Alexander Kanevskiy
Cloud Software Architect, Intel

Advanced platform features in Kubernetes*

2018-11-02

DADDY, WHAT ARE CLOUDS MADE OF?

LINUX SERVERS, MOSTLY

Agenda

- Memory management: Huge Pages
- CPU management: CPU Manager & CMK
- Node features: Node Feature Discovery
- Multiple network interfaces: Multus CNI, SR-IOV CNI
- Hardware Accelerators: Intel Device Plugins (GPU, QAT, FPGA)
- Container Runtimes: Kata Containers

Memory management

Huge Pages

Huge Pages

- Multiple architectures support
  - i386: 4K, 2M
  - x86_64: 4k, 2M, 1G
  - aarch64: 4k, 2M, 1G
- Native Huge page support
  - Alpha in 1.8
  - Beta in 1.10
- First class resources
  - hugepages-2Mi
  - hugepages-1Gi

- Application usages
  - Java*
    - -XX:+UseLargePages
  - Memcached*
    - memcached -L
  - MySQL*
    - [mysqld]
      large-pages

*Other names and brands may be claimed as the property of others.

Huge Pages

- Usage
  - Request resource
  - Volume mount
- Limitations
  - Pre-allocation
  - Pod level resources
  - NUMA locality
- Links
  - https://kubernetes.io/docs/tasks/manage-hugepages/scheduling-hugepages/
  - https://wiki.debian.org/Hugepages

```
containers:
   ...
   volumeMounts:
   - mountPath: /hugepages
     name: hugepage
   resources:
     limits:
       hugepages-2Mi: 100Mi
 volumes:
 - name: hugepage
   emptyDir:
     medium: HugePages
```

CPU Management

CPU Manager & CMK

CPU Manager

- CPU Manager feature
  - Alpha in 1.8
  - Beta in 1.10, enabled by default
- Kubelet configuration
  - --cpu-manager-policy=static
  - --cpu-manager-reconcile-period=5s
  - --kube-reserved=cpu=X
  - --system-reserved=cpu=X
- CPU Pools
  - Reserved
  - Exclusive
  - Shared
- Types of workload
  - Guaranteed
  - Burstable
  - Best Effort

CPU Manager

- Best Effort
  - Resources in Requests and Limits are not specified
- Burstable
  - Limits > Requests
- Guaranteed
  - Requests == Limits
  - Requests not specified, only Limits

- CPU Pools
  - Exclusive
    - Guaranteed with integer CPU requests
  - Shared
    - Guaranteed
    - Burstable
    - Best Effort

CPU Manager

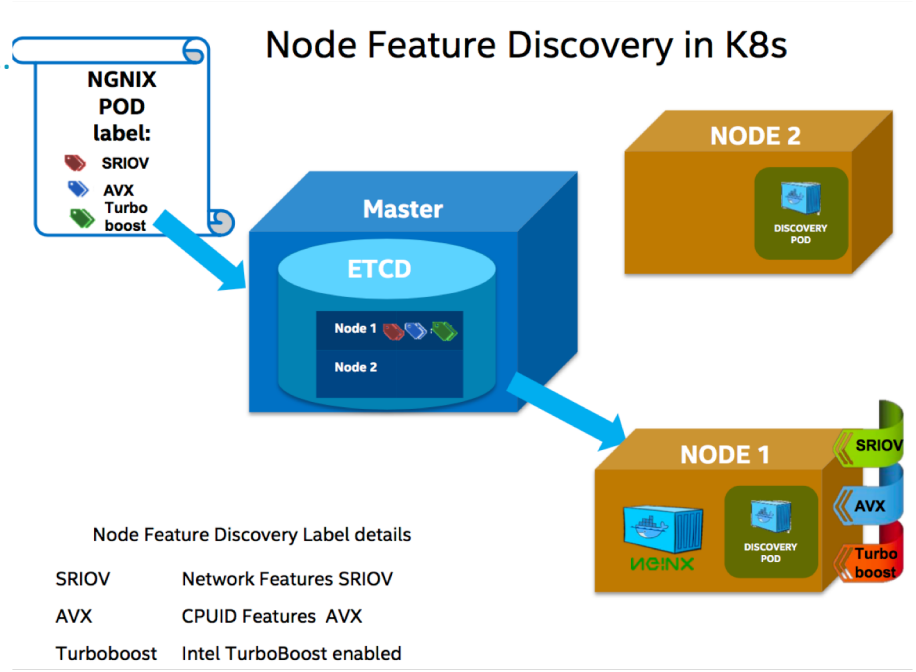
- CPU Manager for Kubernetes
  - CPU manager for NFV workloads
  - More features, off-tree
  - https://github.com/Intel/CPU-Manager-for-Kubernetes
- Links
  - https://kubernetes.io/blog/2018/07/24/feature-highlight-cpu-manager/
  - https://kubernetes.io/docs/tasks/administer-cluster/cpu-management-policies/
  - Topology Manager / NUMA https://github.com/kubernetes/community/pull/1680
  - RDT https://github.com/kubernetes/community/pull/1733

