



LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

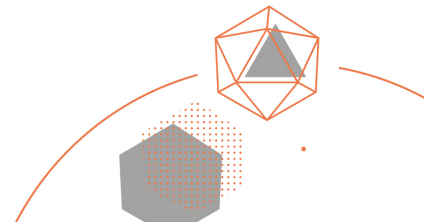
Kubernetes* Resource Management

Alexander Kanevskiy

Krisztian Litkey

Intel, Finland

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



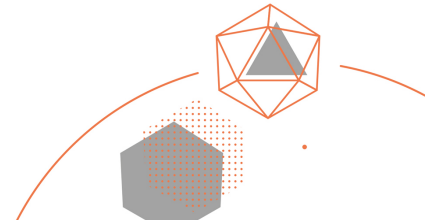


LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Agenda

- The “Noisy neighbor” problem in Kubernetes*
- Small detour into the history of hardware
- Resources in Kubernetes*: what do we have now
- What we can tweak on different levels
- CRI Resource Manager
- Demo



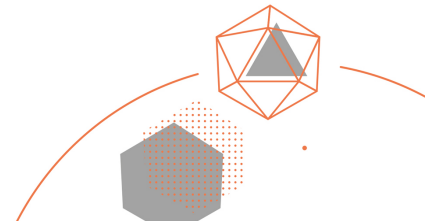
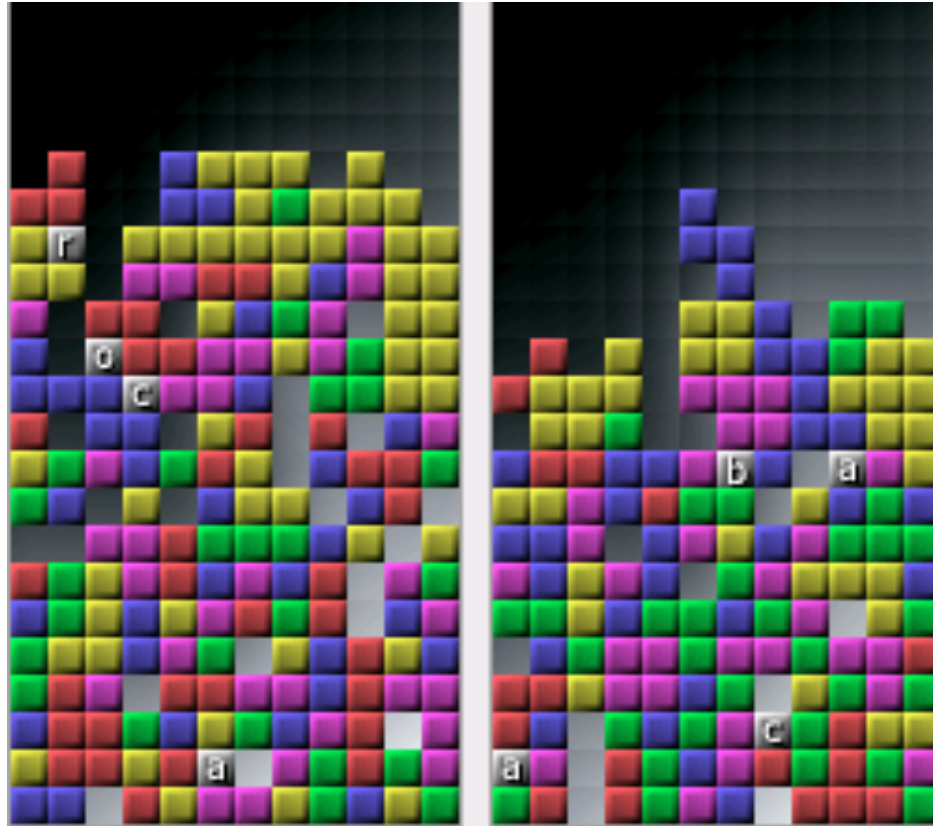
* Other names and brands may be claimed as the property of others.



LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

The “Noisy neighbor” problem

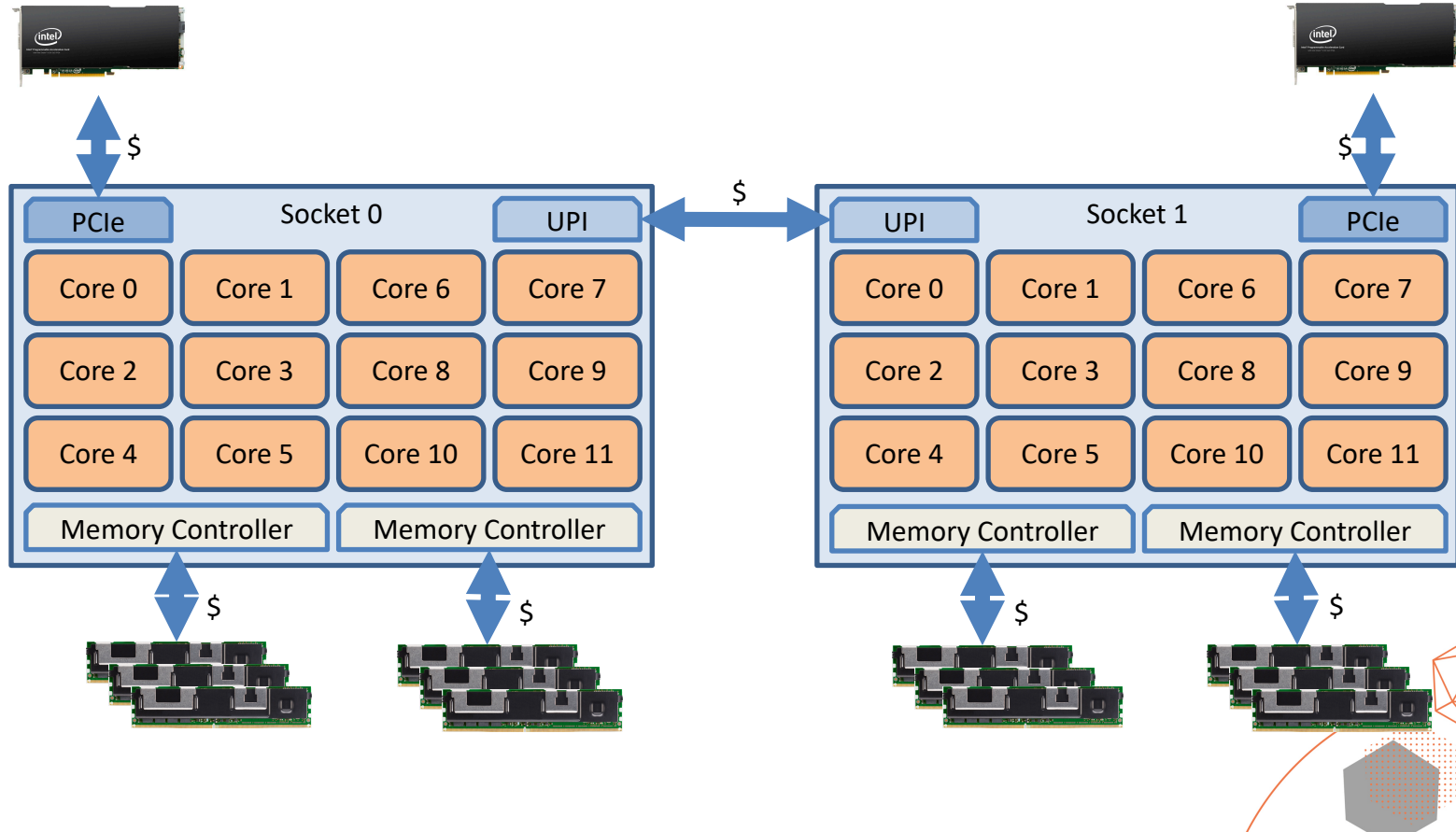




LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

System devices topology

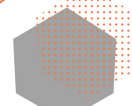
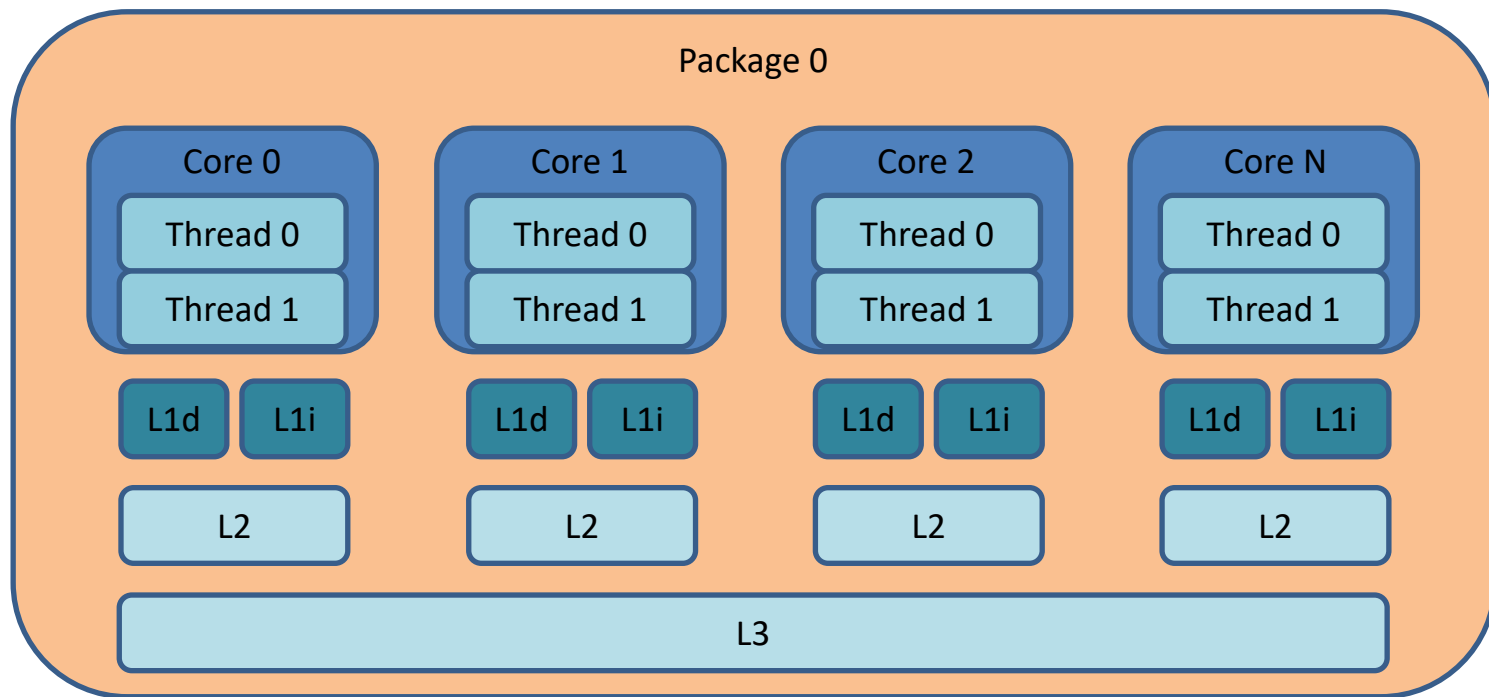




LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

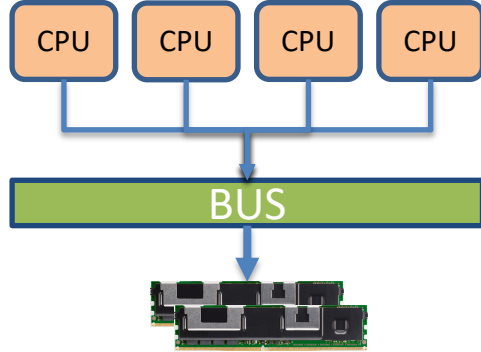
Caches, RDT, MBM





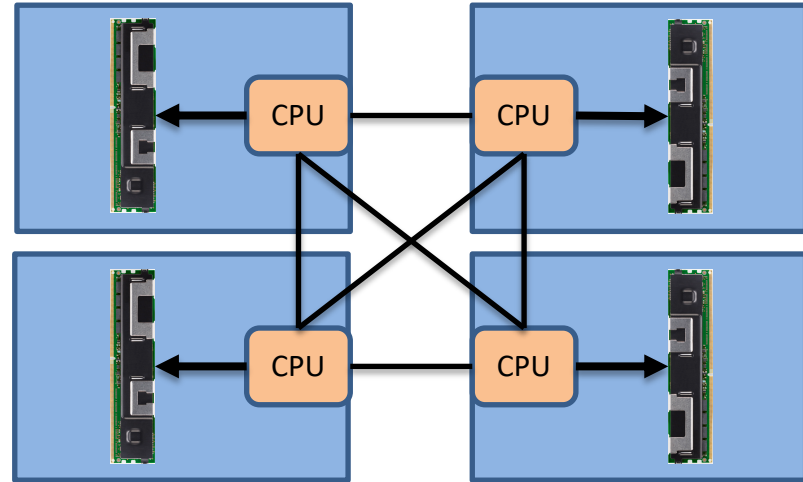
LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG



