Using Pools of Shared Resources to Lower Latency and Improve System Performance

ENDLESS MEMORY
CXL™ FORUM at OCP Summit

George Apostol, Founder & CEO

Elastics.cloud

EMPOWERING OPEN.

OCP GLOBAL SUMMIT

OCTOBER 18-20, 2022
SAN JOSE, CA

Elastics.cloud Profile

Founded in **2020**

1st to demonstrate symmetric memory pooling

5 patents filed

60+ engineers worldwide

Over **1500+** total years of product development experience

Founders

George Apostol

CEO & Founder

- 35 years of experience designing system-on-chip (SoC), hardware, software, and systems.
- Leadership and executive roles at Xerox/PARC, Sun, SGI, LSI Logic, Exar, Samsung, TiVo, BRECIS.
- Holds several patents for interconnect and interface design .

Shreyas Shah

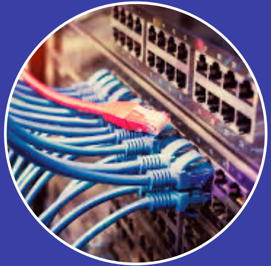
CTO, Chief Scientist & Founder

- 25 years of experience in the design of semiconductor, system design, and architecture in fields of computing, networking, storage technologies, virtualization and Flash based storage.
- Over 15 patents issued and numerous pending.

Data, Compute, & Interconnect Challenges

Data

- Exponentially increasing data
- Need to monetize data while managing TCO
- Increasing need to tier data based on varying compute requirements



Interconnect

- PCIe is higher latency
- Siloed, stranded, and underutilized resources
- Limited architectural options due to PCIe constraints
- Unnecessary bottlenecks

Compute

- Increasingly diverse workloads, latency demands
- Insufficient memory per core
- Excessive copying of data into and out of processor memory

Elastics.cloud Solution Benefits

Distributed CONTROL

Distributed control of resources and resource pools

Workload-centric instead of server-centric

Total COST Savings

Eliminate stranded resources

Memory and power efficiency

Improved system performance

COMPOSABILITY

Heterogeneous environments

Flexibility

Resource pooling

CONNECTIVITY

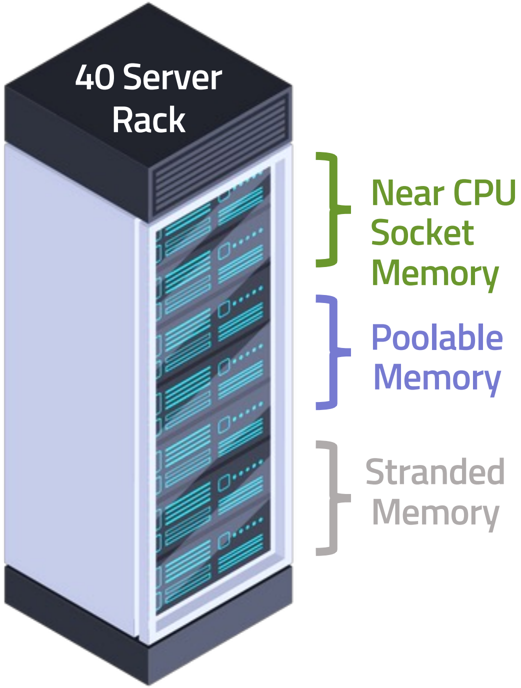
Lower latency

Fewer hops

More operations per second

TCO: Cost Savings with Memory Pooling

ENDLESS MEMORY CXL™ FORUM at OCP Summit



	Without Pooling	With Pooling
Near CPU Socket Memory (6,800 GB/Server = 1/3 of the memory)	$217,600	$217,600
Pool-able Memory (6,800 GB/Server = 1/3 of the memory)	$217,600	$261,120* *20% CXL premium
Stranded Memory (6,800 GB/Server = 1/3 of the memory)	$217,600	$0
Power	$60,000	$40,000
Total Cost	$712,800	$518,720

