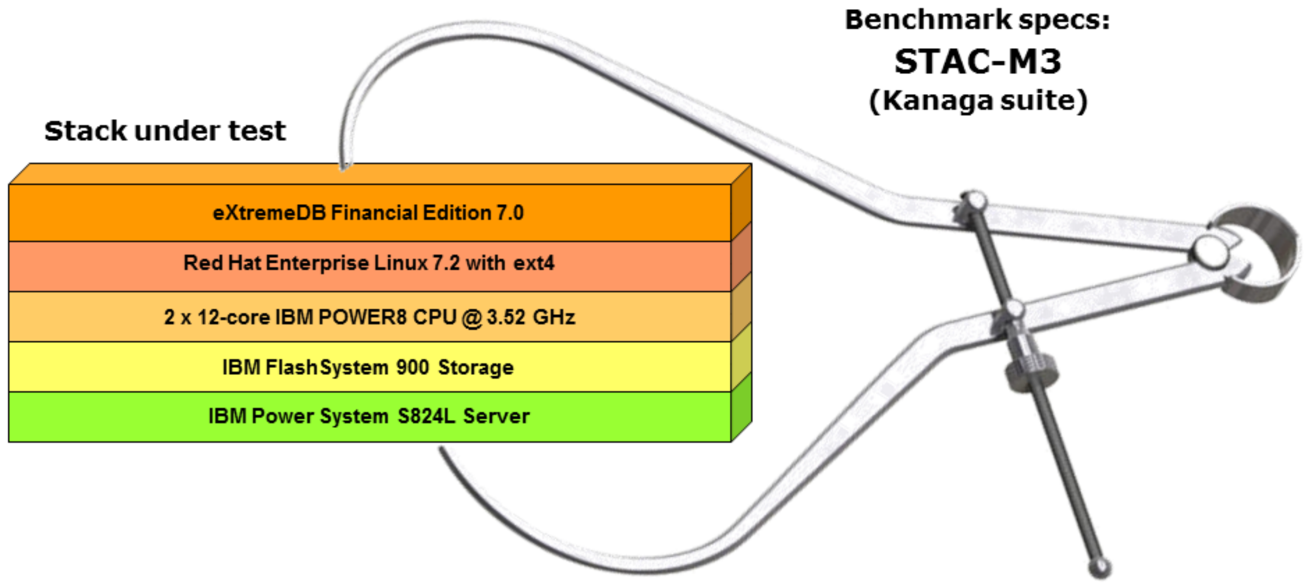
eXtremeDB Financial Edition 7.0 on IBM Power System S824L with IBM FlashSystem® 900 Storage

SUT ID: XTR160413

STAC-M3™ KANAGA BENCHMARK SUITE (Volume-Scaling Benchmarks)

Test date: 21 April 2016

Draft 1.0.0, 5 May 2016



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, Kanaga Suite, Rev H – https://stacresearch.com/stac-m3-benchmark-specs-kanaga-rev-h-optional-scaling. Accessible by qualified members of the STAC Benchmark Council.

Disclaimer

Summary

STAC recently performed the volume- and user-scaling suite of STAC-M3™ Benchmarks (Kanaga) on a stack involving McObject's eXtremeDB Financial Edition 7.0 hosted on an IBM Power System S824L server with two 12-core IBM POWER8 CPUs connected to a 12 x 5.7 TB IBM FlashSystem 900 storage system. This report documents the benchmark results. The McObject code used to execute these benchmarks is available from the STAC Vault for inspection and use by qualified STAC Benchmark Council members.

Test results for this system using the baseline suite of STAC-M3 Benchmarks (Antuco) is available at https://www.stacresearch.com/XTR160413.

The Kanaga suite of STAC-M3 Benchmarks delivers 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:

- *Fastest mean response times and most consistent response times (lowest standard deviation) ever reported, for all combinations of query type, data volume, and concurrent users.*
- *Each mean response time was 5.5x to 212x the previous best result, including:*
  - *21x to 212x the performance of the previous best published result for the market snap benchmarks (10T.YR[n]-MKTSNAP.TIME) **
  - *21x the performance of the previous best published result for year high bid in the smallest year of the dataset (1T.OLDYRHIBID.TIME) **
  - *8-10x the performance of the previous best published result for the 100-user volume-weighted average bid benchmarks (100T.YR[n]VWAB-12D-HO.TIME) ***
  - *5-8x the performance of the previous best published result for the N-year high-bid benchmarks (1T.[n]YRHIBID.TIME) **

* *Previous record held by SUT ID KDB121101.*
** *Previous record held by SUT ID KDB150528.*

A STAC-M3 Kanaga 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.

Getting the most from these results

Any interested party can analyze public STAC Reports to compare the performance of different systems. However, members of the STAC Benchmark Council are able to put these reports to much greater use. Qualified members may:

- Read the detailed system configuration.
- Read the detailed test specifications
- Access additional reports in the confidential STAC Vault™
- Obtain the materials to run the STAC-M3 Benchmarks on their own systems
- Discuss benchmarks, technologies, and related business issues with their peers.

To join the Council or upgrade your membership, please contact council@STACresearch.com.

STAC Report Card

STAC-M3™ (Kanaga) Benchmarks for SUT ID XTR160413: eXtremeDB Financial Edition 7.0 on IBM Power System S824L with IBM FlashSystem 900 Storage

High Bid Over Varying Intervals

Return the high bid for a certain 1% of symbols over a particular range of years in the dataset.

	Raw last-result latency (LAT2) - Milliseconds		Volume-adjusted last-result latency (VLAT2) - Milliseconds		Megabytes read per second*	
Spec ID	MEAN	MAX	MEAN	MAX	MEAN	MAX
STAC-M3.ß1.1T.OLDYRHIBID	68	155	2,923	6,664	2,389	2,977
STAC-M3.ß1.1T.YRHIBID	485	507	485	507	8,788	507
STAC-M3.ß1.1T.2YRHIBID	1,335	1,604	514	617	8,227	9,318
STAC-M3.ß1.1T.3YRHIBID	2,505	2,856	485	553	8,790	9,551
STAC-M3.ß1.1T.4YRHIBID	4,359	4,856	471	525	9,038	9,693

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

Market Snapshot Within Varying Years (10 Client Threads Requesting)

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

	Raw last-result latency (LAT2) - Milliseconds				
Spec ID	MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.10T.MKTSNAP	78	77	66	99	8
STAC-M3.ß1.10T.YR2-MKTSNAP	147	146	131	163	8
STAC-M3.ß1.10T.YR3-MKTSNAP	23	22	5	44	11
STAC-M3.ß1.10T.YR4-MKTSNAP	24	24	5	45	11

	Volume-adjusted last-result latency (VTIME) - Milliseconds				
Spec ID	MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.10T.MKTSNAP	78	77	66	99	8
STAC-M3.ß1.10T.YR2-MKTSNAP	92	91	82	102	5
STAC-M3.ß1.10T.YR3-MKTSNAP	9	8	2	17	4
STAC-M3.ß1.10T.YR4-MKTSNAP	6	6	1	11	3

STAC Report Card (cont'd)

STAC-M3™ (Kanaga) Benchmarks for SUT ID XTR160413

Multi-Day VWAB with Varying Concurrent Requests and within Varying Years

To each of n Client Threads querying a set of symbols on 12 random days in the given year of the dataset, return the 4-hour volume-weighted bid for each date and symbol. Date/symbol combinations are designed with heavy overlap among threads.

