Delivering the Security Service Edge: High-Performance & Zero Trust

April 5, 2022

Notices and Disclaimers

- Intel technologies may require enabled hardware, software or service activation.
- No product or component can be absolutely secure.
- Your costs and results may vary.
-

Welcome

Jeff Sharpe
Senior Director
5G/IoT Edge AI Solutions

5G • IoT • Edge • Cloud
AI/ML/DL • Enterprise
https://linkedin.com/in/jeffbsharpe

Jay Vincent
Senior Solution
Architect

Edge • Cloud • Cybersecurity
• Enterprise • HPC
https://linkedin.com/in/jay-vincent

Dr. Ken Urquhart
Global Vice-President
5G Strategy

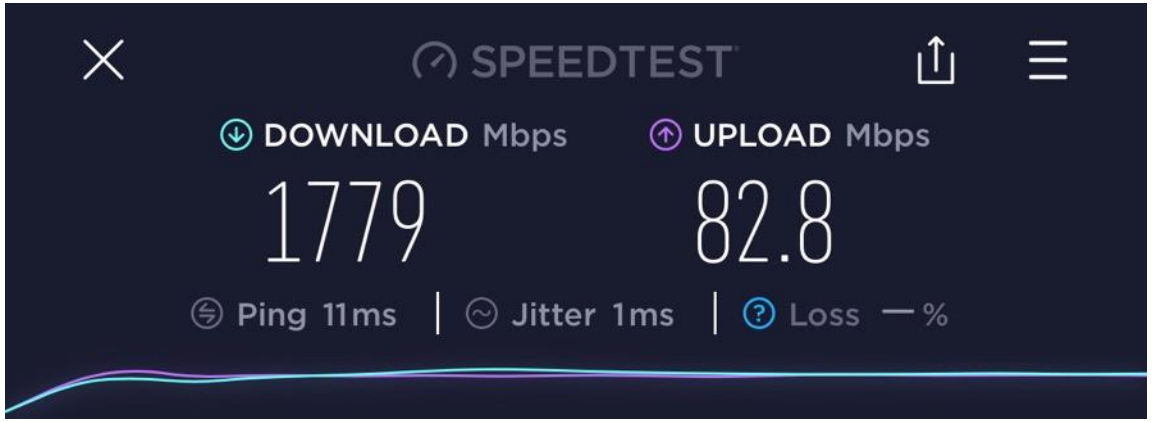
5G • Cybersecurity • AI • Engineering
• Quantum Computing • Cats
https://linkedin.com/in/kenu

5G

1. FAST
2. LIVE
3. MASSIVE

FAST → mmWave

- More bits per second
- Up to 10 Gbps
- 10 to 100x 4G-LTE
- Network slicing



Source: Intel "5G Technology Overview"

5G

1. FAST
2. LIVE
3. MASSIVE

LIVE

- Ultra-low latency (~1 ms)
- Edge compute is the 5G network + cloud upgrade for Mobile & Internet
  - Compute & data resources available at nearest node possible
  - Enables new & existing apps/services to run efficiently and cost effectively

Source: Intel “5G Technology Overview”

5G

1. FAST
2. LIVE
3. MASSIVE

MASSIVE

- 1,000,000 devices per km^2
- 1000x bandwidth/area

AND

- radios: $1/10^{th}$ energy usage
- ultra-precise location info

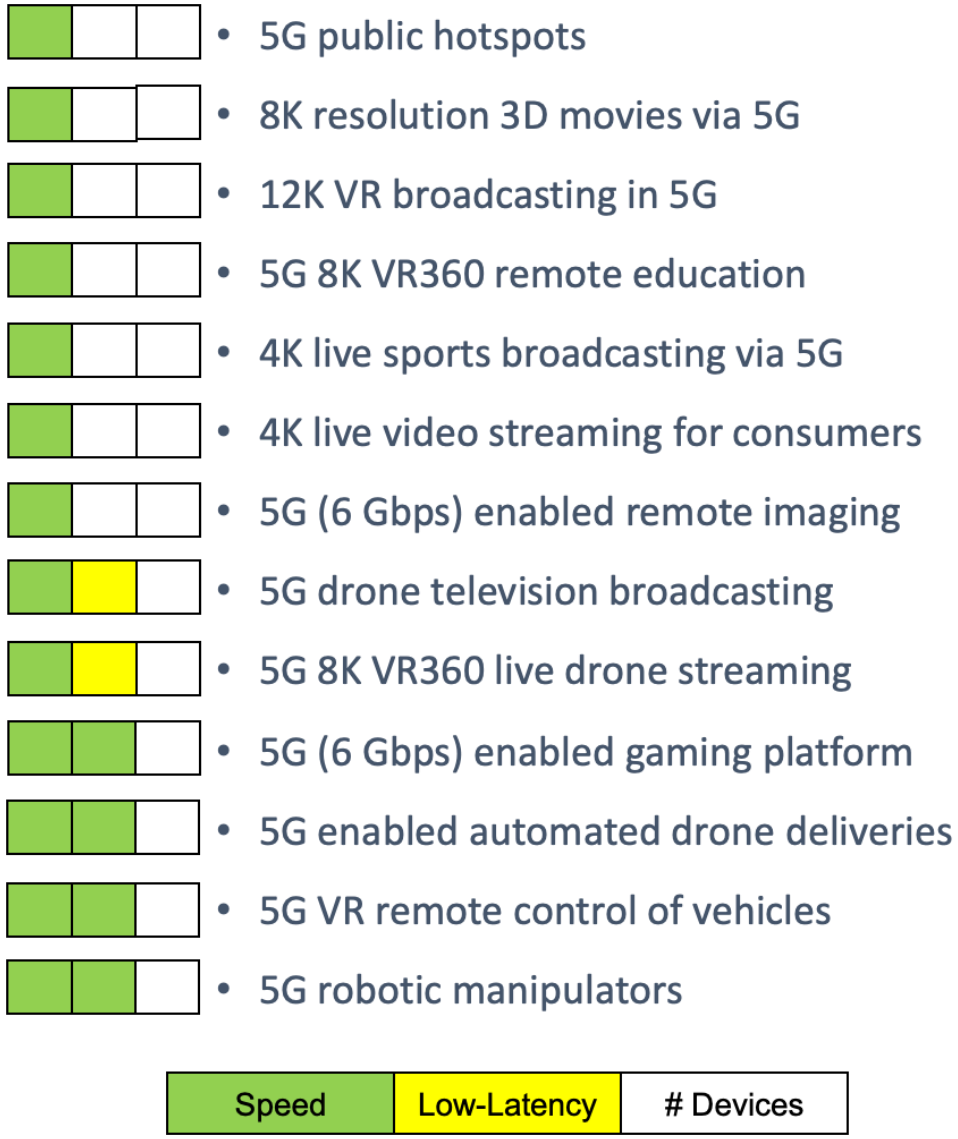
Source: Intel "5G Technology Overview"

Top Enterprise Edge Use Cases

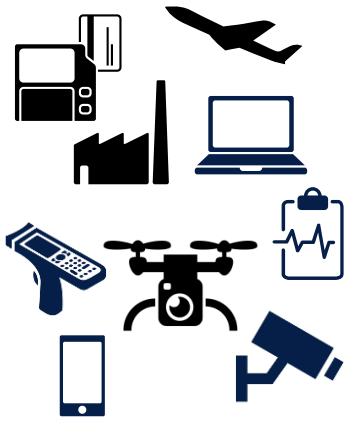
Enterprise edge use cases

- Industrial IoT
- Smart Cities
- Smart Grid
- Smart Metering
- Smartphones
- Connected Health
- Telehealth / Telemedicine
- Retail Immersion
- Disaster Management
- Emergency Management
- Digital Logistics
- Remote Monitoring
- Intelligent Navigation
- Automated Parking Management
- Facilities Management
- Energy Optimization
- Machine to Machine Comms
- Autonomous Driving
- Content Delivery
- Medication Management
- Streaming media
- Connected HVAC
- Precision Asset Tracking
- Surveillance Systems
- Augmented Reality
- Virtual Reality

5G use cases in production today



Where we are today…

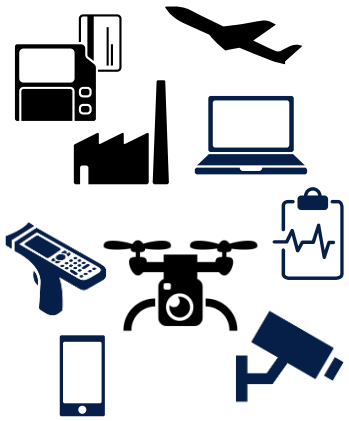


Any Device / Data Source

- The Internet is no longer just for humans – expect a surge of IoT, XR, and other devices
- These devices have limited CPU, GPU, Memory, Storage, Limited Power, etc.
- Ultra-low latency for workloads on the device
- (Mostly) secure when isolated from the Internet

Public/Private Clouds

Where we are today…



Any Device / Data Source

- Private Enterprise Clouds and Hyperscaler Clouds (AWS, Azure, GCP)
- Great toolsets to develop, build, and deploy solutions (to each cloud)
- Lots of compute power, memory, storage, and high-throughput networks
- All the scalability you need

Public/Private Clouds

Where we are today…

Any Device / Data Source

- Security is a concern as your traffic crosses multiple networks you don't control (and can't verify as secure)

Any network(s)

Public/Private Clouds

Where we are today…

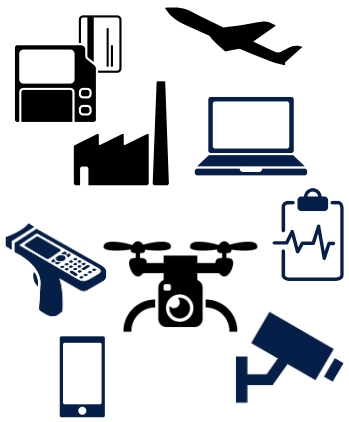
Any Device / Data Source

Real Time

Any network(s)

