

STORAGE DEVELOPER CONFERENCE



*BY Developers FOR Developers*

Virtual Conference  
September 28-29, 2021



# FinTech Data Pipelines and Storage I/O Related Benchmarks in Public Cloud

Reduce Wall Clock Time for "Tick Analytics" with Weka Data Platform

Shailesh Manjrekar, Head of AI, Weka

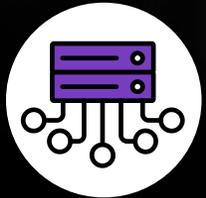
# THE PROBLEM



Exponential growth in modern Pharmaceutical\LS, AI/ML, Financial Services and cloud applications is fueling unheard-of data growth



Organizations desire competitive advantage through digitalization by 10x faster processing of 10x bigger datasets



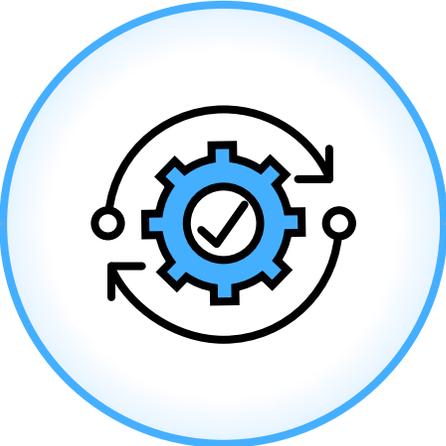
Traditional hardware and cloud storage cannot meet rapidly escalating requirements for *performance AND scale* across core, multi-cloud and edge data centers

---

**...A NEW APPROACH IS NEEDED.**

---

# SEPARATE PRODUCTS SERVICE CUSTOMER NEEDS



SIMPLICITY



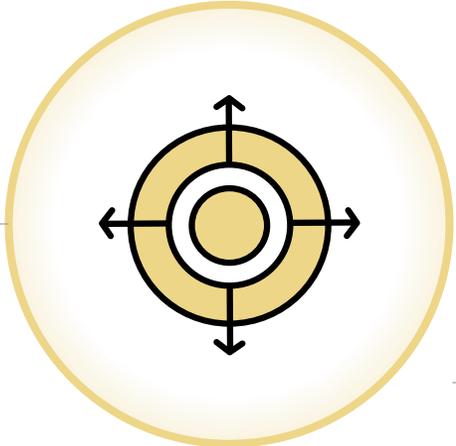
AMAZON EFS  
FSx for Windows  
FSx for Lustre



