Full Scalable Media Cloud Solution with Kubernetes Orchestration

Zhenyu Wang， Xin（Owen）Zhang

LF ASIA, LLC

Agenda

- Media in the Network and Cloud
- Intel Media Server Reference Software Stack
- Container with MSS enablement
- Kubernetes with Container integration
- Kubernetes with Container enabling on VCA2
- Kubernetes device plugin/Intel GPU plugin
- Use Case(1080p): VCA transcoding & k8s scheduling on VCA nodes

Media in the Network and Cloud

Visual Understanding

Object Recognition & Tracking

Indexing / Search

Smart Cities, Security and Surveillance

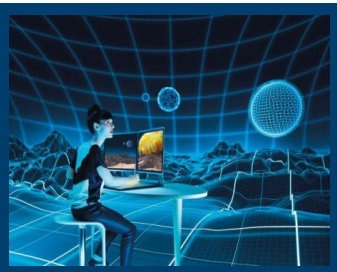
Video Delivery

Cloud and Comms:

Ingest / Storage / Edge

Transcode / Trans-size / Trans-rate

Video Conferencing



Graphics in the Cloud

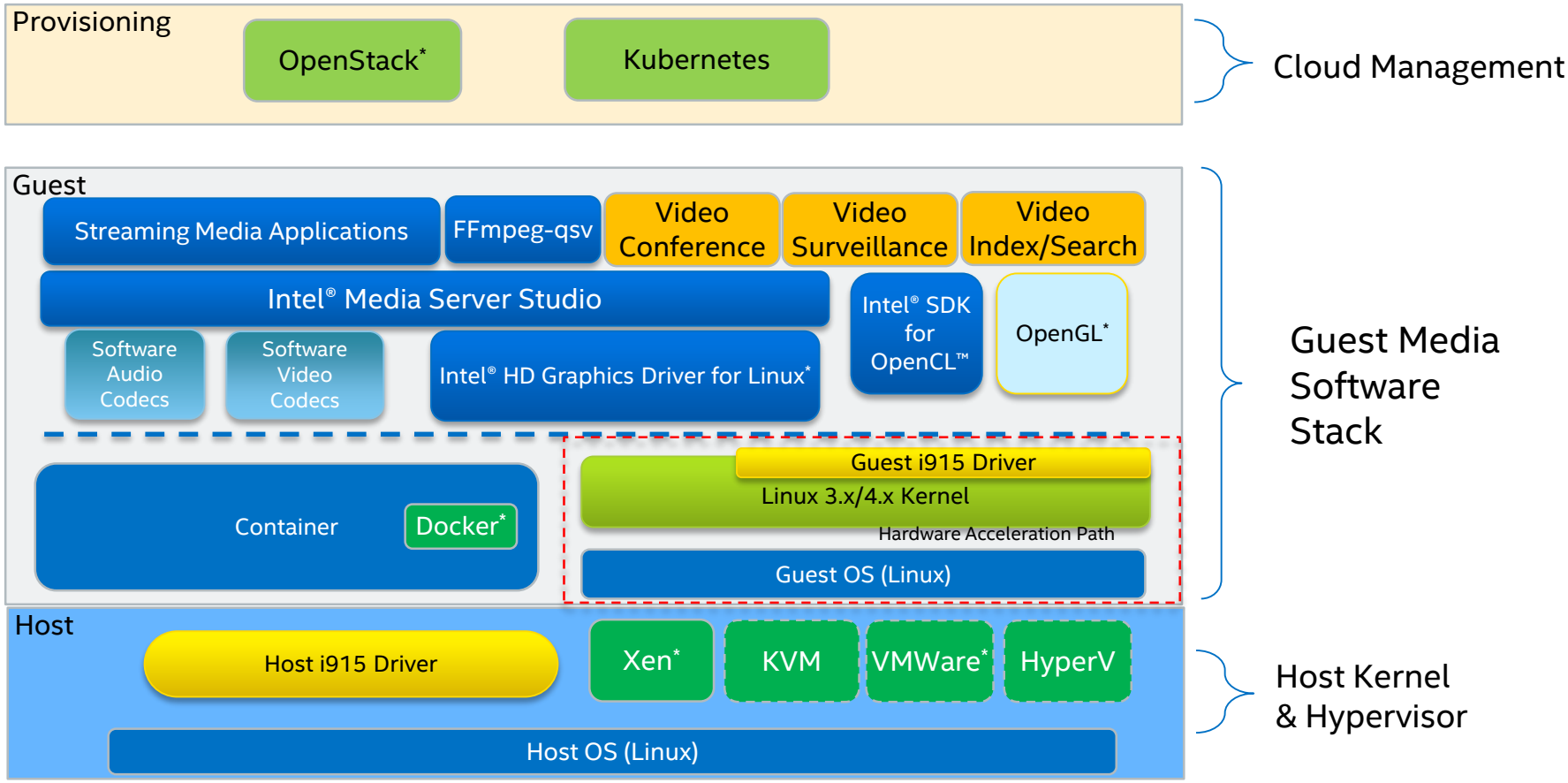
Remote Desktop

Remote Workstation

Cloud Gaming

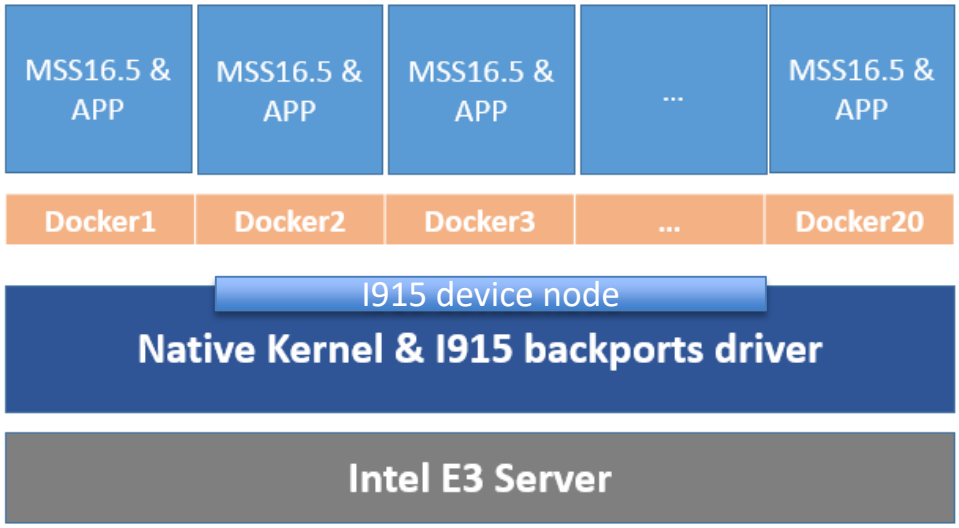
Rendering

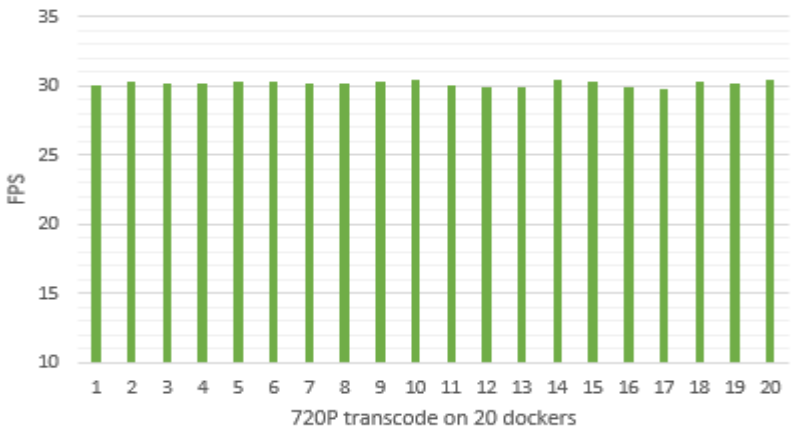
Intel® Media Server Reference Software Stack



