Elastics.
CLOUD

Company Introduction

Composable System Platforms Leveraging CXL

Elastics.cloud
Profile

Founded:	**2020**
Funding to date: *Pre Series A*	**$26m**
Patents filed:	**5**
Total product development experience years:	**1,500+**
Engineers worldwide:	**65+**

Founders

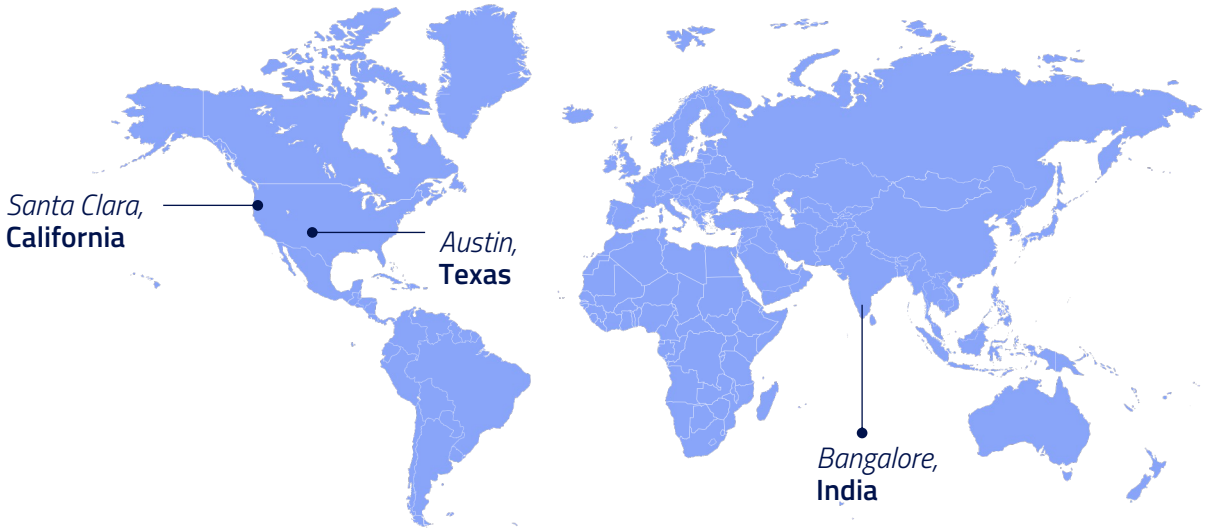
George Apostol, CEO & Founder

- *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, PLX Technology, 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*

Locations

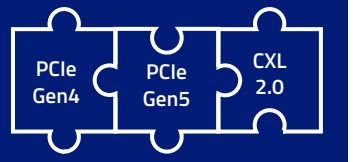


Industry/Market Pain Points

Addressed by Elastics.cloud

Backward Compatible with PCIe

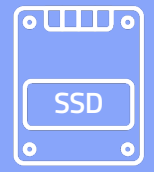
- *Hybrid Switch with support of PCIe Gen1 – Gen5 and CXL 1.1/2.0*
- *CXL 3.0 features with COE*



PCIe/CXL Over PCIe/IB Cables: Bend Radius Limits Scalability

- *COE (CXL Over Ethernet™) to scale inside the Rack and Rack2Rack across 32 racks in aisle*
- *Scalability and resources shared/ pooled across 4,000+ servers*

Slower Media & Disaggregated Memory



- *Pre-fetching and caching to reduce the effect of slower media and dis-aggregated memory over the network*
- *Any CXL ports can be declared as cache port(s)*
- *Applications communicate with caches for faster responses*

Latency Affecting Application Performance

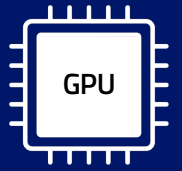
- *Latency one NUMA hop: No change in applications with memory on CXL with switch*

Extending Resource Sharing Across Multiple Fabric Elements

- *Memory Pooling: Inside the Rack and Rack2Rack across 32 racks in aisle*
- *Storage, Networking, GPU accelerators, AI accelerators, Accelerator2directmem™*

GPU Accessing Large Memory Pools Not Enabled with CXL 2.0

- *GPU on PCIe can't access CXL-attached memory without going through CPU*
  - *Increases latency, limited bandwidth through CPU*
- *Elastics.cloud Accelerator2directmem™ enables GPU to directly access CXL-attached memory*

