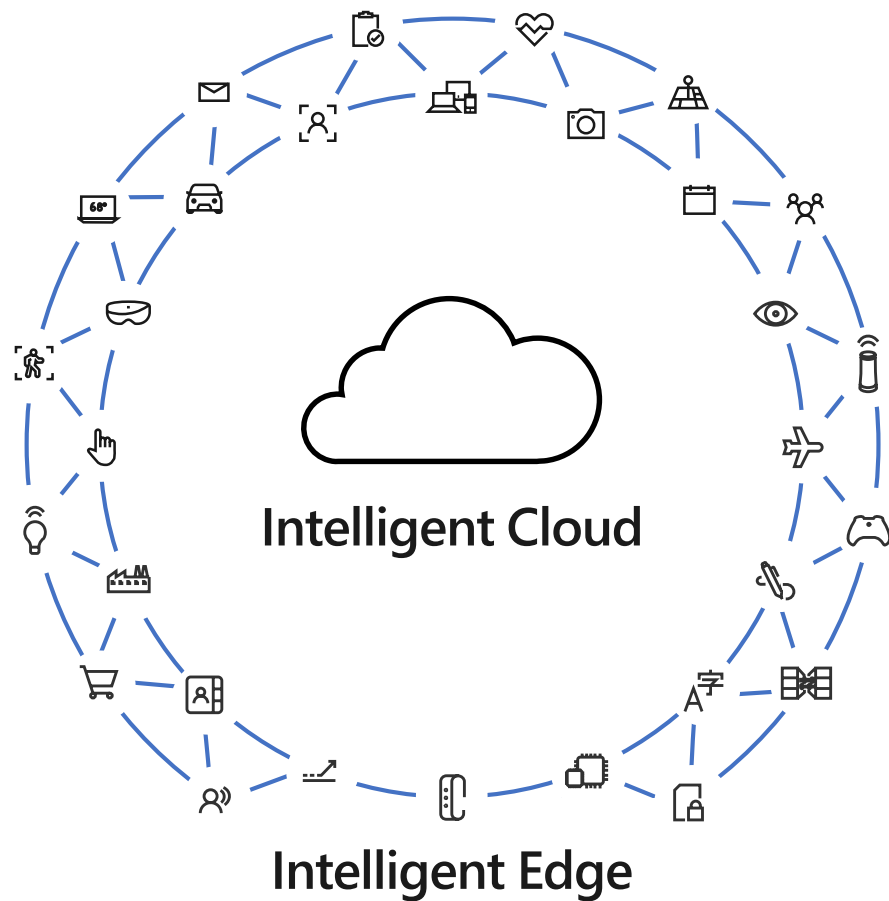


# Update on Microsoft Edge Products

Yongguang Zhang  
Microsoft Research

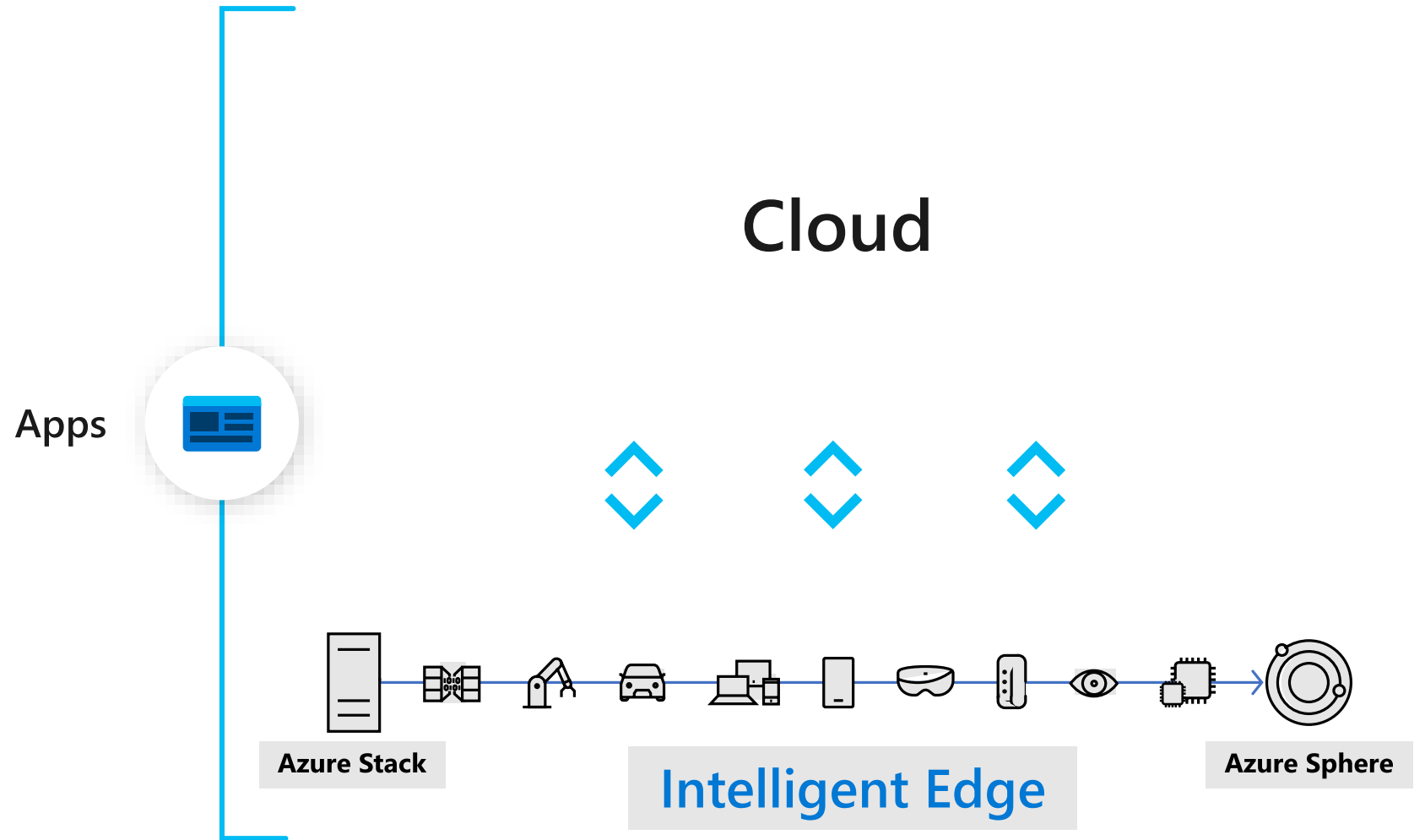


"Azure is being built as the world's computer"

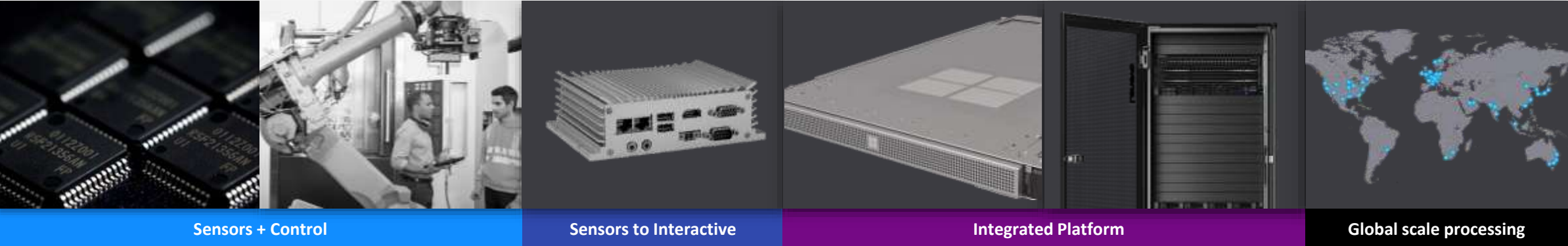
—Satya Nadella, CEO Microsoft



Developer Platform  
DevOps  
Data Platform  
Identity  
Security & Management



# Azure Intelligent Edge + Cloud Taxonomy



## Microcontroller

Azure Sphere

- Integrated Circuit designed to govern a specific operation in an embedded system
- Highly-secured, connected MCU
- Azure Sphere Linux OS for modern MCUs
- Included Azure IoT Device SDK

## IoT Devices

Azure IoT Device SDK

- Endpoint devices such as appliances, vehicles, or factory machines that connect, interact and exchange data
- 1000+ devices
- 250+ partners
- All certified to work great with Azure IoT Hub

## Edge Devices

Azure IoT Edge

- Devices that aggregate, process & provide gateway capabilities for IoT endpoints
- Deploy and manage Azure Services in containers on any IoT device
- AI, AzureML, Azure Stream Analytics and more

## Edge Appliances

Azure Stack Edge

- Integrated appliances that provide a subset of cloud edge roles, such as ML-inferencing
- Stack Edge: AI-Enabled, Storage and compute Azure Edge appliance
- Data Box: Offline, ruggedized data transport, 100 TB – 1 PB

