

# LiveData

For when data is greater than memory

# Agenda for Today

Data Practice

LiveData - MemVerge

Use Cases



# Data Technology Practice

## Data Technology

### LiveData

Powerful, flexible big memory

### ActiveData

Scalable high-performance storage

### DeepData

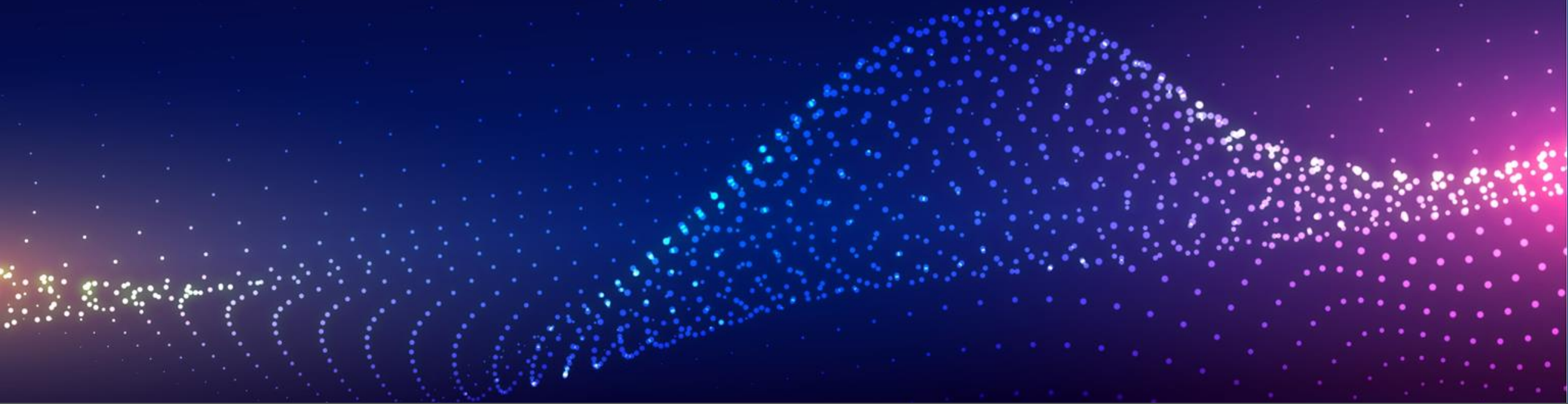
Scalable, resilient, capacity

### DataNexus

Ultimate data portability

Our data technology practice helps our customers **benefit from data-driven insight** by addressing the challenges of increasing data production in every facet of their infrastructure. Optimizing for critical design elements of workload, performance, capacity, locality, and user experience, our carefully architected data platforms will meet your current needs and provide the foundation for your future.

- Power and stability of software defined
- Workload-Optimized platforms
- Enhanced workflows and data protection
- Intelligent tiering architectures
- Unlock the true potential of your in-memory compute
- Liberate your team from data gravity constraints



# LiveData

# DRAM & PMEM



## Capacities

- 16GB / 32GB / 64GB / **128GB**

## Speeds

- 2133MHz / 2400MHz / 2666MHz
- 2933MHz / 3200MHz



## Capacities

- 128GB / 256GB / 512GB

## Speeds

- 2133MHz / 2400MHz
- 2666MHz



# Obstacles Facing PMEM

DRAM is a  
**Higher Performance**  
Memory Tier



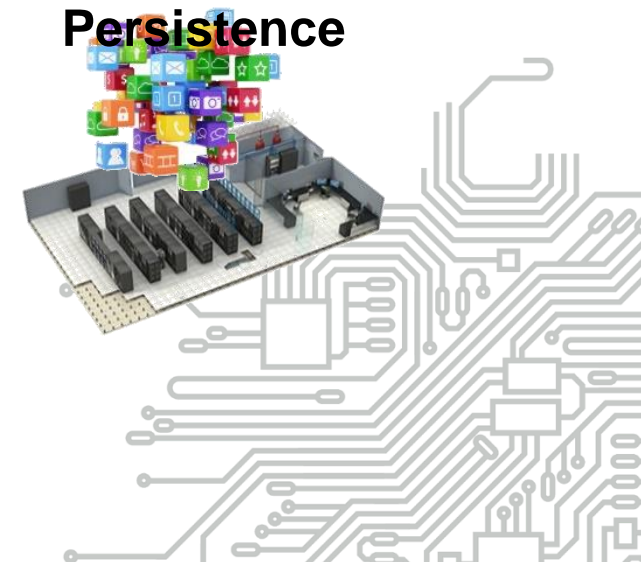
AMD Supports  
**More DRAM**  
Capacity per Server





