

# Introducing the CXL Computing Cell Architecture

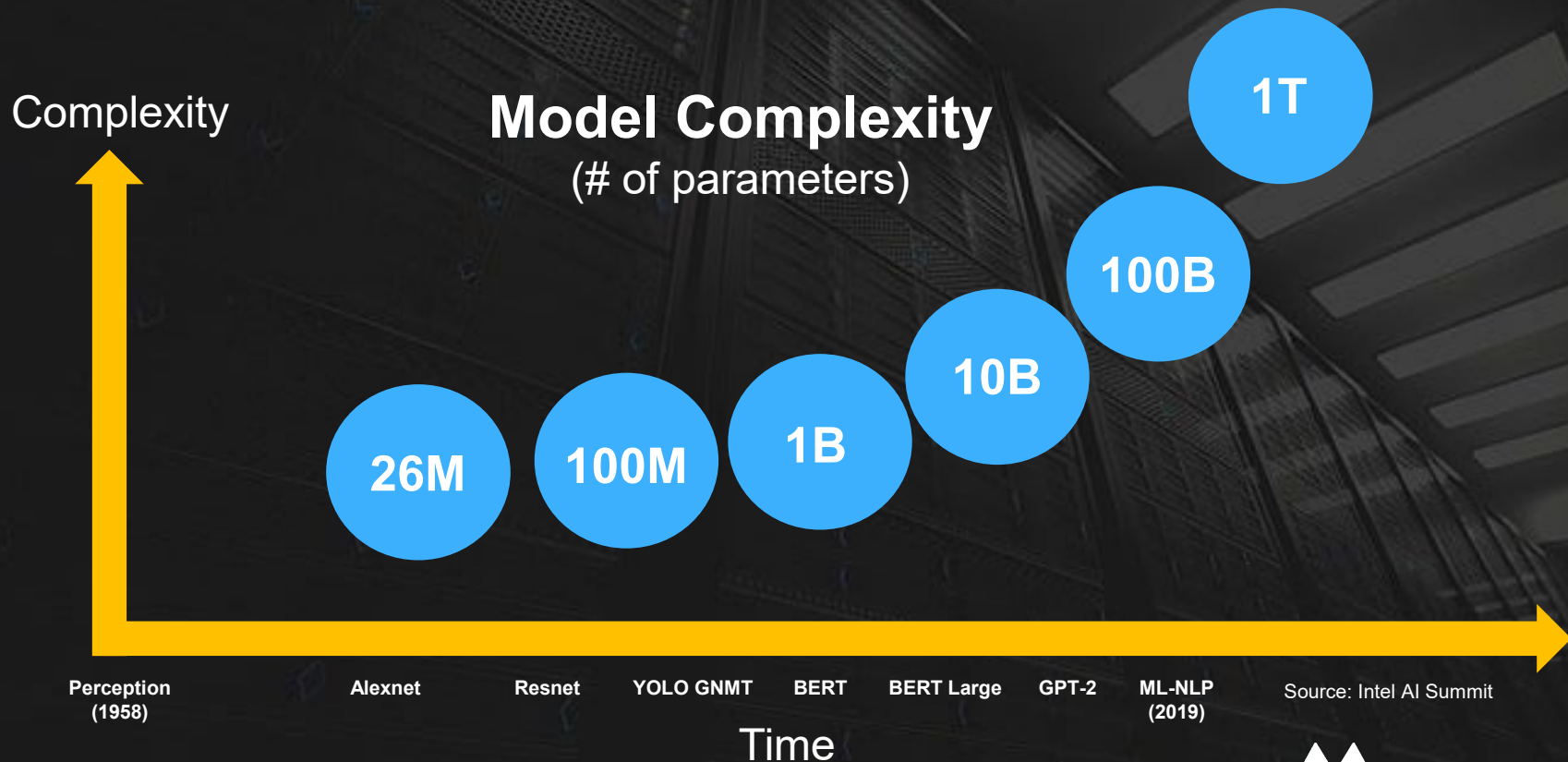


Flash Memory Summit

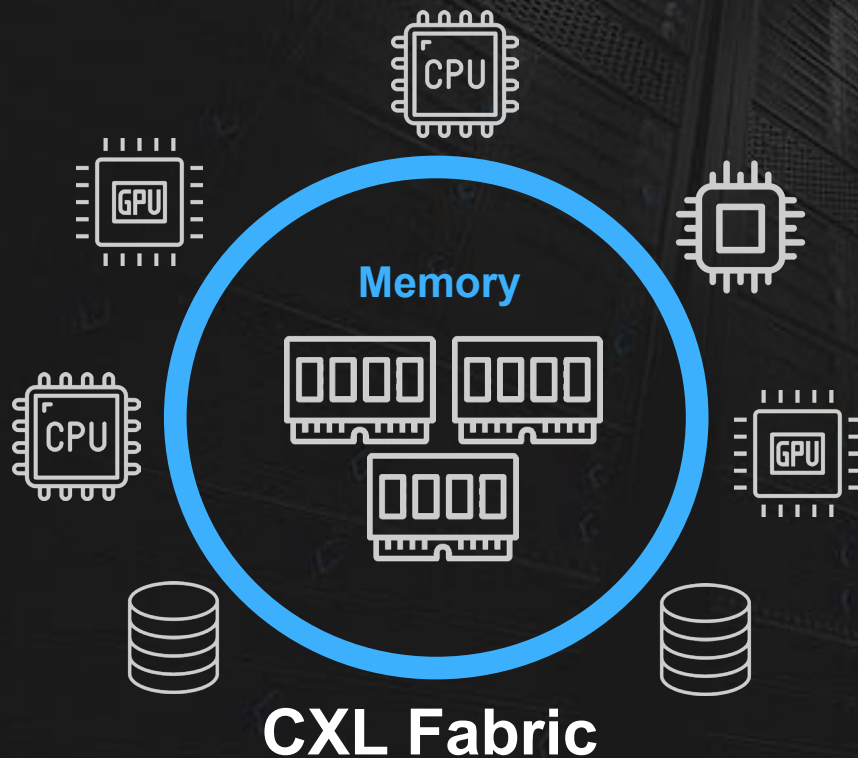
August 2023



# Growth of Data Led by AI/ML



# Emergence of CXL (Compute eXpress Link)



## 1.1 Memory Expansion

- Both capacity and bandwidth
- Spec: 2019; Product: 2023

## 2.0 Memory Pooling

- Disaggregation and composability
- Elastic memory on-demand
- Spec: 2021; Product: 2024

