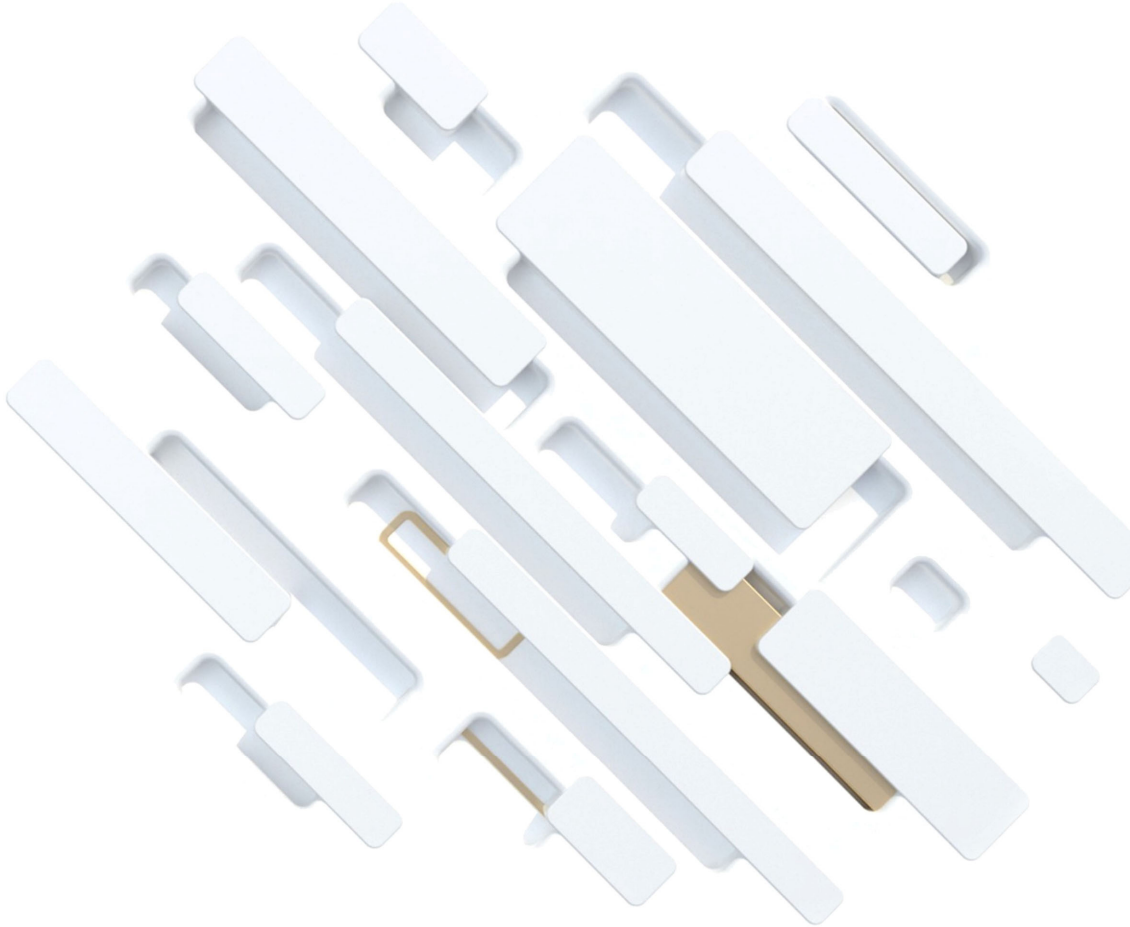
WEKA

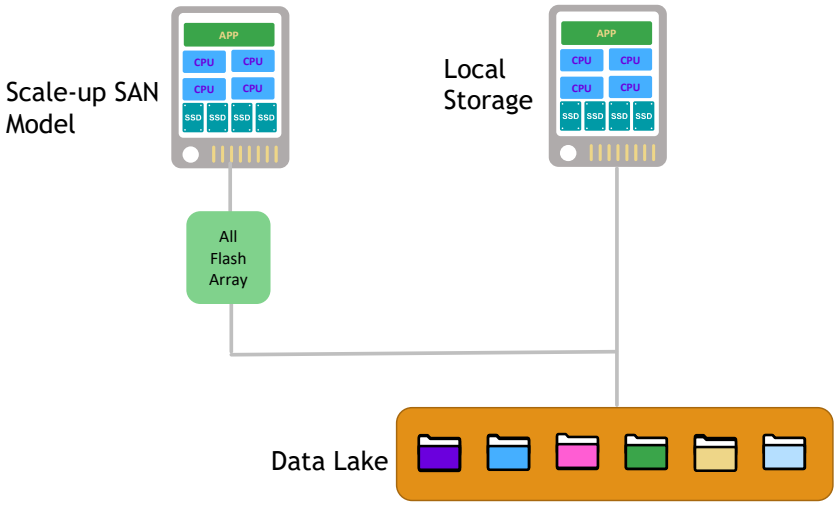
WekaFS™ For Financial Analytics

Barbara Murphy
VP of Marketing



Latency is the Enemy of Technical Trading

What you win with a local FS, you lose with data copy



- Copy data to SAN storage with limited scale
- Copy data to beefy server with lots of local NVMe
  - Data sets are limited to the size of the application server storage
  - Requires very expensive ”beefy” servers
- Adds wall clock time to the overall project
- Nightmare to manage at scale
- Users start “hogging” machines

Weka is an Enterprise POSIX Parallel File System

Introducing WekaFS – the Weka File System	Problems solved With WekaFS
Scalable NVMe-based, high-performance storage • Shared file system with the cache coherency of DAS and SAN • Effortlessly presents petascale data set to applications • Fully saturates compute resources for great efficiency • Seamlessly handles large and small files • Fully cloud enabled for hybrid or public cloud deployments	Faster wall clock time Simpler to manage at scale No performance tuning required Lowers the overall cost of infrastructure Ability to run more on-demand models
Integrated object-based data lake with seamless movement of data between hot and cold data	Best cost for multi-petabyte data sets Data is always available to the applications Models can source far more market indicators

IO500

This is the official list from Supercomputing 2019. The list shows the best result for a given combination of system/institution/filesystem.