Memory is not a  
**High Availability**  
Mass Storage Tier



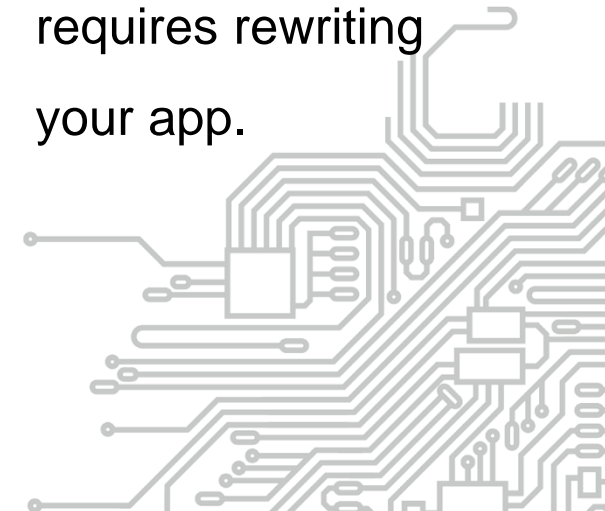
Broad Adoption  
Means Every App  
Must be Modified to  
**Support**  
**Persistence**






# PMEM Mode Options

Access Method	 Memory Mode	 App-Direct Mode
Compatibility	Plug-Compatible with byte-addressable DRAM	Requires rewrite to be byte addressable
Data Services	None	None
Performance	Slightly slower than DRAM	Slightly slower than DRAM
Persistence	Volatile like DRAM	Persistent

- Memory Mode gives up the persistence of Optane.
- App-Direct Mode requires rewriting your app.