## Edge Stack

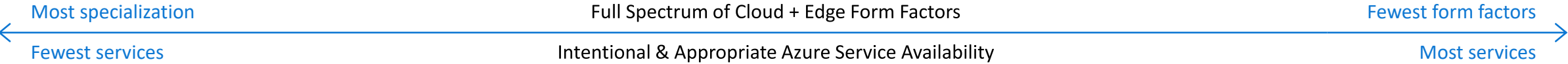
Azure Stack Hub

- Scalable solutions that provide a full cloud stack, including IaaS and PaaS capabilities
- Edge and Disconnected Scenarios
- Regulatory Requirements
- Cloud app model on-premises

## Hyperscale Cloud

Edge Regions

- First-party cloud regions
- Full Range Hyperscale Cloud Services
- Tiered Service availability: Heroes > Hubs > Satellites
- Open Source Based Services & Tools





Business and Mission Critical



Dynamic and Scalable

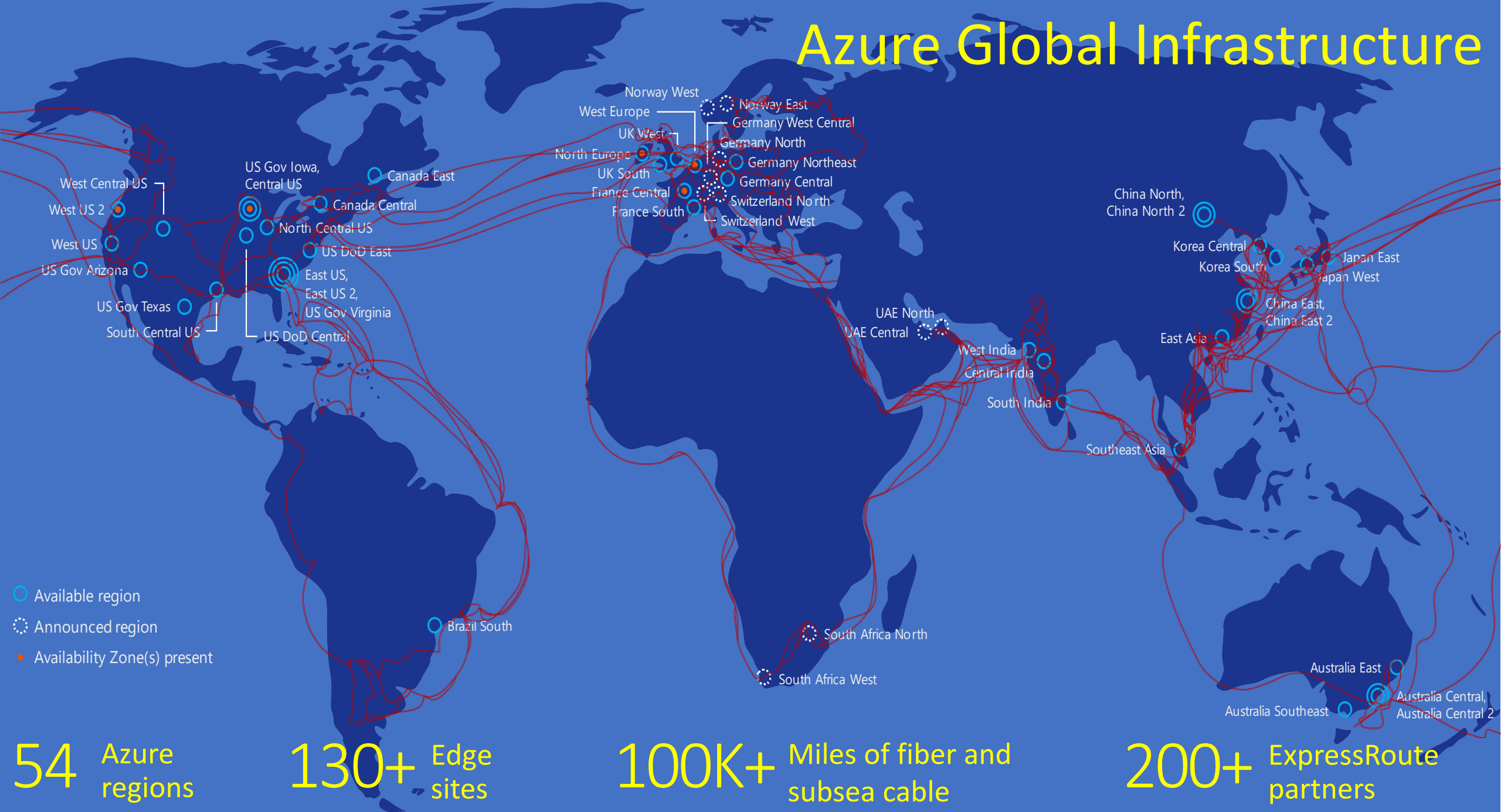


Modern Apps

# Infrastructure designed for every workload

Management | Security + Identity | Networking

# Azure Global Infrastructure



54 Azure regions

130+ Edge sites

100K+ Miles of fiber and subsea cable

200+ ExpressRoute partners

# Growth

Eight new Azure regions launched in 2019

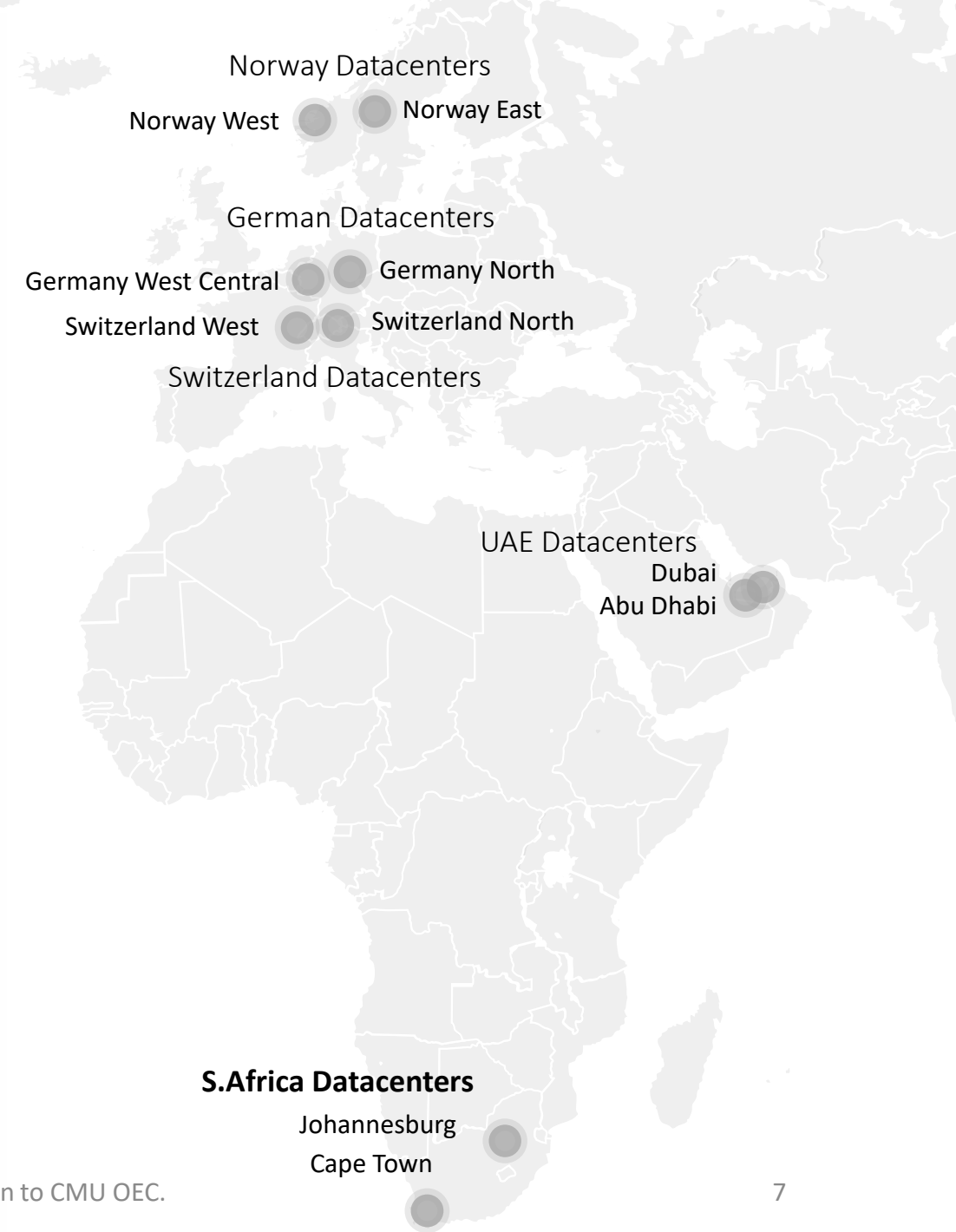
First regions from a global cloud provider on the African continent and in UAE

Expanding at a rapid rate

Actively working on regions in Norway with more to come

Azure Availability Zones

Available in Asia, Europe, and the United States with a total of 30 Availability Zones in 10 regions across the globe



