



Alpha Generation at Exabyte Scale



Over Eight Exabytes of VAST Data

The Fastest Growing In Infrastructure History

GFDL Geophysical Fluid Dynamics Laboratory

BUNGIE

verizon

jumptrading

HARVARD MEDICAL SCHOOL

CHECK POINT

SPEECHMATICS

intel.

Lawrence Livermore National Laboratory

NIH

Los Alamos NATIONAL LABORATORY

Raytheon Technologies

SQUARE POINT

agoda

mercury

U.S. AIR FORCE

in mobileye

U.S. DEPT. OF DEFENSE

COMPANY3

DARTMOUTH

Boston Children's Hospital
Until every child is well

BROWN

GINKGO BIOWORKS™

NATIONAL CANCER INSTITUTE

dug

EMBL-EBI

Yale

FOX

Northeastern University

BROAD INSTITUTE

M Man

RESEARCH G

Department of Veterans Affairs

AQUATIC

tgen

NASA

Carnegie Mellon University

INVITAE

National Heart Lung and Blood Institute

KRYSTAL
Honest. Reliable. Personal.

Sandia National Laboratories

VAST Data 2023 Overview

WEHI
brighter together



“With VAST, we have found a silver bullet: a platform that supports our efforts now and will help to accelerate our roadmap for the future too.

