

The Race To Neural Class Networks

Tina Tsou

Workshop on “Machine Learning
for 5G and beyond”



Building IT that works like the brain



“All significant mental algorithms are learned
except for the learning and reward machinery itself”

AI, Machine Learning & Deep Learning

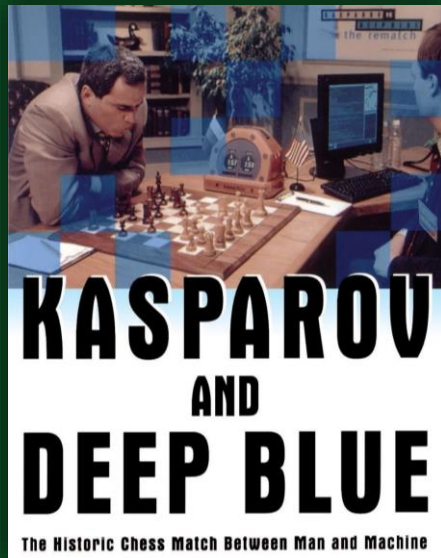


Perform tasks that normally require human intelligence

Replace instructions with a neural network to parse data, learn, and make predictions.

Increase the layers and run massive amounts of data through the system to train it.

Artificial Intelligence



Machine Learning



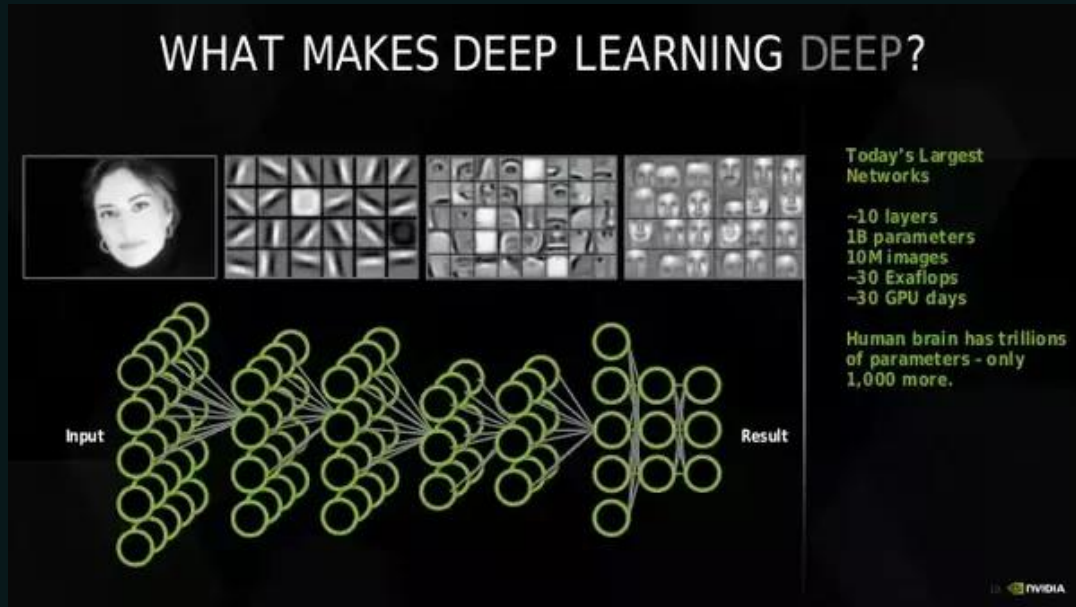
Deep Learning



“Neural-Class” Networks

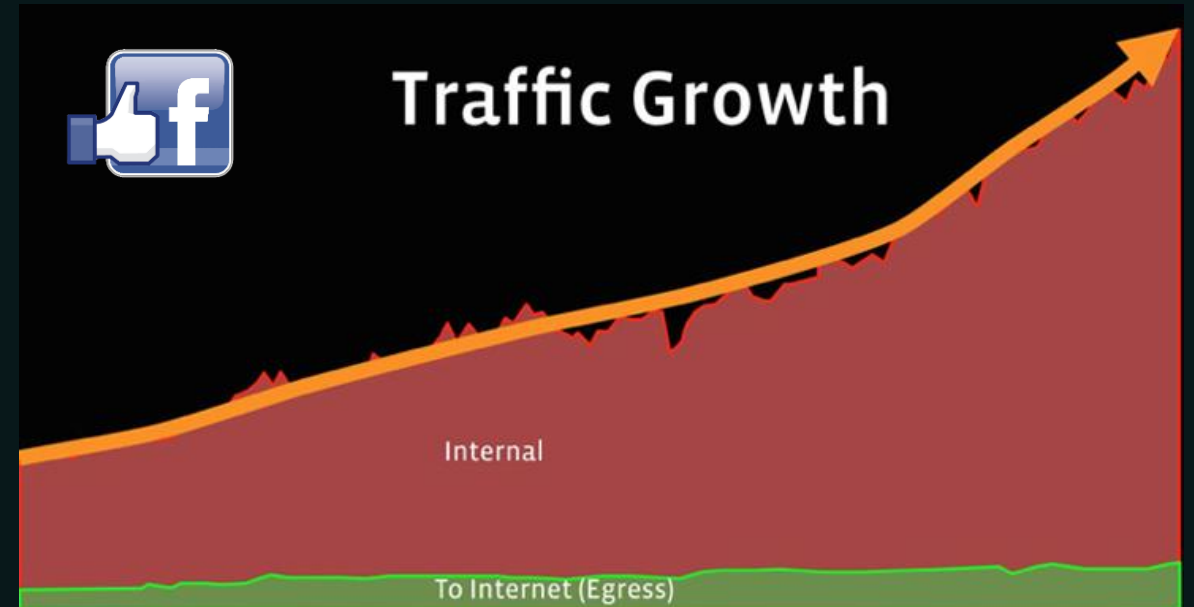


Deep Neural Network



1 Billion Parameters

Hyperscale Environment



1 Billion Users

Year 2000 Server CPU Networking

100 Servers x 1 Core = 100 Cores



SYSTEM SPECIFICATIONS

CPU	1x Pentium
Cores	1

Year 2018 Server CPU Networking

100 Servers x 48 Cores = 4,800 Cores

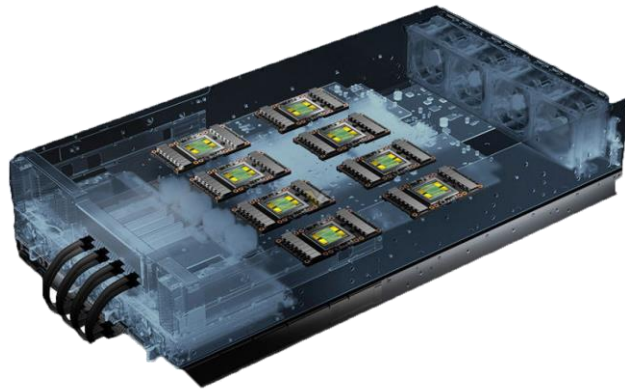
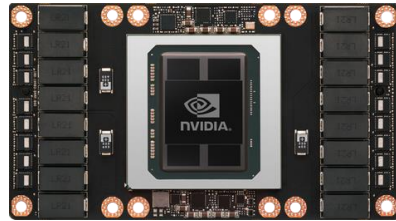