## 3.0 Memory Sharing

- Cache coherence
- Switch cascading
- Spec: 2022; Product: 2026+

# Emerging Products

1

Memory Machine X

Essentials

Essential tool for  
CXL devs & users

CXL 1.1



See  
Your  
Memory

2

Memory Machine X

Pro

CXL fabric & data  
management SW

CXL 1.1



Local  
Memory  
Pool

3

Memory Machine X

Pro

CXL fabric & data  
management SW

CXL 2.0



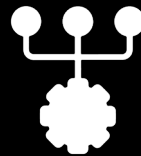
External  
Memory  
Pool

4

Project Gismo

Global IO-Free  
Memory Objects

CXL 2.0



Shared  
Memory  
Objects

5

Computing Cell

Co-Engineered  
HW/SW System

CXL 2.0



Pool  
External  
Memory

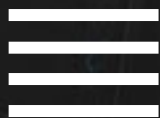
# CXL 1.1 Local Memory Pool



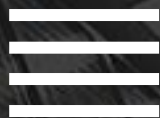
Memory Machine  
Essentials / Pro



CPU  
Socket



CPU  
Socket



C  
X  
L

C  
X  
L

C  
X  
L

C  
X  
L

C  
X  
L

C  
X  
L

C  
X  
L

C  
X  
L

Server



MemVerge



# 1. See Your Memory



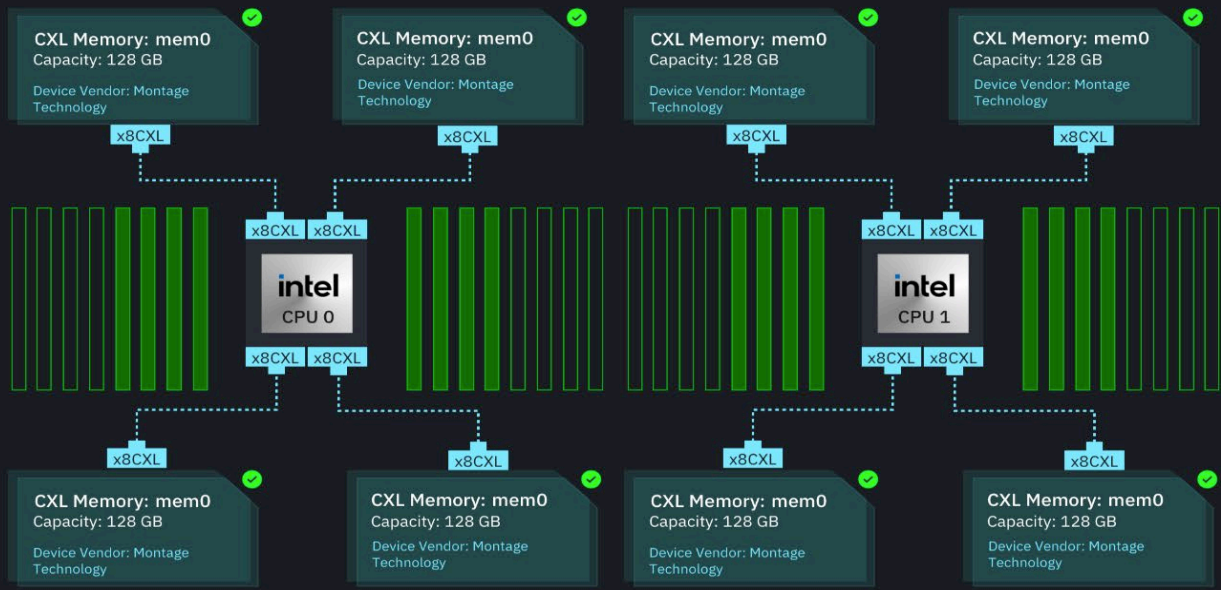
HOSTNAME 001

System Topology

Overview

