Backtesting in the Cloud

STAC Summit: New York

November 2017

For more information contact:

Matt Meinel

mmeinel@levyx.com

615-972-1259

STAC A-3 Results: Backtesting w/ Spark Dataframes on Levyx' Xenon on Google Cloud

- **Google Cloud 5 node cluster (4 workers) (STAC Audited SUT ID: LEVX170603)**
  - ✓ 32.7x improvement in speed compared to previous record on Hadoop Streaming from Intel/Cloudera (SUT ID INTC141220-VI) running on 14 Dell PowerEdge Servers

	INTC141220-VI	SPRK170603	Improvement
STAC-A3.ß1.SWEEP.MAX60	25	200	8x
STAC-A3.β1.SWEEP.SPEED	0.43	14.08	32.7x
Max instrument simulations/second	2.5	12.6	5.0x

- **Levyx STAC-A3 - Results on Amazon Cloud testing in the process of being released to the STAC Vault**

www.STACresearch.com/backtesting

Backtesting Everything …

- Trading Algos (like STAC-A3)
- Compliance Models
- Fraud detection models
- Marketing Models
- Operational Risk Models
- Cybersecurity Models

The popular toolset …

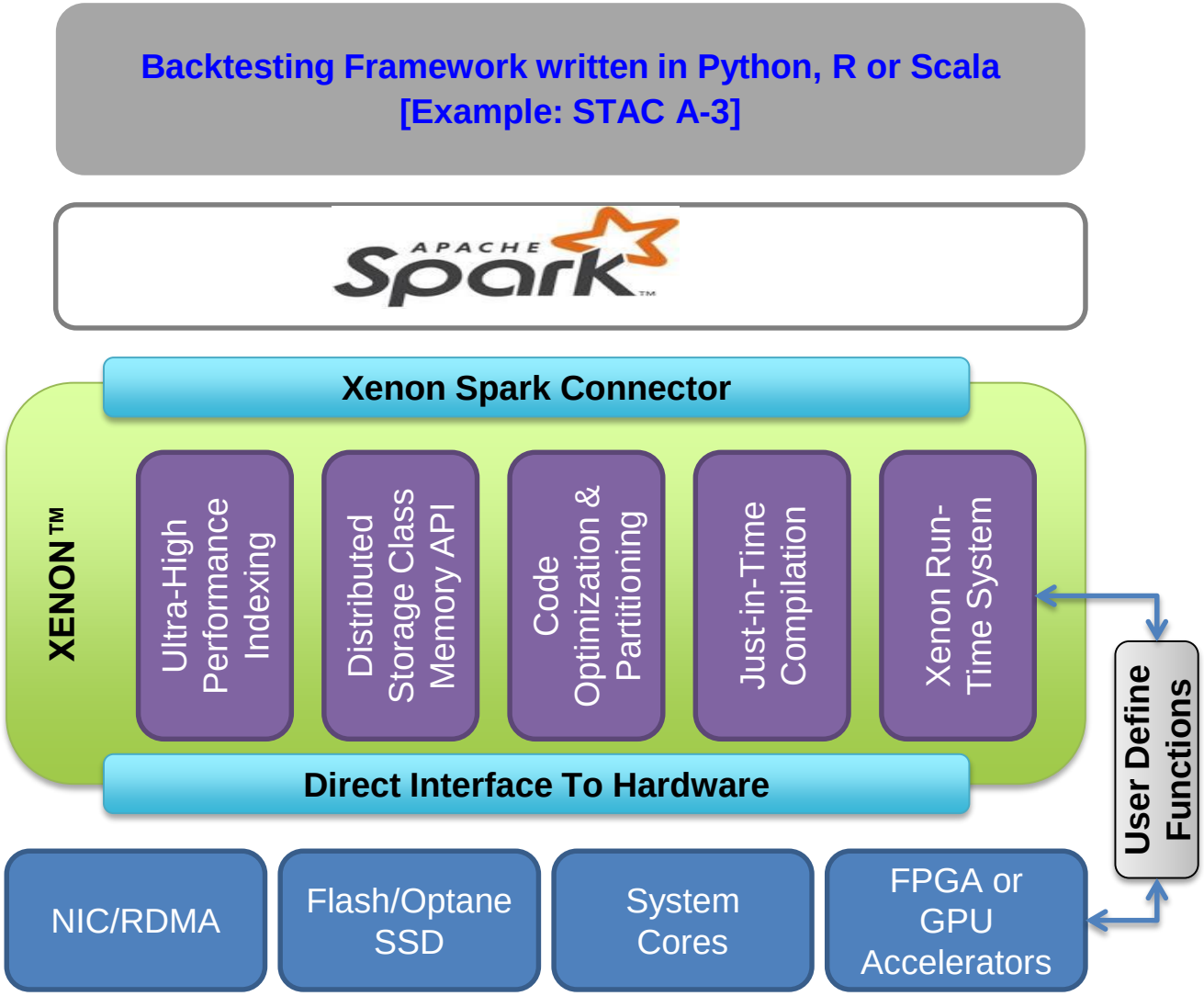
- Languages: Python (PANDAS and sci-kit) & R
- Dataframe Formats: Apache Arrow & Feather
- Compute Frameworks: TensorFlow, Apache Spark (SparkSQL, SparkML, SparkR, pySpark) and Hadoop
- Distributed Times Series DB and Tick Stores: Quasardb, TickSmith,
- Compute Offload: FPGA & GPU

