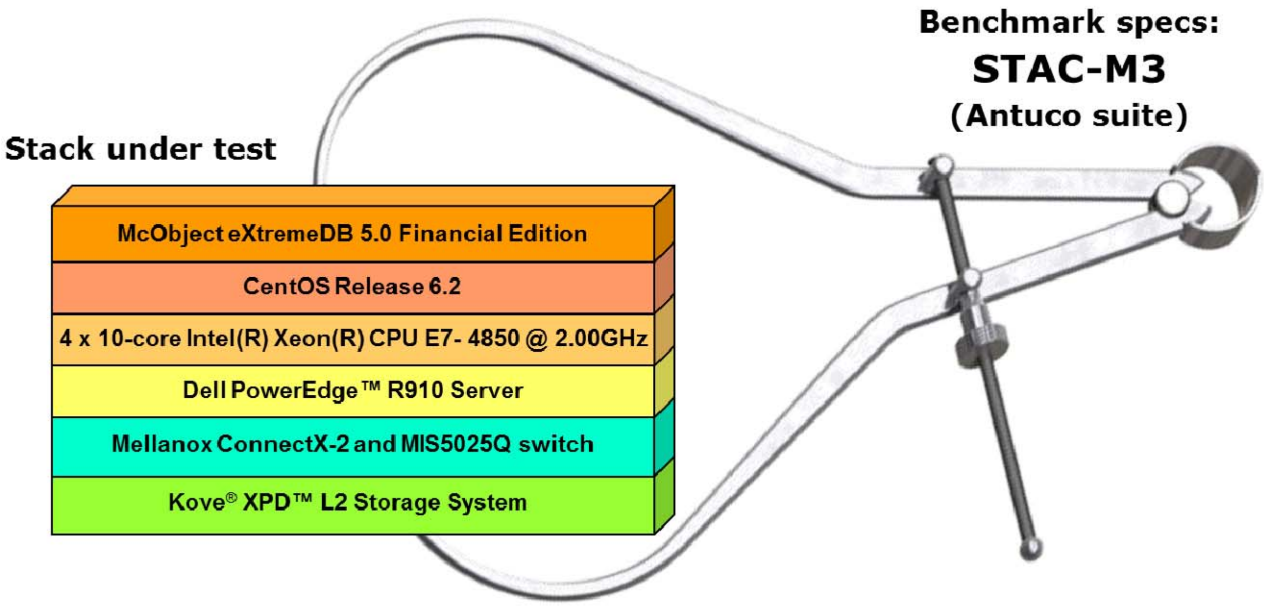
McObject eXtremeDB 5.0 Financial Edition with Kove® XPD™ L2 Storage System, Dell PowerEdge™ R910 Server and Mellanox ConnectX-2 and MIS5025Q QDR InfiniBand Switch

SUT ID XTR121105

STAC-M3™ BENCHMARKS (Antuco Suite)

Test date: 5 November 2012

Draft 1.01, 7 November 2012



STAC BENCHMARK COUNCIL™

THESE TESTS FOLLOWED STAC BENCHMARK SPECIFICATIONS PROPOSED OR APPROVED BY THE STAC BENCHMARK COUNCIL (SEE WWW.STACRESEARCH.COM). BE SURE TO CHECK THE VERSION OF ANY SPECIFICATION USED IN A REPORT. DIFFERENT VERSIONS MAY NOT YIELD RESULTS THAT CAN BE COMPARED TO ONE ANOTHER.

STAC SECURITIES TECHNOLOGY ANALYSIS CENTER

Contents

References

[1] Specifications used for this benchmark: STAC-M3 Benchmark Specifications, Antuco Suite, Rev M – http://www.STACresearch.com/node/8777. Accessible by qualified members of the STAC Benchmark Council.

Summary

STAC recently performed STAC-M3™ Benchmarks on a stack consisting of the McObject eXtremeDB 5.0 Financial Edition and a Kove® XPD™ L2 storage system using purpose-built Dell server technology, mesh-connected via Mellanox ConnectX®-2 InfiniBand adapters through a Mellanox MIS5025Q Switch to a Dell PowerEdge™ R910 Server that used Intel E7-4850 processors. This report documents the benchmark results.

In all, the STAC-M3™ specifications deliver dozens of test results, which are presented through a variety of tables and visualizations in this report. Of these, McObject chose to highlight a few, as follows:

- *Lower mean latencies (response times) than the previously published best results for 15 of the 17 operations*
- *Lower standard deviations of latency than the previously published best results for 13 of the 17 operations*
- *Over 9x the performance of the previously published best result for 10T.THEOPL (SUT ID: KDB110927)*
- *Over 7x the performance of the previously published best result for 1T.YRHIBID-2 (SUT ID: KDB110927)*
- *Over 5x the performance of the previously published best result for 1T.YRHIBID (SUT ID: KDB110927)*
- *Approximately 2x the performance of the previously published best result for 1T.NBBO and 10T.MKTSNAP (SUT ID: KDB110927).*

A STAC-M3 Report Card appears at the top of the report as a convenience for readers who want to get straight to the results. However, we recommend that readers who are not part of the STAC-M3 Working Group first read Section 1 (Overview) to get a feel for the test cases and metrics.

Report Card

STAC-M3™ Benchmarks for SUT ID XTR121105: McObject eXtremeDB 5.0 Financial Edition / Kove® XPD™ L2 / Dell PowerEdge™ R910 / Mellanox ConnectX®-2 and MIS5025Q QDR InfiniBand

Storage efficiency

The reference size of the dataset divided by the size of the dataset as stored by the SUT. The less storage space required, the higher the percentage.	
STAC-M3.v1.1.STORAGE.EFF	**169%**

Light-Compute Benchmarks

High Bid (1 Client Thread Requesting)

Return the high bid for a certain 1% of symbols over varying timeframes. Run the year-high bid a second time (YRHIBID-2) without clearing the cache.

		Latency (milliseconds)	
Spec ID		MEAN	MAX
STAC-M3.ß1.1T.YRHIBID.LAT2	Last-result latency	1,278	1,299
STAC-M3.ß1.1T.YRHIBID-2.LAT2	Last-result latency	663	686
STAC-M3.ß1.1T.QTRHIBID.LAT2	Last-result latency	410	420
STAC-M3.ß1.1T.MOHIBID.LAT2	Last-result latency	336	339
STAC-M3.ß1.1T.WKHIBID.LAT2	Last-result latency	124	129

		MB/second	
Spec ID		MEAN	MAX
STAC-M3.ß1.1T.YRHIBID.BPS	Bytes-read per second*	4,503	4,598
STAC-M3.ß1.1T.YRHIBID-2.BPS	Bytes-read per second*	0	0
STAC-M3.ß1.1T.QTRHIBID.BPS	Bytes-read per second*	3,705	3,799
STAC-M3.ß1.1T.MOHIBID.BPS	Bytes-read per second*	1,522	1,543
STAC-M3.ß1.1T.WKHIBID.BPS	Bytes-read per second*	1,131	1,192

* Bytes read per second from persistent media, according to *iostat*. That is, cache hits do not count as bytes read

Write Test

Perform the Basic Data Generation Algorithm for 1 day's data.

		Latency (milliseconds)	
Spec ID		MEAN	MAX
STAC-M3.v1.1T.WRITE.LAT2	Write-completion latency	26,301	26,788

STAC Report Card (cont'd)

STAC-M3™ Benchmarks for SUT ID XTR121105

Post-Trade Analytics Benchmarks

VWAB on 1 Day's Data
(1 Client Thread Requesting)

Return ~4-hour volume-weighted bid over a single day for certain 1% of symbols

		Latency (milliseconds)	
Spec ID		MEAN	MAX
STAC-M3.v1.1T.VWAB-D.LAT1	First-result latency	198	217
STAC-M3.v1.1T.VWAB-D.LAT2	Last-result latency	198	217

Theoretical P&L
(10 Client Threads Requesting)

For each of 10 Client Threads querying a unique set of 100 trades, find the amount of time until 2x, 4x, and 20x the size of each trade was traded in the market, and return the VWAP and total volume over those times intervals.

		Latency (milliseconds)				
Spec ID		MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.10T.THEOPL.LAT1	First-result latency	151	154	129	167	9
STAC-M3.ß1.10T.THEOPL.LAT2	Last-result latency	151	154	129	167	9

Market Snapshot
(10 Client Threads Requesting)

To each of 10 Client Threads querying a unique date, time, and set of symbols (1% of the total symbols), return the price and size information for the latest quote and trade for each symbol.

		Latency (milliseconds)				
Spec ID		MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.10T.MKTSNAP.LAT1	First-result latency	184	189	159	209	13
STAC-M3.ß1.10T.MKTSNAP.LAT2	Last-result latency	184	189	159	209	13

STAC Report Card (cont'd)

STAC-M3™ Benchmarks for SUT ID XTR121105

Research Analytics Benchmarks

Volume Curves (10 Client Threads Requesting)

To each of 10 Client Threads querying a unique set of 20 dates and set of symbols (10% of the total symbols), return the average proportion of volume traded in each minute interval for each symbol across the date set.

