Overview

The primary objective of 3Forge AMI is to deliver maximum information to humans as fast as possible, allowing them to immediately sort, search and drill into data. Our latencies and related metrics are designed to match that of human user needs.

Requirements For User Experience

- Latency Objective: transmit data to the user at least as fast as the "the human visual system can process images". This is 12 FPS, or 80 milliseconds. ¹
- Query Response Objective: At least as fast as the "human cycle time". This is about 100 milliseconds. 2

Objective For Data

- · Real time tables Objective: 1 million records with full sorting, searching
- Total data Objective : 100 million+ records
- Burst Throughput Objective: 100,000 records per second

Results Overview (details on following pages)

Build Version: 2292

Test 1 – Sustained heavy load of new objects at 25,000/second. Single simulator creating 25,000 objects per second for 30 minutes sustained. The total is 45,000,000 objects being monitored.

Test Results:

- · Avg. Latency from object creation to display: 19.49 milliseconds
- Total objects in display: 45,000,000
- Total memory used in Central Server: 7.2gb

Test 2 – Massive burst of new objects at 500,000/second. Five simulators each bursting 100,000 objects per second for 20 seconds. The total is 10,000,000 objects in just 20 seconds.

Test Results:

- Avg. Latency from creation to display: 84.17 milliseconds
- Total objects in display: 10,000,000

Test 3 – Massive real-time aggregation. This includes Real-time sorting on aggregate column with derived calculations .

Test Parameters:

- Underlying rows: 10,000,000 executions
- Grouping categories: 40,000 symbols
- Max underlying data in one group: 15,071 Executions for top Symbol

Test Results: No visible lag

Test 4 – Real-time search over massive data set. Data is auto-sorted by price. Query includes 4 symbols across 1 billion entries.

Test Results

- Result set: 260,250 market data events
- Time to execute query: 698 milliseconds

[1] Read, Paul; Meyer, Mark-Paul; Gamma Group (2000). Restoration of Motion Picture Film

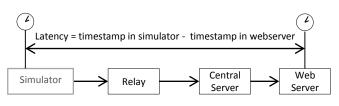
[2] Card, Stuart; Moran, Thomas; Newell, Allen; Lawrence Erlbaum Associates, Inc (1983). The Psychology of Human-Computer Interaction



Performance Statistics AMI build 2292.

Measurement Approach

All tests are end to end, which include the creation, transmission, storage and format of an object at the front end. First, the simulator creates a new "MarketTrade" object which contains a timestamp reflecting current systime (see "exectime" in sample message). This object is passed to the AMI Relay. After the object passes through the Central Server and reaches it's final destination in the Web Server, the system time is compared against the object's timestamp to determine a latency. This , along with the



number of objects, is aggregated and logged 1/sec. This latency will include (a) creation of the message in the simulator (b) transmission from simulator to the Relay (c) Relay processing (d) transmission from Relay to Central Server (e) storage of data in the Central Server (f) transmission from Central Server to Web Server (g) formatting in Web Server.

Message Format:

Each object contains (a) venue [enum], (b) symbol [enum], (c) price [float], (d) time [long], & (e) quantity [int].

Sample: O|T="MarketTrade"|Venue='BATS'|Symbol='MSFT.O'|Price=27.32|exectime=1380479599421L|qty=20200

System Overview (test 1,2,3):

All Tests performed using openjdk 1.6 on a commodity server running centos with 48g & 12 x 3.06GHZ Intel Xeon processors. OS and java are default installations. Simulators, Central AMI server, Relay and Web Server are all running on this single machine, communicating via loopback.

JVM: java version 1.6.0_24 - OpenJDK 64-Bit Server VM (build 20.0-b12, mixed mode) JVM MEM: 20GB for Central Server, 10GBfor Web Server. 1GB for Relay. 1gb for Simulator

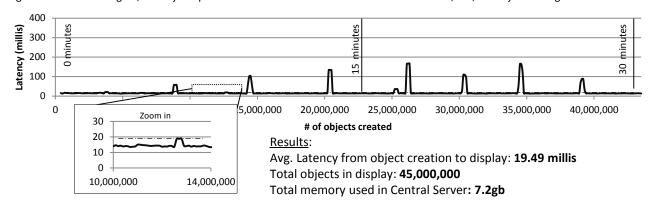
OS: 2.6.32-358.6.1.el6.centos.plus.x86_64

CPU: 2x Intel Xeon X5675 Westmere-EP 3.06GHz 256KB L2CH 12MB L3CH LGA 1366 95W 6-Core

RAM: 6 x Kingston 8GB 240-Pin DDR3 SDRAM ECC Registered DDR3 1333 Server Memory **MB**: ASUS Z8NA-D6C Dual LGA 1366 Intel 5500 ATX Dual Intel Xeon 5500 and 5600 Series

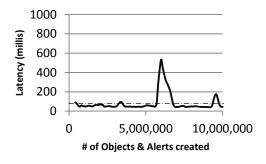
Test 1 – Sustained heavy load of new objects at 25,000/second

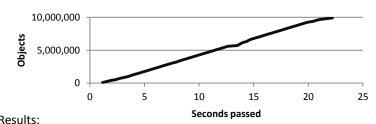
Single simulator creating 25,000 objects per second for 30 minutes sustained. This totals 45,000,000 objects being monitored.



Test 2 – Massive burst of new objects at 500,000/second

Five simulators each bursting 100,000 objects per second for 20 seconds. This totals 10,000,000 objects in just 20 seconds





Avg. Latency from creation to display: **84.17 millis** Total objects in display: **10,000,000**

Test 3 – Massive real-time aggregation

- · Real-time sorting on aggregate column
- Derived calculation for volume weighted Avg Px
- · No visible lag when searching, sorting, filtering

Results

Underlying rows:

10,000,000 executions

Grouping categories:

40,000 symbols

Max underlying data in one group:

15,071 Executions for top Symbol

40,000 Symbols with Executions				
Symbol	→ Count	Avg Price	Total Quantity	Ŀ
UREL	15,071	\$30.33	2,267,250	E
FLRP	13,239	\$30.65	1,984,700	
UTJY	8,955	\$30.43	1,343,150	
SALH	8,362	\$29.99	1,254,950	
XWZL	8,261	\$30.23	1,241,150	
BPJL	5,534	\$30.37	828,200	
TXCR	5,450	\$29.68	817,650	
OORV	5,423	\$30.47	813,100	
AIZDII	E 400	#20 24	040.000	E

Test 4 - Real-time search over massive data set

- Results auto-sorted by price
- Query time includes request, index lookup, data retrieval, sorting and display to front-end
- · No visible lag when searching, sorting, filtering



Results

Size of Dataset at time of query:

1,000,000,000 market data events

Query includes:

4 symbols at once (selected in left panel)

Result set:

260,250 orders

~50,000 per second.

Simulated Market data events inserted while query was running:

Central Server Memory usage:

~200 GB used / 215.0 GB available

Server Location

Amazon Cloud (Virginia)

Time to execute query:

0.698 seconds