SYSTEM SPECIFICATIONS

CPU	1x Qualcomm Centriq 2400
Cores	48

Year 2018 Server GPU Networking

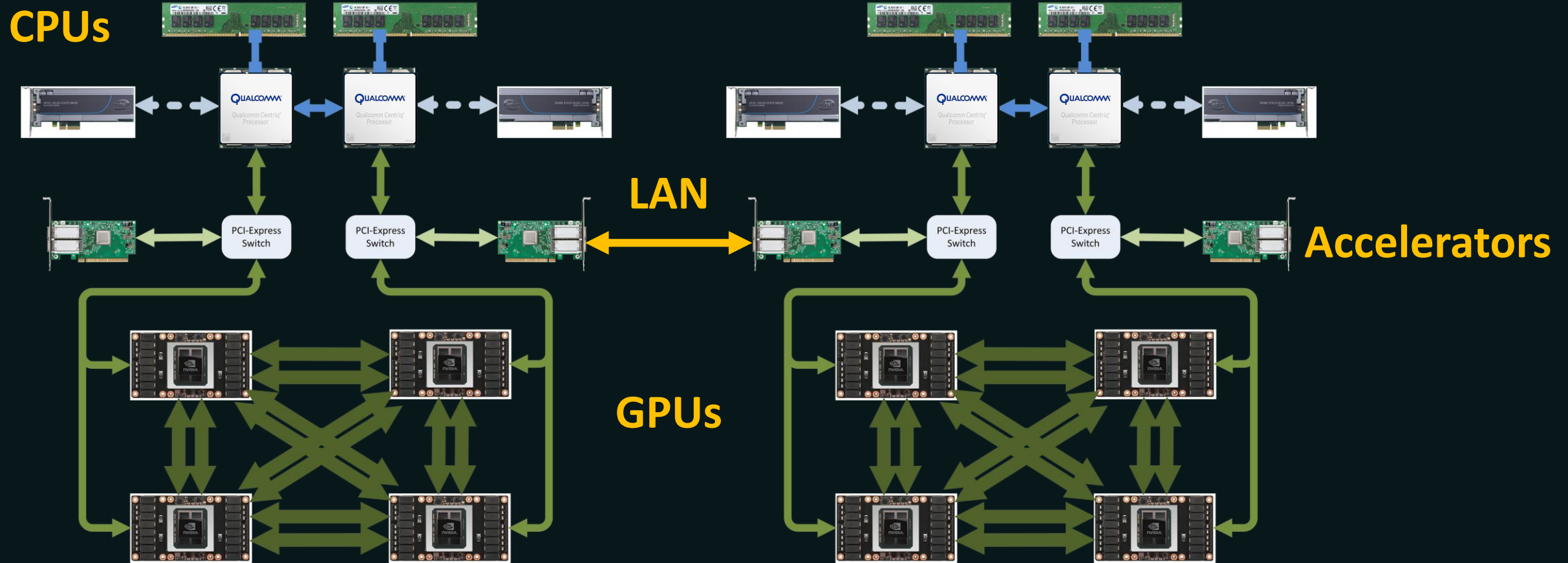
100 Servers x 46,080 Cores = 4,680,000 Cores



SYSTEM SPECIFICATIONS

GPUs	8X Tesla V100	8X Tesla P100
Performance (GPU FP16)	1 petaFLOPS	170 teraFLOPS
GPU Memory	128 GB total system	
CPU	Dual 20-Core Intel Xeon E5-2698 v4 2.2 GHz	
NVIDIA CUDA® Cores	40,960	28,672
NVIDIA Tensor Cores (on V100 based systems)	5,120	N/A

The Future of Networking



A HUGE New Class of Data Centers



NEW

Micro Data Centers
(Edge Computing)