		Latency (milliseconds)				
Spec ID		MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.10T.VOLCURV.LAT1	First-result latency	8,189	8,341	6,582	9,355	623
STAC-M3.ß1.10T.VOLCURV.LAT2	Last-result latency	8,189	8,341	6,582	9,355	623

Aggregated Stats (10 Client Threads Requesting)

For each of 10 Client Threads querying a unique exchange, date, and start time, return basic statistics calculated for the entirety of the 100-minute time range following the start time. Time ranges always cross a date boundary.

		Latency (milliseconds)				
Spec ID		MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.10T.STATS-AGG.LAT1	First-result latency	21,072	21,436	19,221	24,114	1,190
STAC-M3.ß1.10T.STATS-AGG.LAT2	Last-result latency	21,072	21,436	19,221	24,114	1,190

Stats Over Unpredictable Intervals (Variable Client Threads Requesting)

To each of some number of Client Threads querying a unique exchange, date, and start time, return basic statistics calculated for each minute interval in a 100-minute time range following the start time. Start times are offset from minute boundaries by a random amount. Time ranges always cross a date boundary. Tests must be run with 1, 10, 50, and 100 Client Threads. Tests with other numbers of Client Threads are optional.

		Latency (milliseconds)				
Spec ID		MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.1T.STATS-UI.LAT1	First-result latency	6,264	6,422	5,406	6,686	523
STAC-M3.ß1.10T.STATS-UI.LAT1	First-result latency	17,502	17,817	15,112	20,343	1,595
STAC-M3.ß1.50T.STATS-UI.LAT1	First-result latency	23,475	23,275	19,459	29,936	1,960
STAC-M3.ß1.100T.STATS-UI.LAT1	First-result latency	60,181	60,705	51,945	65,268	3,079
STAC-M3.ß1.1T.STATS-UI.LAT2	Last-result latency	6,264	6,422	5,406	6,686	523
STAC-M3.ß1.10T.STATS-UI.LAT2	Last-result latency	17,502	17,817	15,112	20,343	1,595
STAC-M3.ß1.50T.STATS-UI.LAT2	Last-result latency	23,475	23,275	19,459	29,936	1,960
STAC-M3.ß1.100T.STATS-UI.LAT2	Last-result latency	60,181	60,705	51,945	65,268	3,079

STAC Report Card (cont'd)

STAC-M3™ Benchmarks for SUT ID XTR121105

NBBO Benchmark

NBBO

Calculate NBBO across all exchanges for all symbols on one day.

		Latency (milliseconds)	
Spec ID		MEAN	MAX
STAC-M3.ß1.1T.NBBO.LAT2	Write-completion latency	141,288	141,792

Multi-day/Multi-User VWAB Benchmark

VWAB for 12 Days with No Overlap in Interest (100 Client Threads Requesting)

To each of 100 Client Threads querying unique symbol sets, return 4-hour volume-weighted bid for 12 random days per thread for 1% of symbols per thread

		Latency (milliseconds)				
Spec ID		MEAN	MED	MIN	MAX	STDV
STAC-M3.v1.100T.VWAB-12D-NO.LAT1	First-result latency	18,143	18,169	16,922	19,006	374
STAC-M3.v1.100T.VWAB-12D-NO.LAT2	Last-result latency	18,143	18,169	16,922	19,006	374

Chart view

The charts that follow illustrate or elaborate on the results above:

- Figures 1 through 4 plot the mean last-result latency (LAT2) benchmarks for all of the operations.
- Figures 5 and 6 analyze the individual latency observations for the multi-user/multi-day VWAB benchmark (STAC-M3.v1.100T.VWAB-12D-NO.LAT2), first by sorting the results by latency, then by plotting them in a histogram.
- Figure 7 provides a more explicit look at multi-user scaling by plotting the latency for the intervalized statistics benchmark (STAC-M3.ß1.[n]T.STATS-UI.LAT2) against the number of simultaneously requesting client threads (n).
- Figures 8 and 9 take the 100-client-thread case of Figure 7 and analyze the individual latency observations, first by sorting the results by latency, then by plotting them in a histogram.

Refer to Section 1 (Overview) and the tables above for explanations of the benchmark IDs used in the charts.

The axes in the bar charts are fixed, so that results from this SUT may be visually compared to those of other SUTs. Because the results of future SUTs are unpredictable, the axes use a log scale.

