Navigating the Edge

How Cloud-Edge Devices Operate

Alok Shankar

Tuesday, February 13, 2024

About me

Engineering Manager at Oracle Health AI

Worked on building Oracle's Roving Edge Infrastructure from ground up

Worked on building Oracle Cloud Migrations from ground up

Background in Networking, Distributed Systems, Cloud Infrastructure

Past: Sumo Logic, Big Switch/Arista, Microsoft, Cisco

Education: Carnegie Mellon, Cochin University

Occasional Hindi poet and lyricist

Background on Edge Computing

CDN – Introduced by Akamai in 1990s. Uses node at the edge to fetch and serve content

Cloudlet and edge computing itself as a concept was pioneered by **Mahadev Satyanarayanan** at Carnegie Mellon

Small-scale cloud datacenters located at the edge of the internet. Cache data from cloud

Foundational elements in the edge computing architecture

Data can be processed locally at the edge, rather than being sent back and forth to cloud servers.

The architecture has also been called fog computing

3

References: The Emergence of Edge Computing, Mahadev Satyanarayanan

More on Cloudlets

- Proximity matters for IOT/Mobile computing
- Better latency/bandwidth, trust, response time
- Valuable for VR/AR use cases
- Reduce ingress bandwidth to the cloud
- Better privacy – IOT hubs like hubitat conceptually operate like this
- Masking outages
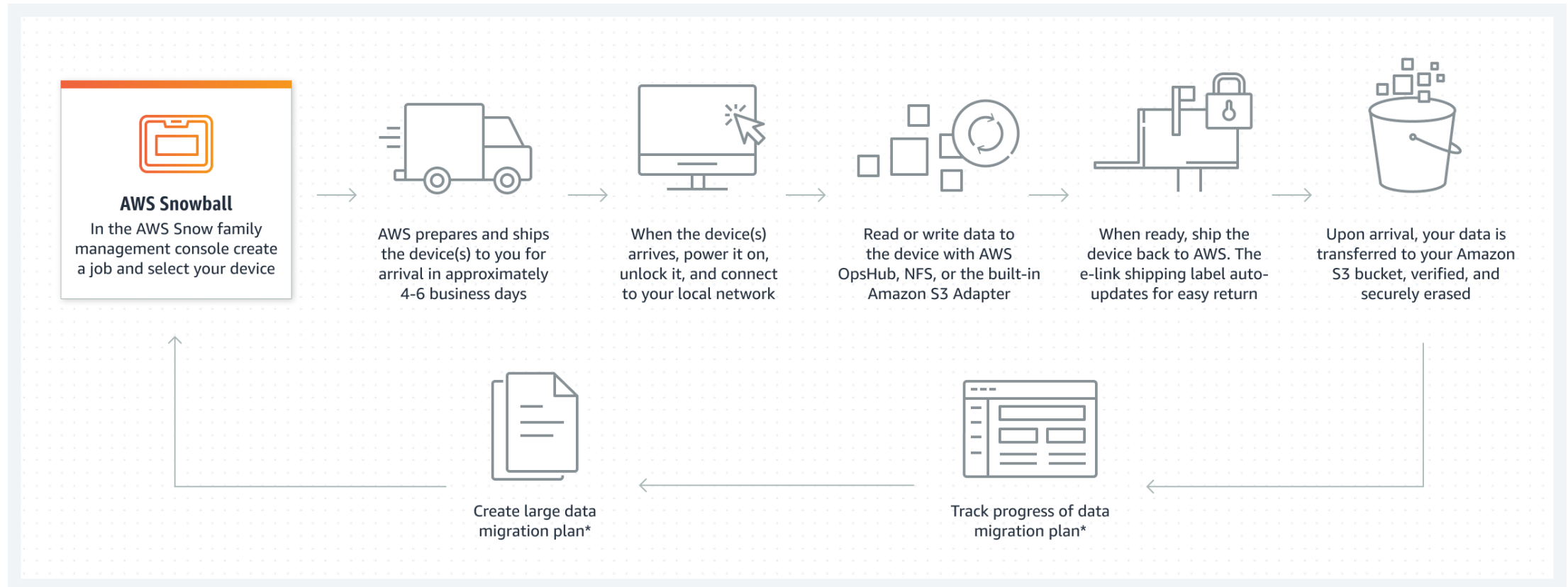- I was involved in a research project using cloudlets in early days