- Better Utilization
- Increased Capacity
- Improved Performance
- Lower TCO

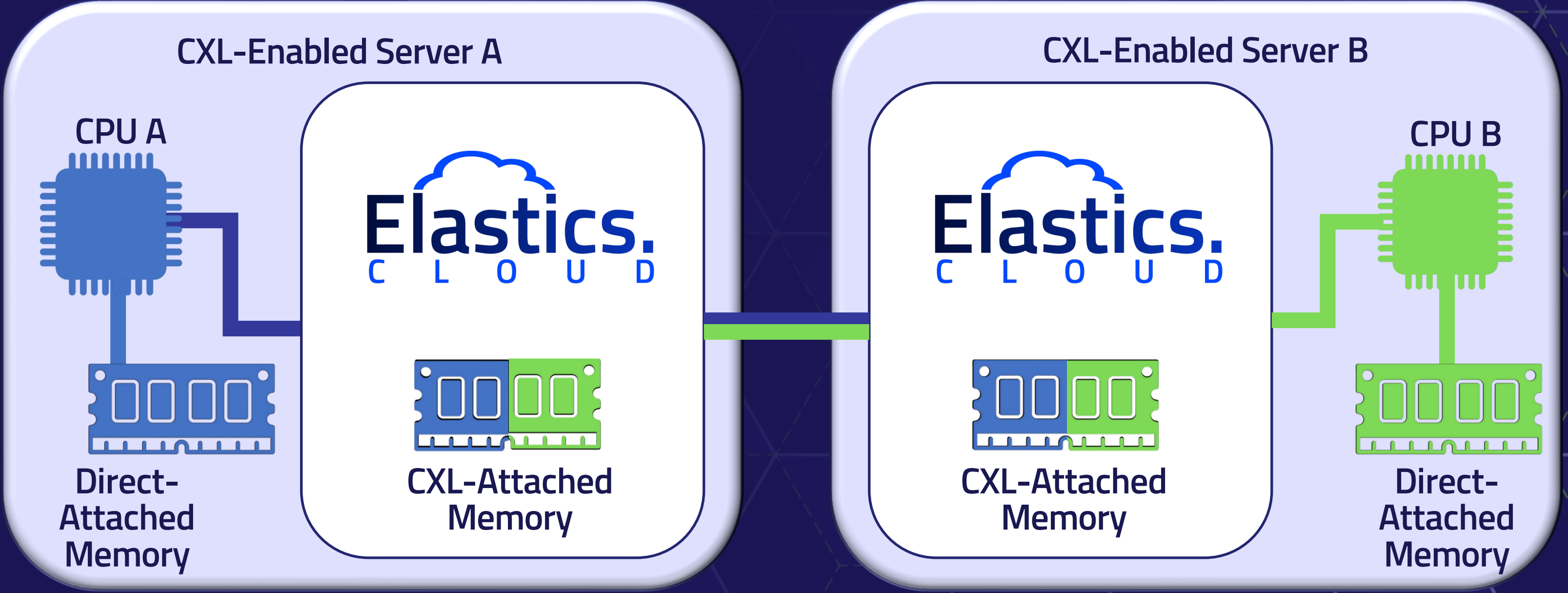
$194,080 Savings

Based on a 40 server rack

$80B memory and power cost savings (400K racks/yr)

- 512 GB/server
- 40 server/rack
- 20 TB total memory per rack
- $32/GB

Elastics.cloud First to Demonstrate Symmetric Memory Pooling



Extendable to multiple servers or resource pool appliances

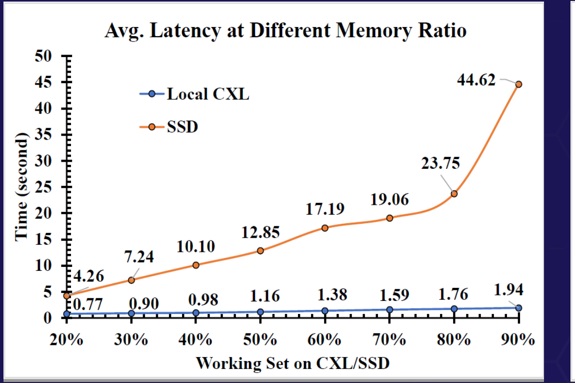
Redis Labs DB: Swap Space SSD vs. CXL-Attached Memory