UMA (aka SMP)
Uniform Memory Access

Memory



NUMA
Non-uniform Memory access

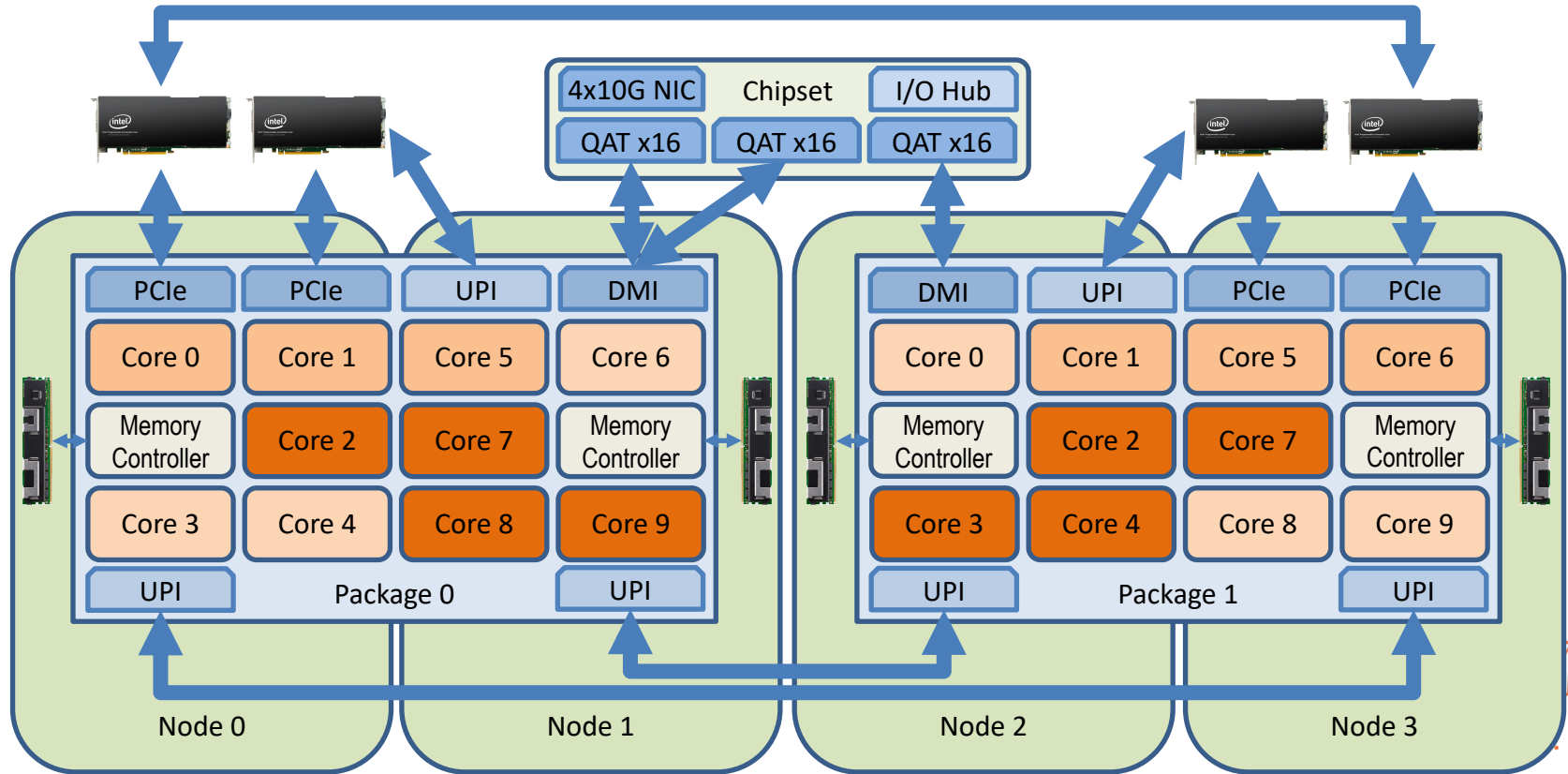




LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

System topology in real world



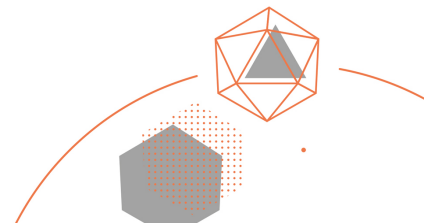


LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Resources in Kubernetes*: Native

- Per container
 - `spec.containers[].resources`
 - requests
 - cpu
 - memory
 - limits
 - cpu
 - memory
- QoS
 - Best Effort
 - Burstable
 - Guaranteed



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



LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Resources in Kubernetes*: Extended

- Extended resources
 - Fully-qualified names
