Uni Freiburg, Web Science Group
Prof. Peter Fischer
Systems Infrastructure for Data Science - Winter 2012/13

Exercise Sheet #13: Data Stream Processing

February 8, 2013

1   Data Stream Processing

Consider the following two data streams from a financial feed provider that reports on the trading activities of stock holders in the New York Stock Exchange (NYSE) and the Swiss Exchange (SWX) electronic trading markets:

```
MarketData_NYSE(time, tid, type, symbol, price, volume)
```

```
MarketData_SWX(time, tid, type, symbol, price, volume)
```

Traders have unique identifiers (`tid`), and they can be involved in three different types of activities in the market (`type`):

  * "bidding", to buy a certain volume of a certain symbol offered below a certain maximum bid price;

  * "asking", to sell a certain volume of a certain symbol at a certain ask price; and

  * the actual "trading" of a certain symbol at a certain price and volume.

A. State the following continuous query in CQL: Monitor the total volume of IBM stocks "traded" on the Swiss Exchange in the last 10 minutes. Which Aurora operators would you use to model this query, in which order, and with which key parameters? (40 points)

B. State the following continuous query in CQL: Monitor the stocks whose "ask" price has been lower at SWX than its "ask" price at NYSE for the last 10 minutes. Which Aurora operators would you use to model this query, in which order, and with which key parameters? (40 points)

C. Are the following two CQL queries equivalent? Please explain briefly. (20 points)

```
Q1: SELECT COUNT(*)
    FROM MarketData_SWX [PARTITION BY symbol ROWS 10]

Q2: SELECT COUNT(*)
    FROM MarketData_SWX [ROWS 10]
    GROUP BY symbol
```