# Azure physical infrastructure

## Geography

- Discrete market with two or more regions
- Meets data residency and compliance requirements
- Fault-tolerant to protect from complete region failure

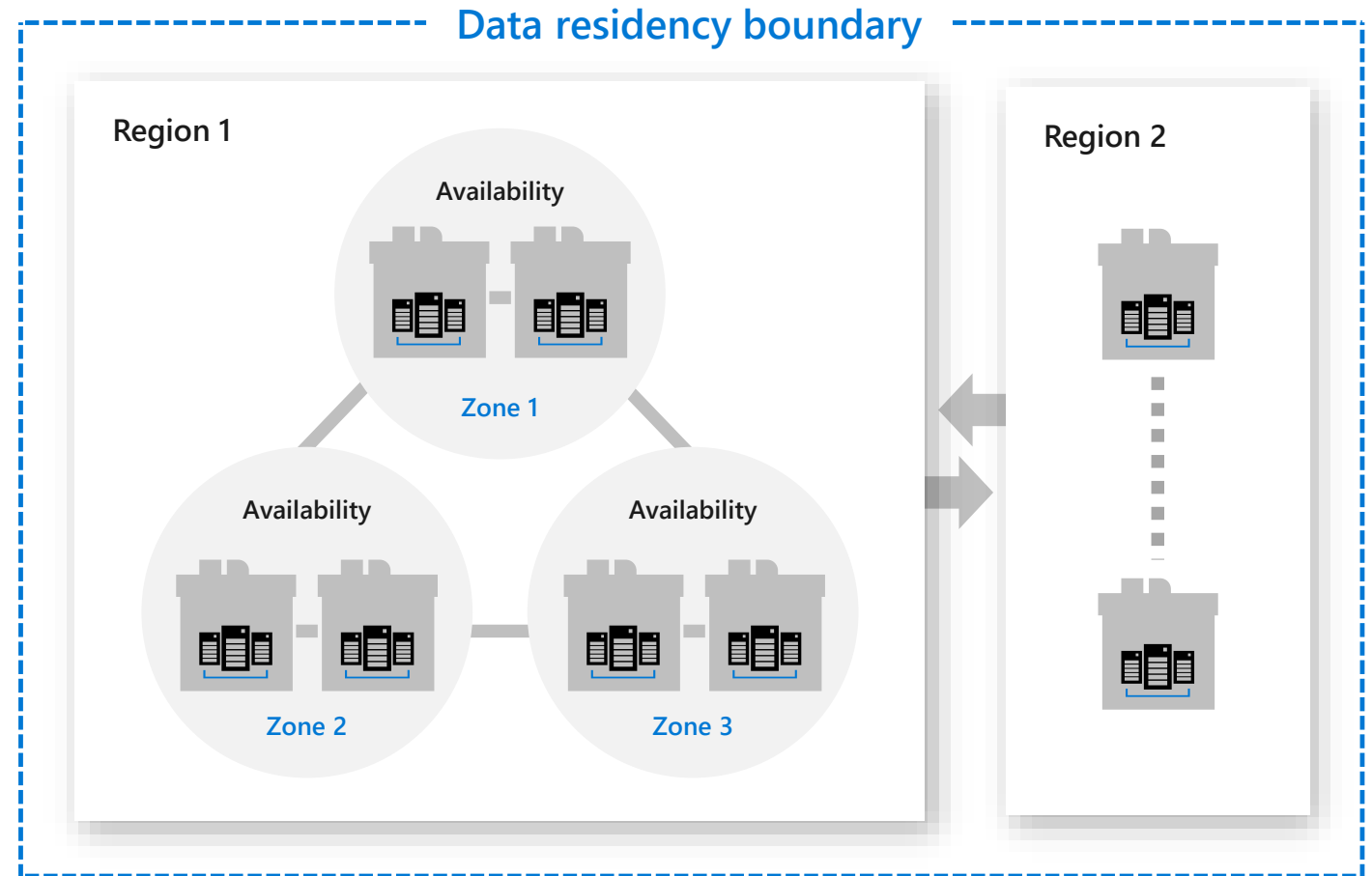
## Region

- Set of datacenters within a metropolitan area
- Network latency perimeter <2ms

## Availability Zones

- Unique physical locations within an Azure region
- Each zone is made up of one or more DCs
- Independent power, cooling and networking
- Inter-AZ network latency <2ms
- Fault-tolerant to protect from datacenter failure

## Geography



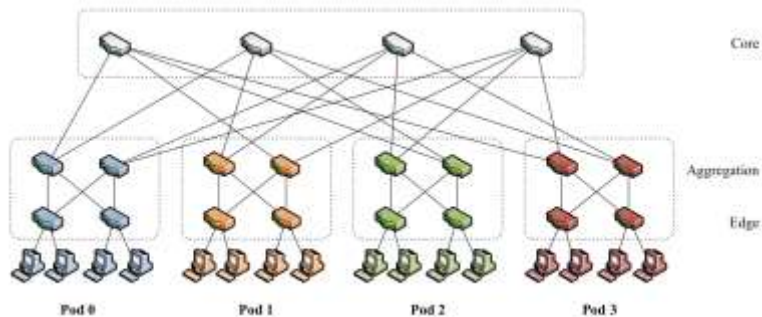


# How modern DC designs address failures

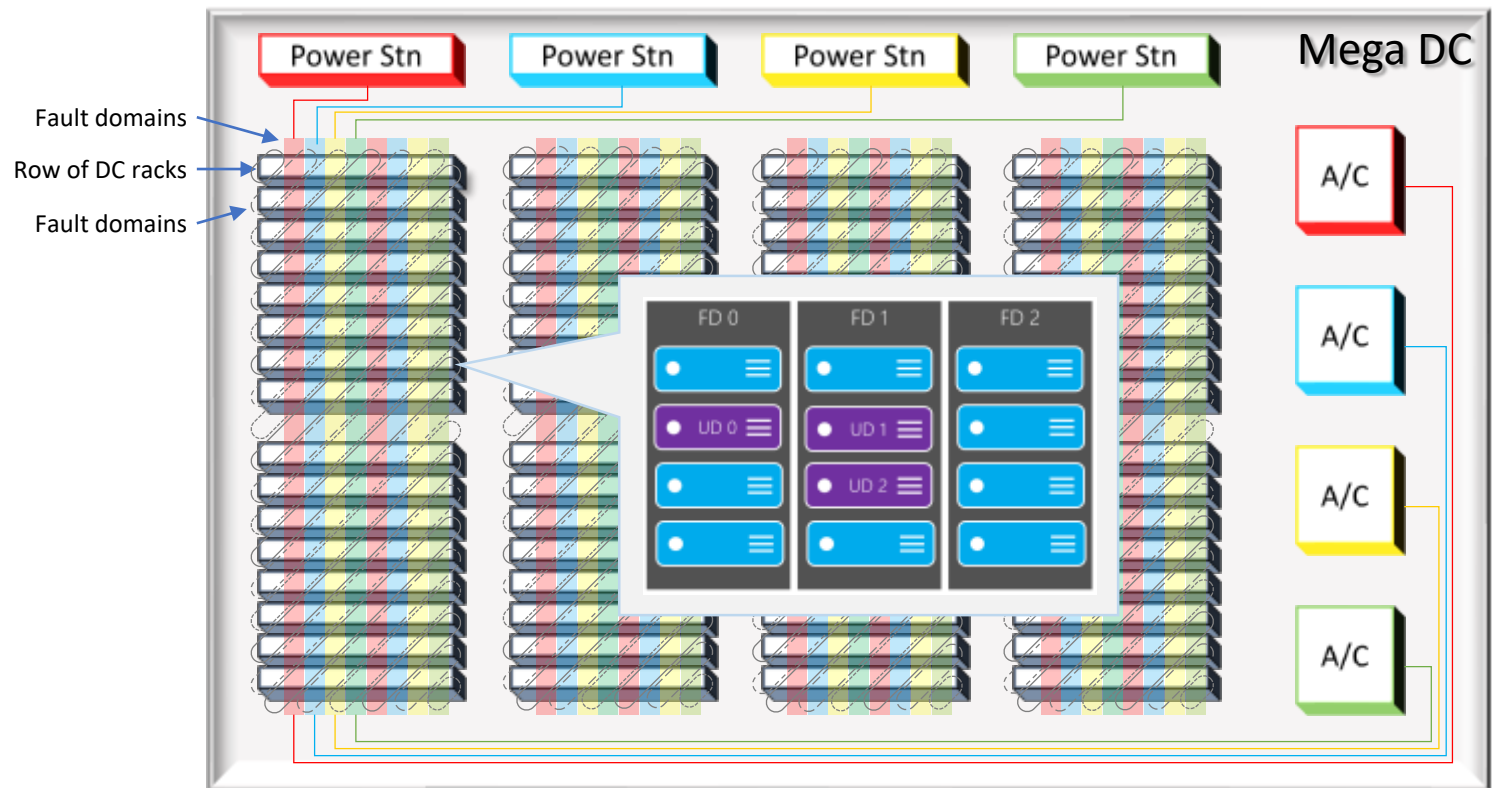
Redundant server components:  
Dual power-supply, dual NIC port