		Raw last-result latency (LAT2) - Milliseconds				
Spec ID	**Concurrent Requests (Client Threads)**	MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.1T.YR1VWAB-12D-HO	1	583	488	396	864	203
STAC-M3.ß1.50T.YR1VWAB-12D-HO	50	4,622	4,751	81	5,653	778
STAC-M3.ß1.100T.YR1VWAB-12D-HO	100	2,339	2,416	174	5,001	1,901
STAC-M3.ß1.1T.YR2VWAB-12D-HO	1	822	680	579	1,531	399
STAC-M3.ß1.50T.YR2VWAB-12D-HO	50	5,845	5,785	112	7,834	1,054
STAC-M3.ß1.100T.YR2VWAB-12D-HO	100	3,046	2,876	269	6,221	2,267
STAC-M3.ß1.1T.YR3VWAB-12D-HO	1	1,320	992	873	2,779	818
STAC-M3.ß1.50T.YR3VWAB-12D-HO	50	7,769	7,965	147	9,314	1,245
STAC-M3.ß1.100T.YR3VWAB-12D-HO	100	4,341	3,954	487	9,178	3,161
STAC-M3.ß1.1T.YR4VWAB-12D-HO	1	2,234	2,019	1,853	3,281	599
STAC-M3.ß1.50T.YR4VWAB-12D-HO	50	10,674	10,886	255	12,279	1,627
STAC-M3.ß1.100T.YR4VWAB-12D-HO	100	7,030	9,851	820	12,932	4,457

		Volume-adjusted last-result latency (VTIME) - Milliseconds				
Spec ID	**Concurrent Requests (Client Threads)**	MEAN	MED	MIN	MAX	STDV
STAC-M3.ß1.1T.YR1VWAB-12D-HO	1	583	488	396	864	203
STAC-M3.ß1.50T.YR1VWAB-12D-HO	50	4,622	4,751	81	5,653	778
STAC-M3.ß1.100T.YR1VWAB-12D-HO	100	2,339	2,416	174	5,001	1,901
STAC-M3.ß1.1T.YR2VWAB-12D-HO	1	321	266	226	598	156
STAC-M3.ß1.50T.YR2VWAB-12D-HO	50	2,283	2,260	44	3,060	412
STAC-M3.ß1.100T.YR2VWAB-12D-HO	100	1,190	1,123	105	2,430	886
STAC-M3.ß1.1T.YR3VWAB-12D-HO	1	515	388	341	1,086	319
STAC-M3.ß1.50T.YR3VWAB-12D-HO	50	3,035	3,111	57	3,638	486
STAC-M3.ß1.100T.YR3VWAB-12D-HO	100	1,696	1,545	190	3,585	1,235
STAC-M3.ß1.1T.YR4VWAB-12D-HO	1	545	493	452	801	146
STAC-M3.ß1.50T.YR4VWAB-12D-HO	50	2,606	2,658	62	2,998	397
STAC-M3.ß1.100T.YR4VWAB-12D-HO	100	1,716	2,405	200	3,157	1,088

Chart view

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

- Figure 1 plots the mean last-result latency (LAT2) and volume-adjusted last-result latency (VLAT2) for all of the HIBID operations.
- Figure 2 plots the median and maximum last-result latencies (LAT2) for all of the MKTSNAP operations. To understand why median and max were chosen for this chart instead of mean, see the Limitations section.
- Figure 3 analyzes the individual latency observations for MKTSNAP benchmarks by sorting the results by latency.

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. The axes in Figure 3 are not fixed.