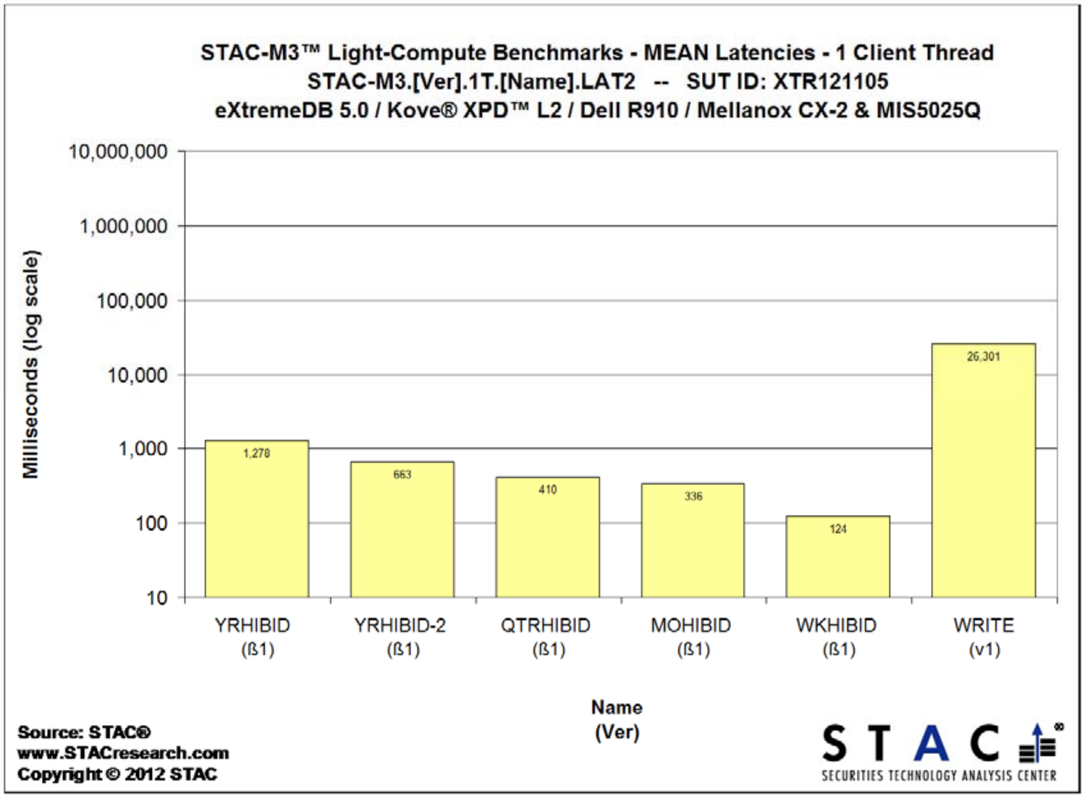


Figure 1

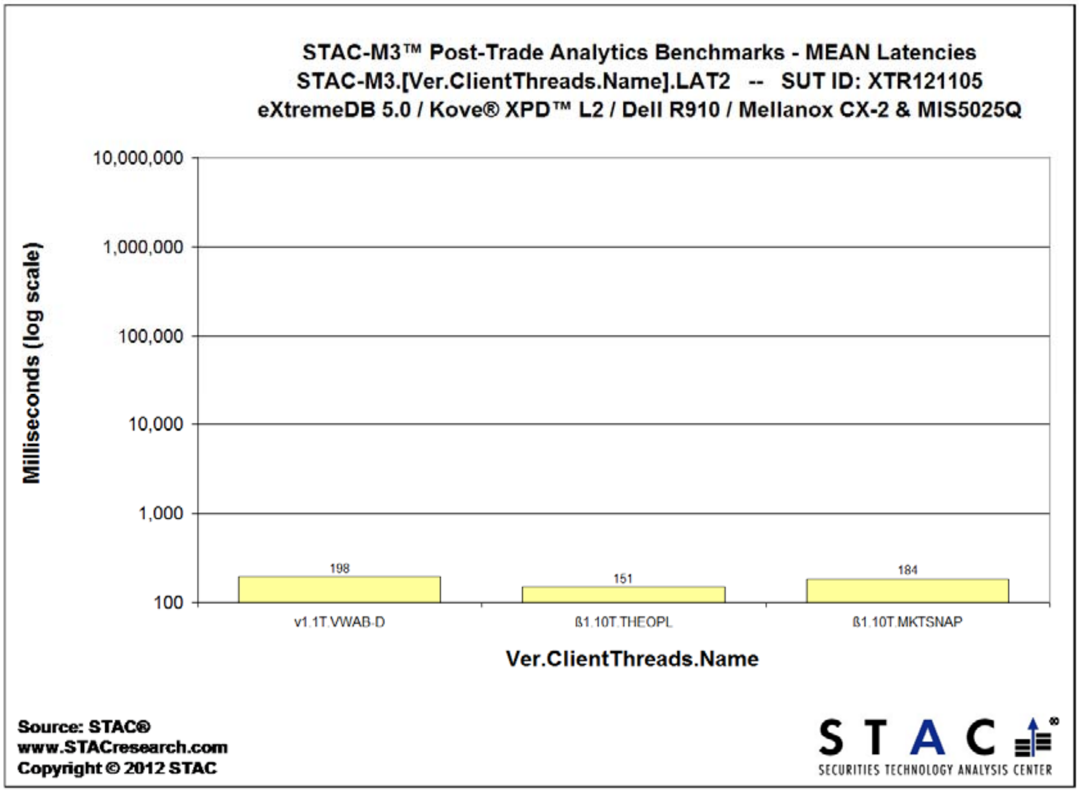


Figure 2

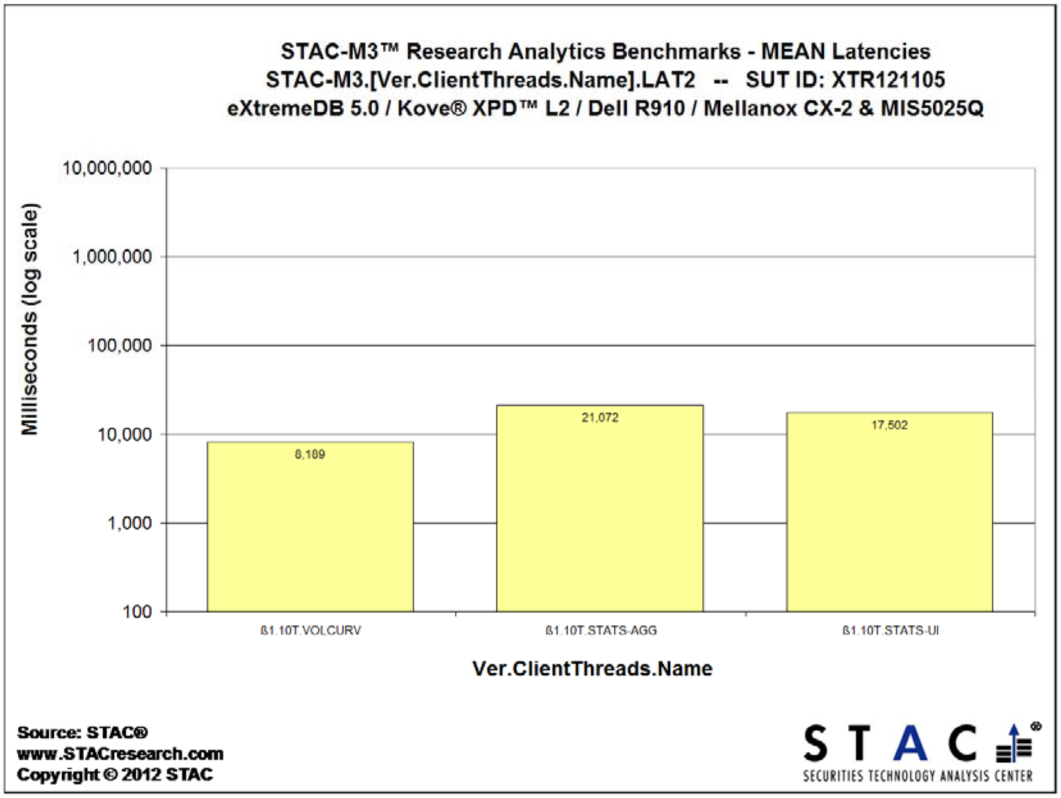


Figure 3

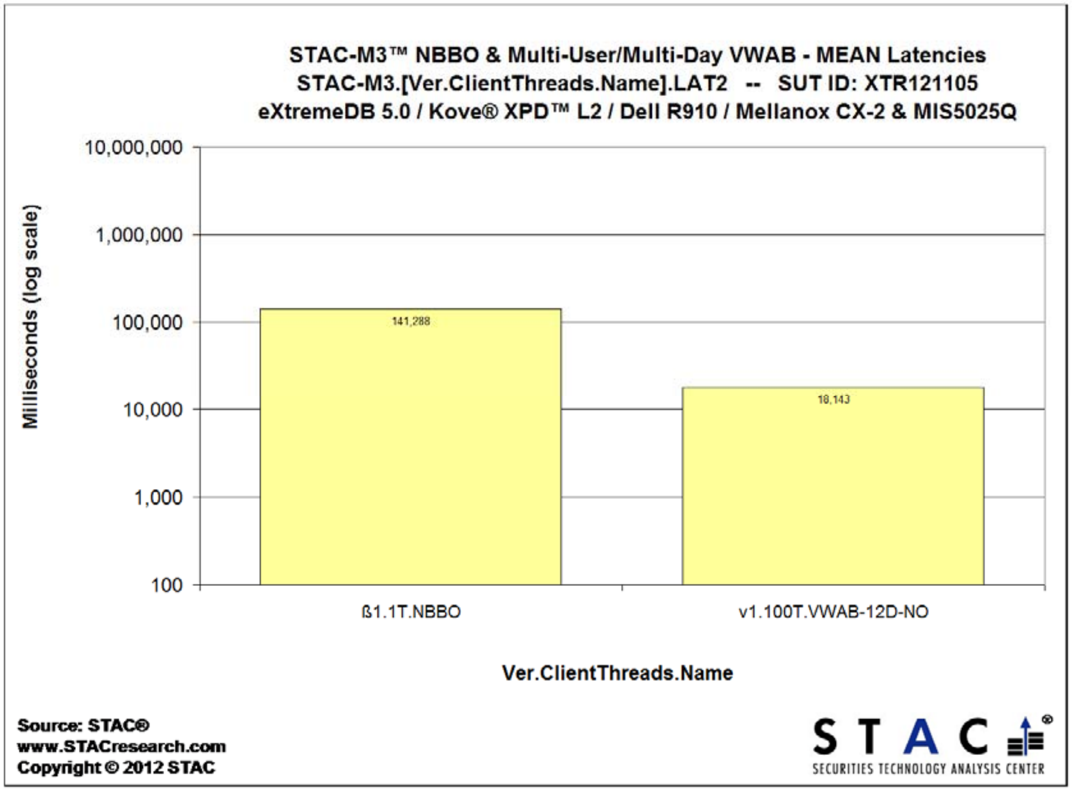


Figure 4

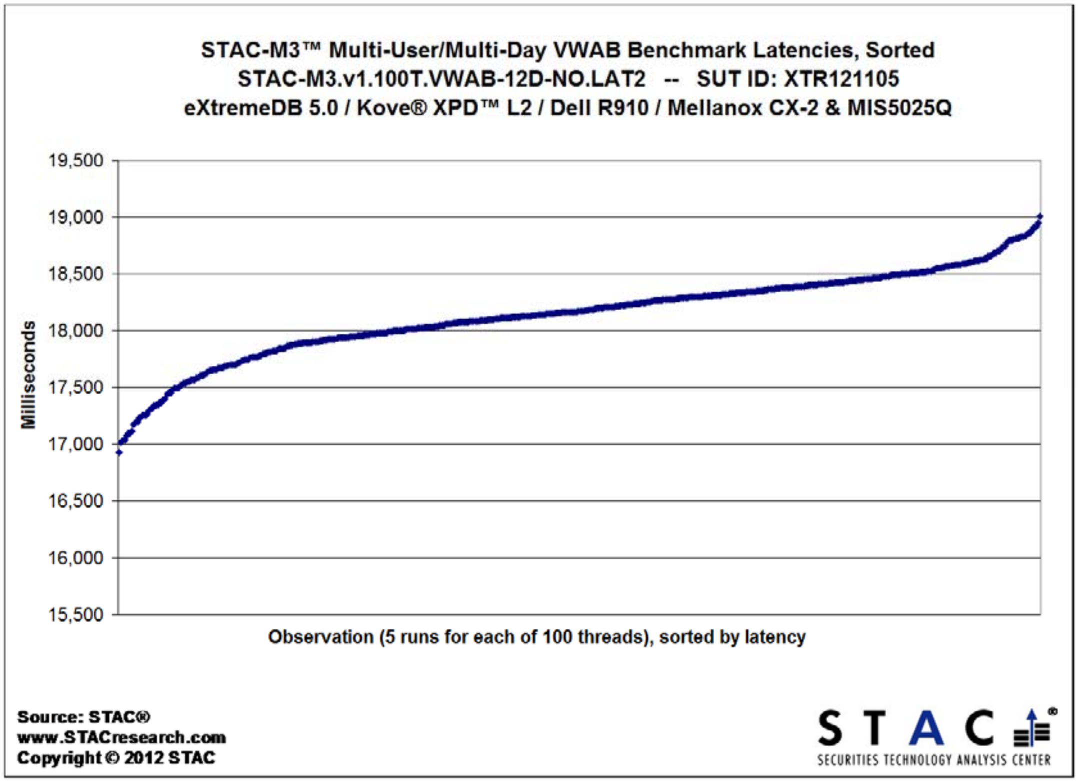


Figure 5

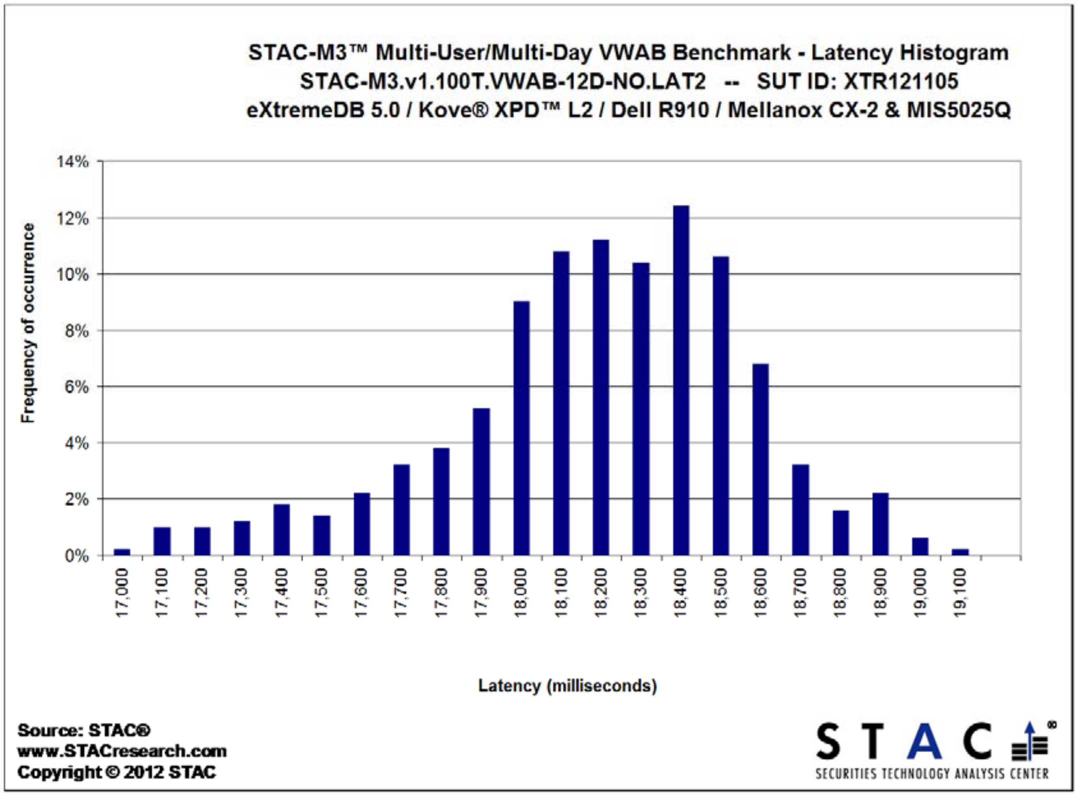


Figure 6

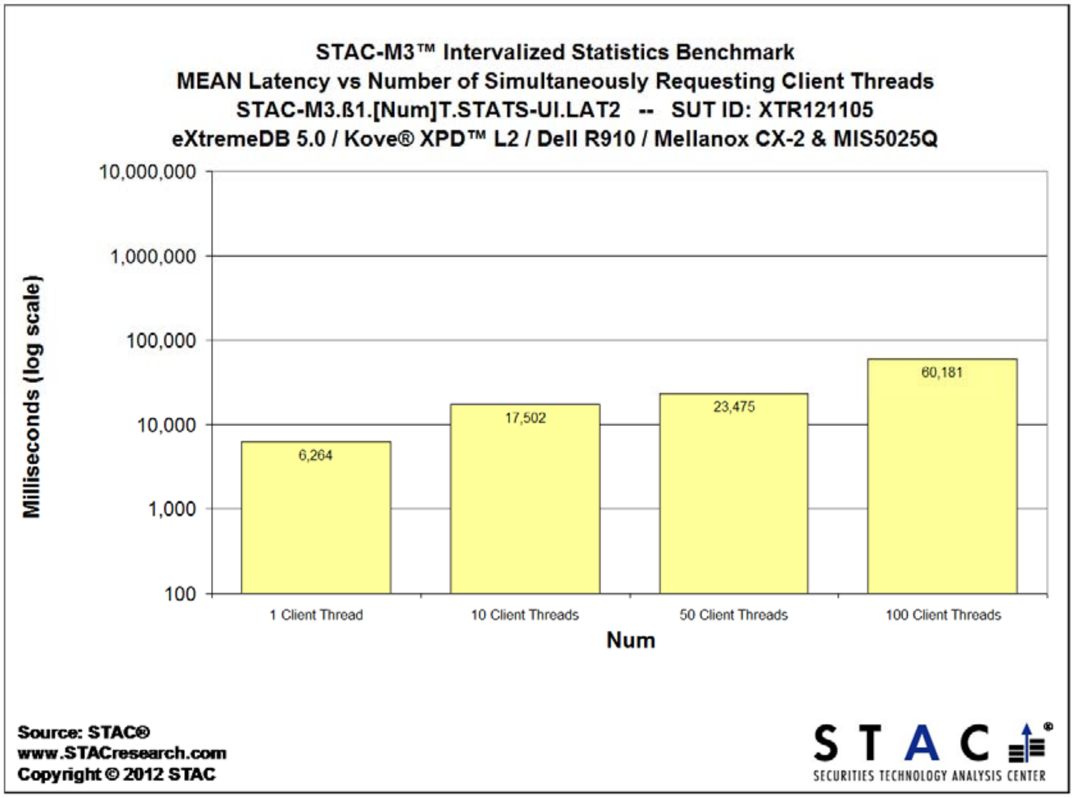


Figure 7

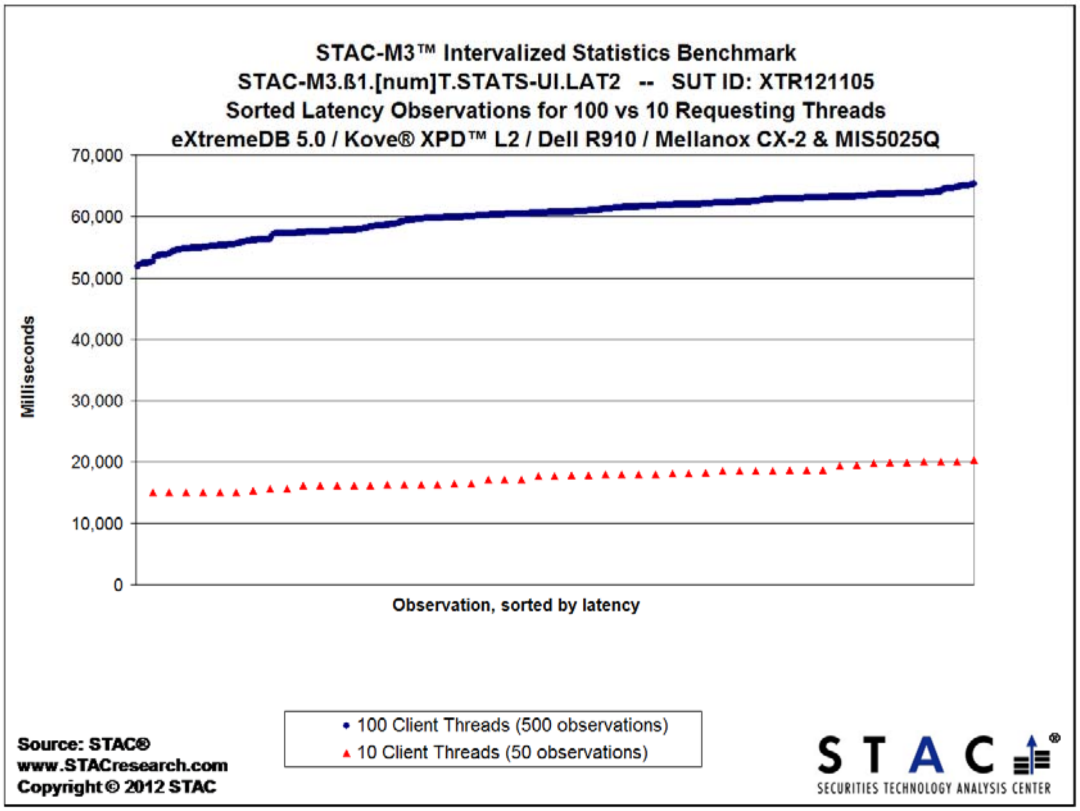


Figure 8

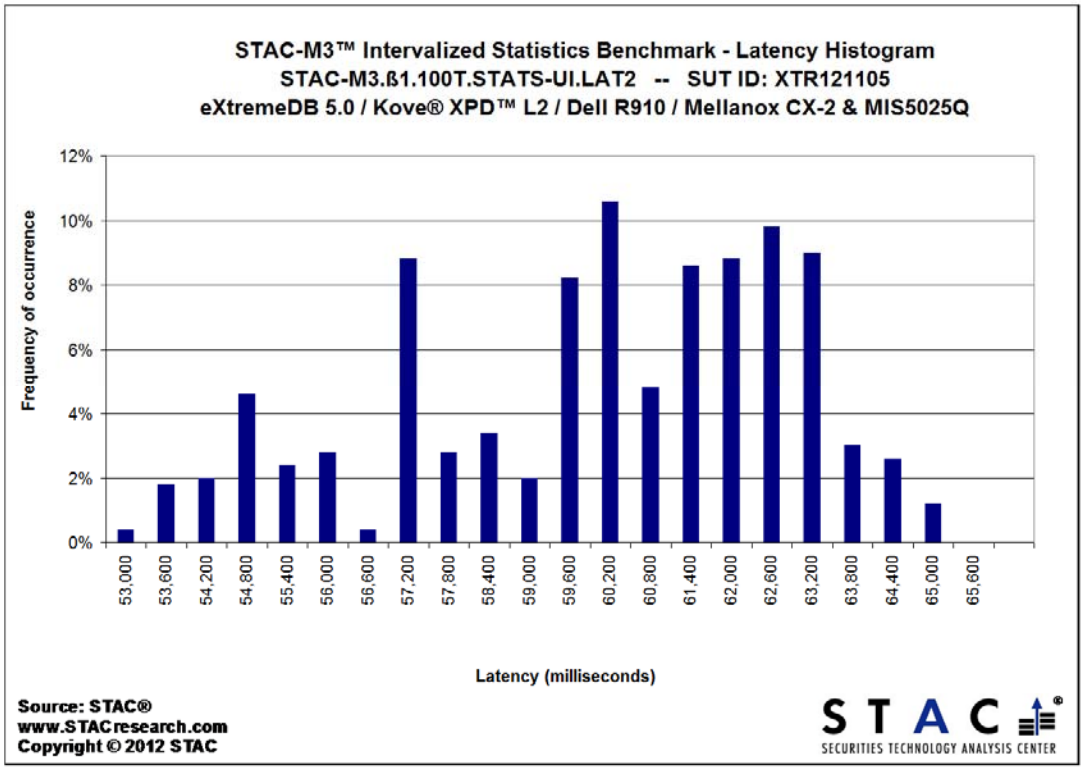

STAC-M3™ Intervalized Statistics Benchmark - Latency Histogram
STAC-M3.ß1.100T.STATS-UI.LAT2 -- SUT ID: XTR121105
eXtremeDB 5.0 / Kove® XPD™ L2 / Dell R910 / Mellanox CX-2 & MIS5025Q
Frequency of occurrence
12%
10%
8%
6%
4%
2%
0%
53,000
53,600
54,200
54,800
55,400
56,000
56,600
57,200
57,800
58,400
59,000
59,600
60,200
60,800
61,400
62,000
62,600
63,200
63,800
64,400
65,000
65,600
Latency (milliseconds)
Source: STAC®
www.STACresearch.com
Copyright © 2012 STAC
STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

Figure 9

1. Overview of the STAC-M3 Benchmark specifications

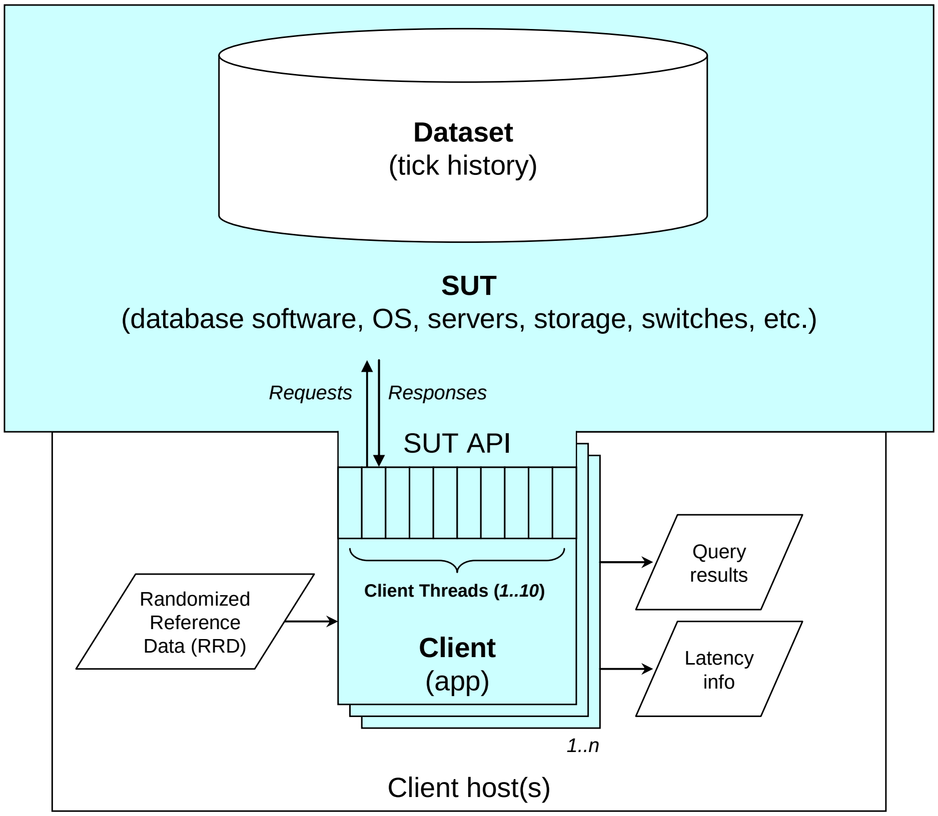
Analyzing time-series data such as tick-by-tick quote and trade histories is crucial to many trading functions, from algorithm development to risk management. But the domination of liquid markets by automated trading—especially high-frequency trading—has made such analysis both more urgent and more challenging. As trading robots try to outwit each other on a microsecond scale, they dish out quotes and trades in ever more impressive volumes. This places a premium on technology that can store and analyze that activity efficiently. For example, the faster an algorithm developer can back-test and discard a haystack of unprofitable ideas, the faster he will find the needle of a winning algorithm, leaving more time to exploit it in the market.