Public/Private Clouds

Where we are today…



Any Device / Data Source

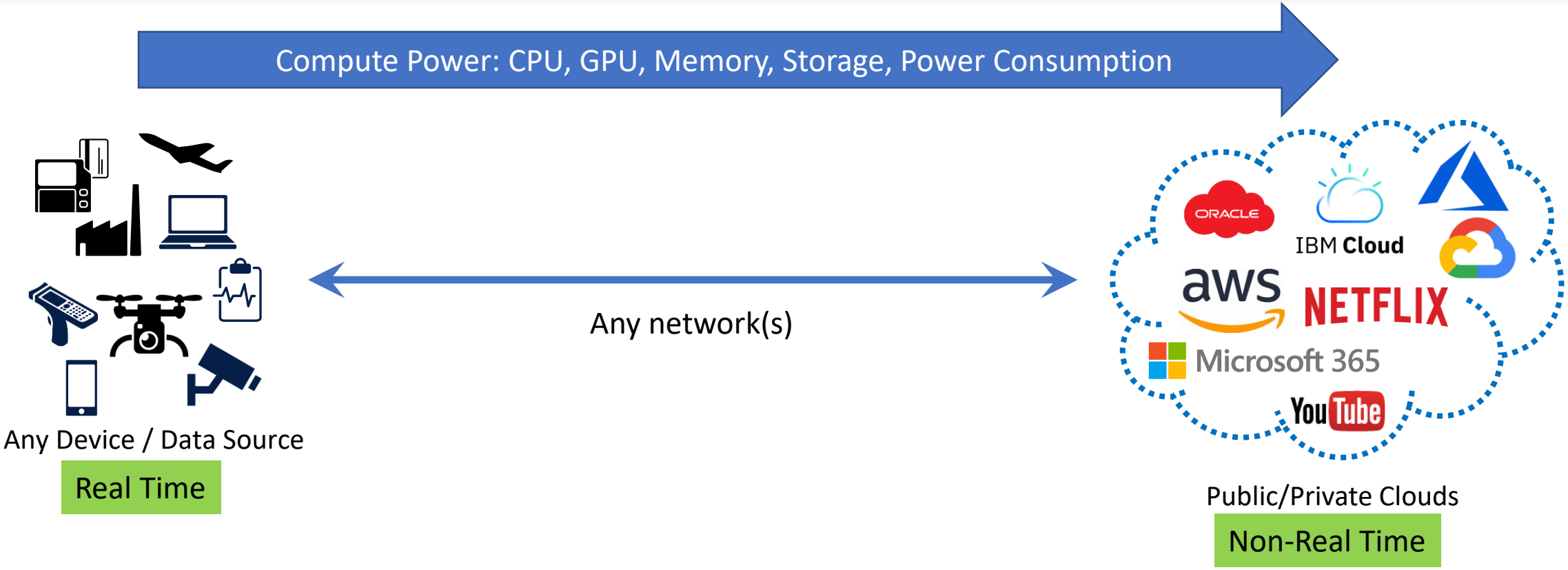
Real Time

Any network(s)

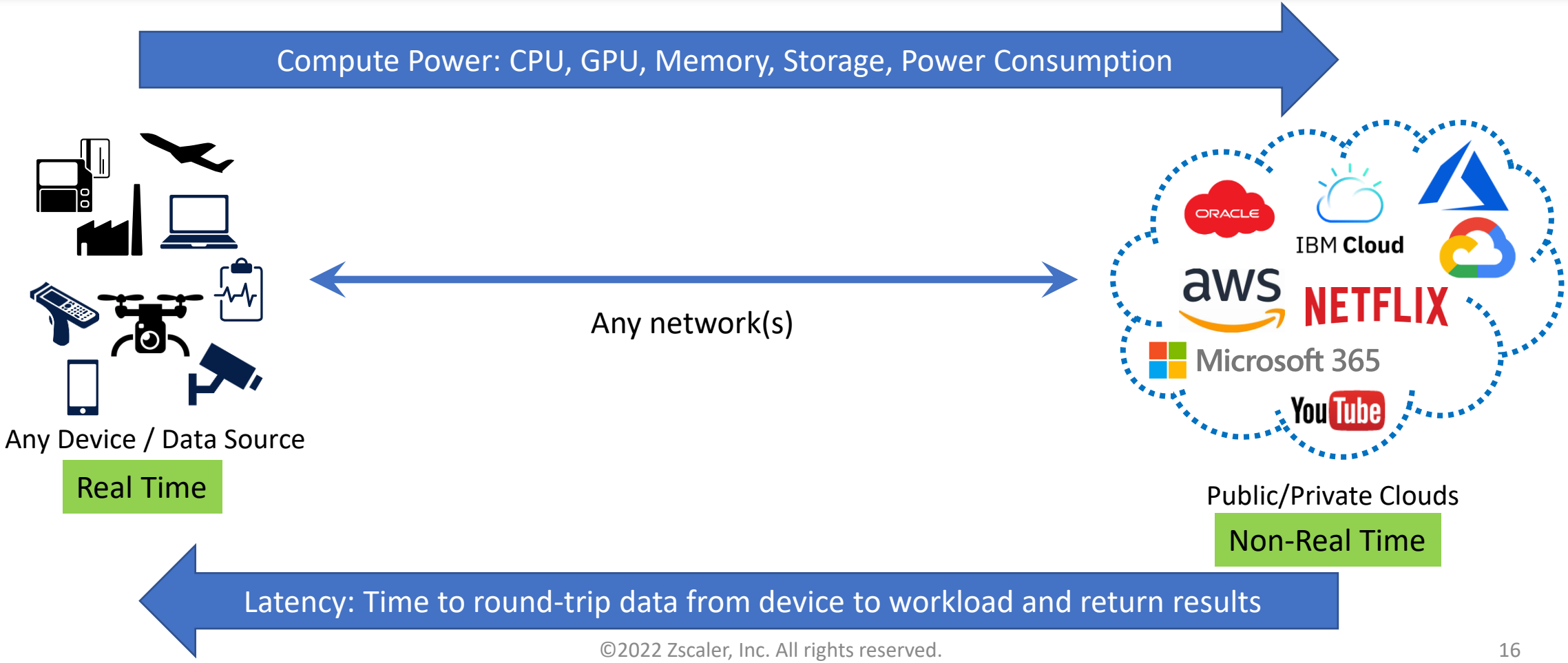
Public/Private Clouds

Non-Real Time

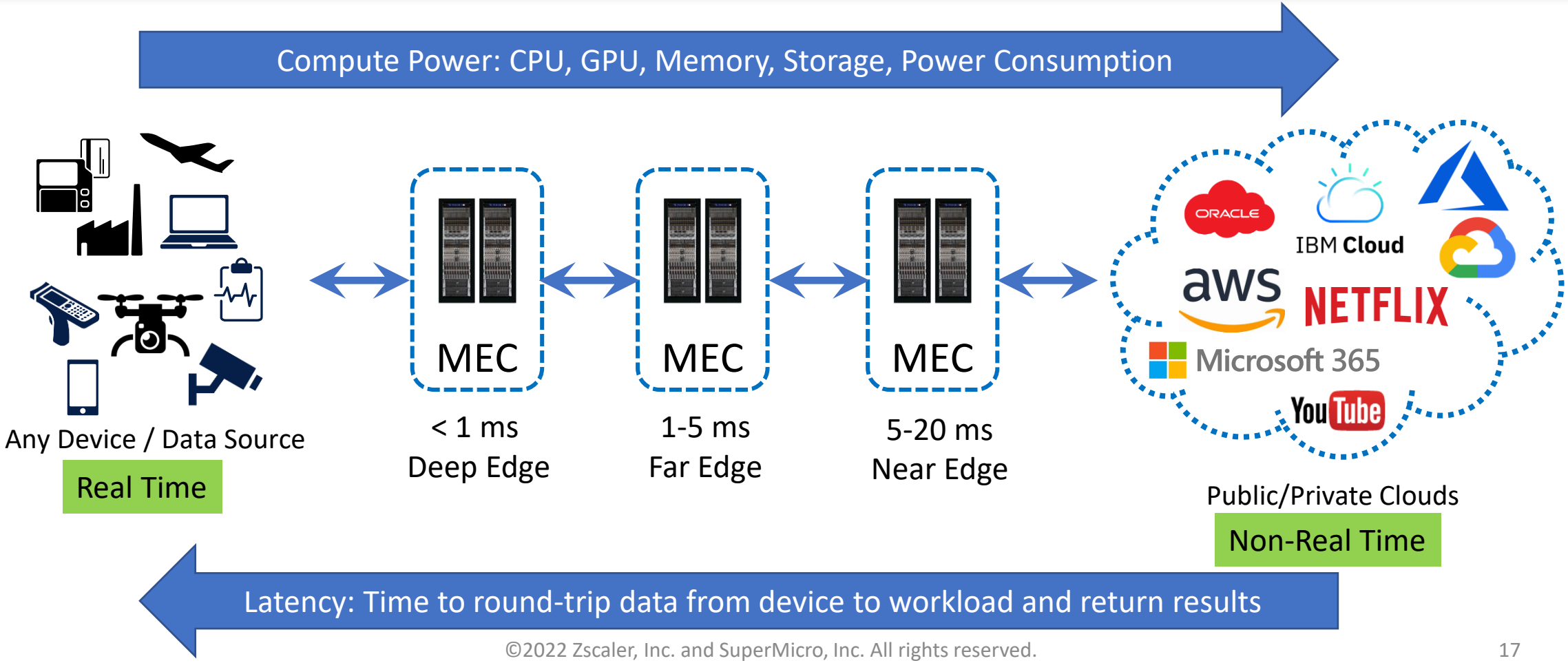
Where we are today…



Where we are today…



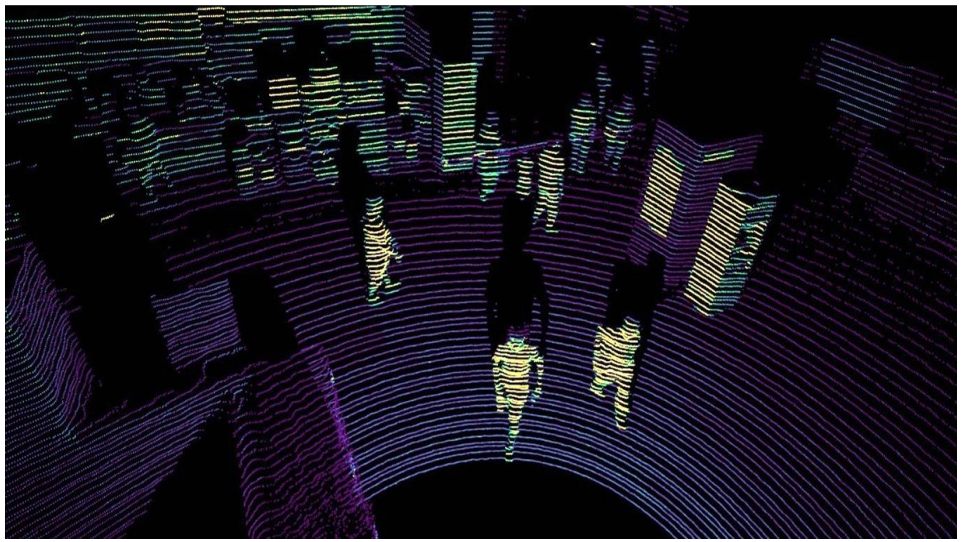
Where we are going…



Use case: Augmented Reality

TAQTILE

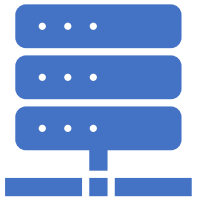
TaqTile AR for Robotics
- No-code robot programing
- Delivers AR for training, preventative maintenance & improved performance
- 12 ms or less latency



Lidar for recognition/tracking:
- Detects objects in a volume of space
- Track objects entering and exiting
- Low latency = Near real-time observation
- Multiple verticals