Fat-tree DCN connectivity architecture  
Multi-path routing



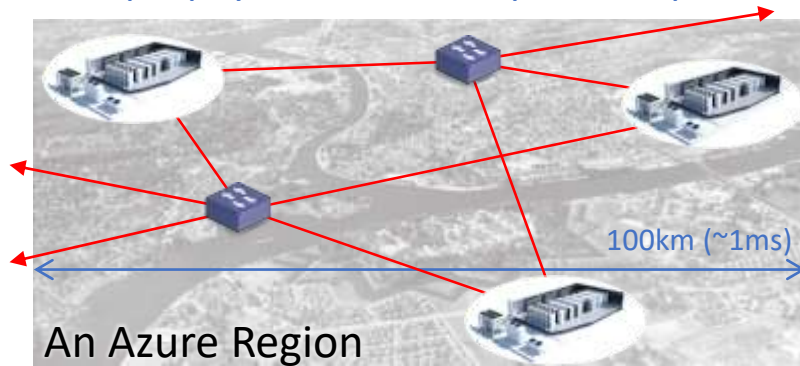
Multifaceted fault isolation:  
Multiple fault domains, update domains



# How modern cloud networks address failures

## Geo-separation

Multiple physical DCs, multiple WAN paths



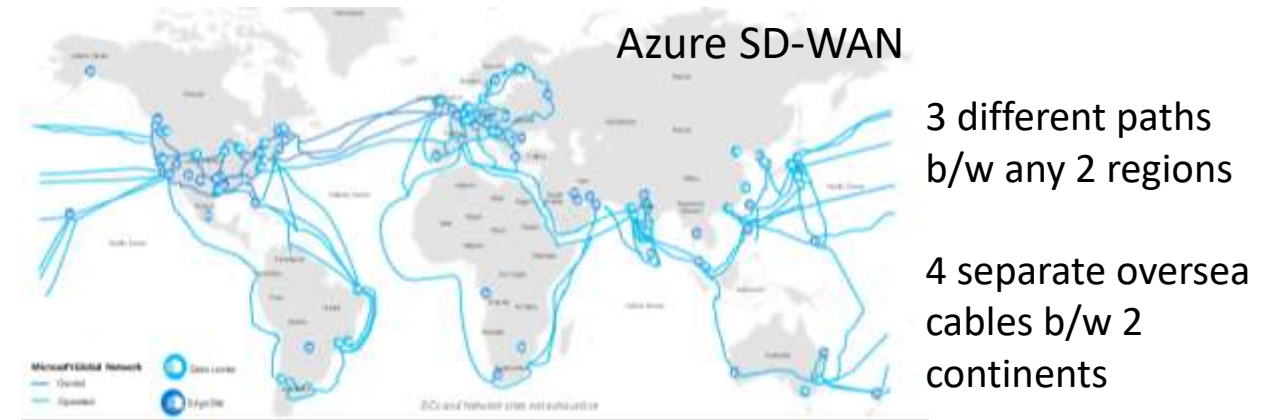
## Active monitoring

Both physical networks and virtual networks



## Redundant global connectivity

k-edge-connected ( $k \geq 3$ )