The STAC Benchmark Council has developed the STAC-M3 Benchmarks in order to provide a common basis for quantifying the extent to which emerging hardware and software innovations improve the performance of tick storage, retrieval, and analysis.

STAC-M3 tests the ability of a solution stack such as columnar database software, servers, and storage, to perform a variety of operations on a large store of market data. The STAC-M3 Working Group designed these test specs to enable useful comparisons of entire solution stacks (i.e., to gauge the state of the art) as well as comparisons of specific stack layers while holding other layers constant. Comparisons can include (but are not limited to):

- Different storage systems, including SSD, DRAM, interconnects, and file systems
- Different server products, processors, chipsets, and memory
- Different tick-database products

As shown below, the test setup for STAC-M3 consists of the "stack under test" (SUT) and client applications. No restrictions are placed on the architecture of the SUT or clients (though members of the STAC-M3 Working Group frequently provide input on architectures they would like to see tested). Threads within the clients take in Randomized Reference Data (RRD) such as dates and symbols, submit requests for the required operations, receive responses, and store the timings and results from these queries. Vendor-supplied code for the operations and latency calculations are subjected to a combination of source-code inspection and empirical validation.



Dataset

STAC-M3 draws from client experience with equities and FX use cases. The database is synthetic, modeled on NSYE TAQ data (US equities). While it is also desirable to test with real data, synthetic data has three advantages that make it compelling for this STAC-M3 suite:

- Synthetic data allows us to control the database properties exactly, which in turn allows us to randomize elements of queries from project to project while keeping the resulting workload exactly the same (for example, we control how much volume is associated with each symbol).