Numa Node 0

Numa Node 1



CXL Memory Capacity	256 GB
DRAM Capacity	256 GB
System Total Capacity	512 GB

Memory

Throughput

CPU

DRAM

Total 256 GB

Used: xxx GB

Free: xxx GB

Usage Trend

Real Time

CXL Memory

Total 256 GB

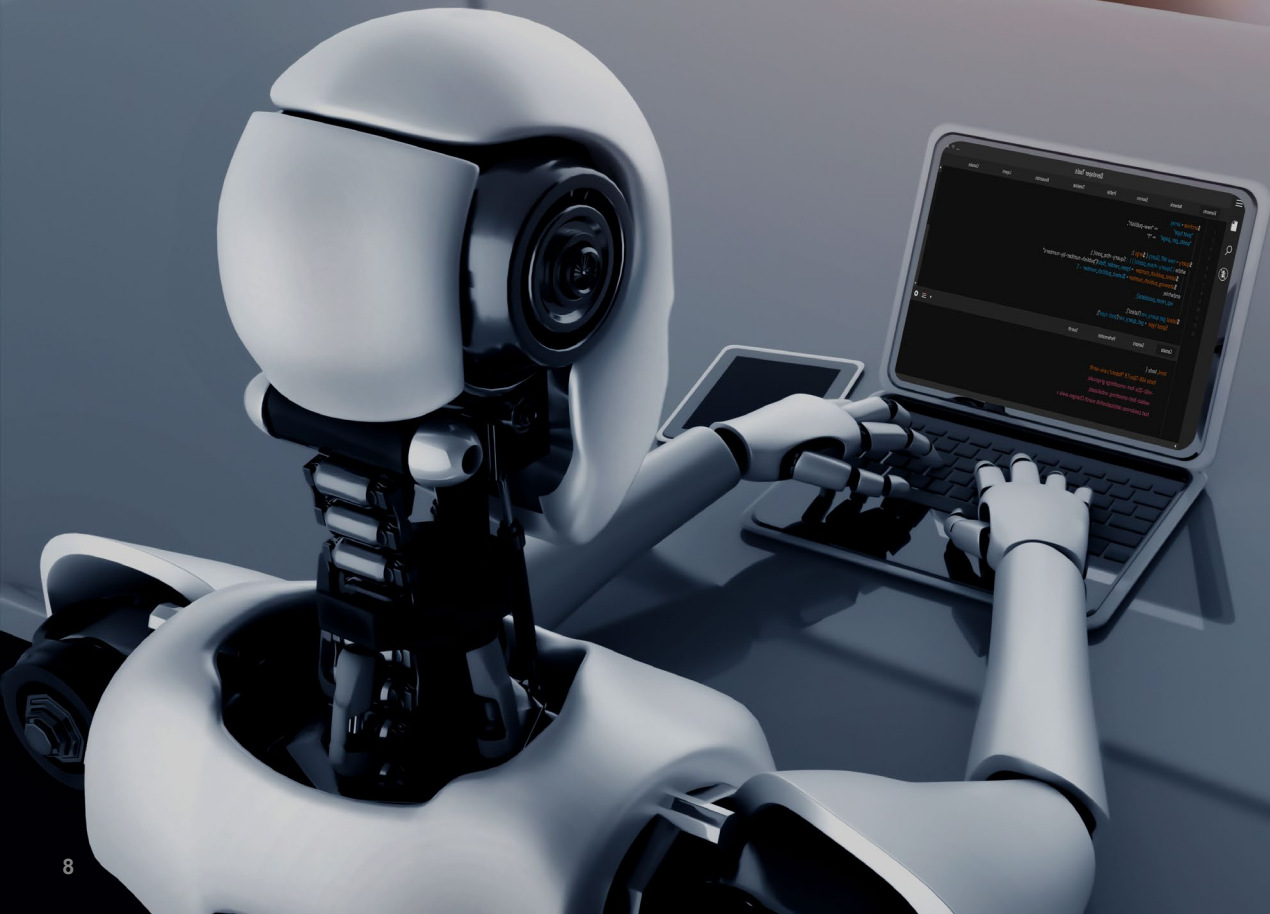
Allocated: xxx GB

Available: xxx GB

Usage Trend

Real Time

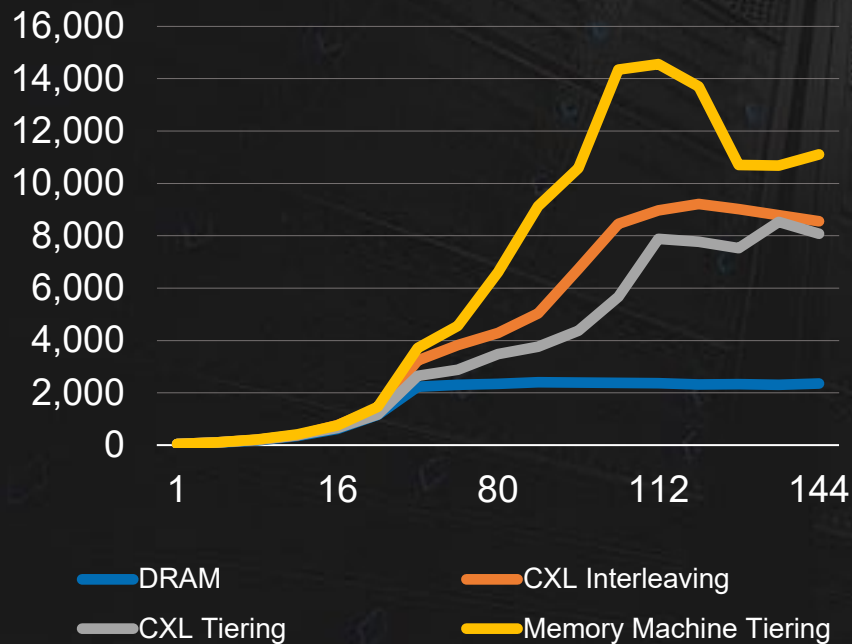
## 2. Local Memory Pooling



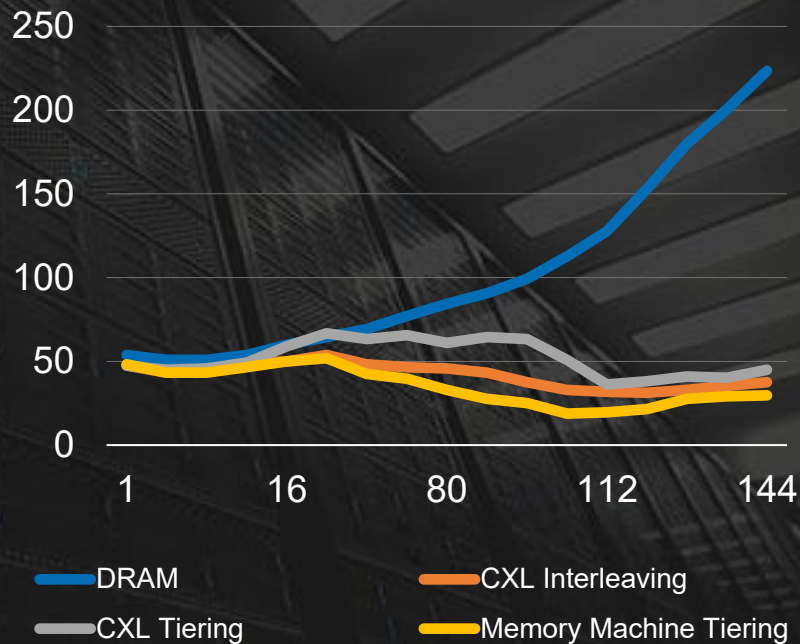