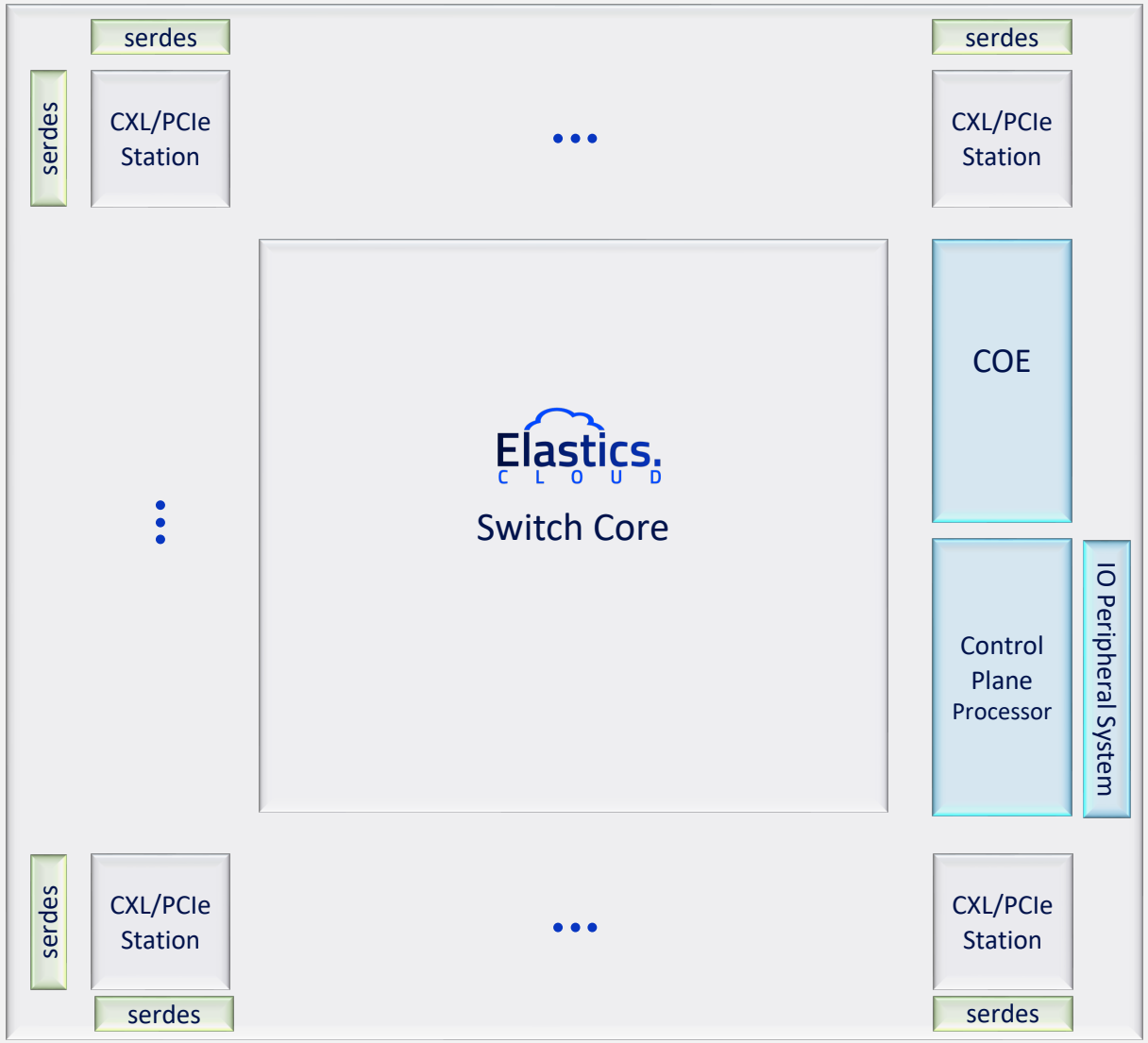


Elastics. CLOUD

Elastics.cloud

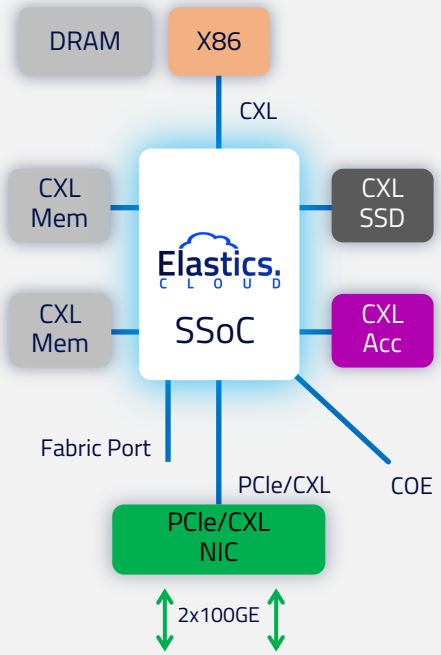
Switch System on Chip (SSoC)

SSoC (Switch System on Chip) - CXL 1.1 and 2.0	*Ethernet – Traffic management*
PCIe 5.0/CXL 1.1/2.0	*2x PCIe, I2C/I3C for FM running on BMC/control plane processor connectivity*