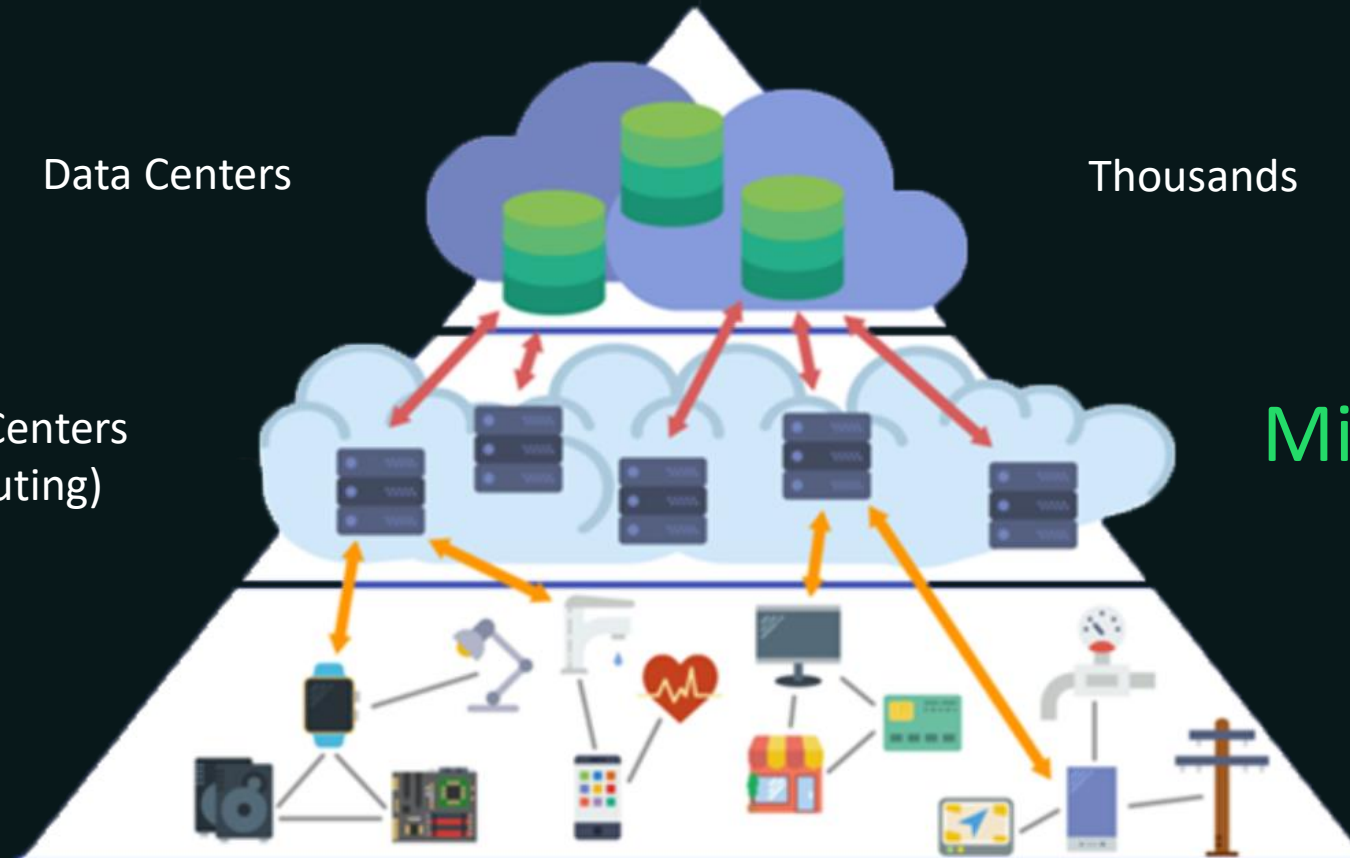
Data Centers

Thousands

Millions

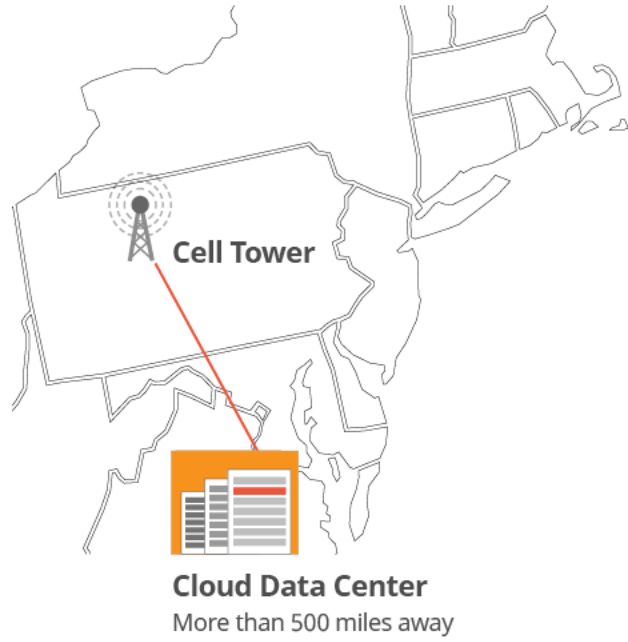
Devices

Billions



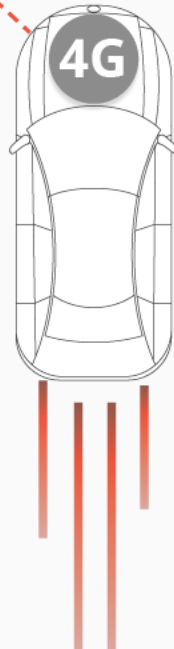
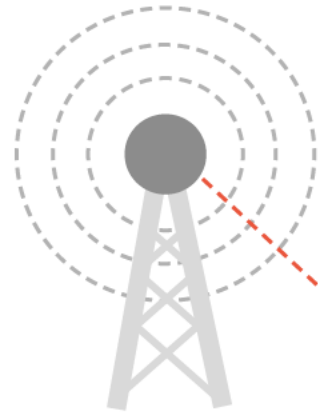
Current: 4G

Only a few large centralized data centers



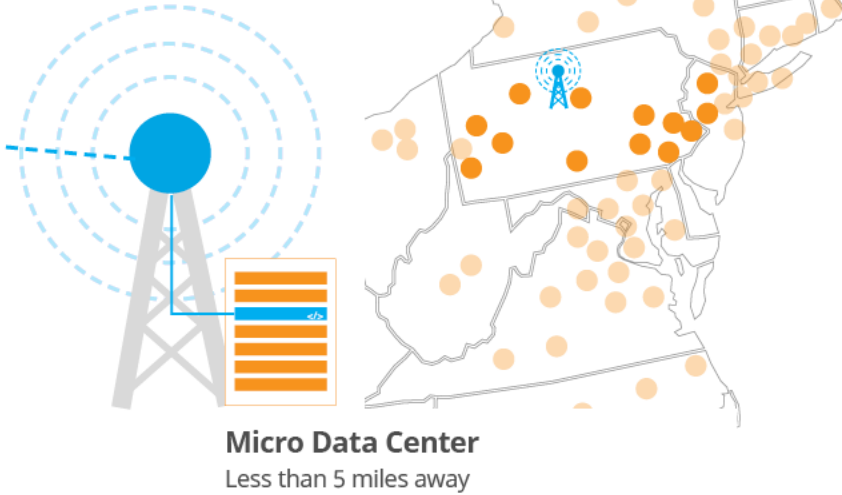
> 80 ms latency

The car moved *over four feet* by the time it received a response due to the large distance from the data center.