SPEED



AMAZON EBS

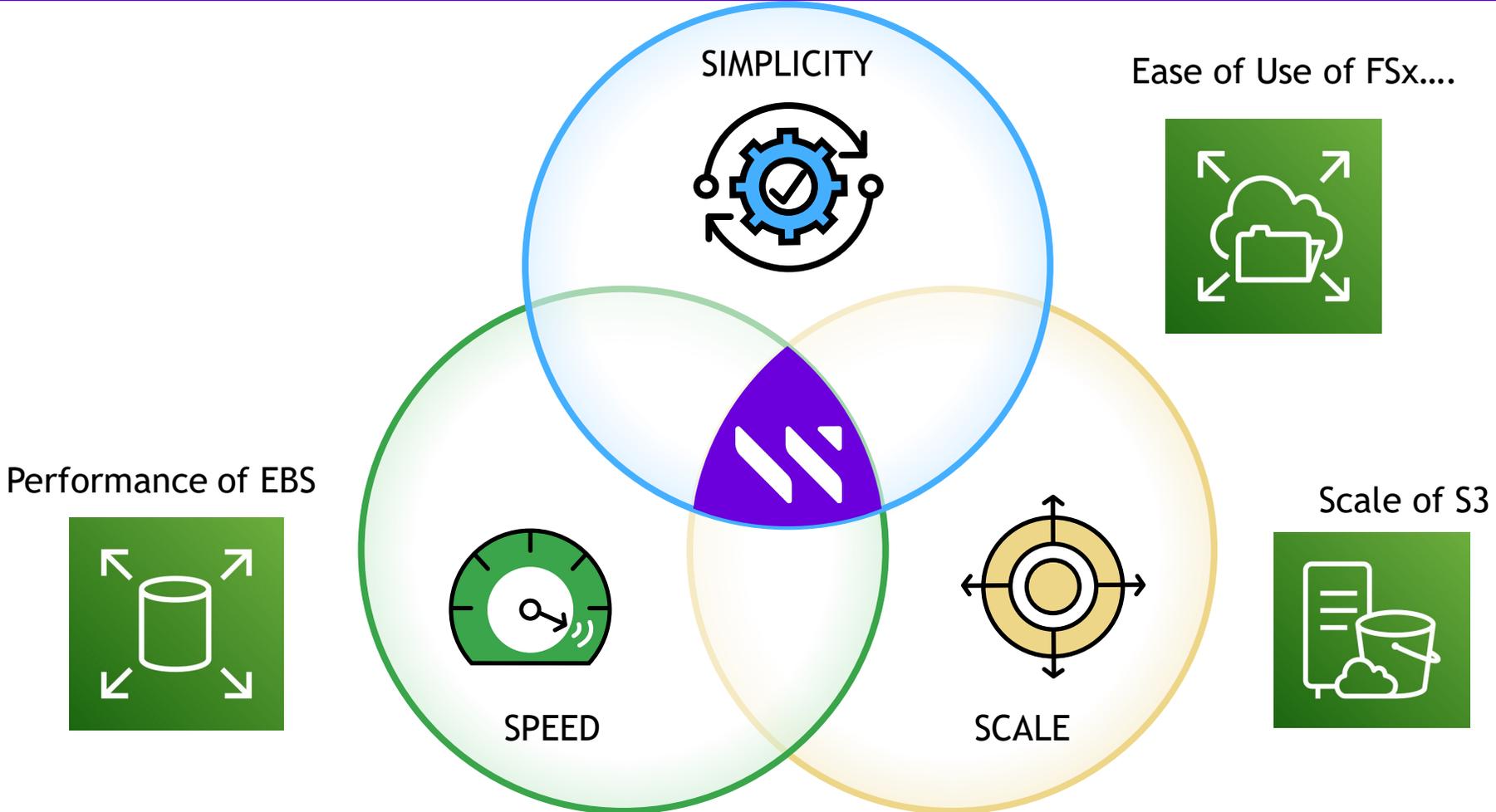


SCALE



AMAZON S3

# WEKA DATA PLATFORM FOR ALL WORKLOADS

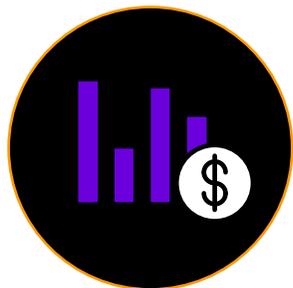


**FOR MODERN ENTERPRISE APPLICATIONS**

# Welcome to WEKA

## Our Mission

Make storage a utility by delivering simplicity, speed, scale, and better economics



8

of the Fortune 50  
are customers



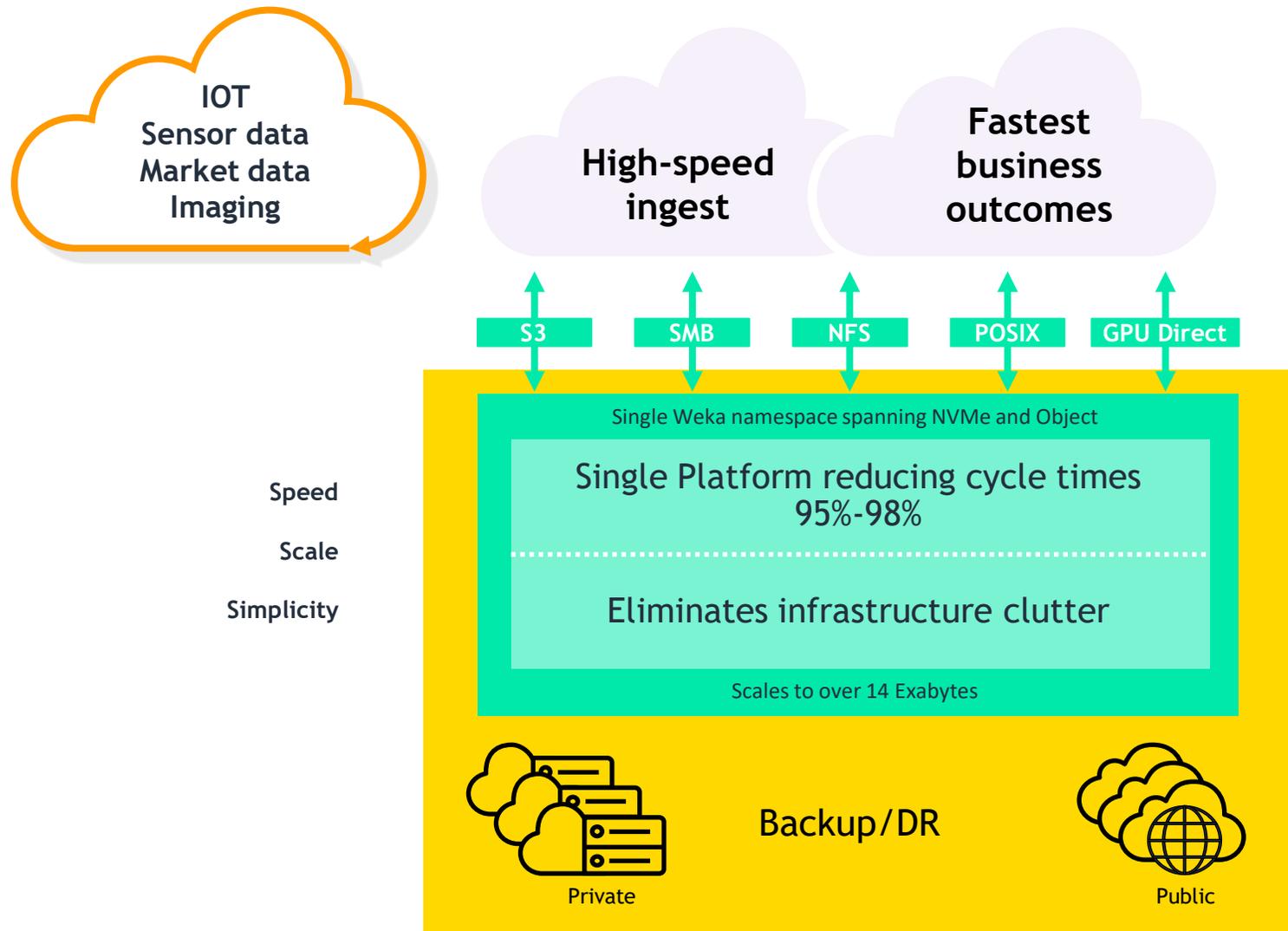
Advanced  
Technology  
Partner

Storage Competency

## Backed by industry leaders



# Solution: zero-copy architecture



## After Weka

1. Collapse into single system, reducing infrastructure costs by up to 75%
2. 95% - 98% reduction in time for data to be actionable (first to market advantage)
3. Immense scalability up to 14EB's in single namespace (no additional software needed)
4. 20x reduction in management overhead