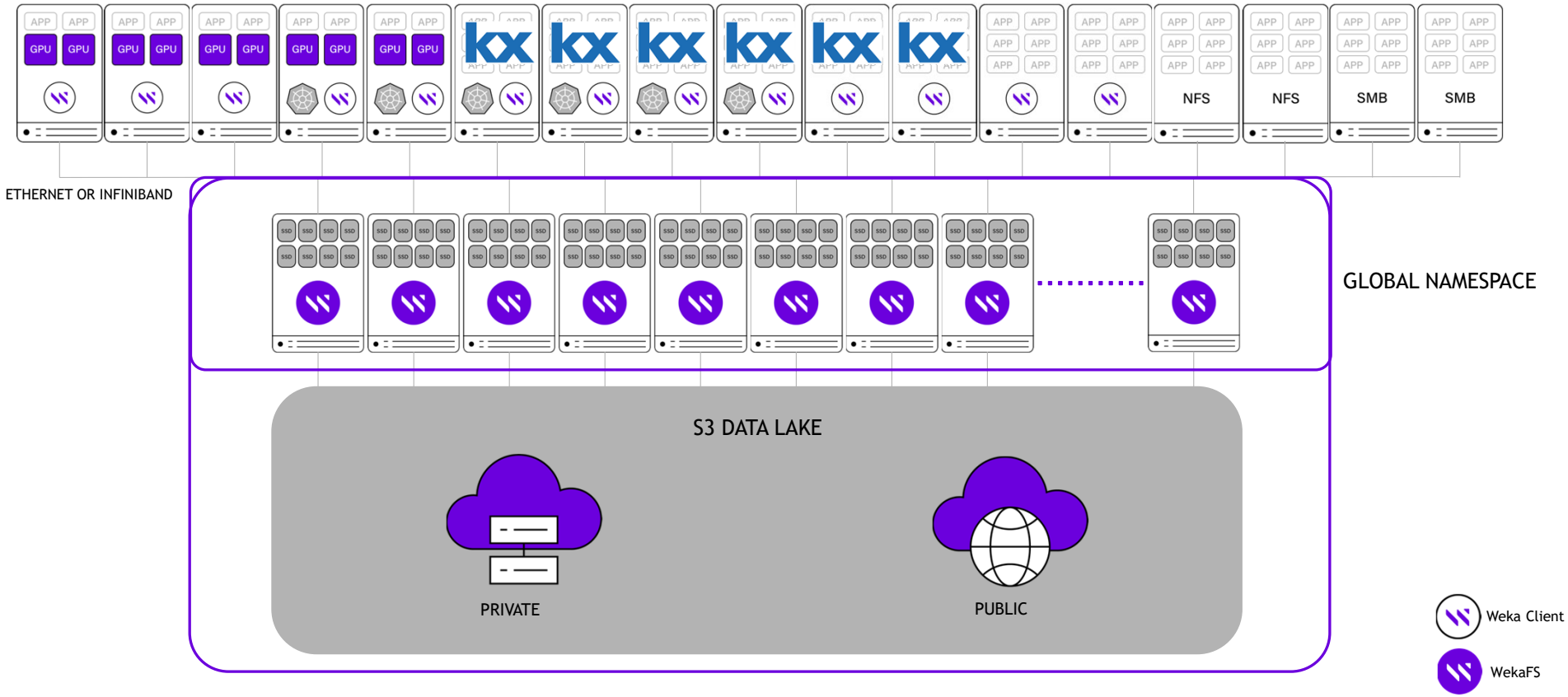
Please see also the 10 node challenge ranked list.

https://www.vi4io.org/std/io500/start

#	information								io500		
	list id	institution	system	storage vendor	filesystem type	client nodes	client total procs	data	score	bw GiB/s	md kIOP/s
1	sc19	WekaIO	WekaIO on AWS	WekaIO	WekaIO Matrix	345	8625	zip	938.95	174.74	5045.33
2	sc19	Intel	Wolf	Intel	DAOS	26	728	zip	933.64	183.36	4753.79
3	sc19	National Supercomputing Center in Changsha	Tianhe-2E	National University of Defense Technology	Lustre	480	5280	zip	453.68	209.43	982.78
4	sc19	NVIDIA	DGX-2H SuperPOD	DDN	Lustre	10	400	zip	249.50	86.97	715.76
5	sc19	University of Cambridge	Data Accelerator	Dell EMC	Lustre	128	2048	zip	229.45	131.25	401.13
6	sc19	CEA	Tera-1000	DDN	Lustre	128	4096	zip	210.26	81.01	545.74

#1 on Overall Score
#1 on Metadata

The Weka File System in a Production Environment



Customer Success

Quantitative Trader	Problem: Needed DAS performance but kept burning out local server NVMe drives With Weka File System - 3x Faster than DAS and scales effortlessly
Quantitative Trader	Problem: Wanted to move market analysis with kdb+ to AWS but no solution was performant WekaFS on AWS provides higher performance than on-premises with easy dev-test environment
Major Bank	Problem: Have a single scale-up expensive Kx kdb+ server with everyone waiting in queue Weka File System enables scale-out of kdb+ time series analysis

What is STAC-M3 Benchmark

Independent audited tests for financial market use cases that use a time-series database

