

VMAP

Verizon Media Acceleration Platform

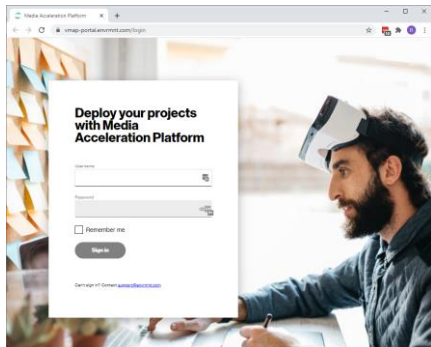
A hybrid edge platform from Verizon XR Labs

Richard Lamb
December 2020



Intro to VMAP

What is VMAP?



DIY container as a service platform

Configure and deploy compute tasks in any edge locations within seconds from a simple web portal.



First class support for GPUs

Virtualized GPUs mean client capacity scales exponentially with each hardware cycle without the need for any more power or space.



Deploy everywhere

Cloud edge - CDN locations
Telco edge - SAP and RAN
On Prem - private 5G or other network.

Problem Statement

- Verizon has a heterogeneous collection of thousands of network edge locations. Verizon Wireless has 50+ SAP sites and 1000s of RAN sites, Verizon Digital Media has 160 worldwide PoPs, Verizon Fios has PoPs in 10+ cities
- These locations have a wide range of network characteristics. Connections to end user devices, backbones and public cloud all have varying bandwidth, reliability, latency and jitter to consider
- We want developers to be able to deploy compute tasks at the optimal location given the available hardware and network resources to give users the very best experience possible

Where do you want to deploy today?



**MS Azure
Regional Cloud**

Quincy, WA, USA



**Equinix
Cloud Edge**

Downtown Toronto Canada



**Verizon RAN
Telco Edge**

Test RAN location in Houston TX

Regional Data Centers vs Cloud Edge vs Telco Edge vs On Prem

		Regional Data Centers	Multi-Cloud Regional	Cloud Edge	Co-located Telco Edge	Telco Edge	Private Edge
Latency	4G	75ms	70ms	55ms (+10%)	50ms	50ms	1ms
	5G	35ms	30ms	15ms (+50%)	10ms	10ms	1ms
	Fios	30ms	25ms	10ms (+100%)	5ms	5ms	1ms
Network Capability Exposure	5G NEF, DNS intercept, URLLC	✗	✗	✗	?	✓	✗
	QoS	✗	✗	✗	?	✓	✓
	Guaranteed Bandwidth	✗	✗	✗	?	✓	✓
Public Cloud Provider Services	AWS Lambda, IAM roles etc..	✓	✗	✗	?	✗	✗
	std. cloud services loadbalancers	✓	✓	✗	?	✗	✗

[Too many] technical details...

Orchestration Stack

Hardware Stack

VMAP federation and orchestration.

Shared cluster Paas.
Orchestration and federation of Kubernetes clusters within Edge Locations and across the whole network.



In cluster container orchestration:

Kubernetes clusters with NVIDIA docker support.
Docker registry provided by Jfrog Artifactory



Virtualization:

Red Hat Enterprise Linux® kernel, Kernel-based Virtual Machine (KVM) technology, and oVirt virtualization management.

NVIDIA GRID virtualized GPUs



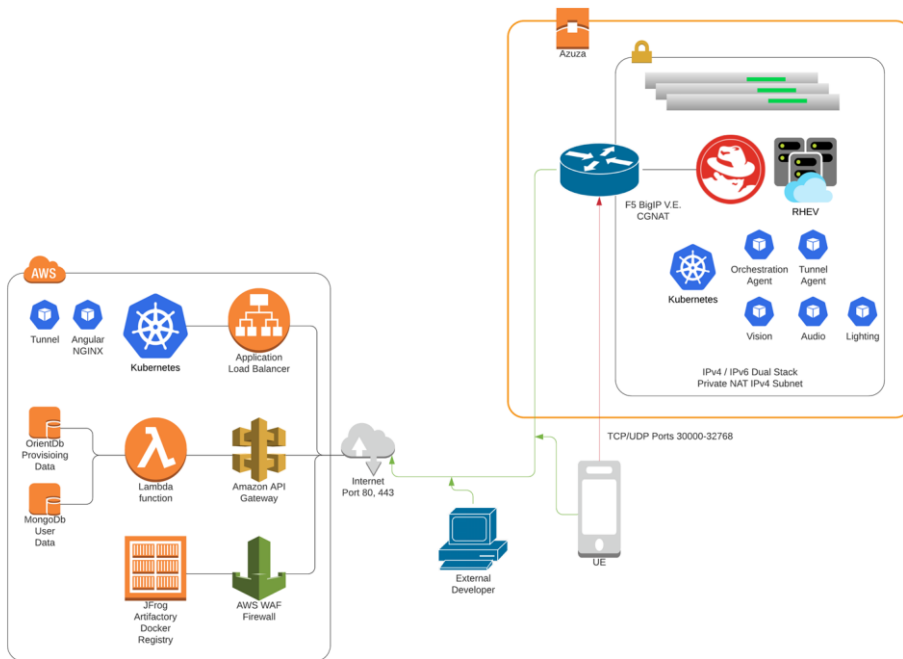
Hardware:

- HPE DL380
- 2x Xeon-S 4110, 128GB RAM, 2x 48GB NVIDIA RTX Quadro RTX8000

Software Architecture.

Architecture is conceptually simple. Edge locations poll central API and respond to changes that have been requested.

- Everything persistently stored in AWS
- Orchestration agent in edge cluster continuously monitors and mutates state to match central source of truth.
- Edge clusters are basically stateless except for auth and identity tokens.
- Software stack can be deployed over the top on any hardware or available VMs
- Ansible scripts automate standing up edge clusters. Around 3 hours to completely tear-down and rebuild.
- F5 BigIP VE is used for ingress and egress to the virtual cluster



Exponentially scalable within the same power and space footprint.

We developed our own custom server with HPE primarily thinking about locations with very limited power and space.

- Fastest chips and most memory we could get.
- 2x Xeon-S 4110, 128GB RAM, 2x 48GB NVIDIA RTX Quadro RTX8000
- KVM with NVIDIA GRID Virtualized GPU
 - Gives more possible clients now
 - Will scale better over time (Moore's law for GPUs)
 - Helps convince security teams unsure of K8s



Testing Results Non-Virtualized vs Virtualized GPUs

Non GPU Sliced Configuration							GPU Sliced Configuration						
Test Cases	FPS	FPS Pass/Fail	Server Response Time (mS)	Server Response within 20 mS	CPU Usage	GPU Usage	Test Cases	FPS	FPS Pass/Fail	Server Response Time (mS)	Server Response within 20 mS	CPU Usage	GPU Usage
1 Device connected to Sliced GPU	29.9	Pass	9.6	Pass	3-7%	2-3%	1 Device connected to Sliced GPU	29.0/29.5	Pass	12.5	Pass	~30%	8-13%
2 Device connected to Sliced GPU	23.2	Pass	10.2	Pass	7-10%	15-16%	2 Device connected to Sliced GPU	29.4/29.4	Pass	11.1	Pass	~30%	10-20%
3 Device connected to Sliced GPU	23.2	Pass	9.3	Pass	10-15%	26-27%	3 Device connected to Sliced GPU	30.0/30.0	Pass	4.5	Pass	~30%	7-16%
4 Device connected to Sliced GPU	8.3	Fail	29.9	Fail	15-18%	35-36%	4 Device connected to Sliced GPU	29.6/29.7	Pass	8	Pass	~30%	10-16%
5 Device connected to Sliced GPU	9.1	Fail	72.5	Fail	18-22%	45-46%	5 Device connected to Sliced GPU	30.0/30.0	Pass	4.3	Pass	~30%	10-20%
6 Device connected to Sliced GPU	9.6	Fail	50.7	Fail	22-25%	52-53%	6 Device connected to Sliced GPU	29.2/30.0	Pass	9.7	Pass	~30%	10-20%
7 Device connected to Sliced GPU	Did not test 7 or more Devices						7 Device connected to Sliced GPU	30.0/30.0	Pass	13.9	Pass	~30%	10-20%
							8 Device connected to Sliced GPU	30.0/30.0	Pass	13.7	Pass	~30%	10-20%
							9 Device connected to Sliced GPU	30.6/31.0	Pass	5.7	Pass	~30%	10-20%
							10 Device connected to Sliced GPU	22.3/26.8	Pass	66.4	Fail	~30%	10-20%
							11 Device connected to Sliced GPU	28.0/30.0	Pass	24.9	Fail	~30%	8-24%
							12 Device connected to Sliced GPU	29.3/28.5	Pass	6.6	Pass	~30%	8-24%
							13 Device connected to Sliced GPU	29.7/29.7	Pass	12.1	Pass	~30%	10-17%
							14 Device connected to Sliced GPU	30.6/30.4	Pass	4.5	Pass	~30%	7-20%
							15 Device connected to Sliced GPU	30.0/30.4	Pass	7.1	Pass	~30%	10-20%
							16 Device connected to Sliced GPU	27.5/27.6	Pass	15.7	Pass	~30%	10-20%
Tested on DL360							Tested on DL380						