Fastest outcomes

Beat competitors to market

Competitive advantage

# AWS + WEKA JOINT HYBRID OPPORTUNITY

- 230 PB Managed in AWS S3 per year across verticals
  - Financial Services
  - Life Sciences
  - Greenfield Machine Learning market
- Available in AWS Marketplace and AWS Outposts certified
  - SaaS listing - BYOL, Marketplace license
  - Cloud Formation front end

<https://start.weka.io/>



02

# STAC Audit Summary

# Who is STAC?

## Securities Technology Analysis Center

STAC® provides technology research and testing tools based on community-source standards

The standards are developed by the STAC Benchmark Council™, a group of 390 of the world's leading banks, hedge funds and financial services technology companies

### Several Research Domains -

A2 - Risk Computation, A3 - Backtesting, M3- Tick Analytics

### Why STAC-M3?

The **STAC-M3** Benchmark suite is the industry standard for testing solutions that enable high-speed analytics on time series data, such as tick-by-tick market data (aka "tick database" stacks)



# STAC-M3 Benchmarks

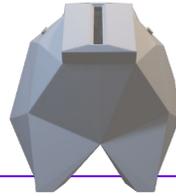
- defines parameters for testing and validates the results through careful audit
- Antucco
  - The baseline suite - Antuco - uses a limited dataset size with constraints to simulate performance against a full-size dataset residing mostly on non-volatile media. It tests a wide range of compute-bound and storage-bound operations to probe the strengths and weaknesses
- Kanaga
  - The scaling suite - Kanaga - uses a subset of Antuco queries without constraints against a significantly larger data set. The key metric for the STAC-M3 benchmark suite is query response time. Latency is the enemy of the finance market
- In total there are 17 mean response time benchmarks in the Antuco suite and 24 mean response time benchmarks in the Kanaga test suite
  - and they test the system in two dimensions - number of concurrent clients and quantity of data under analysis

# Where STAC-M3 matters?



## Risk Analytics

- Banks, Insurance, Exchanges, Retail
- Credit Risk
- Insurance Risk
- Fraud Detection
- Operational Risk



## Trading System

- Brokerages, Hedge Funds, Exchanges
- Quant Trading
- Tick Analytics
- Strategy Back-testing
- Trade Reporting and Processing

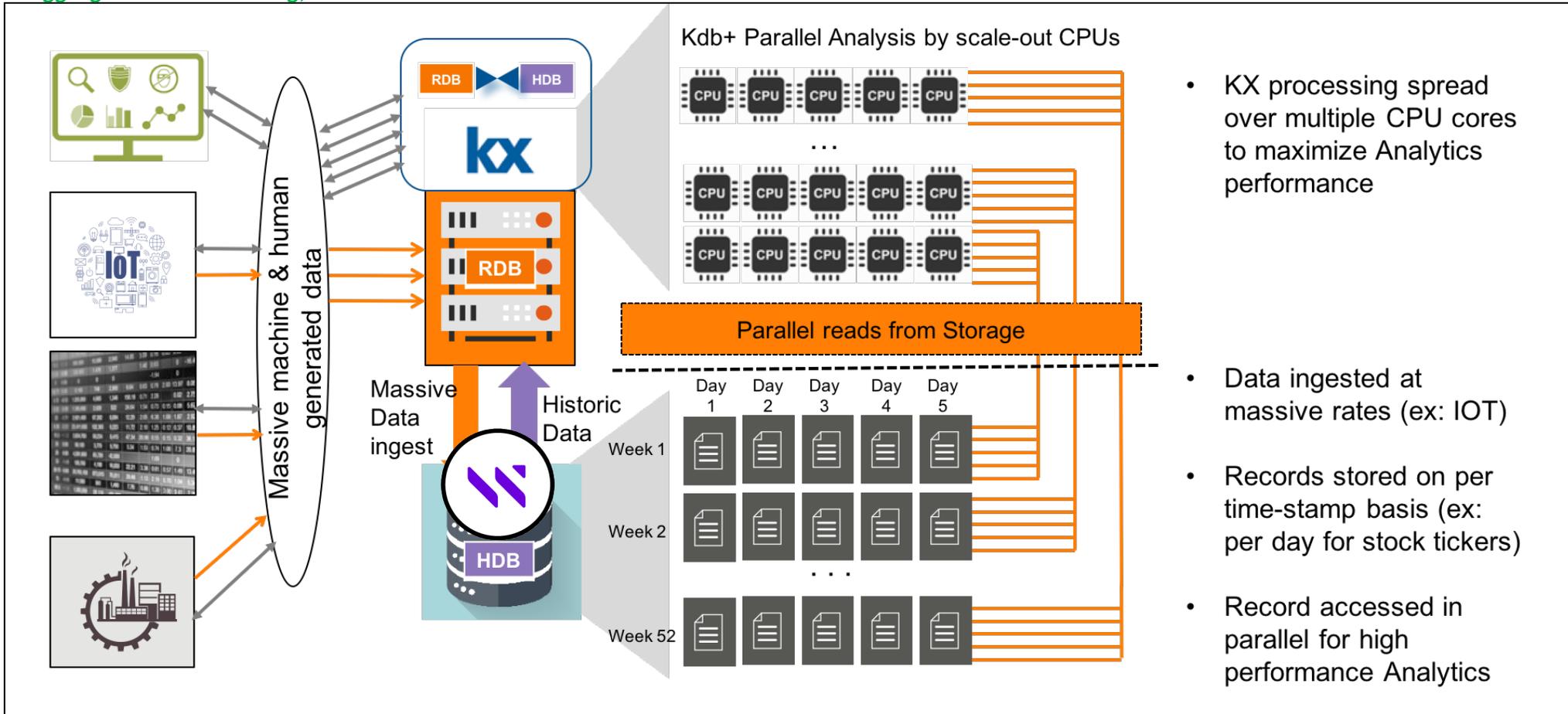


## Banking Systems

- Banks, Clearing Houses
- Electronic payment processing
- Time Bound / Compliance Reporting