Upcoming: 5G

Thousands of new micro data centers under cell towers



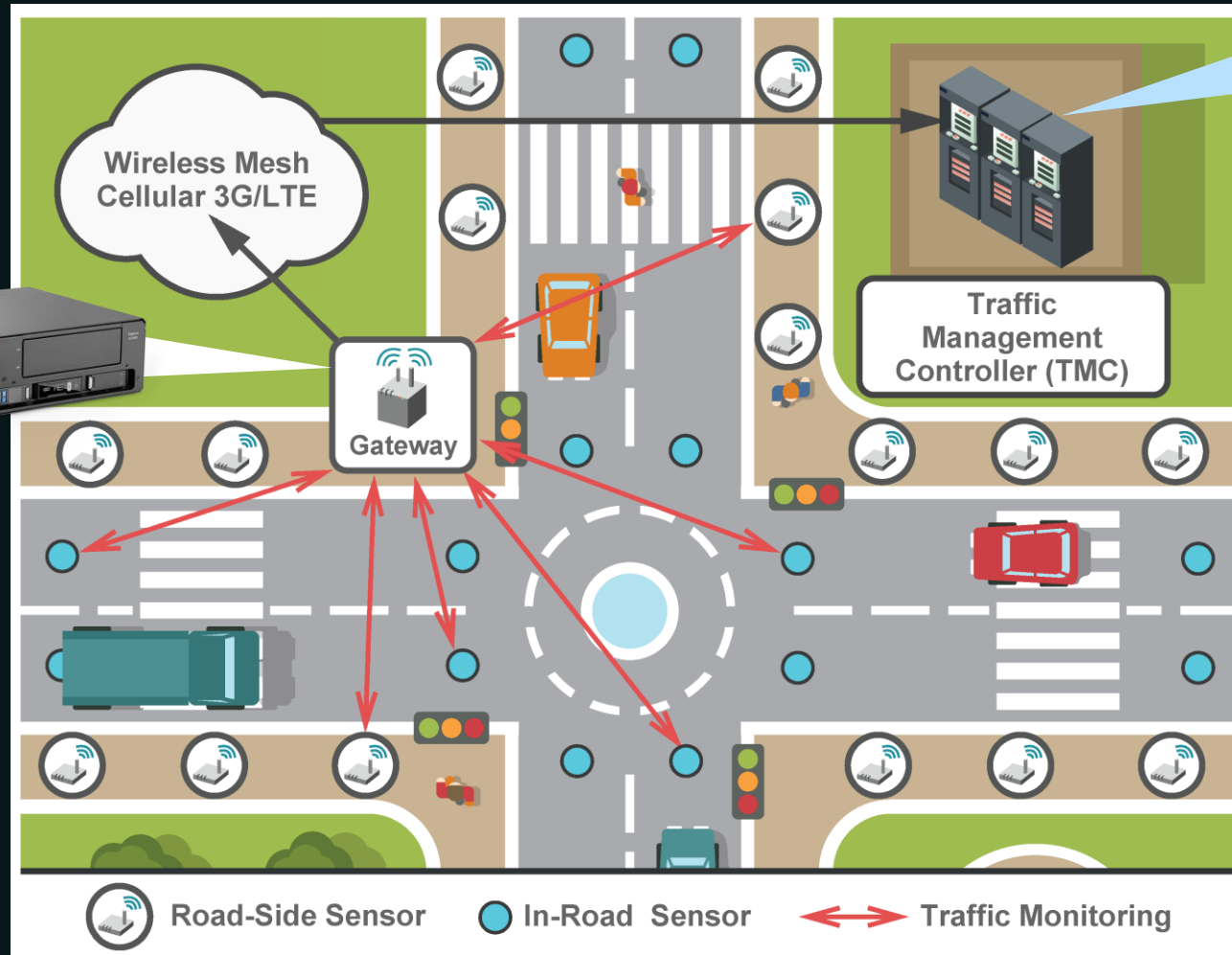
< 5 ms latency

The car moved *less than four inches* by the time it received a response, thanks to the close distance to the Micro data center.

Use Cases



Complex



Edge Gateway



Micro Data Center Rack

Server Configured Like a Human Brain



>100 PetaFLOPS Processing Power

1.3 Billion Terabytes of Storage

3.3 Pounds

20W Power



77 Cubic Inches

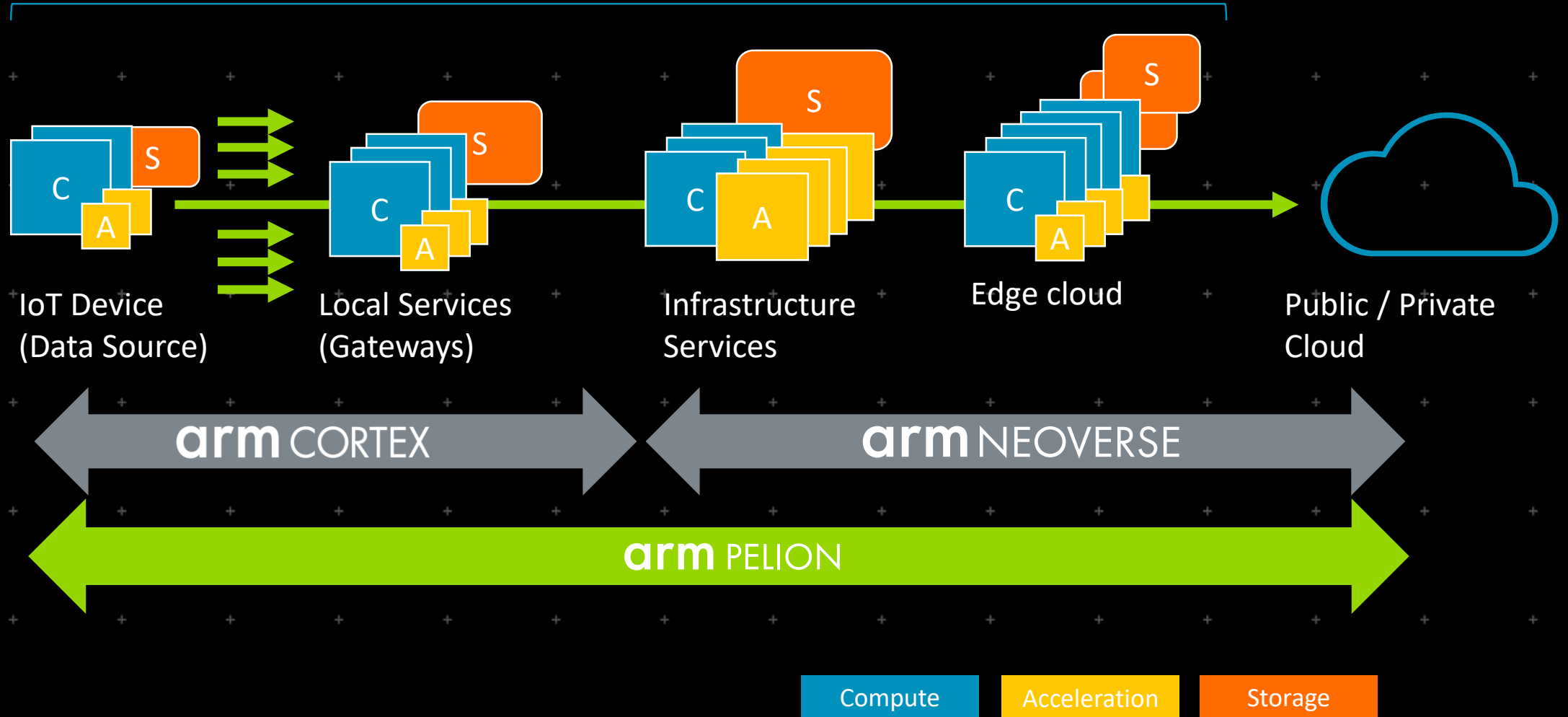
19" Wide x 1.75" (1U) High x 2.3" Deep

10,000 Network Ports

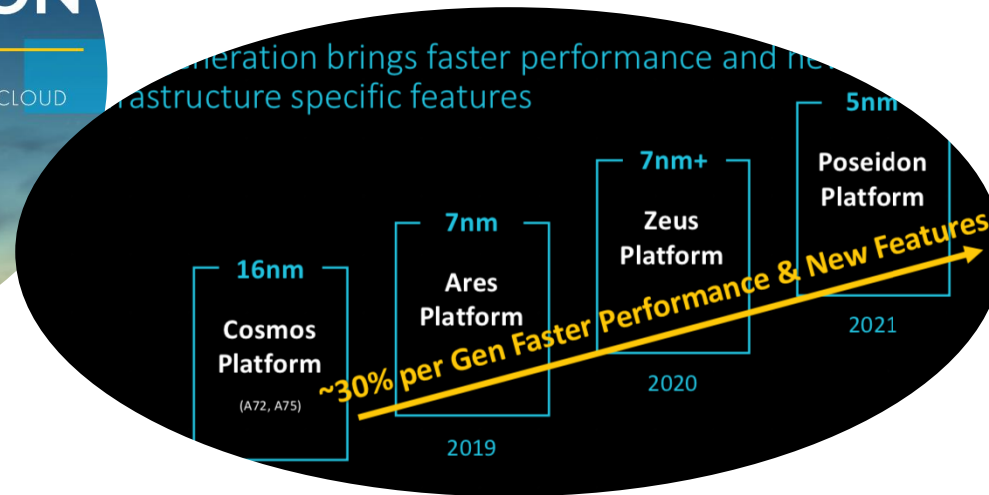
100,000 Miles of Cabling Included

Arm is uniquely positioned in “device-to-cloud”

