Real-Time Analysis for High-Frequency Trading

Your Data Needs to Be Faster, Not Just Bigger

© 2015 Lee A Cole
WHERE BIG DATA INVESTMENT GOES
Real-time data takes a backseat to historical data.

VELOCITY OF DATA
47%

PROCESS IMPROVEMENT

VOLUME OF DATA
16

VARIETY OF DATA
26

COST SAVINGS AND EFFICIENCY
8

SOURCE NEWVANTAGE PARTNERS 2014 BIG DATA EXECUTIVE SURVEY

Harvard Business Review  Feb 2015
Real-Time Analysis

Overview
Specific Example being securities trading

Three Points
• You need to be faster
• Faster is probably faster than you think
• Have a(n automated) strategy for missing or bad data.
Real-Time Analysis

What is High-Frequency Trading?

• Trading, not investing

• Start holding no positions
  – Finish holding no positions
  – Buy/Sell millions of shares during the day

• Move into and out of a position in minutes, seconds, even milliseconds
  – Microseconds soon
  – NASDAQ will respond in ~100 nanoseconds after an order hits its machines
Real-Time Analysis

Data Volumes

Market Data Volumes & Rates

• US Listed Equities ~5 Billion messages/day
• US Options 10-100 changes per equity change
• FOREX higher still
• Futures, commodities, OTC, pink sheet, overseas…
• Other data sources:
  – SEC filings
  – News
  – Social media

© 2015 Lee A Cole
Real-Time Analysis

Data Volumes

Trading Message Volumes

• ~20 orders placed per order executed
  – 95% cancel rates

• Have seen one trader place 20+ million orders / day

• Each order may have 5 or more messages involved
  – Trade messages are far more variable (format) than market data
Real-Time Analysis

Implications

• Short decision timeframes
  – milliseconds at most, microseconds common

• Mistakes can be expensive

• OTS tools generally too slow
  – Custom code, hand tuned
  – FPGAs (trading, not analysis)
  – Custom ASICs (market data servers)
Real-Time Analysis

Implications (cont.)

• Competitive pressures
• Preplanned, automated strategies
  – missing, noisy, or misleading data
    • missing data, if resent, will be stale and of little use
• Network noise, garbage collection times, page swapping delays can be killers
• Databases are slow
Real-Time Analysis

You must be faster than you think

• Queues
  – Arrival times vary
  – And are outside of your control

• Utilization
  \[ p = \frac{\text{average time to process a message}}{\text{average time between messages}} \]

• Rule of thumb:
  – Average queue depth = \( p/(1-p) \)

© 2015 Lee A Cole
Real-Time Analysis

Average Queue Depth vs Utilization

© 2015 Lee A Cole
Real-Time Analysis

Average Queue Depth vs Utilization
(zoomed)

Utilization

Average Queue Depth

© 2015 Lee A Cole
Real-Time Analysis

Keep utilization low (<20-25%)
• A bored cpu can respond quickly
• Applies to network and disk i/o as well

How?
• Algorithms
• Simple design (simple usually runs faster)
• Tight coding
  – Buffers not objects
• Measure/monitor constantly
  – Things always change. Know how, and how it impacts the analysis

© 2015 Lee A Cole
Real-Time Analysis

Wrap Up

• Your data should be faster, not just bigger
  – Competitive pressures will force it
  – IOT => more real-time data

• Utilization rates should be ~20-25% or less

• Automated strategies for data cleaning
  – As part of your normal processing time
  – You won’t have time otherwise
Real-Time Analysis

Real-Time Analysis for

High-Frequency Trading

Your Data Should Be Faster, Not Just Bigger

Lee A Cole, PhD
leecole@gmail.com
+1 720 985-4716

© 2015 Lee A Cole