# Performance with Automated Memory Tiering

## TPS Absolute Perf



## P95 Latency Absolute Perf





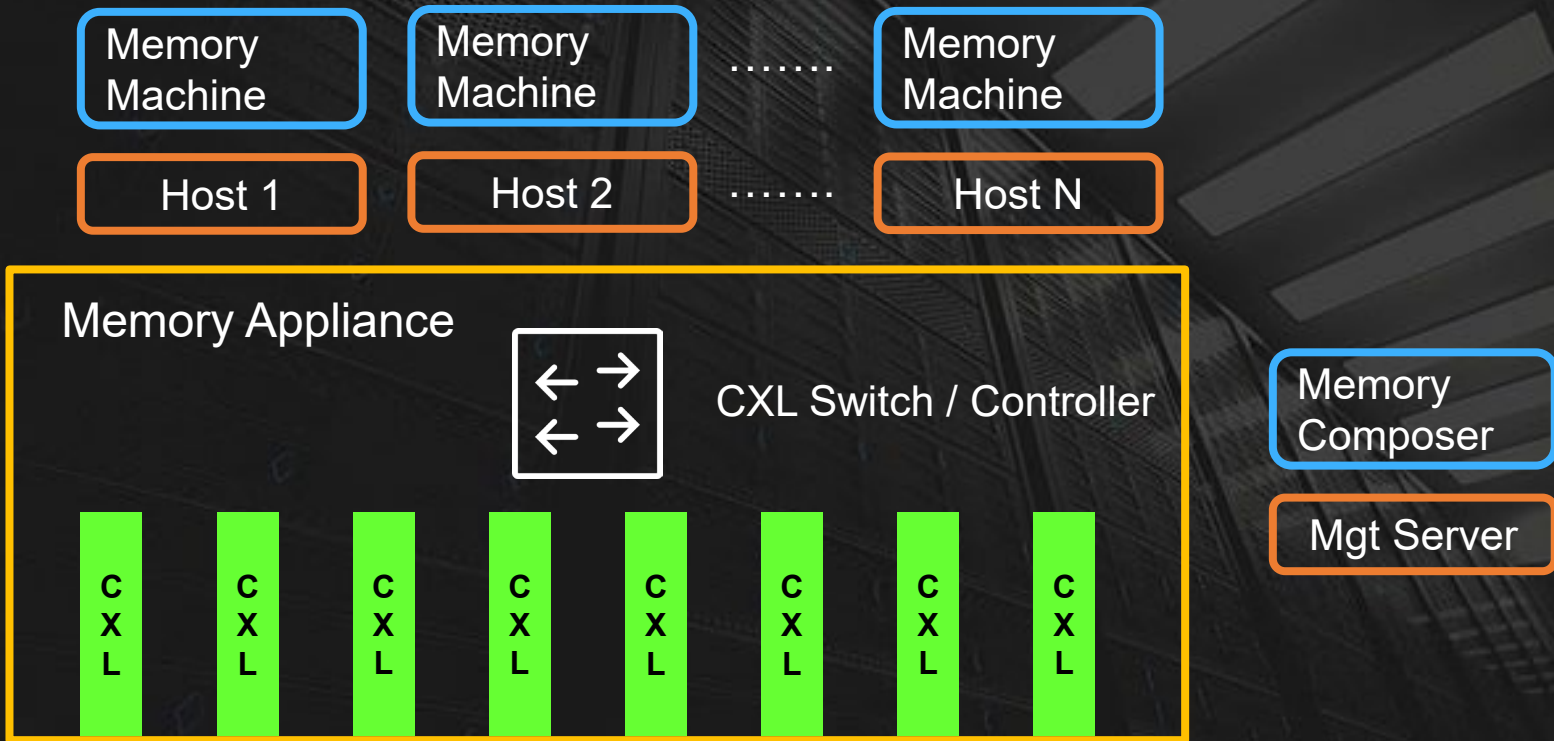
**Target Availability  
in Q3 2023!**



# 3. Elastic Memory



# CXL 2.0 Memory Pool

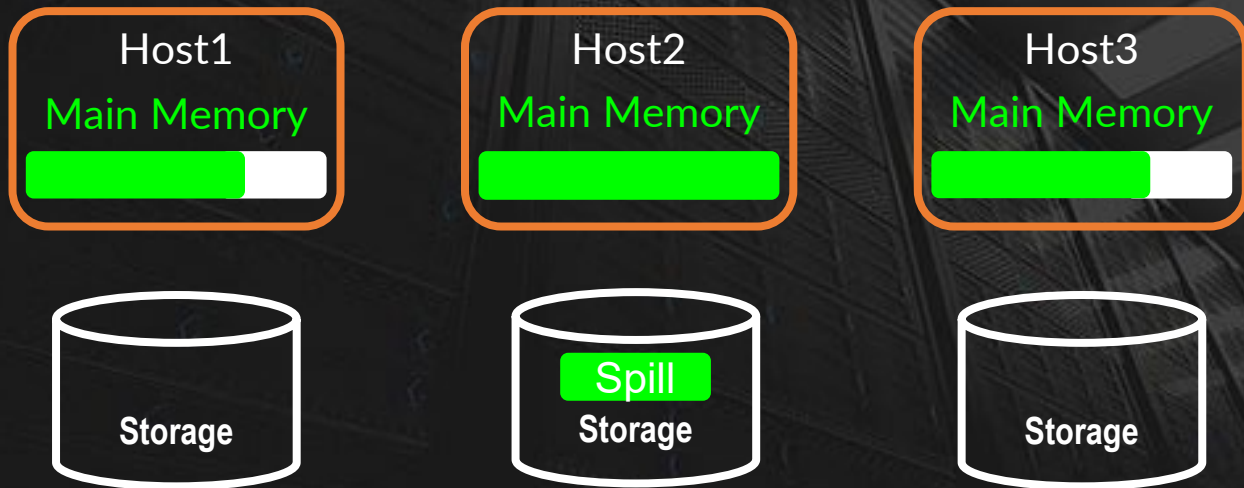




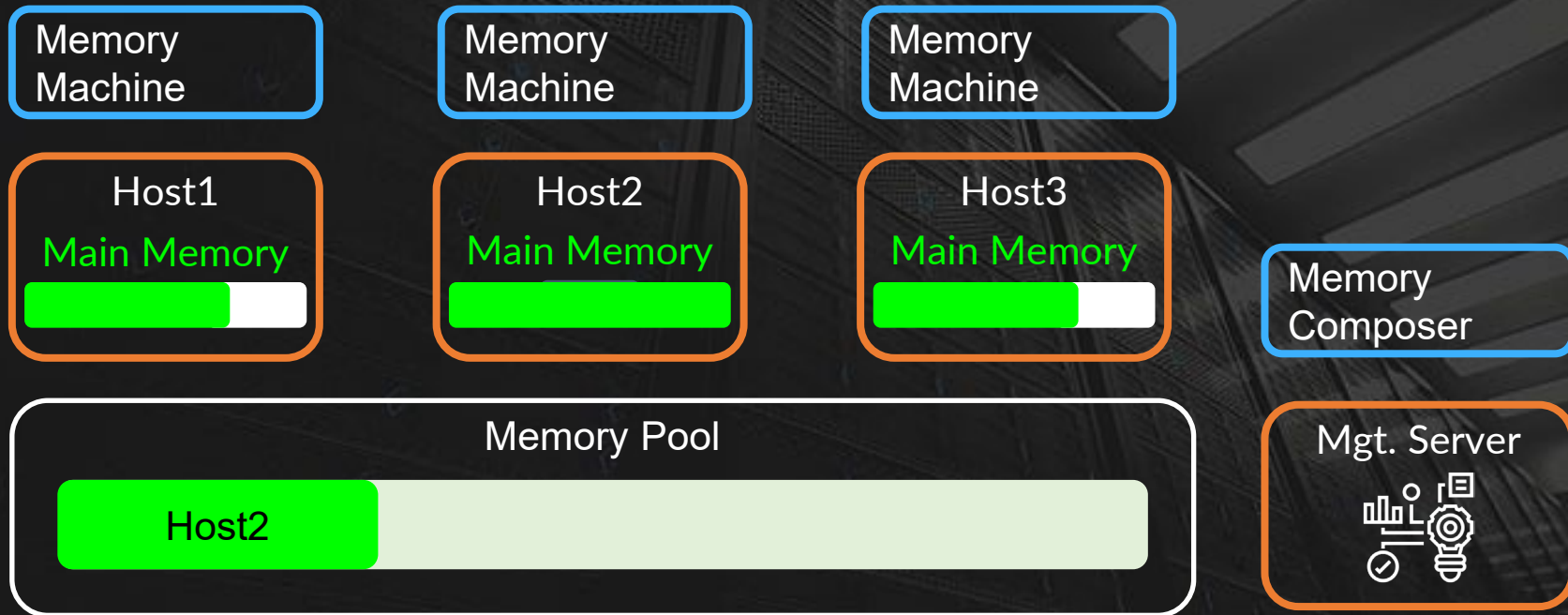
# What Used To Be a Joke is Now Real



# Data Skew, Out-of-Memory, and Spill-to-Storage

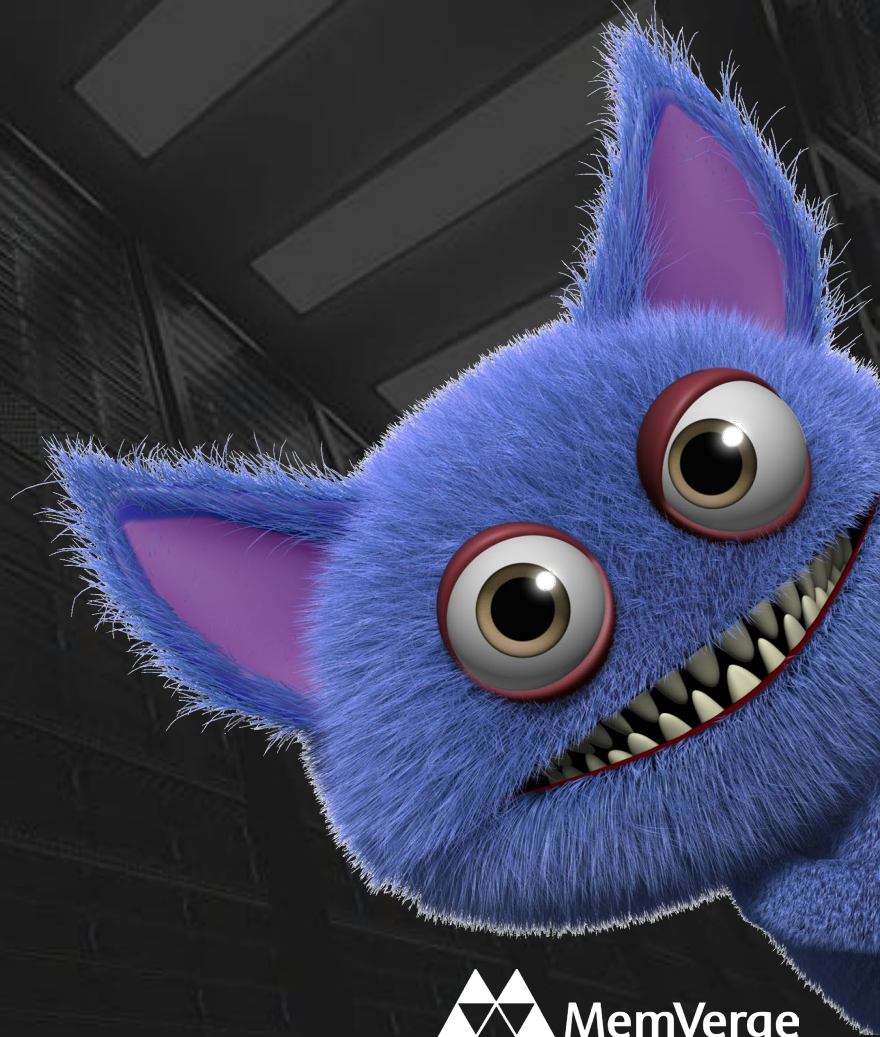


# Elastic Memory with CXL Pooling



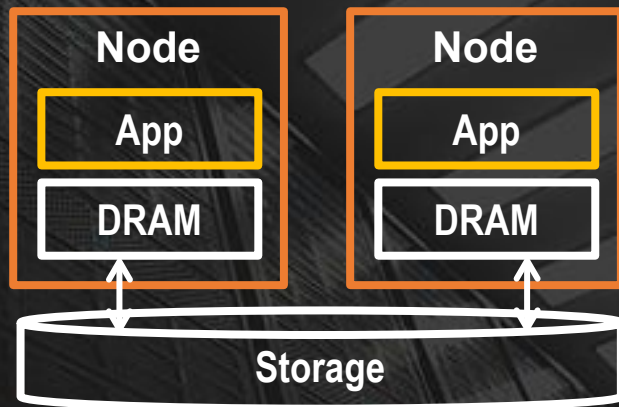
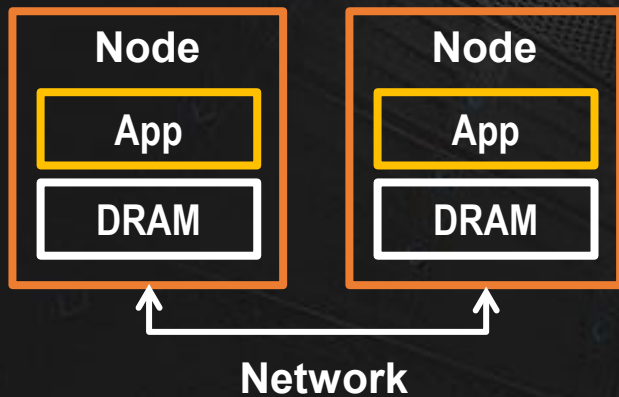