- Synthetic data does not incur fee liability from a third party such as an exchange.
- Synthesizing the data makes it easy to scale the database to an arbitrarily large size and run benchmarks against projected future data volumes.

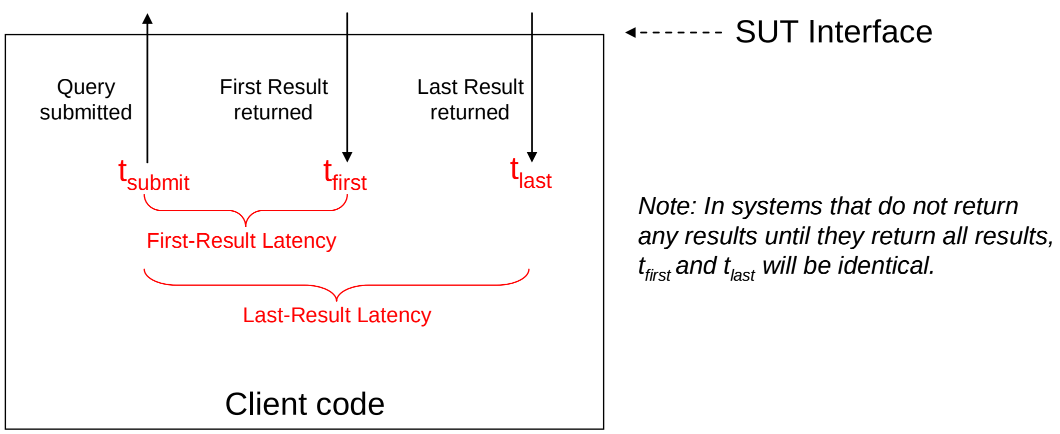
The dataset consists of high-volume symbols and low-volume symbols in proportions based on observed NYSE data. The data volume per symbol was based on doubling the typical volume in NYSE TAQ in 1Q10. The resulting database is considerably smaller than databases in use at customer sites. This was a deliberate choice by the STAC-M3 Working Group to minimize the cost of running benchmarks while still yielding valuable results. Benchmarks that scale the database to the size of existing customer footprints and well beyond are contained in the Kanaga suite of STAC-M3 Benchmark specifications.

Metrics

The key metric in STAC-M3 is the latency of query responses (aka response times). Latency measurements are performed in the clients. A client thread gets a local timestamp (t_{submit}) just before submitting a query. When the first results arrive, the client gets another timestamp (t_{first}). When it receives the complete results (sorted appropriately), the client immediately gets a third timestamp (t_{last}). For systems that return all results in one chunk, the first-result and last-result timestamp are identical. As the diagram below illustrates, latencies are defined as follows:

First-result latency (LAT1) = $t_{first} – t_{submit}$

Last-result latency (LAT2) = $t_{last} – t_{submit}$



Timestamp and latency meanings

The algorithms in all benchmarks are defined so as to keep the result sets small. This ensures that network I/O between the test clients and server(s) is negligible compared to back-end processing times.