Node features

Node Feature Discovery

Node feature discovery for Kubernetes

- node.alpha.kubernetes-incubator.io
  - CPUID for x86 CPU details: AESNI, AVX, SGX, …
  - Intel Resource Director Technology
  - Intel P-State driver
  - Network: SR-IOV
  - Storage: SSDs
  - PCI devices / accelerators
- Links
  - https://github.com/kubernetes-incubator/node-feature-discovery
  - https://github.com/redhat-performance/openshift-psap



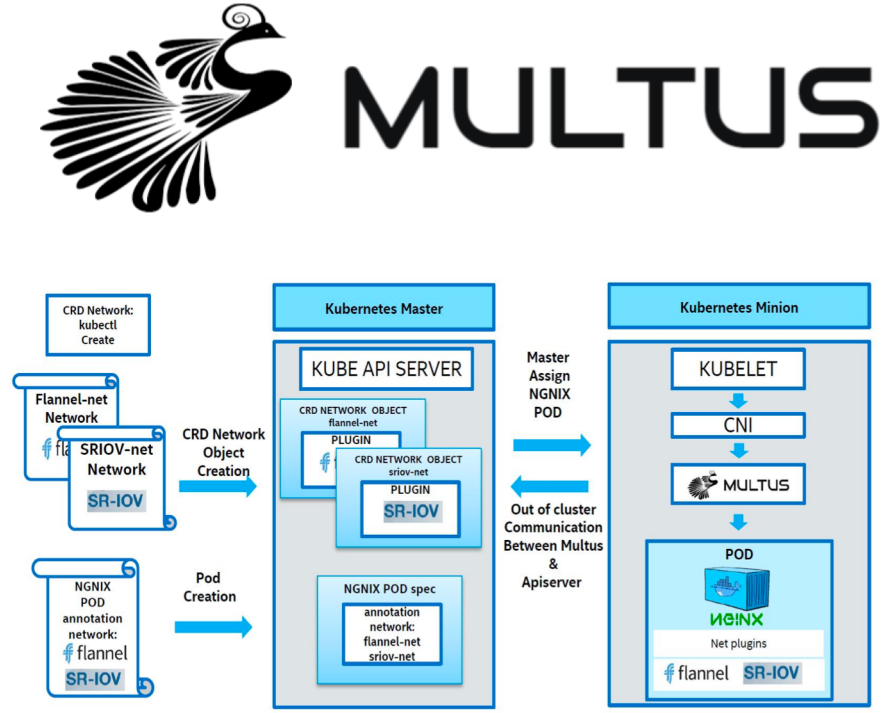
Multiple network interfaces

Multus CNI, SR-IOV CNI

Multus: multiple network interface for Pods

- Compatible with reference (flannel, DHCP,…) and 3rd party plugins (Calico*, Weave*, …)
- Utilizes CRDs for network plugin configurations
- Utilizes Pod Annotations to specify requested networks

- Links
  - https://github.com/Intel/multus-cni
  - https://github.com/Intel/sriov-cni
  - https://github.com/intel/userspace-cni-network-plugin



*Other names and brands may be claimed as the property of others.

