The Future Value of Time

STAC Summits – Spring 2022

Joe Steiner
CTO

DELLTechnologies

My Kids: Sam / Jack

Their Future?

The Value of Time!

D❮LL Technologies

My Other Kids: PowerScale / ECS

Their Future!

The Value of Time?

D&LL Technologies

Isilon / PowerScale / ECS

High Performance Big Data Tick Analytics

Endless Scale

Extreme Concurrency
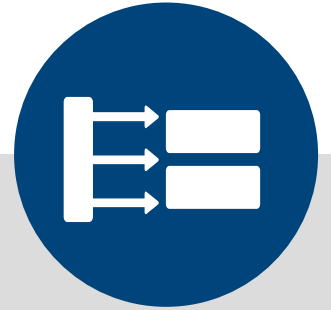
Simplified Architecture

Shorter Model Development Cyles

Score Models Faster

Deeper Analysis

Improved Analyst Agility

Lower TCO

DELLTechnologies

The Ideal Platforms for Tick Data Analytics

PowerEdge R640
The Hot Edge

PowerScale F200

SUT ID KDB200914

Outperformed single-user NBBO (STAC-M3.β1.1T.NBBO.TIME)	Vs. KDB200401

PowerEdge R940
Massively Parallel IO Maximum Density

Isilon F800

SUT ID KDB190430

Outperformed in 15 of 16 mean-response time Kanaga benchmarks	Vs. KDB150528

PowerEdge R640
Extreme Performance
PowerScale F900

SUT ID KDB210929

Outperformed F800 in 14 of 17 mean-response time Antuco benchmarks	Vs. KDB190430

- **High Performance Edge for Data Subsets**   • **Massive Parallel IO / Maximum Density**   • **Extreme Performance for the Entire Data Set**

STAC Benchmark: What does it mean?

The results showed the sweet spot for PowerScale based solutions to deliver real-time performance on smaller data sets (<10TB) and near real-time performance on large data sets (>10 TB) at high concurrency (100s into the millions).

DELLTechnologies

Performance Dynamics: 'The Hot Edge'

"All Data"
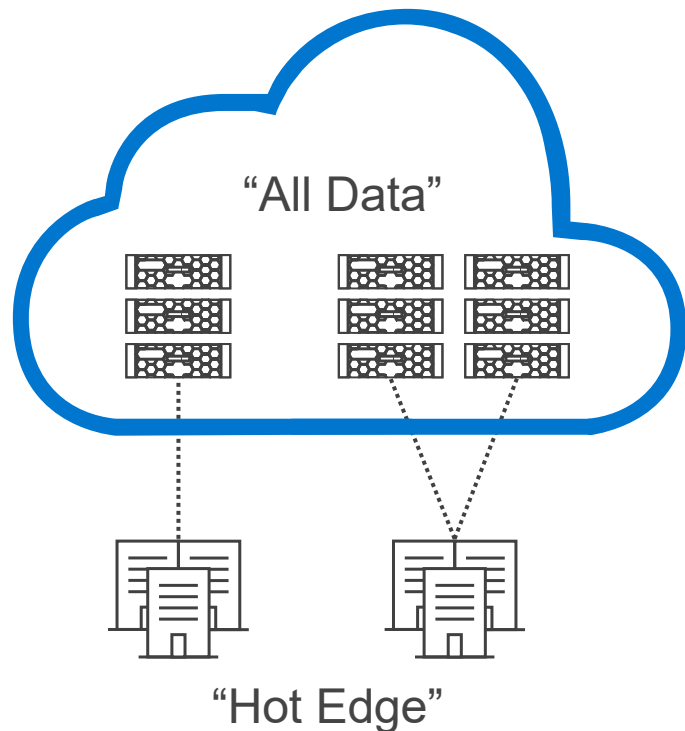
"Hot Edge"

- Deeper and richer data sets are being kept for longer. This is an ever increasing trend with longer history and richer data.