Some of the I/O-focused benchmarks also measure the bytes read per second from persistent storage (i.e., excluding server cache), which is computed from the output of appropriate system utilities.

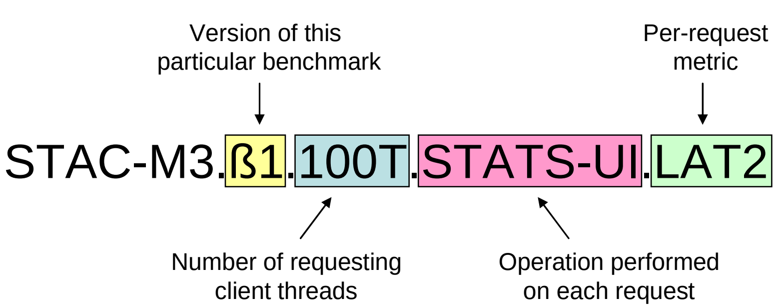
Test cases

The current tests in the STAC-M3 suite are listed in the Summary Table below. As the versioning illustrates, the STAC-M3 Working Group is bringing benchmark specifications to market in phases. The first set of approved specs (those marked as "v1" or, in one case, "v1.1") focused on storage-system performance with respect to heavy historical data loads. These workloads were deliberately light on compute and heavy on I/O. The second phase, (the Antuco suite, see [1]) added benchmark specs that involve more compute-intensive analytics. These new specifications, marked as "ß1", have not been put to a vote by the full STAC Benchmark Council and will become v1 specs if approved. These new benchmarks operate by symbol on many fields of underlying tick data for both trades and quotes across varying time windows.[1] The table classifies each test case as relatively heavy on I/O, compute, or both.

The tests require a client application that is written to a product API and is capable of submitting requests from 10 independent threads. As detailed in the table, some of the benchmarks call for one client instance making requests from a single thread, while others call for one client using 10 threads, and still others require 10 clients each using 10 threads (100 total requesting threads). One set of benchmarks (using the STATS-UI operation) tests multi-user scaling by running with 1, 10, 50, and 100 client threads and allowing the tester to scale to even higher numbers of concurrent threads. In all cases, benchmark results refer to per-request latency. For example, the mean of 10T.MKTSNAP.LAT2 is the mean time to satisfy a single market-snapshot request, not the total time to satisfy requests from all 10 client threads.

The range of dates eligible for querying depends on the benchmark. For example, some algorithms operate on dates randomly chosen throughout the year, some stick to a recent date range, and some always run on the most recent date (see the "Input Date Range" column of the table). The purpose of this is to provide a realistic optimization strategy for systems with multiple storage tiers of different speed, such as solid-state disk (SSD) and spindle-based storage. For example, suppose the tester wanted to demonstrate the benefit of adding an SSD to a system that otherwise relied on spindle-based storage. Furthermore, suppose the SSD was only large enough to store 1/5 of the test database. Benchmarks that accessed dates throughout the entire database would show some performance improvement when the SSD is added, since 1/5 of the queries would enjoy acceleration. In the real world, however, customers don't simply add faster storage and randomly allocate data to it. Rather, they typically reserve the fastest storage tier for the most frequently accessed data or data accessed by especially time-sensitive algorithms. In STAC-M3, the tester can load the most recent data into the fastest storage. In the case above, the benchmarks that operate on, say, the most recent day (1/252 of the database) or the most recent month (1/12 of the database) will show the maximum improvement possible from the SSD, since their entire queries can be satisfied from SSD. This provides the clearest indication of acceleration possibilities for datasets that can fit within the faster tier.

The STAC-M3 Report Card and accompanying charts identify each benchmark unambiguously, as follows:



In charts, the ID is sometimes decomposed, with part of it in the chart title or labels. Each individual STAC Benchmark™ specification has its own version number. The same version of a given spec may appear in multiple benchmark suites. Thus, the code names of the suites are irrelevant when making comparisons. Versioning individual specs enables the reader to compare a discrete result from this "stack under test" (SUT) to the corresponding result from another SUT. When making comparisons, be sure that the identifiers match exactly. If they do not, the benchmark results cannot be fairly compared.

[1] Future phases will supplement these with benchmarks that provide insight into additional aspects of system performance. Contact council@STACresearch.com if you would like to be part of the process.

Summary Table – STAC-M3 Benchmarks in the Antuco Suite

The table below gives a brief overview of each test in this STAC-M3 suite. Version numbers of 1 or greater indicate benchmark specs that have been approved. Versions less than 1 are proposed by the STAC-M3 Working Group but not yet voted on by the full STAC Benchmark Council.

Root ID	Operation name	Ver	Number of requesting Client Threads	Algorithm performed on behalf of each requesting Client Thread	Algorithm I/O intensity	Algorithm compute intensity	Input date range*
VWAB-D	VWAB-Day	1	1	4-hour volume-weighted bid over one day for 1% of symbols (like VWAP but operating on quote data, so much higher input volume).	Heavy read	Light	Last 30 days
VWAB-12D-NO	VWAB-12DaysNoOverlap	1	100	4-hour volume-weighted bid over 12 days for 1% of symbols. No overlap in symbols among client threads.	Heavy read	Light	Full year
YRHIBID	Year High Bid	ß1	1	Max bid over the year for 1% of symbols.	Heavy read	Light	Full year
YRHIBID-2	Year High Bid Re-run	ß1	1	Re-run of YRHIBID (same symbols) without clearing the cache.	Heavy read†	Light	Full year
QTRHIBID	Quarter HighBid	ß1	1	Max bid over the quarter for 1% of symbols.	Heavy read	Light	Most recent quarter
MOHIBID	Month High Bid	ß1	1	Max bid over the month for 1% of symbols.	Heavy read	Light	Most recent month
WKHIBID	Week High Bid	ß1	1	Max bid over the week for 1% of symbols.	Heavy read	Light	Most recent week
STATS-AGG	Aggregate Stats	ß1	10	One set of basic statistics over 100 minutes for all symbols on one exchange. Each 100-minute range crosses a date boundary.	Heavy read	Heavy	Full year

STATS-UI	Stats - Unpredictable Intervals	ß1	1, 10, 50, 100 (more optional)	Per-minute[‡] basic statistics over 100 minutes for all high-volume symbols on one exchange. Each 100-minute range crosses a date boundary.	Heavy read	Heavy	Full year
MKTSNAP	Market Snapshot	ß1	10	Most recent trade and quote information for 1% of symbols as of a random time.	Heavy read	Heavy	Full year
VOLCURV	Volume Curves	ß1	10	Create an average volume curve (using minute intervals aligned on minute boundaries) for 10% of symbols over 20 days selected at random.	Light read	Heavy	Full year
THEOPL	Theoretical P&L	ß1	10	For a basket of 100 trades on random dates, find the future times at which 2X, 4X, and 20X the trade size traded in each symbol. Trade sizes cause up to 5 days of forward searching. Calculate the corresponding VWAP and total volume traded over those periods.	Light read	Heavy	Full year
NBBO	NBBO	ß1	1	Create the NBBO across all 10 exchanges for all symbols on the most recent day. Write to persistent storage.	Heavy read and write	Heavy	Most recent day
WRITE	Write	1	1	Write one day's quote data to persistent storage, following the same algorithm used to generate the randomized dataset used in the other Operations.	Heavy write	Light	n/a
STORAGE.EFF	Storage efficiency	1.1	n/a	Reference Size of the Dataset divided by size of the Dataset in the SUT format used for the performance benchmarks. Expressed as as percentage.	n/a	n/a	n/a

[*] In some cases, one or more dates at the end of the year were excluded from eligibility to prevent an algorithm that crosses days from running out of input data.

[†] Typically this will be reads from DRAM cache.

[‡] In this case, interval start times are offset from minute boundaries by a consistent random amount per test run, so that the SUT cannot rely on pre-calculated minute statistics.

2. Product background

The stack under test (SUT) included the following:

- McObject eXtremeDB 5.0 Financial Edition
- Kove® XPD™ L2 Storage System with Mellanox QDR InfiniBand, dual port
- Dell PowerEdge™ R910 server
- Intel Xeon E7-4850 Processors
- CentOS Release 6.2 Final
- Mellanox MT26428 Connect®X-2 QDR InfiniBand, dual port HCA
- Mellanox MIS5025Q InfiniScale-IV QDR InfiniBand switch

McObject submitted the following information and claims about its products:

eXtremeDB has been used for over 11 years in high performance applications in Military, Defense, Avionics, Netcom, and other markets. There are hundreds of companies using the development software and over 20 million copies deployed in applications around the world. McObject developed the eXtremeDB Financial Edition database management system to break through the financial IT data management bottleneck. The volume of data flowing through today's automated capital markets is skyrocketing, and success for financial technology hinges on acting on this information instantly. But conventional relational database management systems cannot deliver the needed speed and flexibility.

eXtremeDB Financial Edition leverages proven eXtremeDB embedded database strengths – including a streamlined in-memory database system (IMDS) design, multi-core optimization, maximum developer flexibility and high scalability – and adds specialized features to address key financial data management challenges. These features include:

- *Columnar data layout for fields of type 'sequence'. Sequences can be combined to form a time series, ideal for working with tick streams, historical quotes and other sequential data*
- *A rich library of vector-based math functions that accelerate management of time series data by maximizing L1/L2 cache use. Functions can be pipelined to form an assembly line of operations on sequences in support of statistical/quantitative analysis*
- *Improved clustering technology that boosts distributed database performance and scalability via the ability to define tables as "local" (exempt from replication), but shareable through a scatter/gather mechanism*

These enhancements, and eXtremeDB's rich core features and many specialized options, combine to make eXtremeDB Financial Edition the fastest database system with the most flexibility for financial applications.

