XPU Based Cloud Native Server:

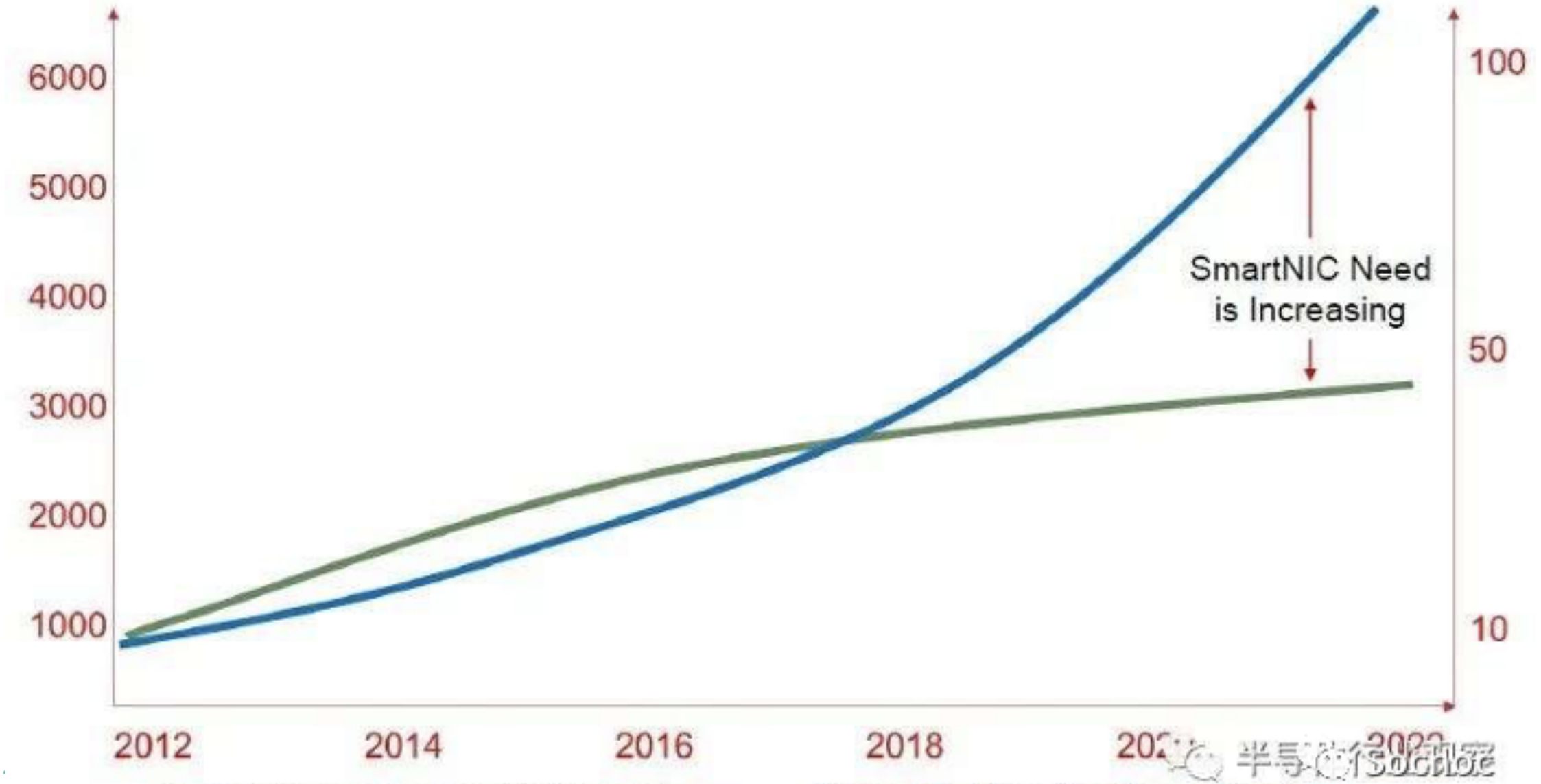
Architecture, Implementation & Applications

Dr. Fu Li (LEO)