Hardware Accelerators

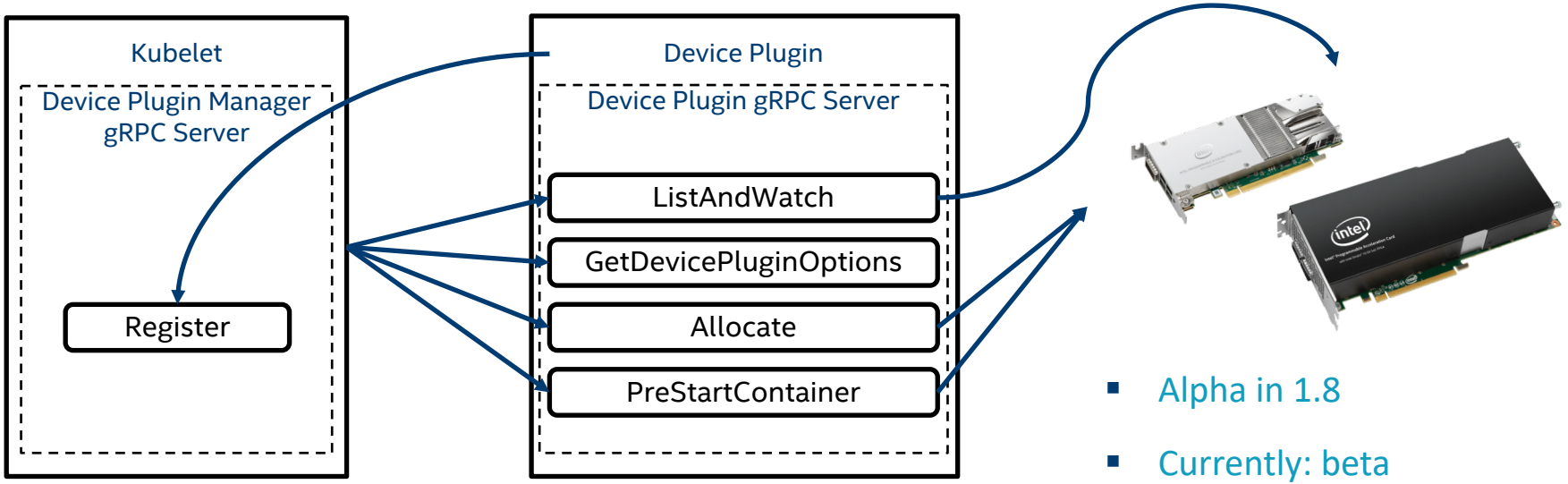
Intel Device Plugins for Kubernetes*: GPU, QAT, FPGA

Hardware Accelerators: “What?” and “Why?”

- Variety of devices
  - GPUs
  - SmartNICs
  - FPGAs
- Highly optimized hardware for specific tasks
- Significant performance gain
- Saves CPU cycles for other workloads
- Require application support

- Links
  - https://github.com/intel/intel-device-plugins-for-kubernetes

Hardware Accelerators in Kubernetes*



- Alpha in 1.8
- Currently: beta

- Links
  - https://kubernetes.io/docs/concepts/extend-kubernetes/compute-storage-net/device-plugins/

*Other names and brands may be claimed as the property of others.

Hardware Accelerators: Intel GPU

- Intel GPUs
  - Integrated in Intel Core and Xeon processors
  - Intel Visual Compute Accelerator

- GPU Acceleration use cases
  - Intel Media SDK
  - OpenCL*

```
containers:
 ...
 - name: demo-container-1
   image: k8s.gcr.io/pause:2.0
   resources:
     limits:
      gpu.intel.com/i915: 1
```

- Links
  - https://github.com/intel/intel-device-plugins-for-kubernetes
  - https://www.intel.com/content/www/us/en/products/servers/accelerators.html

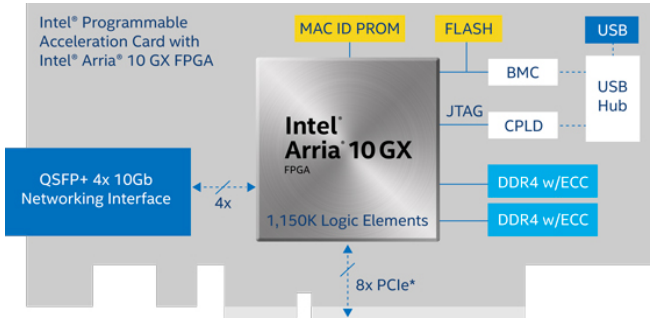
*Other names and brands may be claimed as the property of others.