ENDLESS MEMORY
CXL™ FORUM at OCP Summit

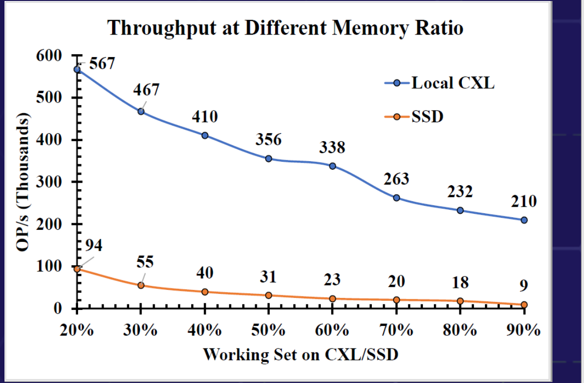
Latency
up to 22x lower

Ops/sec
up to 23x higher

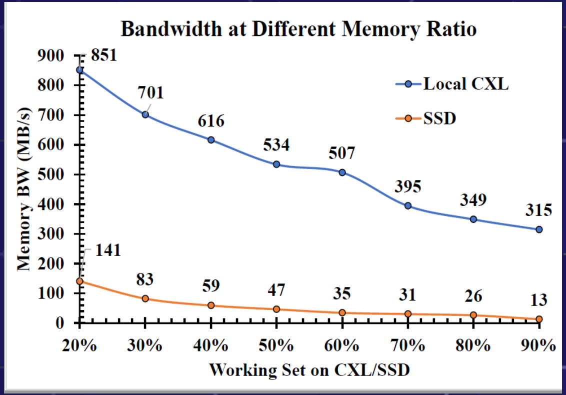
Bandwidth
up to 24x higher



Average latency (seconds)/operation



Operations/sec



Memory bandwidth MB/sec

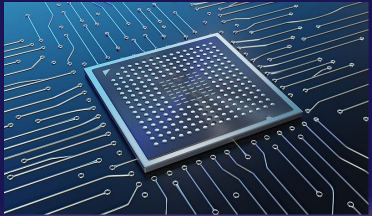
Datasets increase in size, swap space SSD vs CXL - 20% to 90%

Elastics.cloud Composability Solutions

ENDLESS MEMORY
CXL™ FORUM at OCP Summit

Current

AMD/Intel FPGA
running Elastics.cloud IP

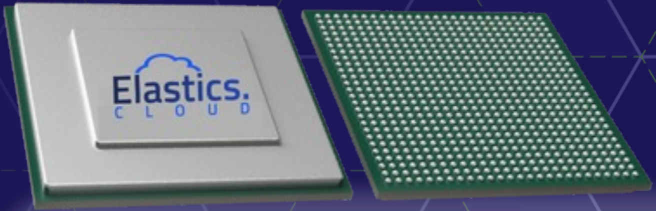


Elastics.
CLOUD
IP

- CXL standards-based interface
- Memory expansion
- Memory pooling
- Accelerator pooling

Future

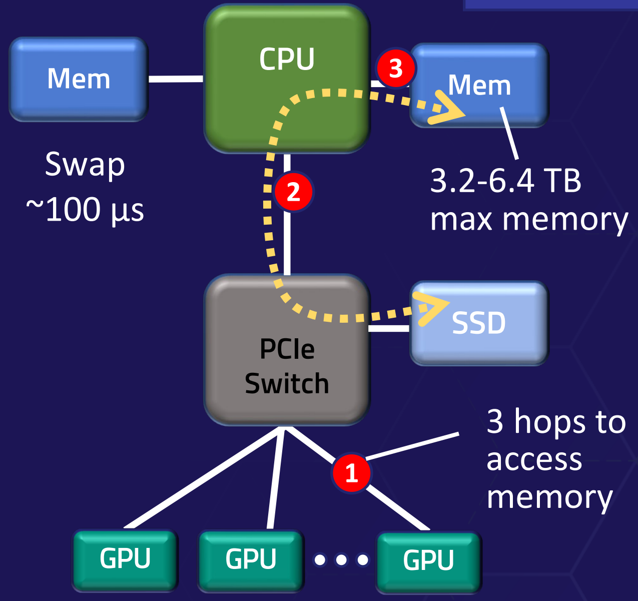
Elastics.cloud SSoC
(Switch System on Chip)



- CXL standards-based interface
- Memory expansion
- Memory pooling
- Accelerator pooling
- Full composability
- Added features & intelligence