# Kdb+ AND WEKA SOLUTION

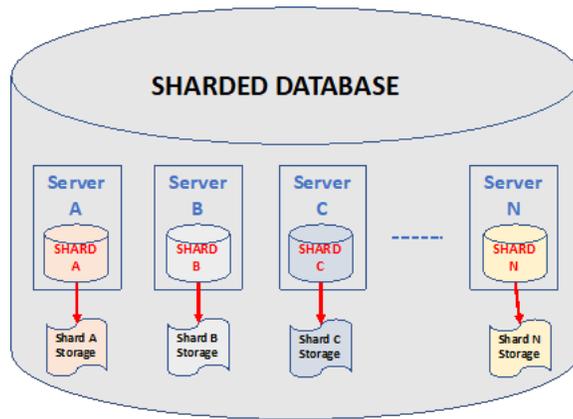
Data feeds - NYSE, NASDAQ  
 Aggregators - Bloomberg, Thomson Reuters



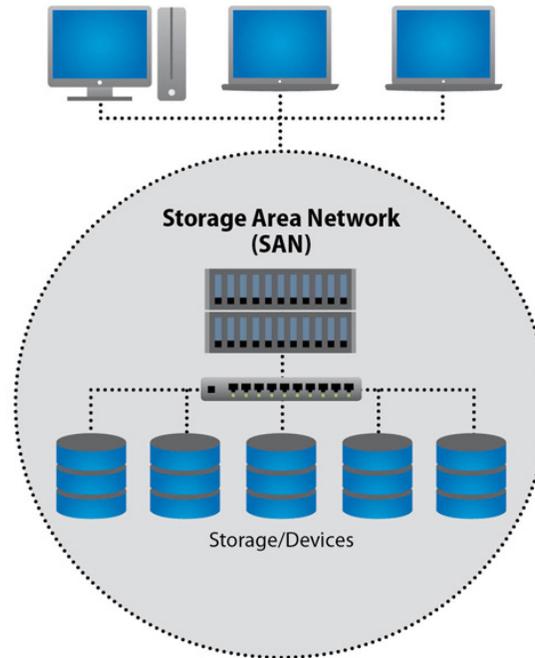
KX RDB - Real time database

KX HDB - Historical database hosted on WekaFS

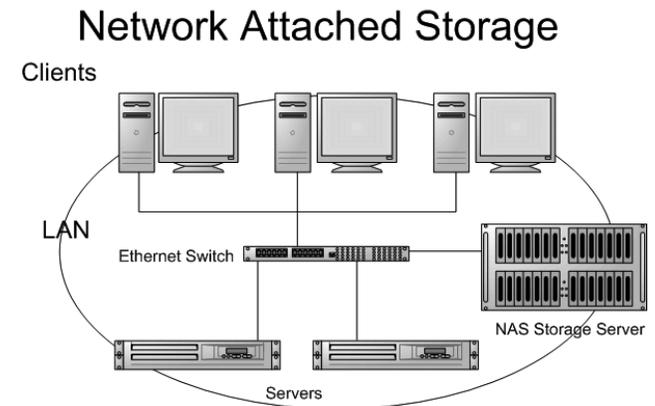
# ALTERNATIVE SOLUTIONS FOR TICK ANALYSIS



- Some queries which fall within a time interval could be served only by single individual server
- Query is not distributed across servers