COE (patent pending)	*Non-blocking low latency switch core*
Control plane processor to run the FM API	*Controllability, observability, and RAS features throughout chip*

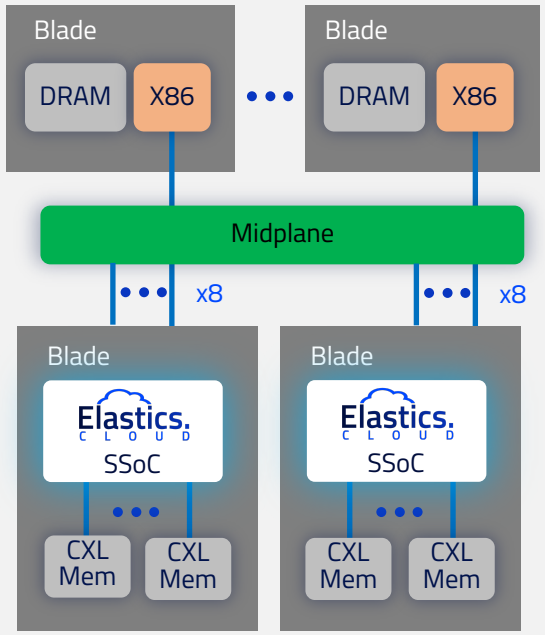


SSoC Memory and Resource Expansion and Pooling Use Cases

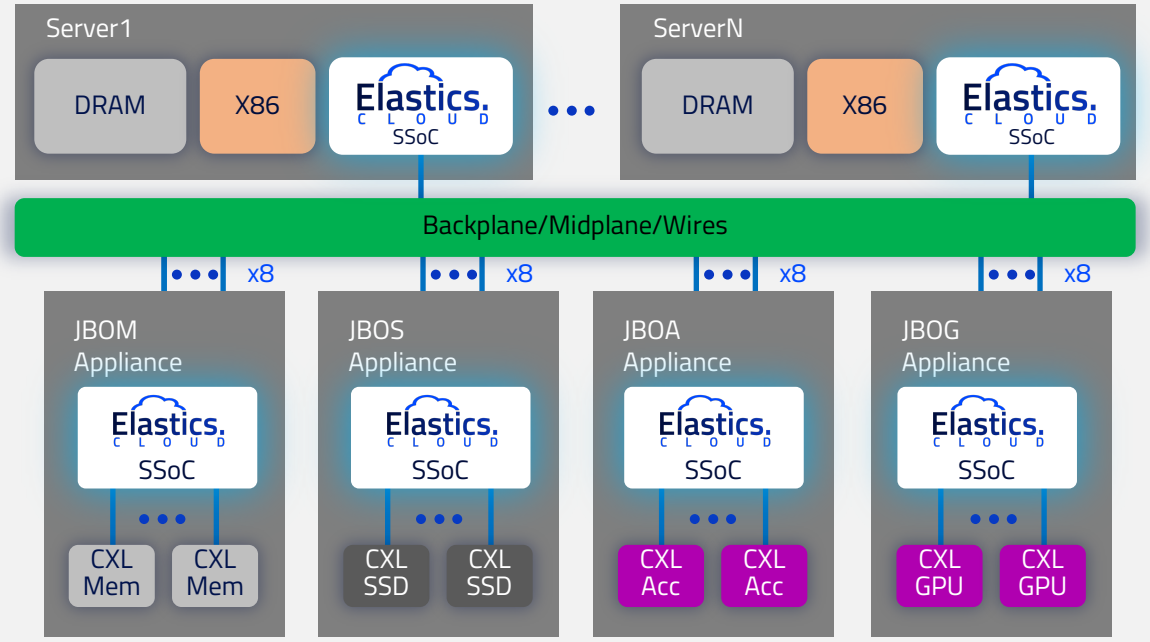
1. Memory Expansion in Single Socket Server