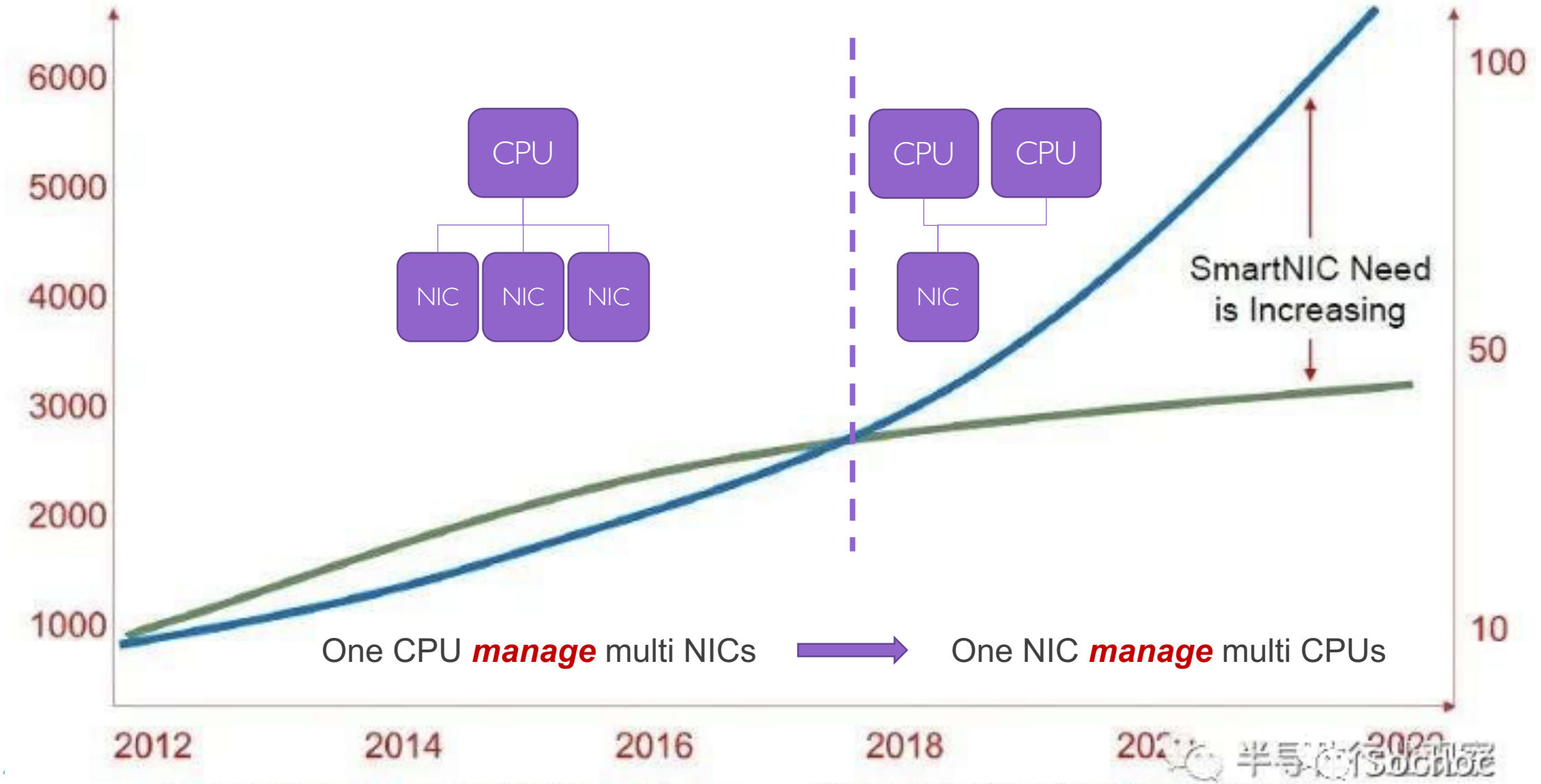
li@socnoc.ai

Socnoc AI Inc. 合肥边缘智芯科技有限公司

Problem to Address?

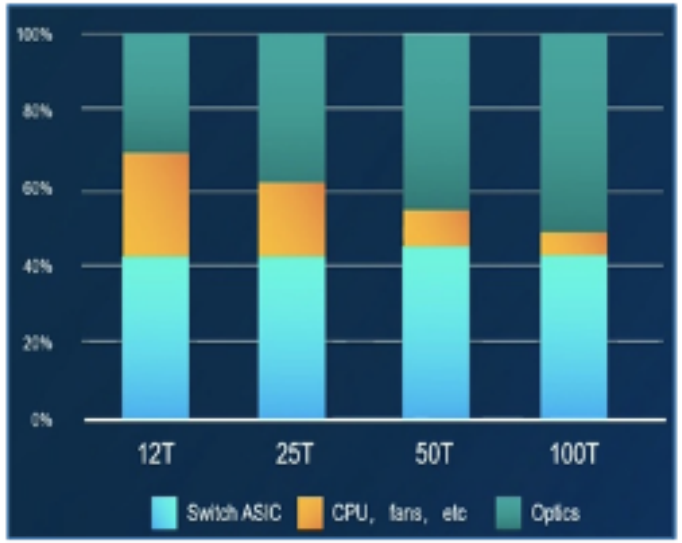


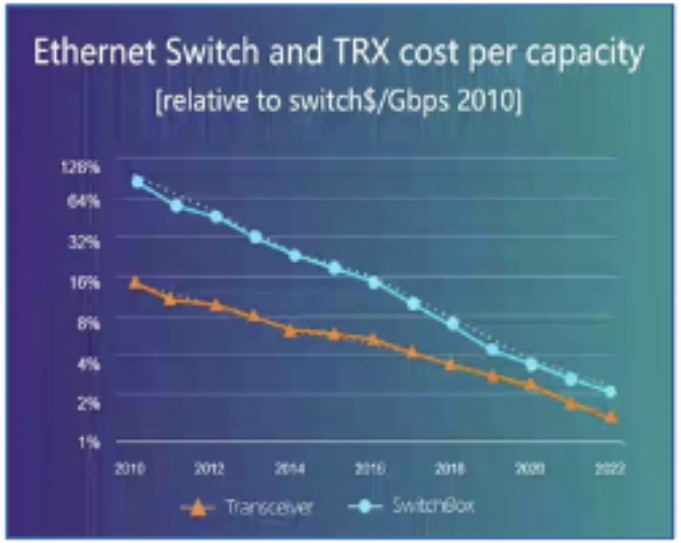
Problem Revisit

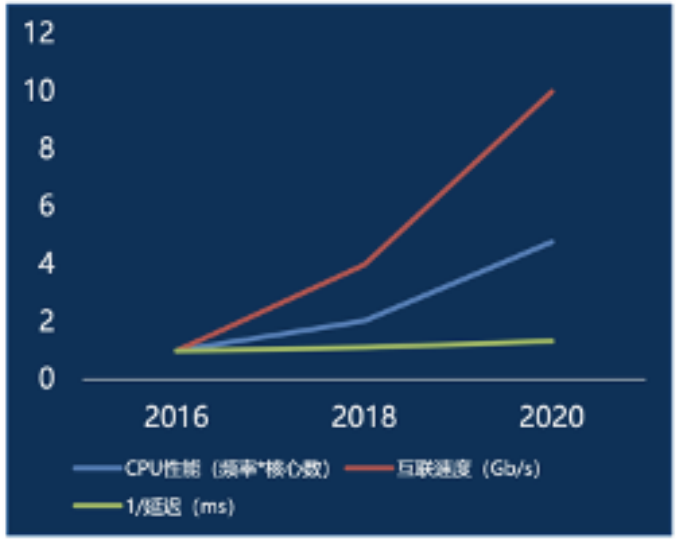


Real Problem Computing Cluster Faced!