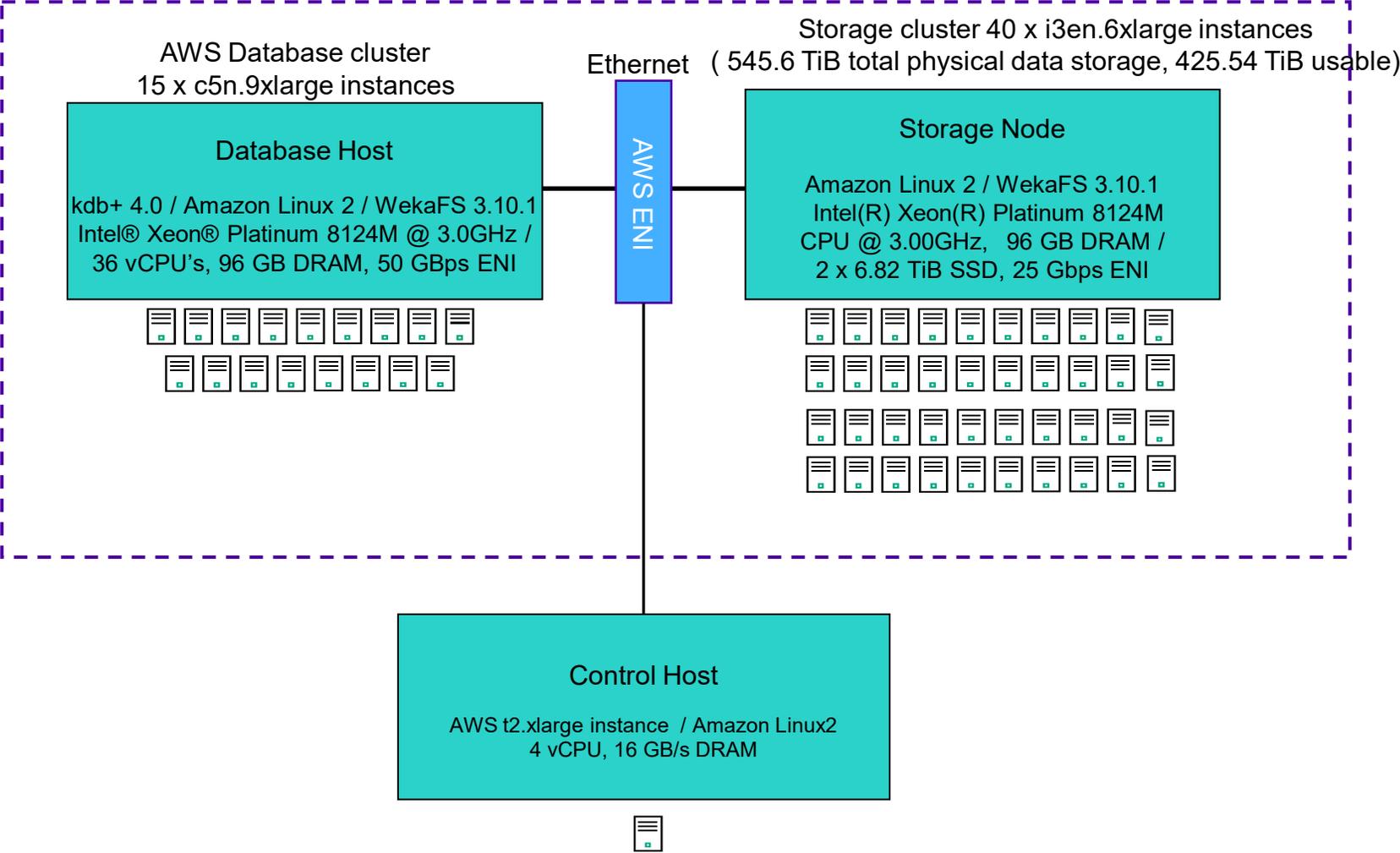


- Block storage with limited scale
- Copy data to beefy server with lots of local NVMe
  - Data sets are limited to the size of the application server storage
  - Multiple copies



- NFS based All-Flash OR
- Throughput optimized parallel filesystems
  - Poor latency and IOPs

# STAC-M3 AWS EC2 SUT – System under Test



1 x client (not part of the SUT)

