Arm HPC Solutions and Beyond

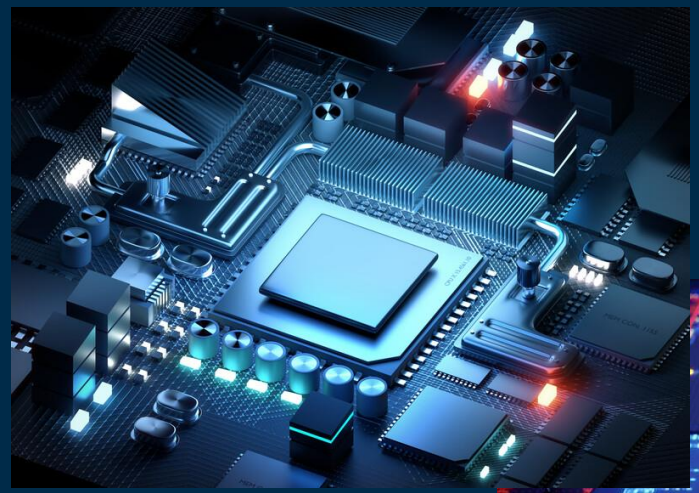
Brent Gorda, Senior HPC Executive, Arm

Arm HPC Platform Workshop @ PCCC| August 26th , 2021

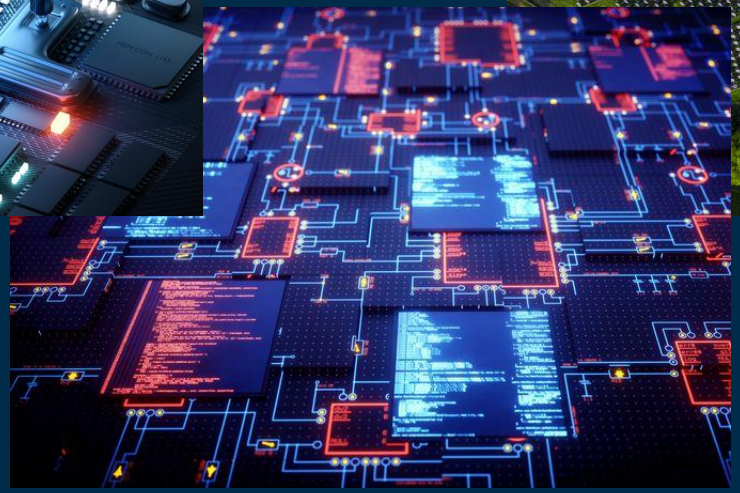
Confidential © 2021 Arm

Extreme Scale Compute System Design Objectives

CPU Per-Core Performance



Performance Efficiency

