WHITE PAPER

TICK DATA ANALYTICS: SCALING CONCURRENCY AND I/O PERFORMANCE WITH DELL EMC ISILON

A Storage Solution for the Financial Services Industry

July 2019

Author: **Boni Bruno, CISSP, CISM, CGEIT**

Chief Solutions Architect, Dell EMC

Abstract

Massive data volumes, low-latency, and complex processing capabilities are common characteristics of high frequency trading and analysis environments. Financial firms have a growing need to collect, store, and analyze more data than ever before. Tick data analysis helps firms keep sight of their investments and quickly react to market change to maximize profits.

Such analysis requires a powerful database and storage solution to handle real-time and historical data sets. This paper describes the Dell EMC Isilon All-Flash Scale-Out NAS Storage solution for Kx Systems' kdb+ database applications commonly used in the financial services sector.

An overview of kdb+ near real-time and historical work flows are presented to understand the suitability of the Dell EMC Isilon Storage solution for tick data analysis. STAC-M3 benchmark results show Isilon performs very well with high concurrency and high I/O workloads for both small and large data sets.

TABLE OF CONTENTS

Introduction 4
Document purpose 4
Solution purpose 4
The Business Challenge 4
Kx Systems (Kdb+) Overview 5
Kdb+Tick Architecture 5
Kdb+ Database 6
Kdb+ Partitioning 7
Dell EMC Isilon F800 All-Flash NAS 9
Isilon nodes 9
Network 10
Back-end (internal) network 10
Front-end (external) network 10
Isilon OneFS File System 10
Kdb+ and Isilon Topology 11
Dell EMC PowerEdge R940 11
Dell EMC Tested Configuration 12
Kdb+ 12
Compute Nodes 12
Isilon F800 (NFS Mounted from each Kdb+ database node) 12
Kdb+ Databases 12
STAC-M3 Benchmark Suite 13
STAC-M3 Performance Test Results 13
High Concurrency Tests 13
Dell EMC Isilon F800 vs Direct-Attached Solution using Optane 3D NAND Flash SSD Drives 13
NBBO Test 17
Dell EMC Isilon All-Flash F800 vs Competitive All-Flash Array 18
Dell EMC Isilon All-Flash F800 vs Lustre-based Solution 19
Why Dell EMC Isilon for Tick Data Analytics? 21
Conclusions 22
References 22

Introduction

Document purpose

The document describes the solution overview of Dell EMC Isilon Scale-Out NAS Storage for Kx Systems kdb+ database and tick data analytics.

Solution purpose

The purpose of this document is to present an overview on the kdb+ database and to understand the suitability of Isilon Scale-Out NAS platform as the storage platform of choice for kdb+ tick databases and high-speed time series data analysis.

The Business Challenge

The financial sector is experiencing massive data growth and a need to support more complex data processing and analysis capabilities. In 2013, the NYSE averaged half a billion trades and quotes per day, in 2018, there were approximately 4 billion trades and quotes per day with peaks going over 8 billion trades per day. As a result, the financial services industry now receives multiple terabytes of streaming tick data per day.

Tick Data Analysis requires an analytics application, a framework for tick data, and a robust storage solution. Data analysis is done on three distinct data sets. The data sets can be categorized as:

1. Real-time data
2. Near Real-time data
3. Historical

A real-time and near real-time database stores today's data. Typically, it would be stored in-memory during the day, and written out to the historical database at end of day. Storing real-time data in-memory results in extremely fast update and query performance.

A historical database holds data before today, and its tables would be stored on disk as being much too large to fit in-memory. Each new day's records would be added to the historical database at the end of day.

A high performance data storage solution is required for the near real-time and historical data sets. Key storage requirements include:

- Highly scalable storage solution to efficiently store billions of files
- High read/write performance
- High user concurrency capability
- Fast interconnect between compute resources and the storage platform

Dell EMC Isilon All-Flash Scale-out NAS is an ideal storage platform for tick data analytics applications. Isilon enables financial organizations to non-disruptively scale performance and capacity with simplicity, high-availability, security, and unmatched efficiency.

This paper describes Dell EMC Isilon All-Flash Scale-Out NAS Storage solution for Kx Systems kdb+ database product commonly used in the financial services industry. Kdb+ is a database solution designed to handle increasing volumes of financial tick data and can support both real-time and historical data sets. An overview of kdb+ workflows is provided in this paper to understand the suitability of the Dell EMC Isilon Storage solution for tick data storage and analysis.

Kx Systems (Kdb+) Overview

Kdb+ is a high-performance, high-volume database designed from the outset to handle tremendous volumes of data. It is fully 64-bit, and has built-in multi-core processing and multi-threading capabilities. The same architecture is used for real-time and historical data. The database incorporates its own powerful query language, **q,** so analytics can be run directly on the data in memory or on disk.

Kdb+Tick Architecture

Kdb+tick is an architecture which allows the capture, processing, and querying of real-time and historical tick data. Tick data comes directly from various feeds (NYSE, NASDAQ, etc., as well as aggregators such as Bloomberg and Thomson-Reuters). The following illustration provides a generalized outline of a typical kdb+ tick architecture, followed by a brief explanation of the various components and the through-flow of data.

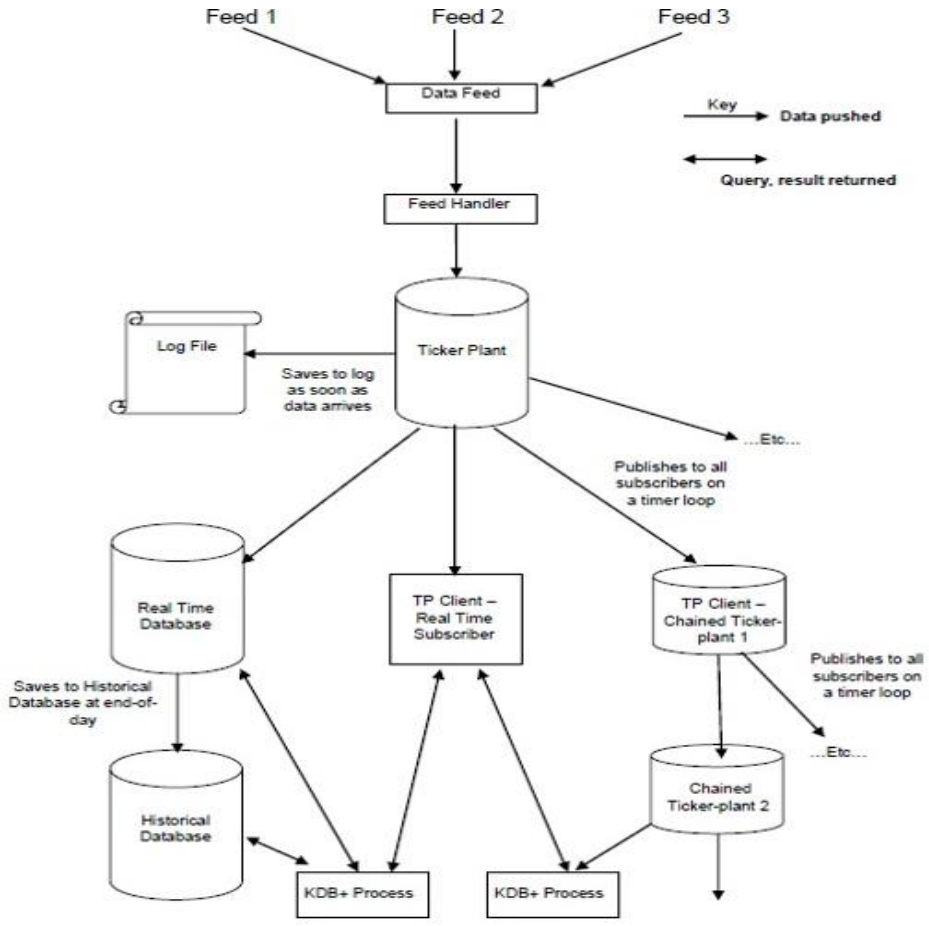


Figure 1 – Kdb+tick Architecture

- The **Data Feeds** are a time series data that are mostly provided by the data feed providers like Reuters, Bloomberg or directly from exchanges.
- To get the relevant data, the data from the data feed is parsed by the **feed handler**.
- Once the data is parsed by the feed handler, it goes to the **ticker-plant**.
- To recover data from any failure, the ticker-plant first updates/stores the new data to the log file and then updates its own tables.
- After updating the internal tables and the log files, the on-time loop data is continuously sent/published to the real-time database and all the chained subscribers who requested data.
- At the end of a business day, the log file is deleted, a new log is created and the real-time database is saved onto the historical database. Once all the data is saved onto the historical database, the real-time database purges its tables.