## Fault prediction

Big-data driven, machine learning technique

Azure DC Availability: >99.999%

# Azure Architecture

Azure Portal

CLI & PowerShell

SDKs

## Azure Resource Manager (ARM)

RBAC

Activity Logs  
& Telemetry

### ARM Configurations

Resource Metadata

Resource Groups

Subscription

Management Groups

Tags

### ARM Resource Providers

Bot  
Framework

IoT  
Hub

App  
Services

Key  
Vaults

Other  
RPs

Service  
Fabric

AKS

SQL

Azure  
Functions

Event  
Grid

Compute RPs

Networking RPs

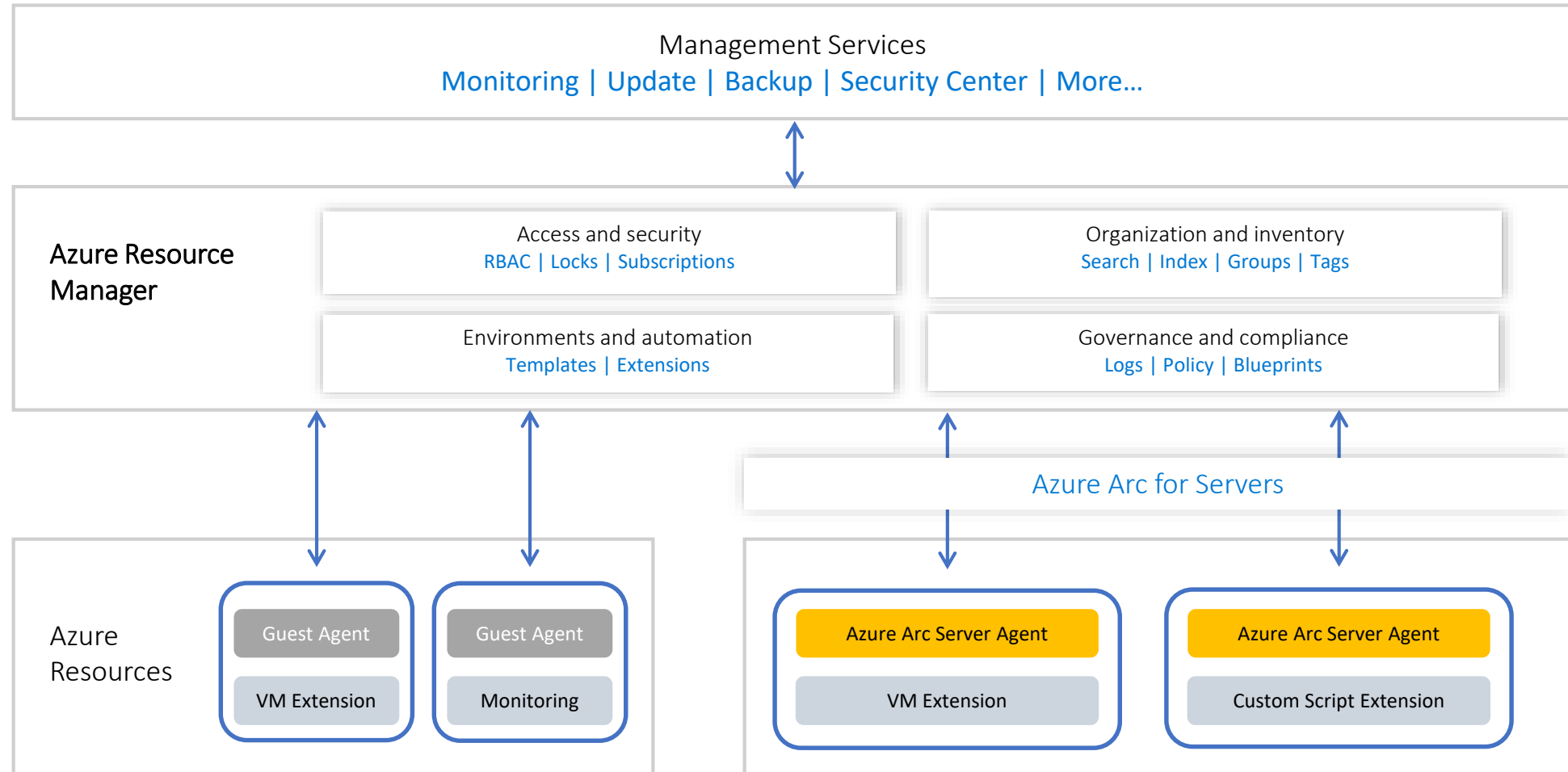
Storage RPs

Azure Fabric Controller

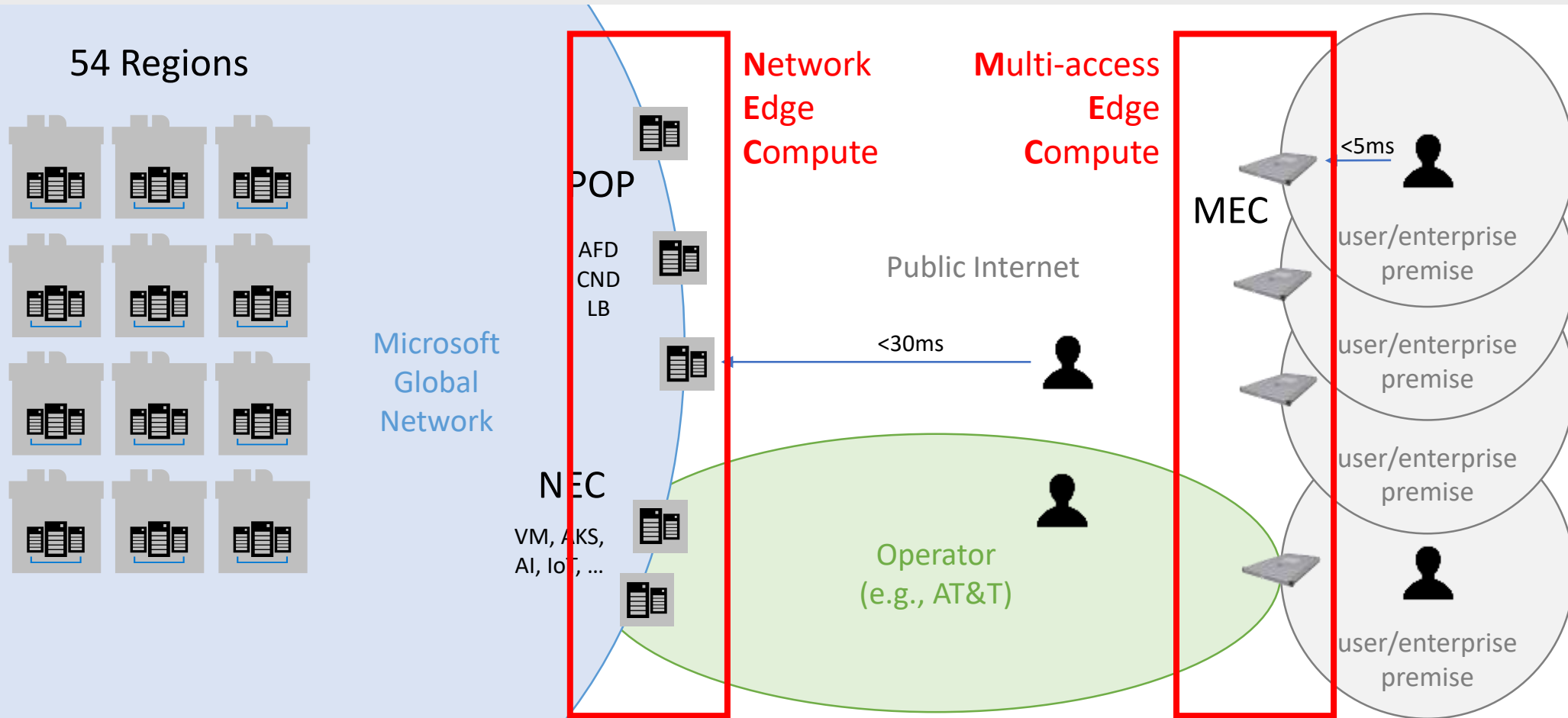
Hardware Manager

Azure Infrastructure

# Server management – Azure Arc



# Microsoft's Edge Product Vision



# Microsoft Azure Edge Compute Offerings



## Network Edge Compute (NEC)

- Azure closer to the end user
- Telcos' DCs & Microsoft edge sites
- One hop away from Telcos' 5G core
- Optimized for multitenant deployments
- Larger scale workloads



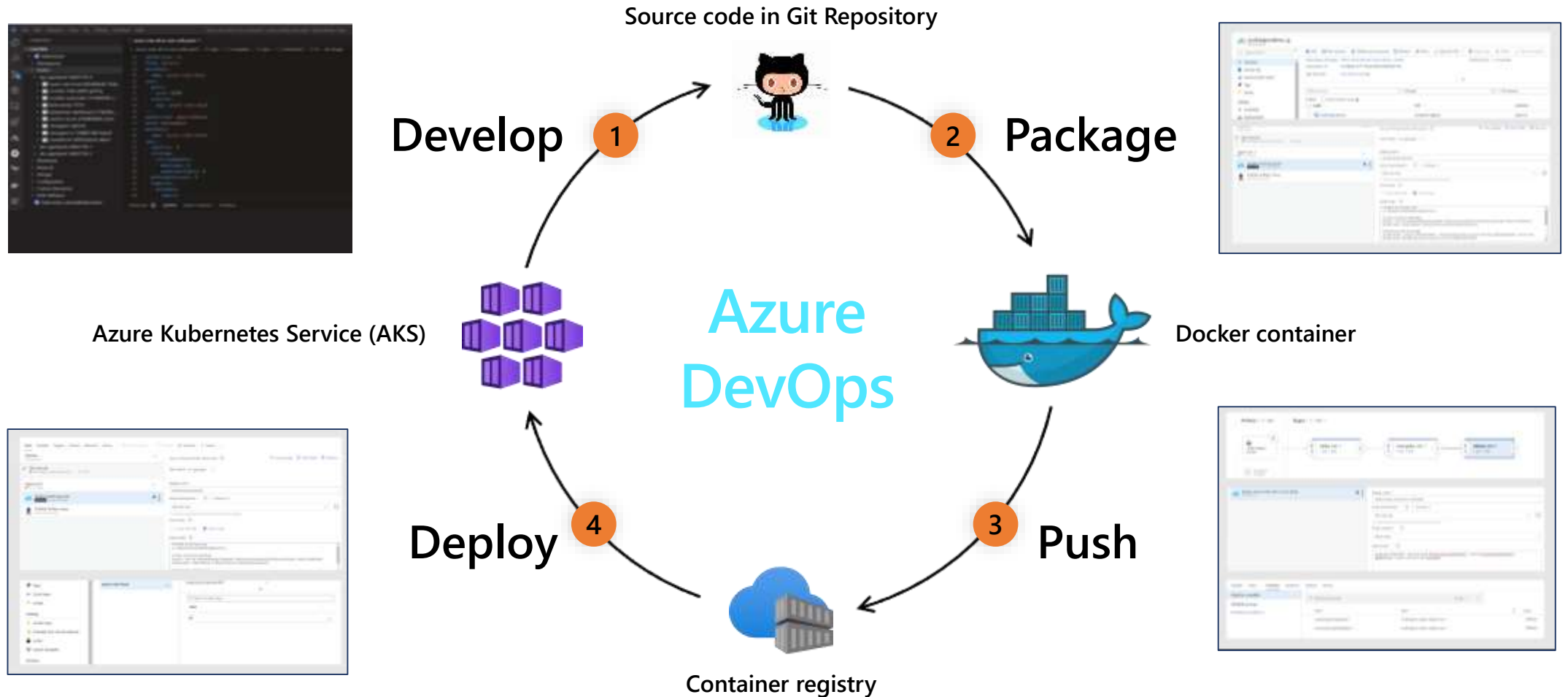
## Multi-access Edge Compute (MEC)

- Based on Azure Stack Edge
- Customers on-premises
- Runs managed VNFs for private mobile networks
- Optimized for single tenant deployments
- Scales from 1U to racks, other formfactors

### Same Azure experience closer to the user

- Managed by Azure
- Same dev-ops and life-cycle tools

# Same DevOps tools and pipeline



# Azure Stack Portfolio

Consistently build and run hybrid apps across on-premises, cloud, and edge



## Azure Stack Edge

Cloud-managed appliance

Machine learning at the edge  
Edge compute and IoT solutions  
Network data transfer to cloud



## Azure Stack HCI

Hyperconverged solution

Scalable virtualization and storage  
Remote branch office  
High-performance workloads