The Cloud's Data Transfer problem

- Moving data over public internet is not always feasible
- Network Costs
- Network connectivity
- Long transfer times
- Security Concerns
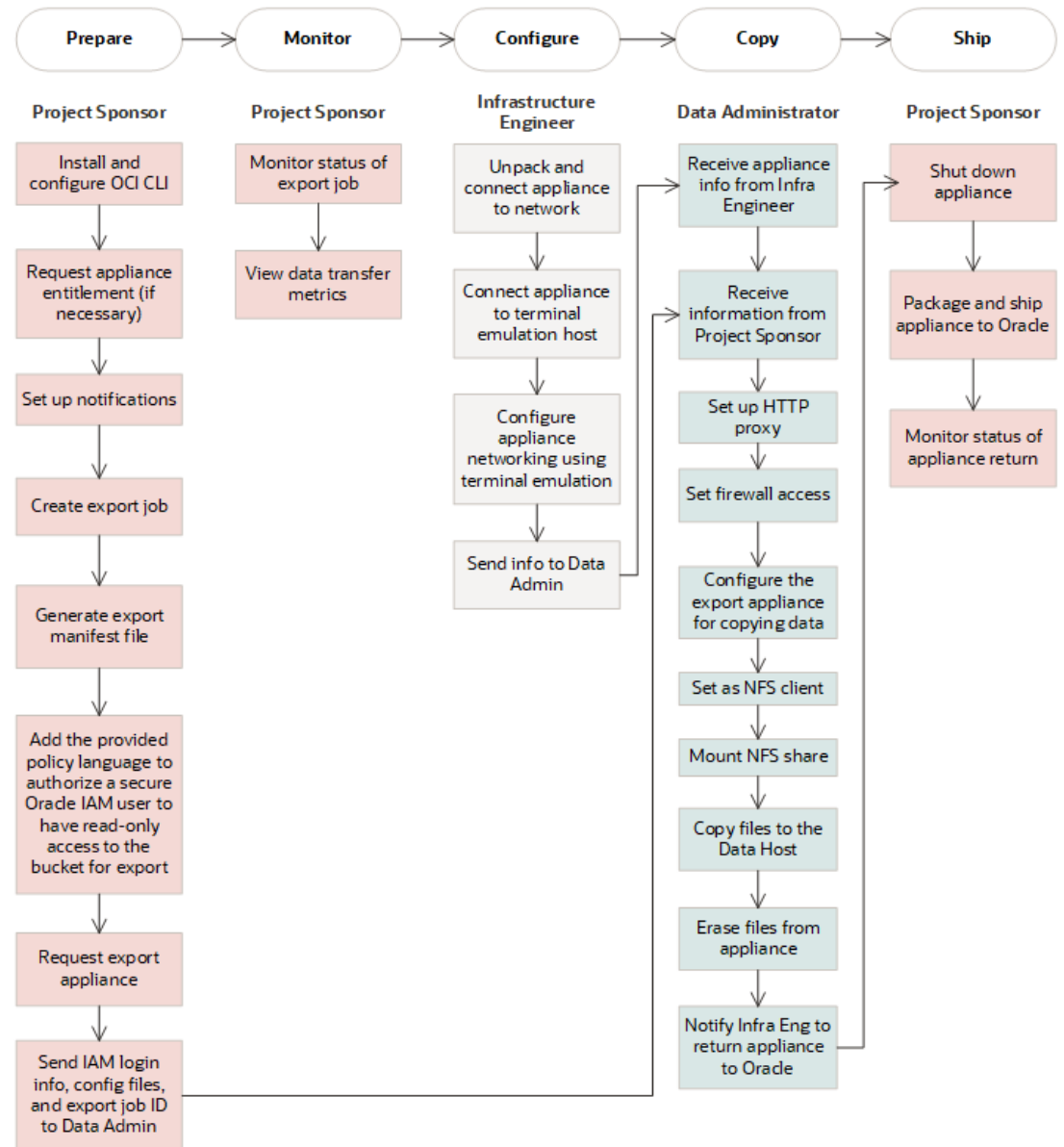- Storage gateway services provide data transfer and application extension

Offline Data Transfer

- Physical devices used to transfer data in and out of cloud
- Typical options have disk, tape and an appliance
- Device is physically shipped to customers
- Users can turn the device on or plug it in
- Return ship it when done
- Faster than online transfer for large data