To keep costs under control most Fintech firms are using cloud service providers for machine learning model exploration and training as well as for on-demand backtesting compute needs.

Most backtesting processes revolve around getting time series data into MEMORY as a DATAFRAME & then running parallel calculations to train and/or test the model.

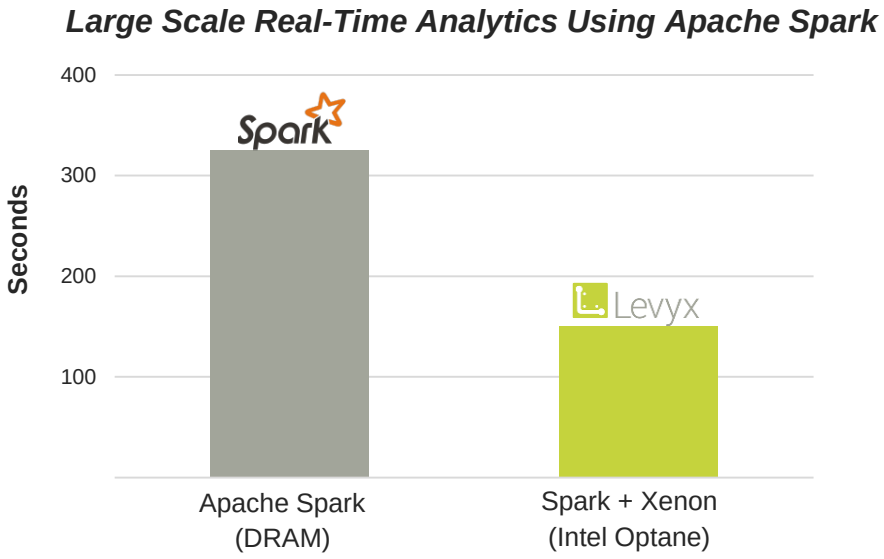
But there is a better way … use SSD or SCM in place of RAM.

Xenon™ Offload Analytics Engine Backtesting Based Solution



- **Levyx's Xenon, offered as a cluster service**
  - ✓ Native C/C++ API may be used directly by big-data applications
  - ✓ Seamlessly integrates with Apache Spark / HDFS / QuasarDB
- Provides a **live persistent DataFrame** abstraction
- Works with off the shelf valuation libraries eg Intel Finlib 1.0 for FPGA
- Provides **indexed operations** as well as efficient scan support
- Provides offload of query/analytics capabilities, including **hardware acceleration**

Xenon Applications: TPC/H-DS Benchmarking



Characteristic	DRAM	Optane drives + Xenon
Price of capacity (750GB)	>$11,000	$3,000
Analytics processing performance (TPCH, TPCDS)	650 MB/sec	1.3 GB/sec Analytics Processing
Time to load (sec)	89	1.3
Execution time (sec)	330	150