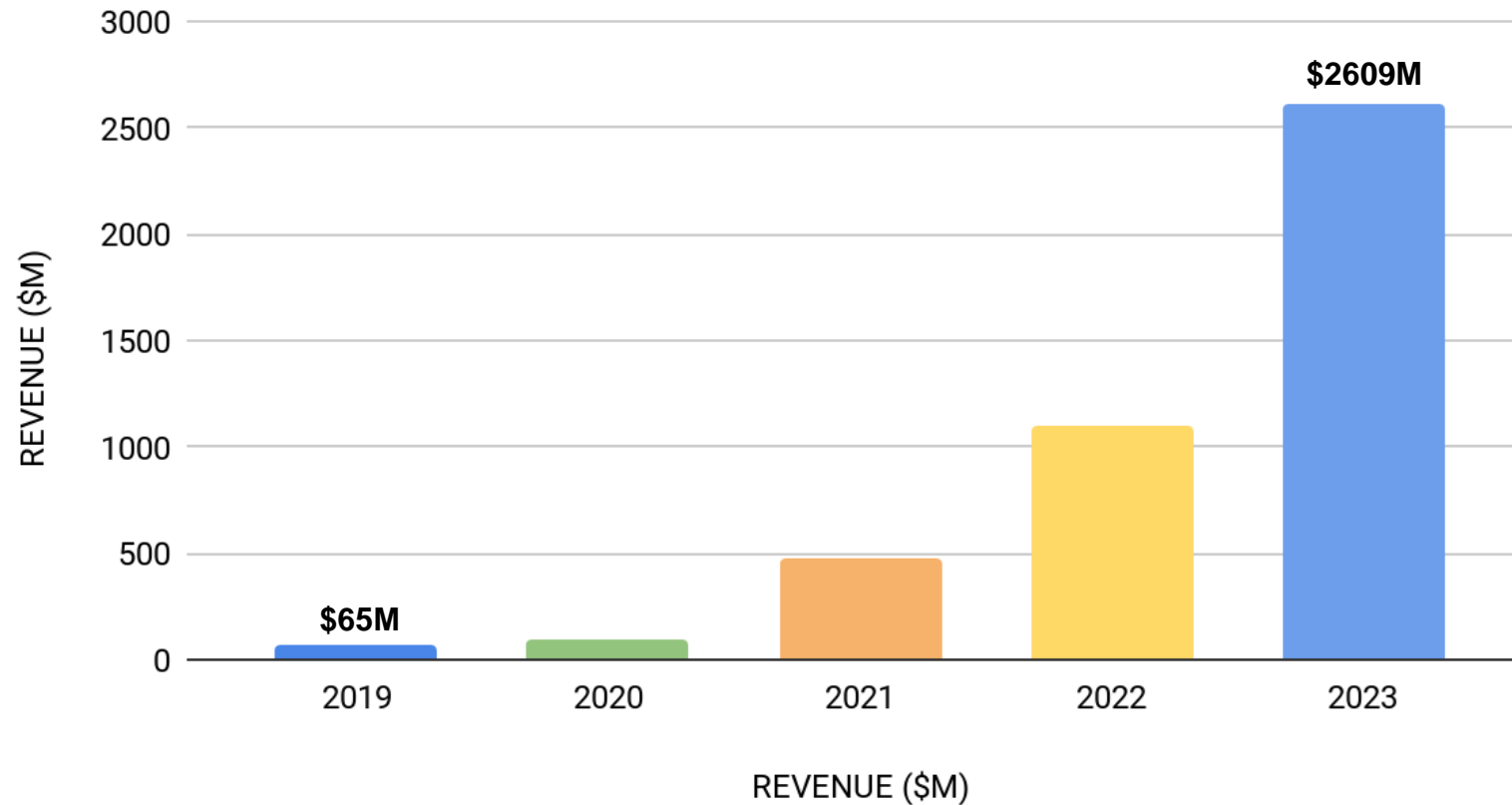


# MemVerge Memory Machine

Access Method	 Memory Mode	 App-Direct Mode	 Memory Machine
Compatibility	Plug-Compatible with byte-addressable DRAM	Requires rewrite to be byte addressable	Plug-Compatible with byte-addressable DRAM
Data Services	None	None	ZeroIO Snapshot, Cloning, Replication
Performance	Slightly slower than DRAM	Slightly slower than DRAM	Fast Hybrid Memory (DRAM+PMEM)
Persistence	Volatile like DRAM	Persistent	Persistent + Snapshot, Replication, & Recovery

# Persistent Memory Market Growth

248% CAGR 2019-2023



# LiveData

## Data Practice

### LiveData Big Memory

Penguin Computing LiveData addresses the DGM roadblock by providing a memory virtualization software layer that delivers software-defined memory services to the applications without application changes. This allows the thousands of applications running in the data center today to take advantage of higher capacity memory at a lower cost.

## Features

- Tier Persistent Memory and DRAM
- Low-latency memory replication
- Virtualize to form a platform for enterprise-class data services.
- Plug-compatible access with existing applications.
- Recover hundreds of GB in seconds with ZeroIO™ memory snapshots
- Clone databases in seconds

## Benefits

- Scale memory capacity and improve system performance.
- Maintain availability - recover in seconds, not hours.
- Enable agility with efficient clone deployment, and fast application rollbacks.
- Avoid application disruption or rewrite.
- Reduce latency.

# LiveData






## Penguin Computing LiveData Components

Data Practice



- AI/ML Workloads
- Realtime Analytics Workloads
- HPC Workloads
- Cloud/VDI Workloads