› Optics are too expensive both in ***power*** and ***cost***







光模块功耗占比已经超过50%

光模块成本已经超过通道成本

延迟降低进展缓慢

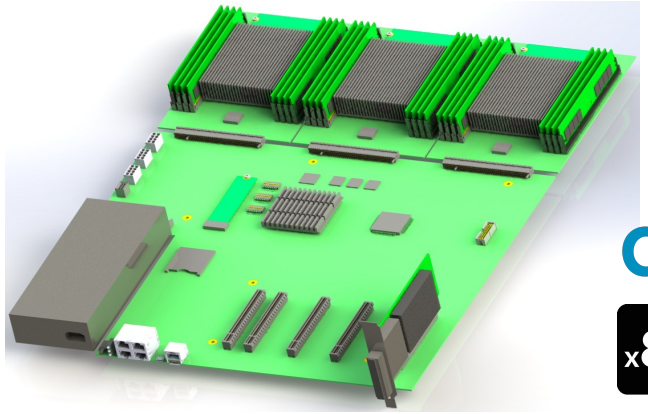
About SOCNOC: Kill Optics in Short Distance!

Socnoc AI (https://www.socnoc.ai) is the startup aiming to provide low-cost, green and high-performance networking solutions to smart clusters and edge cloud computing. Socnoc employ PCIe and CXL based interface with its own RDMA protocol to build the best infrastructure for networked and composable systems.

Our Goal:

- ***Reduce the network cost to $3 per Gbps host-to-host***
- ***Eliminate Optics and moving parts in Edge Clusters***

XPU



XPU Cloud Native Server

arm

x86

Data Era Connected by SOCNOC

SOCNOC consolidate PCIe with Ethernet through combination of SW & HW to change the server & server interconnection architecture of data center, leading the traditional server industry into the New Ear--Data Centered Cloud Native Computing Era.