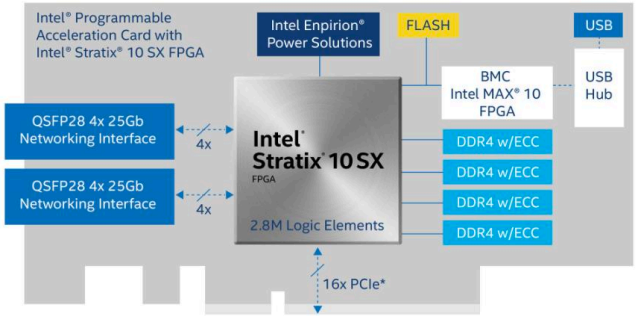
Hardware Accelerators: Intel QuickAssist Technology

- Accelerator for DPDK applications
  - Security
    - SSL/IPSec
    - Symmetric Cryptography: AES, KASUMI,…
    - Asymmetric cryptography: DH, RSA, DSA, ECDSA,…
    - Digest/hash: MD5, SHA1, SHA2, SHA3,…
  - Compression
    - Deflate: LZ77 with gzip or zlib header
    - Stateless compression and decompression
- Links
  - https://github.com/intel/intel-device-plugins-for-kubernetes
  - https://01.org/intel-quickassist-technology
  - https://www.intel.com/content/www/us/en/ethernet-products/gigabit-server-adapters/quickassist-adapter-for-servers.html

Hardware Accelerators: Intel FPGAs

- OPAE and OpenCL*
- Pre-programmed accelerators
- Orchestration programming
- Access control





- Links
  - https://opae.github.io
  - https://github.com/intel/intel-device-plugins-for-kubernetes
  - https://www.intel.com/content/www/us/en/programmable/solutions/acceleration-hub/platforms.html

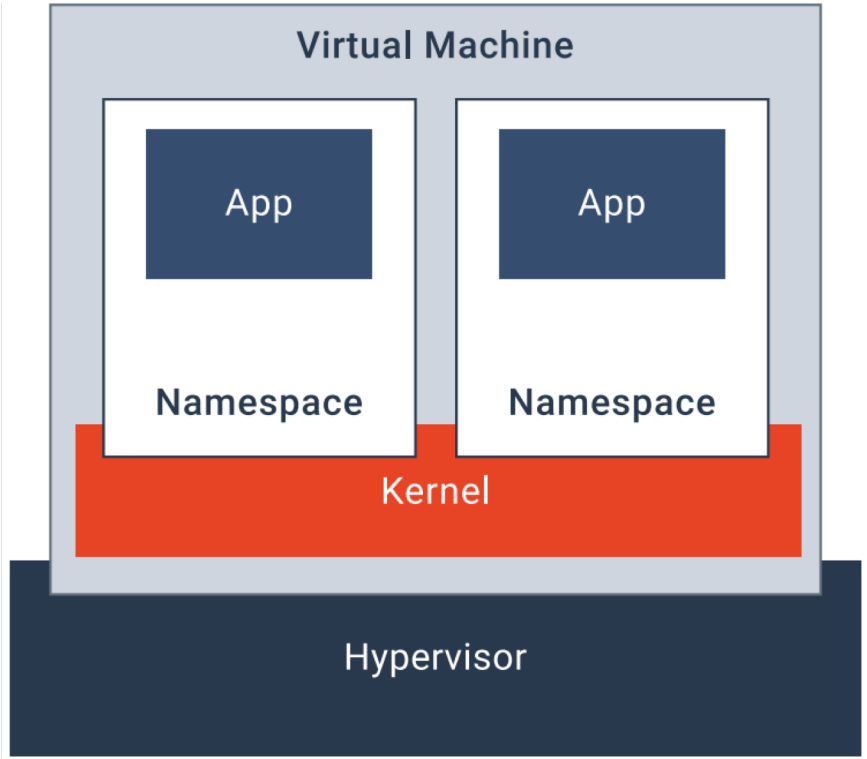
*Other names and brands may be claimed as the property of others.