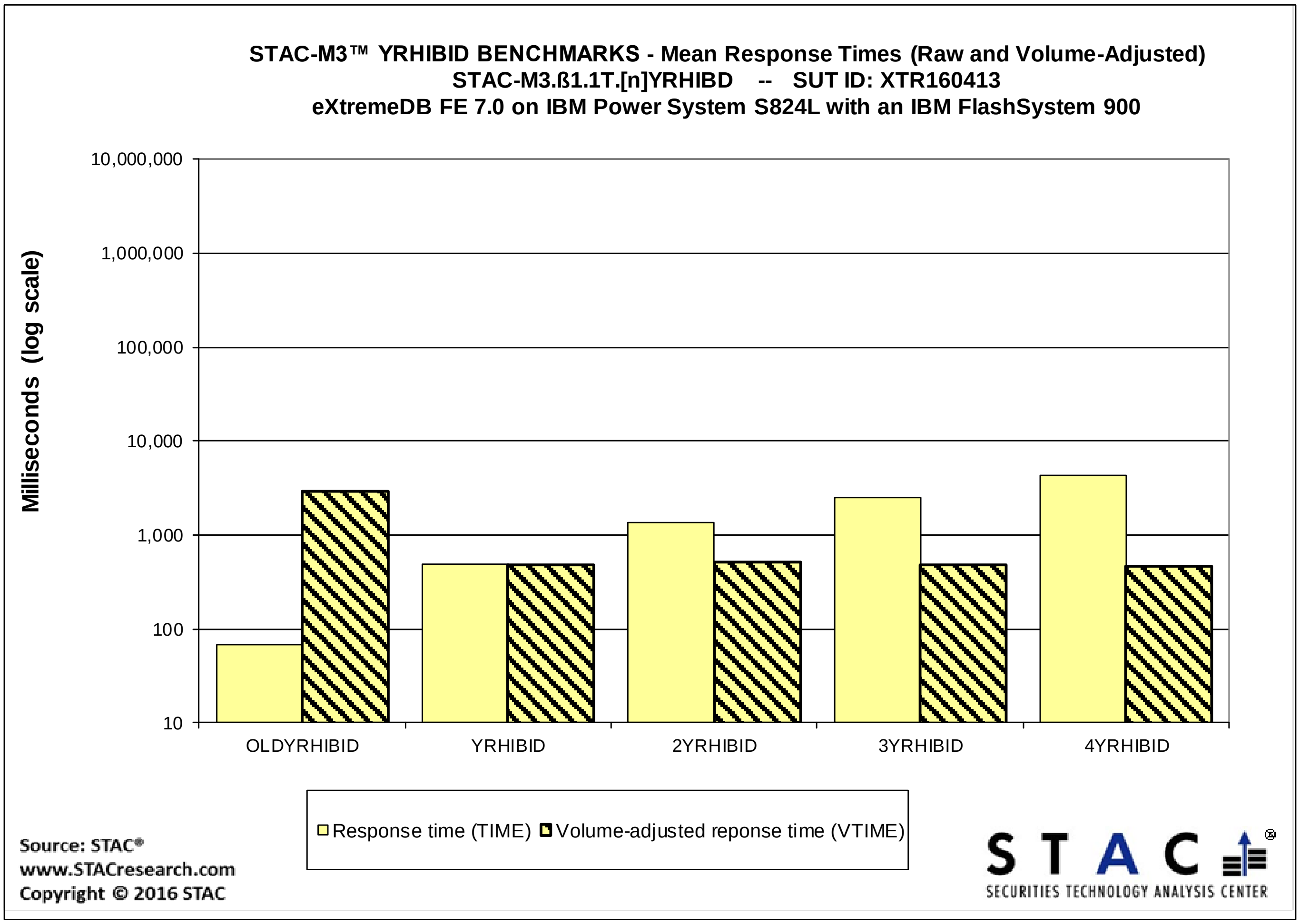


Figure 1

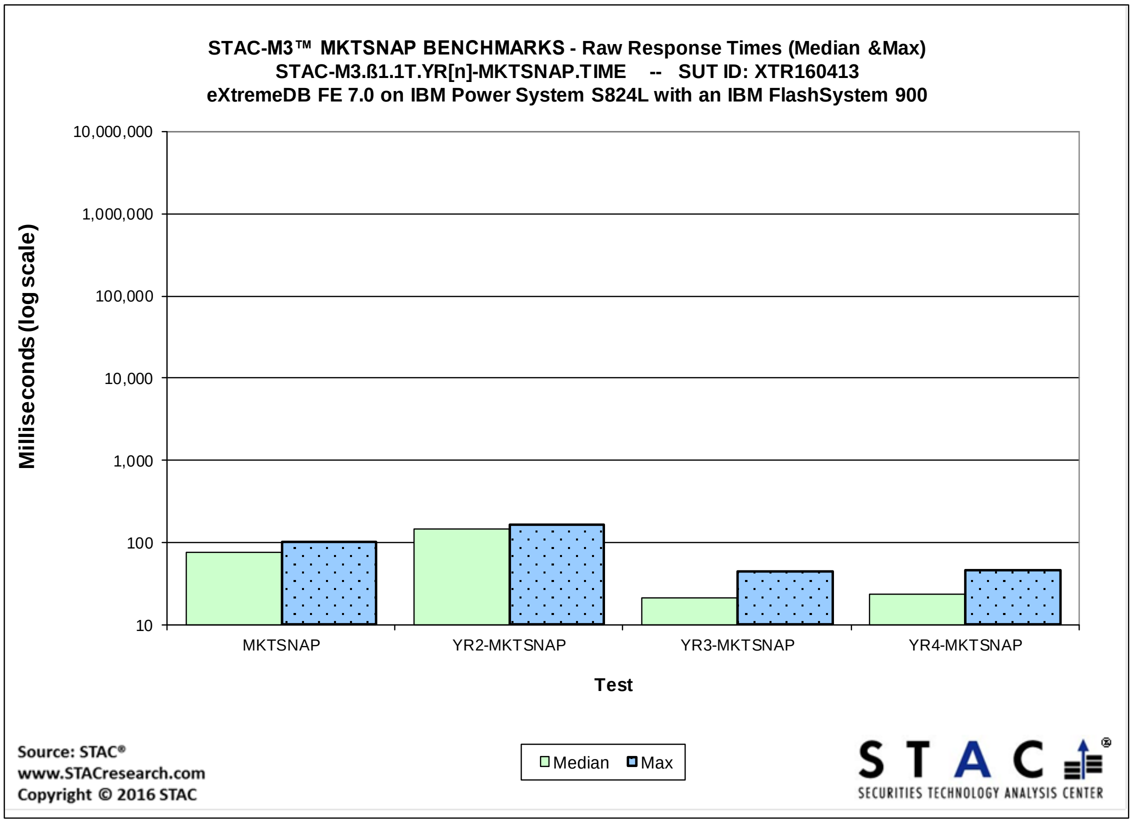


Figure 2

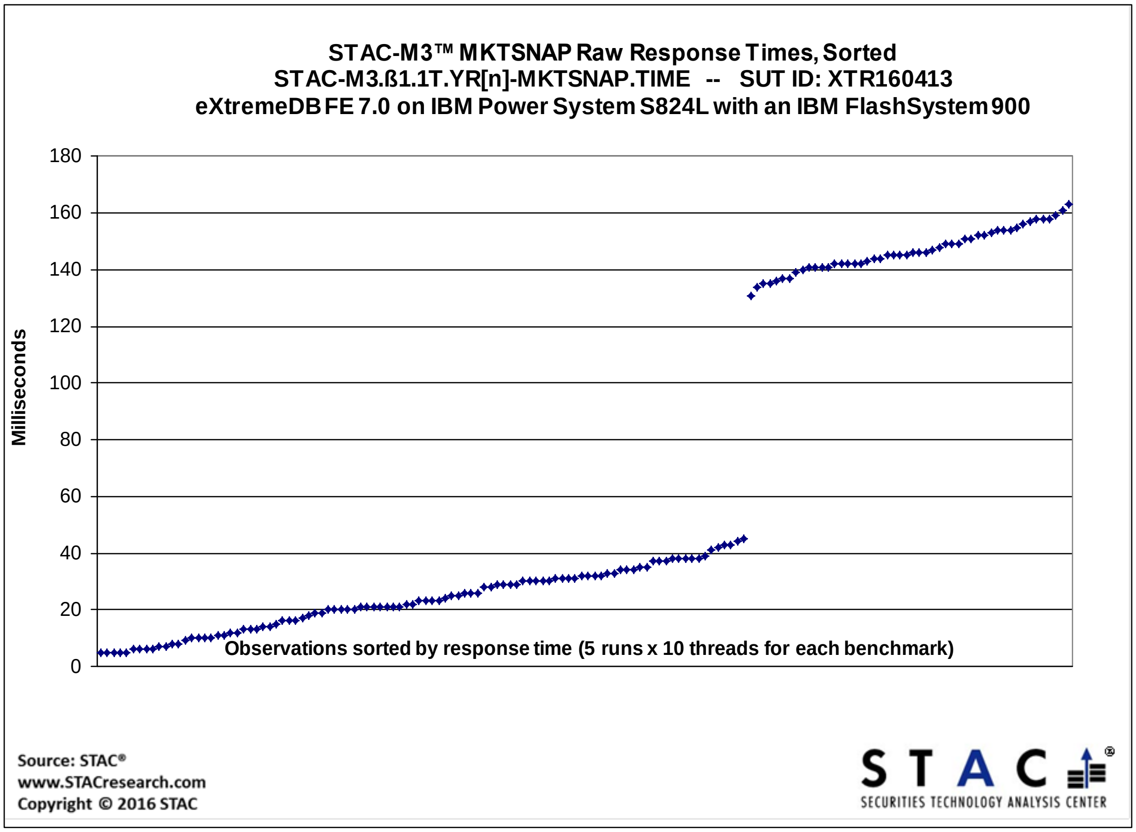


Figure 3

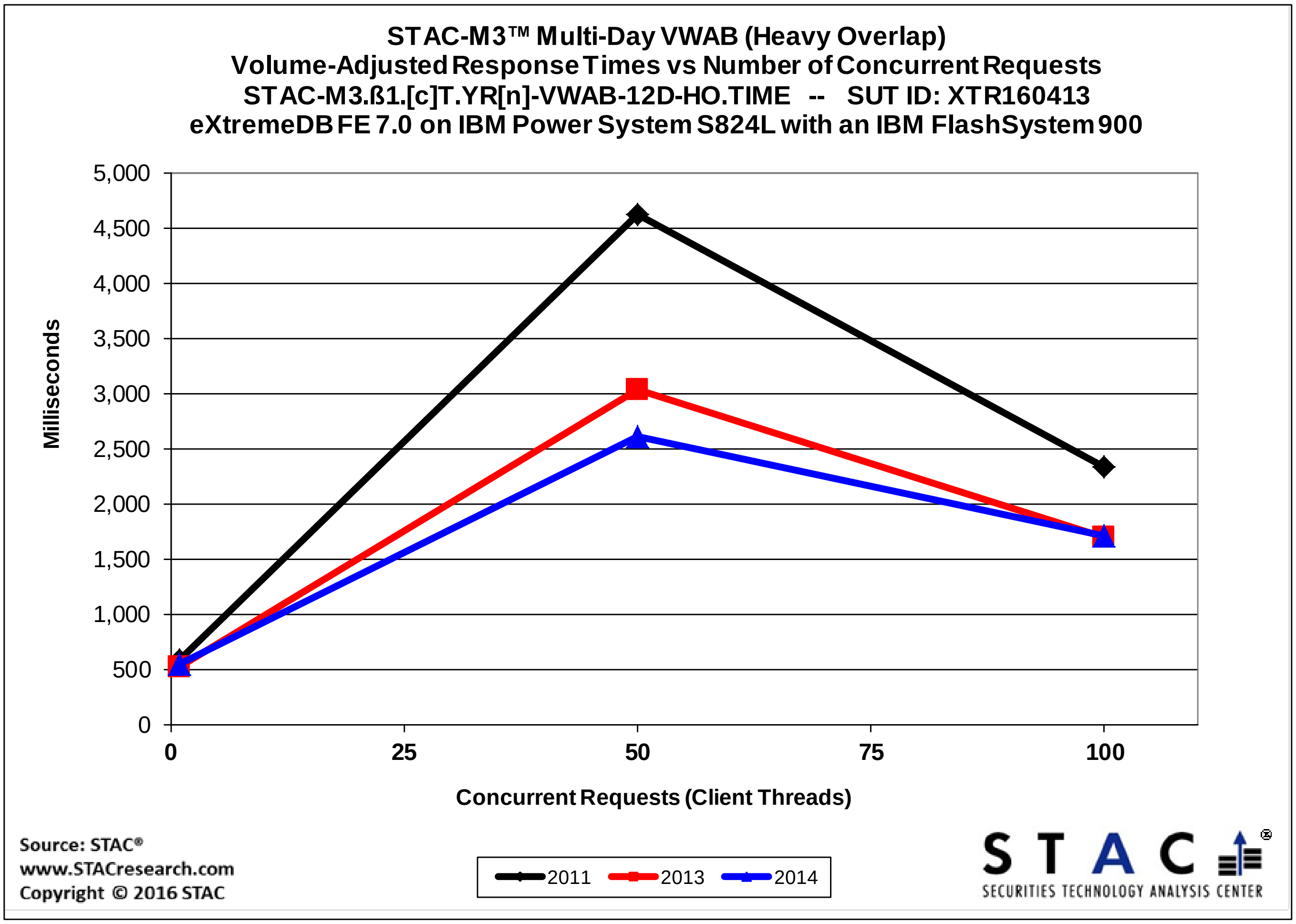

STAC-M3™ Multi-Day VWAB (Heavy Overlap)
Volume-Adjusted Response Times vs Number of Concurrent Requests
STAC-M3.ß1.[c]T.YR[n]-VWAB-12D-HO.TIME -- SUT ID: XTR160413
eXtremeDB FE 7.0 on IBM Power System S824L with an IBM FlashSystem 900
Milliseconds
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
0
25
50
75
100
Concurrent Requests (Client Threads)
Source: STAC®
www.STACresearch.com
Copyright © 2016 STAC
2011
2013
2014
STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