DATA ERA connected by Socnoc

AKRAINO | WeOpen | NextArch FOUNDATION

Data Fabric Landscape

TCP/IP

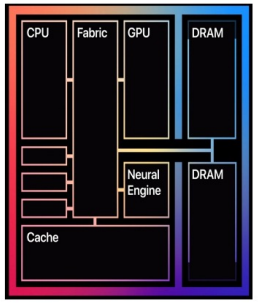
PCIe/CXL
cxl.io/cxl.cache

UCIe

Chiplet

I/O module

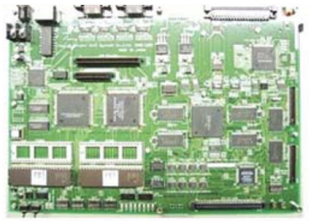
System-in-Package



Interposer Fabric

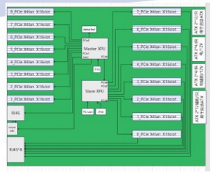
In-Package Links

System-on-Board



PCIe Bus based link

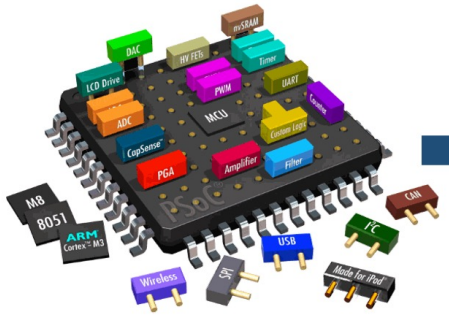
Cloud-on-Board AKRAINO



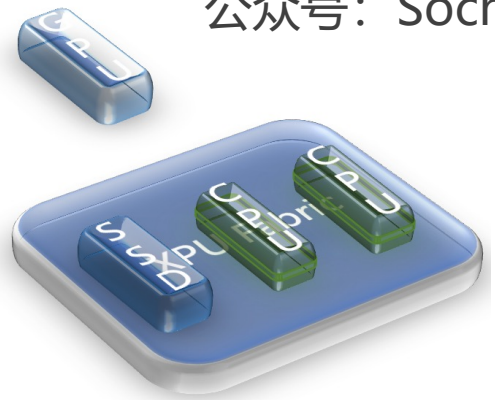
PCIe Net based Fabric

Off-Package Links

XPU: Center of Data Fabric



In Package Technology

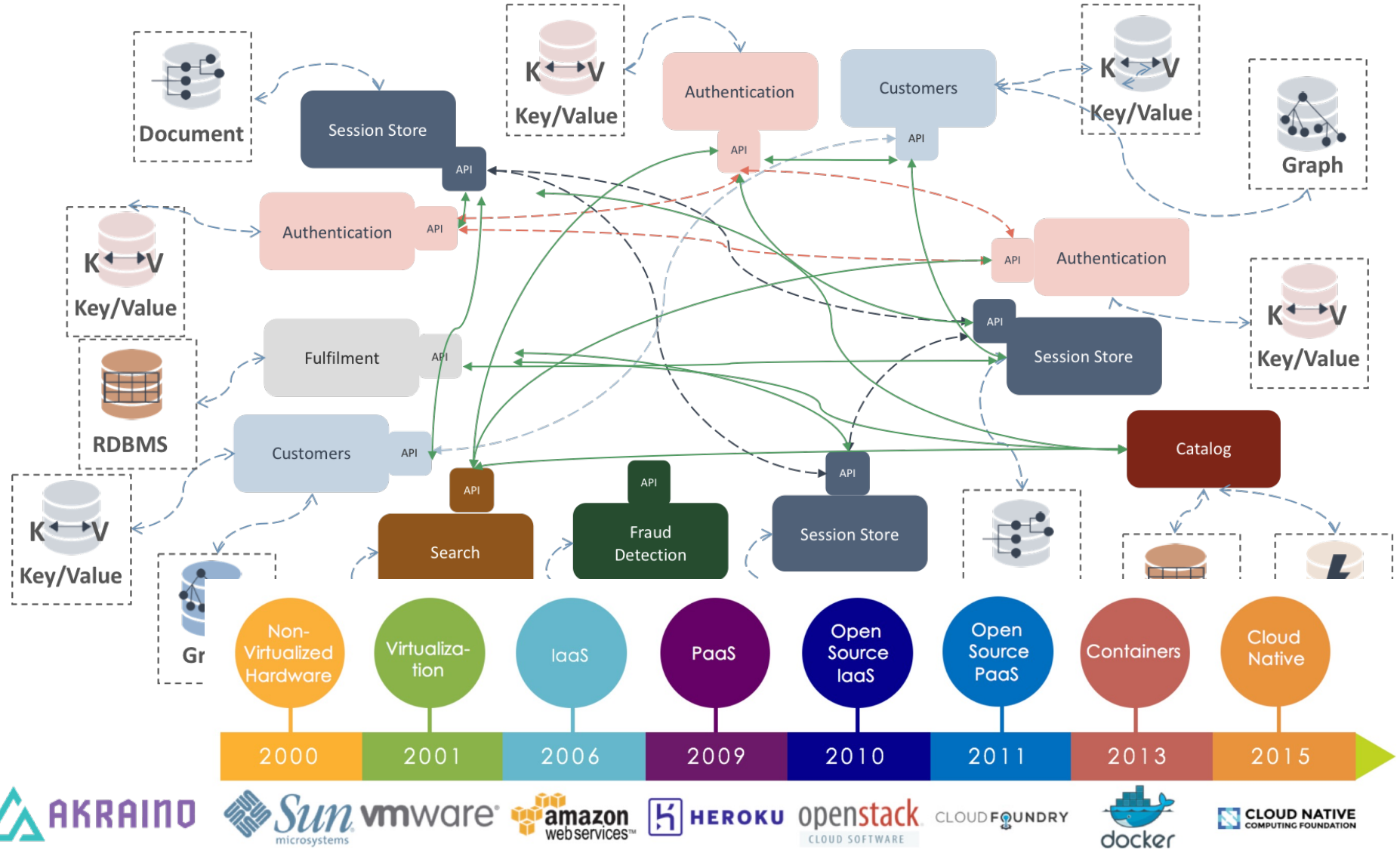


Off Package Technology

公众号：Socnoc

SOCNOC

Application Evolution: From Metal to API Ready!