# 4. Project Gismo

(Global IO-free Shared Memory Object)





# The I/O Wall in Distributed Applications



**Serialization & Deserialization**



**Data Copies**

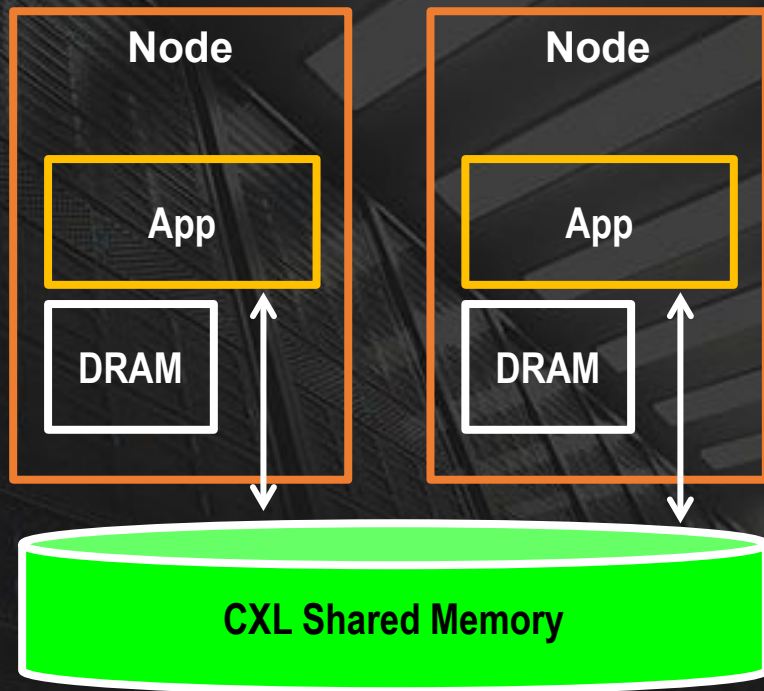


**Media Speed**



# CXL Makes Cross-Node Shared Memory Possible!

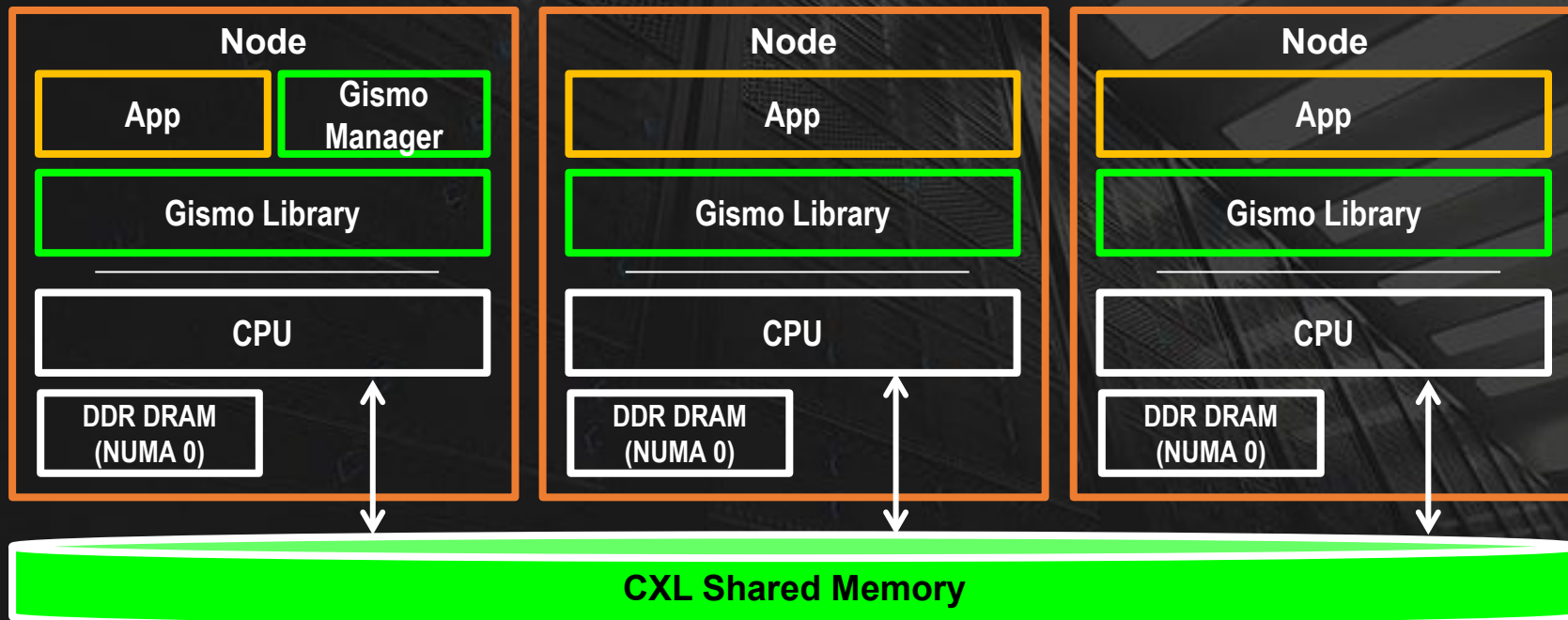
- Multiple nodes have direct memory access to the same CXL memory region
- Cache Coherence is part of CXL 3.0 hardware specification
- Software cache coherence can be implemented on top of CXL 2.0 hardware





