

ACCELERATING Kx kdb⁺ WITH SFA12KX[™] - 40 & INTEL[®] LUSTRE

SHARED STORAGE WITH PARALLEL ACCESS DELIVERS BETTER STRATEGIES, FASTER

Using fast, scalable, shared, external disk systems with massively parallel access to data, researchers can perform analysis against much larger datasets than by batching smaller datasets direct attached storage or serial network attached storage.

Visionary hedge funds, proprietary trading firms and other financial institutions have been changing their infrastructure to take advantage of shared storage with parallel access in order to analyze more strategies faster and develop more effective trading and risk management platforms that can be deployed in less time.

With high performance data access solutions from DDN, Intel and Kx, proprietary trading firms, hedge funds and banks have seen up to 500% faster results than any other storage solution on the market, including SSD-direct attached storage or all flash arrays; 50% better latency characteristics on IO intensive workloads; and, unmatched bandwidth, latency and runtime for large scale workflows.

TIME BASED COLUMNAR DATABASE ACCELERATION WITH A DDN PARALLEL FILE SYSTEM & STORAGE APPLIANCES

ADVANTAGES OF RUNNING kdb+ OVER LUSTRE

DDN High performance parallel file system storage appliance offers several advantages over a single direct attached file system and when used in conjunction with the in-memory database kdb+, the advantages are:

1. A significant decrease in operational latency per kdb+ query, especially when running queries that search through large amounts of historical market information. This is achieved by balancing content around multiple file system servers.
2. Parallelization of kdb+ query “threads” in a single shared namespace, allowing a user to treat any data workload independently from other data workloads.
3. Support for simultaneous read/write operations on a single namespace for the entire database and for any number of kdb+ clients, (i.e.: end of day data consolidations into a Hierarchical Database (hdb) instance)
4. Shared data among different, independent hierarchical database (hdb)/ relational database (rdb) instances. Many instances of kdb can view the same data, meaning that strategies for data sharing and private data segments may be consolidated into the same space. This avoids the need for kdb+ administrators to physically copy data around the network or disks, as automatic space allocation balancing is built-in to the parallel file system.
5. Kdb+ context can be “striped” around all file system servers, or it can be allocated in a round-robin fashion against each server. Striping allows for some files to attain maximal I/O rates for a single kdb+ “object”.

QUALIFICATION OF kdb+ AND DDN EXAScaler® LUSTRE STORAGE

DDN, Kx and Intel collectively tested and qualified the Intel Enterprise Edition for Lustre 2.2 with Kx Systems kdb+ v3.2. The host operating system was CentOS 6.6. The storage infrastructure comprised an SFA12KX-40 based DDN EXAScaler solution. The physical topology of the system under test was kdb service (q is running directly on the kdb+ nodes) was distributed over 22 client nodes (kdb1...22).

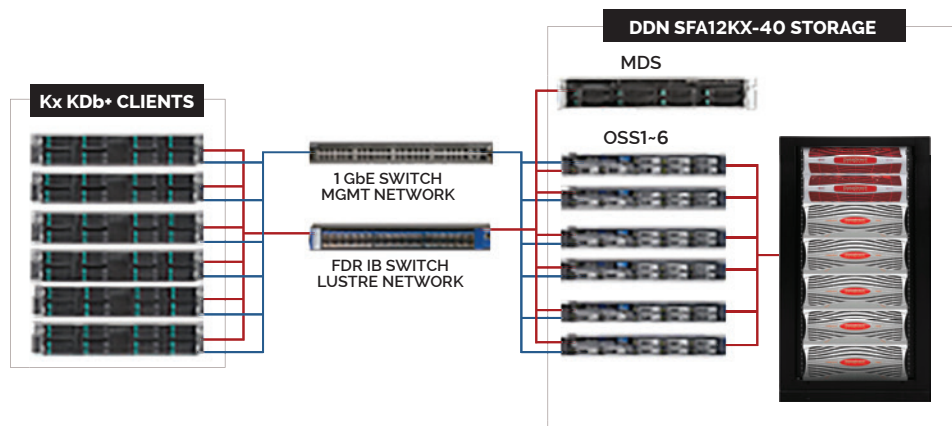
This setup was used to execute two commonly occurring market data workload patterns, a volume and user scaling use case (KANAGA) using data sets between 33TB to 897 TB, and a broad range of read & write operation scenarios (ANTUCO) using dataset sizes of 4.5 TB.

We loaded these kdb+ databases on top of the Lustre file system using a traditional TAQ-like trade/quote allocation – representing 12 months of trade/quote data history.

The purpose of this test was to measure the ingest rate of data from the storage using a query demanding a very large working set of data. Against this, a “year high bid” type of query for a large range of quote symbols was performed. The scale oriented benchmarks executed similar queries across five years of historical data.