Offline Transfer Workflow – Snowball Appliance

AWS Snowball
In the AWS Snow family management console create a job and select your device

AWS prepares and ships the device(s) to you for arrival in approximately 4-6 business days

When the device(s) arrives, power it on, unlock it, and connect to your local network

Read or write data to the device with AWS OpsHub, NFS, or the built-in Amazon S3 Adapter

When ready, ship the device back to AWS. The e-link shipping label auto-updates for easy return

Upon arrival, your data is transferred to your Amazon S3 bucket, verified, and securely erased

Create large data migration plan*

Track progress of data migration plan*

Source: AWS Documentation

OCI workflow for export

Source: OCI Documentation

Bringing Compute to the Edge

- Modern edge cloud devices are typically extensions of the data transfer devices

- A compute data and control plane is added in addition to the storage control/data plane

- A subset of cloud services are available including Identity, Storage, Compute and Networking

- May include container and function orchestration features

The Edge Cloud Device

- Compute or Storage Optimized
- Node or a Cluster
- Large or small
- Ruggedized or Not
- Can handle harsh environments
  - Comes with temperature, decibel
- Encrypted data
- Storage API or mount point
- Cloud Management Console

Major Products

- AWS Snow Family
- Azure Stack Edge
- Oracle Roving Edge Infrastructure
- Google Distributed Cloud Edge

What typical devices look like

Azure Stack Edge Pro Series
Enterprise scale and performance for your edge workloads.

Pro 2
Compact form factor optimized for edge and branch locations. Flexible mounting options.

Configuration options:

- 32 vCPUs, 51 GB RAM, 720 GB
- 32 vCPUs, 102 GB RAM, 1.6 TB, 1 NVIDIA A2 GPU
- 32 vCPUs, 204 GB RAM, 2.5 TB, 2 NVIDIA A2 GPUs

All figures are customer usable capacity.

Pro
1U rack-mountable appliance, optimized for conditions inside a data center or branch location.

Configuration options:

- 40 vCPUs, 102 GB RAM, 4.2 TB, 1 NVIDIA T4 GPU
- 40 vCPUs, 102 GB RAM, 4.2 TB, 2 NVIDIA T4 GPUs

All figures are customer usable capacity.

Pro R
Ruggedized datacenter-grade power with a built-in NVIDIA T4 GPU, in a transportable case for remote locations.