Use Cases

Computer Vision API for XR

Benefits of Edge Based CV

- Real-time horizontal scaling computer vision for 2D and 3D recognition and tracking
- Supports full or hybrid render
- Fully featured web GUI for designing and training targets
- Supports occlusion, classification, and segmentation



Raytracing API

Benefits of Edge Raytracing

- Supports high end desktop quality graphics on mobile devices
- NVIDIA RTX hardware via Vulkan API
- Lightmap atlas stream combines with local rendering pipeline for hybrid lighting effects



Edge XR Lighting API

Benefits of Edge Based Lighting

- Real-time lighting / reflections on mobile devices for XR applications
- Low latency makes real-time reflections in XR possible
- End user setup of live AR portals and subscription could support new types of shopping / tourist / entertainment experiences
- Could also be used for generating in game lighting / reflections, instead of camera, imagery can come from virtual world to provide higher fidelity mobile experience than is available on current hardware



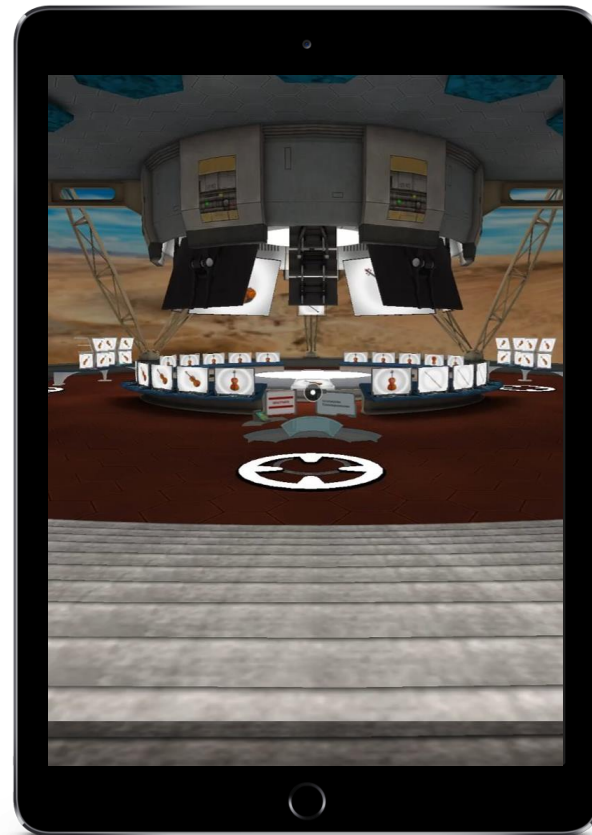
Spatial Audio API

MEC Deployment Requirements

- GPU & CPU architecture variants
- Requires a dedicated container per application
 - Each container capable of supporting multiple concurrent users
- Resource requirements determined by application
- Container launched upon application request and delete on disconnect.