OUSTER

New Edge Servers & New Optimizations from SuperMicro

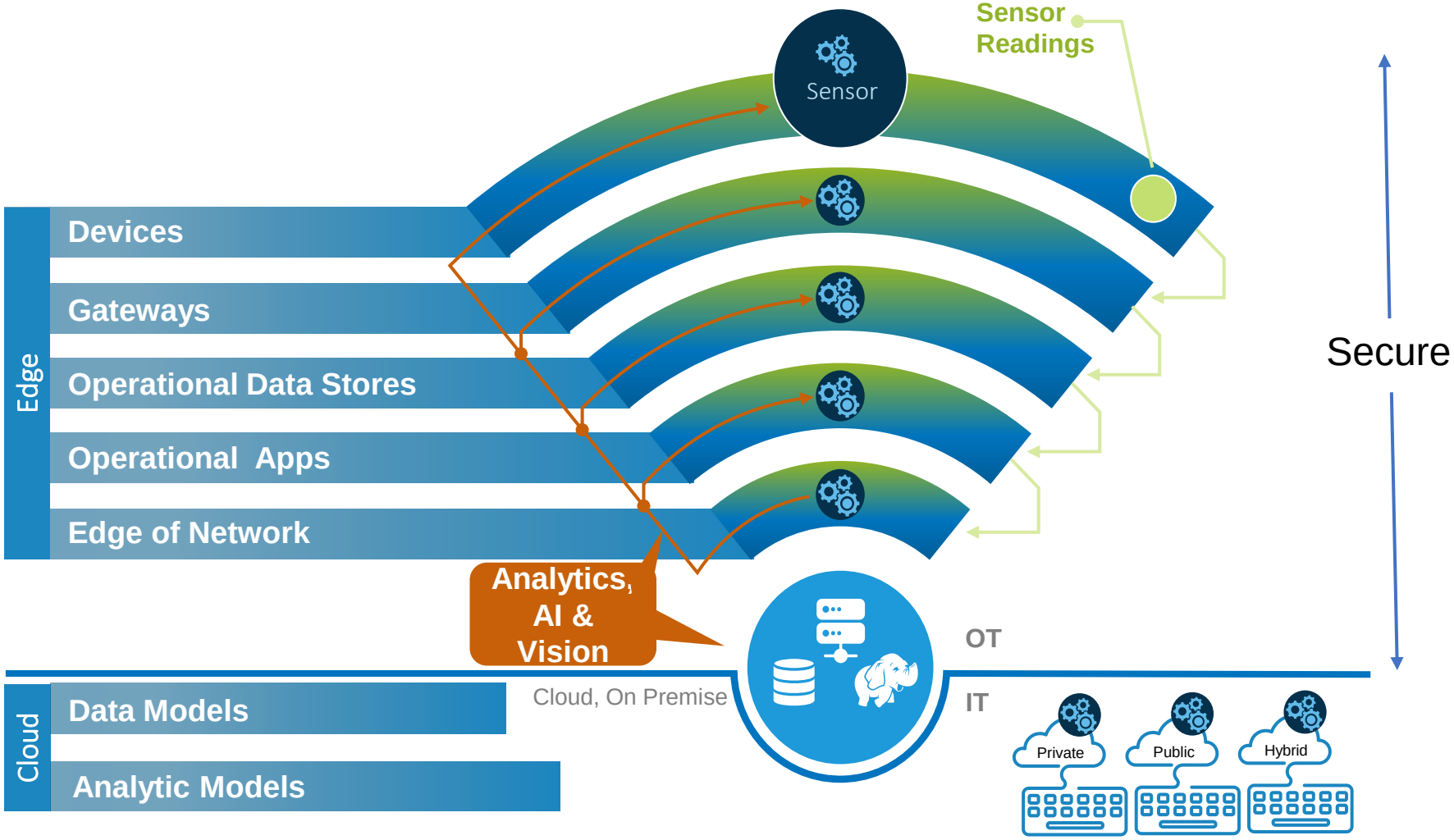


Market Evolution Snapshot

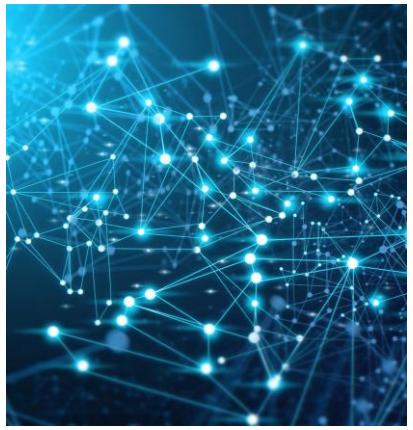
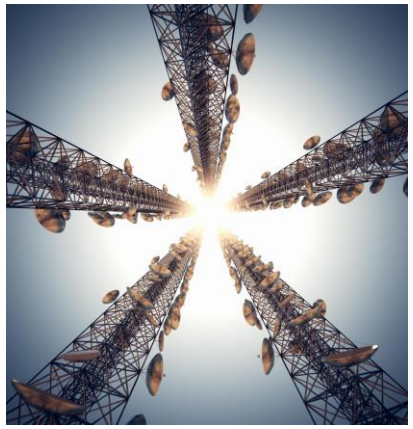
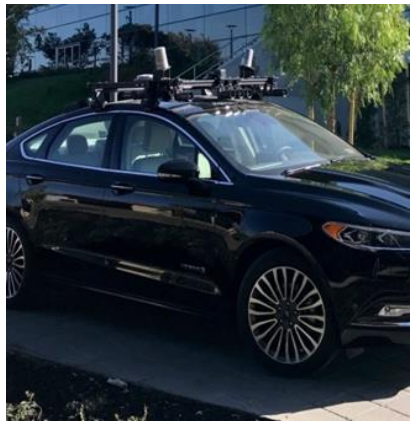
- Momentum into Wireless & High Throughput
- Merging of common elements for heavy workloads
- Network Architecture for Edge-to-Cloud platforms
- Nano-clouds / Edge devices for low latency apps
- Increased use of AI inferencing and training
- Network & Device security for Zero-Trust/SASE network is a critical element
- Video/CDN increases the need for higher bandwidth and edge security
- Distributed Automated, Self Healing/Managing Intelligent platforms
- “Killer-Apps” driving the need for increased AI/ML/DL based systems

Edge-to-Cloud AI Infrastructure

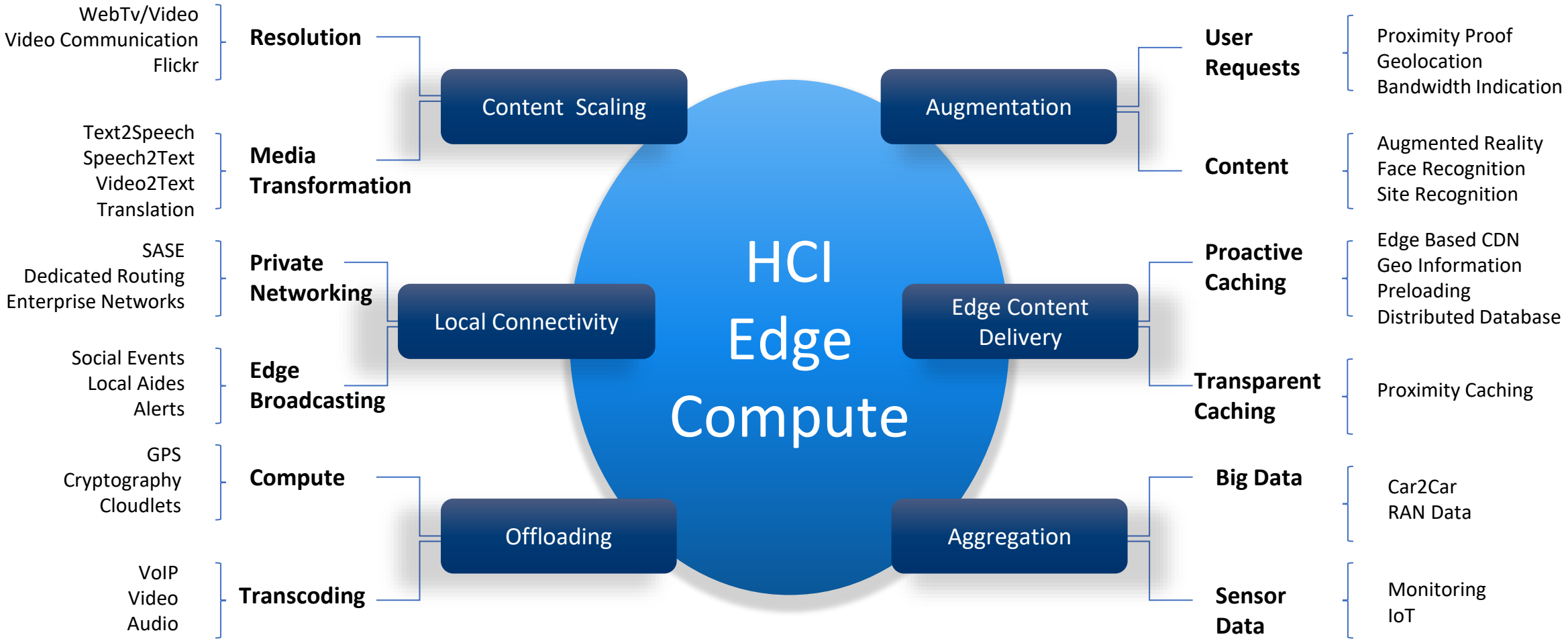
- Industry 4.0
- AR/VR
- Retail
- Venues / Smart Places
- Smart Cities
- Transportation (V2X)
- Healthcare
- Telecom MEC



Core Requirements of Edge-to-Cloud Resources

High Throughput	Privacy & Security	Spectral Efficiency	Low Latency Services	Exponential Data Growth
				
DPI Backhaul Switching Policy	DDOS NGFW IPSec DLP Intrusion	Slicing Mesh DAS Micro/Macro Cells MIMO Private 5G	Autonomous xx AR/VR Recovery/HA Safety/Response Speech Rec.	Transportation Smart Places Renewable Energy Media/Video QoExperience

Edge Service Delivery (GPU/CPU/FPGA)



Edge Appliance Prerequisites

Self Healing and Resiliency at the Edge

Edge Systems are remotely managed and unmanned. Provides low latency and high throughput delivery for VM's & Containers – some of which are expected to be up 99.9999% of the time

Secure and Zero-Trust

Expanding the abilities of multiple compute devices attached in a hybrid environment of public and private networks increases a security risk. Key areas include: assume breach is inevitable or has already occurred; Always limit access to only what is needed; Looks for anomalous/malicious activity everywhere

Cloud native approach and DevOps enablement

Service Providers & NFV vendors are driving a cloud native approach. Service providers acknowledge the evolvement of software in every part of the network allowing cloud-based automation for infrastructure deployment, operations and VM/Container management.

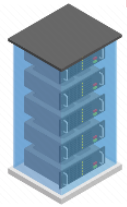
Role of Open Software and Hardware Architectures

Open Architectures for software and hardware infrastructure to develop, deploy and manage vNFs, containers and APIs for services – enablement of multiple, integrated partners providing guidance and standards for delivery.