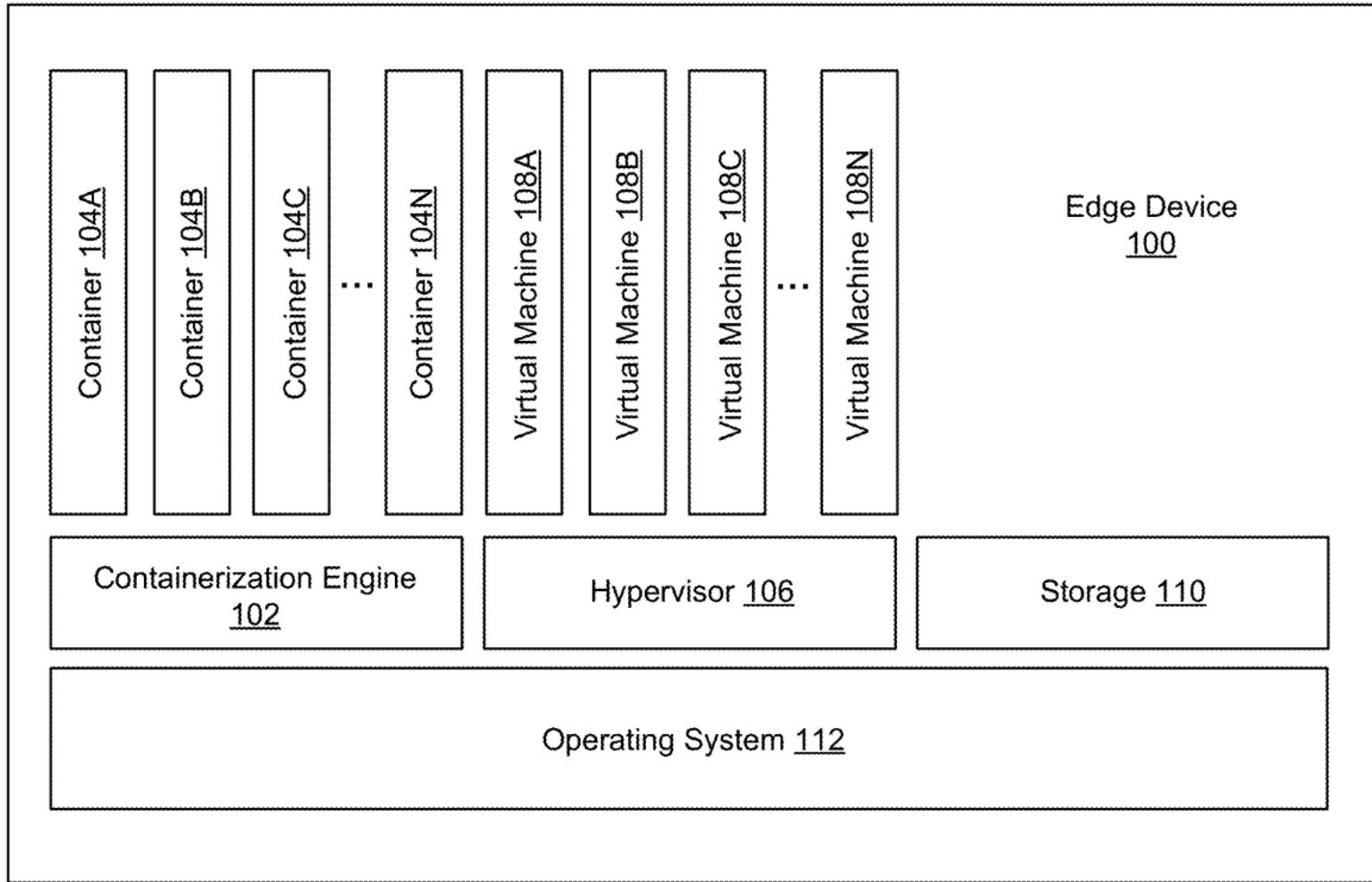Available options: With or without Uninterruptable Power Supply (UPS).

Read datasheet >

Major Products - AWS
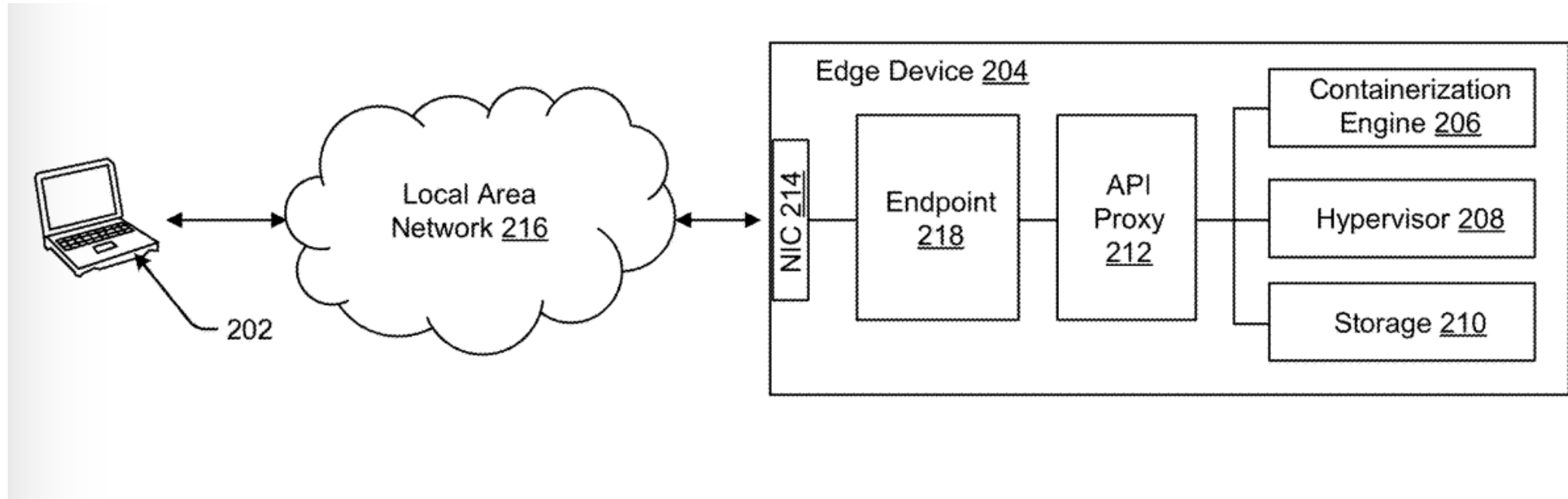
Feature comparison matrix