Edge computing – necessary to handle the massive upstream data



Compute solutions Device-to-Cloud

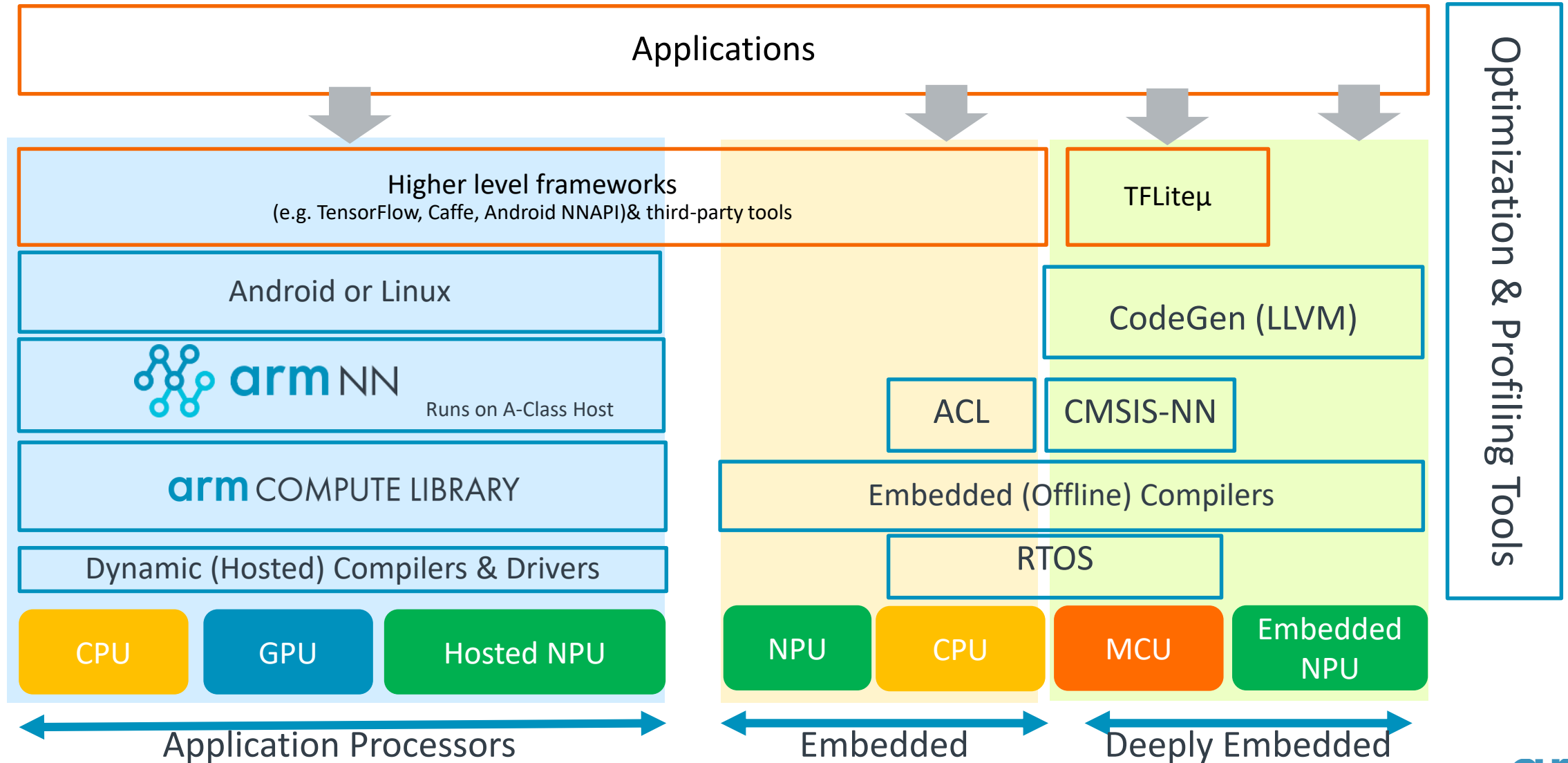


Corstone secure foundations for Cortex processors

Neoverse platforms



Arm's Heterogeneous ML Platform – Hosted & Embedded



Akraino Release 1

Akraino R1: Unifying the Edge

NEWS

Akraino Edge Stack Issues Premier Release, Sets Framework to Enable 5G, IoT Edge Application Ecosystem

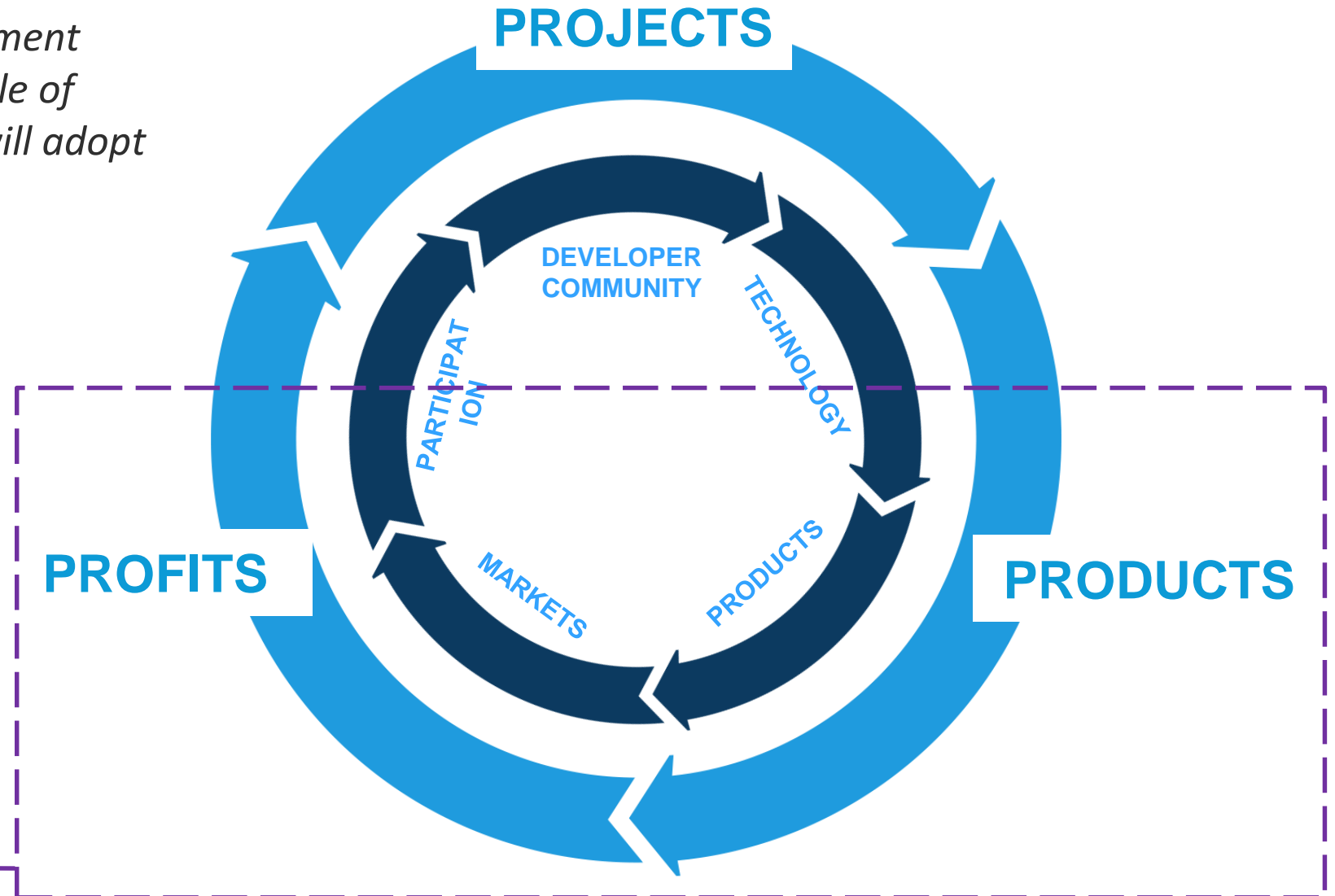
- *Inaugural release unifies multiple sectors of the edge across disciplines, including IoT, Enterprise, Telecom, and Cloud*
- *Delivers tested and validated deployment-ready blueprints*
- *Creates framework for defining and standardizing APIs across stacks, via upstream/downstream collaboration*