Figure 4

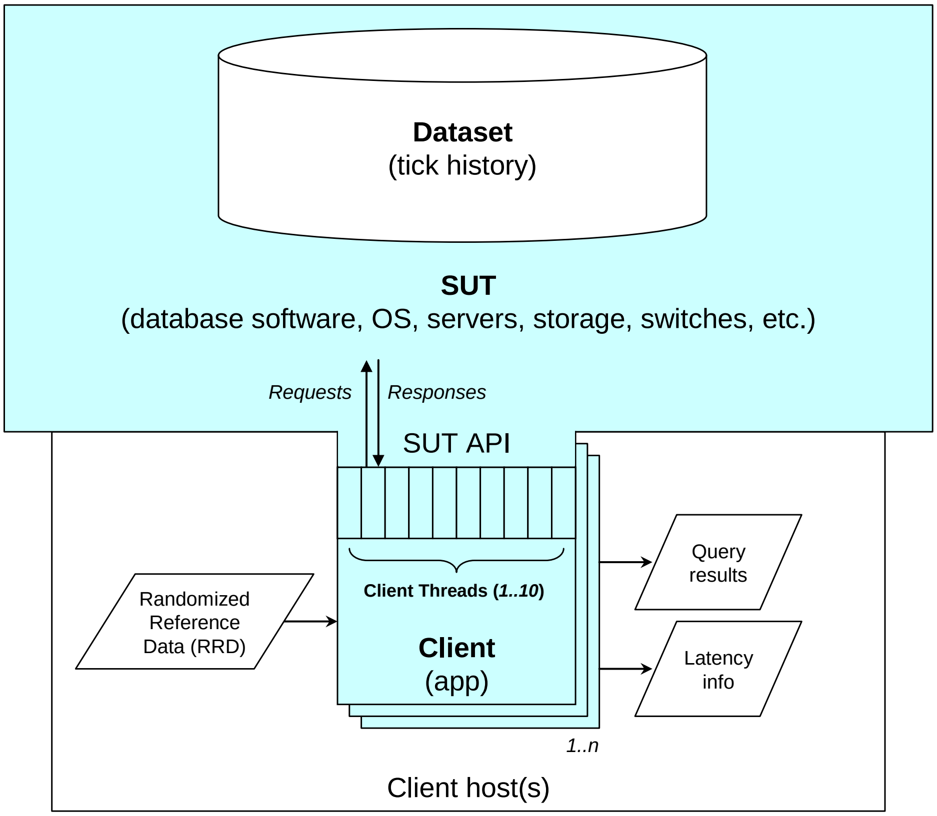
1. Overview of the STAC-M3 (Kanaga) Benchmark Suite

The STAC-M3 Benchmark specifications enable comparative testing of technology stacks suitable for managing financial time-series data such as tick data. The STAC Benchmark Council has developed these 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 Benchmarks are grouped into suites. The base suite, code-named "Antuco," contains a range of test cases with varying levels of CPU and storage-I/O intensity. See http://www.stacresearch.com/STAC-M3_Antuco_Overview.pdf for an overview of the Antuco suite of STAC-M3 Benchmark specifications.