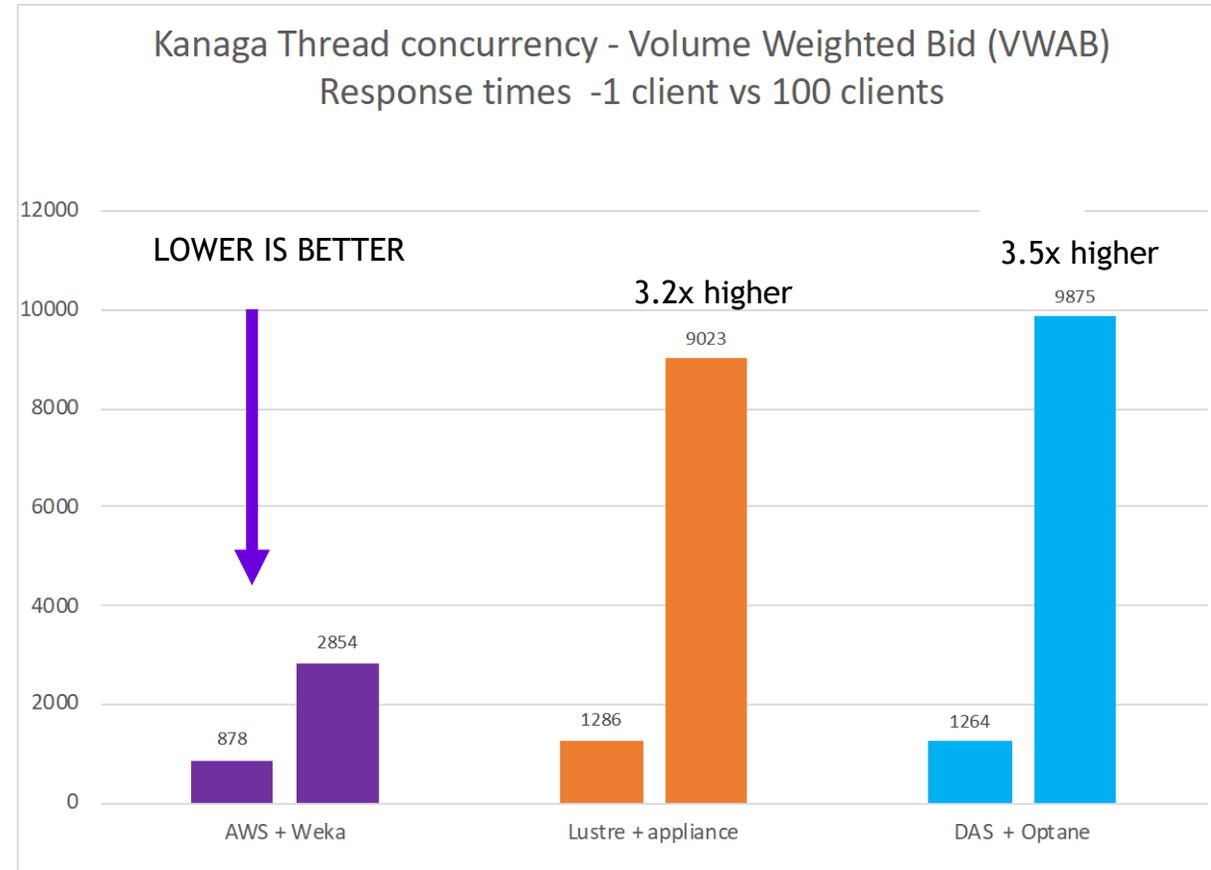
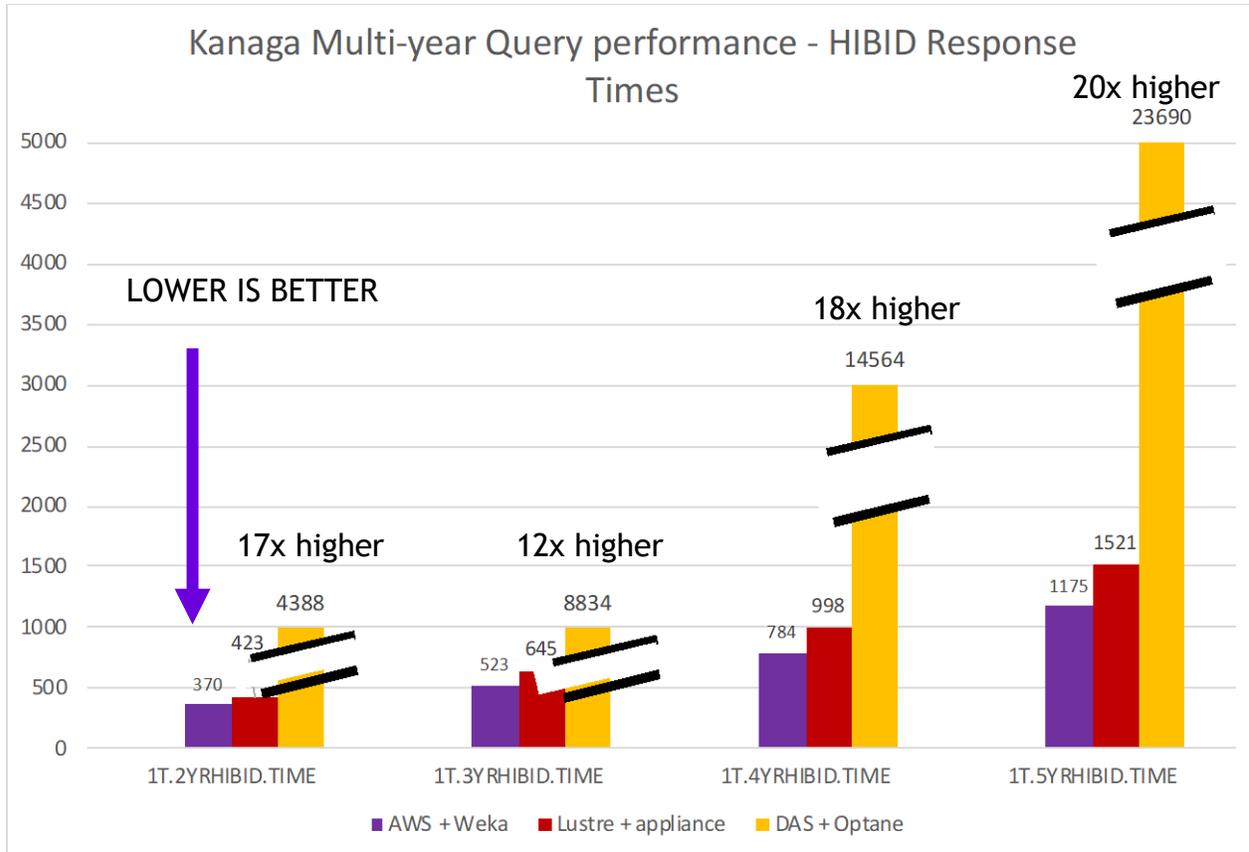
# 6 RECORDS ON STAC-M3 IN AWS – #KDB210507

- 3 RECORDS ON STAC-M3 Kanaga and 3 Records in throughput benchmarks - SUT #KDB210507
- WekaFS v3.10, was faster in several other benchmarks as below -

	STAC System Under Test #	WekaFSv3.10 in Kanaga benchmarks	WekaFSv3.10 in Antuco benchmarks
Lustre + Appliance	<i>#KDB200915</i>	Faster by 20 of 24	Faster in 4 of 17 benchmarks
Direct attached 10 servers with Optane	<i>#KDB200603</i>	Faster by 16 of 24	Faster in 9 of 17 benchmarks
All-Flash NAS	<i>#KDB200914</i>	All-Flash NAS did not submit to this benchmark	Faster in 15 of 17 benchmarks