# Introducing Project Gismo

Global I/O-free Shared Memory Objects

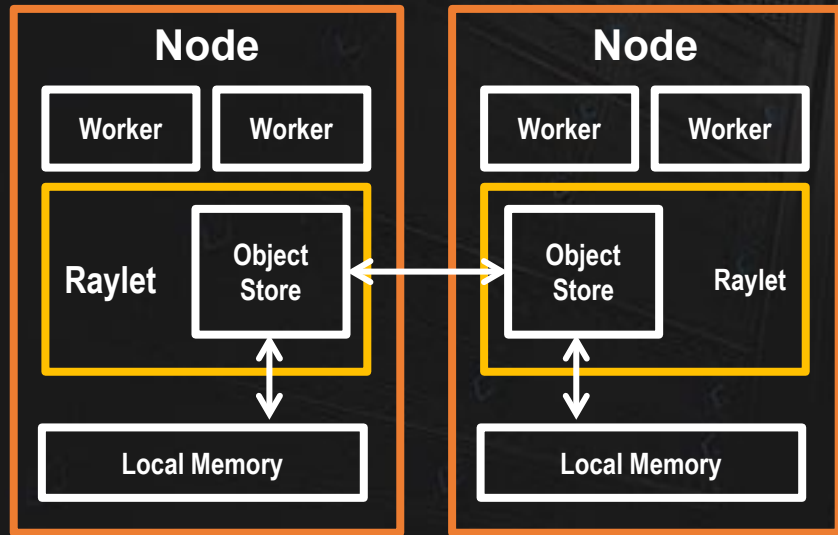


# Use Case 1: AI/ML



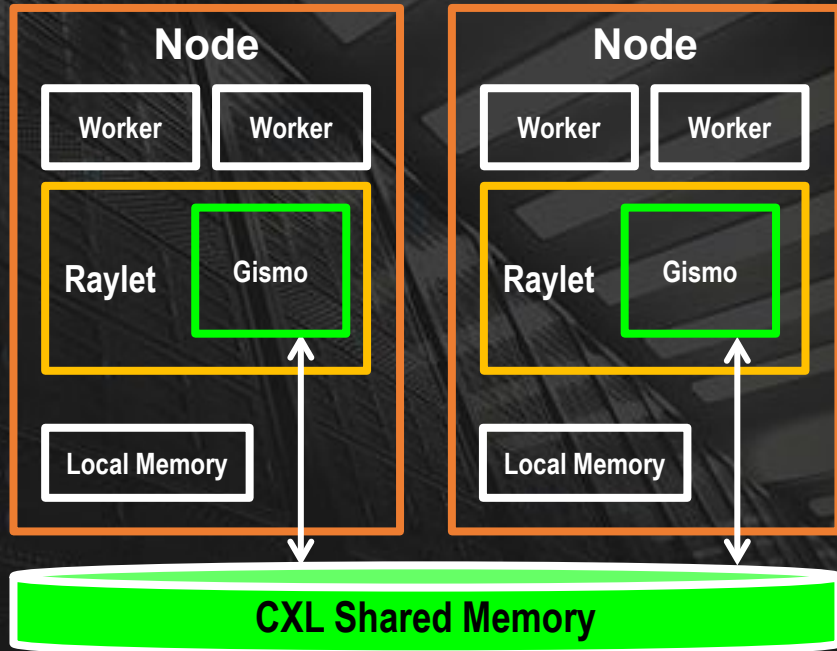
# RAY

Baseline Ray



Serialize & copy over network

Ray + Gismo





# Shuffle Benchmark Results

	Baseline Ray	Ray + Gismo
Local Get 1GB object	0.4 sec	0.4 sec - <b>CXL shared memory as fast as local memory</b>
Remote Get 1GB object	2.7 sec	0.4 sec - <b>675% faster</b>
Shuffle 50GB 4 nodes, each 4 cores, 128 GB object store	515 sec	185 sec - <b>280% faster</b>