1 year (Antuco) data set

5 year (Kanaga) data sets

Varying number of clients to test concurrency

- 1 client
- 50 clients
- 100 clients

Read I/O intensive

https://www.weka.io/blog/what-is-the-stac-m-3-benchmark-and-why-should-you-care/

Weka on STAC-M3 Testing

- Set 17 New Records
- Beat our own prior 2019 records by up to 2x
  - Improvements in our read performance
- Clean Sweep of all STAC-M3 Bandwidth Records
  - Highlighting our Read performance

HPE + Weka Solution (KDB200401): Key products

- STAC-M3™ Packs for kdb+ Rev 3.0 Antuco * Kanaga, Compatibility Rev E
- kdb+ 3.6 running in distributed mode
- 14 x HPE ProLiant XL170r Gen10 database servers
- 18 x HPE ProLiant XL170r Gen10 storage cluster servers
- WekaIO WekaFS Storage Software Release v3.6.2
- Mellanox SB7790 36-port Non-blocking Externally-managed EDR 100Gb/s InfiniBand Switch

* Full STAC® Report report at: SUT ID KDB200401

Record Performance for Read Intensive Benchmarks

- Outperformed all publicly disclosed results in **11 of 24** Kanaga mean-response time (MRT) 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)
- Outperformed all publicly disclosed results in **5 of 5** Kanaga throughput benchmarks
  - STAC-M3.β1.1T.*.BPS

Compared to other solutions

Versus a kdb+ solution involving an all-flash NAS and 4 database nodes (SUT ID KDB190430):

- was faster in all **24 Kanaga MRT benchmarks**; and
- 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)
  - 4.5x speedup in 10-user aggregate stats (STAC-M3.β1.10T.STATS-AGG.TIME)

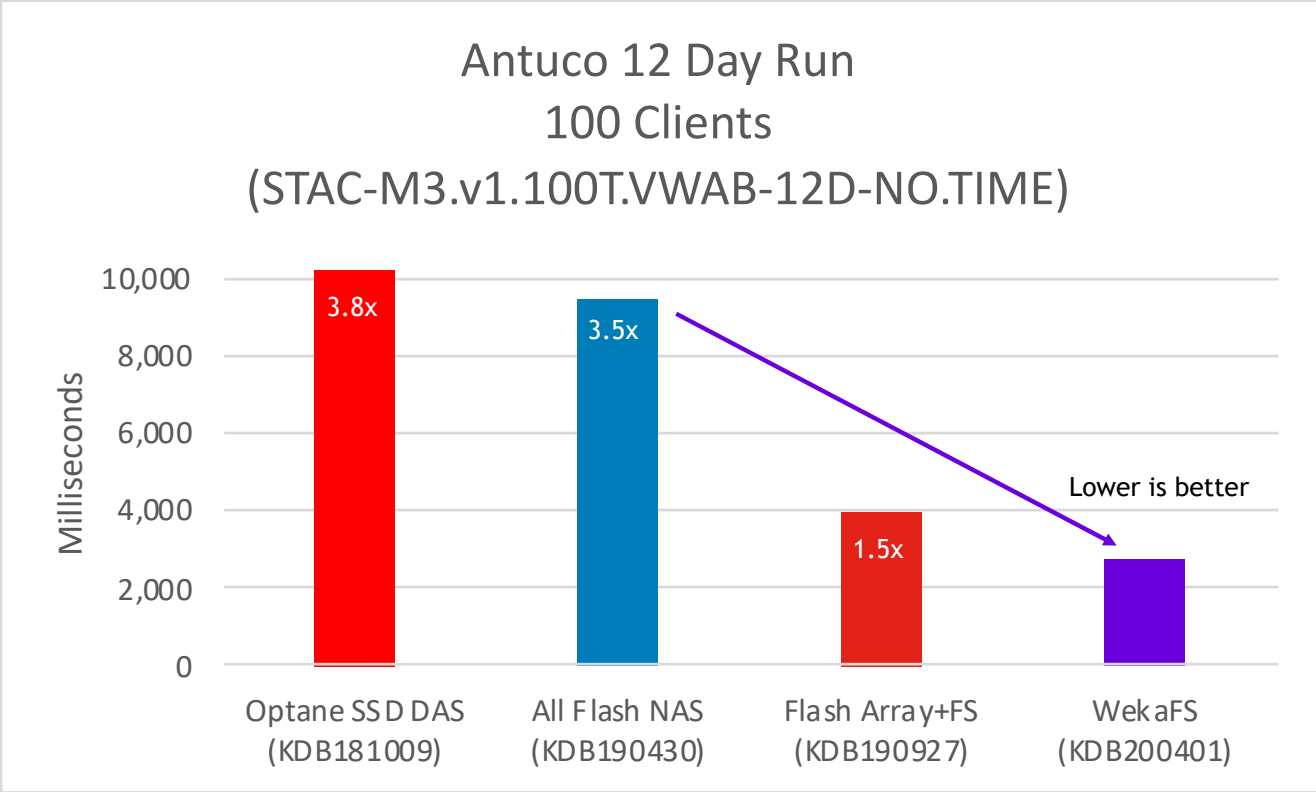
Versus a kdb+ solution involving a single server with direct-attached Intel Optane and 3D-NAND Flash SSD (SUT ID KDB181009):