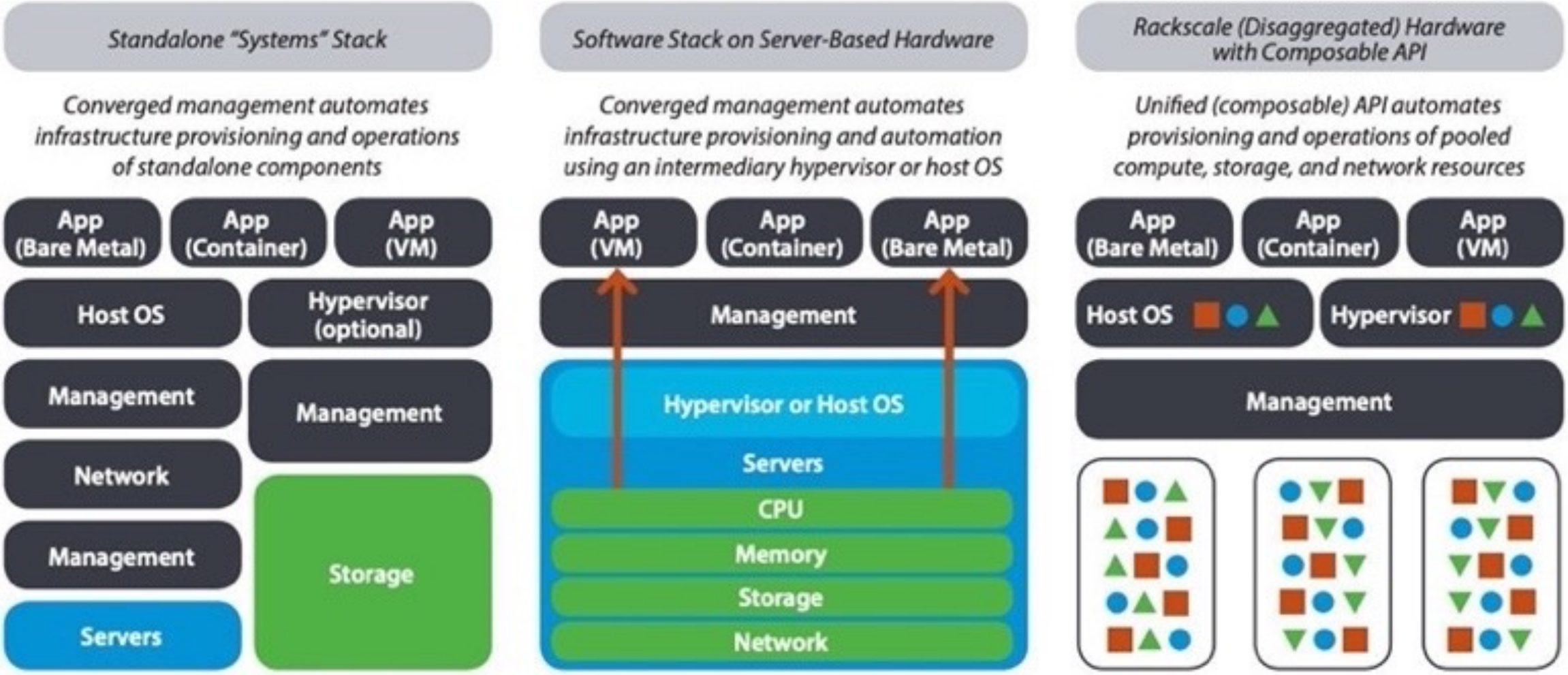
Metal Servers

Hardware-driven
BUS-driven

Serverless API

Software-driven
TCP/IP-driven

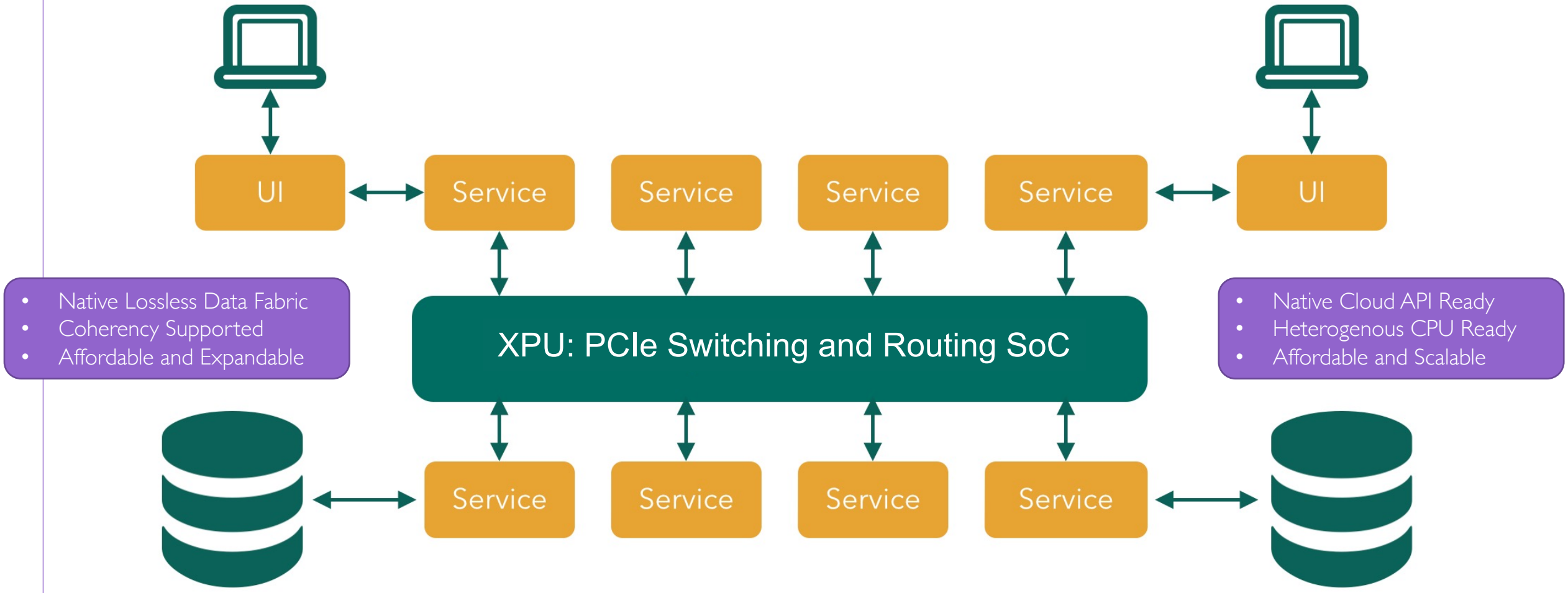
Infrastructure Trends: Turn Metal into API



AKRAINO | WeOpen | NextArch FOUNDATION

SOCNOC

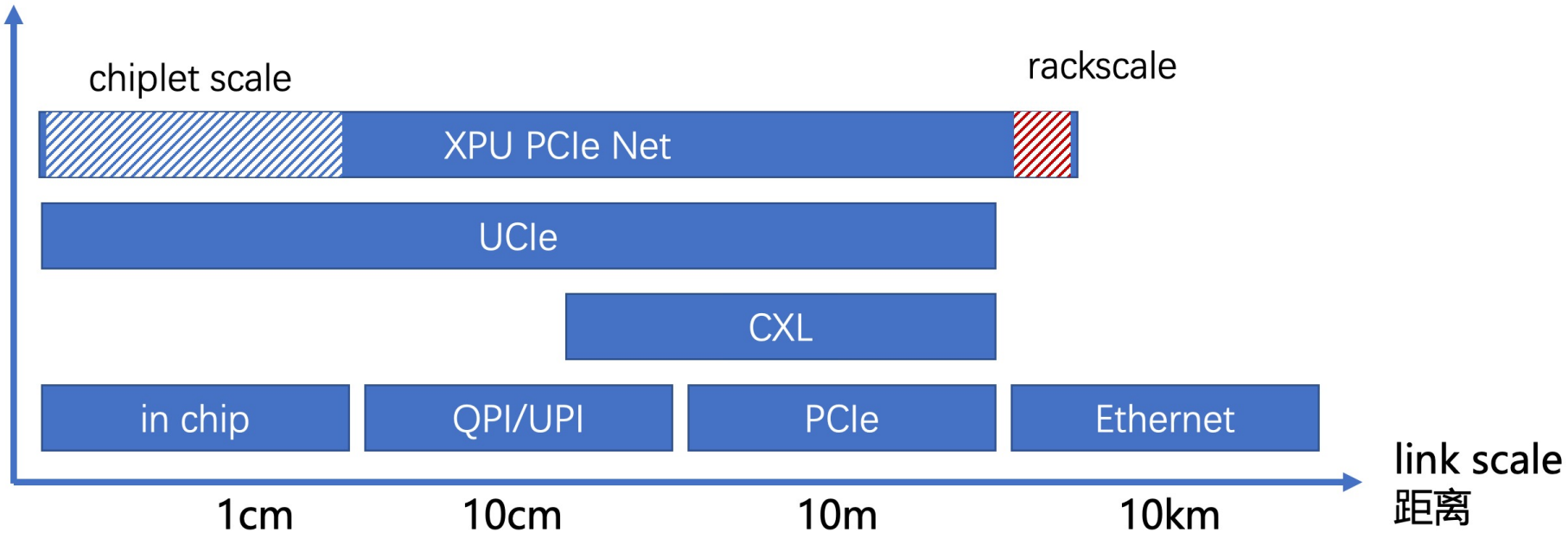
Our Solution: Bridge BUS and TCP/IP with PCIe!