Blog

<https://www.gresearch.co.uk/blog/article/the-search-for-universal-storage/>



A Hyperscale Transactional Architecture

Disaggregated, Shared Everything Architecture

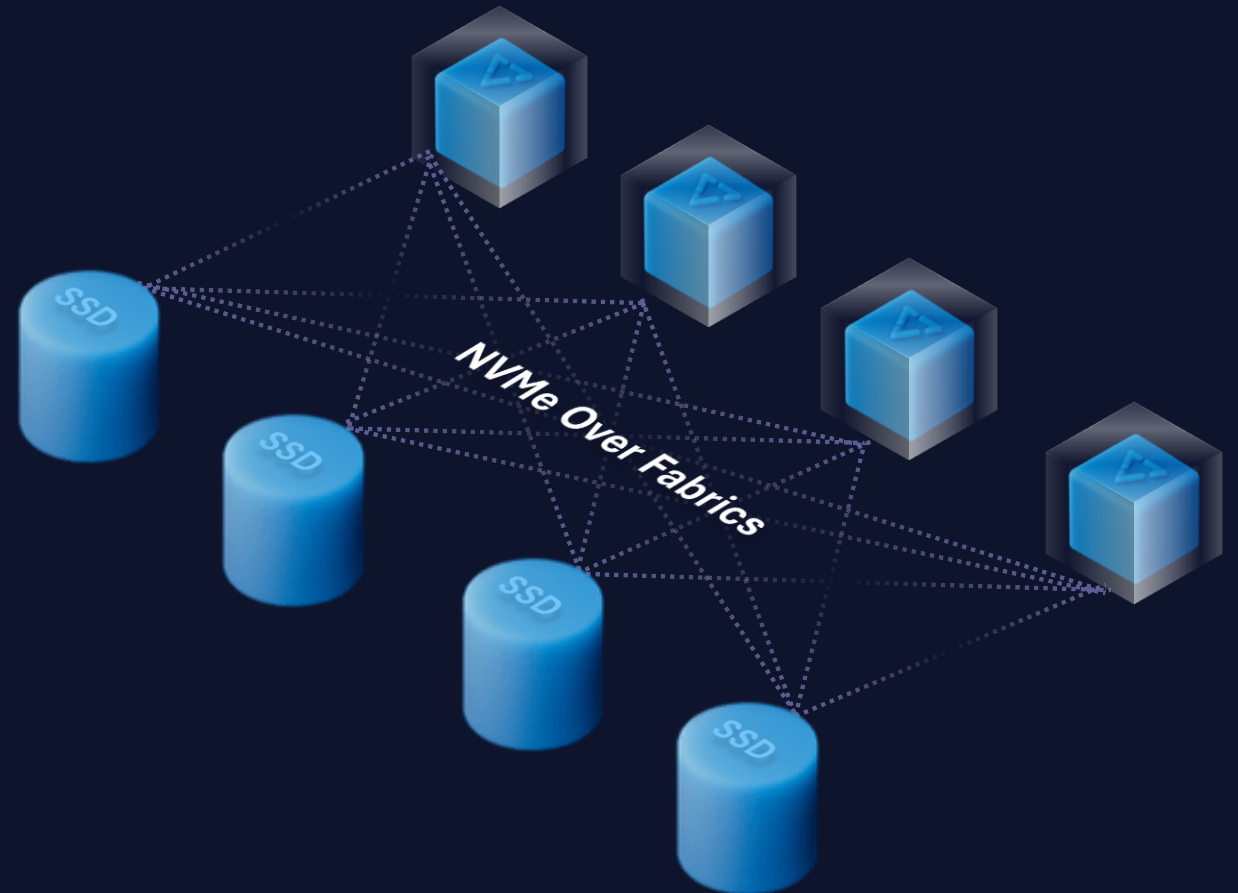
Parallel scale, no east-west traffic

Stateless Platform CPUs

No need for cache protection, scale
CPUs independent of SSDs

Shared-Everything State

All CPUs see all data and metadata over
next-gen storage fabric



Transforming The Calculus of Flash Ownership

Similarity Is Game Changing



Compression

Fine Grained, But Local

Deduplication

Global, But Coarse

VAST Similarity Reduction

Global & Fine Grained

3:1

Pre-Reduced Backups

3:1

Pre-Compressed Log Files

2:1

Life Science Data

3:1

HPC Data

3:1

Animation Data

8:1

Incremental Backups

**Example Savings From Similarity
(Not a STAC Benchmark)**

KDB+ Historical DB with Compression

1. Write Performance

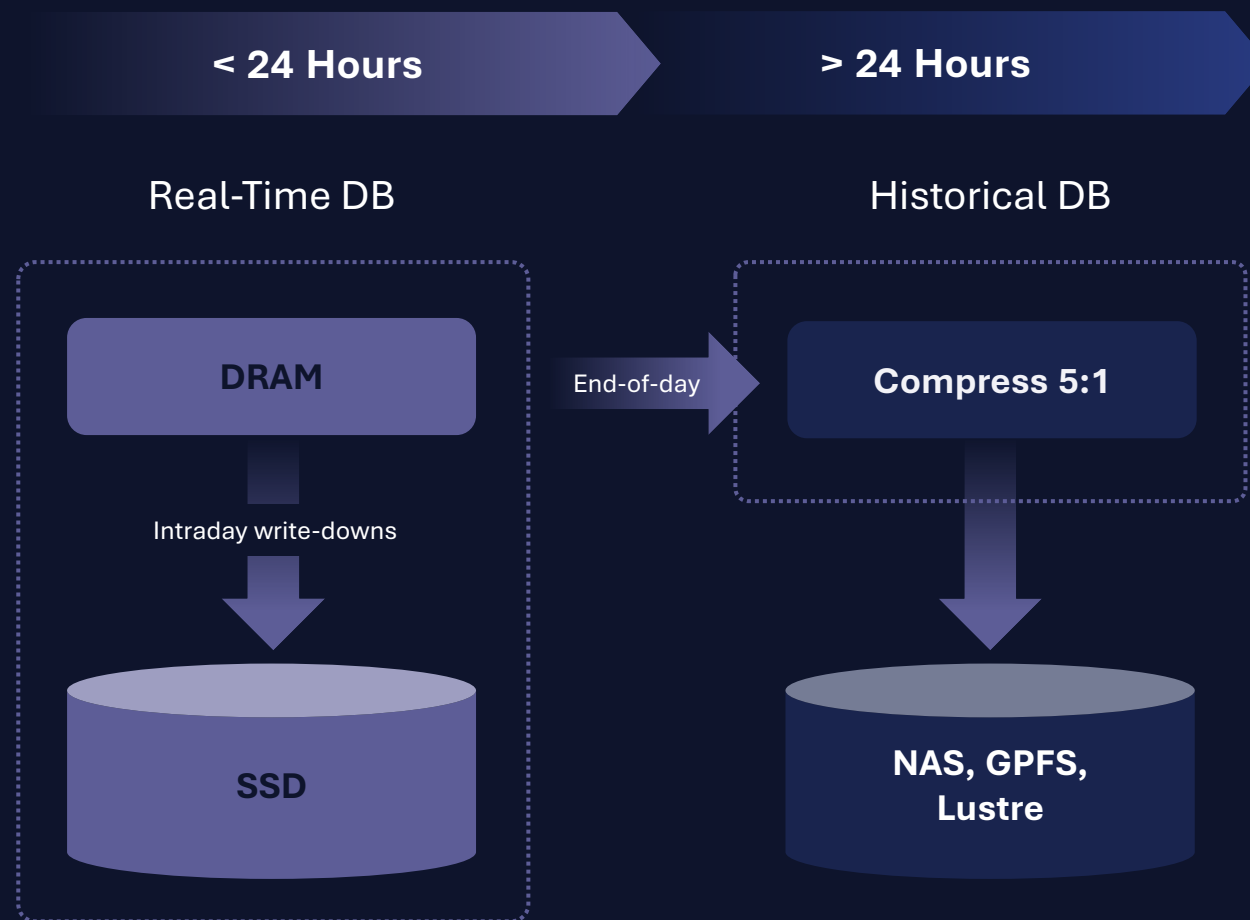
- Non-efficient compression scope is limited to each file
- Requires more CPU and time