- Subset environments (or 'hot edge') are being created with smaller data sets (~1-2PB) with the large performance characteristics (e.g., TB/sec network throughput) and scalability (millions of open files)

- The 'hot edge' can be in the cloud or in data centers

- Reason for Dell's F200 STAC M3 Announcement was (1) increasing interest in high performance bandwidth with limited storage capacity, and (2) a reasonable cost entry solution with exceptional $ / GB of bandwidth ratio on your premise or in the Cloud. Google offering with MPS connections to GCP Compute resources

D&**LL**Technologies

Performance Dynamics: 'Extreme Performance'

"All Data"

SUT ID = KDB210929

"Extreme Performance"

The stack under test was KX's kdb+ 4.0 DBMS running on 9 Dell EMC PowerEdge R640 servers, each with 2 x 18 core Intel® Xeon® Gold 6254 CPU @ 3.10GHz and 384GiB DRAM. Data was accessed via NFS 3 and stored in a Dell EMC PowerScale F900 All-Flash Scale-Out NAS 3-node cluster with 251TiB total physical storage capacity. Dell chose to highlight that this solution was:

• Faster than a solution involving 4 database servers accessing an earlier generation of Dell EMC's flash storage appliance with kdb+ 3.6 (SUT ID KDB190430) in 14 of 17 mean-response time STAC-M3 Antuco benchmarks, including:

- 5.3x speed-up in the 10-user market snapshot (STAC-M3.β1.10T.MKTSNAP.TIME)
- 4.9x speed-up in the 10-user volume curve (STAC-M3.β1.10T.VOLCURV.TIME)
- 2.9x speed-up in the 100-user 12-day VWAB (STAC-M3.v1.100T.VWAB-12D-NO.TIME)
- 2.7x speed-up in the 100-user unpredictable interval stats (STAC-M3.ß1.100T.STATS-UI.TIME)

• Faster than a solution involving a parallel filesystem with 14 database servers and 18 storage servers with kdb+3.6 (SUT ID KDB200401) in 3 of 17 mean-response time STAC-M3 Antuco benchmarks:

- 9.8x speed-up in the single-user NBBO (STAC-M3.β1.1T.NBBO.TIME)
- 3.6x speed-up in the 10-user volume curve (STAC-M3.β1.10T.VOLCURV.TIME)
- 17% faster in the single-user VWAB (STAC-M3.v1.1T.VWAB-D.TIME)

D&ll Technologies

Solutions built for the modern digital organization

Unstructured data solutions

Enable innovation anywhere

Deliver performance & scale

Eliminate burden on end users & simplify their workflows

Cost effective & sustainable storage

Meet the broadest use case needs

Provide coverage across edge, core & cloud

D∉LL Technologies

And simplifies your essential data needs

Match storage costs to your data's business value

Intelligent, automated data access, protection & retention

Policy-based, automated tiering

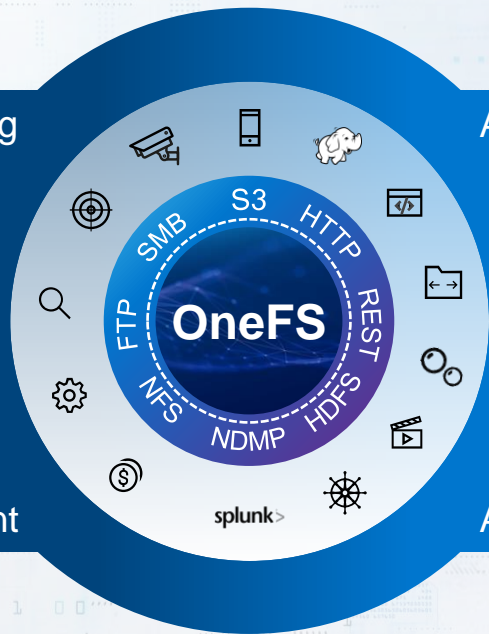Asynchronous replication for DR

Flexible data reduction options

Fast backup and recovery

Cloud tiering, on or off-prem

Policy-based compliance

Granular quota management

Automated load balancing

SMB   S3   HTTP
FTP   **OneFS**   REST
NFS   NDMP   HDFS

splunk>

Any client ●━━━━━━● **Any user**