Kdb+ Database

Kdb+ databases are stored as a series of files and directories on-disk which make it the perfect for Dell EMC Isilon Scale-out High-Performance NAS solution. The Kdb+ design makes handling databases extremely easy as database files can be manipulated as regular files on Isilon. Backing up a kdb+ database can therefore be implemented by using any standard file system backup utility or using Isilon OneFS Snapshot and or SyncIQ features. This is a key difference from traditional databases that use proprietary backup utilities and do not allow direct access to the database files.

Kdb+ also uses standard operating system features for accessing data (memory-mapped files), whereas traditional databases use proprietary techniques in an effort to speed up the reading and writing processes. The typical kdb+ database layout for a tick based system is partitioned by date, although integer partitioning is also possible.

Below is an example of a tick data partition layout on a Dell EMC Isilon NAS for Dec 27, 2015, the Linux program "tree" is issued from a kdb+ database server to show the directory layout. All the directories and files reside on Isilon.

```
[root@master db]# tree -a 2015.12.27
2015.12.27
├── quote
│   ├── asize
│   ├── ask
│   ├── bid
│   ├── bsize
│   ├── .d
│   ├── ex
│   ├── mode
│   ├── sym
│   └── time
└── trade
    ├── cond
    ├── .d
    ├── ex
    ├── price
    ├── size
    ├── stop
    ├── sym
    └── time
```

Figure 2 – Example Kdb+ Tick Database Layout on Dell EMC Isilon

The tick data example above displays the various file system entities which constitute a Kdb+ tick database.

Kdb+ Partitioning

Kdb+ provides a simple method to allow the partitioning of kdb+ data across an array of Isilon NFS mount points. A single text file (par.txt) is populated with a list of directories, one per line. If the storage to be used is DAS or SAN attached, each line represents a discrete file system. However, when using Isilon as the storage platform for tick data, each line can be a separate partition directory within the Isilon OneFS file system. The OneFS filesystem is a single namespace and simplifies data management, security, and scalability.

Kdb+ is intelligent enough to balance new data on ingest to each of the Isilon mount points. Dated directories are effectively round-robin assigned to a given Isilon mount point. A single day's data will live on at most one Isilon mount point.

When configuring a kdb+ par.txt file for an Isilon cluster, the following guidelines should be followed:

- Leverage as many Isilon nodes as possible to maximize I/O throughput.
- Use each Isilon node equally:
    - For example, if there are four Isilon nodes, and eight kdb+ threads, make sure that there are two mount points per Isilon node.
- Each mount point should go to a separate subdirectory.
- If applicable, spread the mount points across more than one client-side 10 or 40 Gigabit Ethernet network interface.

A sample kdb+ par.txt configuration file for a 4 node Dell EMC Isilon All-Flash F800 NAS system is shown below:

Each partition should contain an equal portion of the dataset. For example, if we list the tick data contents of partition 2 and 3 on Isilon:

Partition 2 Listing

```
[root@master p2]# ls
2011.01.04  2011.02.01  2011.03.15  2011.04.06  2011.05.24  2011.06.17  2011.08.04  2011.09.13  2011.10.07  2011.11.08  2011.12.20
2011.01.11  2011.02.09  2011.03.21  2011.04.14  2011.06.01  2011.07.11  2011.08.12  2011.09.21  2011.10.20  2011.11.16  2011.12.29
2011.01.24  2011.02.25  2011.03.29  2011.05.16  2011.06.09  2011.07.27  2011.08.25  2011.09.29  2011.10.31  2011.12.02
[root@master p2]#
```

Partition 3 Listing

```
[root@master p3]# ls
2011.01.10  2011.02.10  2011.03.22  2011.04.15  2011.05.17  2011.06.10  2011.07.28  2011.09.06  2011.10.13  2011.11.17  2011.12.19
2011.01.25  2011.02.18  2011.03.30  2011.04.29  2011.05.25  2011.07.12  2011.08.05  2011.09.14  2011.10.24  2011.11.25  2011.12.27
2011.02.02  2011.03.14  2011.04.07  2011.05.09  2011.06.02  2011.07.20  2011.08.29  2011.09.22  2011.11.01  2011.12.06
[root@master p3]#
```

Both partitions above have 32 entries, this is an even tick data distribution and promotes even performance across all threads. With eight mount points defined in the example par.txt provided, kdb+ should be configured with eight or less threads. When a kdb+ q process is started with slave threads (as is the case most of the time), each of the partitions in par.txt is handed out to slave threads in a round-robin fashion. If there are more par.txt lines than slaves, each slave will get assigned some number of par.txt entries. This is why it is beneficial for the number of entries in par.txt to be divisible by the number of slave threads.

For example, if you have a par.txt file with sixteen mount points, and you have four slave threads, then each slave thread will get four mount points, resulting in even performance across all threads. If the balance of mount points among slave threads is not even (e.g.: if we had fourteen mount points in the above four-thread example), then performance between threads would be uneven.

The runtime of a database query is gated by the speed of the slowest thread. If some threads are responsible for a larger set of data than others, they will slow the entire query down, as idle threads cannot be dynamically reassigned to share work with other threads. The goal of many kdb+ queries is to scan through a date-range worth of ticker history (e.g.: the last 21 days), and return a result, such as ‘find all trades for all symbols where bidPrice was greater than 50'. This results in a full 'table-scan' of all data in this date-range.

Each kdb+ q process would perform scans of data under their jurisdiction. When each slave thread is finished reading data, it reports the result to the parent thread, which then collates the data and presents it back to the user or application that initiated the query. Note however that the parent thread is not able to tabulate the results until ALL slave threads have reported back. This means that the overall runtime can be negatively impacted when a single thread underperforms its peers. This is the primary reason why load should be as evenly balanced among threads as possible. If the data is not evenly distributed across all the partitions, just use the “mv” command to move specific tick data from one Isilon partition to another until you end up with an even tick data distribution, kdb+ will take care of finding the data throughout all Isilon partitions automatically during data scans.

Dell EMC Isilon F800 All-Flash NAS

Dell EMC Isilon F800 all-flash scale-out NAS storage provides up to 250,000 IOPS and 15 GB/s bandwidth per chassis. With a choice of SSD drive capacities, all-flash storage ranges from 96 TB to 924 TB per chassis making the Isilon F800 ideal for demanding storage requirements in high volume tick data applications.

In additional to an all-flash high-performance scale-out hardware design of the Isilon F800, the embedded storage operating system (Isilon OneFS) provides a unifying clustered file system with built-in scalable data protection that simplifies storage management and administration. OneFS is a fully symmetric file system with no single point of failure — taking advantage of clustering not just to scale performance and capacity, but also to allow for any-to-any failover and multiple levels of redundancy that go far beyond the capabilities of RAID. Self-encrypting drives can be used to provide enhanced data security.

OneFS allows hardware to be incorporated or removed from the cluster at will and at any time, abstracting the data and applications away from the hardware. Data is given infinite longevity and the cost and pain of data migrations and hardware refreshes are eliminated.

Isilon nodes

OneFS works exclusively with Isilon scale-out NAS nodes, referred to as a “cluster”. A single Isilon cluster consists of multiple nodes, which are rack-mountable enterprise appliances containing: memory, CPU, networking, Ethernet or low-latency InfiniBand interconnects, disk controllers and storage media. As such, each node in the distributed cluster has compute as well as storage capabilities.

With the new generation of Isilon hardware (“Gen 6”), a single chassis of 4 nodes in a 4U form factor is required to create a cluster, which currently scales up to 252-nodes. Previous Isilon hardware platforms need a minimum of three nodes and 6U of rack space to form a cluster. There are several different types of nodes, all of which can be incorporated into a single cluster, where different nodes provide varying ratios of capacity to throughput or Input/Output operations per second (IOPS). This provides customers the ability to tier data and meet price and performance requirements by using different Isilon storage node types in the storage cluster.