Memory Expansion & Pooling Performance Improvements

Current - PCIe

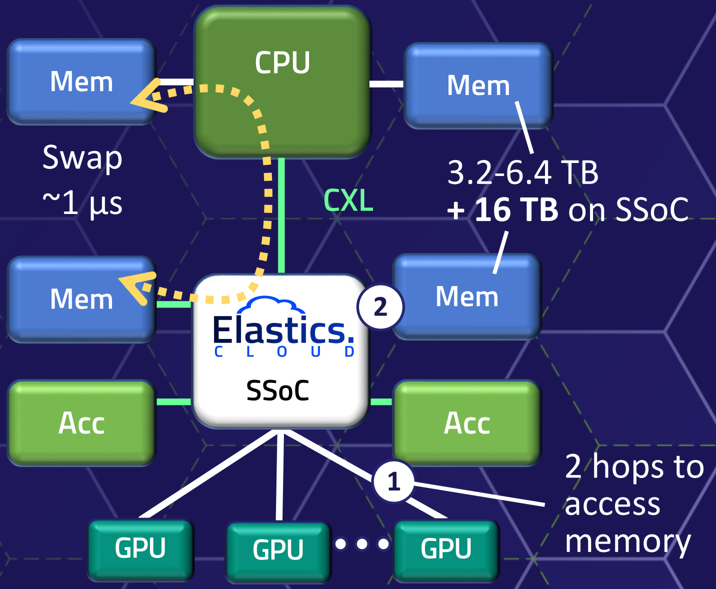


Lower Latency
Higher Ops/Sec
Increased Memory Size & BW
Fewer Hops

Hardware swap latency: ~100 µs to ~1 µs

DB App: up to ~23x more ops/sec, 22x lower latency

Memory Expansion with CXL Peer to Peer Transfers



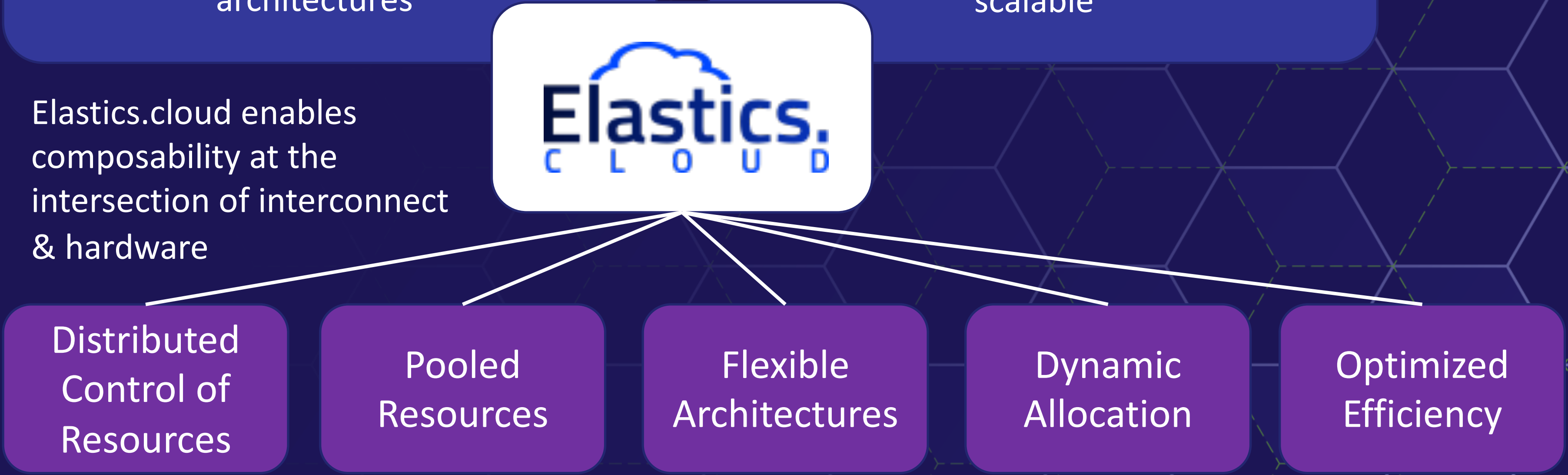
Beyond Interconnects: CXL + OCP

ENDLESS MEMORY
CXL™ FORUM at OCP Summit

CXL is a breakthrough interconnect supporting composable, heterogeneous architectures

OCP is designing hardware that is more efficient, flexible, and scalable

Elastics.cloud enables composability at the intersection of interconnect & hardware