Enabling Edge to Cloud

Edge Platforms

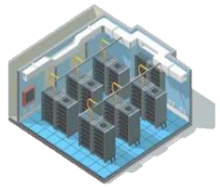
CPE



Edge DC

Network Edge

Central Office / Edge Cloud



Regional DC

Remote DC

<1ms/Realtime | 2-5ms | <10ms | <20ms | >50ms

Customer Premises

Edge-Optimized

Network-Optimized

Resource Management

ML/Analytics/Archive

Edge requirements

- ✔ Built for Net/Comms
- ✔ 2x workloads
- ✔ Power Needs
- ✔ Increased internal connections
- ✔ Redundancy/HA
- ✔ Expandability
- ✔ Flexible
- ✔ Rackmount or standalone
- ✔ NEBS and harsher certifications

- ✔ Cost Effective
- ✔ Enterprise driven
- ✔ Normalized Workloads
- ✔ uCPE & Intel Select solutions
- ✔ SMB capable
- ✔ Lower power consumption

vNF Workloads

SYS-210SE

SYS-220HE

X-Large

(Dual Socket Xeon-SP 3 or more GPU/VPU)

SYS-E403P

Large

(Single Socket Core / Xeon-SP 1 - 3 GPU/VPU & I/O options)

SYS-510D

SYS-110P

Medium

(Single Socket Core / Xeon-D 1 GPU/VPU & I/O options)

SYS-E50/100/302

SYS-E5019A

Small

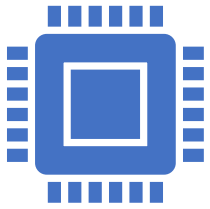
(Atom/Core)

Throughput

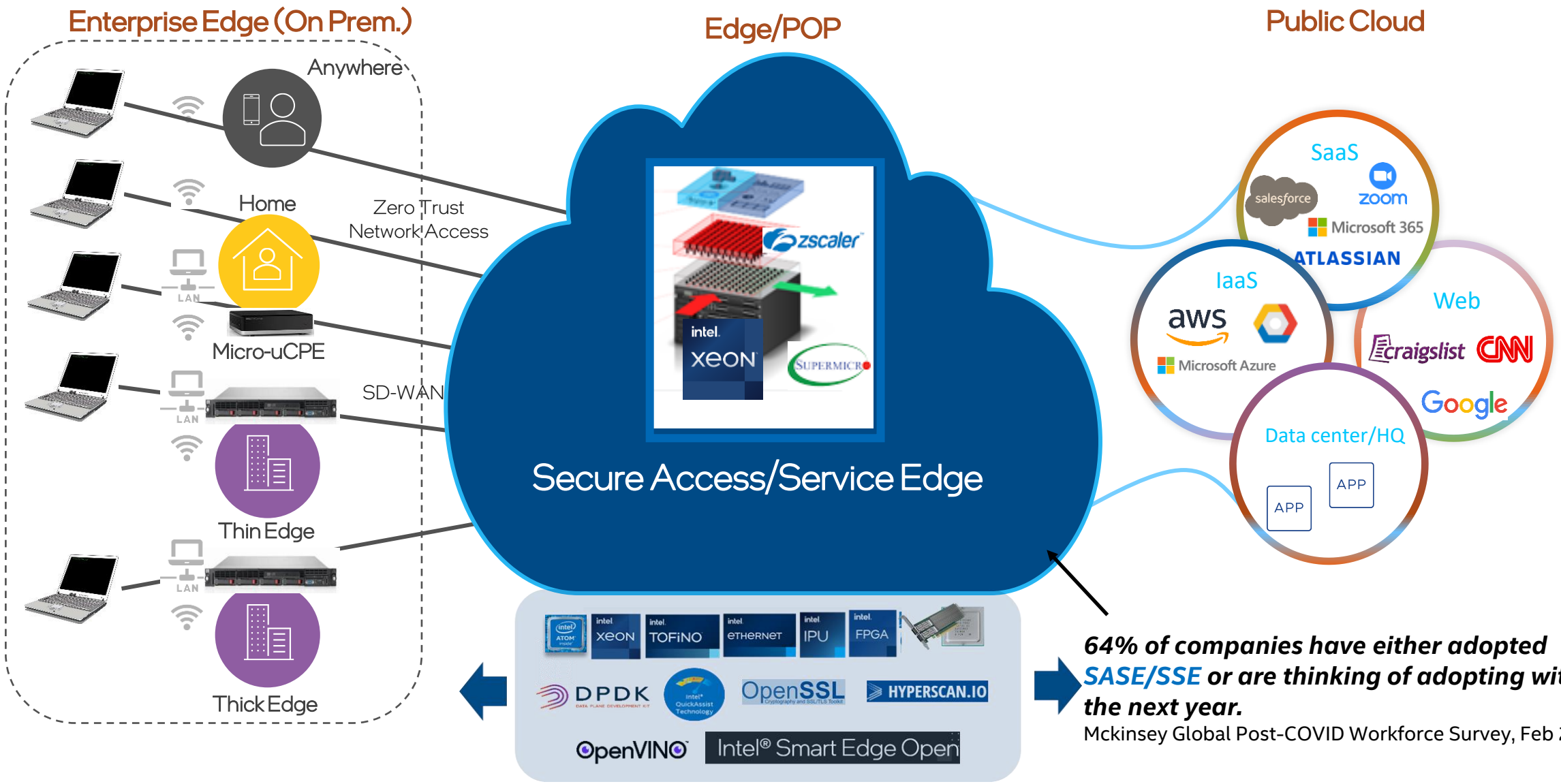
Minimal Port / Throughput

Maximum Port / Throughput

Intel Powers the Edge to Unlock Value of New Use Cases



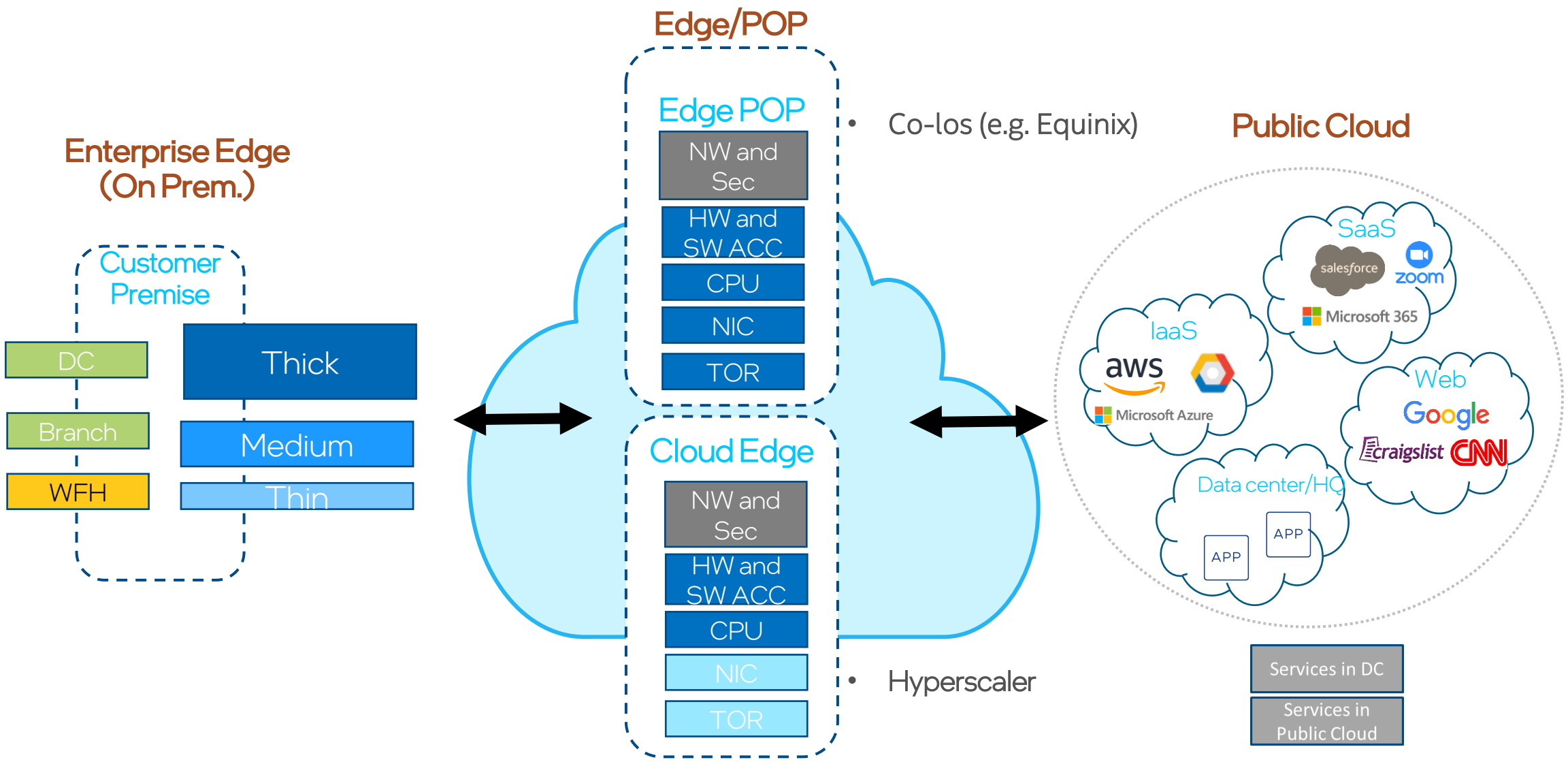
Enterprise Edge-Cloud Landscape



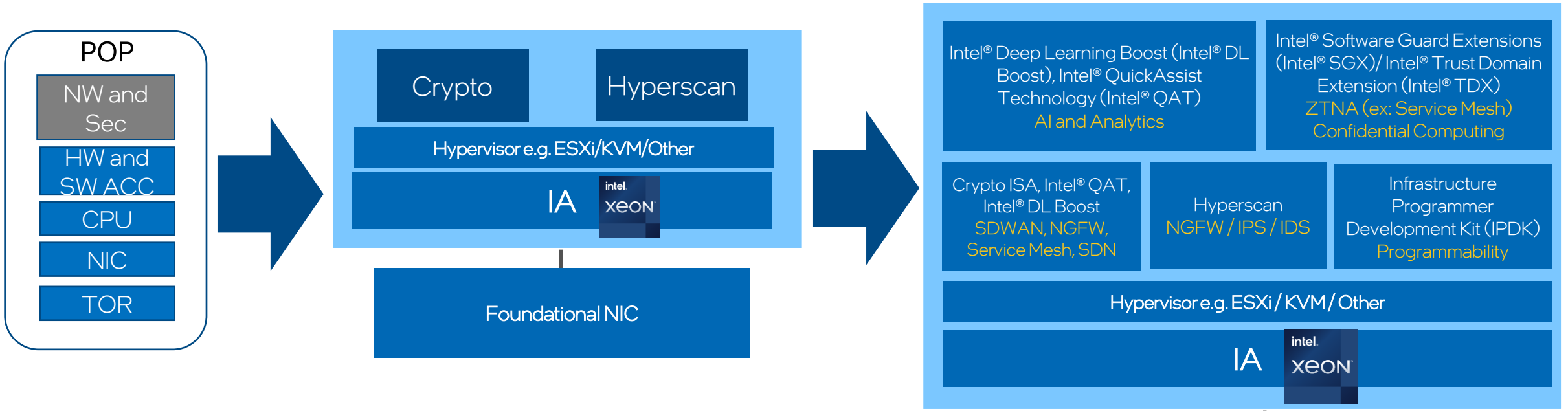
64% of companies have either adopted SASE/SSE or are thinking of adopting within the next year.