User Layer

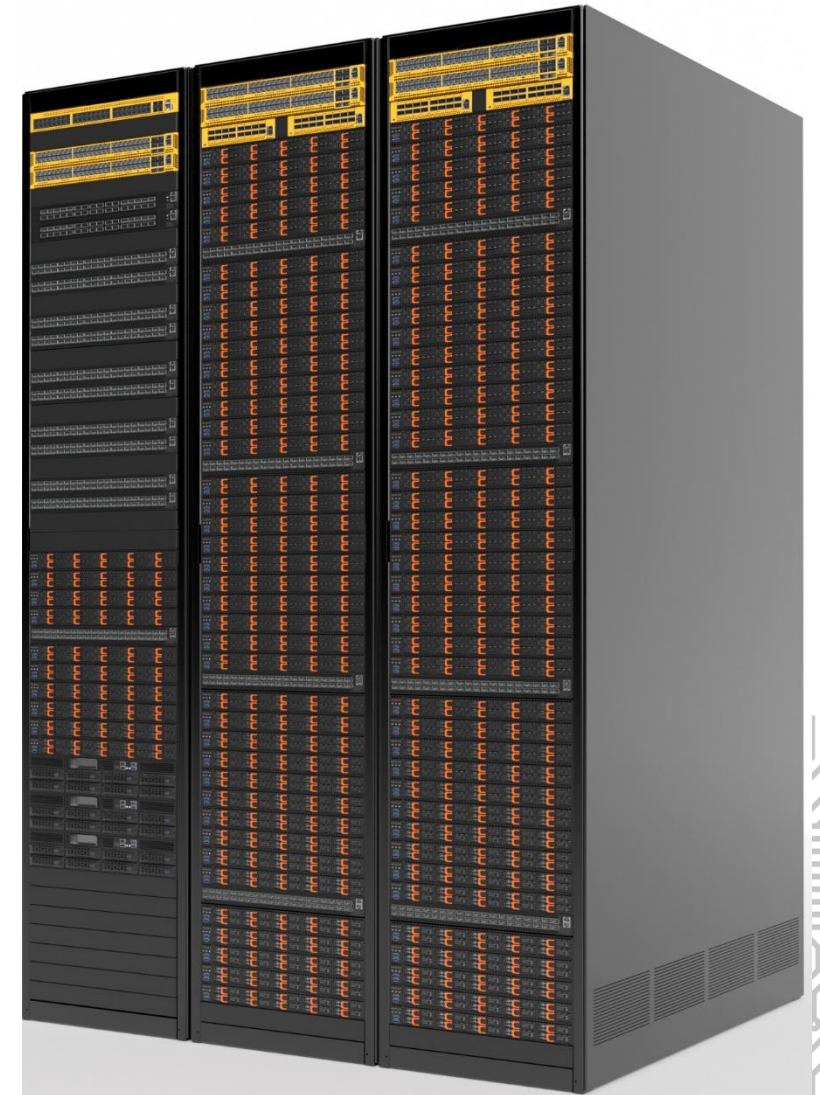
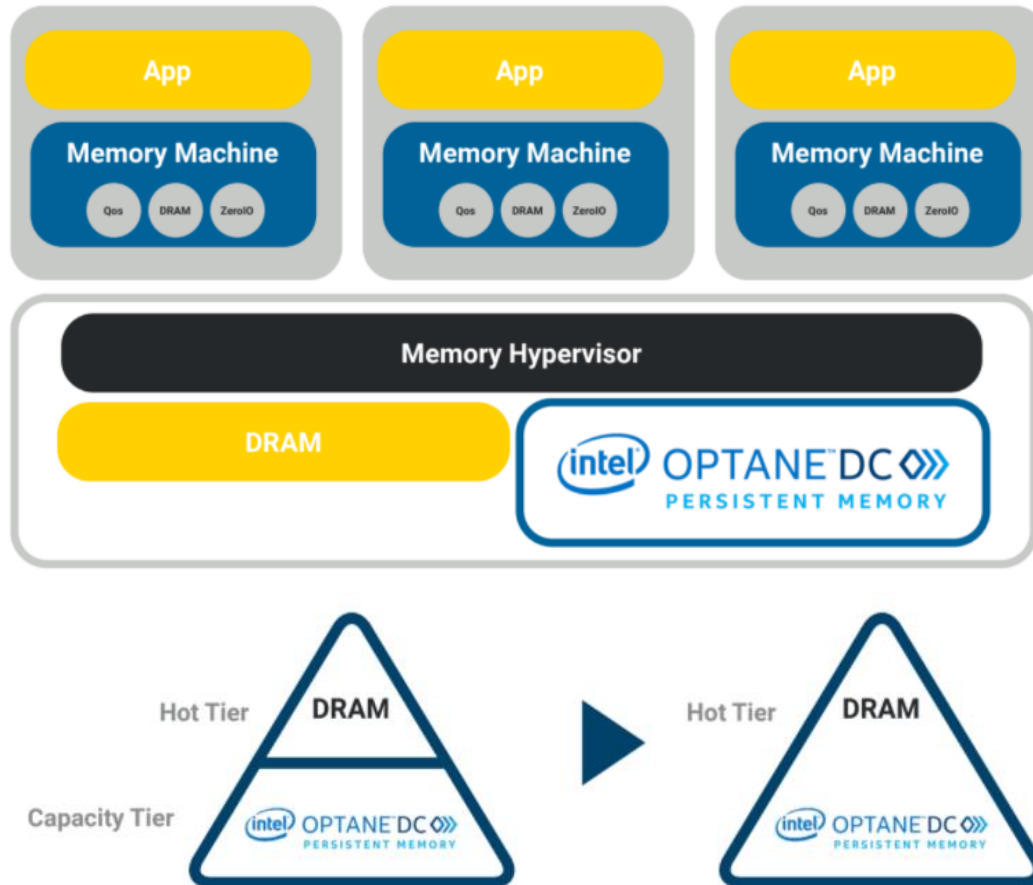
- Software Technologies 
- Compute Technologies 
- Data Technologies 
- Data Center Infrastructure 
- Penguin Computing Services 