Note: Intel Optane-based solution (750GB) vs. DRAM (750GB), both running Apache Spark

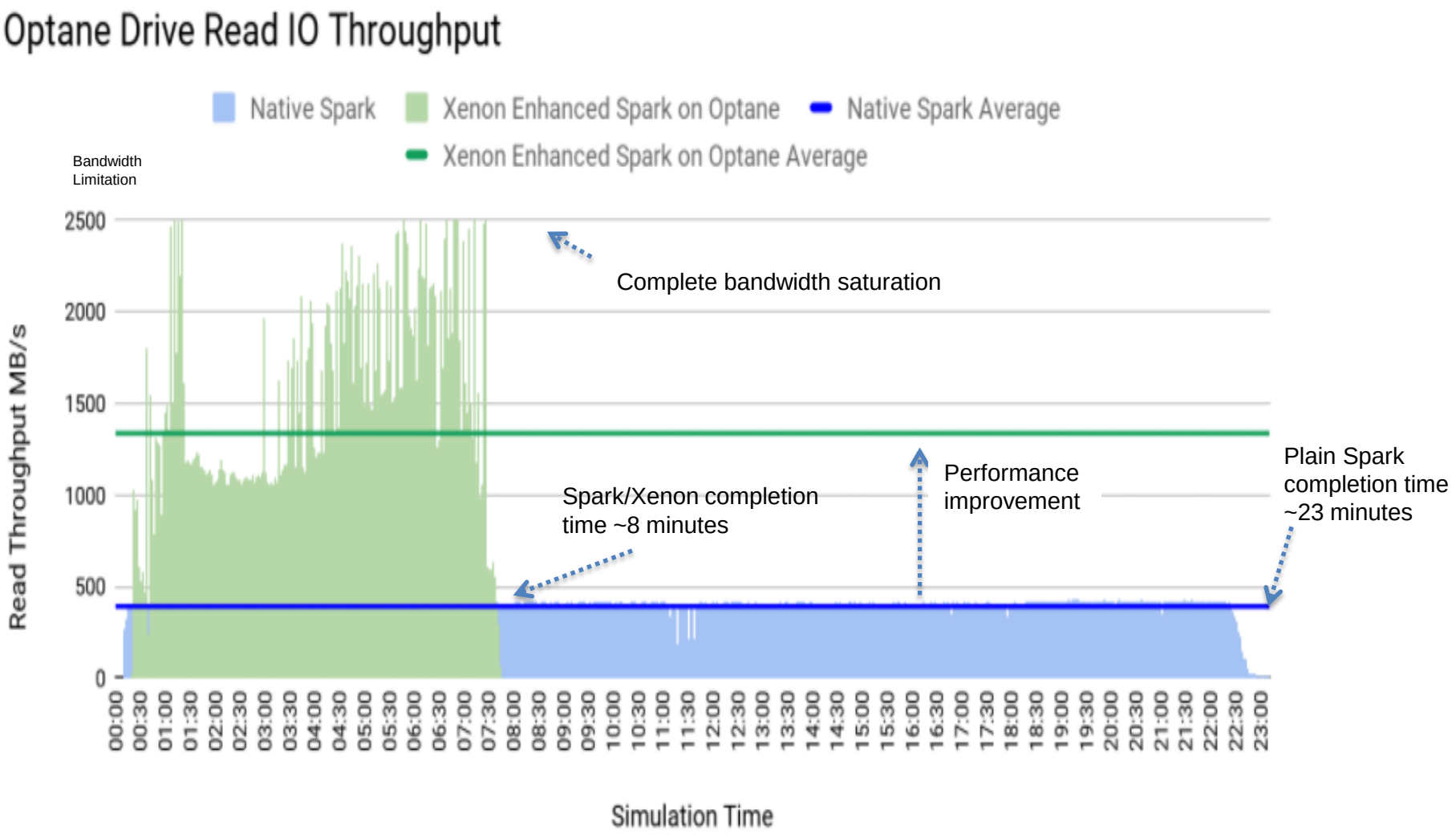
- **On all Platforms tested Spark/Xenon showed 2-3X advantage over plain Spark**
  - ✓ Google Cloud
  - ✓ Amazon Cloud
  - ✓ X-IO Axellio single 2U 2 node cluster
  - ✓ Dell with Intel P3700's