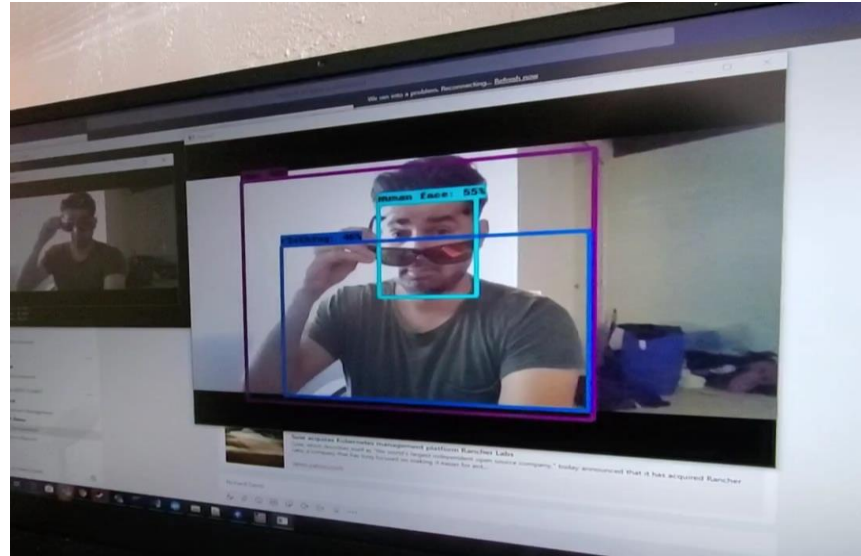
Benefits of MEC

- Ability to offload expensive audio processing to free client resources & enable thin clients
- 15 – 50ms total RTT target, including DSP
- Supports high-end, desktop-quality immersive & social audio experiences, otherwise impossible on mobile
- Supports large number of concurrent OTT voice chat & spatial audio users per application
- Able to do machine learning on voice for captioning, translation, & other data analyses



Object Detection and Video Analytics

- Real-time detection with notifications
- Pipeline management from the edge
- Highly scalable monitoring, alerting and presence management from low cost sensors



Partnerships

Company	
Oculus	Vision, rendering
Epic Games	Vision, raytracing
Nvidia	Stack, ingest, raytracing
Disney	Ingest, raytracing
Sony Music	Ingest, spatial audio
HPE	Stack, HW Spec



THANK YOU

Richard Lamb
richard.lamb@verizon.com



Q&A

Computer Vision API for XR

MEC deployment requirements

- 2 GB vGPU RAM per instance for Object Detection
- Requires dedicated container for each user
- Container launched upon user login and delete on disconnect.
- Minimum 1.5Mbps network link per user upstream and downstream

Benefits of MEC

- Real-time horizontal scaling computer vision for 2D and 3D recognition and tracking
- Supports full or hybrid render
- Fully featured webGUI for designing and training targets
- Supports occlusion, classification, and segmentation



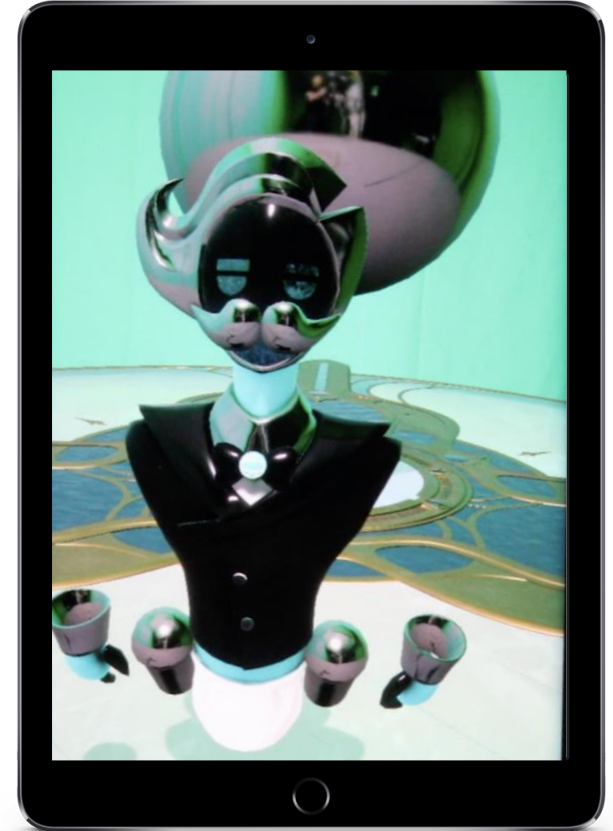
MEC XR Lighting API

MEC Deployment Requirements

- 1GB vGPU per Camera / Environment
- Up to 64 users per container
- Container launched per camera
- 10gbps network total downstream
- 150mbps per user downstream

