



STAC Update:

Big data

Michel Debiche
Director of Analytics Research, STAC

michel.debiche@STACresearch.com

Context: STAC big data analytics benchmarks

Benchmark	Suite	Domain	Computational Complexity	I/O	Input Data Size	Separable	Concurrency
STAC-M3	Baseline (Antuco)	Tick analytics	Low to Moderate	Variable	Small (~4 TB)	Varies	Varies
STAC-M3	Scale (Kanaga)		Low to Moderate	Variable	Large (>50 TB), Variable	Varies	Varies
STAC-M3	Small in-memory (Shasta)		Low to Moderate	Very low	Small (~4 TB)	Varies	Varies
STAC-A3	Sweep	Backtesting	Low	Heavy	Medium (~8 TB)	Yes	Single user
STAC-A3	BLASH		Low	Heavy	Large, Variable	Partially	Single user
STAC-A3	Options		High	Heavy	Medium (~8 TB)	Yes	Single user

STAC-M3

- Performance benchmarks for enterprise tick analytics
 - Language/DBMS neutral
 - Developed by banks and hedge funds
- Workload:
 - Synthetic data modeled on NYSE TAQ
 - Mix of I/O- and compute-intensive operations (read-heavy)
 - Simulates concurrent access with varying number of users
- 3 Suites
 - Baseline (Antuco): Limited dataset forced to access storage
 - Scale (Kanaga): Large dataset
 - Small in-memory (Shasta): Limited dataset not forced to access storage

www.STACresearch.com/m3

STAC-M3 (continued)

- Wide range of implementations
 - Databases: kdb+, eXtremeDB
 - Clustered file systems, parallel file systems, NFS, flash arrays, NVME over Fabric, direct-attached SSD, NAND and post-NAND Flash (e.g. Optane)
 - Single database server, database cluster (bare metal and cloud)

www.STACresearch.com/m3

STAC-M3 / kdb+ / Pavilion Data Hyperparallel Flash Array

- SUT ID: KDB190927
- Stack:
 - kdb+ 3.6
 - Pavilion Hyperparallel Flash Array 2.2.1
 - 6 2S Supermicro DB servers: 4 with Intel Xeon ES-2690 v4 CPUs, 2 with Xeon ES-2678 v3 CPUs, 208-320GB of DRAM each
 - Centos 7.6 with ext4
 - Dell Z9100-ON 100G Ethernet switch
- STAC-M3 Antuco



www.STACresearch.com/KDB190927

STAC-M3 Antuco results (baseline suite)

- The HFA-based solution outperformed all publicly disclosed results in 4 of 17 mean-response time benchmarks:
 - STAC-M3.β1.100T.STATS-UI.TIME
 - STAC-M3.β1.50T.STATS-UI.TIME
 - STAC-M3.β1.10T.STATS-UI.TIME
 - STAC-M3.β1.10T.STATS-AGG.TIME



www.STACresearch.com/KDB190927

STAC-M3 Antuco results (baseline suite)

- The HFA-based solution was faster in 8 of 17 mean-response time benchmarks than all other solutions using kdb+ and flash arrays:
 - STAC-M3.β1.100T.STATS-UI.TIME
 - STAC-M3.β1.50T.STATS-UI.TIME
 - STAC-M3.β1.10T.STATS-UI.TIME
 - STAC-M3.β1.10T.STATS-AGG.TIME
 - STAC-M3.β1.100T.VWAB-12D-NO.TIME
 - STAC-M3.β1.10T.VOLCURV.TIME
 - STAC-M3.β1.1T.MOHIBID.TIME
 - STAC-M3.β1.1T.STATS-UI.TIME



www.STACresearch.com/KDB190927

STAC-M3 Antuco results (baseline suite)

- The HFA-based solution was faster in 10 of 17 mean-response time benchmarks than a previous record-setting kdb+ solution involving a competitive flash array (SUT ID KDB170421), including:
 - 2x the speed for STAC-M3.β1.10T.STATS-AGG.TIME
 - Over 1.7x the speed for both STAC-M3.β1.10T.MKTSNAP.TIME and STAC-M3.β1.10T.STATS-UI.TIME
 - Over 1.5x the speed for both STAC-M3.β1.100T.STATS-UI.TIME and STAC-M3.β1.50T.STATS-UI.TIME



www.STACresearch.com/KDB190927

Vault Report: STAC-M3 / eXtremeDB / Pavilion HFA

- SUT ID: KDB190914
- Stack:
 - eXtremeDB for HPC 8.0
 - Centos 7.6 with ext4
 - 6 2S Supermicro DB servers: 4 with Intel Xeon ES-2690 v4 CPUs, 2 with Xeon ES-2678 v3 CPUs, 208-320GB of DRAM each
 - Pavilion Hyperparallel Flash Array 2.2.1
 - Dell Z9100-ON 100G Ethernet switch
- STAC-M3 Antuco
- Unaudited results will be in the STAC Vault



Vault report coming soon

Back to the context

Benchmark	Suite	Domain	Computational Complexity	I/O	Input Data Size	Separable	Concurrency
STAC-M3	Baseline (Antuco)	Tick analytics	Low to Moderate	Variable	Small (~4 TB)	Varies	Varies
STAC-M3	Scale (Kanaga)		Low to Moderate	Variable	Large (>50 TB), Variable	Varies	Varies
STAC-M3	Small in-memory (Snasta)		Low to Moderate	Very low	Small (~4 TB)	Varies	Varies
STAC-A3	Sweep	Backtesting	Low	Heavy	Medium (~8 TB)	Yes	Single user
STAC-A3	BLASH		Low	Heavy	Large, Variable	Partially	Single user
STAC-A3	Options		High	Heavy	Medium (~8 TB)	Yes	Single user

- Recap:
 - Workloads that emulate real-world backtesting jobs
 - Measure speed, scalability, efficiency of any architecture
- Test harness hands the implementation jobs to execute
- Measures the throughput and efficiency of the SUT

Now is a great time to get involved

- Several proposals are about to come out
- We will discuss in online forums, then occasional telecons
- Not a huge time commitment
- To join the STAC-A3 Working Group:
 - Go to www.STACresearch.com/a3, right side of the page
 - If you see the “Group Members” list, you’re already in it
 - If not, click on the “Enable Me” button

www.STACresearch.com/a3

Get access to this domain

If you'd like to obtain privileged materials from this domain, or if you would like to participate in this group, please click the button below.

Enable me! >