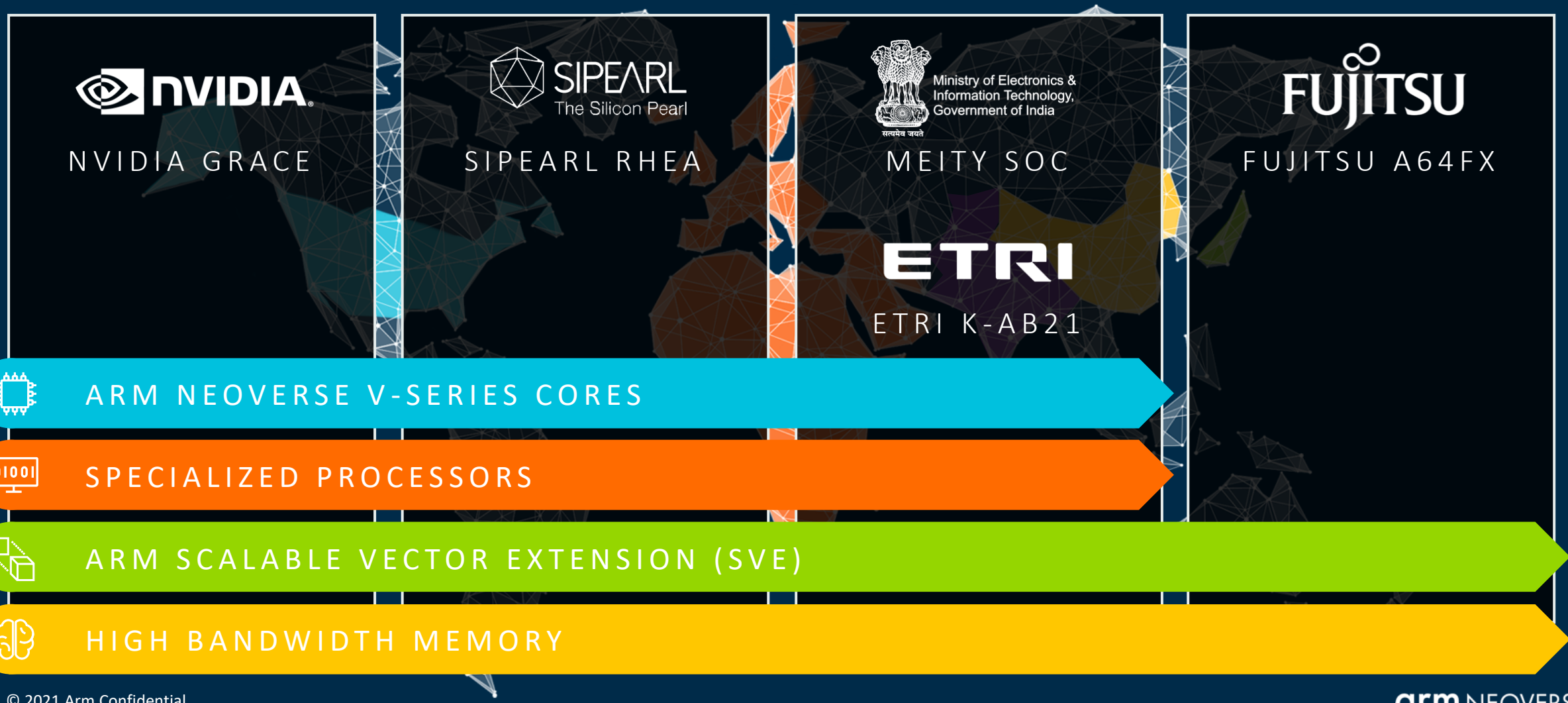


Memory Bandwidth & IO Flexibility

Technical Sovereignty

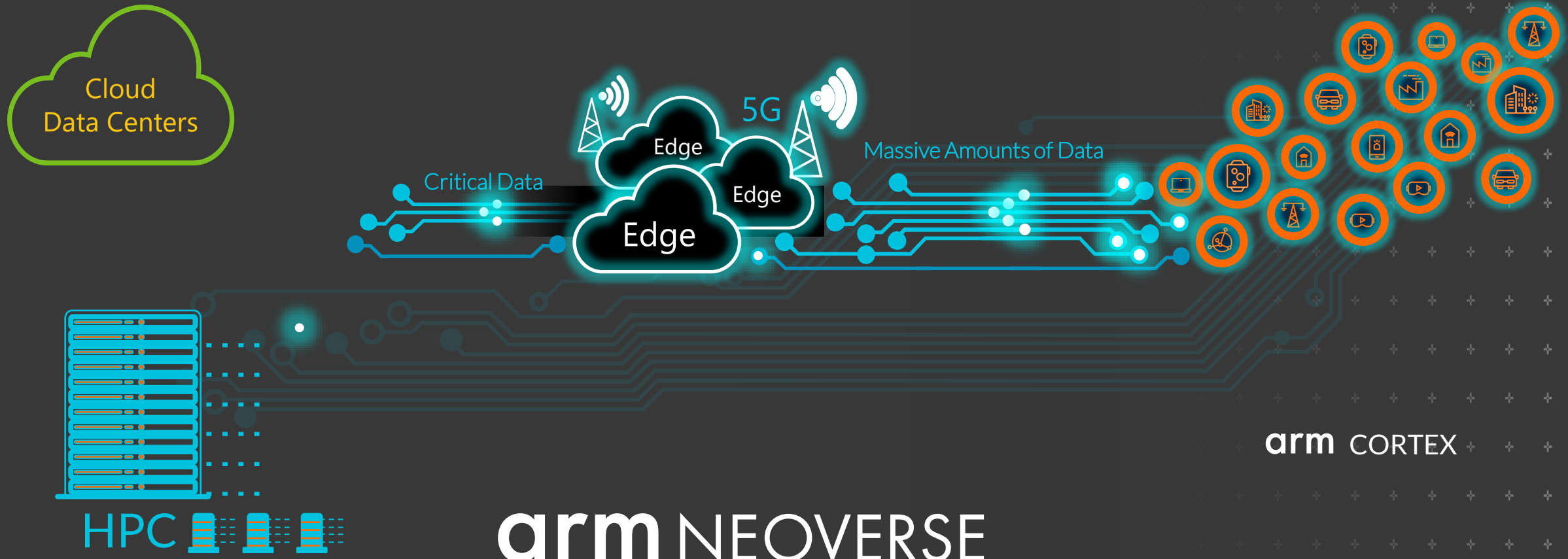
© 2021 Arm Limited

Arm Is the Driving Force Behind Innovation for HPC and AI



© 2021 Arm Confidential

arm NEOVERSE



arm CORTEX

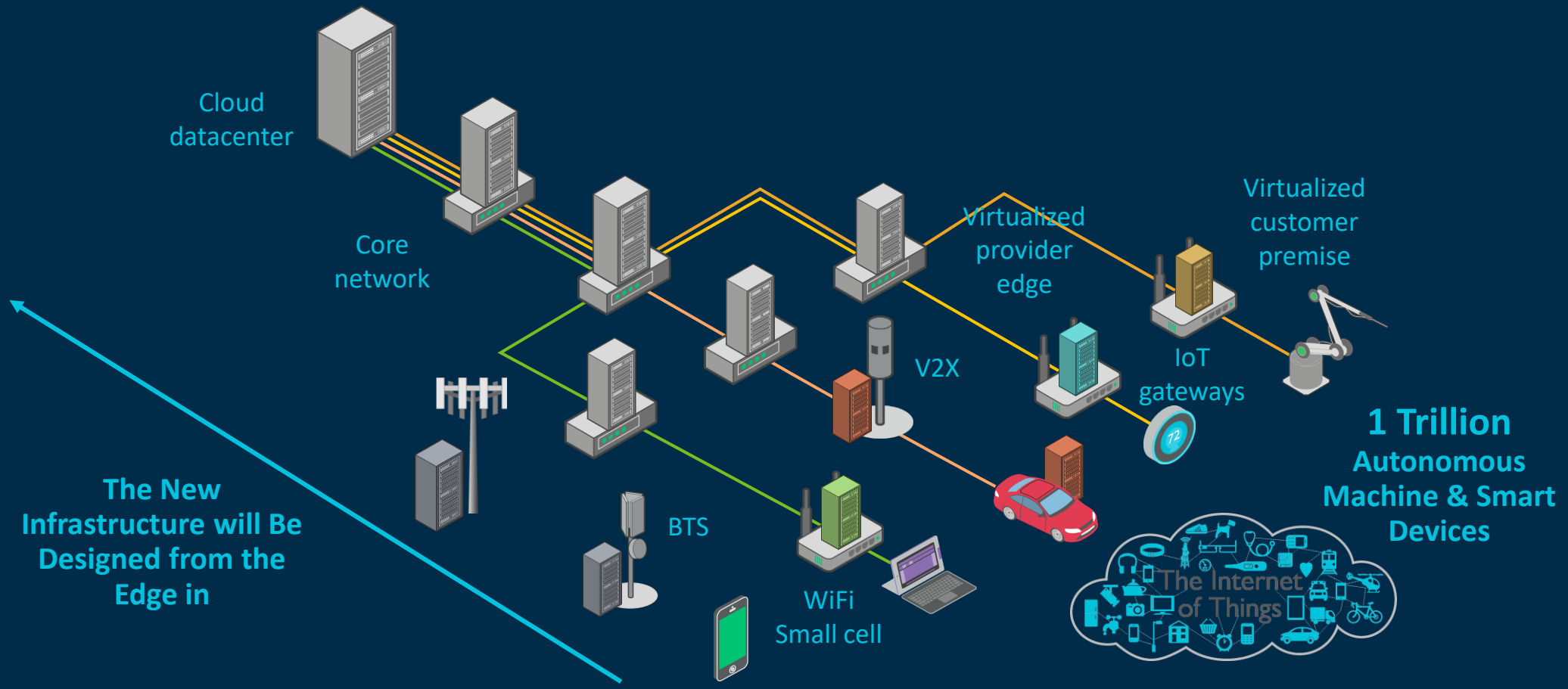
arm NEOVERSE

© 2019 Arm Limited

arm