## Azure Stack Hub

Cloud-native integrated system

Disconnected scenarios  
Data sovereignty  
Application modernization



# Azure Stack Edge – Announcements



Virtual machines



K8S



High availability



GPU support

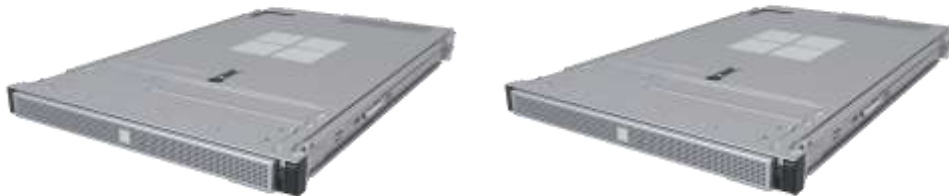


Ruggedization



New form factors

## Commercial (C) series



Enterprise-ready form factors  
With Intel FPGA or NVIDIA GPU

## Rugged (R) series



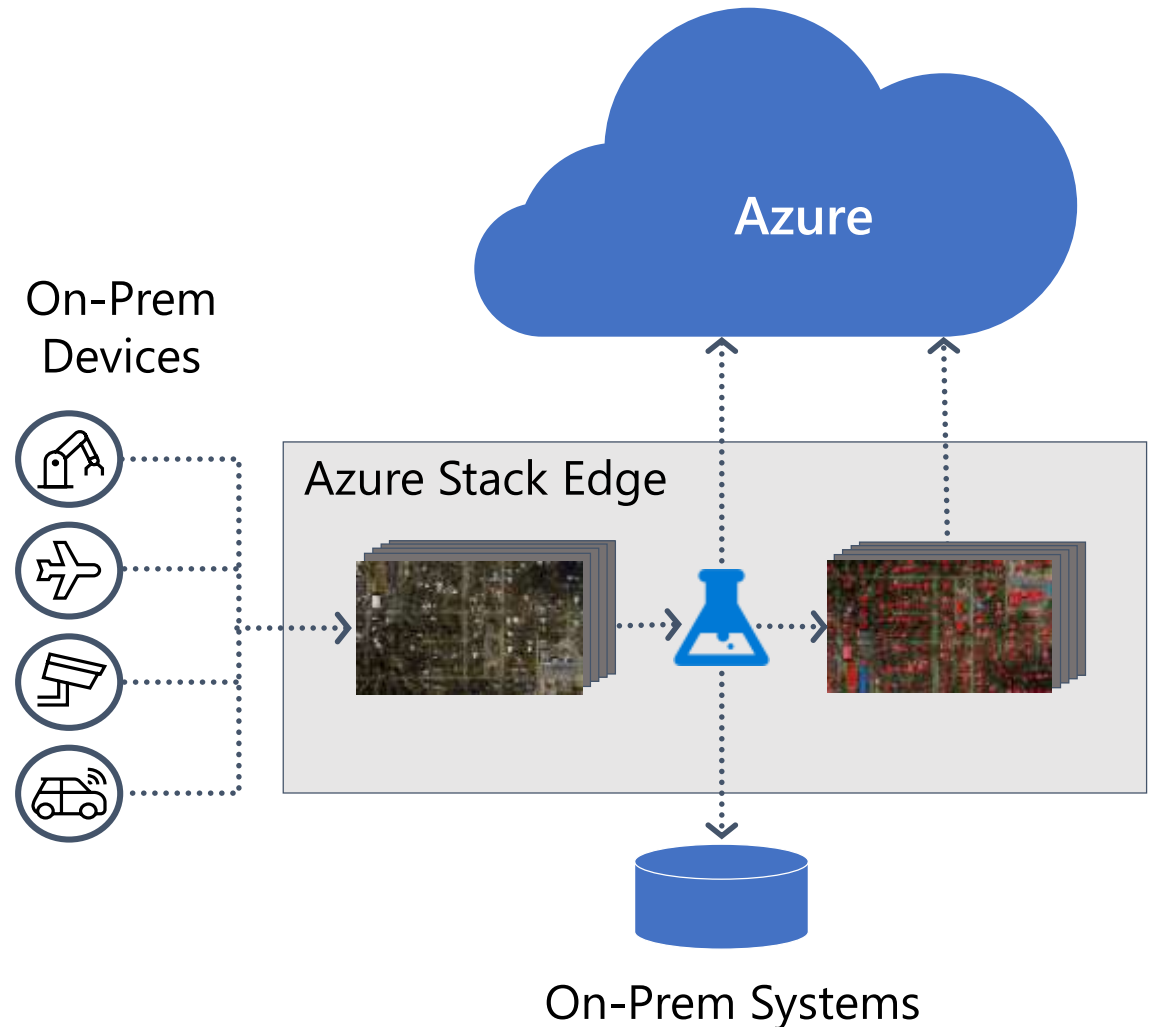
Ruggedized, portable, battery-operated  
form-factors for harsh field conditions

# Hardware Accelerated Machine Learning (ML)

Process data locally, where it is generated, for immediate results.

Use the results to take action in the cloud or on local systems. Transfer or keep only the processed data.

From the cloud, get visibility across operations on all your locations.



# Hardware Acceleration Options



## FGPA with Azure Machine Learning

- Use Azure Machine Learning's supported models and train with your data via transfer learning in the cloud.
- Automatically accelerated on the FPGA.



**NVIDIA**

## Nvidia T4 GPU

- Supports full GPU ecosystem. Use Azure ML, ONNX, Nvidia EGX and Deepstream, tensorflow, and more.
- Preview coming soon



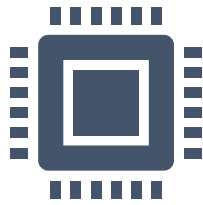
# Kubernetes on Azure Stack Edge

End-to-end first-party, cloud-managed, at-scale support for Kubernetes at the edge



## Create a cloud-managed Kubernetes cluster of Azure Stack Edge appliances in a few clicks

In under an hour, go from plugging in your Azure Stack Edge appliances to running applications in your Azure Stack Edge Kubernetes cluster.



## Scale workloads for more powerful edge solutions

**Scale compute:** Leverage hardware acceleration (FPGA or GPU) across the cluster

**Scale storage:** Persistent storage volumes across the cluster

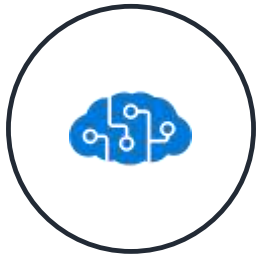
**Improve resiliency:** Build for high availability in a cluster of two or more appliances



## Deploy and manage applications via cloud or edge

- Azure Arc for Kubernetes
- Azure IoT Edge
- Native Kubernetes tools (kubectl) over your local network

..... Many Azure services offer containerized edge versions .....



Azure Cognitive  
Services



Azure Machine  
Learning



Azure Stream  
Analytics



Azure  
Functions



AI  
Toolkit



SQL Database  
Edge

Preview coming soon!

# Virtual Machines

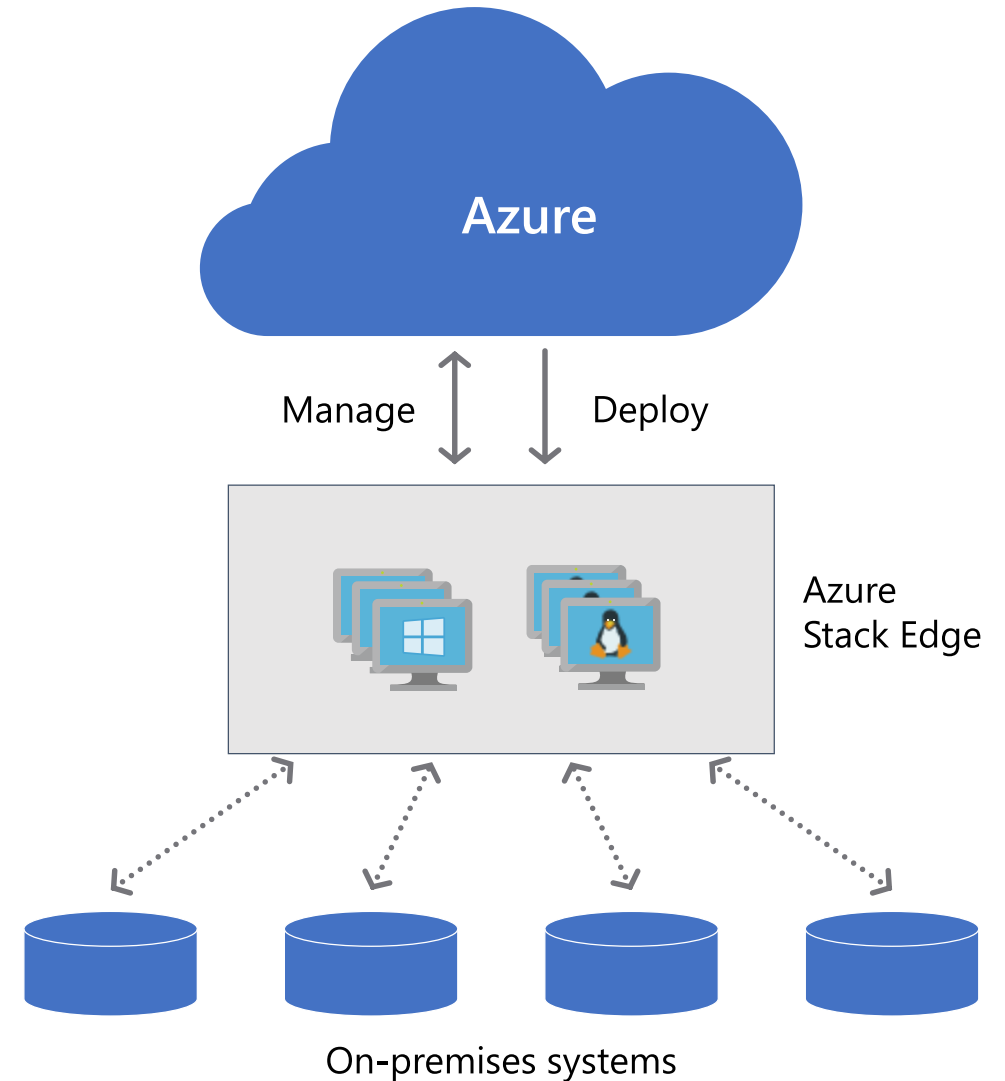
Run VMs at your location, deployed and managed from Azure

