STAC Update:
Big data

Michel Debiche
Director of Analytics Research, STAC

michel.debiche@STACresearch.com

STAC-M3

- Performance benchmarks for enterprise tick analytics
  - Language/DBMS neutral
  - Developed by banks and hedge funds

  www.STACresearch.com/m3

- 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

STAC ®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 Suites

Suite	Computational Complexity	I/O	Input Data Size	Pre-load data?	Concurrency
Baseline (Antuco)	Low to Moderate	Variable	Small (~4 TB)	No	Varies
Small in-memory (Shasta)	Low to Moderate	Very low	Small (~4 TB)	Yes	Varies
Scale (Kanaga)	Low to Moderate	Variable	Large (>50 TB), Variable	Yes	Varies

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 Suites

Suite	Computational Complexity	I/O	Input Data Size	Pre-load data?	Concurrency
Baseline (Antuco)	Low to Moderate	Variable	Small (~4 TB)	No	Varies
Small in-memory (Shasta**)**	Low to Moderate	Very low	Small (~4 TB)	Yes	Varies
Scale (Kanaga)	Low to Moderate	Variable	Large (>50 TB), Variable	Yes	Varies

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 Database Scaling for Kanaga Suite

STAC-M3 Kanaga dataset size by 'year,' using kdb+
(Triangles = annual, Squares = cumulative)
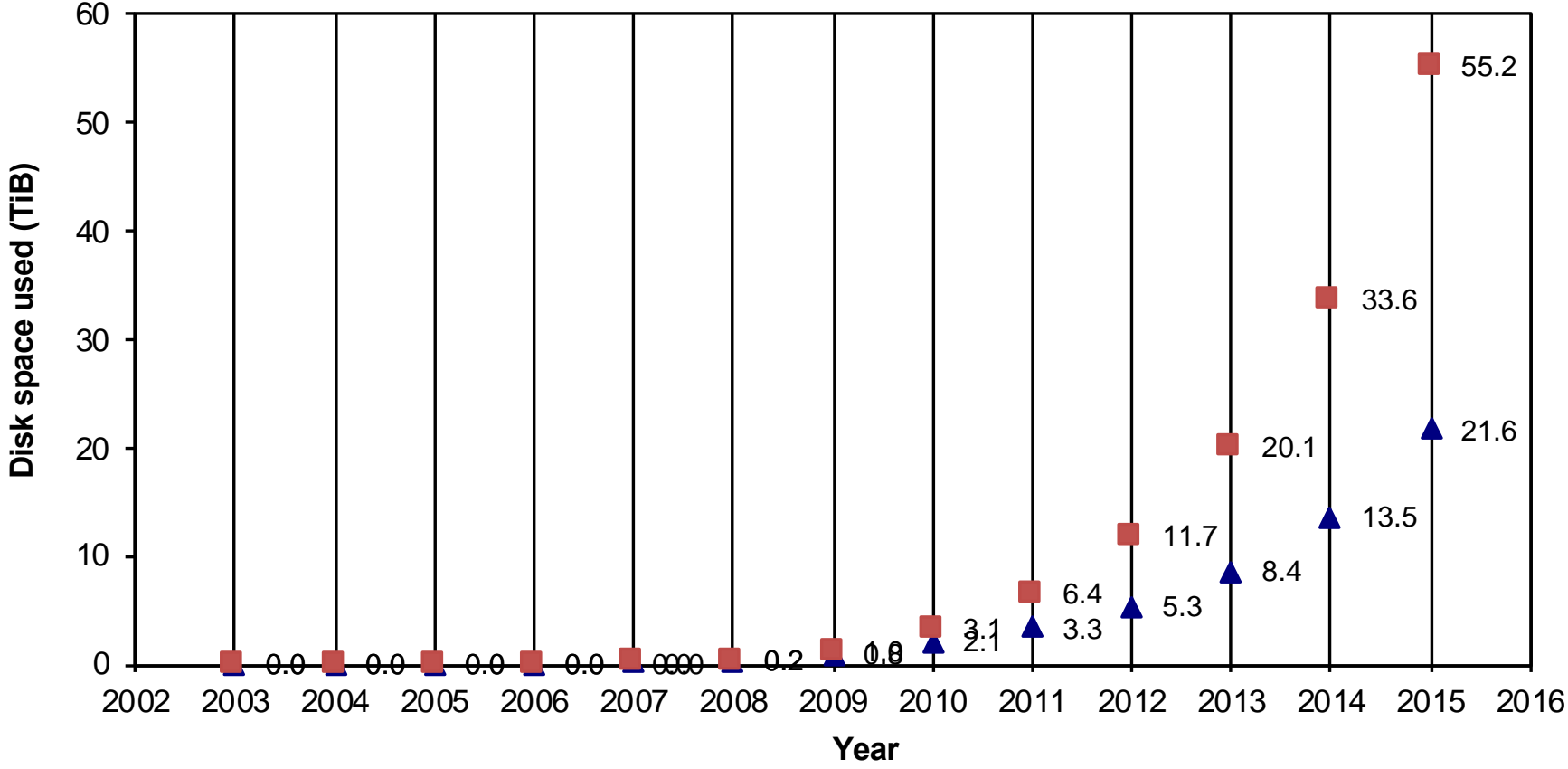
Note: Values indicate size of the dataset itself and may not completely reflect database or filesystem overhead, particularly for years prior to 2011.