Container with MSS enablement

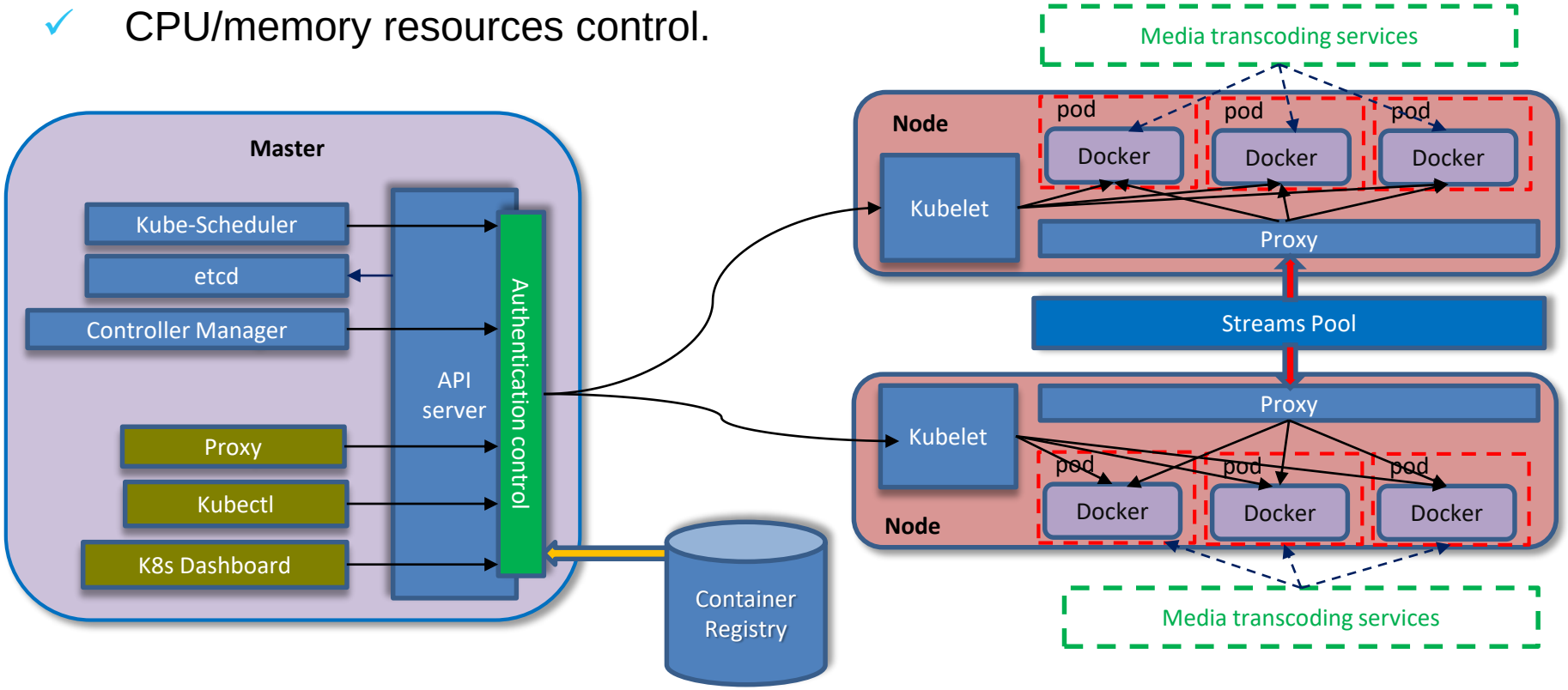
- More Containers can be run than VMs
- Almost same performance with Native
- Package application and dependencies integrated
- Share same kernel as the host
- No need providing hardware based on the isolation