Great option for compute tasks that aren't in containers yet

Both Linux and Windows

Designed for basic configurations

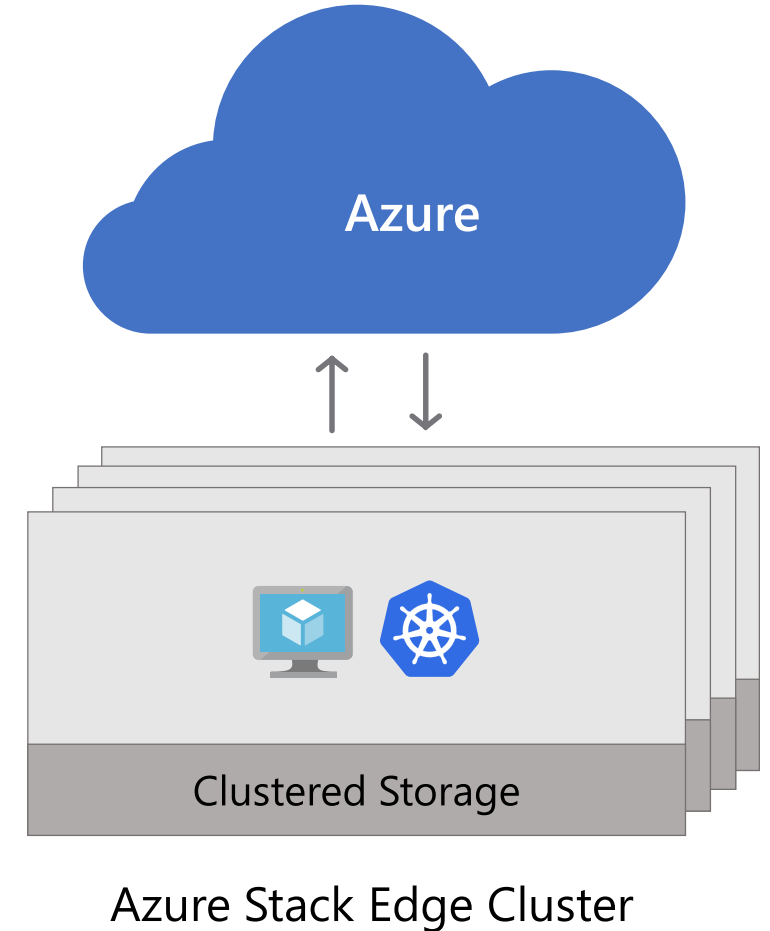
More details coming soon



# Cluster Azure Stack Edge devices

Clustering helps your edge infrastructure stay reliable or scale up to workloads that require more than one appliance.

Clusters are easy to create and managed from the Azure cloud.

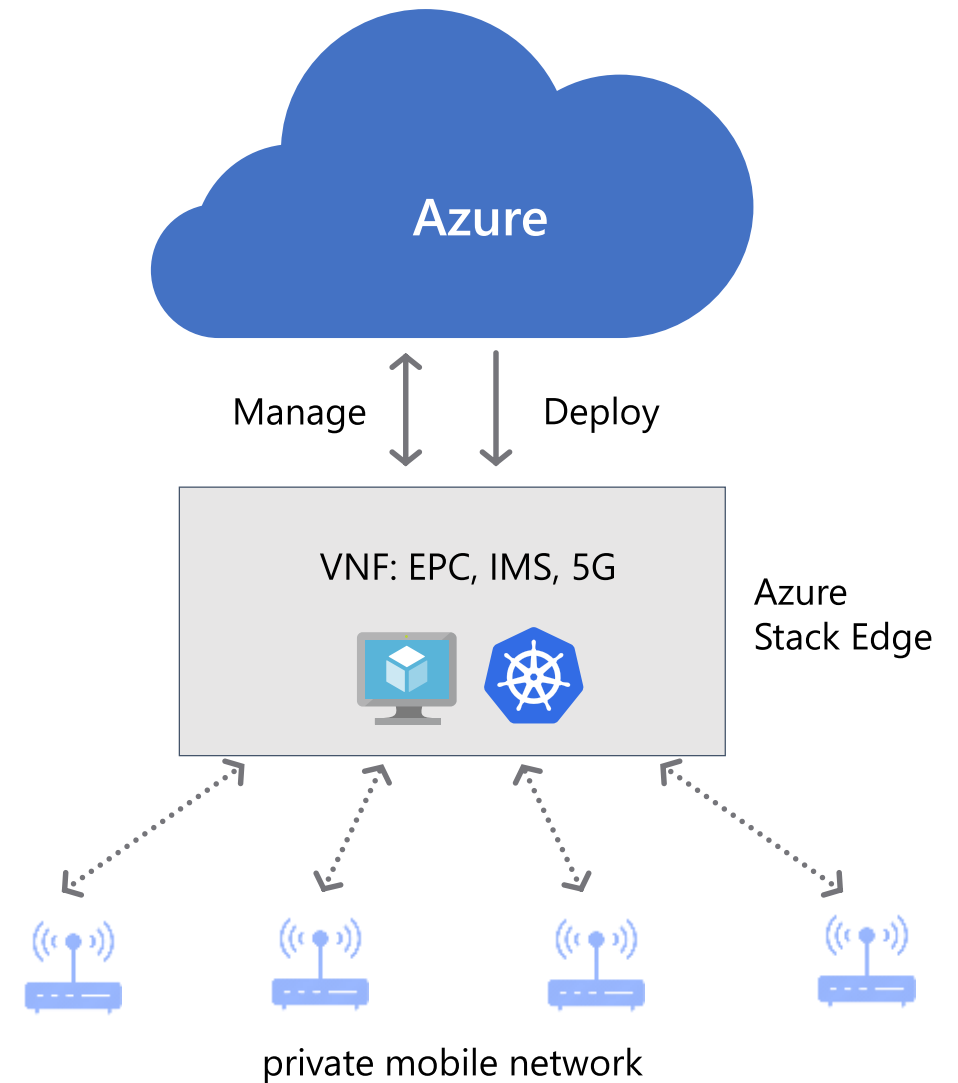


# Private Mobile Networks

Managed Compute and Network as a Service.

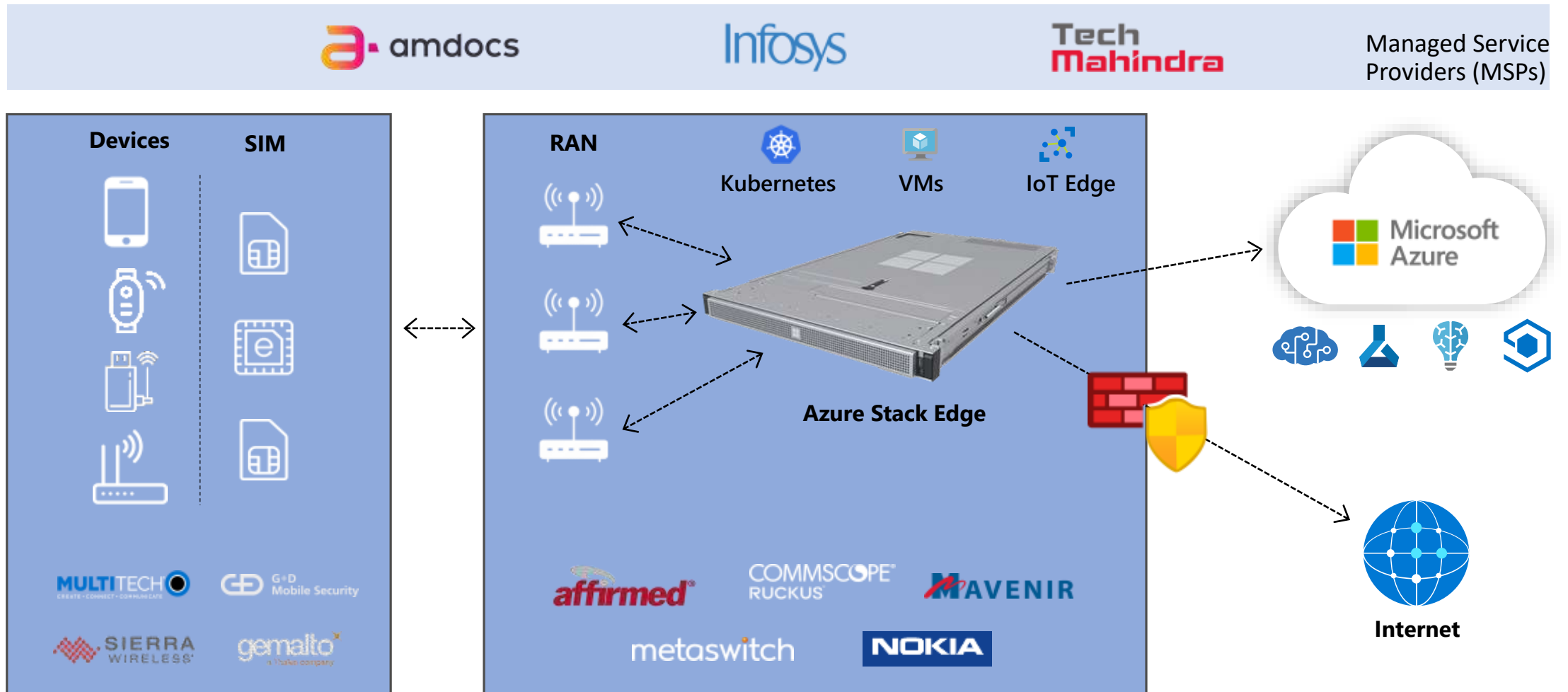
Integrated with other Azure Services to enable new use cases.

Supports 4G/5G VNFs. Supports private LTE and private 5G.