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®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 / kdb+ / WekaIO WekaFS / HPE Proliant clusters
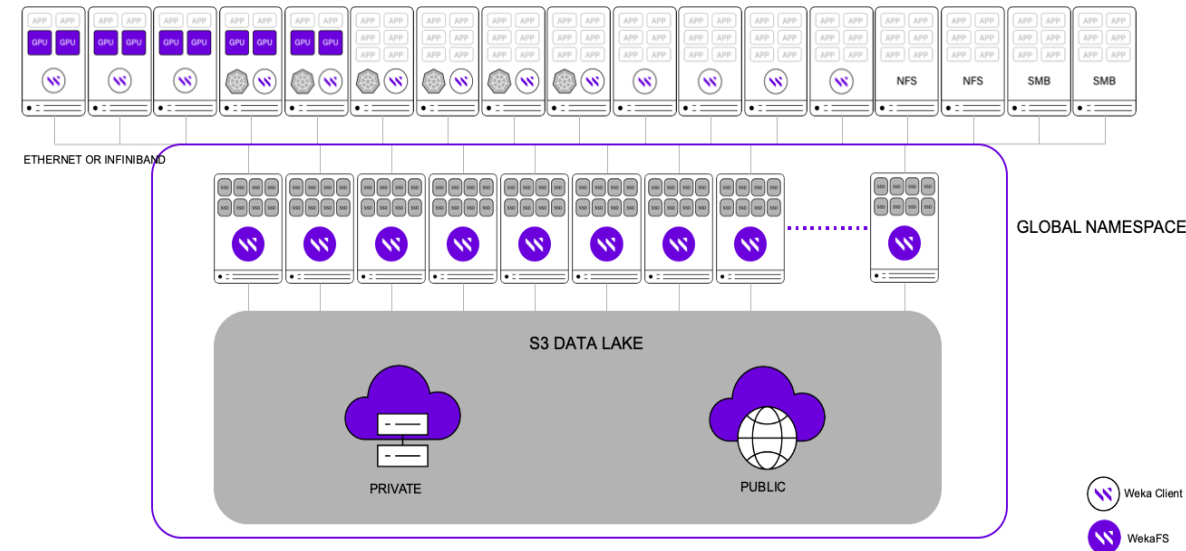
- ## SUT ID: KDB200401

- ## Stack:

  - ### kdb+ 3.6

  - ### DB server cluster: 14 x HPE Proliant XL 170r Gen10, 2 x Intel® Gold 6134 CPUs each