2. Query Performance

- Data must be decompressed on query
- Non-partitioned query must decompress lots of data
- Up to 2-10X performance penalty¹

3. Performance Tuning

- Must constantly tune to satisfy business user and cost requirements



KDB+ Historical DB with VAST

1. Write Performance

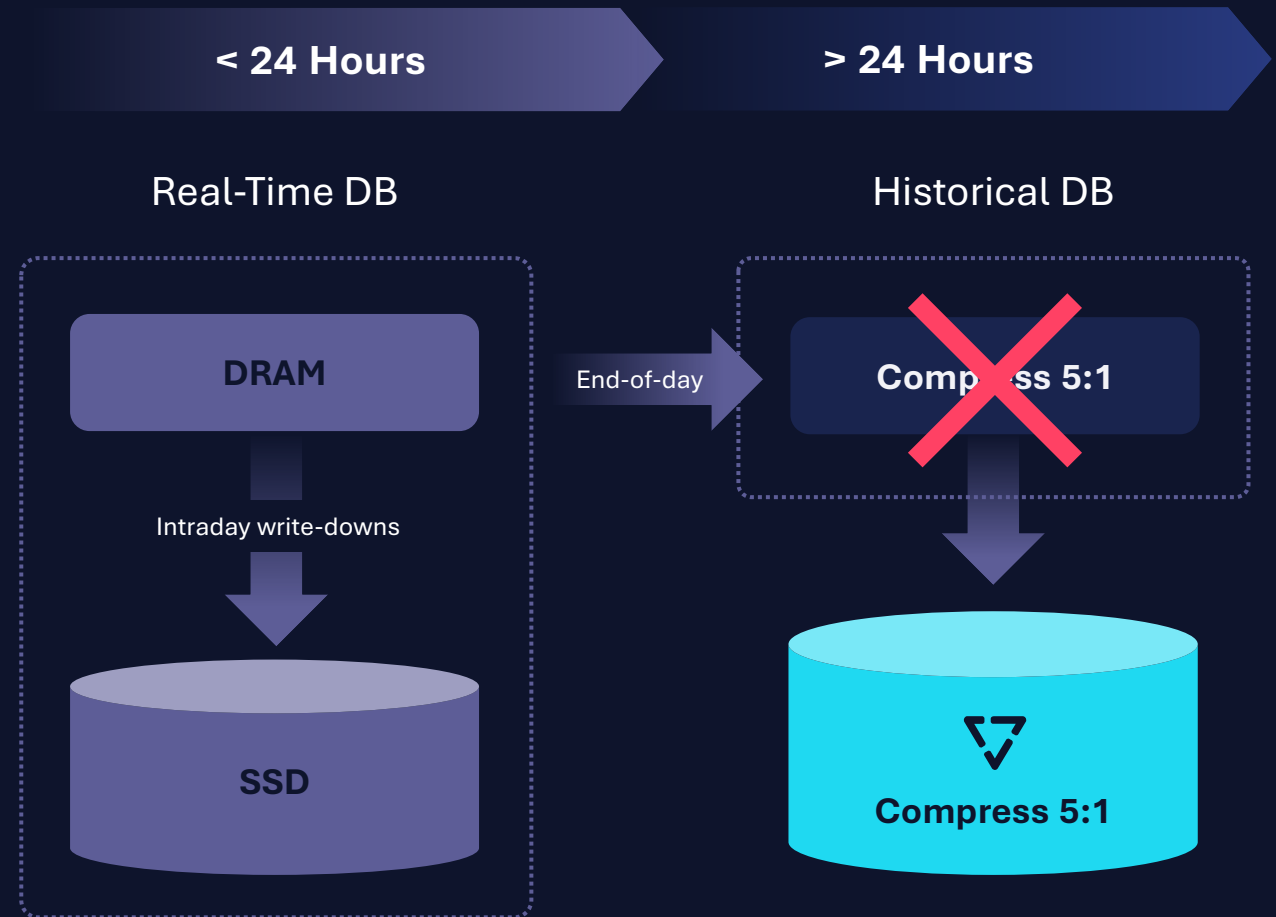
- Compress across **entire** dataset achieving the most efficient data reduction possible
- Requires **no** CPU or additional time

2. Query Performance

- All queries can be natively read without any performance impact
- All data stored on Storage-Class Memory and SSDs resulting in improved performance

3. Performance Tuning

- *What performance tuning?*



Achieve better data reduction ratios **AND** remove all the trade-offs!

**All-Flash Market
Data Archives
Are What We Do**





Thank you