arm NEOVERSE

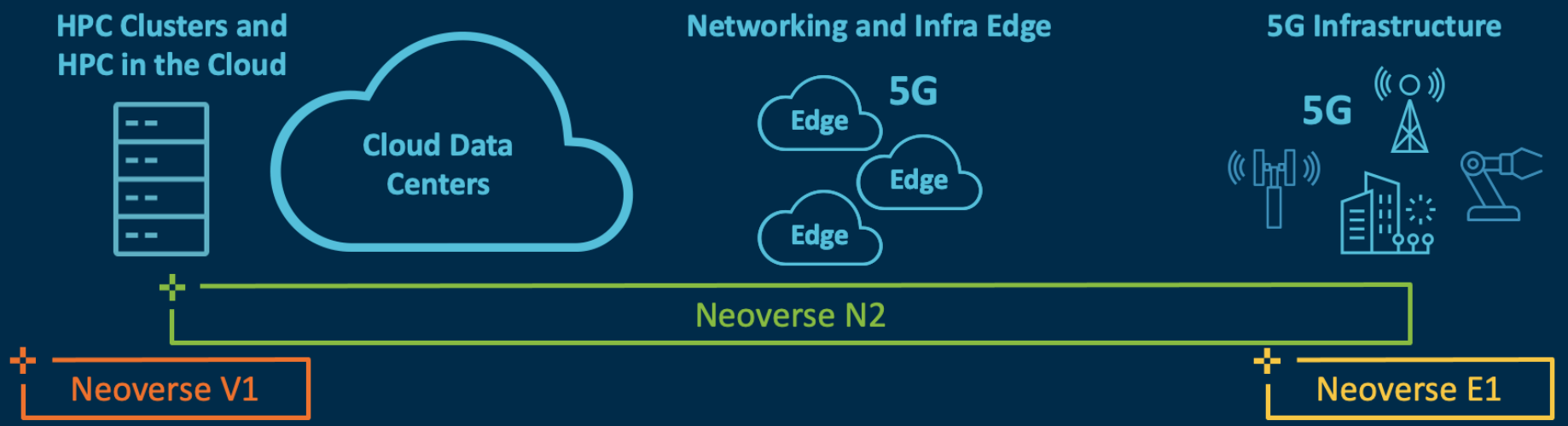
New infrastructure for support of 1T Arm devices

… will be designed from the edge in



© 2021 Arm Limited

Today's Server Silicon Design with Arm Neoverse



32-192 cores
80-350W TDP

High core count, high power HPC, Enterprise & Cloud CPUs

12-36 cores
30-80W TDP

Mid core count, mid power Switch and SmartNIC CPUs

8-16 cores
20-35W TDP

Low core count, low power Gateway and Router CPUs

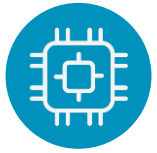
arm NEOVERSE

Cloud Native Software Development on Arm

- **100s of OSS Projects**
  Native Build
- **100+ ISVs**
  Commercial Support
- **100K+**
  Docker Hub Images
- **1M+**
  CI/CD Build Minutes per Month