Kubernetes with Container integration

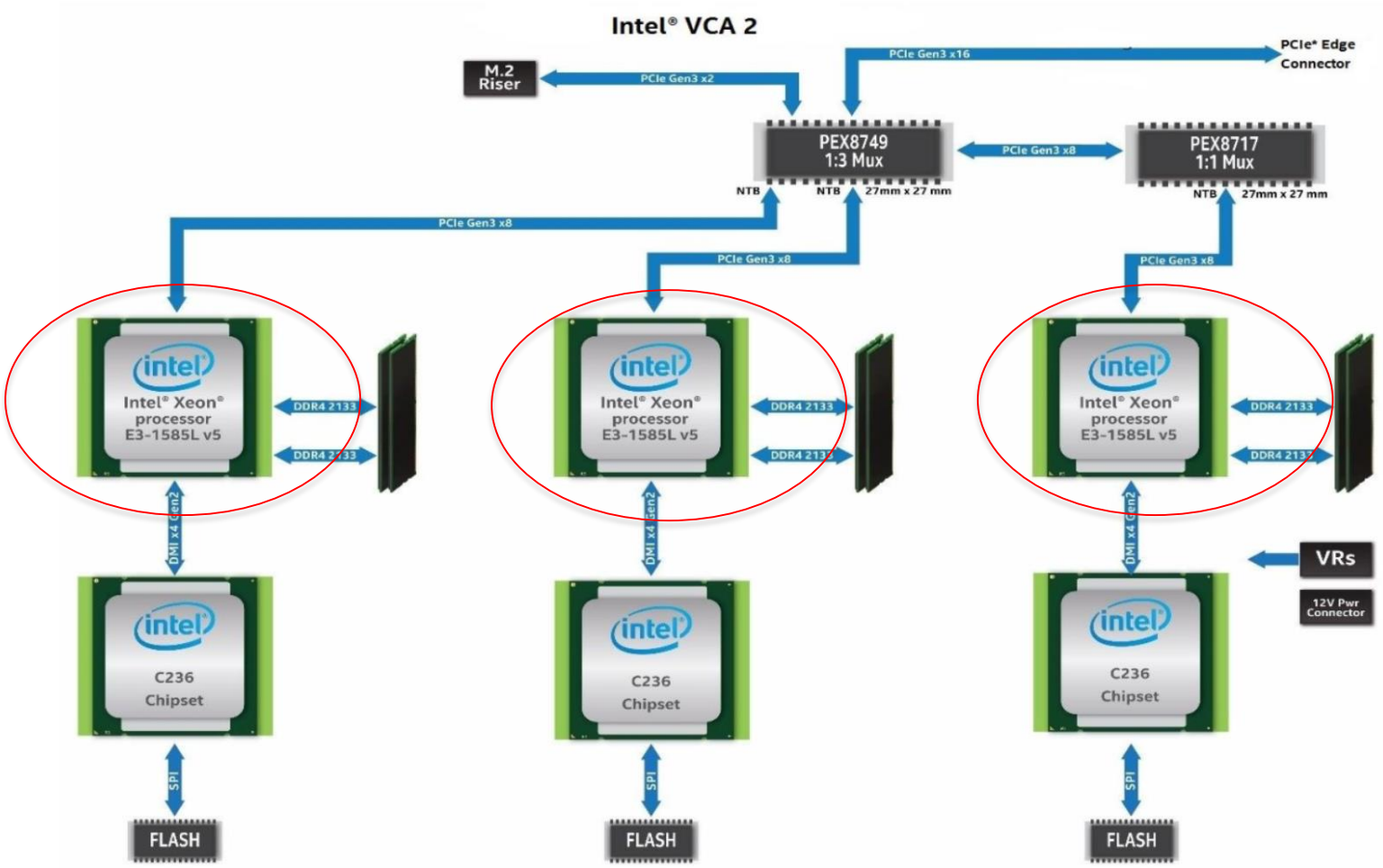
- ✓ Orchestrate intelligent scheduling, self-healing, and horizontal scaling for rapid delivery of apps.
- ✓ Easy to customize the worker nodes.
- ✓ CPU/memory resources control.