Mckinsey Global Post-COVID Workforce Survey, Feb 2021

Edge to Cloud Architecture



Edge/POP Platform Architecture



Lower TCO with a consistent cloud -focused platform approach across edge locations

- Applications and Services
- Middleware
- Resource and service orchestration
- Operating system
- Edge hardware platform

Key challenges to overcome

- Deliver platform consistency & scalability across diverse edge location requirements
- Optimize cloud native frameworks to meet stringent edge KPIs and network complexity
- Leverage a broad ecosystem and evolving standards for edge computing

Intel® Xeon® Scalable Processors

DATA CENTER CPU OPTIMIZED FOR AI

INTEL® ADVANCED VECTOR EXTENSIONS 512 (INTEL® AVX-512)
INTEL® DEEP LEARNING BOOST (INTEL® DL BOOST)
INTEL® OPTANE™ DC PERSISTENT MEMORY

Intel® DL Boost Technologies			
Microarchitecture	AVX512_VNNI	AVX512_BF16	AMX
Client			
Core 10th Gen	✓	X	X
Server			
Xeon SP Gen 2	✓	X	X
Xeon SP Gen 3H	✓	✓	X
Xeon SP Gen 3	✓	X	X
Next Gen Xeon SP	✓	✓	✓

2019

XEON SP GEN 2

14NM
NEW AI ACCELERATION (VNNI)
NEW MEMORY STORAGE HIERARCHY

2020

XEON SP GEN 3 H

14NM
NEXT GEN INTEL® DL BOOST (BFLOAT16)

2021

XEON SP GEN 3

10NM
SUPPORTING INTEL® DL BOOST (VNNI)

2022

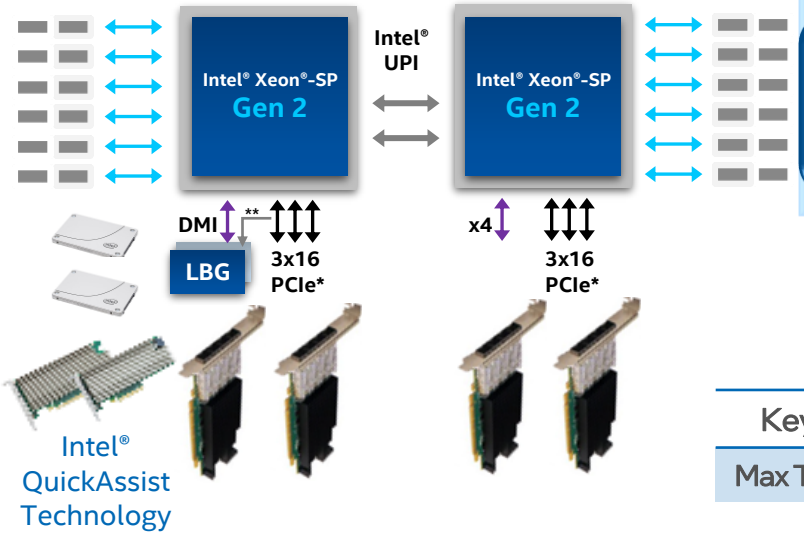
NEXT GEN XEON SP

NEXT-GENERATION TECHNOLOGIES
(Intel® Advanced Matrix Extensions (Intel® AMX)

LEADERSHIP PERFORMANCE

IA Powers the SASE

2S Intel® Xeon® Scalable processor Gen 2 Platform



NICs: XXV710-DA2 2x25Gbe
IO Density: 400Gbps Network IO Per 2RU

DDR4 DIMMs

DDR4/Intel® Optane™ Persistent Memory

PCIe* uplink connection for Intel® QuickAssist Technology and Intel® Ethernet

Up to 38cores/CPU
up to 112 logical cores

8x Memory channels @3200MTs

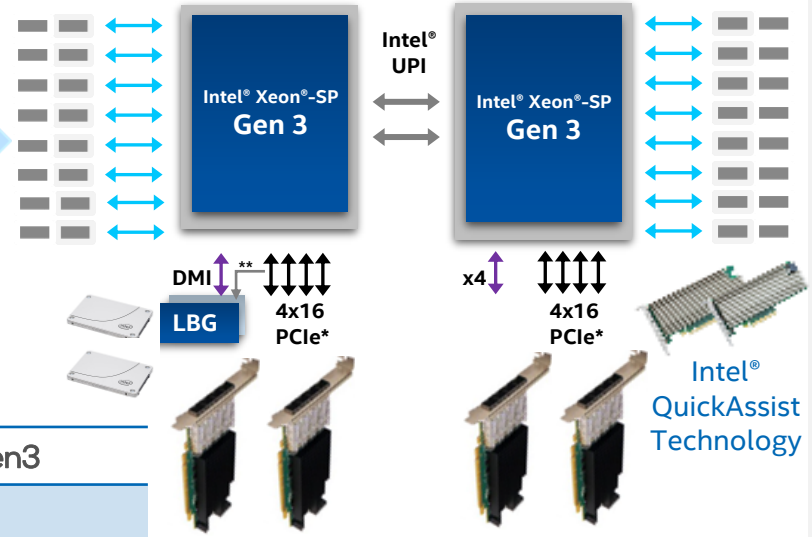
Next Gen Intel® Optane™ DCPMEM

128 Lanes PCIe Gen 4 IO BW 2048Gbps

Higher Performance Solution

Total Memory Encryption

April'2021 Launched
2S Intel® Xeon Scalable processor Gen 3



Key features	Xeon-SP Gen2	Xeon-SP Gen3
Max TDP	205W	270W
New instructions	VNNI	VNNI, Vector AES, VBMI, SHA extensions, VPMADD52
Intel® SGX	No	Yes, Confidential Computing
# DDR4 channels / Max Speeds	6 x 2933	8 x 3200
Optane memory support	Intel® Optane™ DC Persistent Memory Module (Apache Pass)	Next generation Intel® Optane™ DC persistent memory module
PCIE IO	PCIe 3.0, 48 lanes (x16, x8,x4)	PCIe 4.0, 64 lanes (x16, x8,x4)