Elastics.cloud

Distributed Control of Resources

Pooled Resources

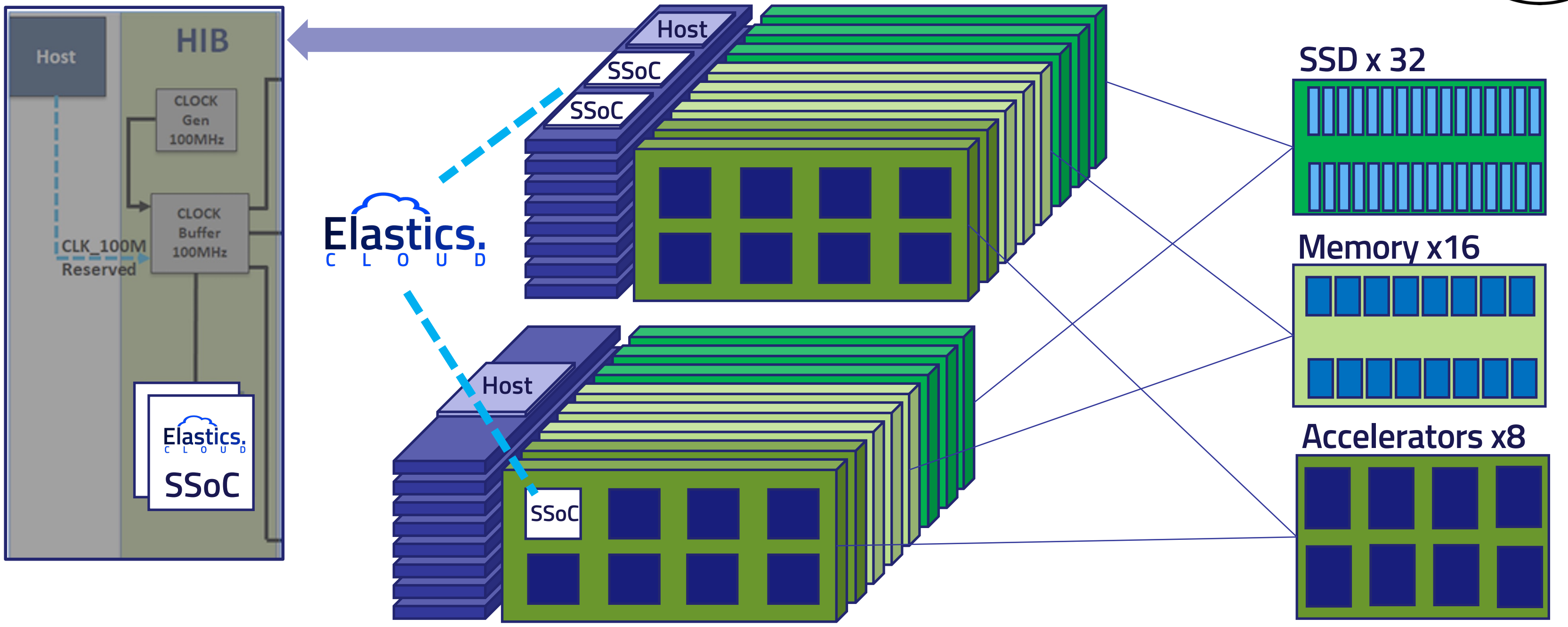
Flexible Architectures

Dynamic Allocation

Optimized Efficiency

Composable Chassis with Elastics.cloud SSoC and OCP Building Blocks

ENDLESS MEMORY CXL™ FORUM at OCP Summit



Conclusions

1. CXL protocol combined with OCP standards will provide new levels of hardware and interconnect composability, driving lower TCO and better system utilization

2. CXL will provide improved performance and lower latency than PCIe

3. Advantages are achieved above what the connectivity provides for systems applications like in-memory databases

4. Resource sharing (i.e. memory, accelerators, storage) improves utilization and performance

5. Elastics.cloud patent-pending solutions provide additional systems-level features further improving performance and lowering TCO at scale

Elastics.
CLOUD

ENDLESS MEMORY
CXL™ FORUM at OCP Summit

Thank you

EMPOWERING OPEN.

OCP
GLOBAL
SUMMIT

OCTOBER 18-20, 2022
SAN JOSE, CA