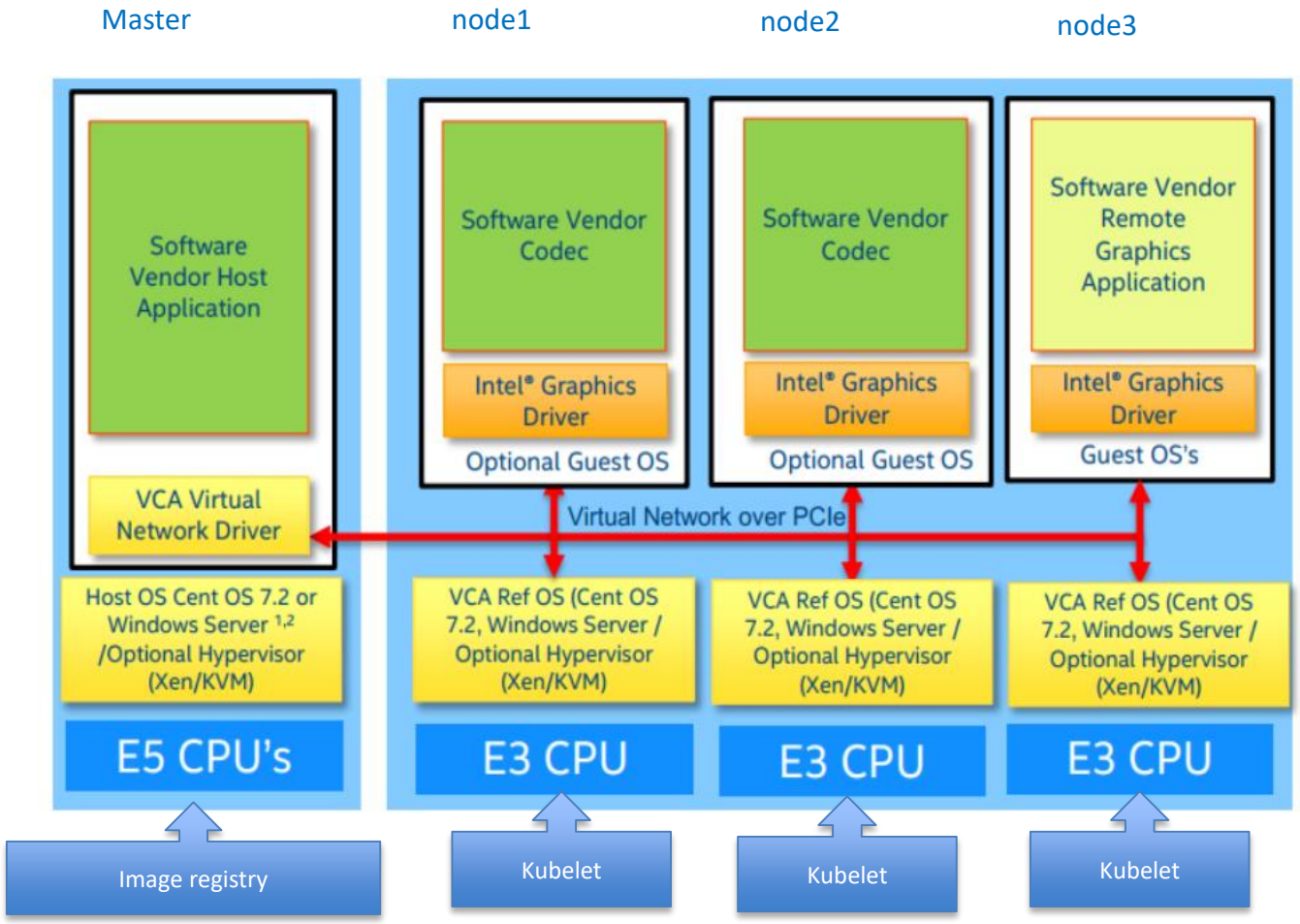
Kubernetes with Container enabling on VCA2

- Intel® Visual Compute Accelerator 2 (Intel® VCA 2) (VCA1585LMV)
  - *Intel® Xeon® Scalable processor and Intel® Xeon® processor E5-based platforms with Iris® Pro Graphics and Intel® Quick Sync Video media transcode capabilities*
- Applications include:
  - *Broadcast – Ultra-high channel density with high visual quality.*
  - *Remotely rendered graphics – High video quality, low latency graphics for enterprise productivity and anytime anywhere gaming.*
  - *Multi-party communication – Video-enabled B2B, B2C, and C2C communication with massive scaling*