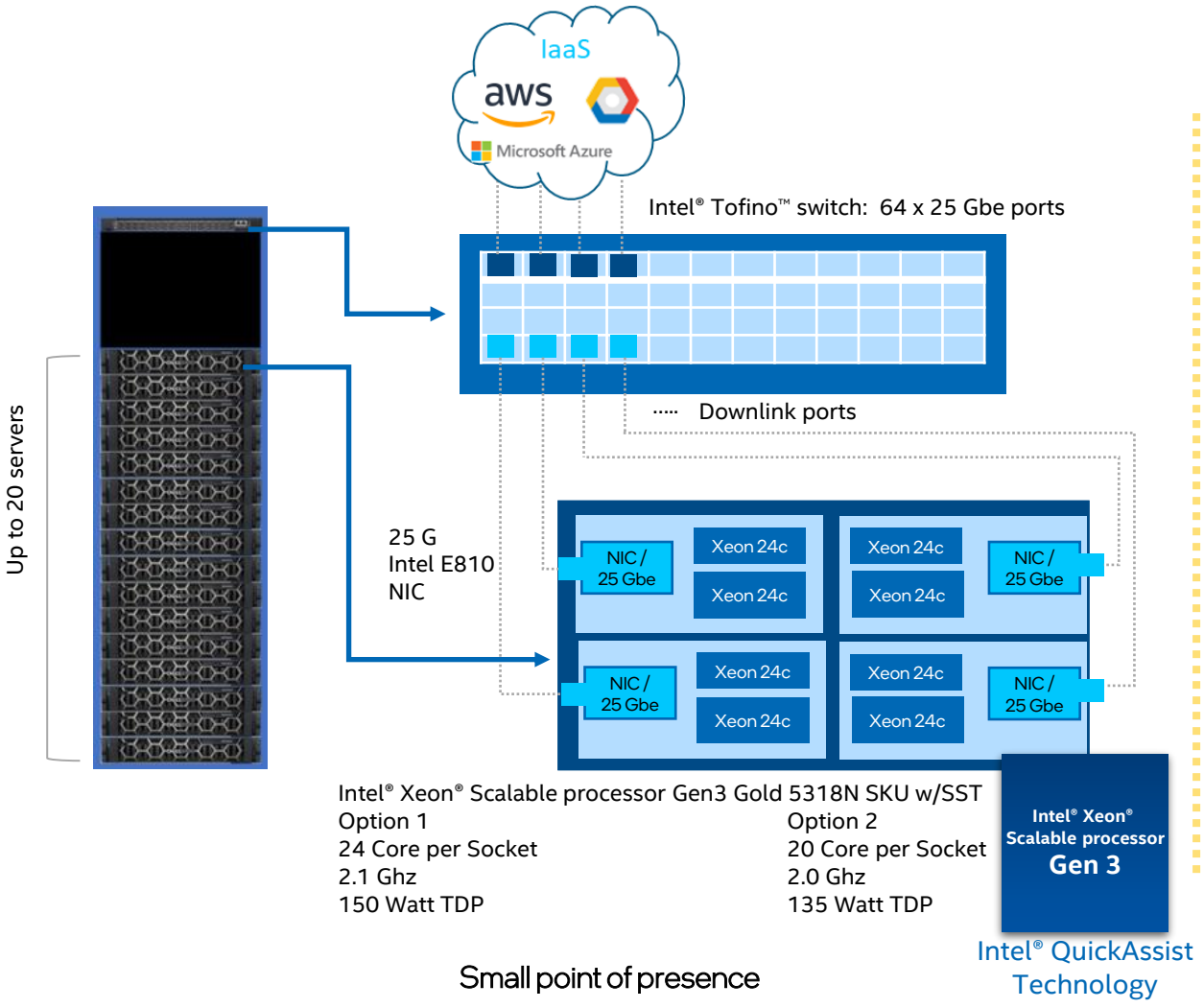
NIC's: E810 1x100, 2x50, 4x25 & 2X 100

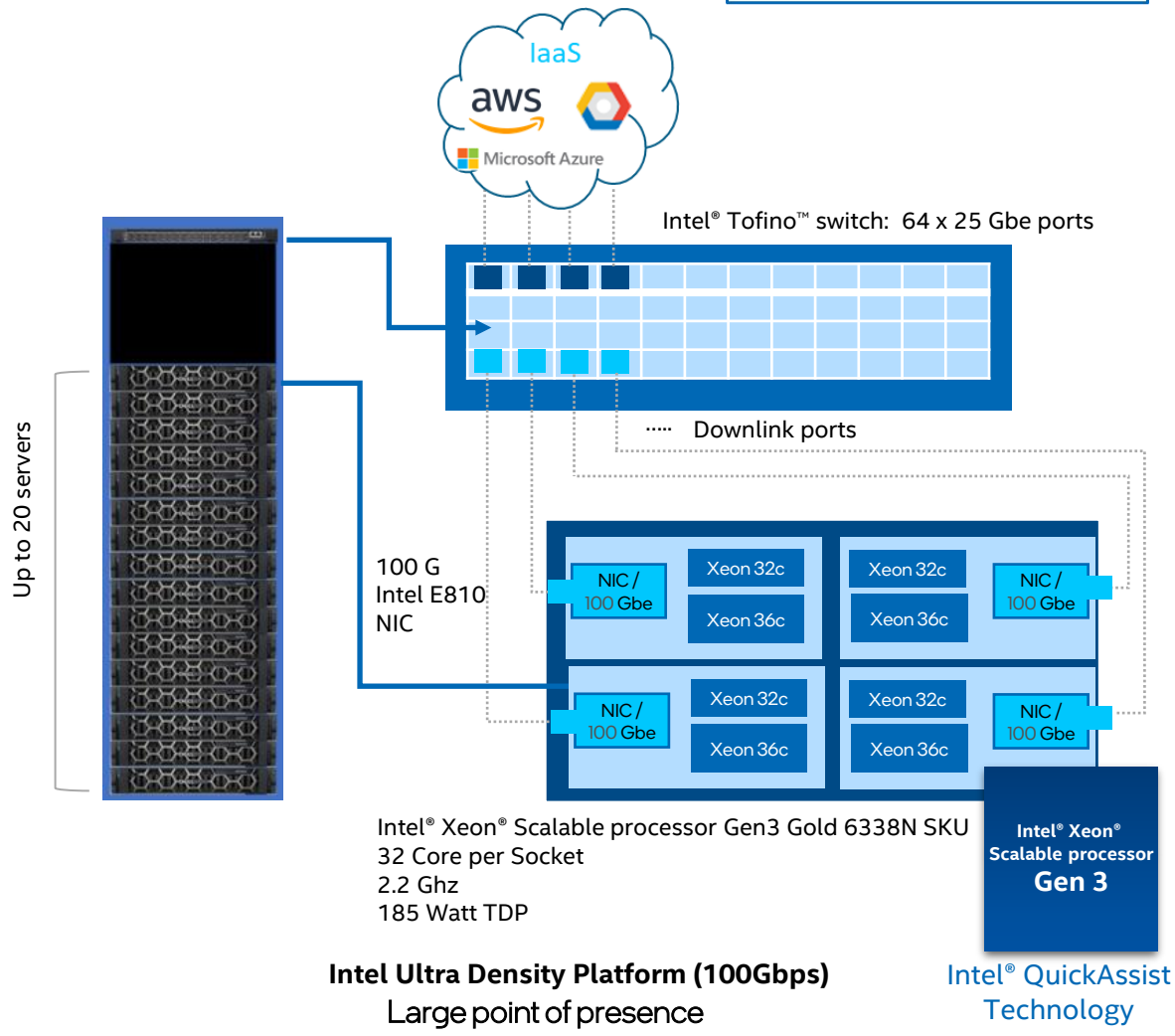
IO Density Possible: 800Gbps Network IO Per 2RU

Xeon-Sp Hero features

- SGX : Confidential Computing
- Crypto: IPsec/TLS/QUIC/WireGuard
- Intel® DL Boost: Analytics and AI use case

SSE POP Reference Configurations





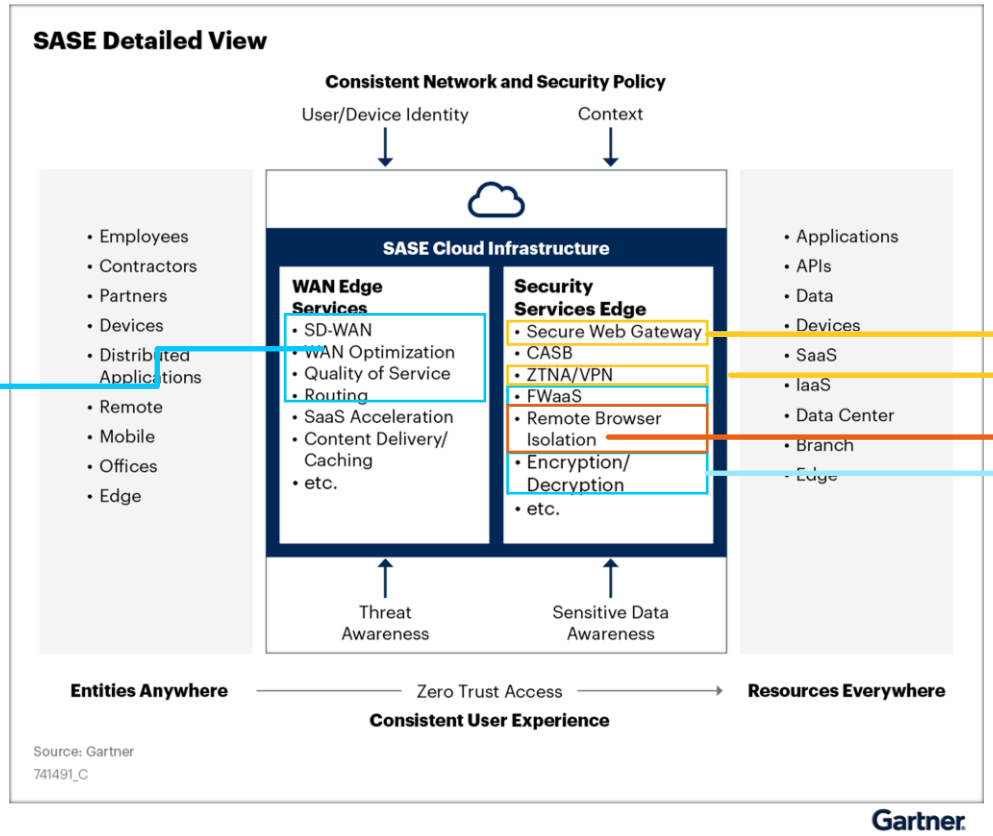
Extensive Intel Software Value

Pattern matching for DPI, IPS, WAF — HYPERSCAN.IO

User plane Stack — .io

I/O, Libs — DPDK

Load balancing — HDSLB



intel XEON GOLD

- uService/Service Mesh
- VM/IaaS, Power Saving
- Platform/Data Security
- Confidential Computing

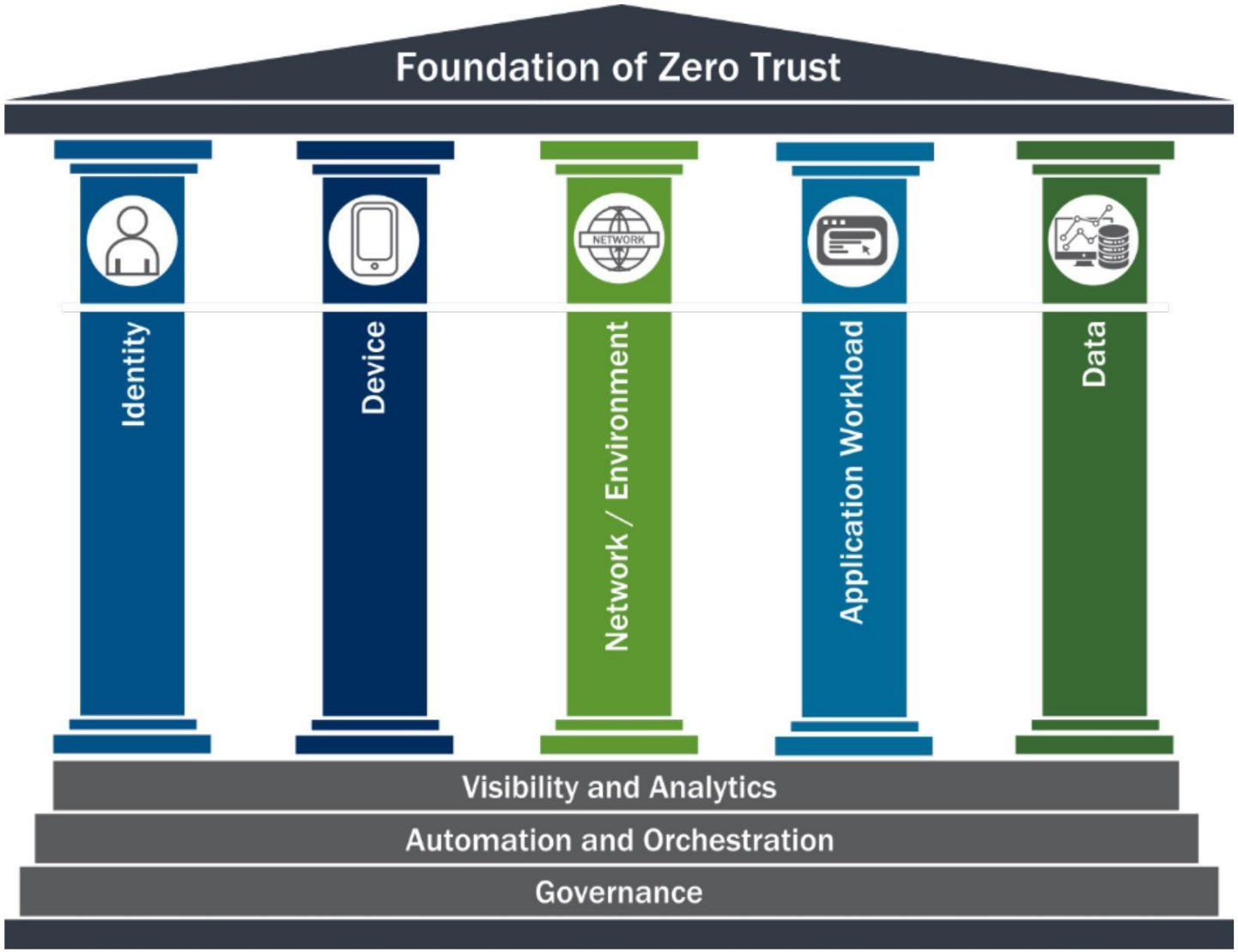
SWG/DNS	DNS Security, DNSoHttps URL filtering (+ Hyperscan), FQDN
ZTNA/ VPNaaS	Vault, IAM, Wireguard/IPSec VPNaaS, VPP/IPSec
Security AI / CPU Inference	TADK, Intel OneDNN, TensorFlow, OpenVINO, ONNX TensorFlow Lite / XNNPack
Encrypted traffic analysis,Https2.0 / 3.0, L7BS	OpenSSL, BoringSSL/QUIC Nginx
Remote Browser Isolation	Hyperscan, AES

Intel® QuickAssist Technology

What About Cyber-Security?