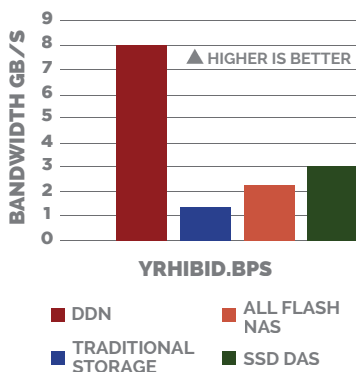
The DDN EXAScaler high performance parallel file system storage appliance excelled in each aspect of the benchmark testing by:

- Demonstrating 2.4x improvement in time sensitive & I/O intensive financial analytics (querying the maximum bid over the year for 1% of the symbols)
- Setting world record performance for large scale KANAGA benchmarks
- Proving 50 percent better latency characteristics on 13 of the most I/O intensive workloads versus previous benchmarks



CONCLUSIONS

DDN IS 2.7-6X HIGHER BANDWIDTH THAN AFA, SSD, DAS, & TRADITIONAL STORAGE



INDEPENDENT STAC AUDITED - WORLD RECORD PERFORMANCE FOR KDB+

Within a short span of two years, 40% of the largest financial services companies have deployed DDN solutions to address performance, scalability and TCO challenges in their organizations. These independently audited benchmarks show why - providing definitive quantification of DDN's unmatched performance for key data and scale intensive workflows in financial services.

These results show how storage systems powered by parallel file systems can greatly influence the benefits users can obtain from their server and software investments. Noteworthy attributes of the independently audited benchmarks include:

- The results demonstrate an effective use of a parallel file system as a performance at scale platform for a distributed query model based on kdb+ 3.2 database
- This benchmark was achieved with a very similar configuration of the SFA12K appliance as was done before (KDB121101), albeit with this validation we used lower cost industry standard near line SAS drives
- Even with the lower cost variant of disk drives being used, the infrastructure demonstrated an average 50% better performance (latency) characteristics for 13/17 workload patterns when compared to the previous benchmark

The STAC-M3 Kanaga suite simulates the size of data sets used in real production settings with up to 4 years' worth of historical data in this case. It also extrapolates the growth that many users are seeing in their tick data, making this a particularly useful study for users who need to assess their IT investments' performance over time. These test results show the importance of storage system scalability in achieving better than linear results (after adjusting for volume) as the data set sizes increased by more than 9x. Furthermore, the very small impact on random-access query times despite an increase in the queryable dataset by 4x, in an algorithm with significant compute intensity, shows how storage systems alone can greatly influence the benefits that users can obtain from their server and software investments. These results conclusively validate the extreme scale performance delivered using DDN and Intel Enterprise Edition for Lustre. Throughout this and all of the STAC-M3 benchmarks, the DDN solution has proven to give consistent and predictable results for each of the individual tests.

This further demonstrates DDN's capability of achieving industry leading results using a high performance file system. The benefit of this approach is it enables a shared data model, empowering both consolidation of data and massively parallel data queries in market data analytics, delivering optimized CAPEX and OPEX without compromising on performance.

REAL WORLD RESULTS AND BENEFITS

Leading hedge funds, proprietary trading firms and other financial institutions have been leveraging parallel file systems, like Lustre, to deliver several important advantages when compared to a single direct attached traditional file system. The primary advantages behind parallel storage are sustained high performance and the ability to easily scale upward to support larger workloads. When used in conjunction with in-memory databases like kdb+, some examples of the advantages are:

- Up to 5x improvement in algorithm development speeds
- Scaling storage IO performance linearly or near-linearly as Kx servers are added
- Shared access to large volumes of data over multiple, internal teams
- Eliminate data silos and simplify data management infrastructure
- Minimizing datacenter footprint and TCO and eliminating siloed infrastructure

REAL WORLD RESULTS AND BENEFITS

For more information about DDN, Intel and Kx joint solutions visit ddn.com or intel.com/Lustre.

ACKNOWLEDGEMENTS

The storage equipment and parallel file system consulting support for the benchmark was from DDN; kdb+ support was directly from Kx systems; server infrastructure, networking and Lustre technical support was supplied by Intel's High Performance Data Division.

ABOUT DDN®

DataDirect Networks (DDN) is the world's leading big data storage supplier to data-intensive, global organizations. For more than 15 years, DDN has designed, developed, deployed and optimized systems, software and solutions that enable enterprises, service providers, universities and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud. Organizations leverage the power of DDN technology and the deep technical expertise of its team to capture, store, process, analyze, collaborate and distribute data, information and content at largest scale in the most efficient, reliable and cost effective manner. DDN customers include many of the world's leading financial services firms and banks, healthcare and life science organizations, manufacturing and energy companies, government and research facilities, and web and cloud service providers. For more information, visit our website www.ddn.com or call 1-800-837-2298.