Kubernetes with Container enabling on VCA2



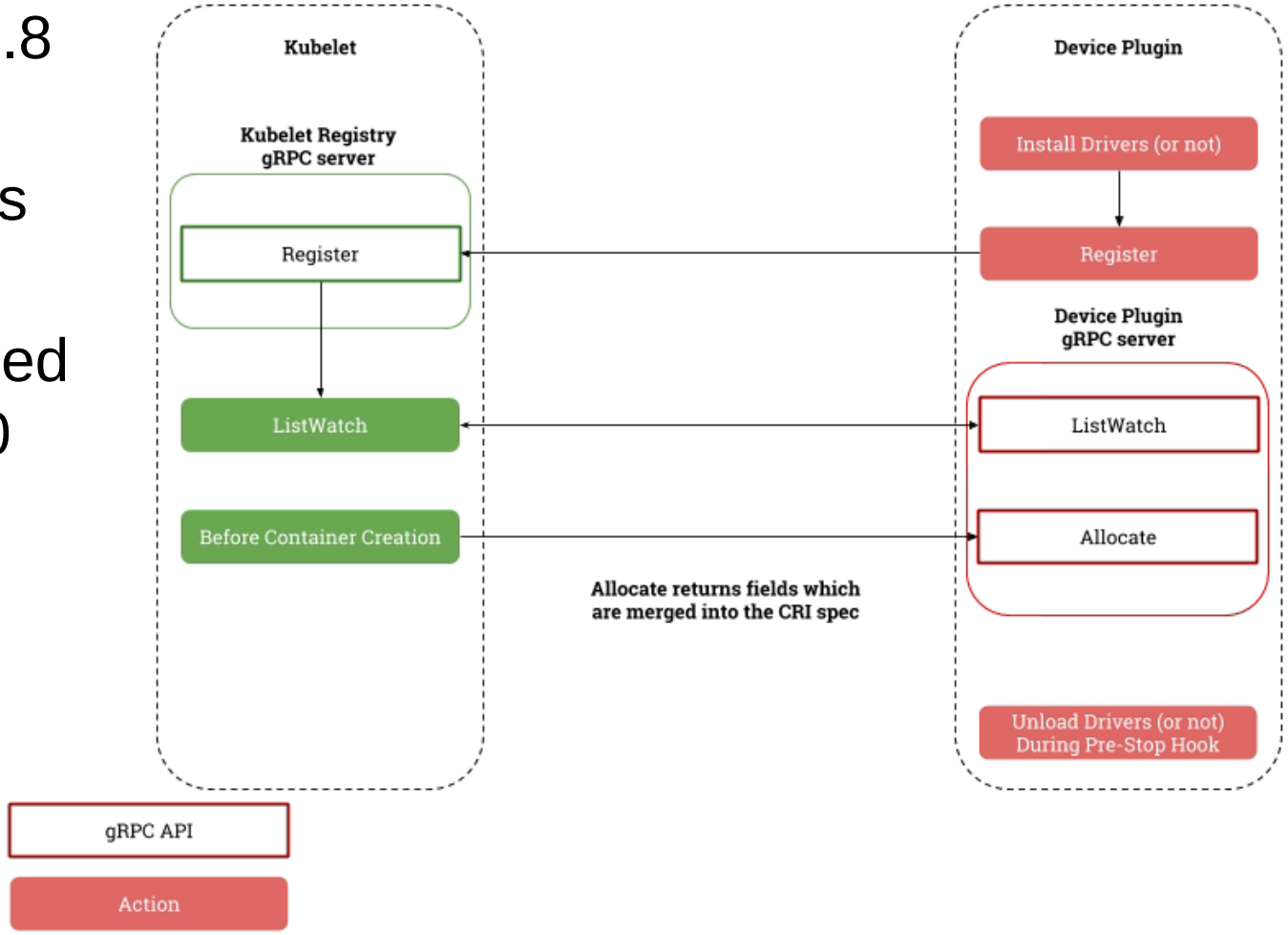
Kubernetes with Container enabling on VCA2



- 2-4 VCA2 cards
- 12 k8s nodes

Kubernetes device plugin

- k8s version 1.8 above
- DevicePlugins feature gated
- Default enabled after k8s 1.10