✓ **PERSISTENT** ✓ **LOWER TCO**

✓ **LOW POWER** ✓ **SHARABLE**

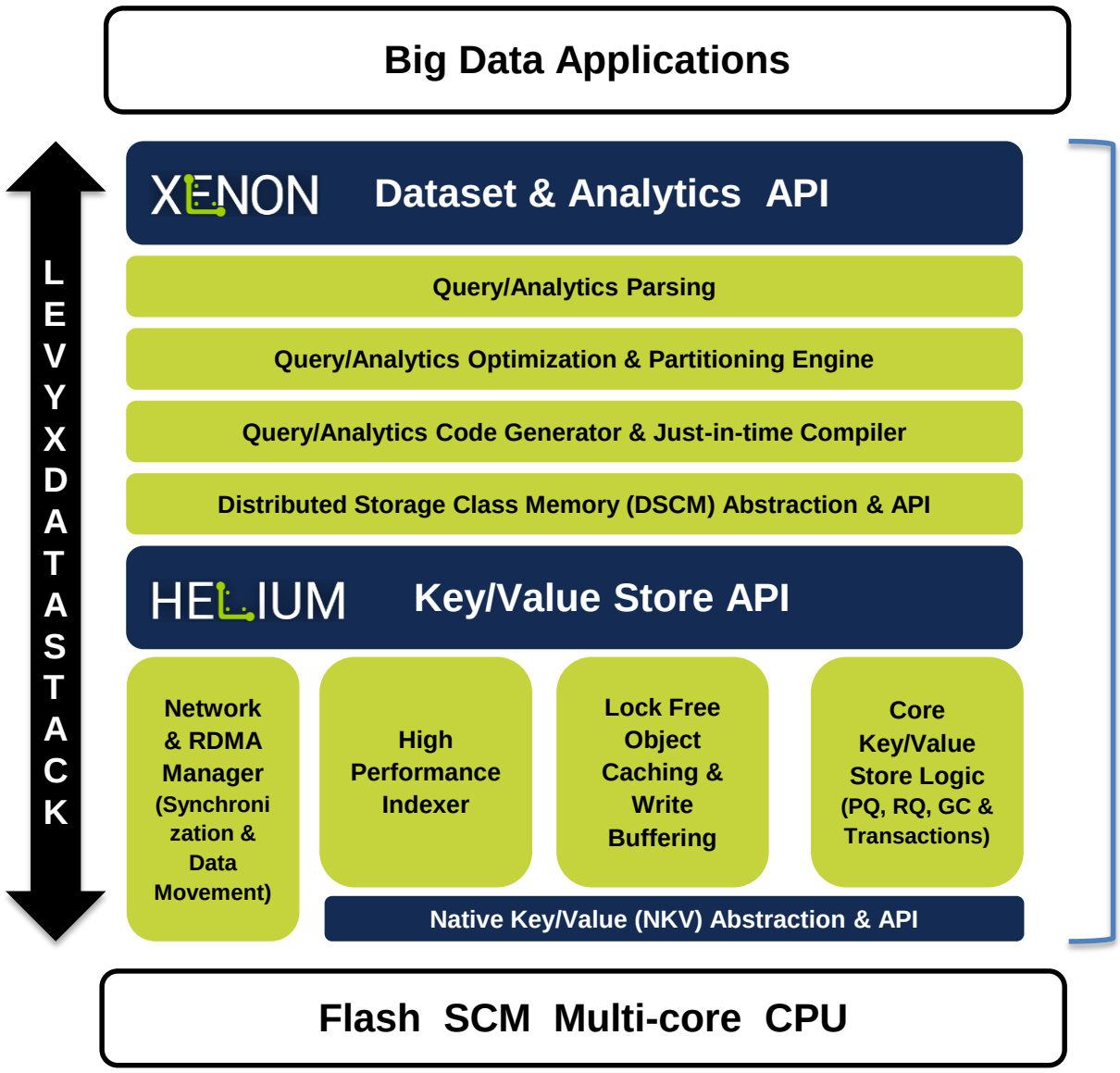
* Not STAC BENCHMARKS

Levyx' Optimal Bandwidth Utilization Improves Spark-based Backtesting Programs



* Not STAC BENCHMARKS

Flash-Optimized, Kernel Bypass Data Stack

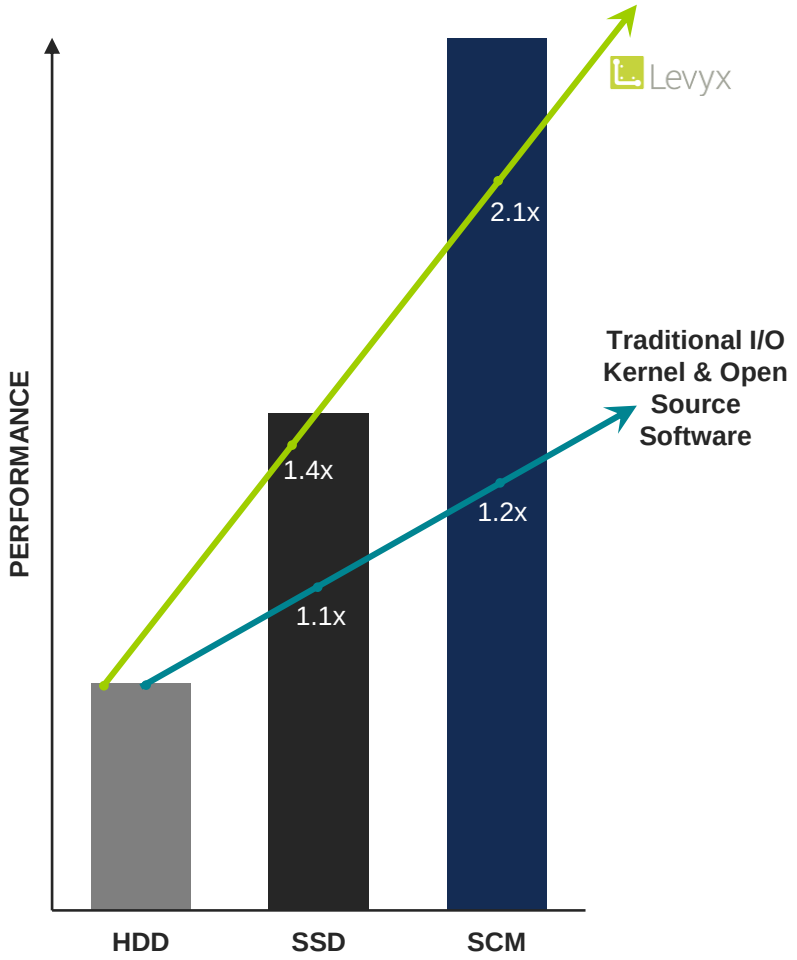


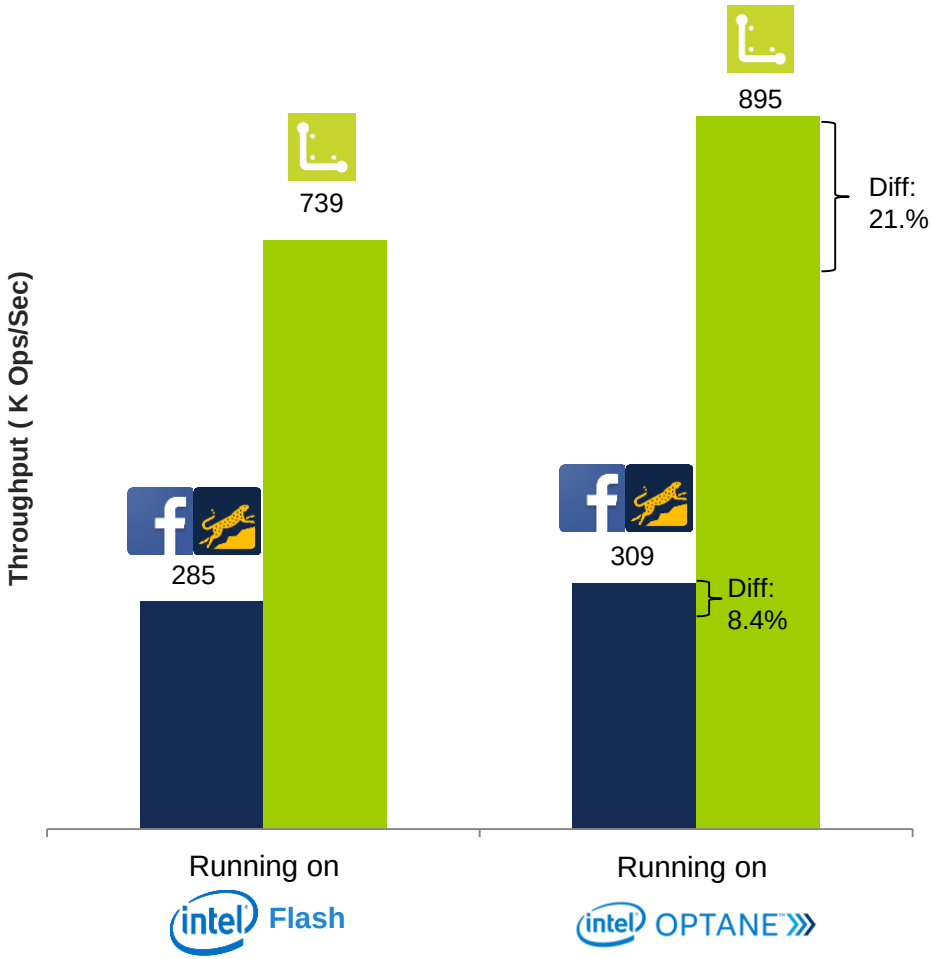
Ultra Low Latency — 10X Faster

Index Billions of Objects on a Single Node

Unmatched Price/Performance

Cost Challenge: Taking Advantage of New Technology





* Not STAC BENCHMARKS

New Levyx Features Since June

- **Community Version of He-RocksDB:**
  - RocksDB users can download today by emailing herocks@levyx.com.
- **Xenon Support for User Defined Functions**: Allows Xenon to push down calculations to FPGA's, GPU's and C/C++ libraries.