Zero Trust Security Model

- **May 2021: Executive Order on Improving the Nation's Cybersecurity**
  - https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/
- **March 2022: Statement by President Biden on our Nation's Cybersecurity**
  - https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/21/statement-by-president-biden-on-our-nations-cybersecurity/
- **Zero Trust approach**
  - Assume breach is inevitable or has already occurred
  - Constantly limit access to only what is needed
  - Looks for anomalous/malicious activity everywhere



From: U.S. Cybersecurity and Infrastructure Security Agency (CISA)

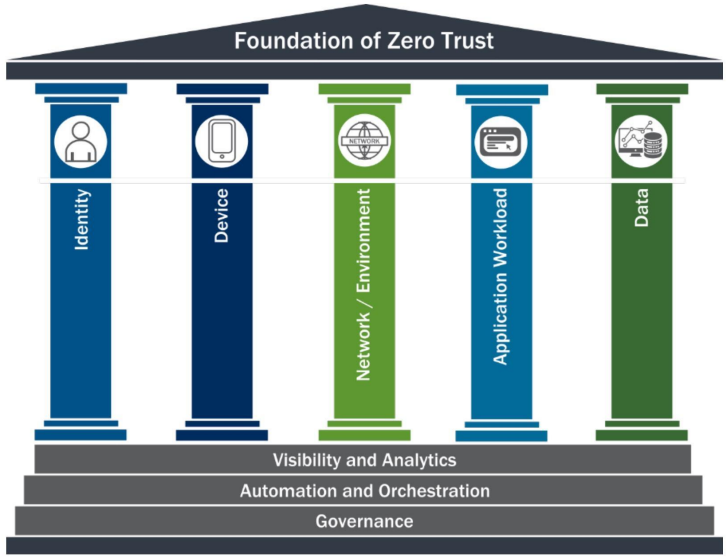


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user
2. from their device
3. over any network
4. to a specific application
5. to exchange data

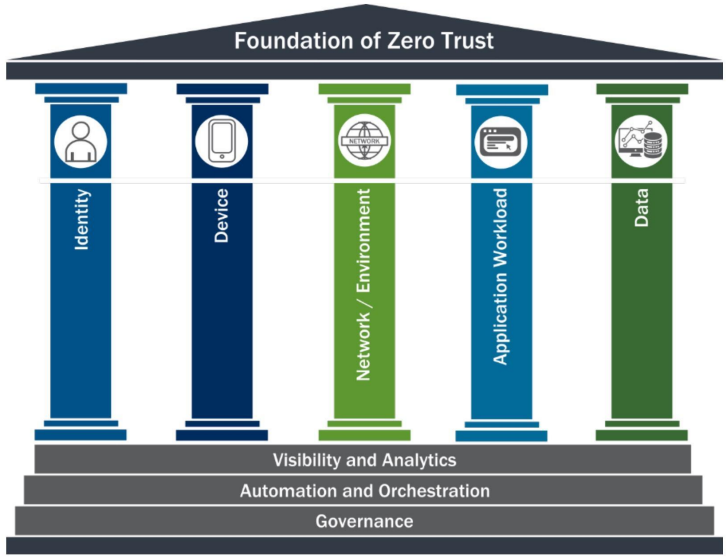


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

IDENTITY ← **Do I know you?** Employee/partner/device identity (Microsoft, Okta, Ping)

2. from their device

3. over any network

4. to a specific application

5. to exchange data

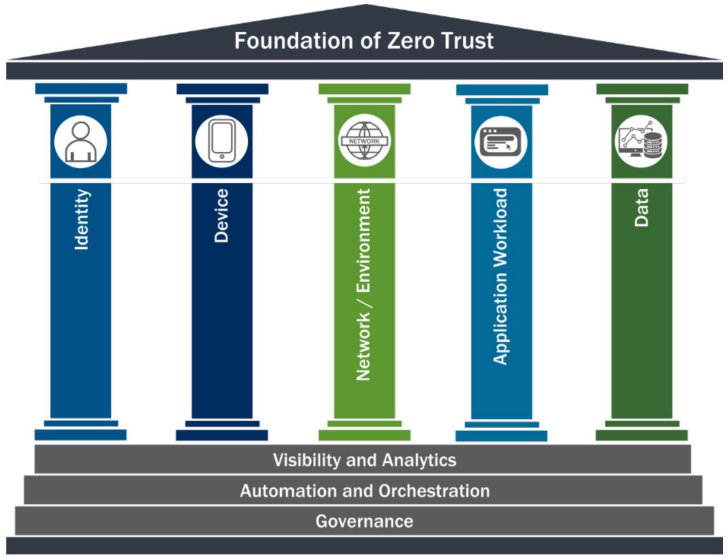


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

IDENTITY ←

Do I know you?
Employee/partner/device identity
(Microsoft, Okta, Ping)

2. from their device

DEVICE ←

Do we trust the device?
Device type/posture, managed/BYOD
(Microsoft, CrowdStrike)

3. over any network

4. to a specific application

5. to exchange data

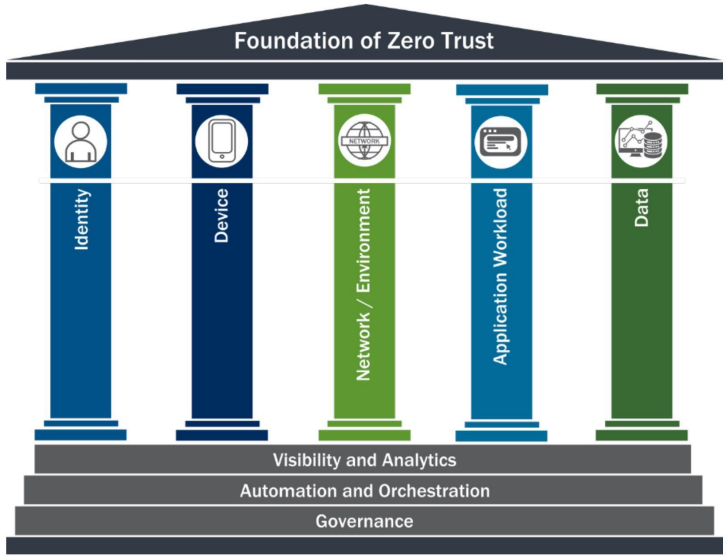


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

IDENTITY ←

Do I know you?
Employee/partner/device identity
(Microsoft, Okta, Ping)

2. from their device

DEVICE ←

Do we trust the device?
Device type/posture, managed/BYOD
(Microsoft, CrowdStrike)

3. over any network

NETWORK ←

Where are you going?
Internal, external, sanctioned SaaS, or other destination

4. to a specific application

5. to exchange data

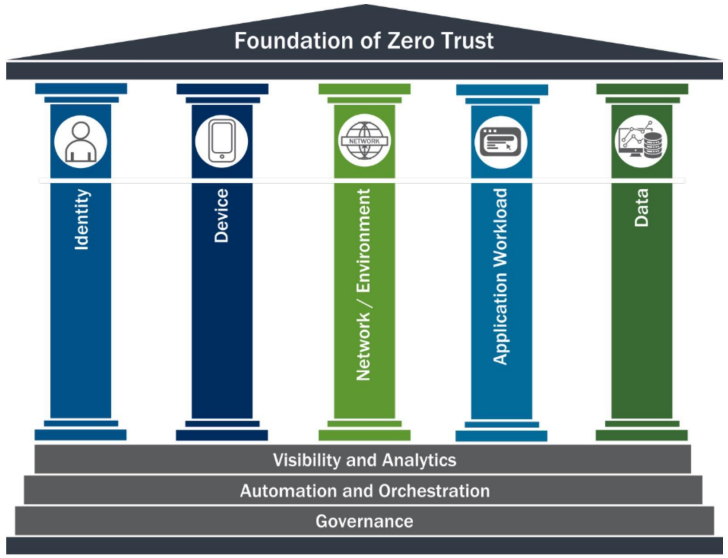


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

IDENTITY ←

Do I know you?
Employee/partner/device identity
(Microsoft, Okta, Ping)

2. from their device

DEVICE ←

Do we trust the device?
Device type/posture, managed/BYOD
(Microsoft, CrowdStrike)

3. over any network

NETWORK ←

Where are you going?
Internal, external, sanctioned SaaS, or other destination

4. to a specific application

APPLICATION ←

Do we trust the application?
Expected or anomalous behavior?
Usual ports and protocols?

5. to exchange data

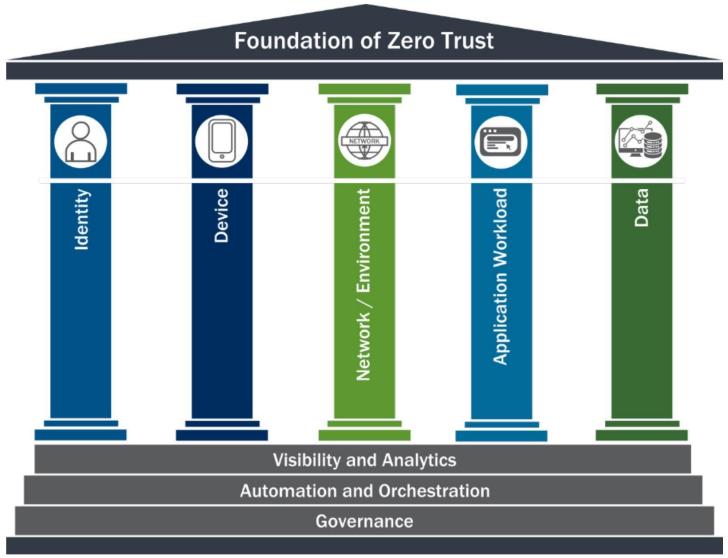


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

IDENTITY ← **Do we know you?** Employee/partner/device identity (Microsoft, Okta, Ping)

2. from their device

DEVICE ← **Do we trust the device?** Device type/posture, managed/BYOD (Microsoft, CrowdStrike)

3. over any network

NETWORK ← **Where are you going?** Internal, external, sanctioned SaaS, or other destination

4. to a specific application

APPLICATION ← **Do we trust the application?** Expected or anomalous behavior? Usual ports and protocols?

5. to exchange data

DATA ← **Do we trust the data?** Are you exchanging good or bad things? Content inspection (SSL)?