- was faster in **19 of 24 Kanaga MRT benchmarks**, including:
  - 20.3x speedup in STAC-M3.β1.100T.YR2VWAB-12D-HO.TIME; and
- was faster in 4 of 17 MRT Antuco benchmarks.

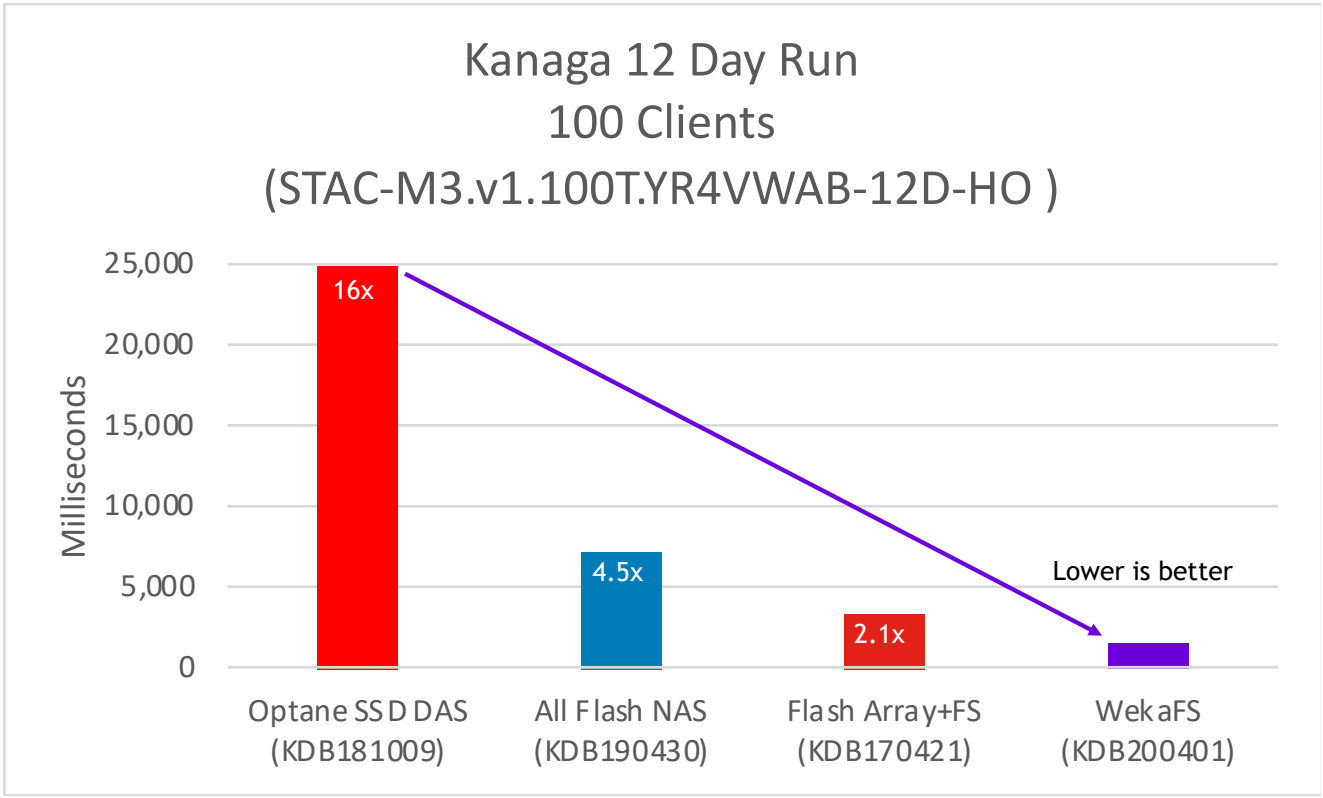
Versus a kdb+ solution involving a Fibre Channel-connected flash array and 4 database nodes (SUT ID KDB170421):