	AWS SNOWCONE	AWS SNOWBALL EDGE STORAGE OPTIMIZED 80 TB	AWS SNOWBALL EDGE STORAGE OPTIMIZED 210 TB	AWS SNOWBALL EDGE COMPUTE OPTIMIZED	AWS SNOWMOBILE
Usable HDD Storage	8 TB	80 TB HDD	N/A	N/A	100 PB
Usable SSD Storage	14 TB	1 TB	210 TB NVMe	28 TB	No
Usable vCPUs	2 vCPUs	40 vCPUs	104vCPUs	104 vCPUs	N/A
Usable Memory	4 GB	80 GB	416 GB	416 GB	N/A
Device Size	9in x 6in x 3in 227 mm x 148.6 mm x 82.65 mm	548 mm x 320 mm x 501 mm	548 mm x 320 mm x 501 mm	548 mm x 320 mm x 501 mm	45 ft. shipping container
Device Weight	4.5 lbs. (2.1 kg)	49.7 lbs. (22.3 kg)	49.7 lbs. (22.3 kg)	49.7 lbs. (22.3 kg)	N/A
Storage Clustering	No	No	No	Yes, 3-16 nodes	N/A
256-bit Encryption	Yes	Yes	Yes	Yes	Yes
HIPAA Compliant	No	Yes, eligible	Yes, eligible	Yes, eligible	Yes, eligible

Source: AWS Documentation

Typical Workflow – Roving Edge

- Create an edge device using cloud console
- Attach VM snapshots and Storage buckets to the device
- Order the device
- Monitor the status
- Cloud vendor provisions device
- Ships the device via common carrier
- Receive and deploy
- Contact support

High Level Architecture

Source: Cloud edge device virtualization, Alok Shankar, Zach Meyer

Source: Cloud edge device virtualization, Alok Shankar, Zach Meyer

Architecture

- Typically contains control/data plane for each major feature
- Physical network devices are mapped to virtual ones for VMs/Containers to use
- API proxies/gateways route requests to appropriate service
- A substrate, private network exists within device for services to communicate internally
- Imported data and images are stored as files and brought up as needed
- New VMs can be created on-device and exported to cloud

Ordering

- Typically done through cloud console UI
- Customer needs prior agreement with Cloud Service with specific entitlements to be able to order
- Can specify existing Cloud resources to move to device
- After submitting order, status can be tracked on cloud portal
- Order is handed over to provisioning

Provisioning

- Cloud services gather data and images specified by customer from their cloud resources
- Data moved to physical device
- Validation/Sanity checks are done
- Device shipped to customer and tracking information updated to cloud portal in order
- Replacements handled in similar way

Shipping and Tracking

- Typically sent via common carrier like FedEx/UPS
- Return labels are created by cloud providers and customer sends them using those
- Data can be brought from customer to cloud on the way back

Deployment

- Devices can operate as their own cloud region
- Support connectivity to cloud
- Typically plugged in to customer's private network
- Support typical cloud APIs
- For clusters, all devices need to typically run same services/stacks
- Need a local network connectivity. Vendors may ship networking equipment for clusters

Devops

- Device detects updates from cloud
- Download updates and install them
- Disconnected upgrades done via downloaded bundles
- Devices can be rolled back to previous versions
- Relevant cloud services follow typical CI/CD devops model
- Troubleshooting done via diagnostic bundles and sent to support
- Basic metrics for Edge device services are available

Trends: Proliferation of IOT Devices

- Exponential growth
- Smart Everything
- Energy Efficiency
- Defense and other Industries

AI and Machine Learning

- AI at the edge – Advanced models trained in cloud and running in the edge devices

- Analytics – Healthcare monitoring, Smart Cities, Predictive Maintenance

- Low latency AI services- Vehicles, Drones, Robotics

- Smart Infrastructure

Enhanced Security and Privacy

- Decentralized security
- Privacy by design
- Regulatory Compliance

Thank you!