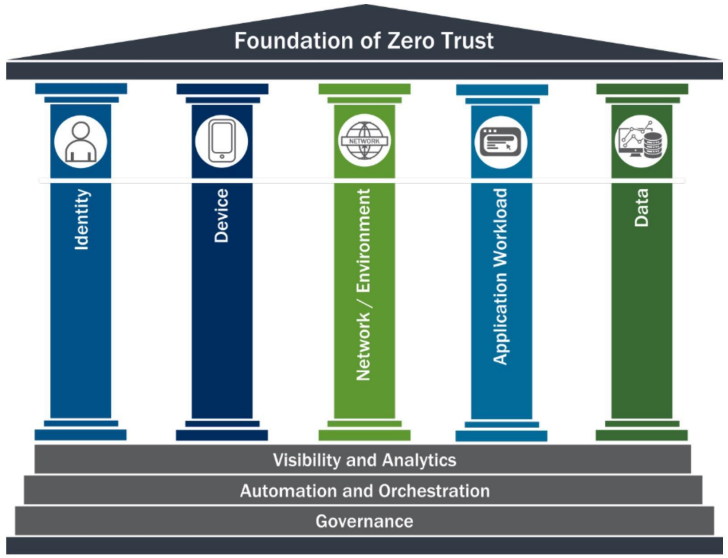


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

IDENTITY

2. from their device

DEVICE

3. over any network

NETWORK

4. to a specific application

APPLICATION

5. to exchange data

DATA

A. Governance

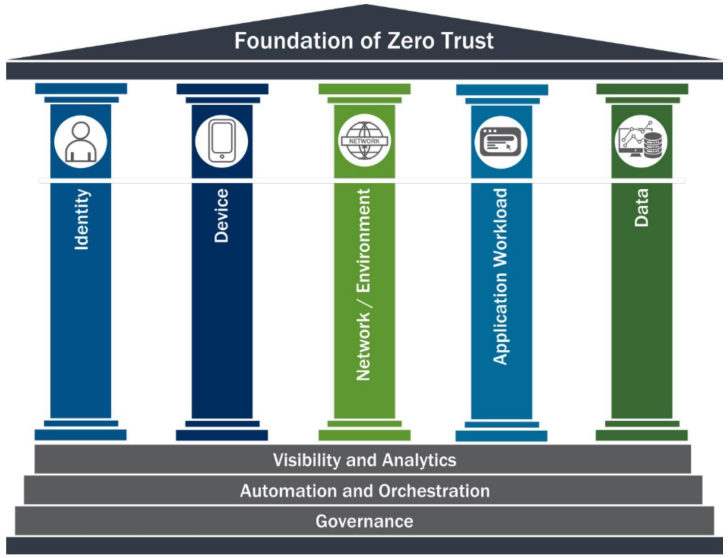


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

   IDENTITY

2. from their device

   DEVICE

3. over any network

   NETWORK

4. to a specific application

   APPLICATION

5. to exchange data

   DATA

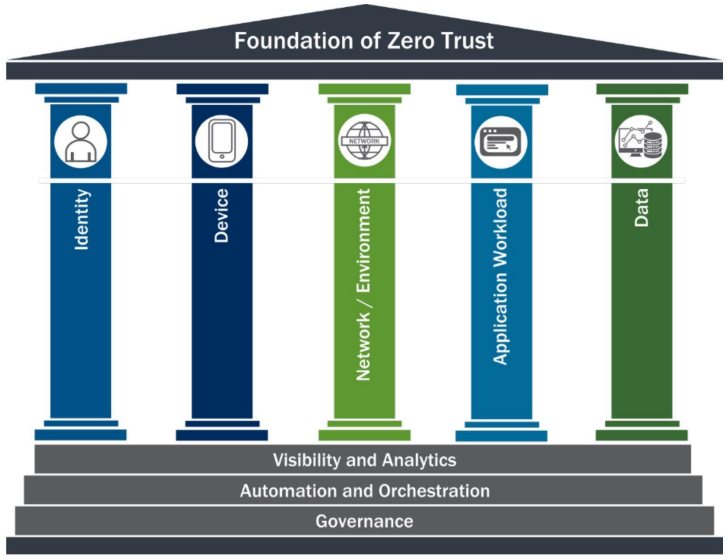


Illustration by U.S. Cybersecurity and Infrastructure Security Agency (CISA)

What does Zero Trust really mean?

1. connect a user

   IDENTITY

2. from their device

   DEVICE

3. over any network

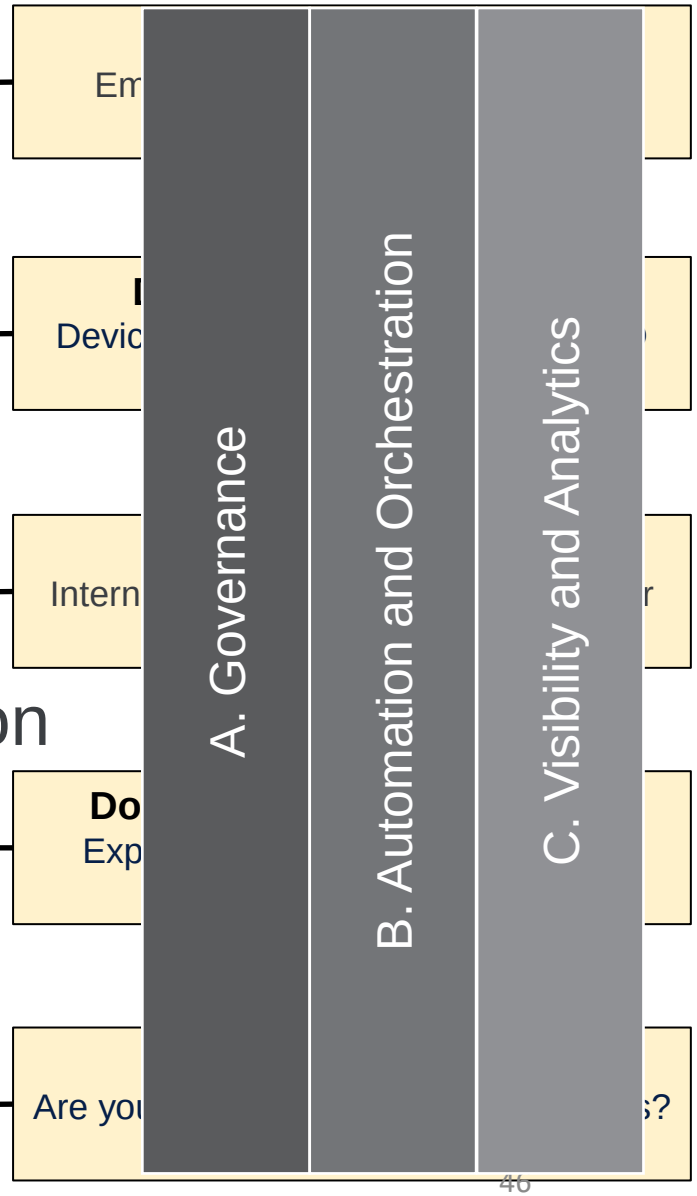
   NETWORK

4. to a specific application

   APPLICATION

5. to exchange data

   DATA



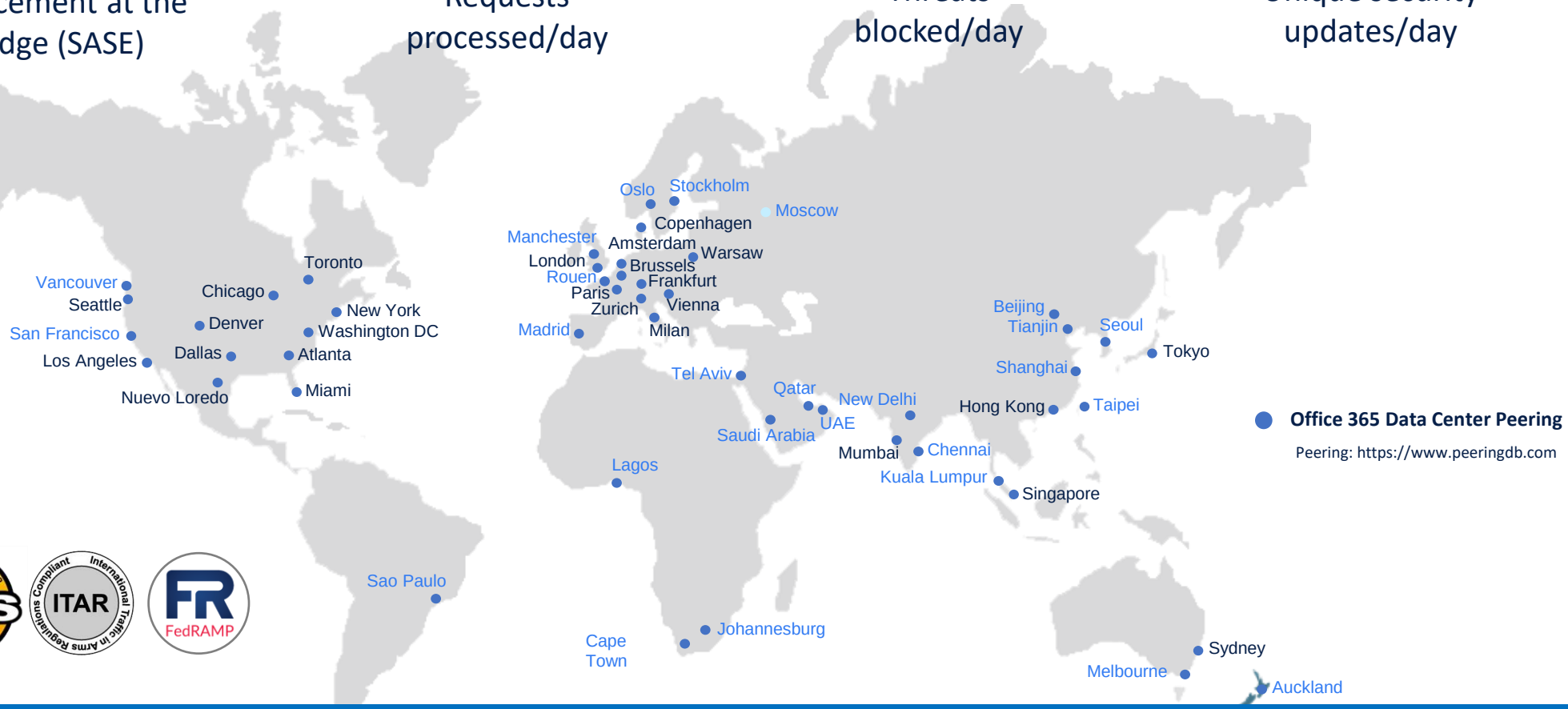
Zscaler: Security at the Edge – for over 10 years

150 Data Centers
Policy enforcement at the service edge (SASE)

250B+
Requests processed/day

100M+
Threats blocked/day

175K+
Unique security updates/day



Peering in internet exchanges

Office 365 Data Center Peering
Peering: https://www.peeringdb.com

Delivering exceptional User Experience + Reliability, Availability, and Scalability

Q&A

Jeff Sharpe
Senior Director
5G/IoT Edge AI Solutions

5G • IoT • Edge • Cloud
AI/ML/DL • Enterprise
https://linkedin.com/in/jeffbsharpe

Jay Vincent
Senior Solution
Architect

Edge • Cloud • Cybersecurity
• Enterprise • HPC
https://linkedin.com/in/jay-vincent

Dr. Ken Urquhart
Global Vice-President
5G Strategy

5G • Cybersecurity • AI • Engineering
• Quantum Computing • Cats
https://linkedin.com/in/kenu

Delivering the Security Service Edge: High-Performance and Zero Trust

April 5, 2022