- was faster in **12 of 24 Kanaga MRT benchmarks**; and
- was faster in 3 of 17 Antuco MRT benchmarks.

Weka is over 3.5x better than DAS or All Flash NAS

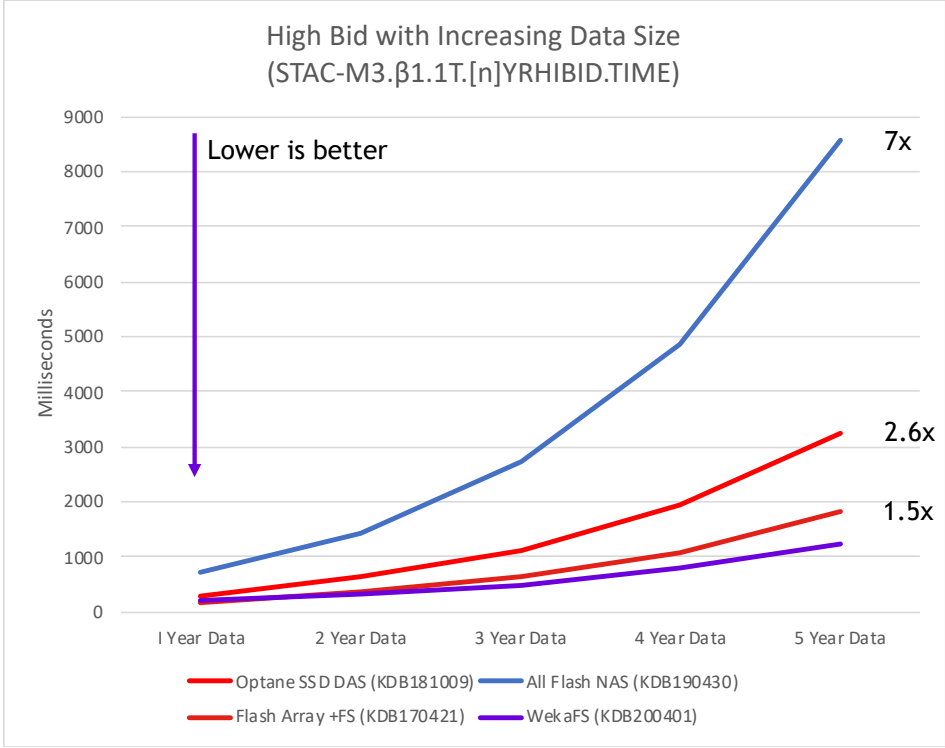


Weka Performs Best when Workload has High Concurrency



WekaFS gives users the ability to scale to high concurrency for small and large data

Weka Performs Best when Data Sets Increase



A shared file system can deliver better than what was previously only possible with DAS or SAN


WEKA