SAN FRANCISCO – June 6, 2019 – [LF Edge](#), an umbrella organization within the [Linux Foundation](#) that aims to establish an open, interoperable framework for edge computing independent of hardware, silicon, cloud, or operating system, today announced the availability of [Akraino Edge Stack](#) Release 1 (“Akraino R1”). Created via broad community collaboration, Akraino’s premiere release unlocks the power of intelligent edge with deployable, self-certified blueprints for a diverse set of edge use cases.

Role of Akraino Blueprints in Open Source Lifecycle

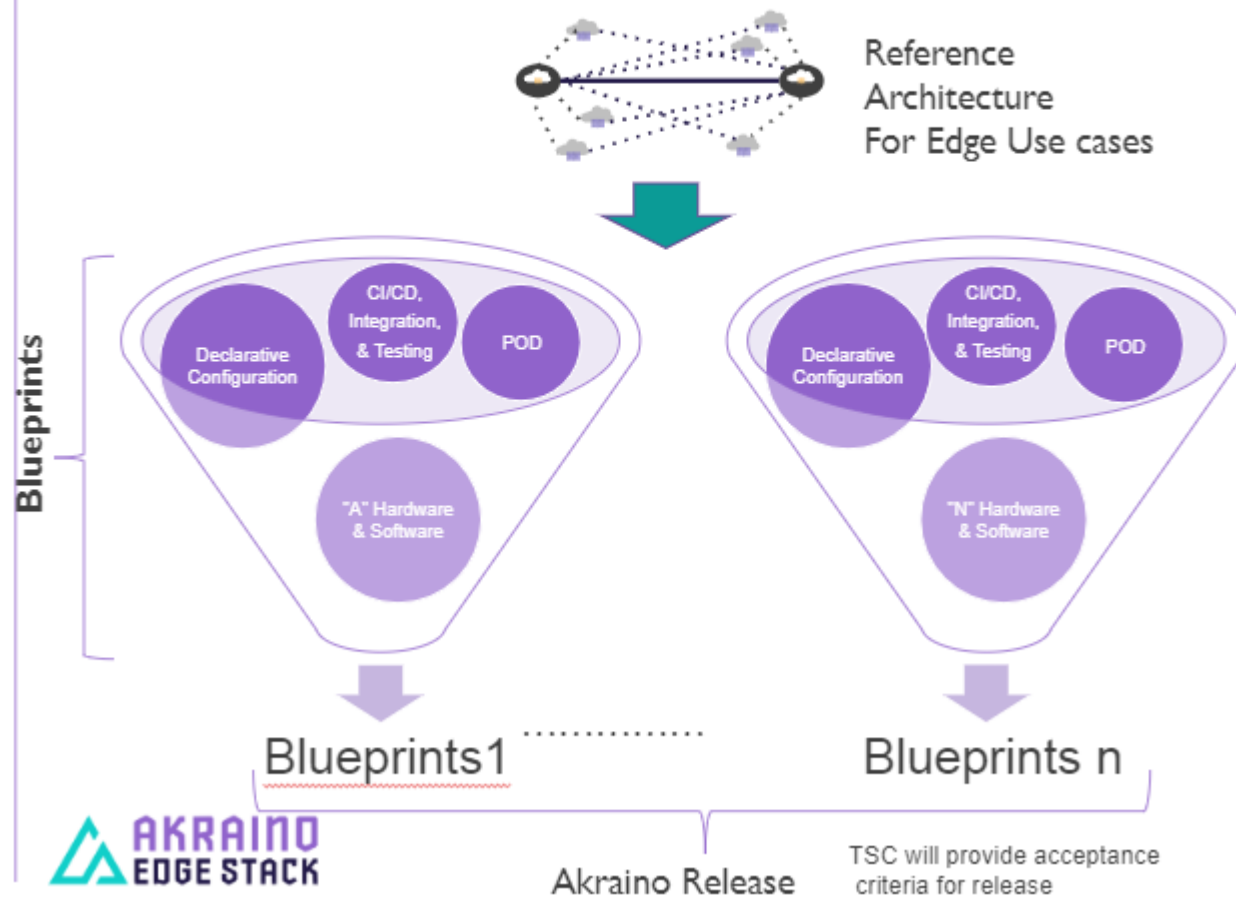
Successful Open Source Development depends on the complete life cycle of projects, products that market will adopt and deploy

- Akraino's blueprint model enables rapid transition from projects to PoC/deployment, cutting down integration time significantly.
- Blueprints augment Open Source Projects to address & accelerate Interoperability, Packaging, and Testing under Open and neutral governance.



Akraino R1: Tested & Validated Blueprints

Akraino Blueprints & release



- 11+ Blueprint families, 20 Blueprints under development
- Community-tested & validated on real hardware, Akraino Labs by members and community.

Blueprints - approved & tested declarative configuration based on use cases, set of hardware, POD & software

