3/13/2016

MTH 9898 – BIG DATA IN FINANCE
ASSIGNMENT B – "IN-MEMORY TIME-SERIES DATA ANALYSIS WITH KDB AND Q"
ELDAR URMANOV

INTRODUCTION

In this assignment we are using kdb+/q in-memory database to generate synthetic financial time-series data and to compute the daily volume-weighted average prices for several tradable equities.

DECRIPTION OF METHODOLOGY

In order to generate the synthetic time series data, we set the following parameters:

- Number of trading days (30). Our time horizon for VWAP computation is 1 month which seem reasonable for testing purposes. Parameter can be changed at the function call in the code.
- Number of stock symbols (10). We decided to use 1-letter real financial tickers for extra simplicity. Number of tickers chosen appears to be reasonable as we would like to show how VWAP can be computed for several assets at once (not too many, not too few).

Below is the list of tickers used:

C: Citigroup, F: Ford, K: Kellogg, L: Loews, M: Macy's, P: Pandora, S: Sprint, T: AT&T, V: Visa, Z: Zillow

- Number of trades per day (1000). This is a reasonable assumption given how real-life low-frequency trades occur.

The synthetic data values for price and volume were created using built-in Q randomizer (?) so there is not assumption of normality. Values were generated uniformly: for price in the range of 100 (very reasonable as most stock prices are within this range), for volume in the range of 100,000 (also reasonable, judging from real-life trading activity).

For VWAP function definition, we used simple formula: (sum of price*volume)/(sum of volume).

RESULT

The following two files are produced as an output: trades.csv (which contains generated synthetic data) and vwap.csv (which contains daily VWAP for each of the 10 stocks for 30 days).