Intel GPU device plugin

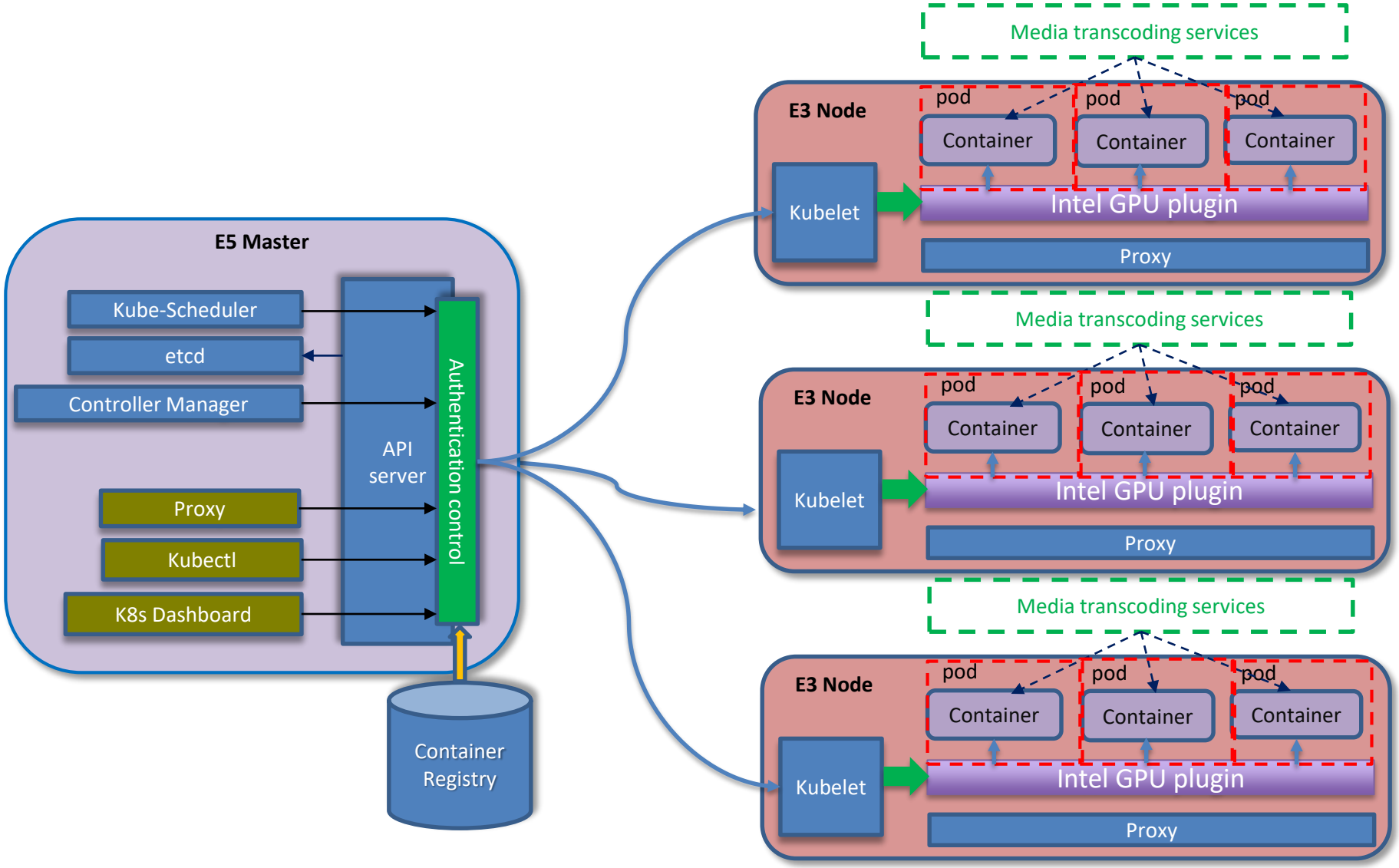
- Enumerate all DRM devices for Intel GPU
  - /dev/dri/card[0-9]*
  - /dev/dri/renderD[0-9]*
- Create server and sock
  - /var/lib/kubelet/device-plugins/intelGPU.sock
- Register resource to kubelet
  - Resource name: “intel.com/gpu”
- Serve “ListAndWatch” request
  - Send device id and healthy
- Serve “Allocate” request
  - Return device id’s HostPath/ContainerPath to assign

Use of Intel GPU plugin


```
apiVersion: v1
kind: Pod
metadata:
  name: intelgpu-demo-pod
spec:
  containers:
    -
      name: intelgpu-demo-container-1
      image: ubuntu-demo:latest
      imagePullPolicy: IfNotPresent
      command: [ "..." ]
      resources:
        limits:
          intel.com/gpu: 1
```

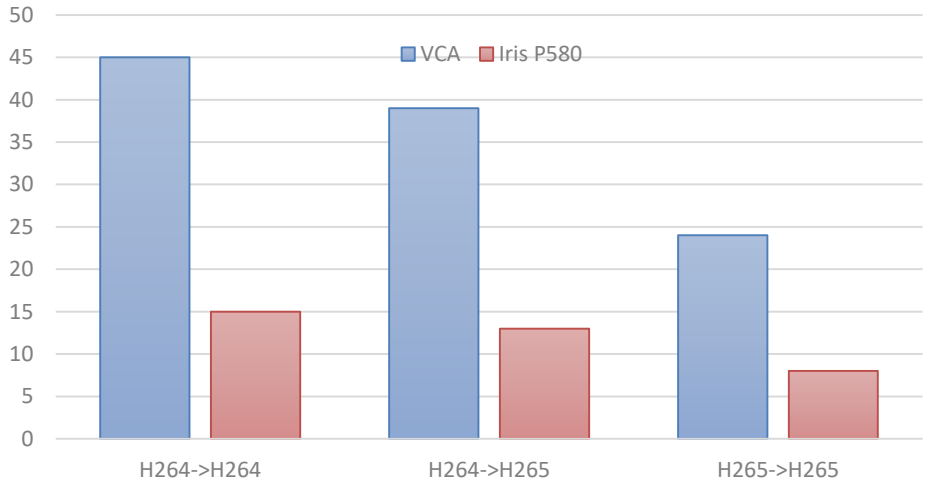
- A yaml file for pod
- https://github.com/intel/intel-device-plugins-for-kubernetes