2. Memory Pooling in Single Chassis



3. Resource Pooling in Disaggregated Rack



PCIe Wire — CXL Wire ⟷ Ethernet

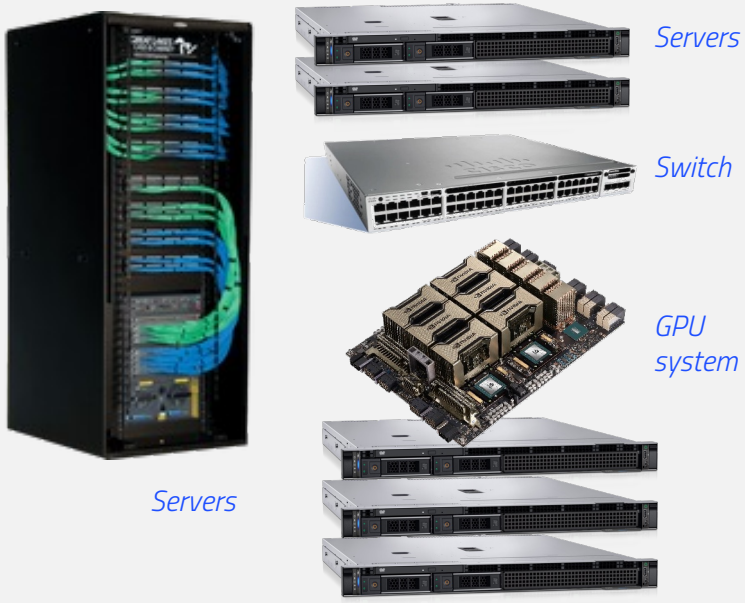
- *SSoC enables pooling of any resource type*
- *Disaggregation and pooling allow dynamic composition of server resources for a given workload*
- *Multiple hosts can share resources from resource pools*
- *Managed by a single management host/BMC*

Elastics. CLOUD

CXL Over Ethernet (COE)