\* Running in emulation environment

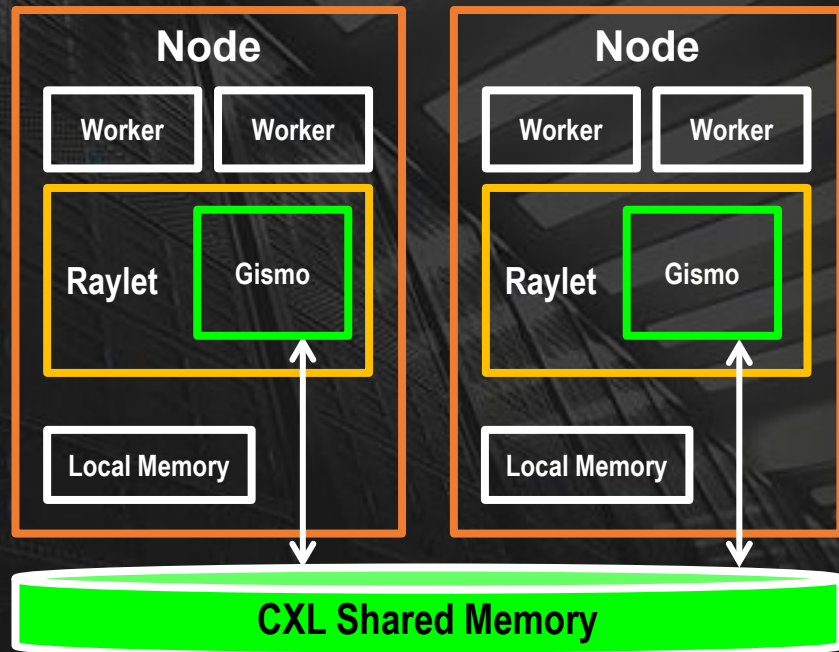
# Benefits of Ray + Gismo



**IO-free:** Eliminates object serialization and transfers over network for remote object access

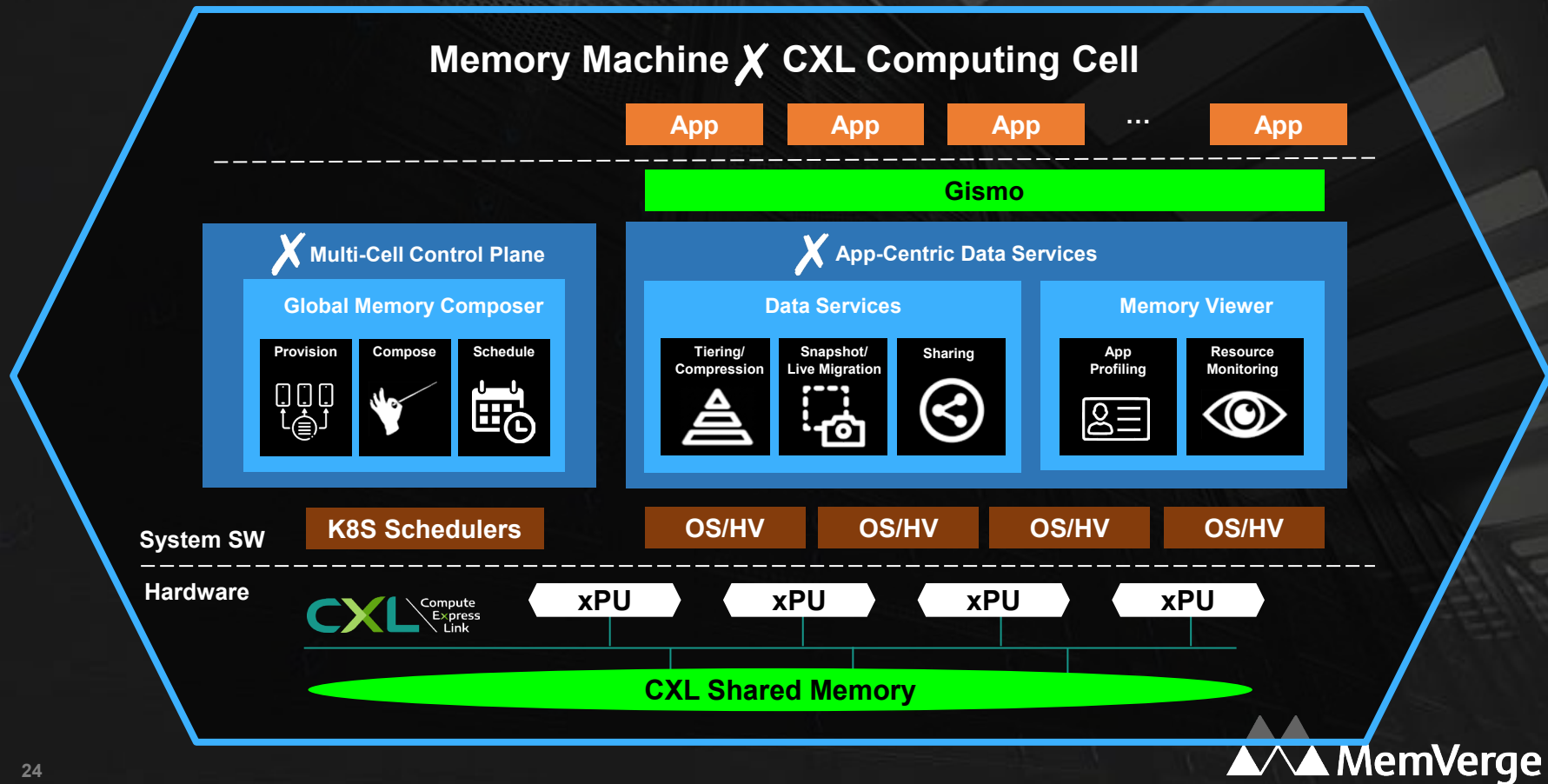
**Zero Copy:** No more duplicate object copies on different nodes

**No Spilling:** Reduces object spilling and data skewing with each node accessing memory pool



# 5. Computing Cell

# Post-Server Memory-Centric Infrastructure





# Collaboration with the CXL Ecosystem



SAMSUNG



intel

arm



Memory Machine



Computing Cell

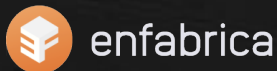


Lenovo

Rambus



NetApp



Alibaba Cloud



# Joint Solution Announcement



**SAMSUNG**



**Memory Machine X**  
Computing Cell

# Joint Solution Announcement



**Memory Machine X**  
Computing Cell



# Joint Solution Announcement



LIGHTELLIGENCE

**Memory Machine X**

Computing Cell

# MemVerge: Leading Big Memory Software Since 2017

**Built World's 1st Software-Defined Memory Product**

**Memory Machine™**

---

**Founders were creators of \$B product lines**



**Raised ~\$100M of Investment**

**SAMSUNG**

intel  
capital

cisco  
investments

SK hynix

NetApp

---

**Global Customers**

**VISA**



Tencent Cloud

**MEMX**



ByteDance

INTESA



SANPAOLO

tgen  
part of City of Hope



Analytical  
Biosciences



PennState



**MemVerge**



# Thank you



Flash Memory Summit