# RECORD-BREAKING STAC-M3 (Kanaga)

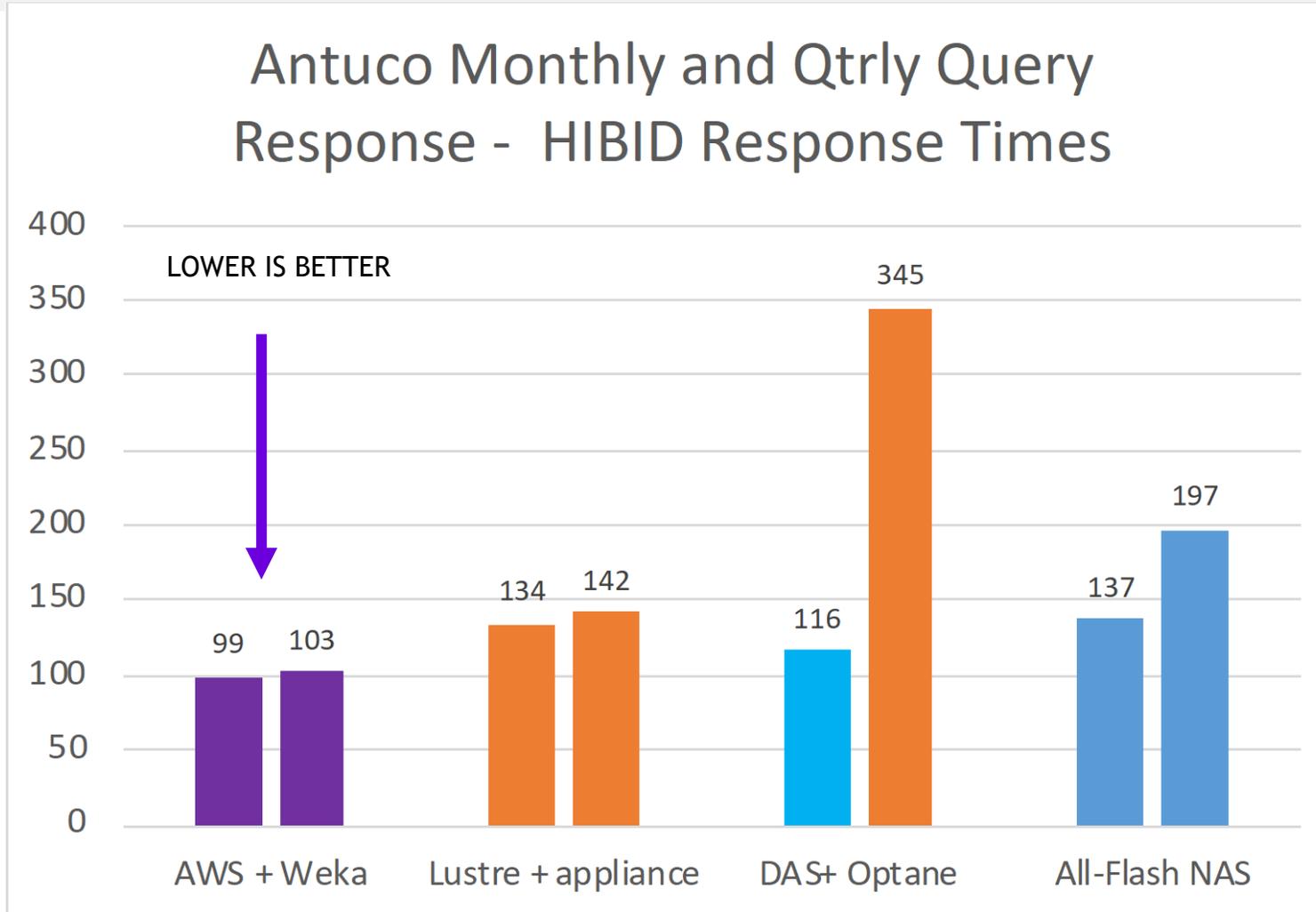
## PERFORMANCE IN AWS BEATS ON-PREMISES SOLUTIONS



- AWS + Weka used v3.10.1
- HPE + Weka SUT - KDB200401, AWS + Weka SUT - KDB210507, WEKA SUT - KDB210507
- Lustre + appliance SUT - KDB200915, DAS + Optane - 200603

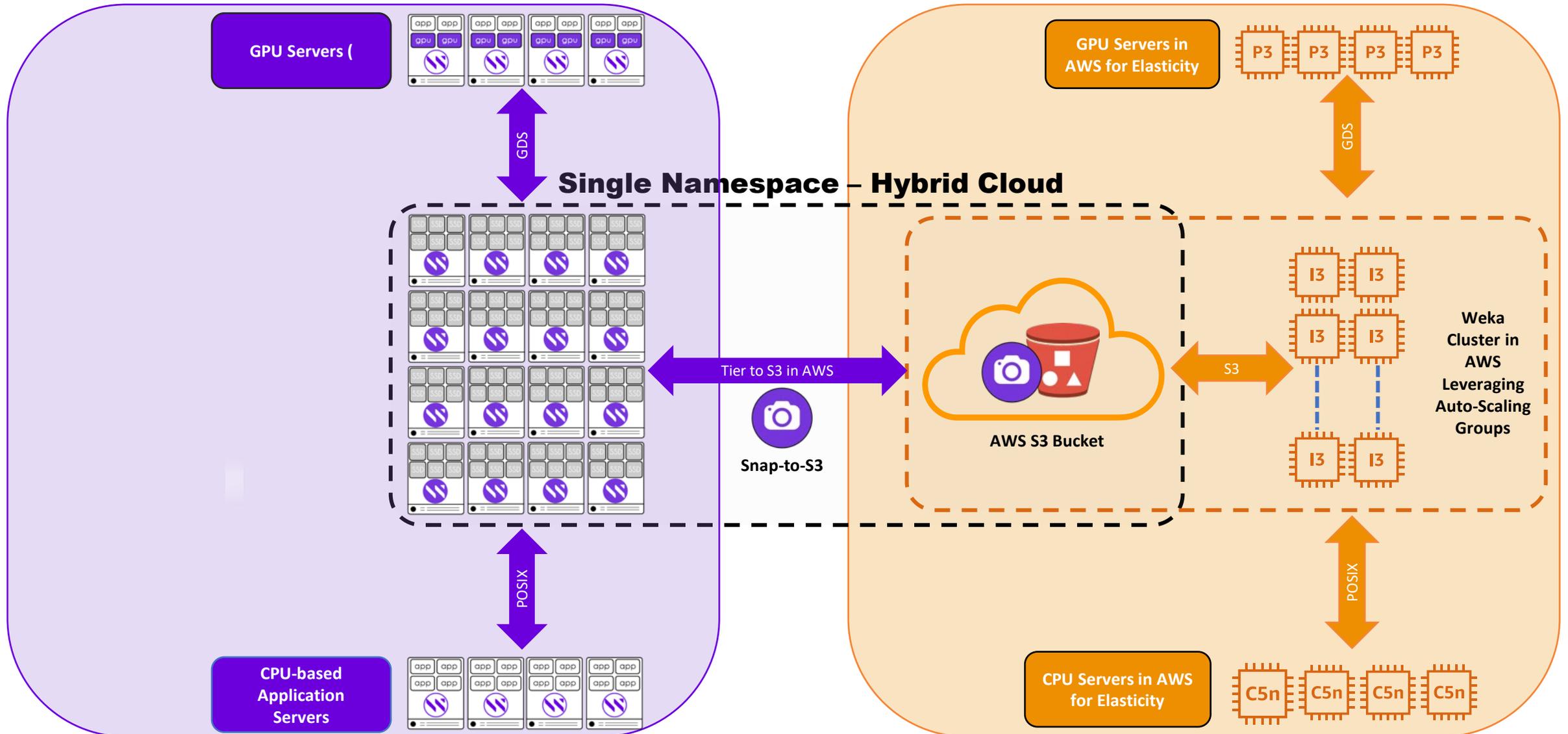
# RECORD-BREAKING STAC M3 (Antuco)