The optional "Kanaga" suite of STAC-M3 consists of two test sequences that extend Antuco benchmarks across larger quantities of data in order to measure the volume-scalability of a database stack. The ability of a tick database solution to handle increasing volumes of historical data is important to today's trading organizations. Market data volumes continue to grow quickly, sometimes in step-function increments. Engineering a solution that delivers consistent, high performance across ever larger datasets can be a challenge. STAC-M3 Kanaga provides some insight into how well a given tick database solution scales.

The test setup for the STAC-M3 Kanaga suite is identical to that for the STAC-M3 Antuco suite. It 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.



The STAC-M3 Kanaga dataset is an extension of the STAC-M3 Antuco dataset, which is a synthetic store of quotes and trades modeled on NSYE TAQ data (US equities). In Antuco, the data volume per symbol was determined by doubling the daily NYSE TAQ volume observed over a few busy days in 1Q10. The resulting dataset was assigned to dates in 2011, constituting a hypothetical year's worth of data. Since it represents only one year of one dataset (level 1 equities), the Antuco database is considerably smaller than the full datasets in use at customer sites. This was a deliberate choice by the STAC-M3 Working Group to minimize the cost of running the baseline benchmarks while still yielding valuable results.

The Kanaga suite is designed to address the question of dataset size. It calls for additional (future) years to be added to the Antuco database, using the same data structures, where each year's size is a multiple of the previous year's. This allows re-use of Operations from Antuco by adjusting them to operate over different time intervals.

Using a fairly aggressive rule of thumb that assumes total market data volumes double roughly every 18 months, Kanaga sets the annual scaling factor at 1.6. A firm that wishes to publish STAC-M3 Kanaga results is required to scale the dataset at least through 2013, but the specifications are defined as far forward as 2020. In addition to scaling forward in time, the specs also call for a 2003 dataset (1.6^8 times smaller than the 2011 dataset) in order to test how the technology stack handles small datasets.

Figure 4 shows the dataset sizes that result when applying these rules to the tick database tested in this report. The left y-axis shows the size in terabytes, while the right y-axis shows the daily total of quotes and trades in the database. This daily total is the same for all 252 days in the database for the given year.

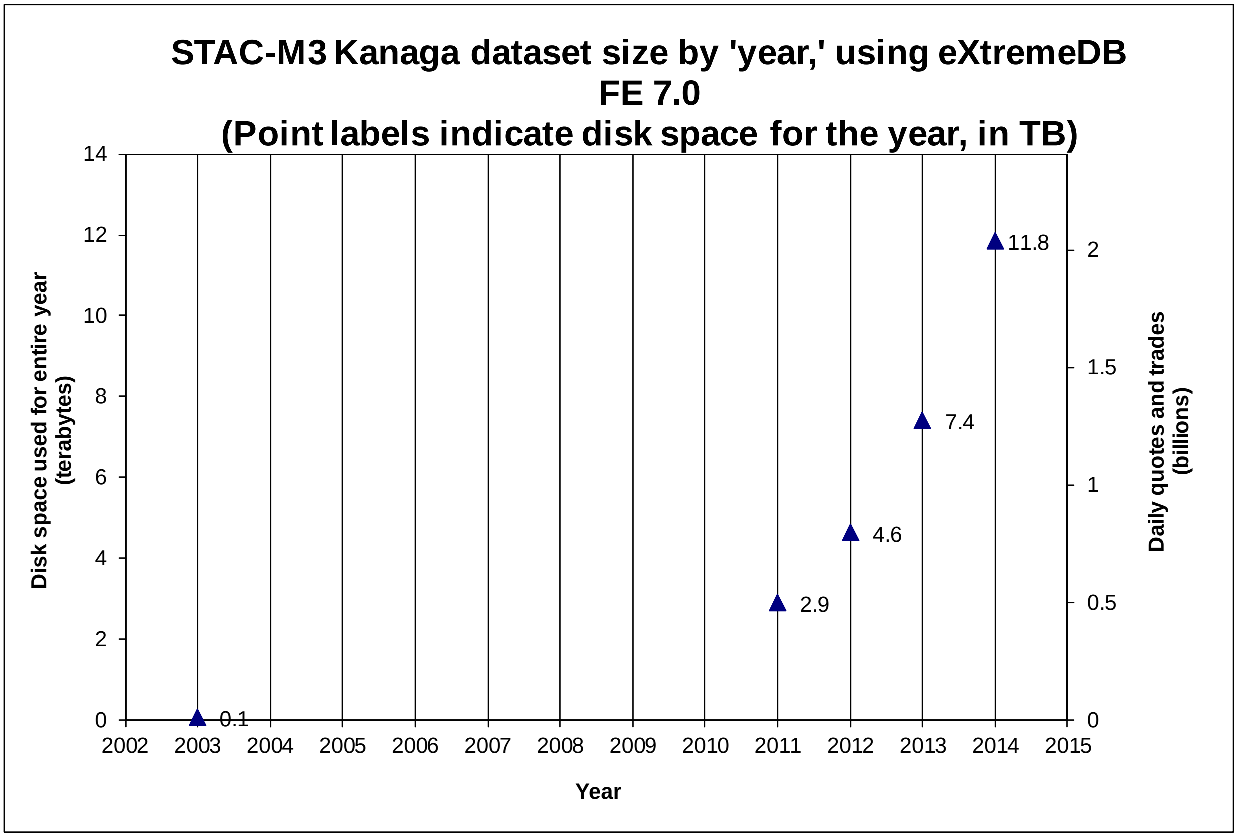


Figure 4

These sizes are not meant to predict future NYSE TAQ volumes. Using years as the basis for increasing the volume of quotes and trades in the benchmarks is simply a convenience. Scaling up a single level-1 dataset is not necessarily the most realistic way to emulate the large tick data stores in trading institutions today. Real deployments typically also involve level 1 and level 2 data from many asset classes. However, the STAC-M3 Working Group determined that scaling the existing dataset would provide significant insight without the expense and complexity of specifying additional trade and quote record formats and designing entirely new queries to operate on them.