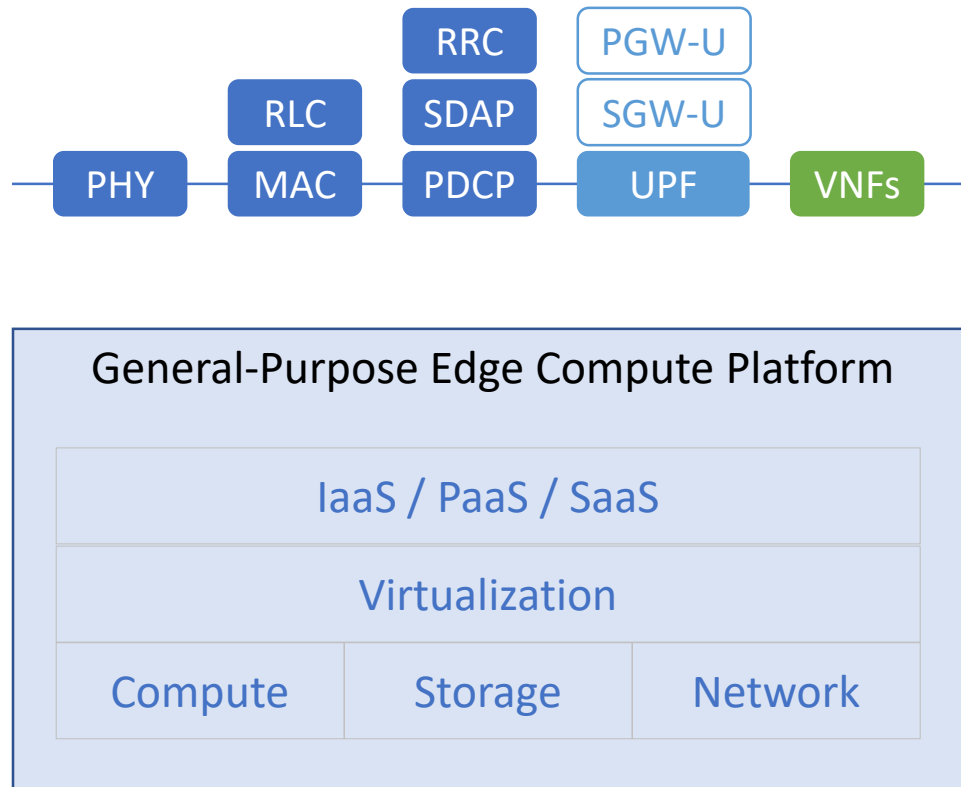
# Building blocks and partners



# Edge Infrastructure Research at Microsoft

- 5G RAN is the killer app for edge
  - At least 1B DAU
- Rearchitect telecom stack to be true cloud-native
  - Scale to any cloud and any edge
- Edge compute API and programming model
  - It is about partition of computation/data between cloud, edge, and device

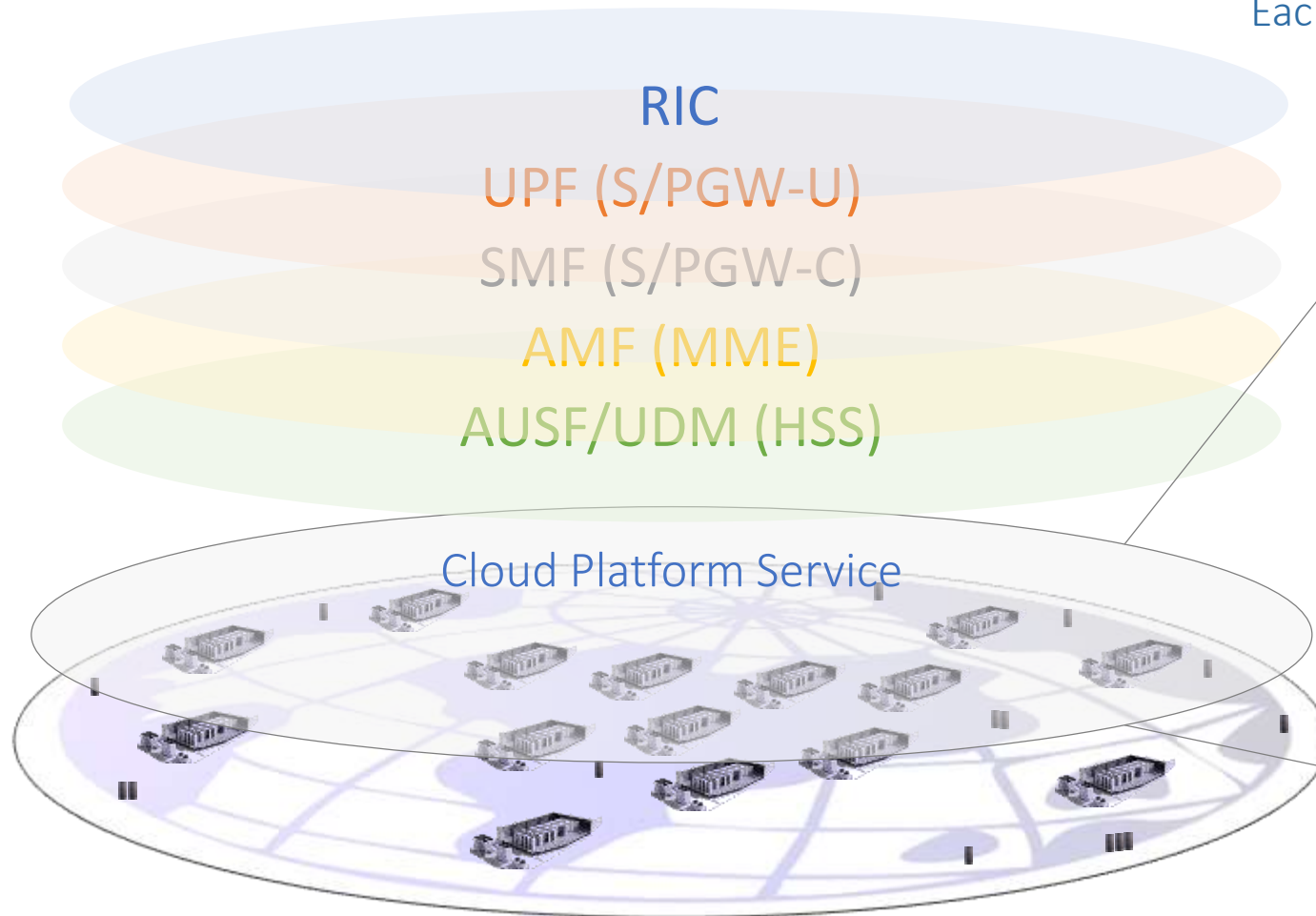
# 5G RAN on the Edge



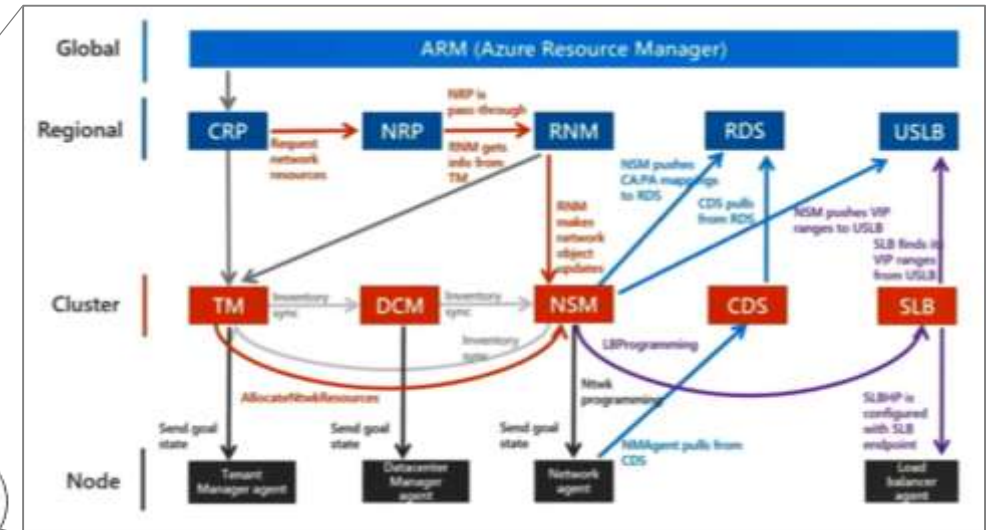
- Architect RAN to run well on general edge platform
  - Cloud-native, edge-native
- Edge platform improvement
  - Virtualization for real-time software
  - CPU core pinning, etc.
  - Fast and reliable state management
- Data-plane acceleration
  - Edge platform options: GPU, TPU, FPGA, etc.

# VNF as a planetary-scale cloud service

Each VNF is built as a planetary-scale cloud service/micro-service



Example: Azure platform service



Automatic resource allocation  
Automatic placement for high availability  
Automatic scale up/down

# Lack of Edge Application Models

- Microsoft first-party apps
  - xcloud (game streaming)
  - HoloLens (AR)
- Azure services at the edge
  - Many Azure services offer containerized edge versions
  - But we don't have any tools to help develop edge-enabling applications

# Thank you!