Architecture for Minimizing Latency
eXtremeDB's roots are in real-time embedded systems. How does it deliver a response that is fast and predictable enough for mission-critical avionics and defense electronics, industrial controllers and telecom/networking gear, as well as for ultra-low latency trading?

In-memory and/or persistent tables. *eXtremeDB's design is based on a core in-memory database system that eliminates performance-draining I/O, cache management, data transfer, and other sources of latency that are hard-wired into traditional disk-based relational database management systems (RDBMSs). McObject's relentless focus on efficiency extends to persistent table management.*

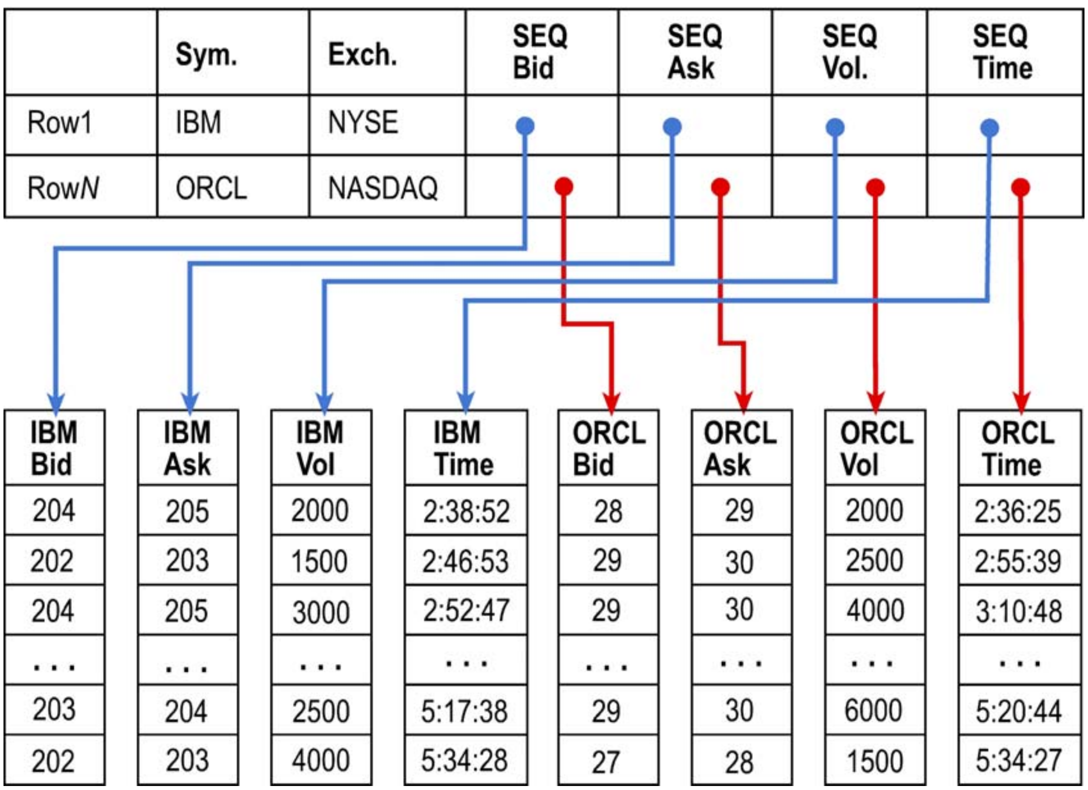
Embedded. *eXtremeDB runs entirely within the application process, eliminating inter-process communication (IPC) between client and server modules. In contrast, IPC messaging is an inherent source of latency in RDBMSs and in some in-memory database systems based on client/server design.*

<u>Short execution path</u>. eXtremeDB is written in C, and its code path is very short: a "footprint" of approximately 150K for the core database system points to McObject's unrelenting focus on eliminating even small potential sources of latency.

<u>Programming interfaces</u>. While eXtremeDB supports standard SQL, developers can use a direct access C/C++ API for faster and more predictable performance (i.e. to minimize latency spikes). eXtremeDB also provides the ability to store records directly as C/C++ data types (including structures, vectors and arrays), eliminating the overhead of conversion to SQL data types. Date, time and timestamp (down to the nanosecond) are fully supported with all eXtremeDB Financial Edition programming interfaces.

<u>Memory management</u>. eXtremeDB performs its own memory management instead of relying on the programming language run-time. This contributes to its performance advantage and scalability for multi-core configurations.

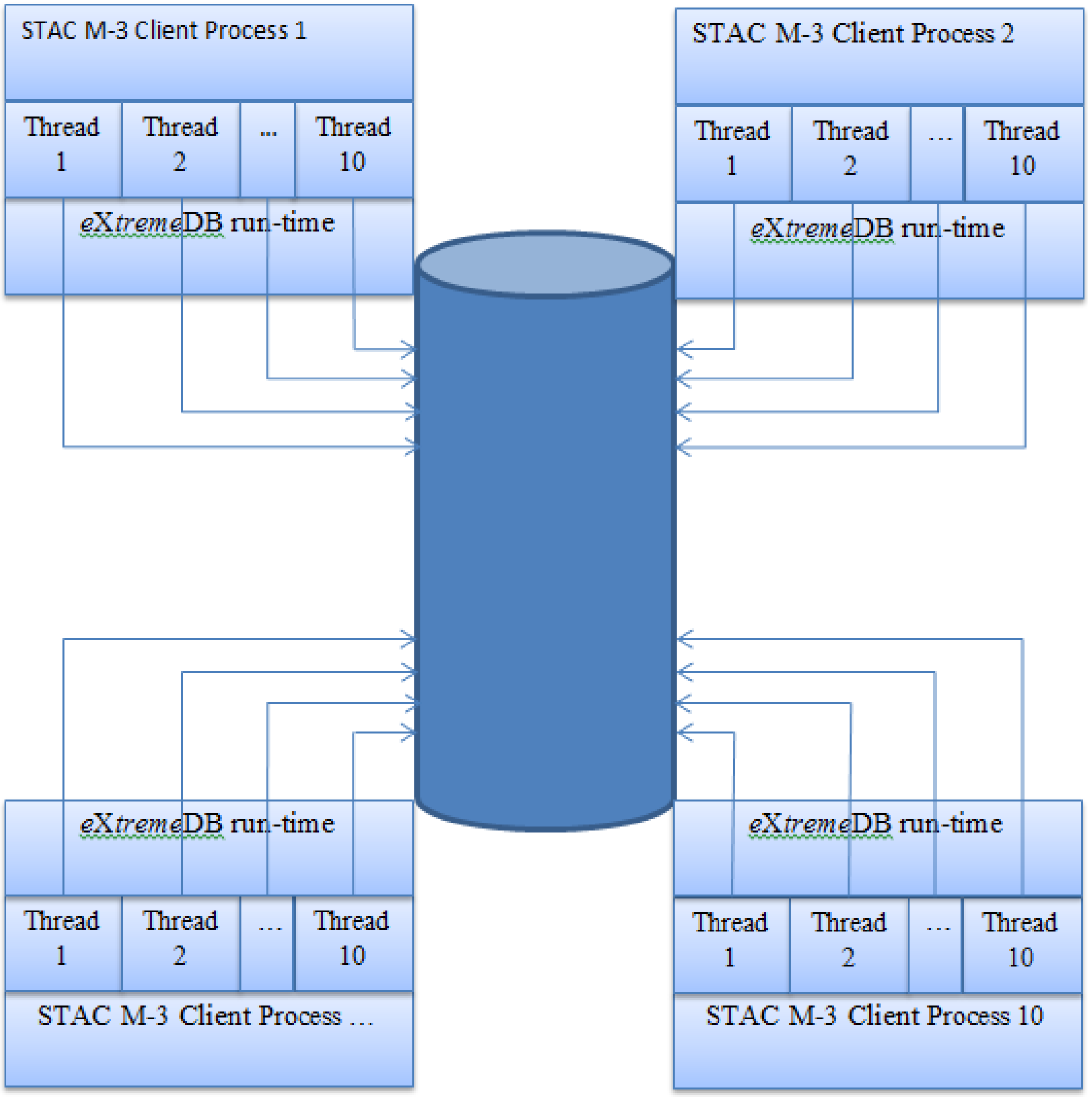
<u>CPU cache efficiency</u>. Keeping code and data in CPU (L1/L2) cache eliminates costly (in performance terms) fetches. eXtremeDB Financial Edition's column-based data layout for sequences maximizes L1/L2 cache efficiency when working with tick and quote data, and its small code size maximizes the likelihood that the entire code path for a given operation is loaded into the cache at once.