The STAC-M3 Kanaga suite has just three Operations:

- [n]YRHIBID is the same as the YRHIBID operation in the Antuco suite except that it operates over multiple years in the Kanaga dataset rather than just the single year of Antuco (2011). [n]YRHIBID returns the highest bid price for each of a certain 1% of symbols over a particular range of years in the dataset. The range for 2YRHIBID is from the first day of 2011 through the last day of 2012. The range for 3YRHIBID is from the first day of 2011 through the last day of 2013, and so on.
- YR[n]-MKTSNAP is the same as the MKTSNAP operation in the Antuco suite except that it operates in each of the years in the Kanaga dataset rather than just within the Antuco year (2011). YR[n]-MKTSNAP returns the price and size information from the latest quote and trade for each of a certain 1% of symbols at a unique time on a unique date in the given year of the dataset. YR2-MKTSNAP queries dates and times in 2012, while YR3-MKTSNAP queries dates and times in 2013, and so on.
- [c]T.YR[n]VWAB-12D-HO is a VWAB operation similar to the VWAB-12D-NO operation in the Antuco suite (4-hour volume-weighted average bid for 12 randomly-selected days) except for three things: 1) it varies the number of concurrent requests (client threads, the "c" in the benchmark ID); 2) it operates in multiple years of the Kanaga dataset rather than just within the Antuco year (2011); and 3) the dates and symbols are chosen so as to ensure heavy overlap among requests, since this is a common pattern in the real world. The tester chooses three scale points in terms of client threads and three years to test. These scale points must top out at the maximum points to be tested for the SUT.

Note that some of the tables and charts in this report also display results of the corresponding tests from the Antuco suite (STAC-M3.ß1.1T.YRHIBID, STAC-M3.ß1.10T.MKTSNAP). These test cases were run on this exact SUT and are included as the baseline for scale comparisons.

The fundamental metric in STAC-M3 Kanaga is "last result latency" (LAT2) of an operation, as defined in Antuco. This is the time that it takes for the operation to return all of its results. Latency measurements are performed in the clients. A client thread gets a local timestamp (t_{submit}) just before submitting a query. When it receives the complete results (sorted appropriately), the client gets another timestamp (t_{last}). Last-result latency (LAT2) = $t_{last} – t_{submit}$.
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 mean time to satisfy requests from all 10 client threads.

[n]YRHIBID is a good test of sequential read performance, while YR[n]-MKTSNAP is a good test of random read performance. As with YRHIBID, [n]YRHIBID has an additional metric: the bytes read per second from persistent storage (i.e., excluding server cache), which is computed from the output of appropriate system utilities.

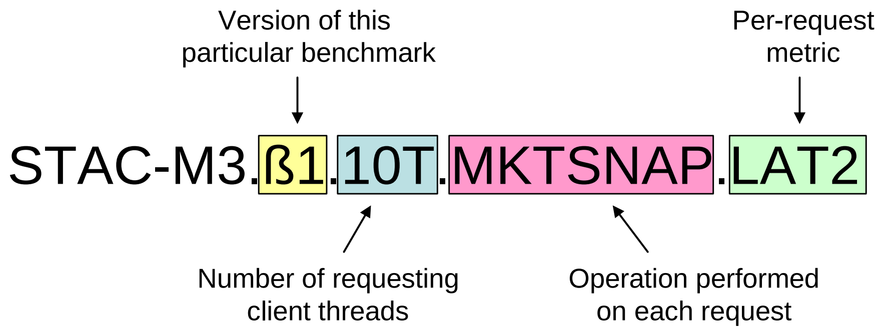
[c]T.YR[n]VWAB-12D-HO is a good test of user scaling and how that varies with the size of the dataset.

Kanaga also includes a volume-adjusted latency metric (VLAT2), which divides the latency (LAT2) by a weighted-volume factor (WVF). The (WFV) is a relative measure of the number of quotes and/or trades in the dataset subject to the given query compared to the number of quotes and trades subject to the same query in the base year (2011). For example, since 2YRHIBID covers 2.6 times the quotes and trades of YRHIBID (2012 is 1.6 times the size of 2011, and 2YRHIBID covers both 2011 and 2012). Volume-adjusted latency normalizes latencies in order to see how the response time per quote or trade changes with the size of the dataset.

The tests require a client application that is written to a product API and is capable of submitting requests from 10 independent threads. The [n]YRHIBID operation is called by one client instance making requests from a single thread, while YR[n]-MKTSNAP is called by one client using 10 threads to submit queries simultaneously. The [c]T.YR[n]VWAB-12D-HO operation is called by a varying number of client instances, depending on the scale points to be tested.

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.

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.

2. Product background

Key products in this SUT included:

- McObject eXtremeDB Financial Edition 7.0
- Red Hat Enterprise Linux 7.2
- 2 x IBM Power System S824L with POWER8 CPU @ 3.52 GHz
- IBM FlashSystem 900 with 12 x 5.7TB flash modules

McObject submitted the following information and claims about its products:

eXtremeDB Financial Edition achieves record-breaking performance due to its highly efficient architecture, hybrid row/columnar storage, database sharding, powerful indexing capabilities, extensive vector-based math functions, and calculation pipelining for on-chip analytics. As a true ACID compliant DBMS for Big Data, eXtremeDB Financial Edition scales predictably and is both robust and highly reliable, which is critical for a wide variety of capital market applications where high performance, low latency, and minimal jitter are important.

eXtremeDB has been used for more than 15 years in high performance applications in capital markets, defense, avionics, netcom, and other markets. There are hundreds of companies using the development software and more than 30 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 alternatives today either lack the performance required by low latency applications, are too difficult to scale, are too hard to use and/or are too expensive.

eXtremeDB Financial Edition leverages proven eXtremeDB embedded database strengths – including a streamlined hybrid in-memory/persistent storage database system 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 cache use. Functions can be pipelined to form an assembly line of operations on sequences in support of statistical/quantitative analysis*
- *Use of common languages including SQL, C/C++, JAVA, C#, Python, ODBC/JDBC*
- *Sharding and distributed query processing, which support high scalability*

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 or Client/Server. eXtremeDB can run entirely within the application process, eliminating inter-process communication (IPC) between client and server modules. To exploit sharding, multiple CPUs/cores and distributed/parallel query execution, eXtremeDB can run in client/server mode.