Each node or chassis added to a cluster increases aggregate disk, cache, CPU, and network capacity. OneFS leverages each of the hardware building blocks, so that the whole becomes greater than the sum of the parts. The RAM is grouped together into a single coherent cache, allowing I/O on any part of the cluster to benefit from data cached anywhere. A file system journal ensures that writes are safe across power failures. Spindles and CPU are combined to increase throughput, capacity and IOPS as the cluster grows, for access to one file or for multiple files. A cluster’s storage capacity can range from a minimum of 18 terabytes (TB) to a maximum of ~ 58 petabytes (PB). The maximum capacity will continue to increase as disk drives and node chassis continue to get denser.

Isilon nodes are broken into several classes, or tiers, according to their functionality:

Tier	I/O Profile	Drive Media	Isilon Node Type
Extreme Performance	High Perf, Low Latency	Flash	F810, F800
Performance	Transactional I/O	SAS & SSD	H600, S210
Hybrid/Utility	Concurrency & Streaming Throughput	SATA & SSD	H5600, H500, H400, X410, X210
Archive	Nearline & Deep Archive	SATA	A2000, A200, NL410, HD400

This paper focuses on the F800 node type for kdb+. As with most data, only a portion of a historical kdb+ instance will be accessed frequently. Generally, the most recent thirty days' worth of trading data is queried frequently, with older data being read periodically. For example: one back history strategy may involve scanning the past thirty days of trade data on a daily basis to validate new and existing algorithms. Another strategy may involve scanning the past 180 days of trade data on a weekly basis, and so on.

For this reason, it is most cost effective to develop a tiered approach to storing this data. Isilon provides this functionality inherently within the OneFS™ file system using SmartPools™, moving data from tier to tier without the need to change mount points or directory structure. All data, regardless of storage tier, continues to be accessible to clients, without downtime or application reconfiguration.

Network

There are two types of networks associated with a cluster: internal and external.

Back-end (internal) network

All intra-node communication in a storage cluster is performed across a dedicated backend network, comprising either 10 or 40 GbE Ethernet, or low-latency QDR InfiniBand (IB). This back-end network, which is configured with redundant switches for high availability, acts as the backplane for the storage cluster. This enables each storage node to act as a contributor in the storage cluster and isolating node-to-node communication to a private, high-speed, low-latency network. This back-end network utilizes Internet Protocol (IP) for node-to-node communication.

Front-end (external) network

Clients connect to the storage cluster using Ethernet connections (1GbE, 10GbE or 40GbE) that are available on all storage nodes. Because each storage node provides its own Ethernet ports, the amount of network bandwidth available to the storage cluster scales linearly with performance and capacity. The Isilon storage cluster supports standard network communication protocols to a customer network, including **NFS, SMB, HTTP, FTP, and HDFS**. Additionally, OneFS provides full integration with both IPv4 and IPv6 environments.

Isilon OneFS File System

The OneFS file system is based on the UNIX file system (UFS) and, hence, is a very fast distributed file system. Each cluster creates a single namespace and file system. This means that the file system is distributed across all nodes in the cluster and is accessible by clients connecting to any node in the cluster. There is no partitioning, and no need for volume creation.

Because all information is shared among nodes across the internal network, data can be written to or read from any node, thus optimizing performance when multiple Kdb+ users or applications are concurrently reading and writing to the same set of tick data.

For more details on Isilon and OneFS please see Isilon Technical Overview.

Kdb+ and Isilon Topology

The kdb+ and Isilon F800 cluster tested in this paper uses NFS v3 as the network communication protocol with a 40GbE front-end and 40GbE back-end network. The kdb+ compute cluster consists of four Dell EMC PowerEdge R940 servers with 40GbE interfaces that connect to the front-end network over 40GbE using NFS. The compute cluster runs the kdb+ databases, the Isilon cluster stores and serves all the tick data for the environment.

A high-level example of a kdb+ and Isilon topology with combined hardware, software, and networks is shown below:

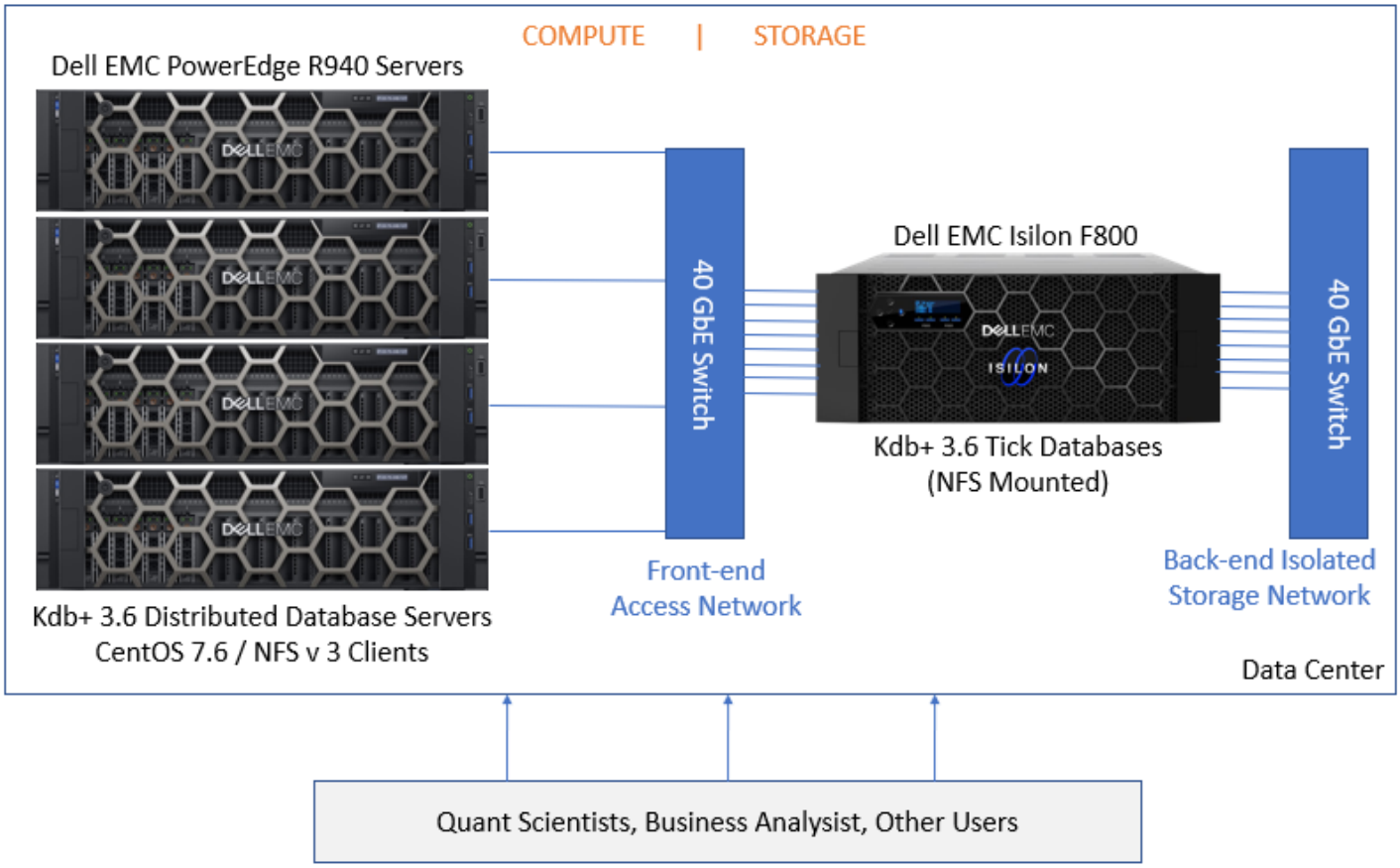


Figure 3 – Example KDB+ and Isilon Topology for Tick Data Analytics

Dell EMC PowerEdge R940

Data is the most precious commodity of our times. A modern IT infrastructure is necessary to process data in a usable way across an organization. PowerEdge four-socket rack servers can address demanding, large data sets that require high performance and large capacity to deliver consistent and fast results.

The PowerEdge R940 rack server is designed to accelerate mission critical applications. With four sockets powered by the latest Intel® Xeon® Scalable processors and up to 12 NVMe drives, the PowerEdge R940 provides high performance in just 3U. Combined with up to 15.36TB of memory, large storage and 13 PCIe Gen 3 slots, the PowerEdge R940 has all the resources to maximize performance and scale to meet future demands.