Benefits of MEC

- Real-time lighting / reflections on mobile devices for XR applications
- Low latency makes real-time reflections in XR possible
- End user setup of live AR portals and subscription could support new types of shopping / tourist / entertainment experiences
- Could also be used for generating in game lighting / reflections, instead of camera, imagery can come from virtual world to provide higher fidelity mobile experience then is available on current hardware



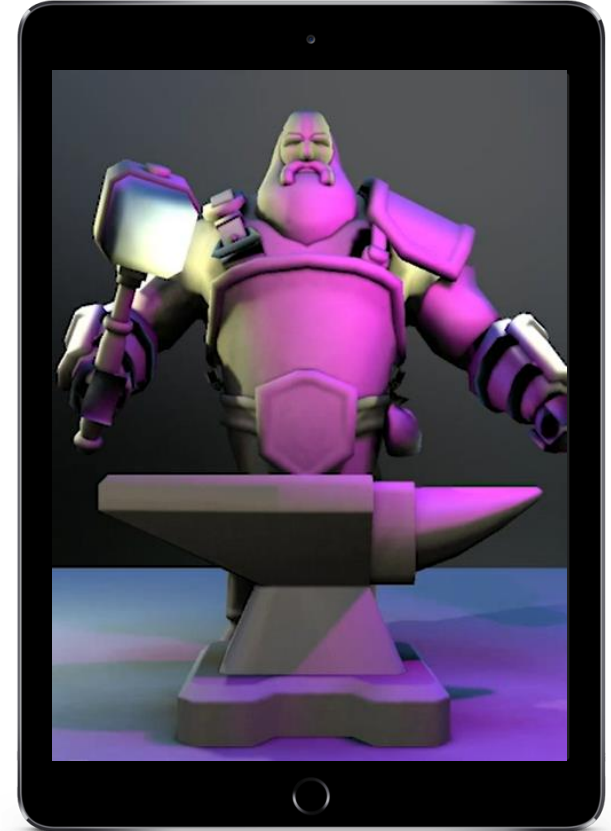
Ray tracing API

MEC Deployment Requirements

- Up to 2 GB vGPU RAM per instance
- Requires dedicated container for each user
- Container launched upon user login and delete on disconnect.

Benefits of MEC

- Supports high end desktop quality graphics on mobile devices
- End user can make their own rendering requests
- End user gets dedicated container for custom rendering



Safe harbor statement

NOTE: In this presentation we have made forward-looking statements. These statements are based on our estimates and assumptions and are subject to risks and uncertainties. Forward-looking statements include the information concerning our possible or assumed future results of operations. Forward-looking statements also include those preceded or followed by the words “anticipates,” “believes,” “estimates,” “expects,” “hopes” or similar expressions. For those statements, we claim the protection of the safe harbor for forward-looking statements contained in the Private Securities Litigation Reform Act of 1995. We undertake no obligation to revise or publicly release the results of any revision to these forward-looking statements, except as required by law. Given these risks and uncertainties, readers are cautioned not to place undue reliance on such forward-looking statements. The following important factors, along with those discussed in our filings with the Securities and Exchange Commission (the “SEC”), could affect future results and could cause those results to differ materially from those expressed in the forward-looking statements: adverse conditions in the U.S. and international economies; the effects of competition in the markets in which we operate; material changes

in technology or technology substitution; disruption of our key suppliers’ provisioning of products or services; changes in the regulatory environment in which we operate, including any increase in restrictions on our ability to operate our networks; breaches of network or information technology security, natural disasters, terrorist attacks or acts of war or significant litigation and any resulting financial impact not covered by insurance; our high level of indebtedness; an adverse change in the ratings afforded our debt securities by nationally accredited ratings organizations or adverse conditions in the credit markets affecting the cost, including interest rates, and/or availability of further financing; material adverse changes in labor matters, including labor negotiations, and any resulting financial and/or operational impact; significant increases in benefit plan costs or lower investment returns on plan assets; changes in tax laws or treaties, or in their interpretation; changes in accounting assumptions that regulatory agencies, including the SEC, may require or that result from changes in the accounting rules or their application, which could result in an impact on earnings; the inability to implement our business strategies; and the inability to realize the expected benefits of strategic transactions.

As required by SEC rules, we have provided a reconciliation of the non-GAAP financial measures included in this presentation to the most directly comparable GAAP measures in materials on our website at www.verizon.com/about/investors