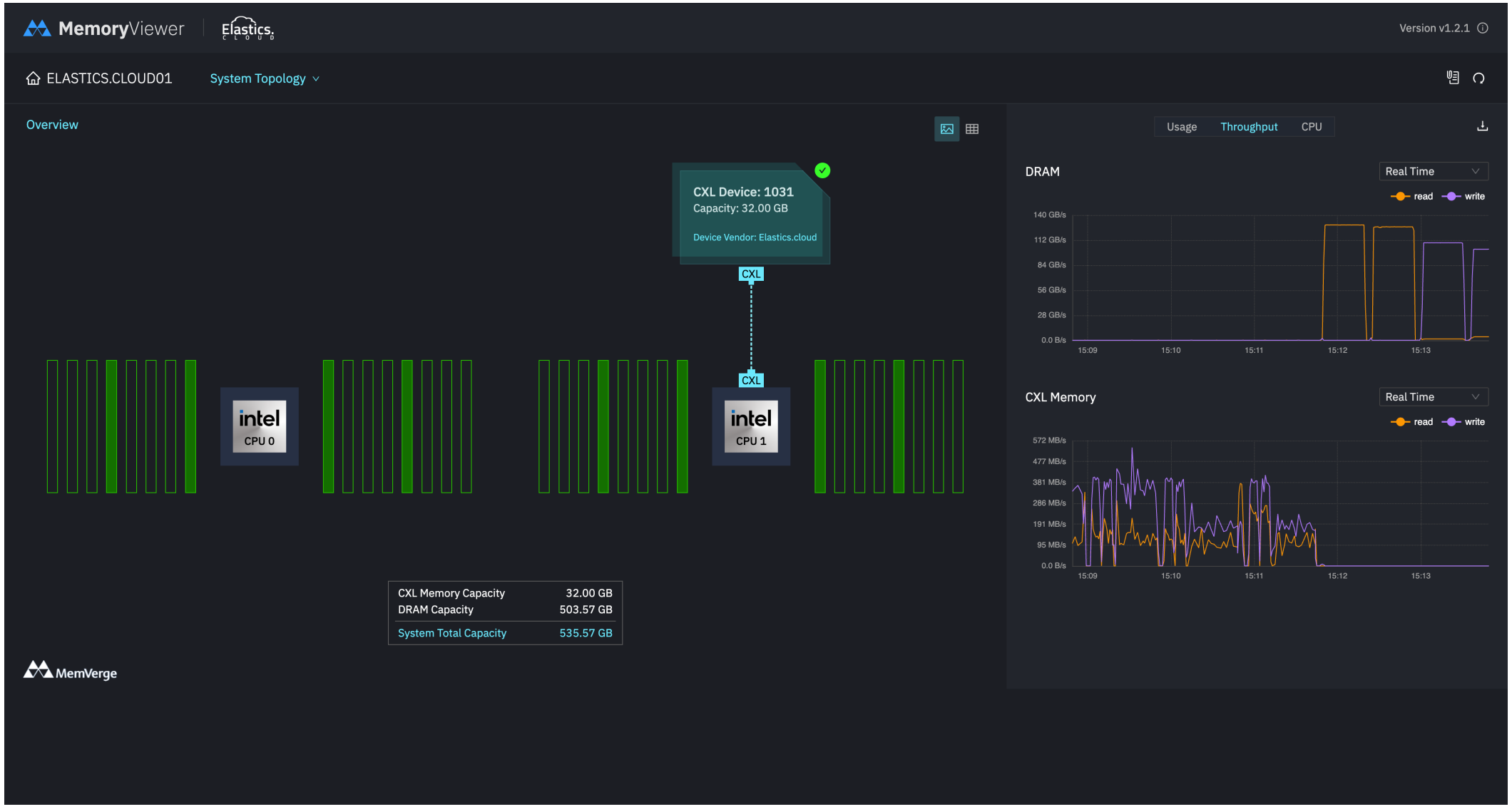
- *COE: CXL protocols over Ethernet*
- *Advantages*
  - *Seamless support for existing Ethernet infrastructure*
  - *Intra-rack and inter-rack reach*
  - *Supports existing Ethernet switches in the market and future low latency Ethernet switches*
- *Support for DCBx, QCN and replay buffers on each side of COE implementation for lossy Ethernet network*

