AI/ML and AR/VR applications at Edge

Sukhdev Kapur

Vikram Siwach

Project Technical Lead: Vikram Siwach

Project Committers detail:

Committer	Committer Company	Committer Contact Info	Committer Bio	Committer Picture	Self Nominate for PTL (Y/N)
Sukhdev Kapur	Juniper Networks	Sukhdev Kapur			
Vikram Siwach	MobiledgeX	Vikram Siwach			

Presentation:

Blueprint species:

Attributes	Description	Informational
Type	New Blueprint for enabling AI/ML and low latency AR/VR capabilities at the Edge	New submission for integrating Tungsten Fabric for Edge Cloud for AI/VR applications for developers in a distributed environment
Industry Sector	Edge Cloud, Enterprise, 5G, and IoT	The blueprint empowers Integrated Edge Cloud architecture to use single SDN Controller (Tungsten Fabric) for all type of Containers, Virtual Machines, and Bare Metal Servers.

Blueprint Family - Proposed Name	Integrated Edge Cloud	
Use Case	Programmability on switches and I/O Accelerations on programmable NICs & embedded FPGAs to deliver AI/ML workload placement and low latency demands of AR/VR applications onboarding edge stack	
Blueprint proposed Name	AI/ML and AR/VR applications at Edge	
Initial POD Cost (capex)	Leverage white boxes, standard NICs: The cost of POD will depend upon the hardware profiles and peripherals desired	
Scale & Type	The deployment will employ baremetal as well virtual machines and containers. The smallest footprint for minimal Cloudlet is 9VMs to a large Cloudlet could span to 8-Cluster Cloudlets	
Applications	AI/ML streaming workloads and AR/VR applications	
Power Restrictions	Less than 10Kw	
Infrastructure orchestration	OpenStack Queens or above Docker 1.13.1 or above Container Orchestration –K8s 1.10.2 or above OS - Ubuntu 16.x, CentOS	
SDN	Tungsten Fabric Kernel vRouter, DPDK vRouter, SR-IOV and SmartNIC	
Workload Type	Containers over VM or baremetal	
Additional Details	Runs on Commodity HW x86, ARM, SoC, ... Multiple options for partial or full NIC offloads (Intel, Netronome, Mellanox) Future Supports: eBPF/XDP offload	