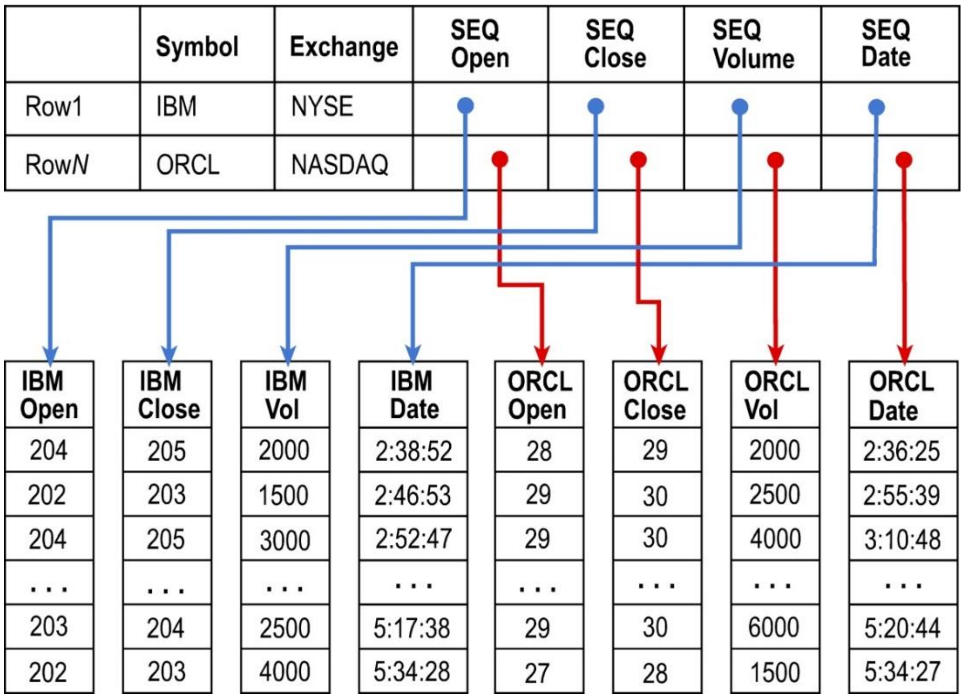
Short execution path. eXtremeDB is written in C/C++, and its code path is very short: a "footprint" of approximately 150K for the core database system points to McObject's unrelenting focus on eliminating even small potential sources of latency.

Programming interfaces. 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.

Memory management. eXtremeDB manages memory with its own custom allocators instead of relying on the programming language run-time's default, generic memory manager. This memory management, tailored for specific data layouts and usage patterns, enhances speed as well as scalability for multi-core configurations.

CPU cache efficiency. Keeping code and data in CPU cache eliminates costly (in performance terms) fetches. eXtremeDB Financial Edition's column-based data layout for sequences maximizes 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 remains in the cache.

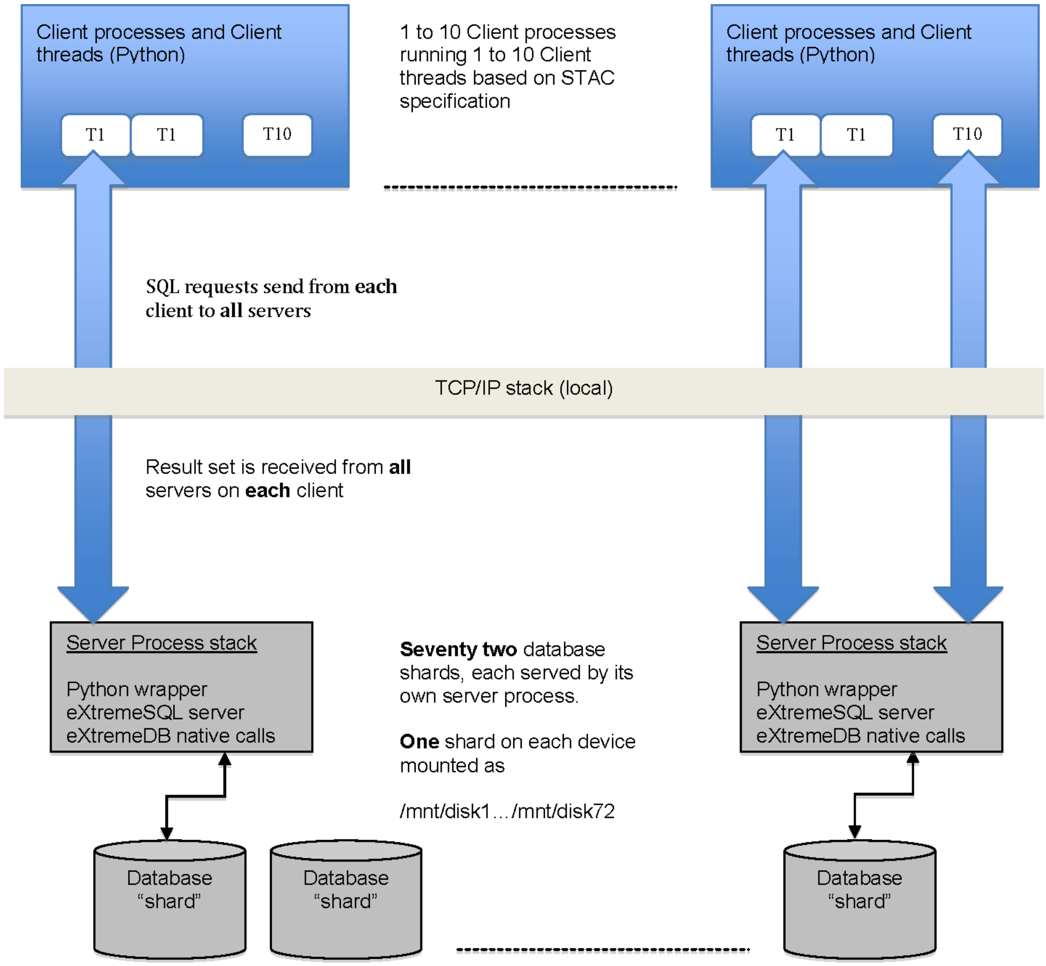
Pipelining vector-based statistical functions. eXtremeDB Financial Edition includes a library of vector-based statistical functions that operate on sequences of market data. These functions are typically pipelined so that the output of one serves as input for the next. Pipelining causes market data to remain in CPU cache during an entire operation (i.e. while it is "worked on" by multiple functions), eliminating latency-inducing transfer of interim results to temporary tables in main memory.



<u>Traditional DBMSs bring rows of data into 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 cache speed and avoids excess fetching of data from RAM.

STAC-M3 Implementation

As shown in the diagram below, the core of the STAC-M3 implementation using eXtremeDB (the STAC-M3 Pack for eXtremeDB) is a set of Python language scripts that utilize eXtremeDB SQL to access a distributed eXtremeDB database, distributed over multiple physical storage locations. "eXtremeDB native calls" are user-defined functions written in C++ and Python. User-defined functions written in C++ and Python are processed by SQL for four STAC-M3 test cases. The rest simply made standard eXtremeDB SQL calls.