The PowerEdge R940 can drive in-memory databases, ERP, e-commerce and other demanding, large data sets. It can run large virtualized corporate applications or be the foundation supporting a multi-tiered infrastructure. Optimized workload tuning can speed and simplify configuration processes. The PowerEdge R940 streamlines the management of routine tasks with intelligent automation. Built-in layers of security help prevent cyber-attacks and keep data safe.

Dell EMC Tested Configuration

Kdb+

The kdb+ version tested for this paper is version 3.6. The Antuco database represents a year of tick data, approximately 3 TB is size. The Kanaga database represents multiple years of tick data, approximately 54 TB is size.

Compute Nodes

All the compute nodes are identical Dell PowerEdge R940 servers each with 4 x Intel Xeon Platinum 8168 @ 2.7 GHz, 30727 GB RAM, and 40G NIC running CentOS Linux release 7.6.

A total of 4 x PowerEdge R940 servers were used for various test scenarios that are described in detail in the *Performance Test Results* section of this paper.

Isilon F800 (NFS Mounted from each Kdb+ database node)

A single 4U Isilon F800 Chassis (4 nodes total) with a total of 60 x 3.2 TB SSD drives were used for Kdb+ Isilon testing. Each Isilon F800 node has 2 x 40GbE connections to the front-end access network and 2 x 40GbE connection to the private back-end storage network.

The specific Isilon Model tested: Isilon F800-4U-Single-256GB-1x1GE-2x40GE SFP+-24TB SSD

The Isilon OneFS release tested: OneFS v 8.1.2.0.

Kdb+ Databases

Two kdb+ databases were tested – **antuco** and **kanaga**.

The **antuco** database size is 3.3TB and consist of all the tick data for the year 2011. This database is used for real-time performance testing and testing a varying number of concurrent users.

The **kanaga** database size is 54TB and consist of all the tick data for the years 2003-2015. This database represents historical tick data and is used for performance testing large data sets and testing a varying number of concurrent users.

STAC-M3 Benchmark Suite

The Securities Technology Analysis Center (STAC) provides technology research and testing tools based on standards developed by the STAC Benchmark Council – a consortium of over 300 financial institutions and more than 50 vendor organizations that develop benchmark standards useful to financial organizations. The STAC-M3 Benchmark suite is the industry standard for testing solutions that enable high-speed analytics on time-series tick data.

The performance test results included in this paper reference the audited Dell EMC Isilon F800 STAC-M3 report (SUT ID KDB190430).

The complete STAC-M3 report is available at DellEMC.com. The test results cover both antuco and kanaga kdb+ databases that reside on a single 4U Isilon F800 All-Flash NAS cluster.

STAC-M3 Performance Test Results

High Concurrency Tests

The STAC-M3 benchmark specifications marked as “v1” focus on storage system performance, the v1 workloads are deliberately heavy on I/O. The STAC-M3 benchmark tests can submit requests from independent threads from multiple clients, e.g. 10 clients each using 10 threads will create 100 total requesting threads to the storage system.

The prominent v1 storage stress test from the STAC-M3 test suite is the VWAB-12DaysNoOverlap benchmark with 100 client threads. This test has a 100 client threads with each thread requesting a 4-hour volume-weighted bid over 12 days for 1% of symbols with no overlap in symbols for a date range of 1 year to 5 years. Isilon can easily support thousands of concurrent connections, the STAC-M3 test suite does not come close to stress testing the Isilon storage cluster as shown later in this paper.

Dell EMC Isilon F800 vs Direct-Attached Solution using Optane 3D NAND Flash SSD Drives

Below are the high concurrency VWAB-12DaysNoOverlap (100 clients) test results between the STAC audited Dell EMC Isilon F800 tested configuration and a recently STAC audited 4-socket server with direct-attached Intel Optane and 3D NAND flash SSDs solution. The results cover both Antuco and Kanaga (Year 1-5) data sets. Lower response times are better.

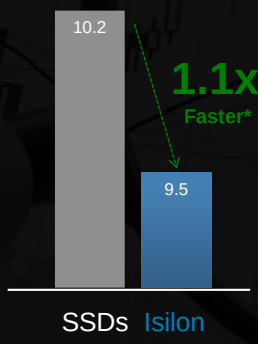
The Dell EMC Isilon F800 STAC SUT ID KDB190430 report can be obtained at dellemc.com, the same report is also available at stackresearch.com.

The Intel Optane and 3D NAND flash SSD STAC report can be obtained from stacresearch.com, SUT ID 181009.

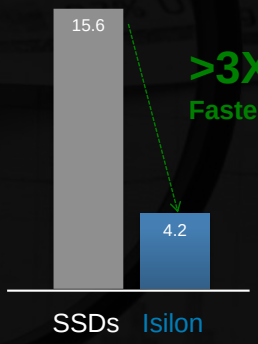
Isilon F800 STAC-M3 Results

High Concurrency Tests vs Direct Attached SSD Storage

Antuco 12 Day Run
100 Clients (sec)
(STAC-M3.v1.100T.VWAB-12D-NO.TIME)



Kanaga 12 Day Run – Year 1
100 Clients (sec)
(STAC-M3.ß1.100T.YR1VWAB-12D-HO)

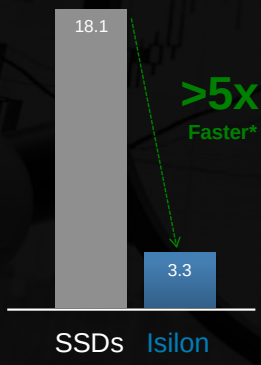


*Based on a STAC-M3™ Benchmark commissioned by Dell EMC, 'STAC Report: kdb+ 3.6 on 4 x Dell EMC PowerEdge R940 servers with Dell EMC Isilon F800 All-Flash NAS: SUT ID: KDB190430', June 2019, compared to SUT ID KDB181009. Results obtained in a 100-user, 12-day VWAB operation on each year of the STAC-M3 Kanaga dataset. Full report here: https://www.stacresearch.com/news/2019/06/16/KDB190430, https://www.dellemc.com/resources/en-us/asset/analyst-reports/products/storage/stac-m3-report-kdb.pdf. "STAC" and all STAC names are trademarks or registered trademarks of the Securities Technology Analysis Center, LLC.