  - ### Storage cluster: WekaIO™ WekaFS™ v3.6.2 Parallel Filesystem (software defined) running on 18 x HPE Proliant XL 170r Gen10 server nodes

- ## STAC-M3 Antuco & Kanaga suites
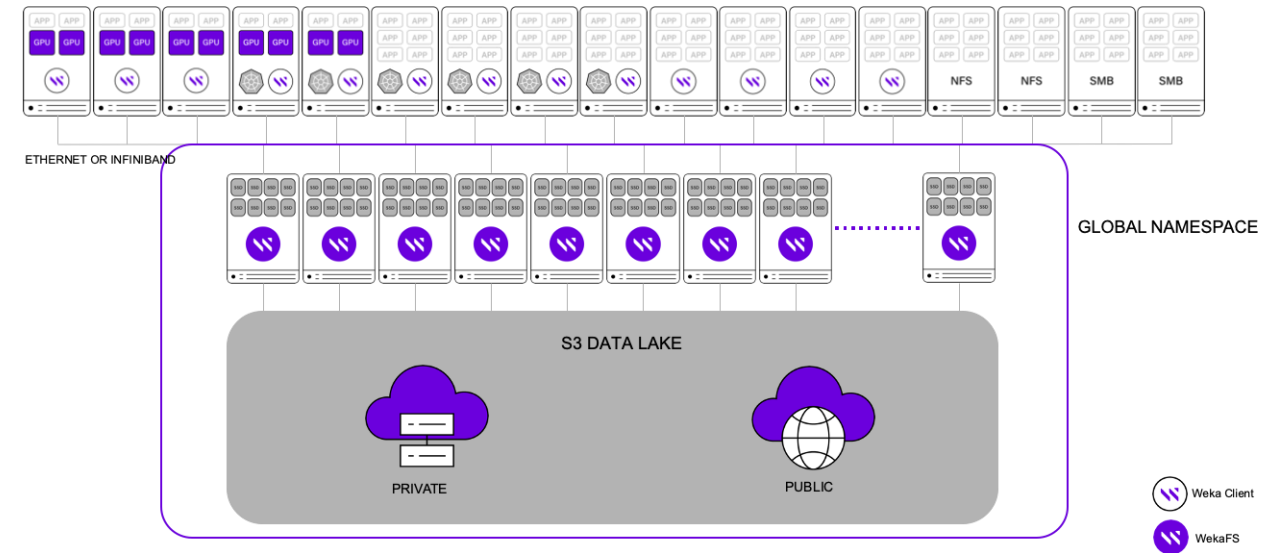
www.STACresearch.com/KDB200401

STAC-M3 Kanaga results (scaling suite)

- The Weka + HPE based solution outperformed all publicly disclosed results in 11 of 24 mean-response time & 5 throughput benchmarks:

  - **100-user 12-day VWAB**: all 5 benchmark years (STAC-M3.β1.100T.YR[n].VWAB-12D-HO.TIME)

  - **50-user 12-day VWAB**: benchmark years 4 and 5 (STAC-M3.β1.50T.YR4VWAB-12D-HO.TIME and STAC-M3.β1.50T.YR5VWAB-12D-HO.TIME).