From the standpoint of database access, the STAC benchmark is logically divided into client and server parts. The client part creates 1 to 10 Python processes and each process creates 1 to 10 Python threads (based on the benchmark requirements). Python threads, in turn, make calls to the eXtremeDB SQL engine client ("SQL client") that connects to the eXtremeDB SQL server for a database partition ("Database shard") through the TCP/IP-based network channel integrated with the SQL engine.

The configuration used in this STAC-M3 implementation creates 72 shards, with each shard occupying its own logical partition. Once the client-server connection is established, the Python applications send SQL requests to servers and receive results sets. Each client sends SQL requests to all SQL servers, and consolidates the results received from them.

This was the first eXtremeDB STAC-M3 implementation for the STAC-M3 Kanaga suite. It used direct I/O and a high degree of parallelization.

For more information on eXtremeDB Financial Edition, please visit http://financial.mcobject.com

IBM submitted the following information and claims about its products:

The STAC-M3 test with McObject eXtremeDB demonstrates that the IBM Power S824L server and FlashSystem 900 provide an exceptional low-latency and low jitter implementation of tick data warehousing and analysis. The extreme cache/memory hierarchy provided by the IBM Power System S824L and the performance of FlashSystem storage rival even the most optimized systems based on other CPU architectures *with extreme storage clustered arrays. The jitter advantages of this configuration and the McObject eXtremeDB demonstrate an enterprise ready tick data warehouse never before available.*

The IBM Power System S824L server is built on advanced microprocessor technology. The POWER8 microprocessor is the 4th generation of Power technology announced in 2001 as the POWER4, the first multi-core server technology in the industry. IBM Research has contributed chip-manufacturing technology that delivers the high reliability, high quality, high-density memory, and high-bandwidth interconnects. IBM's patented embedded DRAM (eDRAM) enables the POWER8 processor to increase cache density to 96K L1 (64K data and 32K instruction), 512K private L2, and 96MB of non-uniform cache access (NUCA) shared low-latency L3, all on the processor. That is double the per core on-chip capacity of competitive processors. In addition, the IBM POWER8 servers can deliver up to 128MB of L4 cache per socket, which is shared among the processor memory interfaces. This feature alone dramatically improves the performance of in-memory columnar databases like eXtremeDB. The memory interfaces to the on-DIMM memory controllers provide 8 channels of memory per socket delivering 192GB/s sustained memory bandwidth. POWER8 also provides Simultaneous Multi-Threading (SMT) with up to 8 threads per core. Additional scale-up POWER8 servers increase that number to 192 cores in a single OS image of a nearly flat SMP model. The IBM Power S824L is the ideal system for distributed scale-out tick data workloads.

The IBM FlashSystem 900 array can deliver high performance and low latency while providing several terabytes of usable multi-dimensional Flash RAID protected data storage in just a few rack units of space. FlashSystem storage systems have enterprise-level availability and reliability with no single point of failure, multiple layers of data correction, chip redundancy, and redundant hot swap components.

Extreme performance, IBM® MicroLatency™, macro efficiency, and enterprise-grade reliability make IBM FlashSystem a powerful and cost effective tool for accelerating OLTP systems, meeting and exceeding online customer expectations, and gaining competitive advantage in eCommerce environments. FlashSystem performance also powers online analytics processing (OLAP) tools that help businesses better understand and more swiftly respond to customer needs and preferences. Just as importantly, the extraordinary capabilities and capacity of FlashSystem arrays enable eCommerce enterprises to address multiple compute challenges in current 24/7/365 operational environments while at the same time empowering growth and innovation into the future.

For more information please visit https://www.ibm.com/products

3. Project participants and responsibilities

The following firms participated in the project:

- McObject
- IBM
- STAC

The Project Participants had the following responsibilities:

- McObject implemented the STAC-M3 Clients and Operations using the STAC-M3 Benchmark specifications.
- IBM supplied the lab and support; server and storage hardware.
- McObject and IBM configured and optimized the full stack under test.
- McObject sponsored the Audit.
- STAC conducted the STAC-M3 Benchmark Audit, which included validating the database; inspecting any source-code revisions to the STAC Pack; validating the Operation results; executing the tests, and documenting the results.

Note that STAC did not physically inspect the system. Configuration details that could not be verified through independent interrogation of the system were accepted from declarations by one or more of the parties above.

4. Contacts

- McObject: Chris Mureen, COO, chris.mureen@mcobject.com , +1 425 888 8505 x211
- IBM: Brent Sesselmann, Power System Sales, bmsessel@us.ibm.com +1 231-557-1487
- IBM: Cedric Cook, Solutions Offering Manager, cdcook@us.ibm.com, +1 713-278-6226
- STAC: info@STACresearch.com

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 Kanaga suite of STAC-M3 Benchmark specifications. Full members of the STAC Benchmark Council can access these specifications at www.STACresearch.com/m3. Premium subscribers can download the programs used in these benchmarks in order to run the same tests on systems in the privacy of their own labs. Results for this same system using the baseline tests of the STAC-M3 Antuco suite are also available at www.STACresearch.com/XTR160421.

7. Limitations

- MKTSNAP is a random-access test, and because systems tend to have areas of storage that differ in access time, the MKTSNAP latencies can vary considerably. STAC-M3 requires a limited number of test runs, which means that the standard deviation of latency can be quite large relative to the mean. The mean YR[n]-MKTSNAP.TIME is therefore, in general, not the best statistic to use from these tests (i.e., two systems with the same performance could get quite different mean latencies simply by chance). Median and max are probably more instructive indicators, which is why these are used in Figure 2.

8. Stack under test

A detailed STAC Configuration Disclosure for the SUT in this report is available to premium members of the STAC Benchmark Council at www.STACresearch.com/XTR160413. That document provides the exact product version numbers, detailed tuning options, and other important information. Additional configuration details such as a SOS report may also be available, depending on the SUT platform. If you are unable to access these materials and would like to learn how to, please contact us at www.STACresearch.com/contact

The STAC Pack used in this SUT (i.e., the STAC-M3 Clients and supporting scripts) was the STAC-M3 Pack for eXtremeDB Rev C Pkg20160413.

9. Vendor Commentary

None

10. STAC Notes

None

About STAC

STAC® provides technology research and testing tools that are based upon community-source standards. STAC facilitates the STAC Benchmark™ Council (www.STACresearch.com/council), an organization of leading financial institutions and technology vendors that specifies standard ways to assess technologies used in finance. The Council is active in an expanding range of low-latency, big-compute, and big-data workloads.

STAC helps end-user firms relate the performance of new technologies to that of their existing systems by supplying them with STAC Benchmark reports as well as standards-based STAC Test Harnesses™ for rapid execution of STAC Benchmarks in their own labs. User firms do not disclose their results. Some STAC Benchmark results from vendor-driven projects are made available to the public, while those in the STAC Vault™ are reserved for qualified members of the Council (see www.STACresearch.com/vault).

To be notified when new STAC Reports™ become available, please sign up for free at www.STACresearch.com.

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