<u>Traditional DBMSs bring rows of data into L1/L2 cache for processing</u>. But financial data – such as ticks, trades and quotes – are better handled by a column-based layout that maximizes efficiency in fetching needed information. The result is higher performance: the database system benefits from L1/L2 cache speed and avoids fetching data via the much slower front-side bus.

STAC-M3 Implementation

The core of the STAC-M3 implementation using eXtremeDB (the STAC-M3 Pack for eXtremeDB[2]) is a single C program using eXtremeDB's native (direct access) API, as shown in the diagram below.



These direct access database function calls provide near-memory performance. eXtremeDB requires no knowledge of a proprietary language. Alternatively, C++, Java, or C# could also have been chosen (though performance may vary). Columnar storage of sequence (time series) data contributed to the high performance numbers. Use of SQL/ODBC/JDBC is also an option, though columnar sequences are not yet supported with that API and thus would have necessitated a different (less efficient) database design.

For more information please visit http://financial.mcobject.com.

2 Premium members of the STAC Benchmark Council may request access to the STAC Pack code at http://www.stacresearch.com/node/13103.

Kove submitted the following information and claims about its products:

Kove, a leading high performance storage vendor, produces the storage industry's highest-density, highest performance memory-disk SSD storage. Kove® XPD™ L2 delivers exceptional price-to-performance ratios, functionality, scalability and ease of use for any storage user. What does this mean for Financial Services Trading Firms? Competitive advantage and profitability.

Organizations using Kove® storage run I/O applications faster than those who do not, plain and simple. If speed matters, investing in Kove® storage means immediate profitability advantages – regardless of the I/O profile. Kove® storage delivers deterministic performance without concern for media wear, capacity fill, sequential versus random usage patterns, small or large data, burst, bandwidth, latency, or IOPS. Rather than using flash or spinning disk, Kove® DRAM performance is deterministic, sustainable, predictable, and maintains consistent performance even with 100% occupancy. It never wears out.

The Kove® XPD™ L2 easily connects directly to existing storage fabrics via Fibre Channel and InfiniBand while increasing performance metrics beyond any other storage option. Since the Kove® XPD™ L2 can operate as a standard mesh sharable device in addition to direct-connect, it is easy to deploy, manage, and scale. Sharing data that exceeds any single unit's capacity is deterministically accommodated without performance compromise.

For more information please visit www.kove.com.

Dell submitted the following information and claims about its products:

Dell Inc. (NASDAQ: DELL) listens to customers and delivers innovative technology and services that give them the power to do more.

For enterprises in the banking and securities industries, Dell provides solutions that solve customers' IT challenges in performance, storage, security, and innovation in ways that are cost-effective and efficient. The Dell PowerEdge™ R910 server is ideal for workloads needing the highest performance, reliability and I/O scalability including Tick Data/Tick Analytic databases and Market Data analytic requirements. The Dell PowerEdge™ R910 is available today and is supported through your Dell Account Teams. For more information, please contact Dell through the contact listed in Section 4.

For more information please visit www.dell.com.

Mellanox submitted the following information and claims about its products:

Mellanox Technologies (NASDAQ: MLNX, TASE: MLNX) is a leading supplier of end-to-end InfiniBand and Ethernet connectivity solutions and services for servers and storage. Mellanox products optimize data center performance and deliver industry-leading bandwidth, scalability, power conservation and cost-effectiveness while converging multiple legacy network technologies into one future-proof architecture. The company offers innovative solutions that address a wide range of markets including HPC, financial services, enterprise, mega warehouse data centers, cloud computing, Internet and Web 2.0.

Mellanox offers the lowest latency networking solutions for high frequency trading with its unique end-to-end solution (adapters, switching platforms and software), leveraging both InfiniBand and 10/40 Gigabit Ethernet technologies. As a pioneer in the field of high frequency trading networking, Mellanox has deployed its solutions at a large array of investment banks, hedge funds and exchanges. These solutions have been certified and deployed with the leading software providers in this industry.

Founded in 1999, Mellanox Technologies is headquartered in Sunnyvale, California and Yokneam, Israel.

For more information, please visit www.mellanox.com.

R-HPC submitted the following information and claims about its services:

> *R Systems NA, Inc. is an on-demand, High Performance Computing cluster resource provider serving in the commercial and academic research communities under the R-HPC brand of products. Every cluster configuration is designed to fit the client's specific usage requirements and packaged with world-class "R-Team" systems administration support. With this support, users avoid tedious systems management, making every project a value-added experience. The results are reduced time to solution, saving costs and increased efficiency.*
>
> *R Systems' R-HPC configurations include optimal core count, networking, memory aggregation and security for cluster usage on a "utility" or "dedicated" basis. Customers utilize over 1600 servers in two separate locations under flexible Service Level Agreements. In designing customer solutions, R Systems partners with both leading edge hardware and software companies to bring the most productive solution to optimizing, scaling-up or revolutionizing the customers' HPC applications.*
>
> *Networking and physical security is a primary component in R Systems' project approach. We are glad to provide complete details on the various network access methods employed and security apparatus in-place.*
>
> ***For more information, please visit www.r-hpc.com.***

3. Project participants and responsibilities

The following firms participated in the project:

- McObject
- Kove
- Dell
- Mellanox
- R Systems
- STAC

The Project Participants had the following responsibilities:

- McObject implemented the STAC-M3 Clients and Operations using the STAC-M3 Benchmark specifications.
- Kove supplied the Kove® XPD™ L2 storage hardware for the test.
- Dell provided SUT host (Dell PowerEdge™ R910 server) for the test.
- Mellanox provided the InfiniBand interconnect solutions for this test.
- R Systems provided the datacenter infrastructure, SUT host (in cooperation with Dell), and network, as well as the setup, configuration, system-optimization services and comprehensive integration expertise for this test.
- McObject sponsored the Audit.
- STAC conducted the STAC-M3 Benchmark Audit, which included validating the database; conducting a code review with end-user members of the STAC Benchmark Council; validating the Operation results; executing the tests; and documenting the results.

4. Contacts

- McObject: Chris Mureen, COO, chris.mureen@mcobject.com, +1 425 888 8505 x211
- Kove: Jim Hetherington, Director of Business Development, jim.hetherington@kove.com, +1 505 379 3125, www.kove.com
- Dell: Mark King, Global Strategy & Business Development Director, mark_king@Dell.com, +1 203 840 1631, www.dell.com
- Mellanox: Asaf Wachtel, Director of Business Development FSI, asafw@mellanox.com, +972 74 7129306, www.mellanox.com
- R Systems: Dan Sholem, Project Director, dan@r-hpc.com, +1 217 954 1056

5. Results status

- These benchmark specifications were developed by the STAC-M3 Working Group of the STAC Benchmark Council. Benchmarks with a “v1” or higher have been approved by the full Council. Those with a “ß” designation have been proposed by the STAC-M3 Working Group but have not yet been approved.
- These test results were audited by STAC or a STAC-certified third party, as indicated in the Responsibilities section above. As such, they are official results. For details, see www.STACresearch.com/reporting.
- The vendors attest that they did not modify the SUT during the Audit.

6. Specifications

This project followed the Antuco suite of STAC-M3 Benchmark specifications. Full members of the STAC Benchmark Council can access these specifications at www.STACresearch.com/m3. Premium subscribers to the Market Data or Trade Flow STAC Tracks can download the programs used in these benchmarks in order to run the same tests on systems in the privacy of their own labs.

7. Limitations

- As discussed in Section 1, this suite of STAC-M3 Benchmarks was designed to test operations on a limited amount of purely historical data. Tests involving much larger amounts of historical data are available in the Kanaga suite of STAC-M3 Benchmark specifications.
- As discussed in Section 1, the dataset used in this version of STAC-M3 is synthetic. The algorithm to generate the dataset creates random values for prices and sizes that can vary widely from tick to tick. In the real world, by contrast, there is significant correlation of successive prices (i.e., large differences from tick to tick are relatively rare). Compression algorithms often take advantage of this fact, such as by focusing on deltas between successive values. Hence, the storage efficiency of a SUT may be higher when working with real data than with the synthetic dataset of this version of STAC-M3.

8. Stack under test

Detailed configuration information is available to premium members of the STAC Benchmark Council. If this document is not in your myVault account (www.STACresearch.com/myVault), and you would like to request access, use the request form at: www.STACresearch.com/node/13076

9. Vendor Commentary

McObject provided the following comments:

- *eXtremeDB returned the complete results of queries at one time, so the concept of "first result" (LAT1) had no meaning in this context.*
- *Although not required by STAC-M3, eXtremeDB also supports transactions with the ACID properties (ACID is an acronym for Atomicity, Consistency, Isolation and Durability). This protects the integrity of the data being managed, preventing corruption and facilitating concurrent access from multiple processes/threads while maintaining a consistent view of the database for each thread/process.*

10. STAC Notes

STAC-M3 calls for the Linux “iostat” utility to be used to measure the number of bytes read per second in the various HIBID benchmarks. However, this SUT configuration did not support iostat. In order to measure bytes read, we therefore made separate test runs with a configuration that did support iostat, using the same randomized reference data (RRD). But the latencies used to calculate bytes per second were those measured with the original SUT configuration. (See the Configuration Disclosure for details).