 - Outside of “kubernetes.io” domain
 - Node level
 - Device Plugin managed resources
 - Arbitrary advertised by node capacity



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



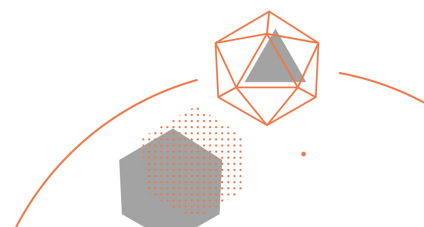
LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Resources in Kubernetes*: Metadata

- Pod's Metadata
 - spec.metadata.labels
 - spec.metadata.annotations

```
apiVersion: v1
kind: Pod
metadata:
  annotations:
    kubernetes.io/ingress-bandwidth: 1M
    kubernetes.io/egress-bandwidth: 1M
...
```



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

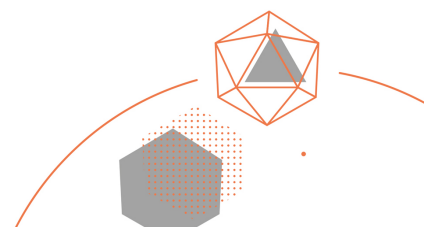
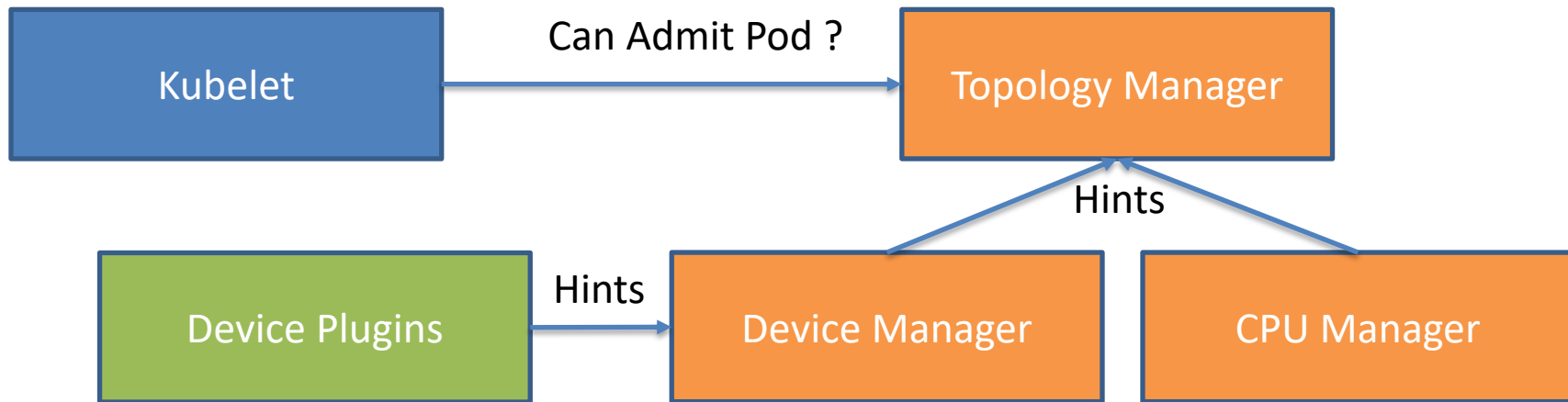


LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Resources in Kubernetes*

- CPU Manager (1.10+) & Topology Manager (1.16+)



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

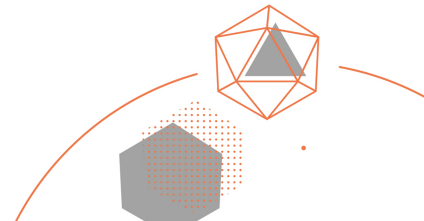


LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

What we can do: CRI

- CPU
 - CFS: Completely Fair Scheduler:
 - period, quota, shares
- Memory
 - Limit, OOM Score
- cgroup
 - cpus, mems





LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

What we can do: OCI

- **runc**
 - blkio: weight
 - CPU realtime period
 - Kernel memory
 - Memory reservation
 - L3 cache schema
 - Memory Bandwidth schema
- **OCI spec**
 - blkio: IOPS / bps throttling
 - HugePage limits
 - Intel® RDT class
 - Hooks

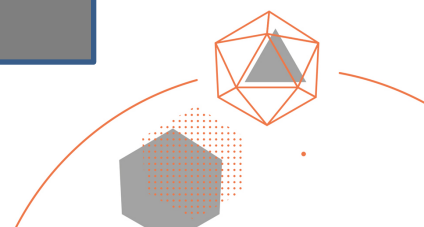
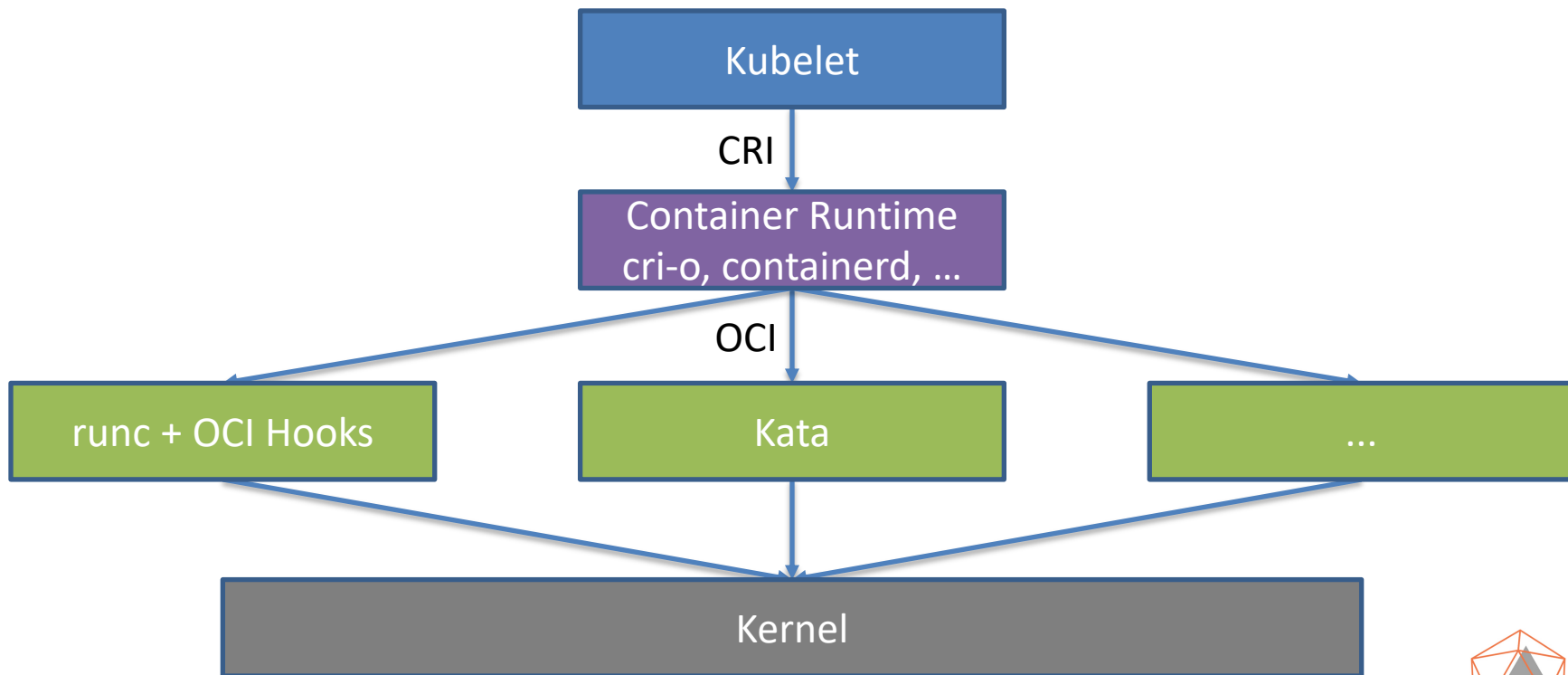




LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Hooking into Kubernetes*



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



LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

CRI Resource Manager

- A new Intel project
 - Container Runtime Interface proxy, sits between CRI Clients and the CRI Runtime
 - Applies (hardware) resource policies to containers by
 - modifying proxied container requests, or
 - generating container update requests, or
 - triggering extra policy-specific actions during request processing
- Benefits
 - Enable easy prototyping of resource allocation policies
 - Instrumentation of CRI interface
 - Drive resource management improvements in upstream Kubernetes
- <http://github.com/intel/cri-resource-manager>