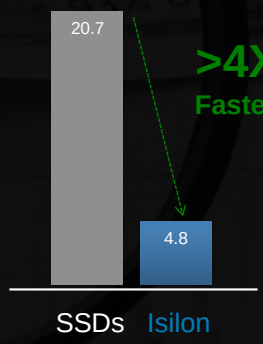
Isilon F800 STAC-M3 Results

High Concurrency Tests vs Direct Attached SSD Storage

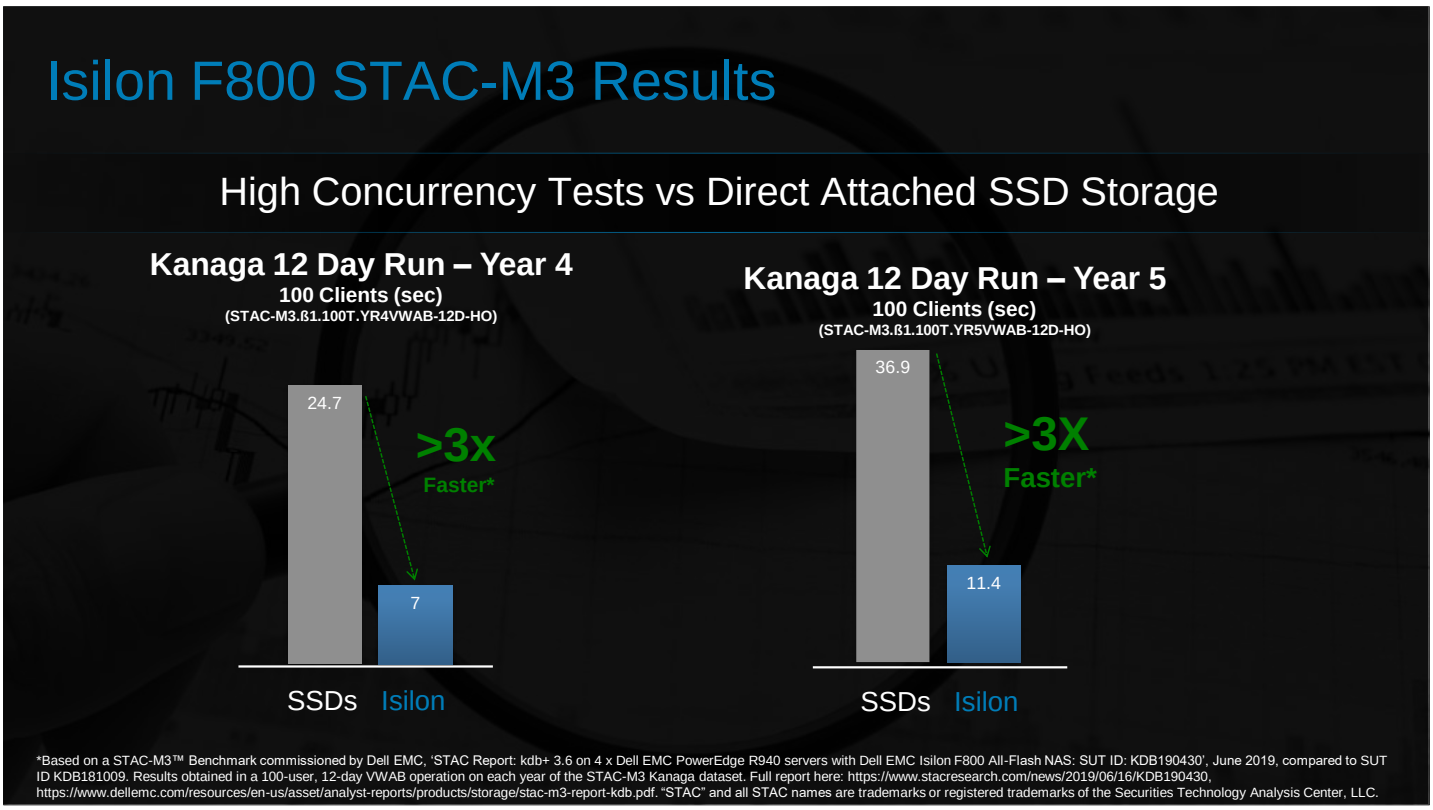
Kanaga 12 Day Run – Year 2
100 Clients (sec)
(STAC-M3.ß1.100T.YR2VWAB-12D-HO)



Kanaga 12 Day Run – Year 3
100 Clients (sec)
(STAC-M3.ß1.100T.YR3VWAB-12D-HO)

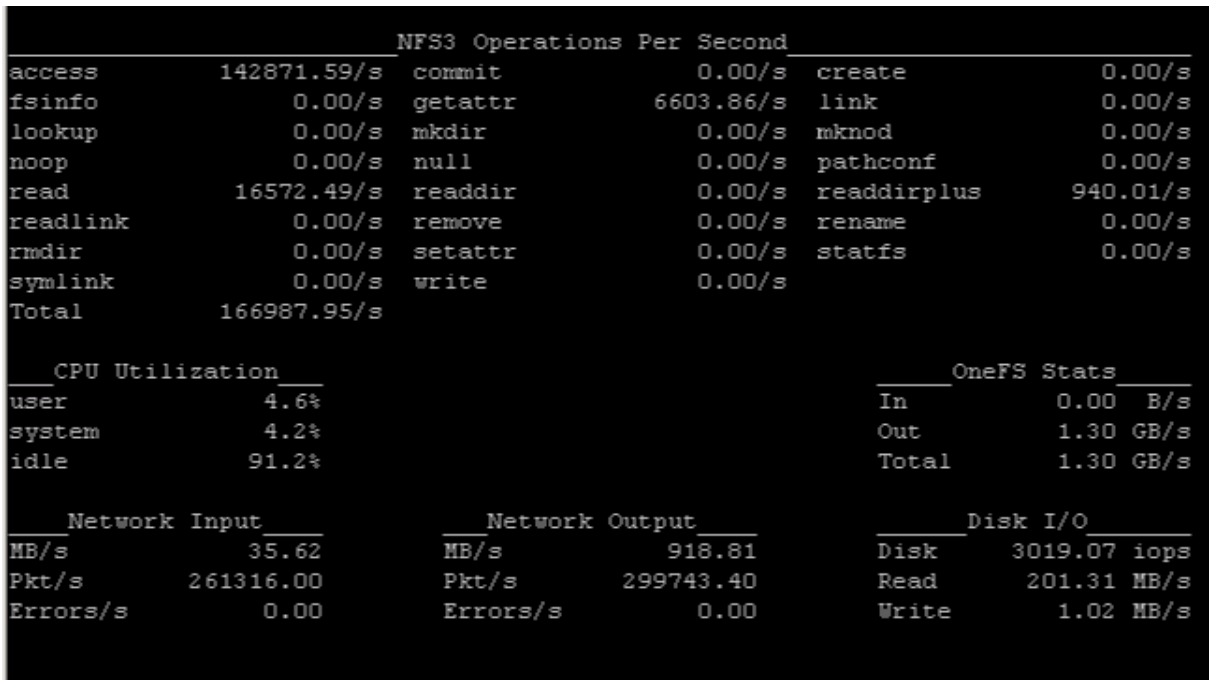


*Based on a STAC-M3™ Benchmark commissioned by Dell EMC, 'STAC Report: kdb+ 3.6 on 4 x Dell EMC PowerEdge R940 servers with Dell EMC Isilon F800 All-Flash NAS: SUT ID: KDB190430', June 2019, compared to SUT ID KDB181009. Results obtained in a 100-user, 12-day VWAB operation on each year of the STAC-M3 Kanaga dataset. Full report here: https://www.stacresearch.com/news/2019/06/16/KDB190430, https://www.dellemc.com/resources/en-us/asset/analyst-reports/products/storage/stac-m3-report-kdb.pdf. "STAC" and all STAC names are trademarks or registered trademarks of the Securities Technology Analysis Center, LLC.



It is clear from the high concurrency high I/O STAC-M3 results that Dell EMC Isilon F800 NAS performs very well with both the Antuco and Kanaga datasets. As the data volume increased from year one through year five during the Kanaga tests, the direct-attached storage system shows significant increases in I/O response times while the Dell EMC Isilon system maintained a steady increase in response times. This is quite impressive considering the 4-socket server is using local storage with fast Optane (3D NAND) SSD drives. The Dell EMC Isilon F800 outperformed this STAC audited direct-attached storage solution in all the VWAB 100 thread tests.

An interesting observation to note during the high concurrency high I/O VWAB tests pertains to the monitored load on Isilon, below are captured statistics on Isilon during the STAC-M3.v1.100T.VWAB.12D-NO.TIME test:



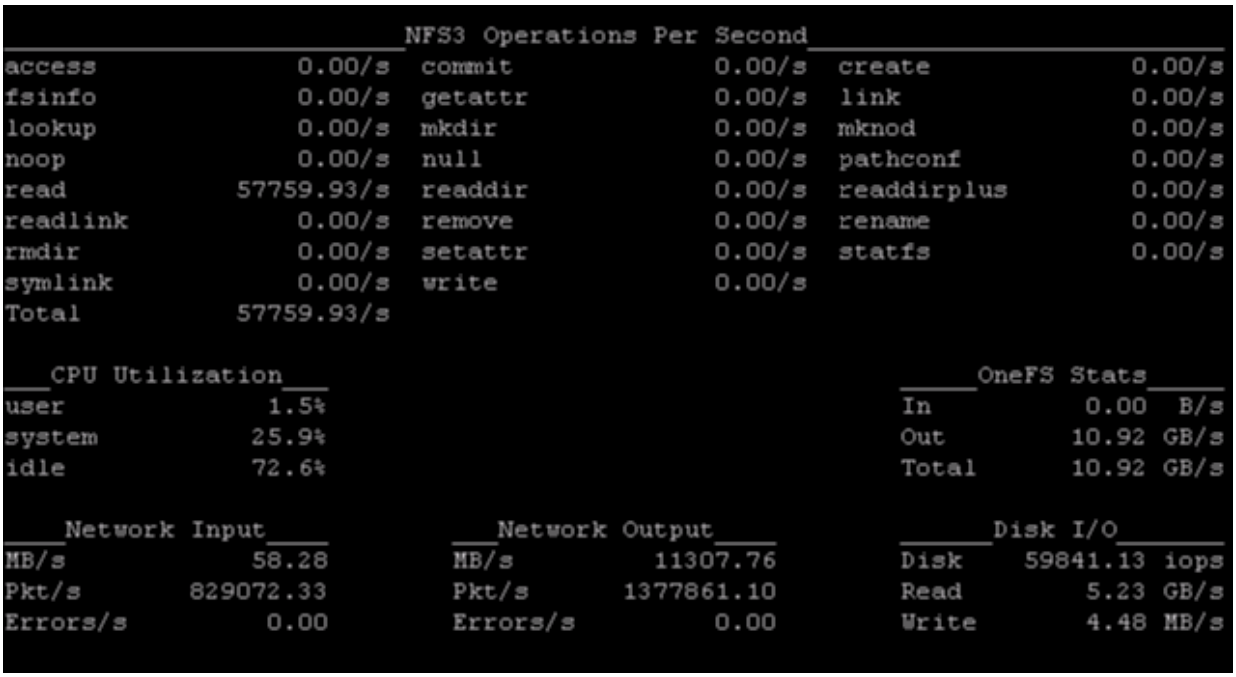
NFS3 Operations Per Second					
access	142871.59/s	commit	0.00/s	create	0.00/s
fsinfo	0.00/s	getattr	6603.86/s	link	0.00/s
lookup	0.00/s	mkdir	0.00/s	mknod	0.00/s
noop	0.00/s	null	0.00/s	pathconf	0.00/s
read	16572.49/s	readdir	0.00/s	readdirplus	940.01/s
readlink	0.00/s	remove	0.00/s	rename	0.00/s
rmdir	0.00/s	setattr	0.00/s	statfs	0.00/s
symlink	0.00/s	write	0.00/s		
Total	166987.95/s				