Reference Architecture - defines Akraino building blocks

Declarative Configuration - hides lower layer complexity to user

CI/CD, Integration & Testing Tools - drive product quality

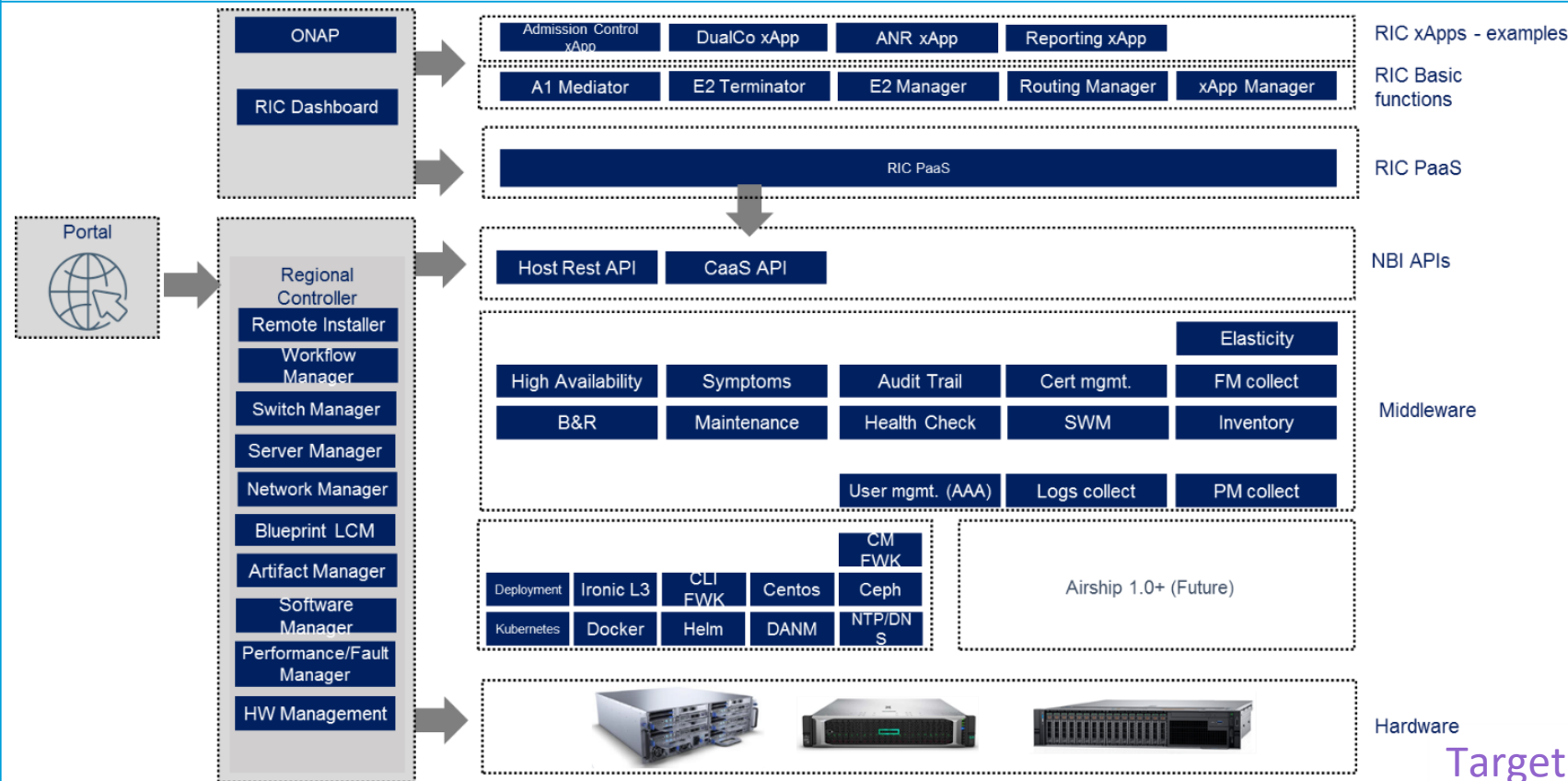
Tested & Validated Deployment-Ready Blueprints Details

Akraino R1 Blueprints Details

Blueprint Family	Blueprint	Primary Use Case	Industry Target	Blueprint Summary
Network Cloud (NC)	Unicycle with SR-IOV	Telco Edge use cases (Multi Server). Multiple applications	Telco, Enterprise	NC blueprint family enables hardware configuration and automated deployment of telco grade multiple edge sites from a remote regional controller.
	Rover	Telco Edge use cases (Single Server). Multiple applications	Telco, Enterprise	
	Unicycle with OVS-DPDK	Telco Edge use cases (Multi Server). Multiple applications	Telco, Enterprise	
Telco Appliance	Radio Edge Cloud (REC)	Appliance for Radio Access Network (RAN), RAN Intelligent Controller and Near realtime Edge MEC Appliance	Telco 5G, Enterprise	Appliance tuned to support the O-RAN Alliance and O-RAN Software Community's Radio Access Network Intelligent Controller (RIC)
Integrated Edge Cloud (IEC)	Type 1 (small Edge)	Telco or enterprise application deployment on Arm servers	Telco, IOT and Enterprise	IEC enables the new functionalities and deployment model on the network edge. It supports ARM processors and architecture.
	Type 2 (Medium Edge)	Telco or enterprise application deployment on Arm servers	Telco, IOT and Enterprise	
StarlingX	Far Edge Distributed Cloud	Enterprise edge and Far edge. Multiple applications	Enterprise & IOT	Addresses edge and Far edge use cases at high density locations such as malls, airports and sports stadiums to support value added services at these events and locations.
Kubernetes- Native Infrastructure for Edge	Provider Access Edge	vRAN and MEC (AR/VR, Machine learning, etc.,)	Enterprise & Telco	Blueprints in the Kubernetes-Native Infrastructure for Edge family leverage the best-practices and tools from the Kubernetes community to declaratively and consistently manage edge computing stacks from the infrastructure up to the workloads.
Edge Lightweight and IOT blueprint (ELIOT)	IOT Gateway	IOT	IOT & Enterprise	ELIOT targets on making the edge node a lightweight software stack which can be deployed on limited hardware capacity.
	uCPE	uCPE	Enterprise & Telco	

Akraino R1 Blueprint: Radio Edge Cloud

Radio Edge Cloud for 5G use cases



Purpose/Features

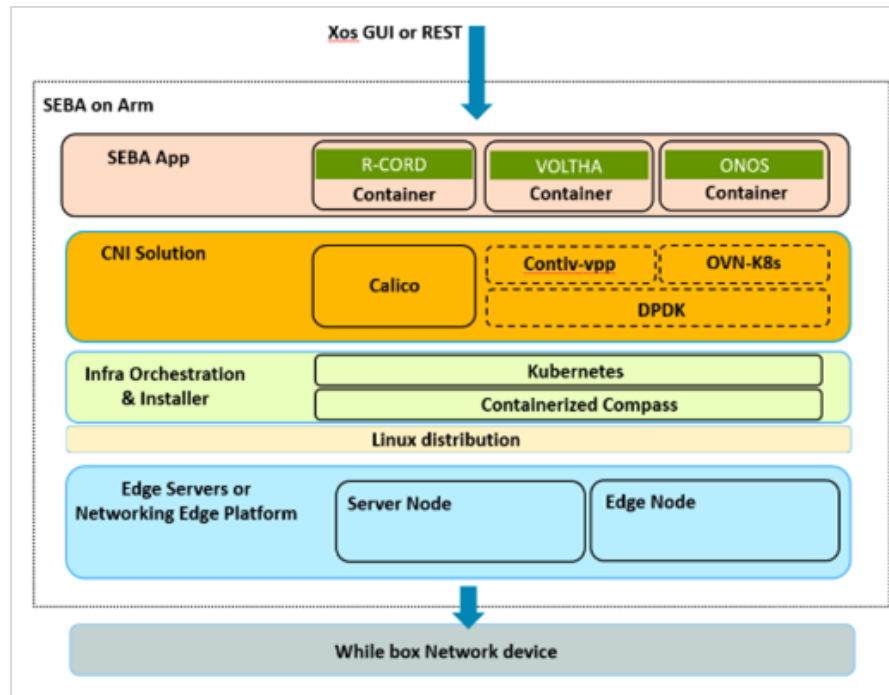
- Telco- grade edge cloud platform for near-real time container workloads.
- Integrated with open-source RAN Intelligent Controller (RIC) – helps deploy customized, tailored apps with Zero touch
- Automated CD pipeline testing the full software stack