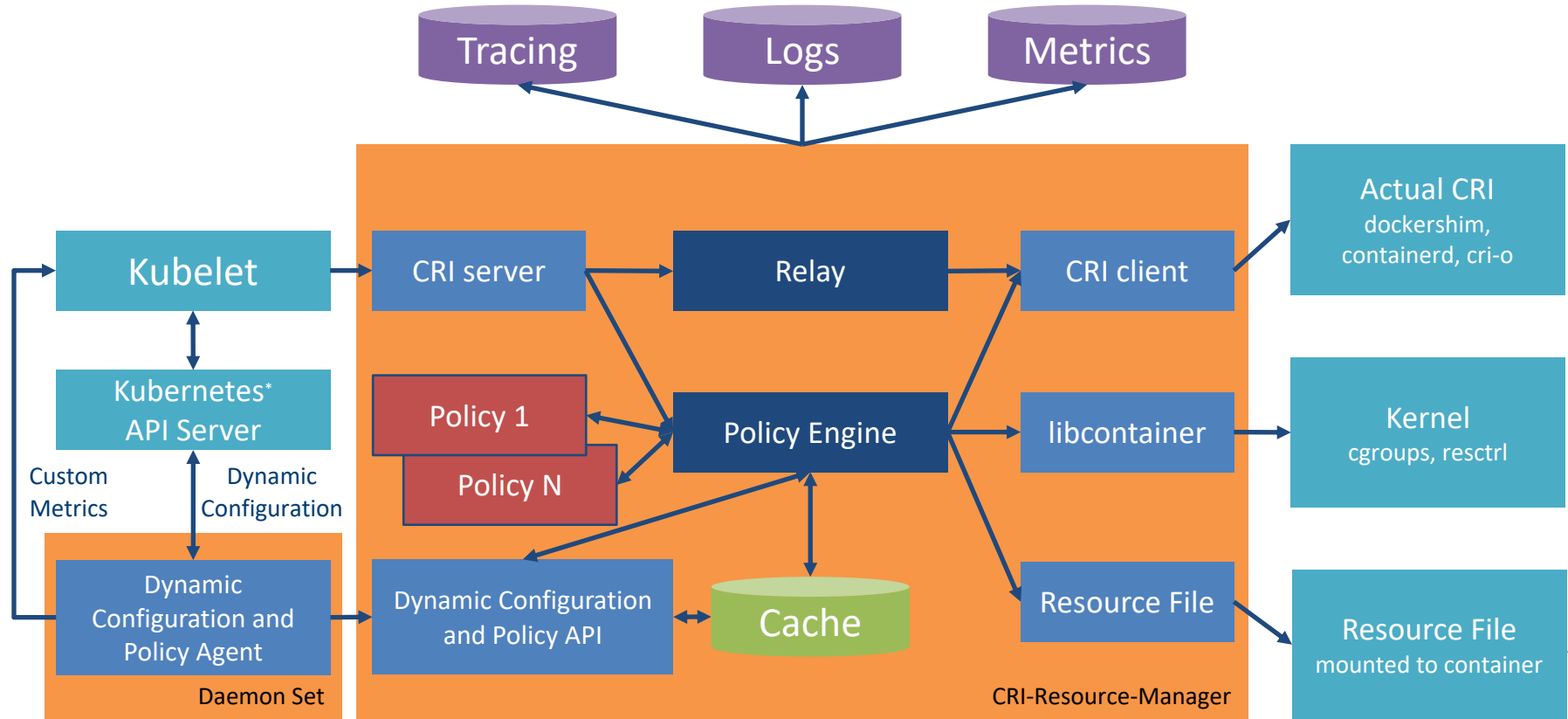




LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

CRI Resource Manager



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

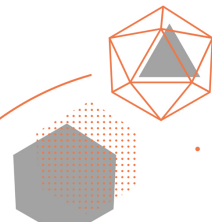


LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

CRI Resource Manager

- What do we have now:
 - Policies:
 - Static
 - Static+
 - Topology-aware
- Work in progress
 - Intel® RDT: Cache and Memory allocation and monitoring
 - Block I/O classification and tuning
 - Better monitoring of resources usage and dynamic rebalancing
 - Dynamic Configuration and Policy APIs





LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

CRI Resource Manager

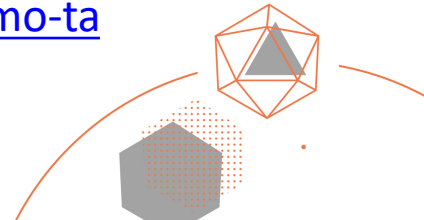


DEMO



<http://bit.ly/cri-r-m-demo-sp>

<http://bit.ly/cri-r-m-demo-ta>

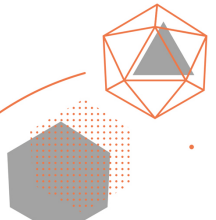




LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Q&A





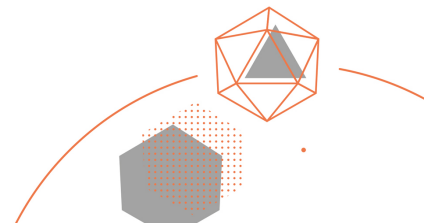
LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

Thank you!

alexander.kanevskiy@intel.com

krisztian.litkey@intel.com



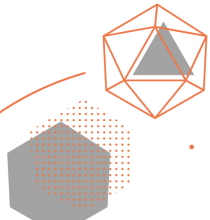


LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

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





LINUX PITER

2019 OCTOBER 4-5
SAINT PETERSBURG