CPU Utilization		OneFS Stats	
user	4.6%	In	0.00 B/s
system	4.2%	Out	1.30 GB/s
idle	91.2%	Total	1.30 GB/s

Network Input		Network Output		Disk I/O	
MB/s	35.62	MB/s	918.81	Disk	3019.07 iops
Pkt/s	261316.00	Pkt/s	299743.40	Read	201.31 MB/s
Errors/s	0.00	Errors/s	0.00	Write	1.02 MB/s

The Isilon statistics show the storage system to be 91.2% idle during the test, Isilon has plenty of resources to process much more I/O.

Below are captured statistics on Isilon during the largest data volume test STAC-M3.ß1.100T.YR5VWAB-12D-HO that covers 5 years:



NFS3 Operations Per Second					
access	0.00/s	commit	0.00/s	create	0.00/s
fsinfo	0.00/s	getattr	0.00/s	link	0.00/s
lookup	0.00/s	mkdir	0.00/s	mknod	0.00/s
noop	0.00/s	null	0.00/s	pathconf	0.00/s
read	57759.93/s	readdir	0.00/s	readdirplus	0.00/s
readlink	0.00/s	remove	0.00/s	rename	0.00/s
rmdir	0.00/s	setattr	0.00/s	statfs	0.00/s
symlink	0.00/s	write	0.00/s		
Total	57759.93/s				

CPU Utilization		OneFS Stats	
user	1.5%	In	0.00 B/s
system	25.9%	Out	10.92 GB/s
idle	72.6%	Total	10.92 GB/s

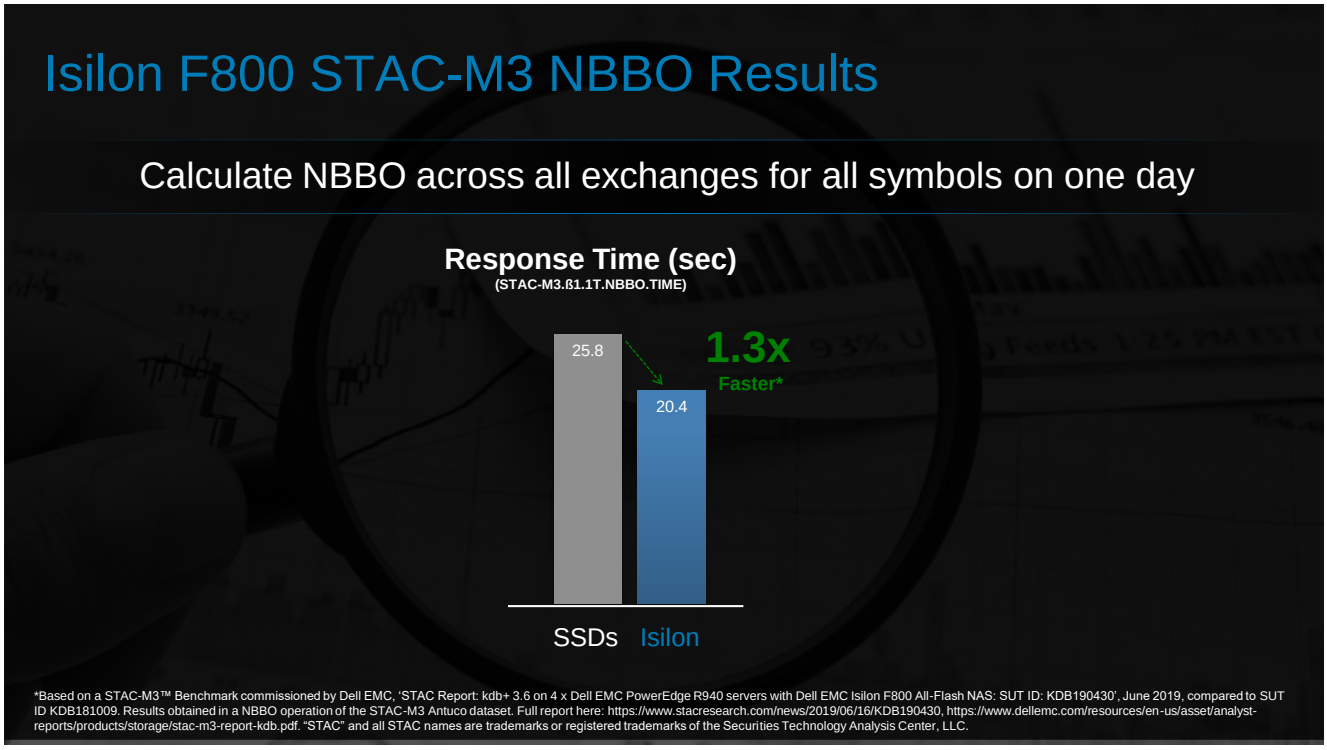
Network Input		Network Output		Disk I/O	
MB/s	58.28	MB/s	11307.76	Disk	59841.13 iops
Pkt/s	829072.33	Pkt/s	1377861.10	Read	5.23 GB/s
Errors/s	0.00	Errors/s	0.00	Write	4.48 MB/s

The statistics show that Isilon has plenty of resources to spare during the largest data volume test with the storage system being 72.6% idle during the test. Regardless of the available resources on Isilon during the VWAB tests, the results show that Isilon F800 All-Flash NAS is indeed a good storage solution to handle high concurrency and high I/O load for both small and large data sets used for tick data analysis. By separating storage from compute, the amount of context switching within the operating system is reduced and more CPU and Memory resources can be used for kdb+ data processing. Therefore, Isilon can outperform local attached SSD storage for high concurrency and high I/O workloads with a single 4U Isilon F800 chassis.