Target Industry: Telco 5G, Enterprise

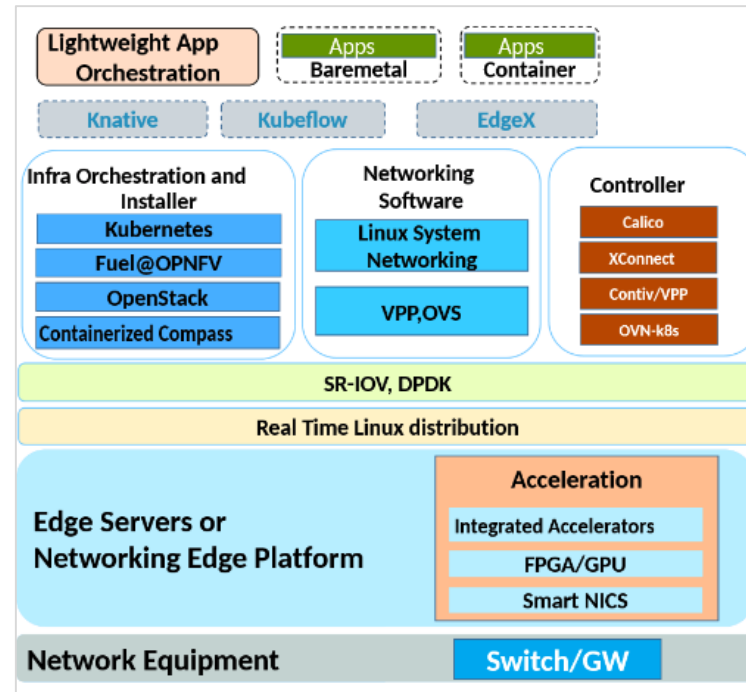
* Target architecture shown here and the fully installable building blocks is in R1

Akraino R1 Blueprint: IOT & Remote Edge

Integrated Edge Cloud



Type 1



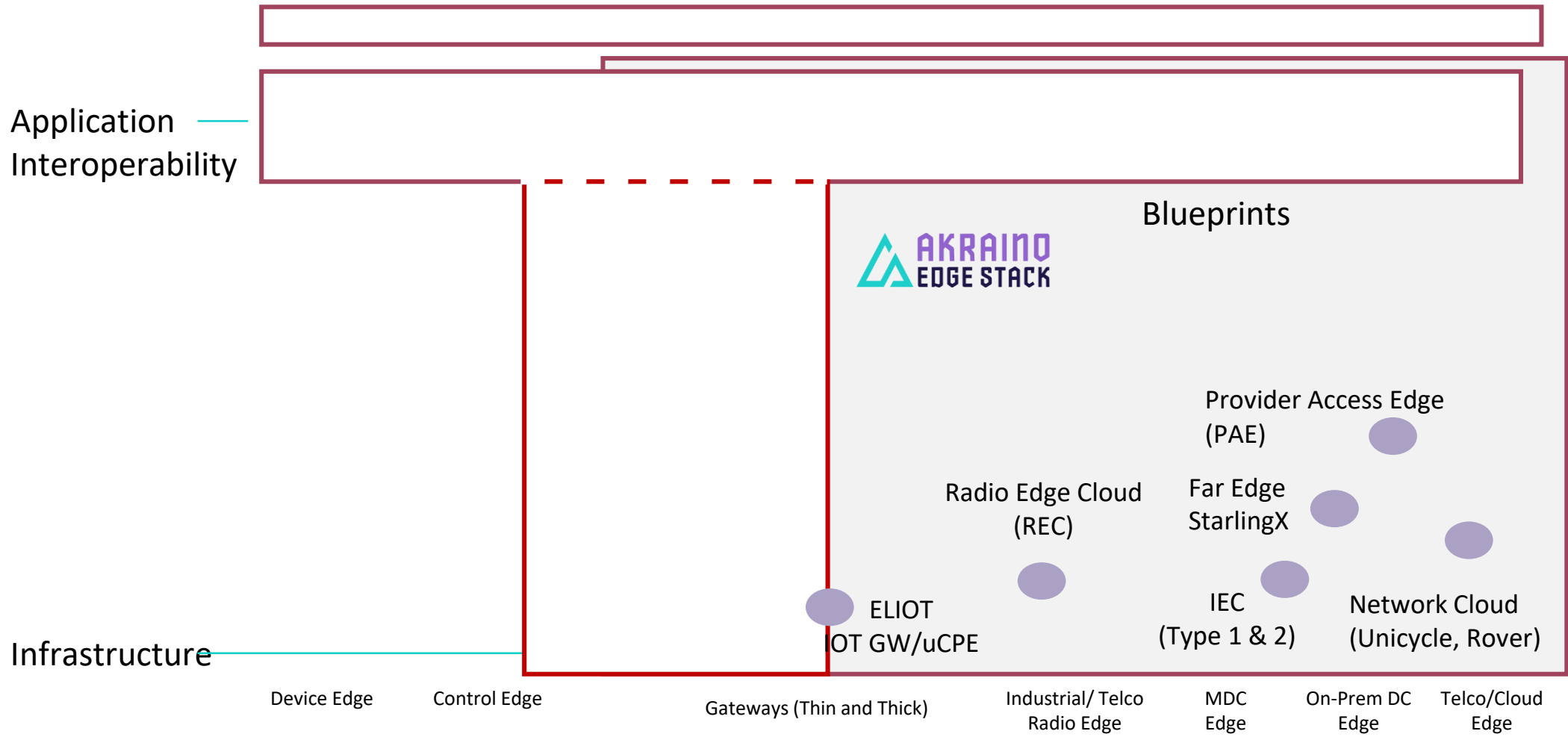
Type 2

Purpose/Features

- Addresses IOT use cases
- Targets telco edge applications & medium edge cloud deployments with Arm, based on Kubernetes and Calico
- Automated installation, integrated with SDN-Enabled Broadband Access (SEBA) use case

Target Industry: Telco, IoT, Enterprise

Functional View: R1 Blueprints in Akraino Edge Stack



Akraino R1: Key Takeaways & What's Next in 2019

1. LF Edge Projects gaining community support with Akraino aimed at accelerating time to deployment -> Projects to Products & Production
1. Akraino establishes unified framework for Edge collaboration & validation across projects & community with Blueprints
1. Akraino's R1 releases 10+ Blueprints for IOT, Enterprise and Telco Edge Cloud



On the Horizon

- › New blueprints (Gaming, Connected Cars...) + enhancements to existing blueprints
- › Tools for automated blueprint validations
- › Edge API's in collaboration with LF Edge projects
- › New community lab hardware

The Race to Neural-Class Networks

Tina Tsou

Workshop on “Machine Learning for 5G and beyond”