Performance
Scalable Neoverse CPU Cores

Efficiency
Best $/throughput in the industry



Platform Diversity
Widest possible choice of Edge Platforms

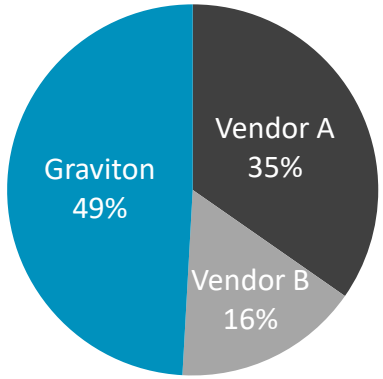
Optimization
Purpose built use cases

Confidential © 2021 Arm Limited

AWS Graviton2 is Quickly Expanding its EC2 Footprint

Steady growth and regional expansion since mid-2020

2020 AWS Instance Type Additions



49% of AWS EC2 instance additions in 2020 are based on AWS Graviton2

AWS Graviton2 Customer Adoption

powered by aws

X2gd

NEW

R6g

+ R6gd

C6g

+ C6gd, C6gn

T4g

M6g

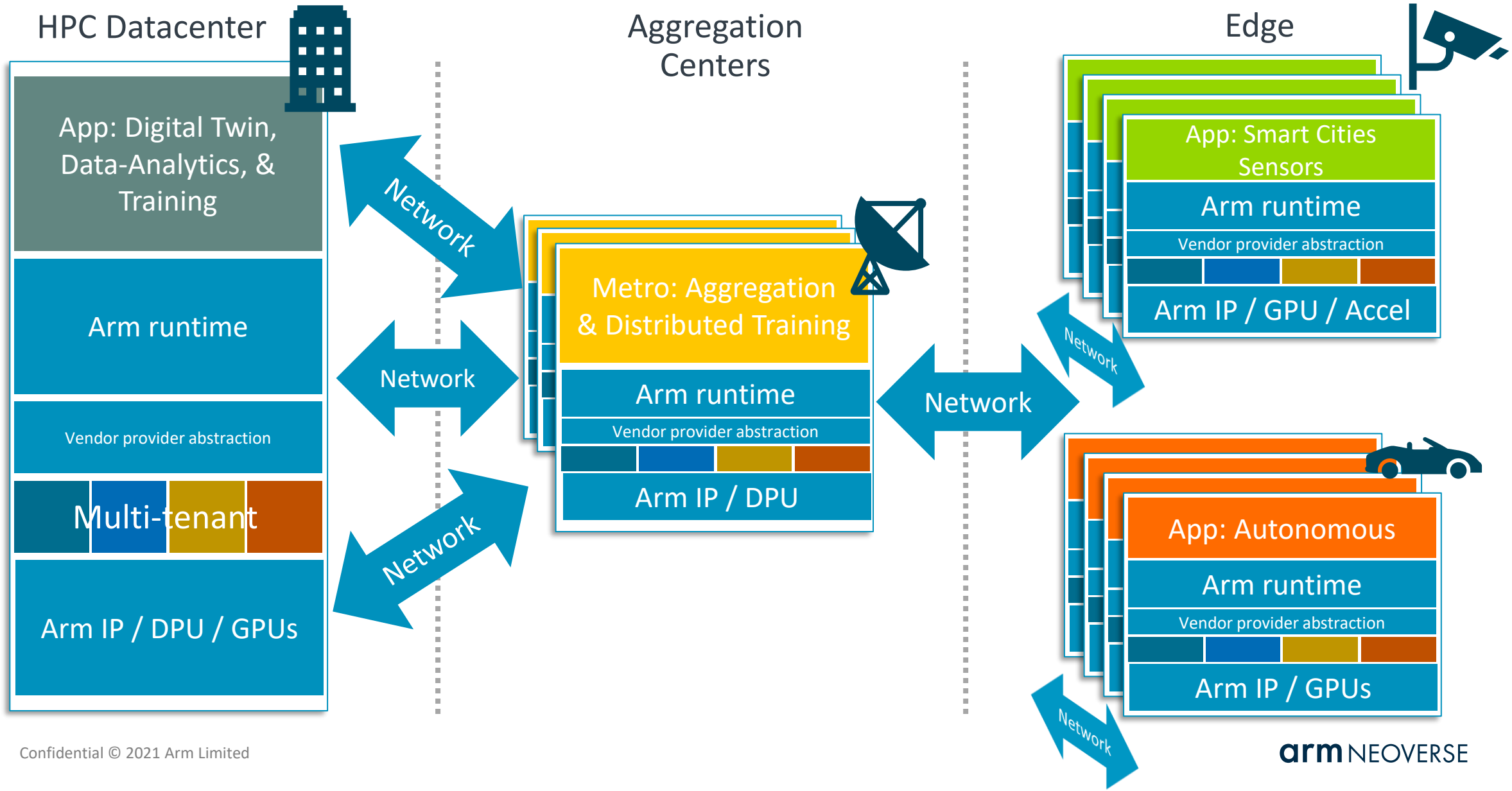
+ M6gd

Source: Liftr Insights

Confidential © 2021 Arm Limited

arm NEOVERSE

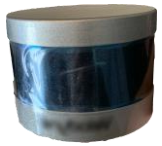
Securely scaling ML application across heterogenous HW