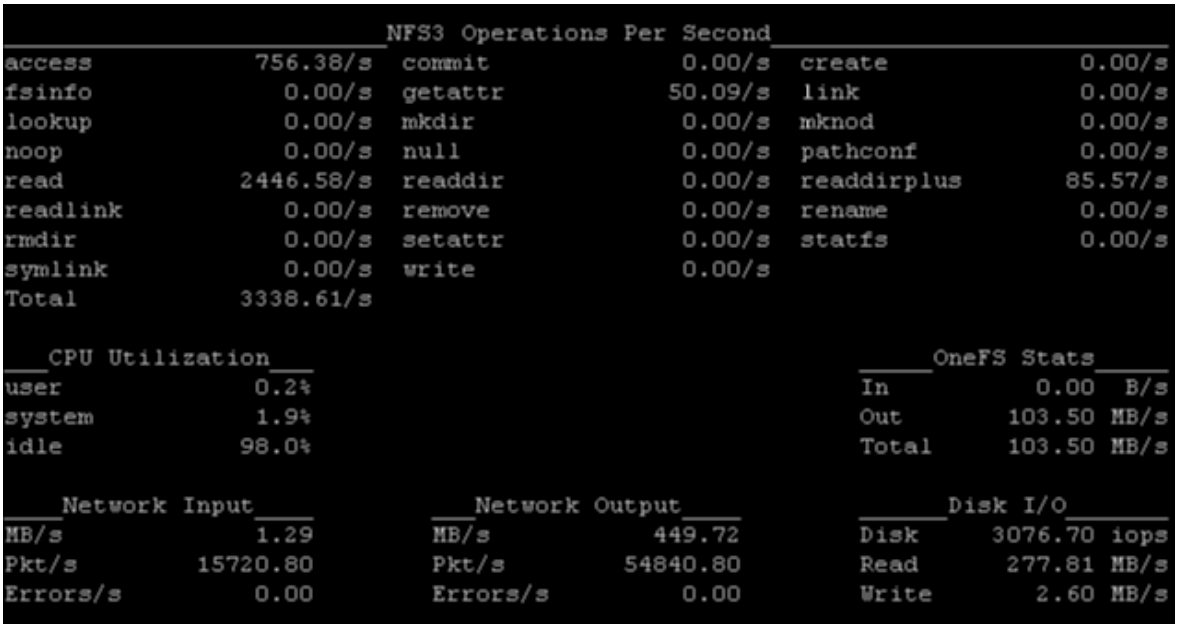
NBBO Test

The STAC-M3 benchmark can query the National Best Bid and Offer (NBBO) across all 10 exchanges for all symbols on the most recent day. This test has a heavy read and write I/O profile and is also heavy on compute intensity.

Below are the results of the NBBO (STAC-M3.ß1.1T.NBBO.TIME) test for the STAC audited direct-attached storage solution and the tested Dell EMC F800 configuration:



The NBBO result is faster with Isilon. Below are the Isilon storage system statistics during the NBBO test:



NFS3 Operations Per Second

access	756.38/s	commit	0.00/s	create	0.00/s
fsinfo	0.00/s	getattr	50.09/s	link	0.00/s
lookup	0.00/s	mkdir	0.00/s	mknod	0.00/s
noop	0.00/s	null	0.00/s	pathconf	0.00/s
read	2446.58/s	readdir	0.00/s	readdirplus	85.57/s
readlink	0.00/s	remove	0.00/s	rename	0.00/s
rmdir	0.00/s	setattr	0.00/s	statfs	0.00/s
symlink	0.00/s	write	0.00/s		
Total	3338.61/s				

CPU Utilization

user	0.2%
system	1.9%
idle	98.0%

OneFS Stats

In	0.00 B/s
Out	103.50 MB/s
Total	103.50 MB/s

Network Input

MB/s	1.29
Pkt/s	15720.80
Errors/s	0.00

Network Output

MB/s	449.72
Pkt/s	54840.80
Errors/s	0.00

Disk I/O

Disk	3076.70 iops
Read	277.81 MB/s
Write	2.60 MB/s

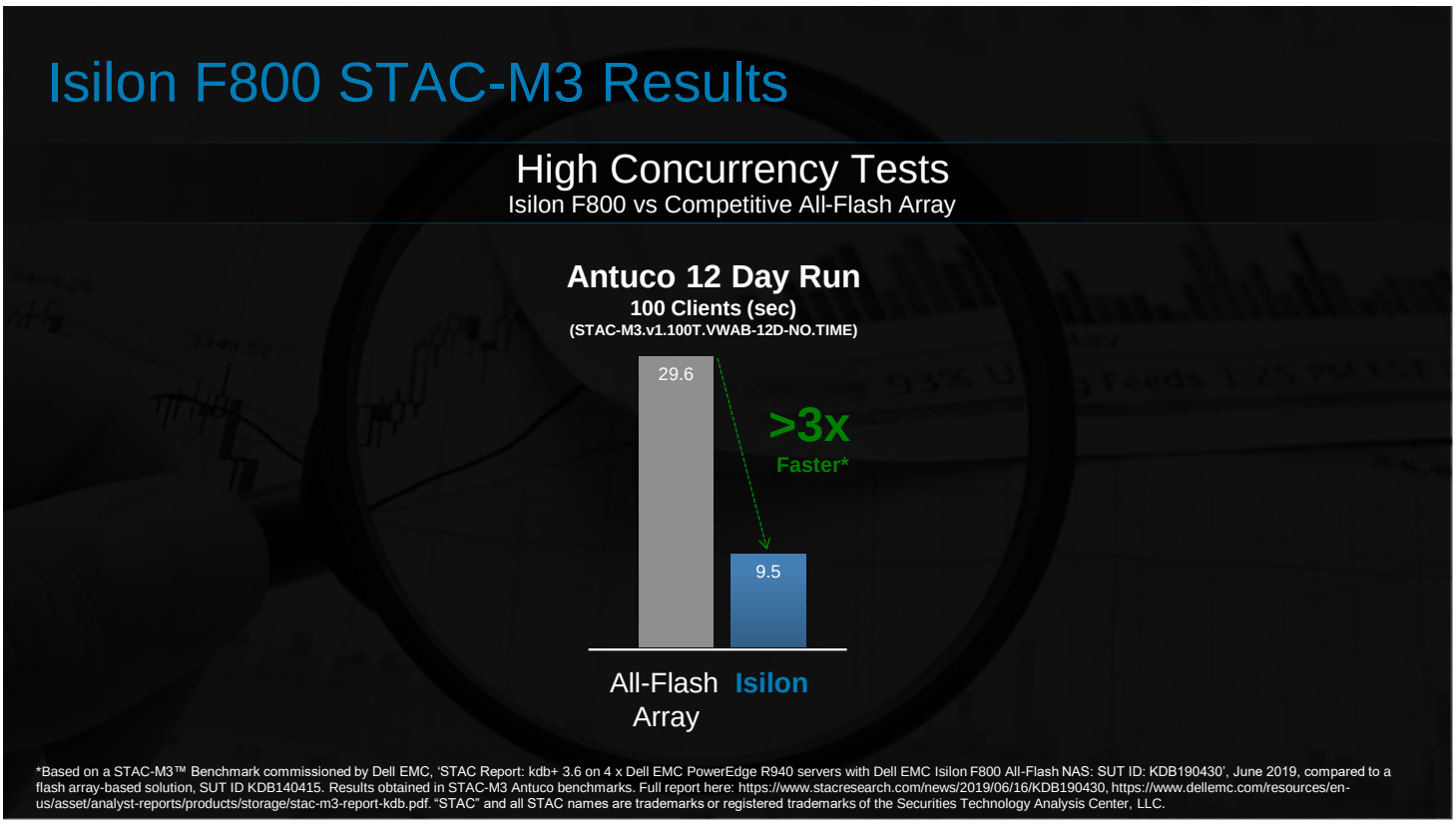
Isilon is 98% idle during the NBBO test. Again, Isilon has plenty of resources while processing STAC-M3 database queries.

What about competing storage solutions to Dell EMC Isilon that also separate storage from compute - like a Netapp All-Flash array or a DDN Lustre-based storage solution? STAC has also audited those storage solutions as well. Let's see how a single 4U Dell EMC Isilon F800 performed against Netapp and DDN for high concurrency and high I/O VWAB 100 client thread STAC-M3 tests.

Dell EMC Isilon All-Flash F800 vs Competitive All-Flash Array

The STAC report for a competitive All-Flash storage array can be obtained from stacresearch.com, SUT ID KDB140145. The competitor chose to only test the Antuco data set, which is the smaller of the two data sets. Again, the prominent STAC-M3 storage test is the VWAB for 12 Days with 100 Client Threads.

Below are the performance comparison results for VWAB (100 Client Threads) between a competitive all-flash array and Dell EMC Isilon:



Again, Dell EMC Isilon is the faster storage solution for the STAC-M3 high concurrency VWAB test.

Let's see how a Lustre-based storage solution compares with Dell EMC Isilon F800.