Infrastructure Layer



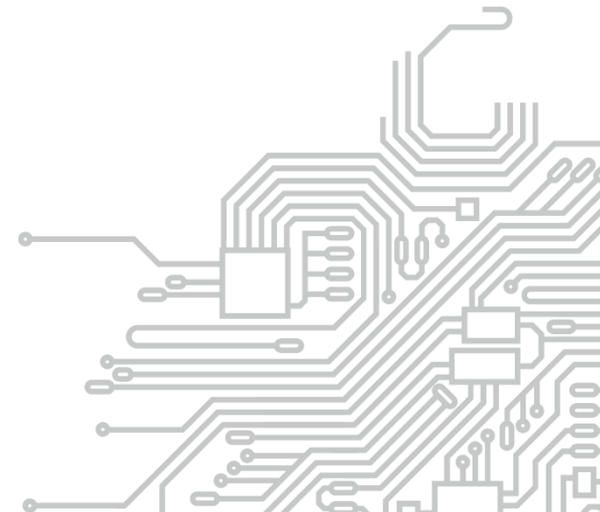
# LiveData

## MemVerge Memory Machine



# LiveData - Value Proposition

- Memory Tiering allows **100% of memory capacity** to be addressable
- **Flexible DRAM to PMEM ratios** to optimize mix for cost savings
- Highest memory **density** per server
- **Enhanced HA** capabilities from MemVerge data services



# LiveData - Server Support



## Relion XE1112e Server

### *Memory Capacity*

- 1.5 TB DDR4 DRAM
- 6 TB Intel Optane PMEM
- 7.5 TB Total Memory



## Relion XE2142e Server

### *Memory Capacity*

- 1.5 TB DDR4 DRAM
- 2 TB Intel Optane PMEM
- 3.5 TB Total Memory



## Relion XE4118GT/S Server

### *Memory Capacity*

- 1.5 TB DDR4 DRAM
- 3 TB Intel Optane PMEM
- 4.5 TB Total Memory

# LiveData - Server Support



## Relion XO1132e Server

### *Memory Capacity*

- 1.5 TB DDR4 DRAM
- 2 TB Intel Optane PMEM
- 3.5 TB Total Memory



## Relion XO1114GT Server

### *Memory Capacity*

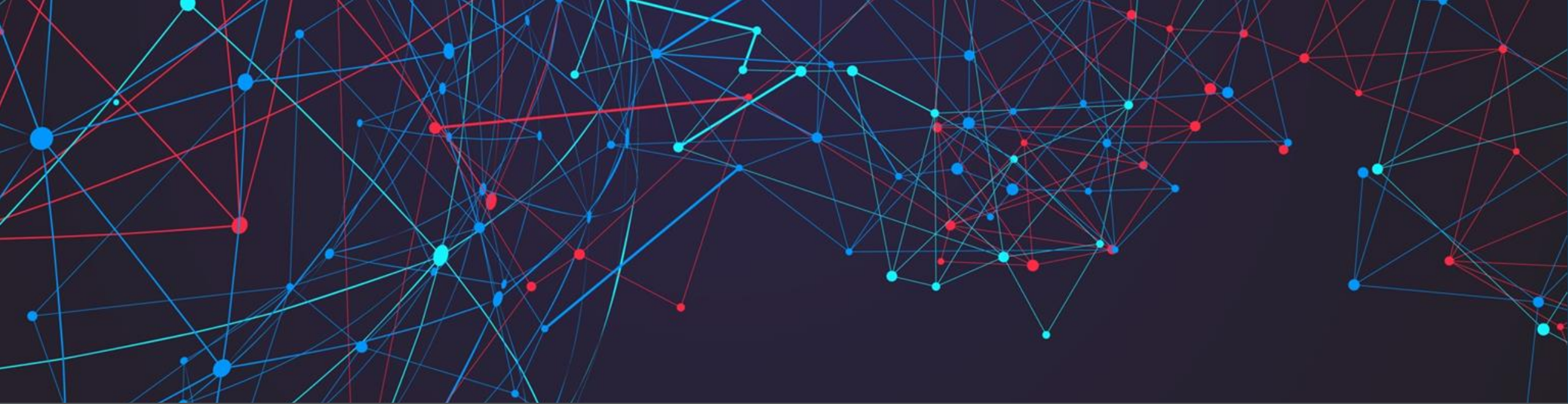
- 1.5 TB DDR4 DRAM
- 3 TB Intel Optane PMEM
- 4.5 TB Total Memory



## Relion XO1114GT/S Server

### *Memory Capacity*

- 1.5 TB DDR4 DRAM
- 3 TB Intel Optane PMEM
- 4.5 TB Total Memory



# Use Cases

# LiveData - Real-Time Workloads

According to IDC, by 2021, 60-70% of the Global 2000 organizations will have at least one mission-critical real-time workload. Below are just a few examples of use cases that are implementing Big

Memory now.



Latency-Sensitive  
Transactional Workloads



Real-Time Big Data  
Analytics  
in financial services, healthcare,  
retail, Federal Govt, and security.