AKRAINO | WeOpen | NextArch FOUNDATION

SOCNOC

Roadmap and Ecosystem



PCIe-based → CXL-based → UCIe-based

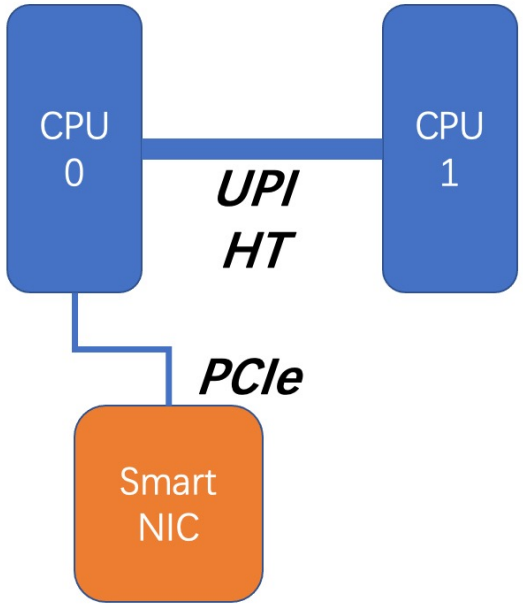
arm x86

AKRAINO | WeOpen | NextArch FOUNDATION

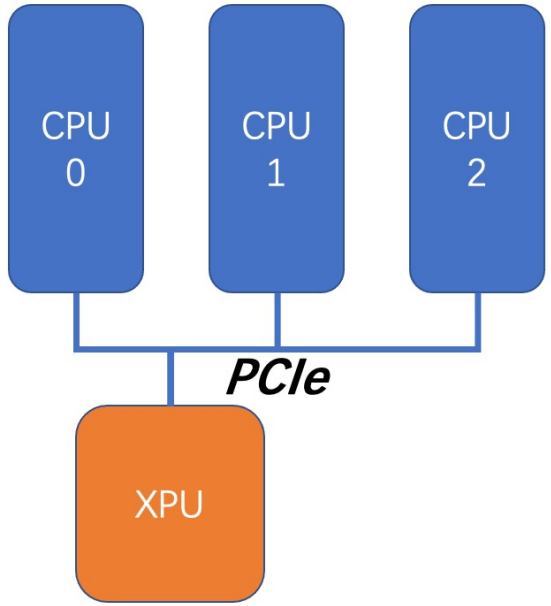
XPU Implementation & Applications

XPU Server: All in PCIe!