Container Runtimes

Kata Containers

Containers in the cloud

- VMs on top of server hardware
- VM kernel shared for all containers
- One VM to one Kubernetes* control plane

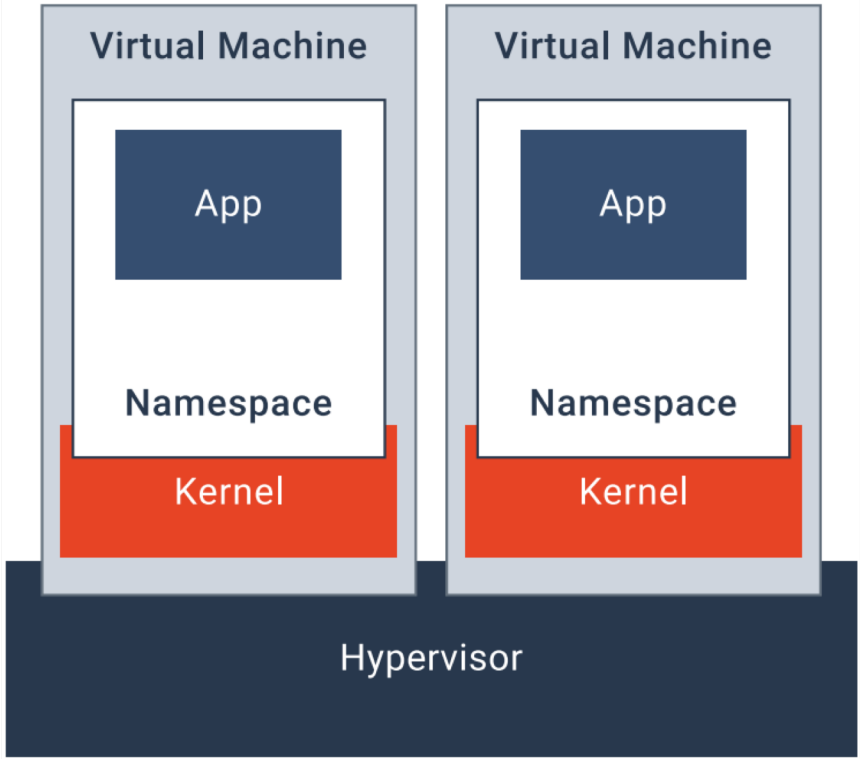


*Other names and brands may be claimed as the property of others.

Kata Containers

- The speed of containers, the security of VMs
- Small as a container
  - Minimal RootFS and Kernel
  - VM template
  - NVDIMM
  - De-duplicate memory across VMs

- Links
  - https://katacontainers.io
  - https://github.com/kata-containers/



What now ?

Join communities and contribute!

Questions ?

Links

- Huge Pages
  - https://kubernetes.io/docs/tasks/manage-hugepages/scheduling-hugepages/
  - https://wiki.debian.org/Hugepages
- CPU Manager & CMK
  - https://kubernetes.io/blog/2018/07/24/feature-highlight-cpu-manager/
  - https://kubernetes.io/docs/tasks/administer-cluster/cpu-management-policies/
  - Topology Manager / NUMA https://github.com/kubernetes/community/pull/1680
  - RDT https://github.com/kubernetes/community/pull/1733
  - https://github.com/Intel/CPU-Manager-for-Kubernetes
- Node Feature Discovery
  - https://github.com/kubernetes-incubator/node-feature-discovery
  - https://github.com/redhat-performance/openshift-psap
- Multus & SR-IOV
  - https://github.com/Intel/multus-cni
  - https://github.com/Intel/sriov-cni
  - https://github.com/intel/userspace-cni-network-plugin
- Device Plugins
  - https://kubernetes.io/docs/concepts/extend-kubernetes/compute-storage-net/device-plugins/
  - https://github.com/intel/intel-device-plugins-for-kubernetes/
  - https://www.intel.com/content/www/us/en/architecture-and-technology/intel-quick-assist-technology-overview.html
  - https://01.org/intel-quickassist-technology
  - https://opae.github.io
- Kata Containers
  - https://katacontainers.io
  - https://github.com/kata-containers/
  - Slack: https://katacontainers.slack.com/

Thank you!

Email: alexander.kanevskiy@intel.com

GitHub*: https://github.com/kad

Kubernetes* Slack*: @akanevskiy

INTEL OPEN SOURCE TECHNOLOGY CENTER | 01.org

Legal notices and disclaimers

- Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.
- Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.
- *Other names and brands may be claimed as the property of others.
- © Intel Corporation