMBASE

Purpose

The purpose of this lab is for the students to get familiar with designing applications for domains in which the data arrives in a streaming manner. The additional practice-aspect consists of getting familiar with the whole paradigm of Visual Programming via TIBCO's Streaming BI tool.

Lab Guide:

Tool: TIBCO Streaming:

To download a 90 days free trial version:

https://www.tibco.com/resources/product-download/tibco-streaming-trial-windows

(or, for MAC or Linux: https://www.tibco.com/products/tibco-streaming)

Lab Units:

- There are plenty of resources – however, for the purpose of the learning objective of this lab, we will use the first five videos at: https://vimeopro.com/streambase/streambase-eventflow-tutorials
- The first two are quick "welcome" videos
- The "meat" of this exercise/lab is in videos #3 (longest of "Hello World", 9:29), #4 and #5.
  "Try-as-you-go" will be employed: at the end of the video (or, with stops along the way), try every example introduced.
- ASIDE:
  - A summary-document with the main steps is also available on Canvas
  - Everyone is more than welcome to explore the rest of the videos in the tutorial (e.g., #6 would tell you how to integrate reusable components for projects…).

  END_ASIDE.

Project Specification:

You are required you to develop a Streambase application for decision-making in a financial world, extending the settings from the lab. Following are the specifications:

1. The format of each data-item[1] in the stream will be:
   (stock_name, value, market_location, time)

[1] When testing, you should try to provide at least 4 different stocks (e.g., IBM, Intel, Google, Microsoft) and at least 4 different market-locations (e.g., New York, Tokyo, London, Beijing). If you would like to "play" with multiple input streams, then you can create one input stream, say, per continent and add cities/locations for that stream.

2. The first analytics-task is to separate a couple of categories:
    a. First separation is based on IBM (vs. non-IBM)
        i. for the IBM ones, you are interested in the average value over a sliding window of size 4 (per count/number of *values*, not time-units);
    b. For all the other (non-IBM stocks) you want to:
        i. separate the ones reported from London and the ones who have originated anywhere else but London;

Extra-credit:

3. You want to detect whether there exists a non-London/non-IBM stock which has a value that is within 10% of the value of an IBM stock, and you do not want the occurrence/detections of the that "other" stock (non-IBM) to be more than 60 seconds apart from the IBM stock (most recent one).
4. You want to use the stock detected in "3." above for your final output.

ASIDE: While the above steps illustrate what to do, it is in order to give a more "semantic-oriented" description of the object. Namely, the project that you will develop corresponds to the implementation of the request/policy specified as: *"Report the stocks which had higher value in the London exchange than the value in any non-London exchange and which matched the boundary with respect to IBM stock (cf. "3." above)– making sure that the two occurrences in the respective markets (London and non-London) were no longer than 60 seconds apart."* END_ASIDE.

Notes:

- For the extra-credit, you will need to use *aggregations* and *joins* (or, if you want to "pre-trim" any data (which you can post-trim after the join too), "f(x)" – the *map operator* too);
- Should you need additional info about joins, you can consult:

  https://docs.streambase.com/sb15/index.jsp?topic=%2Fcom.streambase.sb.ide.help%2Fdata%2Fhtml%2Fauthoring%2Fjoinoperator.html

  (at the bottom of the page, there is a link for "Join Operator Sample")

What to turn in: The best way is if you simply export your project (after you have saved it).

Then, put that together with a pdf file of a screenshot of your implementation from the streambase canvas in a folder called Lab-8; zip the folder and upload it at the Canvas.

You are welcome to work in teams of two students for this project.