AI/ML Analytics and  
Inferencing  
like fraud detection, image and  
video recognition, autonomous  
vehicles and healthcare.

# LiveData - Big Memory for Financial HPC

Financial Services

LiveData  
Big Memory



## Problem - IMDB Cluster takes 24 Hours to Restore

- Hazelcast in-memory database distributed across 40 servers.
- A cluster crash takes over 24 hours to completely restore
- Each server limited to 100GB of DRAM because scaling DRAM is extremely expensive
- Deploying lower cost PMEM sacrifices performance

# LiveData - Big Memory for Financial HPC

Financial Services

**LiveData**  
Big Memory



## Solution - IMDB Cluster Restored in Seconds

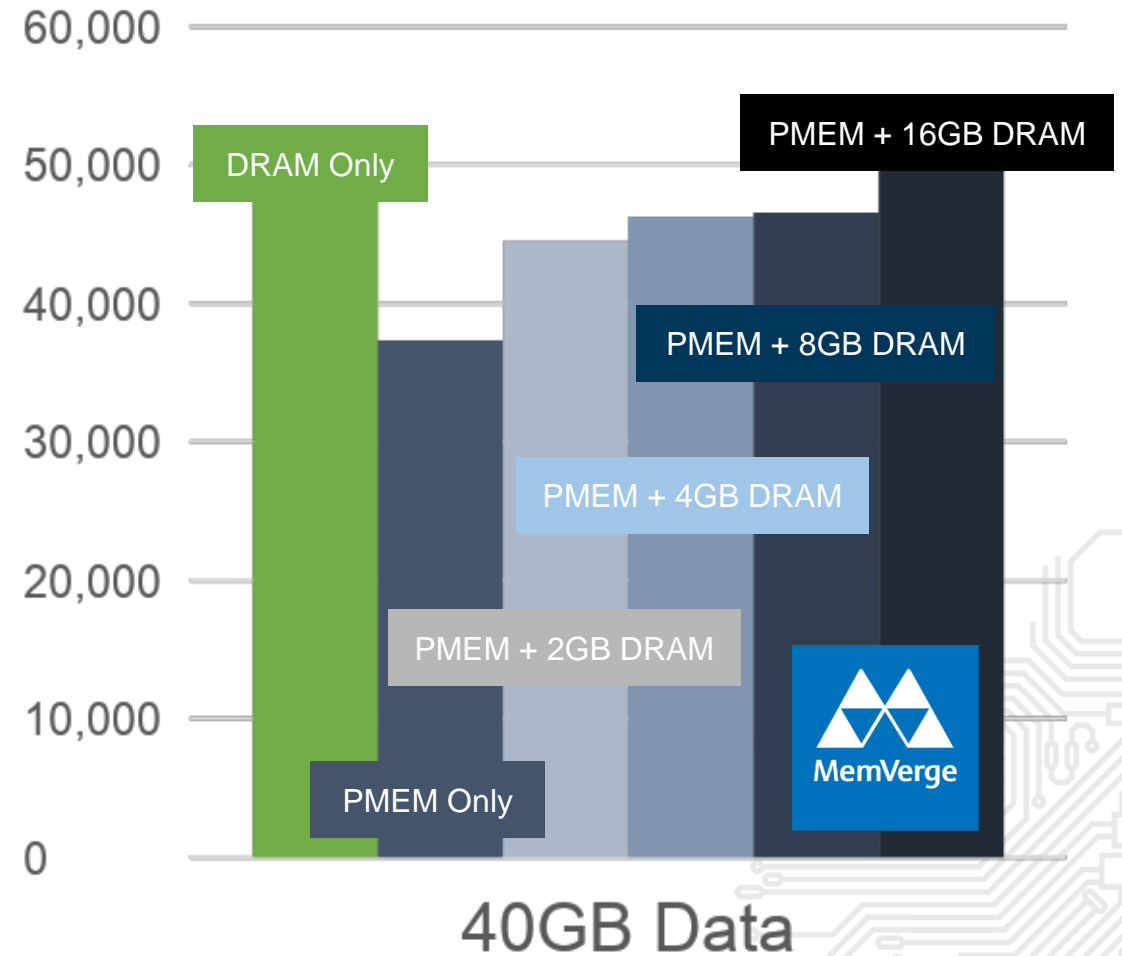
- MemVerge allows for Hazelcast cluster to be restored in seconds
- TCO - PMEM provides twice the memory for the same cost
- 10x Server Consolidation
- MemVerge memory allocator technology allows PMEM to go as fast or faster than DRAM