- **Locality-Aware**: A new feature in Xenon which enables Spark to assign jobs to proper nodes based on hints about data locality. For example, assigning a symbol's P&L job to the node where most of its time series data is resident. Reduces network shuffle overhead drastically.
- **Efficient DataFrame Read Path**: Improved the performance of DataFrame materialization within Spark-Xenon.
- **Support for multiple Spark jobs accessing the same Xenon DataFrame:** eg 10 machine learning models can be training on the same data set in parallel without loading the data set into memory 10 times.

Please Tick the Box and Download! Thank You!

- A free version of the Levyx' He-RocksDB allows users to easily install and experience the benefits of our implementation with just a few clicks and non-disruptively improve the customer's underlying RocksDB deployment. RocksDB users can download today by contacting herocks@levyx.com.
- Levyx's HeRocks solution is ideal for data center environments that run RocksDB to process extensive key-value sets (billion-plus) in real-time.
- Using dbbench, RocksDB's own benchmarking tool, and running on a system with 88 cores and Intel's NVMe DCP3700 SSDs, HeRocks test results were more than an order of magnitude better than the latest open-source version of RocksDB. Value sizes of 100 bytes were used representing workloads such as system logs or network logs for data centers and cybersecurity applications, stock quotes for high-frequency trading and analysis, Facebook status updates or Twitter messages. For a mixed load of 70% read/ 30% write, RocksDB could achieve 330K operations/second (ops/sec) and HeRocks could achieve **5.1 million ops/sec**. For random reads, RocksDB topped at 1.3 million ops/sec while HeRocks numbers were at **16 million ops/sec**. For "Read while Writing" test RocksDB was at 2.0 million ops/sec, and HeRocks was at **16.4 million ops/sec.**

* Not STAC BENCHMARKS