Intel GPU plugin in Kubernetes on VCA2



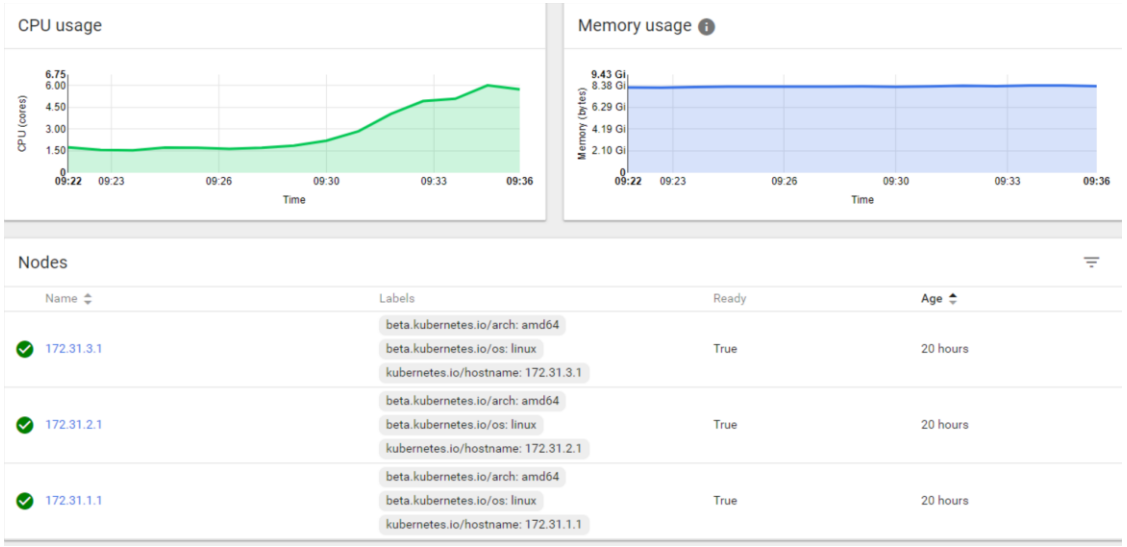
Use Case(1080p): Media transcoding on VCA2

- *the performance comparison of number of real time streams between per VCA card and per Iris P580*



MEDIA TRANSCODING		
Number of real time streams per VCA card @30fps		
1080p	H.264 -> H.264	45
	H.264 -> H.265	39
	H.265 -> H.265	24

Use Case(1080p): k8s scheduling on VCA2




```
[root@localhost demo]# kubectl get nodes
NAME          STATUS    AGE       VERSION
172.31.1.1    Ready     9m        v1.7.3
172.31.2.1    Ready     9m        v1.7.3
172.31.3.1    Ready     15s       v1.7.3
[root@localhost demo]# kubectl get componentstatus
NAME                 STATUS    MESSAGE              ERROR
scheduler            Healthy   ok
controller-manager   Healthy   ok
etcd-0               Healthy   {"health": "true"}
etcd-3               Healthy   {"health": "true"}
etcd-1               Healthy   {"health": "true"}
etcd-2               Healthy   {"health": "true"}
```

```
[root@vca_node_01 demo]# docker images
REPOSITORY                                  TAG                 IMAGE ID
10.67.117.154:5000/centos7-3                local_trans_ffmpeg  8fe8575d8816
10.67.117.154:5000/pod-infrastructure       latest              f66f4bd9b894
```

```
root@vca_node_02 demo]# docker images
EPOSITORY                                   TAG                 IMAGE ID
0.67.117.154:5000/centos7-3                 local_trans_ffmpeg  8fe8575d8816
0.67.117.154:5000/pod-infrastructure        latest              f66f4bd9b894
0.67.117.154:5000/kubernetes-dashboard-amd64 v1.6.1             71dfe833ce74
```


LINUXCON
containercon
CLOUDOPEN
CHINA 中国
THINK OPEN
开放性思维
LF ASIA, LLC