- **Transform UPI/HT based multi-CPU Server into PCIe connected Server!**

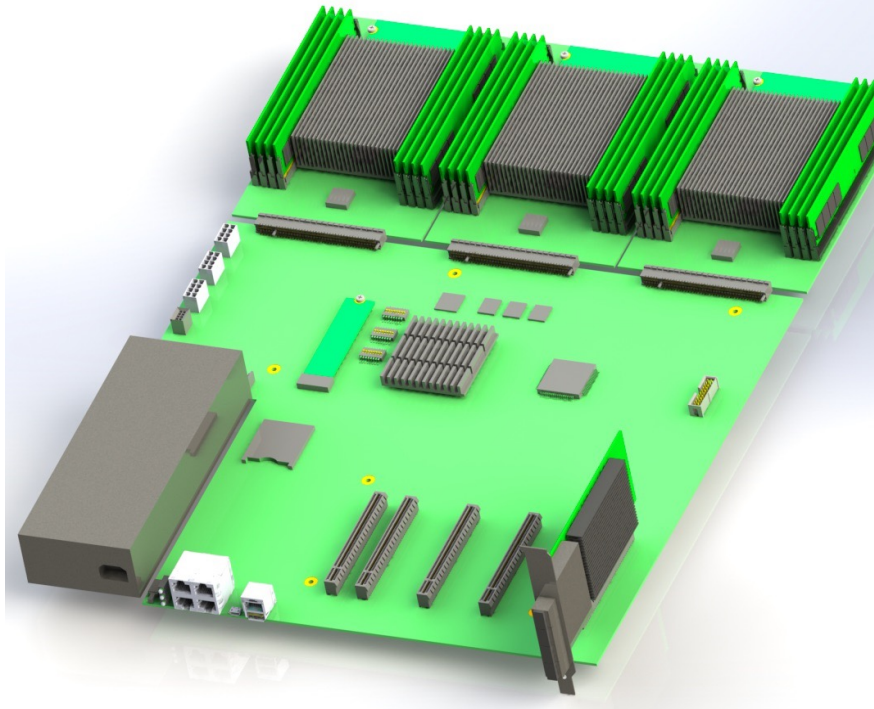


VS



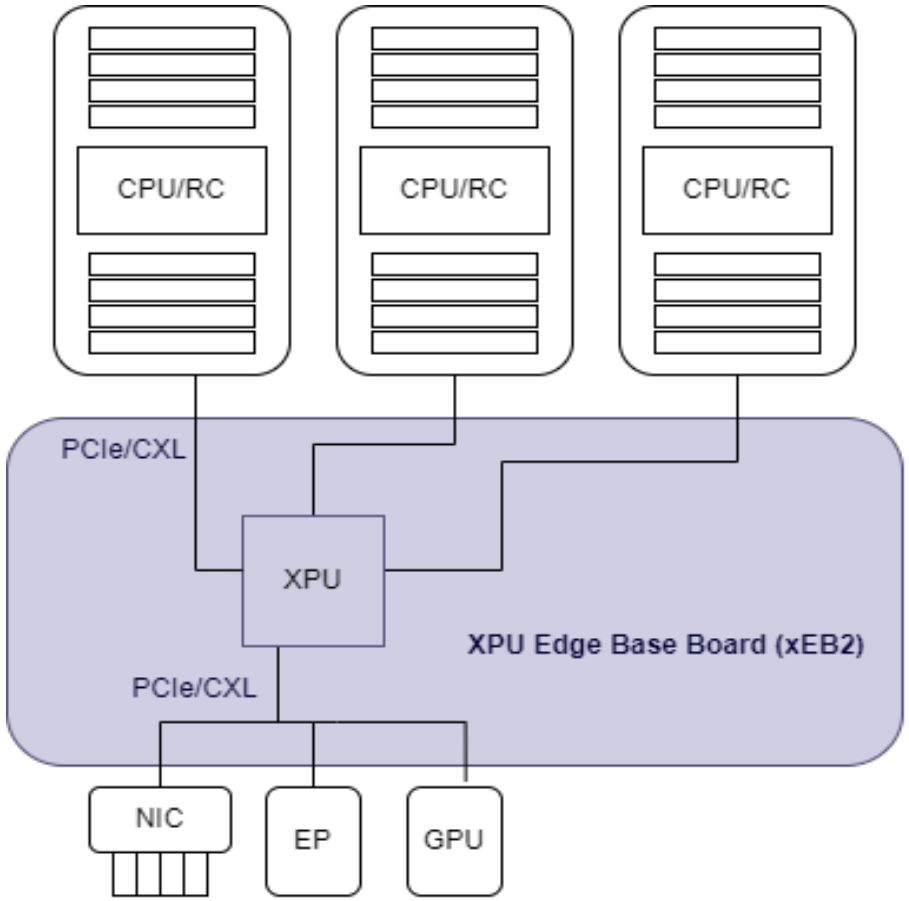
Dual CPU Server

XPU-centered Server



SOCNOC

Cloud Native Drives PCIe-Centered Possible & Essential



PCIe-centered

Cloud Native Architecture, serverless API, every application demonstrates by docker, communication based on TCP/IP

Scale Out Computer Clusters Technology

Data processing in a data center but also low latency in application processing and optimization of communication traffic by providing a high-speed interconnect function