# LiveData - Big Memory for AI/ML Inferencing

## Financial Services

### LiveData Big Memory

- DRAM = 48.3K Transactions Per Second
- PMEM + 16GB DRAM = 50K Tps (+3.5%)



# LiveData - Big Memory for Workstation HPC

Media & Entertainment

**LiveData**  
Big Memory



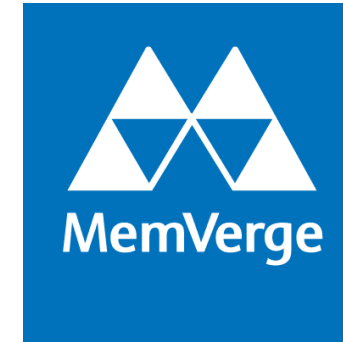
## Problem - 83,000 Hours of Creative Artist Time Lost Per Year

- "We need more RAM" is the top request from artists
- Creative artist time is the biggest cost item for any movie production!
- Creative artist's lose on average 1-2 hours of productivity per week
- 800 Artists x 1,600 hours per week = 83,200 hours per year

# LiveData - Big Memory for Workstation HPC

Media & Entertainment

**LiveData**  
Big Memory



## Solution - Time Equivalent to 40 Creative Artists Restored

- Lower TCO: Workstations with double the Memory + MemVerge for the same cost or less
- “Autosave” improves productivity by potentially 40 Creative Engineers annually
- \$8M potential savings annually
- Higher quality movies & faster time to market

# LiveData - Big Memory for Workstation HPC

Media & Entertainment

**LiveData**  
Big Memory

Snapshot Time

- 99.66% Reduction

Restore Time

- 99.98% Reduction

**Before**

Done every few hours because it's disruptive

Manual  
Snapshot

Snapshot  
takes Minutes

Restore  
takes Hours

10 min

3 hrs

**After**

Automatically done every few minutes

Auto Save

Snapshot  
takes 1-2 Sec

Restore  
takes 1-2 Sec



# LiveData - Big Memory for AI/ML Inferencing

AI/ML Inferencing

LiveData  
Big Memory



## Problem - Data Greater than Memory

- When data is greater than the size of DRAM, AI/ML performance slows down dramatically
- Memory-intensive inference jobs take a long time to load and restart

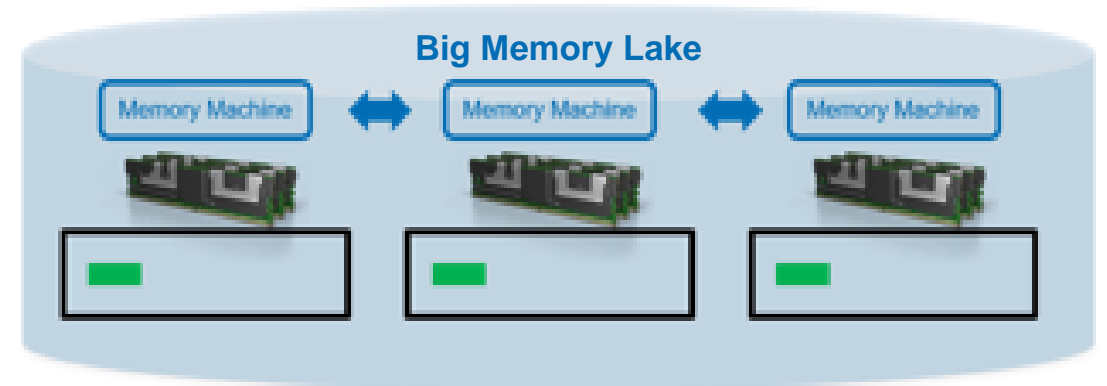
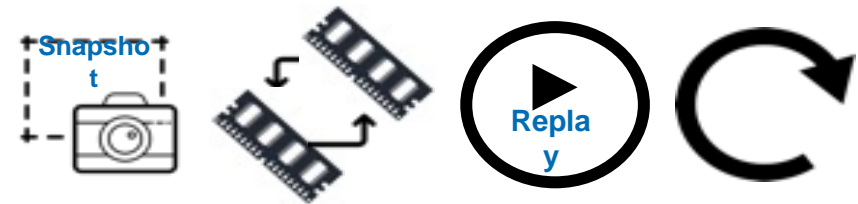
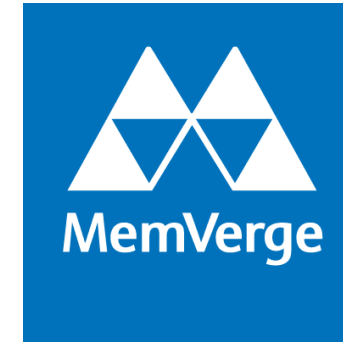
# LiveData - Big Memory for AI/ML Inferencing