  - **Multi-year high bid**: all 4 multi-year spans (STAC-M3.β1.1T.[n]YRHIBID.TIME)
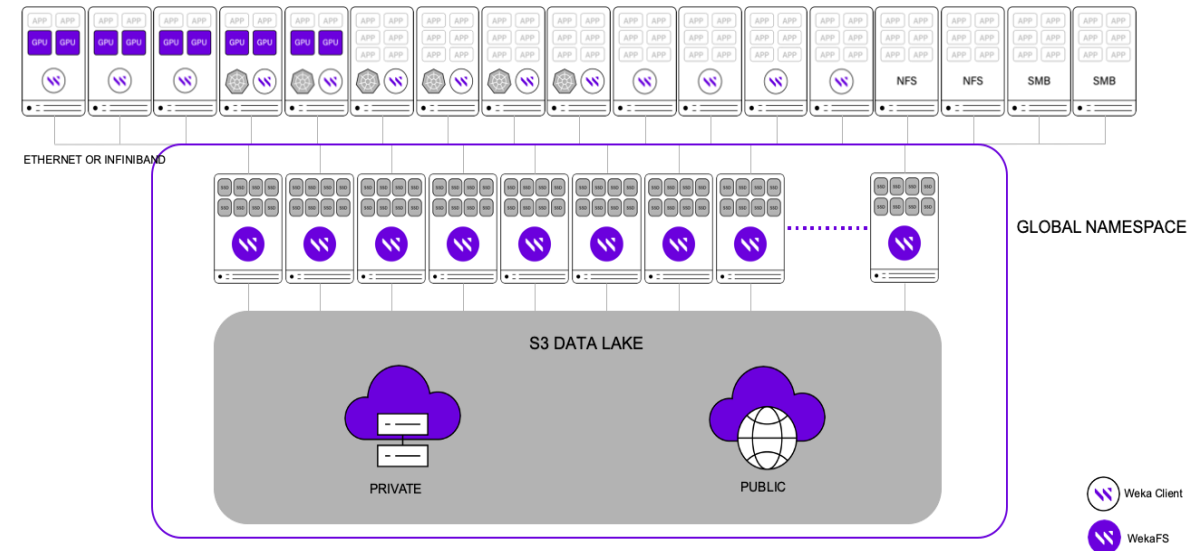
  - **Throughput:** STAC-M3.β1.1T.*.BPS

www.STACresearch.com/KDB200401
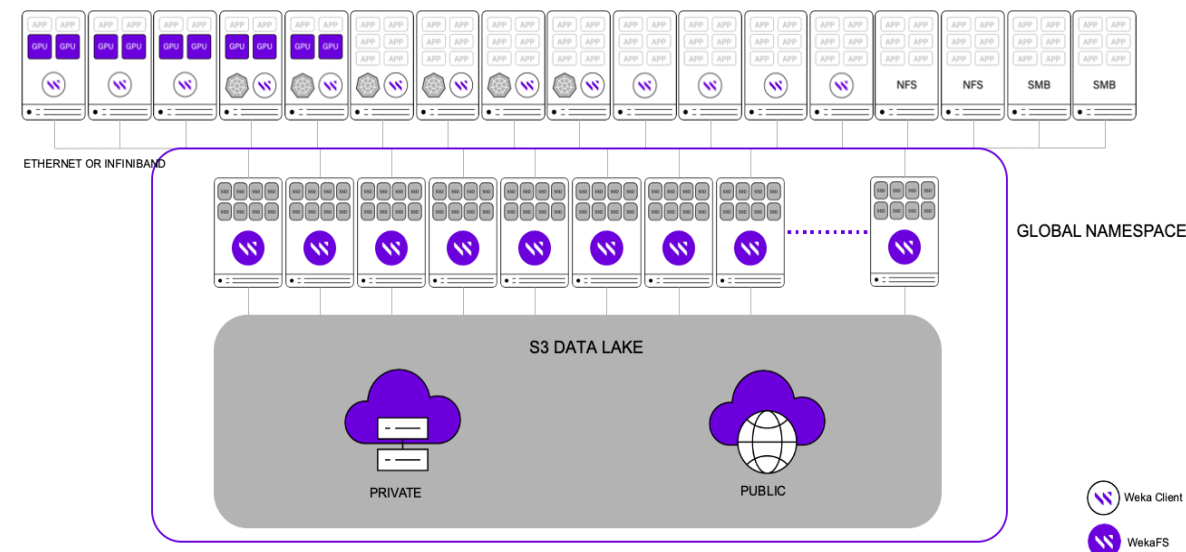
S T A C
SECURITIES TECHNOLOGY ANALYSIS CENTER