Confidential © 2021 Arm Limited
arm NEOVERSE

AI @ Edge

Sensors



LIDAR

Software Defined Radios

Hyperspectral Imaging

Facilities

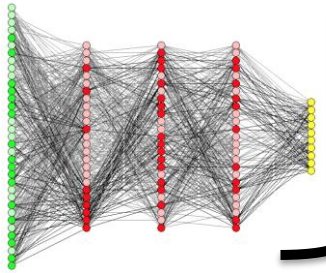
Actuators

Servos

Dynamic adaptation

Powerful

Parallel Edge Computing



Artificial Intelligence
Deep Learning Inference
Lightweight Training

Edge computing and deep learning with feedback for continuous improvement

5G

ESnet

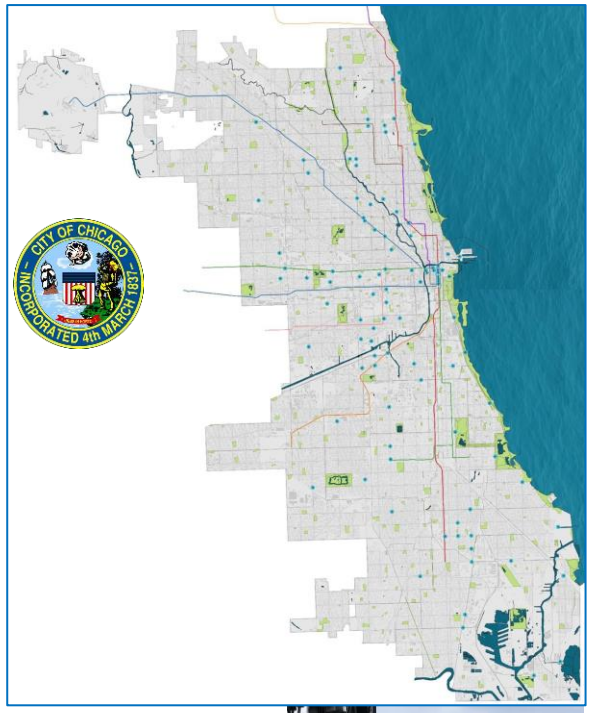
Reduced, Compressed data

New inference (model)
Adaptive steering

HPC

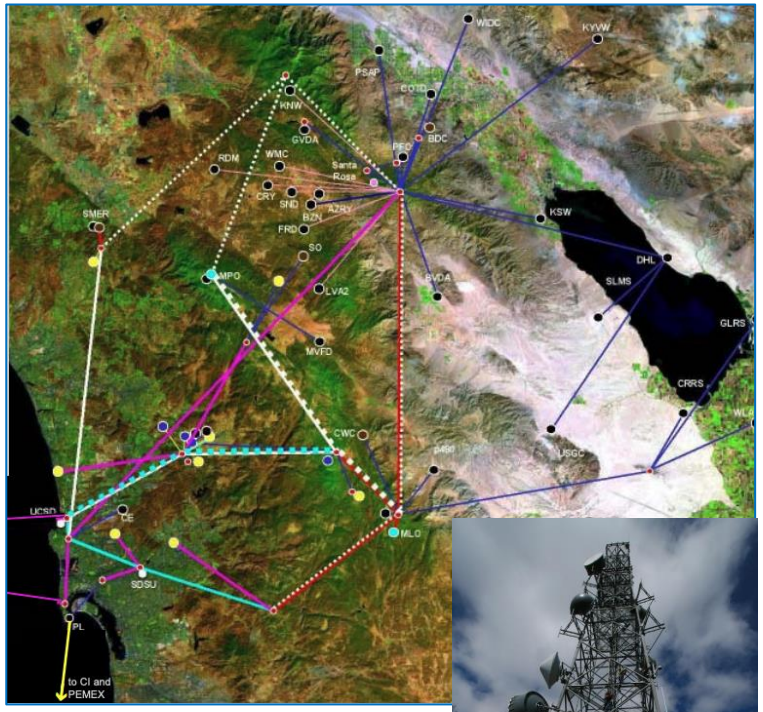
Deep Learning Training
Simulation / Forecast

