KX Improved Performance for KDB+ HDB in AWS

kx

> **Matrix offers kdb+ a combination of good read performance and metadata operational latency, being one or two orders of magnitude better than EFS, storage gateways and all open-source products we tested.**
>
> *Glenn Wright, Systems Architect at Kx*

USE CASE
Kdb+ in the AWS Cloud

COMPANY
Kx Systems

BUSINESS CHALLENGES
The need for on-demand or instant storage capacity scaling, burst processing of the compute load, and data analytics processing without the need to provide extra hardware in the data center.

SOLUTION
WekaIO Matrix™ on AWS

RESULTS
- Suitable model for an on-demand or a spot-pricing service
- Reduce overall costs for hosting the market data pool
- Flex to the desired performance levels
- Drive warm backups of data from in-house to EC2, or across instances/regions in EC2 – spin up for backups, then shut down

Kdb+ is a high-performance technology, and users have been slow to adopt the Cloud because it was thought that kdb+ could not provide a level of performance, storage and memory access commensurate with dedicated or custom hardware implementations. Kx and WekaIO have extensively assessed the functional performance constraints of kdb+ both in EC2 and in the supporting storage layers. If you are currently migrating, or considering migrating an historical kdb+ database (HDB) to the Cloud, you will want to read this case study which assesses the WekaIO Matrix storage solution available within the Amazon Web Services (AWS) Cloud.

THE CHALLENGE

Three key areas which should be considered when migrating kdb+ HDB and analytics workloads to the Amazon Elastic Compute Cloud (EC2) are:

- Performance and functionality attributes expected from using kdb+, and the associated HDB, in EC2
- The capabilities of available storage solutions working in the EC2 environment
- Performance attributes of EC2, and benchmark results

It should be noted that kdb+ can be successfully migrated to AWS EC2 with WekaIO Matrix regardless if the use case is kdb+ on demand or for those who are starting out on kdb+ for the first time and are choosing to host it in the Cloud.

THE SOLUTION: WekaIO MATRIX ON AWS

WekaIO Matrix can be used to store HDB data on AWS and it is fully compliant with kdb+. With Matrix, metadata operational latency is one or two orders of magnitude better than EFS, Storage Gateway, and all of the open-source products. In fact, Matrix exhibits block-like low operational latencies for some metadata functions, and with kdb+, it delivers good aggregate throughputs for small random reads.

With Matrix, kdb+ can be run in one of two ways. Firstly, it can run on the server nodes of the Matrix cluster, sharing the same namespace and same compute components. This eliminates the need to create an independent file system infrastructure under EC2. Secondly, the kdb+ clients can run on clients of the Matrix cluster, the client/server protocol elements being included as part of the Matrix solution and are installed on both server and client nodes. WekaIO tiers its namespace to S3, and it includes operator selectable tiering rules, that are based on age of file and time in cache. The performance is at its best when running from the cluster's erasure-coded SSD tier, exhibiting good metadata operational latency. Matrix requires server and client nodes to dedicate one or more cores

(vCPU) to the file system function. These dedicated cores run at 100% of capacity. When running kdb+ directly on the cluster, this fact is an important consideration when determining CPU core requirements for optimal performance.

THE RESULTS

WekaIO Matrix running kdb+ workloads on the AWS Cloud proves to be a suitable model for cost-effective on-demand, or instant storage capacity scaling and data analytics processing with impressive performance. Streaming reads running in concert across multiple nodes of the cluster achieve 4.6 GB/sec transfer rates, as measured across eight nodes running kdb+ on a single file system. Tests concluded that there was no decline in scaling rate between one and eight nodes. The tested cluster had twelve nodes, running a 4+2 data protection scheme across the nodes, each of instance type r3.8xlarge.

For the S3 tier, a single kdb+ thread on one node will stream reads at 555 MB/sec. (Figure 1) This rises to 1596 MB/sec across eight nodes, continuing to scale, although not linearly.

If you are considering moving your kdb+ workloads to the cloud, WekaIO's Matrix coupled with kdb+ in AWS represents a powerful platform. For a free trial of Matrix, go to https://start.weka.io/. To read the Kx white paper on this topic, go to https://code.kx.com/q/cloud/aws/app-i-wekaio-matrix/
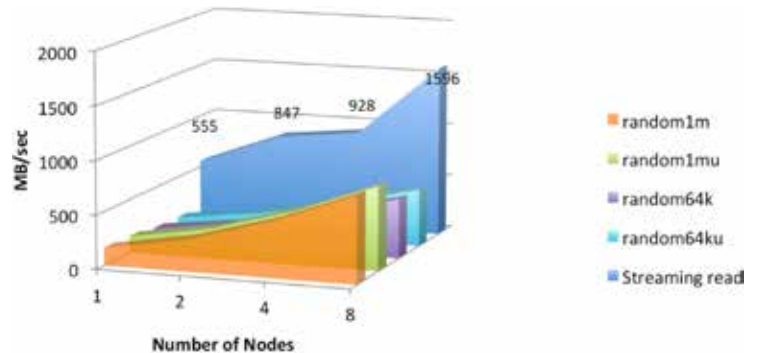

Figure 1. Performance results when forcing the 8xr3 cluster running Matrix on multiple kdb+ nodes to read from the data expired to S3.

About Kx and FD

Kx is a division of FD, a global technology provider with 20 years of experience working with some of the world's largest finance, technology, retail, pharma, manufacturing and energy institutions. Kx technology, incorporating the kdb+ time-series database, is a leader in high-performance, in- memory computing, streaming analytics and operational intelligence. Kx delivers the best possible performance and flexibility for high-volume, data-intensive analytics and applications across multiple industries. For more information about Kx please visit www.kx.com.

About AWS

For 10 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers more than 90 fully-featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 42 Availability Zones (AZs) across 16 geographic regions in the U.S., Australia, Brazil, Canada, China, Germany, India, Ireland, Japan, Korea, Singapore and the UK. AWS services are trusted by millions of active customers around the world monthly — including the fastest growing startups, largest enterprises, and leading government agencies — to power its infrastructure, make it more agile and lower costs. Learn more about AWS, visit aws.amazon.com.