Challenge 1 – Event Streaming

In the first Challenge of the SAP Coding Masters 2013 you will have to deal with a stream of events containing messages from *stocktwits.com* as well as *twitter.com*. The target of this challenge is to fetch all data from *stocktwits.com* that talk about a set of stock symbols and load them into your group's SAP HANA database.

Basics

There is a server which will provide your team with a mixture of *stocktwits.com* and *twitter.com* tweets. All events are JSON-formatted. The server implements the Server-Sent-Events-Pattern. However, there will be no outlying data-tag. In order to register a client at the server, you can use the following url:

http://10.20.41.175:5500/CMStreaming/stream

with POST-encoded parameters:

{"group": <YOUR GROUP NUMBER>,

"password": <YOUR PASSWORD> }

For all tasks you will be scored by the f-measure. For ten percent score you will get 5 points – resulting in 50 points for 100% correctness (f-measure = 1.0).

F-measure = 2 x ((precision x recall) / (precision + recall))

- **precision** is the number of correct results divided by the number of all returned results
- **recall** is the number of correct results divided by the number of results that should have been returned

Task 1 – 50 points

Explore the schema used by *stocktwits.com* and *twitter.com* and insert the solution into your group's HANA Database table "CODING_MASTERS"."STOCKTWITS_SCHEMA"/"TWITTER_SCHEMA". The root of all messages should be "message". Both tables should be filled in the following manner:

PARENT	CHILD
message	user
user	username
user	bio

Task 2 – 100 points

Find all tweets from stocktwits.com that talk about a symbol from a list that is persisted in your group's HANA Database. The table containing the target symbols is "CODING_MASTERS"."SYMBOLS". *Stocktwits.com* uses a "$"-sign to escape their symbols, which you will need to change to a "#" sign. Also you will need to parse the symbols into upper-case.

Example: Symbols = "XYZ"

Source	Target
Test Text text $XYZ!!	Test Text text #XYZ!!
Another test text $xYz	Another test text #XYZ
$Xyz increases for $22!	#XYZ increases for $22!

Insert the found messages into the table "CODING_MASTERS"."STREAMING" in your group's HANA Database to solve this task.

Task 2b

Fill the fields for PARENT_USER_ID and REPLY_TO_USER_ID in the table "STREAMING" for each message. Try to find out which user a message is replying to, and fill this information into the corresponding fields.