Server Components Connected through COE



Elastics.cloud solution with MemVerge software

Memory Expansion and Pooling Demo with FPGA attached Memory



Elastics.
CLOUD

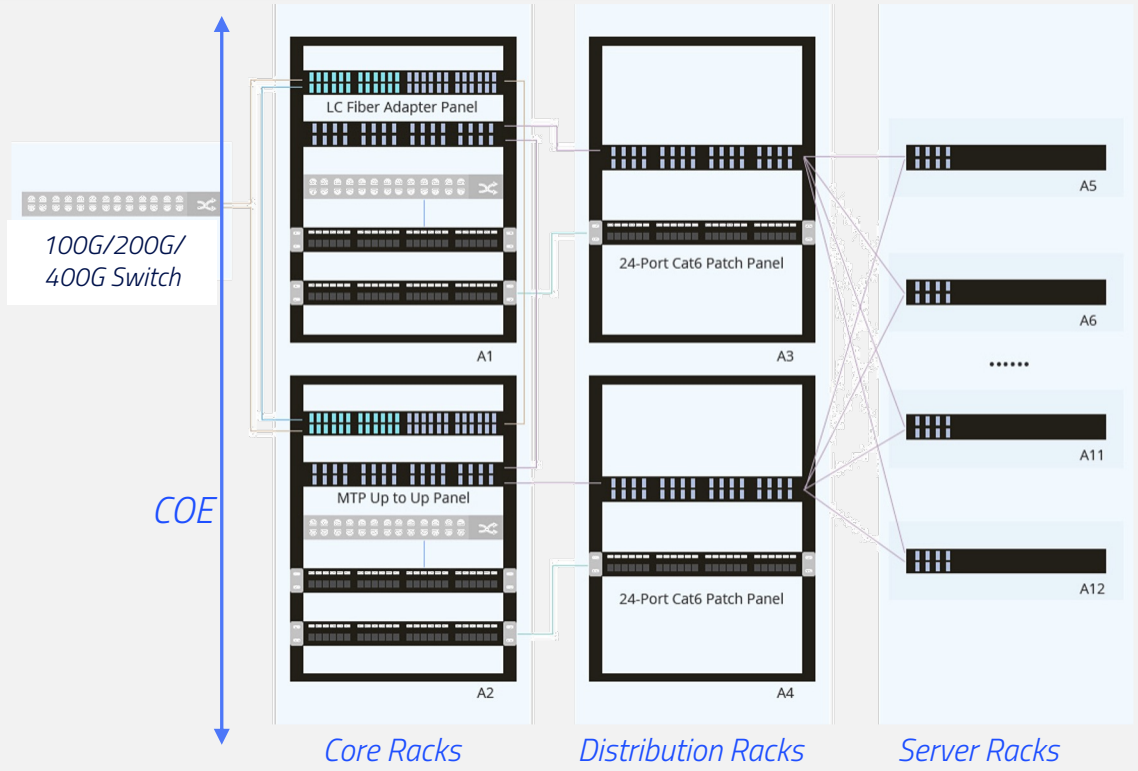
COE Implementation in Rack Systems

Composability within the rack

GPUs & Memories Used for ML/AI Training Clusters

Elastics.cloud's COE technology enables sharing of GPUs & Memories

Intra-Rack Communication with Elastics.cloud COE

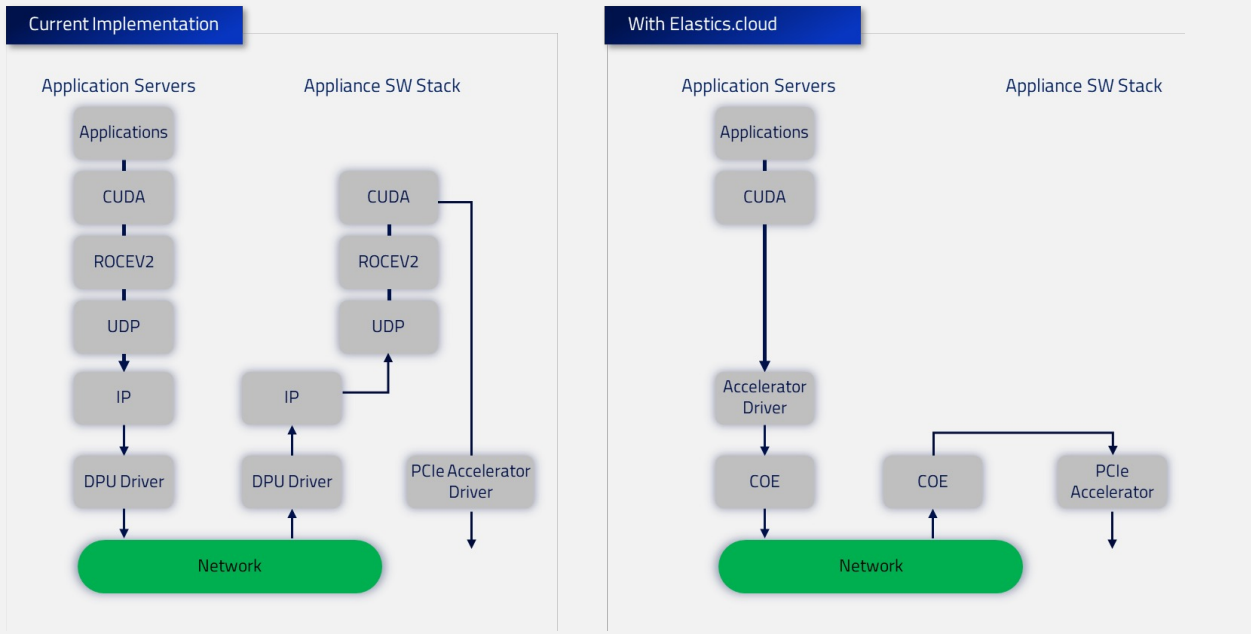


Ethernet wire eliminates cable bend radius limitations
COE enables resource sharing/memory pooling within the rack