## PERFORMANCE IN AWS BEATS ON-PREMISES SOLUTIONS

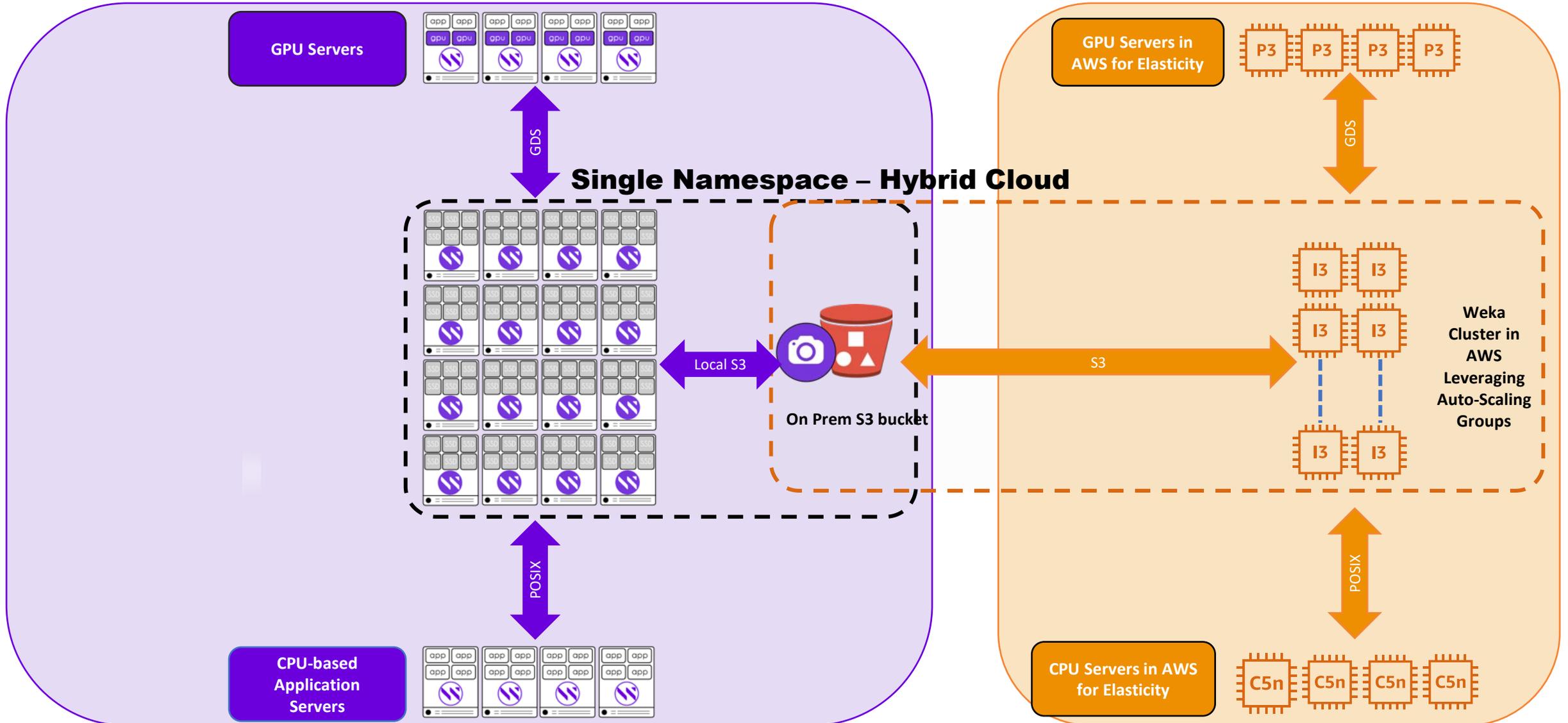


- AWS + Weka used v3.10.1
- HPE + Weka SUT - KDB200401, AWS + Weka SUT - KDB210507,
- Lustre + appliance SUT - KDB200915, DAS + Optane - 200603, All-Flash NAS - KDB200914

# Hybrid Cloudnative Workflow



# Hybrid Cloudnative Workflow





Please take a moment to rate this session.

Your feedback is important to us.