AI/ML Inferencing

LiveData  
Big Memory

## Solution - Big Memory Lake

- Create big memory lakes consisting of DRAM and PMEM to provide capacity needed for all data including models and embeddings
- Fast data recovery and restart by using in-memory data snapshot



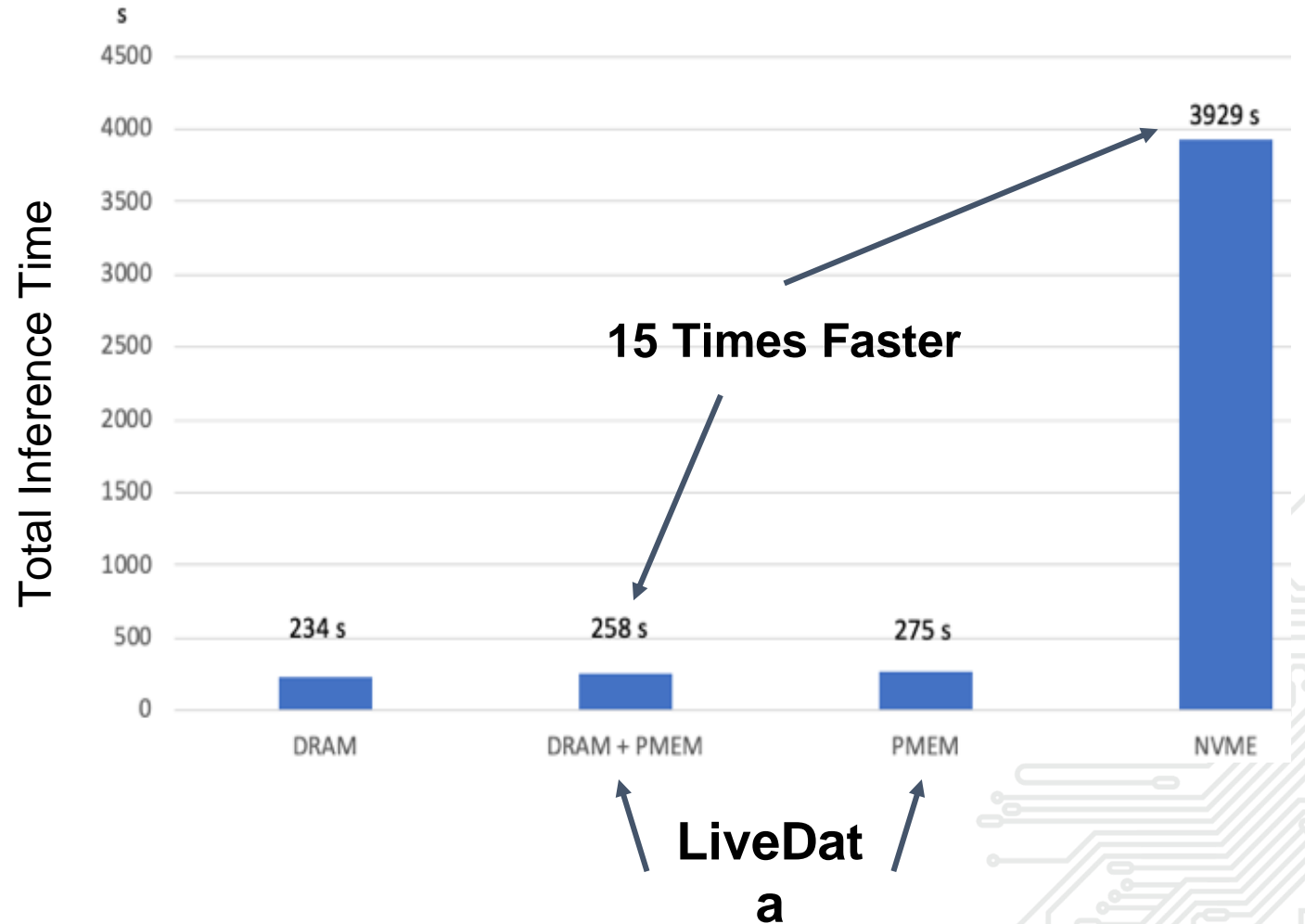
# LiveData - Big Memory for AI/ML Inferencing

## AI/ML Inferencing

**LiveData**  
Big Memory

### Inference Time

- DRAM+PMEM is 10% Slower than DRAM
- DRAM+PMEM is 15X Faster than NVMe



# Thank-You