- The Weka + HPE based solution outperformed all publicly disclosed results in the 100-user 12-day no-overlap VWAB mean-response time benchmark (*STAC-M3.β1.100T.VWAB-12D.NO.TIME*)
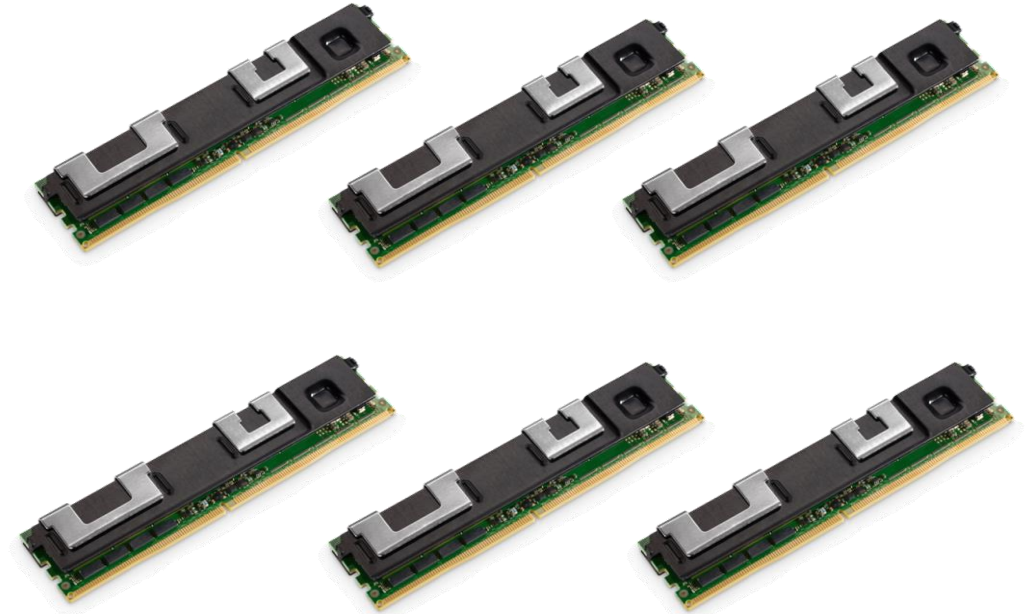

www.STACresearch.com/KDB200401

- Versus a solution involving an all-flash NAS and 4 database nodes (KDB190430)

  - was faster in all 24 Kanaga MRT benchmarks;

  - was faster in 15 of 17 MRT Antuco benchmarks, including:

  - 8.8x speedup in **100-user interval stats** (STAC-M3.β1.100T.STATS-UI.TIME)

www.STACresearch.com/KDB200401

- ## SUT ID: KDB200515

- ## Stack:

  - ### kdb+ 4.0

  - ### RHEL 7.6 with XFS V5

  - ### Intel® Data Management Platform with Red Hat OpenShift v3.11

  - ### 10 x Supermicro SYS-6019U-TN4RT with 2x Intel® Xeon® Platinum 8280L 2.70GHz

  - ### 10 x 6TB Intel® Optane DCPMM

- ## STAC-M3 Antuco & Kanaga suites

Coming soon to www.STACresearch.com/KDB200515

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC Streaming: Review

- Working on benchmarks for streaming ingest and streaming analytics
  - Capacity
  - Latency
  - Impact on historical queries

- Starting with market data ingest

- Will add variant for telemetry data

- Did a market data POC (see Fall '19 STAC Summits)

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC Streaming: Updates

- Working group has agreed on spec requirements for market data

- Tools are nearly ready

- Testing with open source and proprietary time series databases

- Expect initial results this summer

- Will progress telemetry data and streaming analytics in parallel

S T A C ®
SECURITIES TECHNOLOGY ANALYSIS CENTER