COE Implementation in Rack Systems

Composability Rack2Rack

ML/AI Training Software Stack



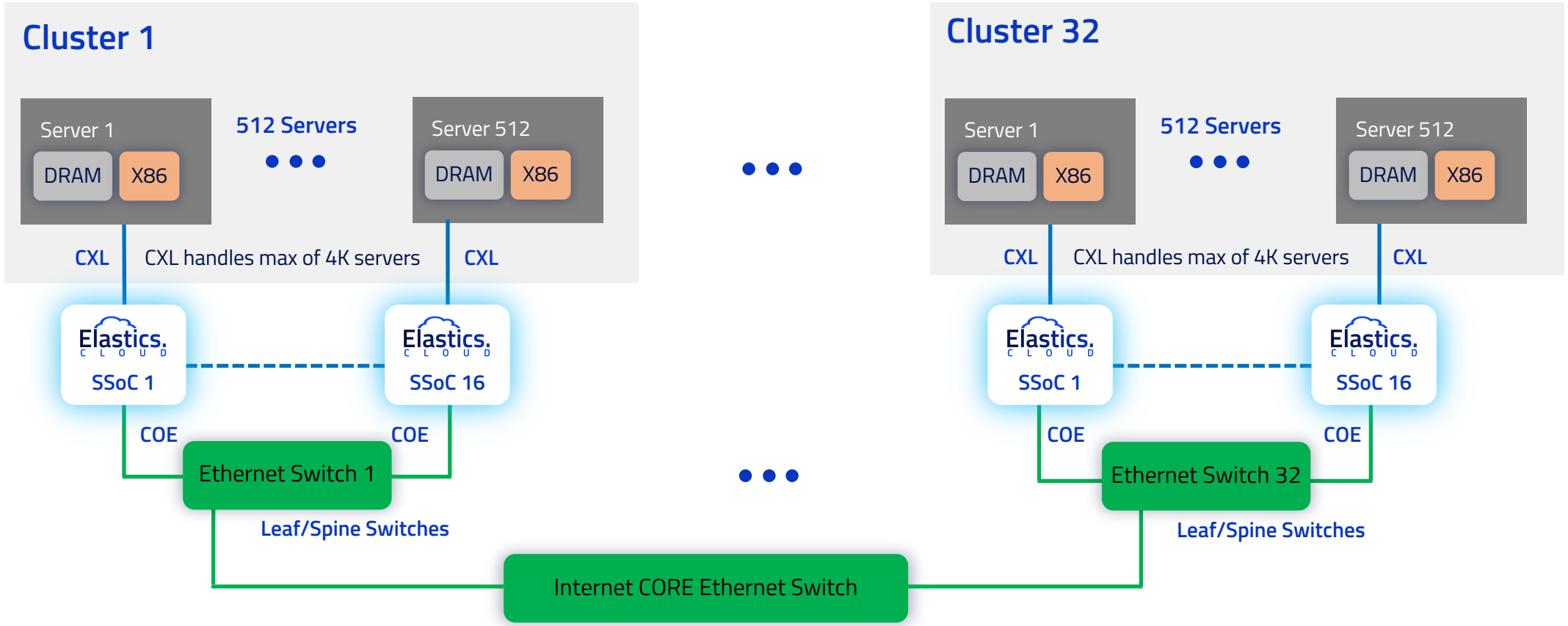
- *Reduces complexity of software – not going through CUDA over ROCEV2/V3*
- *Significantly reduces latency through software stack*
- *ReST API with Elastics.cloud extensions support multiple switches connected over COE*

Inter-Rack Communication with Elastics.cloud COE

- *Elastics.cloud's technology reduces the effect of slower media and disaggregated memory over the network*
- *FM API with Elastics.cloud extensions can be easily integrated with orchestrators to create specific configurations for dynamic workloads, e.g., Kubernetes*

Cluster Connectivity Using CXL and COE

16K servers connected



COE™ extends to hundreds of thousands of servers seamlessly

Summary

Elastics.cloud is highly focused on bringing CXL-based composable infrastructure to market

Key features of Elastics.cloud's solutions:

- *Backward compatibility with PCIe Gen5*
- *Lowest ball-to-ball latency in industry*
- *COE scales CXL inside the rack and rack-to-rack*
- *Solves industry pain points*

Elastics.cloud's solutions enable system level optimizations

- *System level build of JBOMs, JBODs, JBOx*
- *Dynamic assignment of pools of resources (CPU, memory, storage, accelerators) connected via CXL*
- *Simplified management infrastructure*

Elastics. CLOUD


Elastics.
CLOUD
Thank you
http://elastics.cloud/