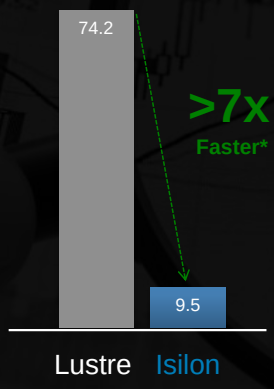
Dell EMC Isilon All-Flash F800 vs Lustre-based Solution

The STAC report for a Lustre-based solution can be obtained from stacresearch.com, SUT ID KDB150528. The Lustre-based audited solution includes both the Antuco and Kanaga data sets, however the Kanaga data set only included 4 years of tick data whereas the Kanaga dataset on audited Isilon solution included 5 years of tick data. Even with an additional year of tick data on Isilon, Isilon significantly outperformed the Lustre-based storage solution as shown below.

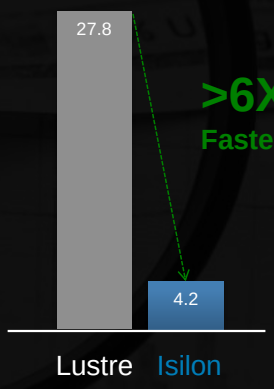
Isilon F800 STAC-M3 Results

High Concurrency Tests vs Lustre Storage Solution

Antuco 12 Day Run
100 Clients (sec)
(STAC-M3.v1.100T.VWAB-12D-NO.TIME)



Kanaga 12 Day Run – Year 1
100 Clients (sec)
(STAC-M3.ß1.100T.YR1VWAB-12D-HO)

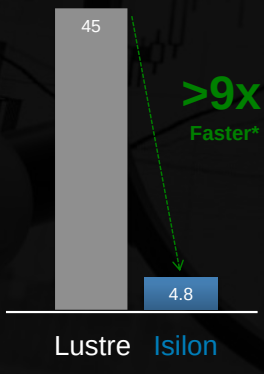


*Based on a STAC-M3™ Benchmark commissioned by Dell EMC, 'STAC Report: kdb+ 3.6 on 4 x Dell EMC PowerEdge R940 servers with Dell EMC Isilon F800 All-Flash NAS: SUT ID: KDB190430', June 2019, compared to a Lustre-based solution, SUT ID KDB150528. Results obtained in STAC-M3.β1.100T.VWAB-12D-NO.TIME (Antuco). Full report here: https://www.stacresearch.com/news/2019/06/16/KDB190430, https://www.dellemc.com/resources/en-us/asset/analyst-reports/products/storage/stac-m3-report-kdb.pdf. "STAC" and all STAC names are trademarks or registered trademarks of the Securities Technology Analysis Center, LLC

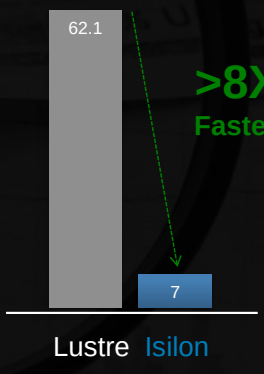
Isilon F800 STAC-M3 Results

High Concurrency Tests vs Lustre Storage Solution

Kanaga 12 Day Run – Year 3
100 Clients (sec)
(STAC-M3.ß1.100T.YR2VWAB-12D-HO)



Kanaga 12 Day Run – Year 4
100 Clients (sec)
(STAC-M3.ß1.100T.YR3VWAB-12D-HO)



*Based on a STAC-M3™ Benchmark commissioned by Dell EMC, 'STAC Report: kdb+ 3.6 on 4 x Dell EMC PowerEdge R940 servers with Dell EMC Isilon F800 All-Flash NAS: SUT ID: KDB190430', June 2019, compared to a Lustre-based solution, SUT ID KDB150528. Results obtained in STAC-M3.β1.100T.VWAB-12D-NO.TIME (Antuco). Full report here: https://www.stacresearch.com/news/2019/06/16/KDB190430, https://www.dellemc.com/resources/en-us/asset/analyst-reports/products/storage/stac-m3-report-kdb.pdf. "STAC" and all STAC names are trademarks or registered trademarks of the Securities Technology Analysis Center, LLC

VWAB results from Year 2 and 5 are excluded from the STAC-M3 Lustre-based solution – SUT ID 150528. In total, Dell EMC Isilon outperformed the Lustre-based solution in 15 of 16 STAC-M3 Kanaga benchmarks. This is based on the STAC audited Dell EMC Isilon Report (SUT ID KDB190430) and the STAC audited Lustre-based Report (SUT ID KDB150528).

Why Dell EMC Isilon for Tick Data Analytics?

Dell EMC Isilon F800 all-flash scale-out NAS storage provides up to 250,000 IOPS and 15 GB/s bandwidth per chassis. With a choice of SSD drive capacities, all-flash storage capacities ranges from 96 TB to 924 TB per chassis making the Isilon F800 ideal for demanding storage requirements in high-speed time series analysis use cases.

Besides the high-performance capabilities of Dell EMC Isilon, financial services customers value the ease of scalability and the enterprise features of the embedded OneFS storage operating system. Snapshots and SyncIQ features make disaster recovery with kdb+ easy. Isilon's multi-protocol support avoids complexity and any need for client-side kernel modifications or installation of any proprietary drivers.

Dell EMC is also the largest OEM partner for NVIDIA. Dell EMC Isilon has been vetted and certified with NVIDIA DGX systems and offers additional AI Ready Solutions. As more and more financial services organizations expand the use of machine learning algorithms and artificial intelligence in their enterprise, Dell EMC Isilon is the leading storage vendor publishing AI benchmarks that scale beyond 32 GPU's, see the Dell EMC Isilon NVIDIA DGX Whitepaper.

Furthermore, a key differentiator between Dell EMC Isilon and other competitive storage solutions is the engineering design of both the Isilon hardware and storage software. We can support high concurrency and high I/O performance and scalability because of how the hardware and software is engineered together.

File reads in an Isilon cluster are all distributed on multiple nodes, and even across multiple drives within nodes. When reading or writing to an Isilon cluster, the node a client attaches to manages data access for the client. In a read operation, the managing node gathers all the data from various nodes in the cluster and presents it in a cohesive way to the requestor.

Due to the use of cost-optimized industry standard hardware, Isilon clusters provide a high ratio of cache to disk (multiple GB per node) that is dynamically allocated for read and write operations as needed. This RAM-based cache is unified and coherent across all nodes in the cluster, allowing a client read request on one node to benefit from I/O already transacted on another node. These cached blocks can be quickly accessed from any node across the low-latency backend network, allowing for a large, efficient RAM cache, that accelerates read performance.

As an Isilon cluster grows larger, the cache benefit increases. For this reason, the amount of I/O to disk on an Isilon cluster is generally substantially lower than it is on traditional storage platforms, allowing for reduced latencies and a better user experience. For files marked with an Isilon access pattern of concurrent or streaming, Isilon OneFS can take advantage of pre-fetching data based on heuristics used by the Isilon SmartRead component. SmartRead can create a data pipeline from L2 cache, prefetching into a local L1 cache on the given node. This greatly improves sequential-read performance across all protocols and means that reads come directly from RAM within milliseconds. For high-sequential cases, SmartRead can aggressively prefetch ahead, allowing reads or writes of individual files at very high data rates.

SmartRead's intelligent caching allows for very high read performance with high levels of concurrent access. This is why Isilon performs very well with the high concurrency VWAB STAC-M3 benchmarks as well as various other AI related benchmarks.

Conclusions

The kdb+ product has been designed in anticipation of vast increases in data volumes. The ability of OneFS to scale to multi-petabytes in a single file system while delivering high performance I/O makes Isilon storage ideal for kdb+ near real-time and historic workflows.

This paper shows that Dell EMC Isilon F800 All-Flash NAS storage solution performs very well under high concurrency and I/O load with kdb+ for both small and large tick data sets. The STAC-M3 benchmark results for all VWAB and NBBO tests are included in this paper and show excellent results for Isilon when compared to recently published benchmarks using local attached SSD storage for the same tests. The STAC-M3 benchmark results for VWAB 100 thread tests for a competitive all-flash array and a Lustre-based solution are also included, again Isilon results are better than these competing shared storage solutions for high concurrency 100 thread tests.

With Dell EMC Isilon, financial organizations and kdb+ administrators can effortlessly scale from 10s of terabytes to 10's petabytes within a single file system, single volume, and with a single point of administration. Dell EMC Isilon delivers high-performance and high-throughput without adding management complexity.

References

KX Systems: https://kx.com/

STAC: https://stacresearch.com/

Dell EMC STAC Report: https://www.dellemc.com/resources/en-us/asset/analyst-reports/products/storage/stac-m3-report-kdb.pdf