docker

Linux

arm x86

AKRAINO | WeOpen | NextArch FOUNDATION

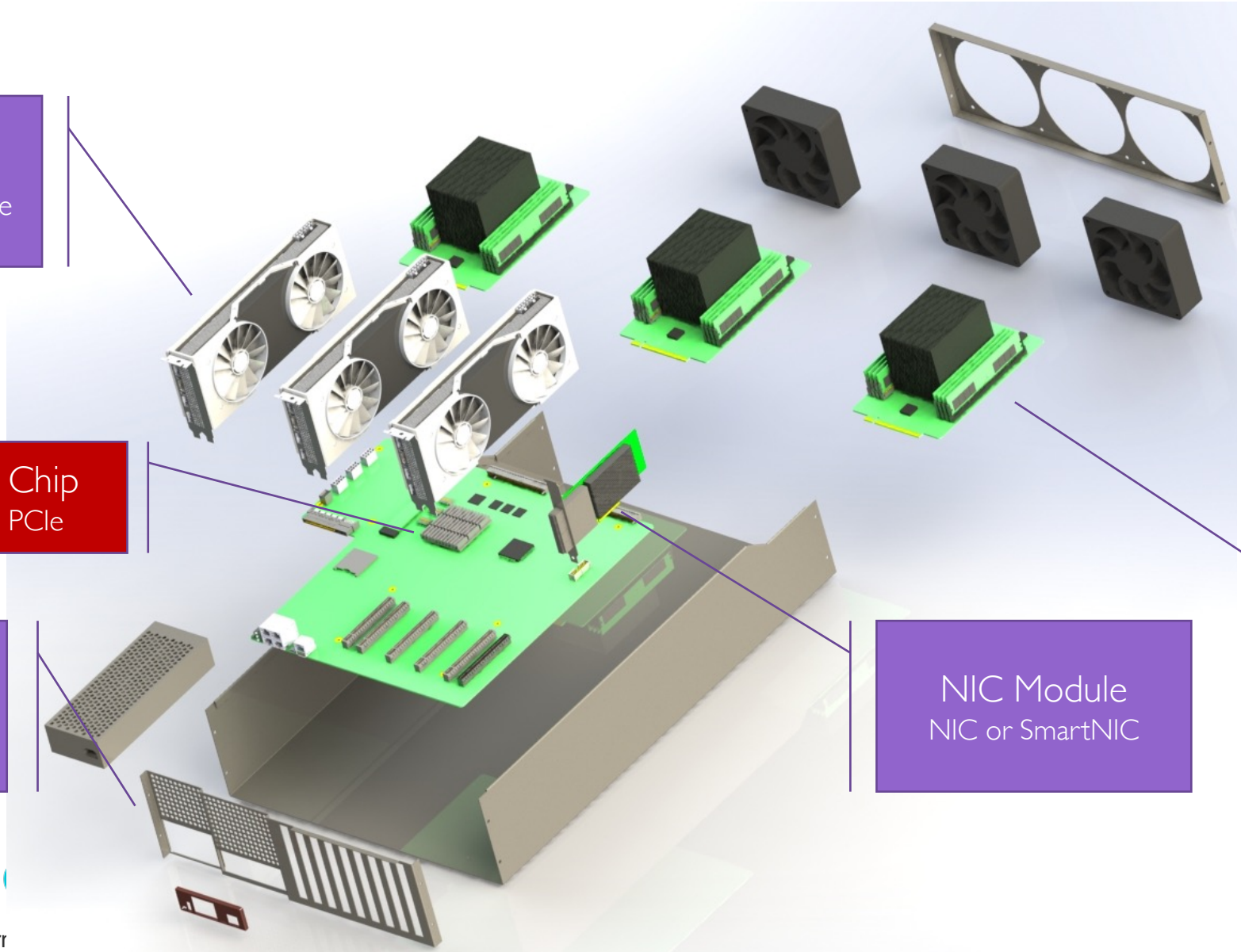
SOCNOC

Cloud Native Server for Integrated Edge Computing

GPU Module
Standard PCIe Device

XPU Chip
All in PCIe

Standard Server Formfactor
1U, 2U, 3U, 4U supported



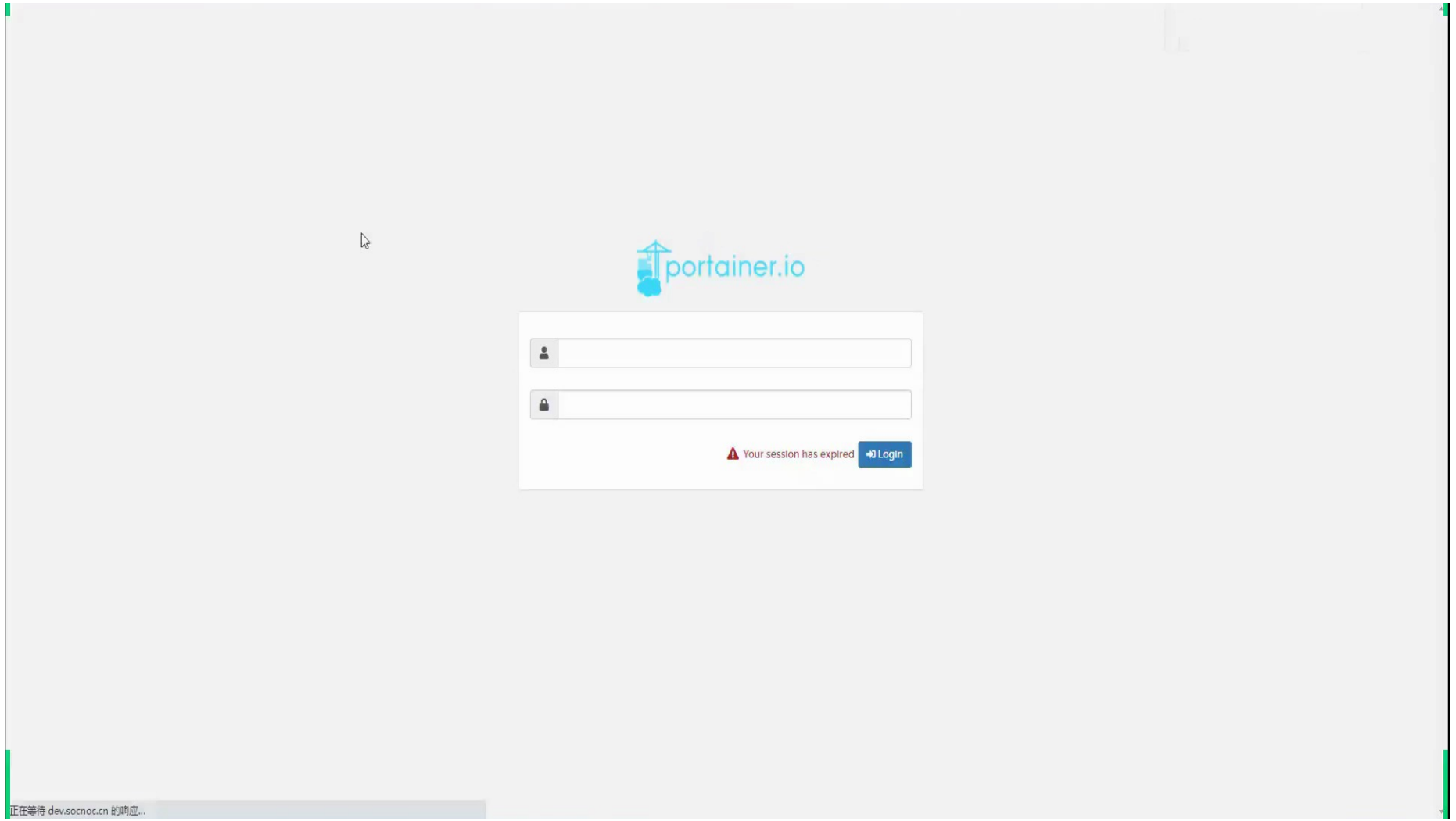
NIC Module
NIC or SmartNIC

CPU Module
CPU, RAM, SSD combo

AKRAINO

SOCNOC

Online Demo



AKRAINO | WeOpen | NextArch FOUNDATION

Thank you !

AKRAINO | WeOpen | NextArch FOUNDATION

www.socnoc.ai