SAGE Partner Instruments



ARRAY THINGS

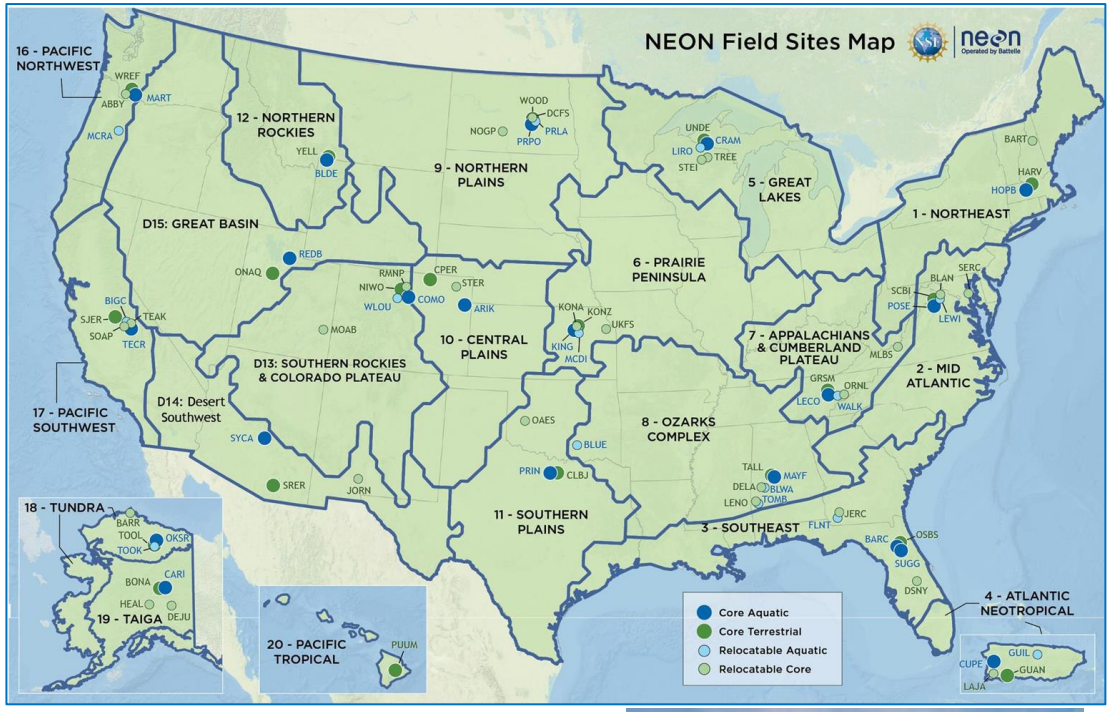
AoT: ***Neighborhood*** scale urban environment and activity.



HPWREN/WIFIRE: ***Regional*** Environmental Conditions and Events.

HPWREN

SDSC
SAN DIEGO SUPERCOMPUTER CENTER



NEON: ***Continental*** scale ecology and environment.

neon
National Ecological Observatory Network

Closing Remarks

- Arm brings mature SW stack (community driven) and competitive HW (licensees)
- Arm's business model empowers customers to design solutions for their markets
  - The A64FX is designed by Fujitsu to focus on RIKEN applications
  - Takes advantage of the software ecosystem (redhat, gnu, etc) from the community
  - Graviton2 & Ampere Altra leverage Arm IP: quick-to-market solutions that compete (+ use software)
- Arm has a very healthy roadmap and will continue to create ISA and IP solutions for the data center. Indeed growth in the hyperscalar market is a significant part of our future
- Arm is growing into cloud & hpc, from the device where we enjoy a healthy business
- The need of hyperscalers to innovate provides ever-more powerful Arm options for HPC
- We continue to invest in software in collaboration with the community
  - A-HUG.org is the HPC community

© 2021 Arm Limited

arm NEOVERSE
The Cloud to Edge Infrastructure Foundation
for a World of 1T Intelligent Devices
Thank You!