K8s Labels Everywhere!

Decluttering With Node Profile Discovery

Dave Cremins & Conor Nolan, Intel

May 7th, 2021

Agenda

- Node Feature Discovery - Background and Problem statement
- Complex Node Spec Example
- Conceptual Overview of Node Profiles
- Demo
- Summary

Problem Statement

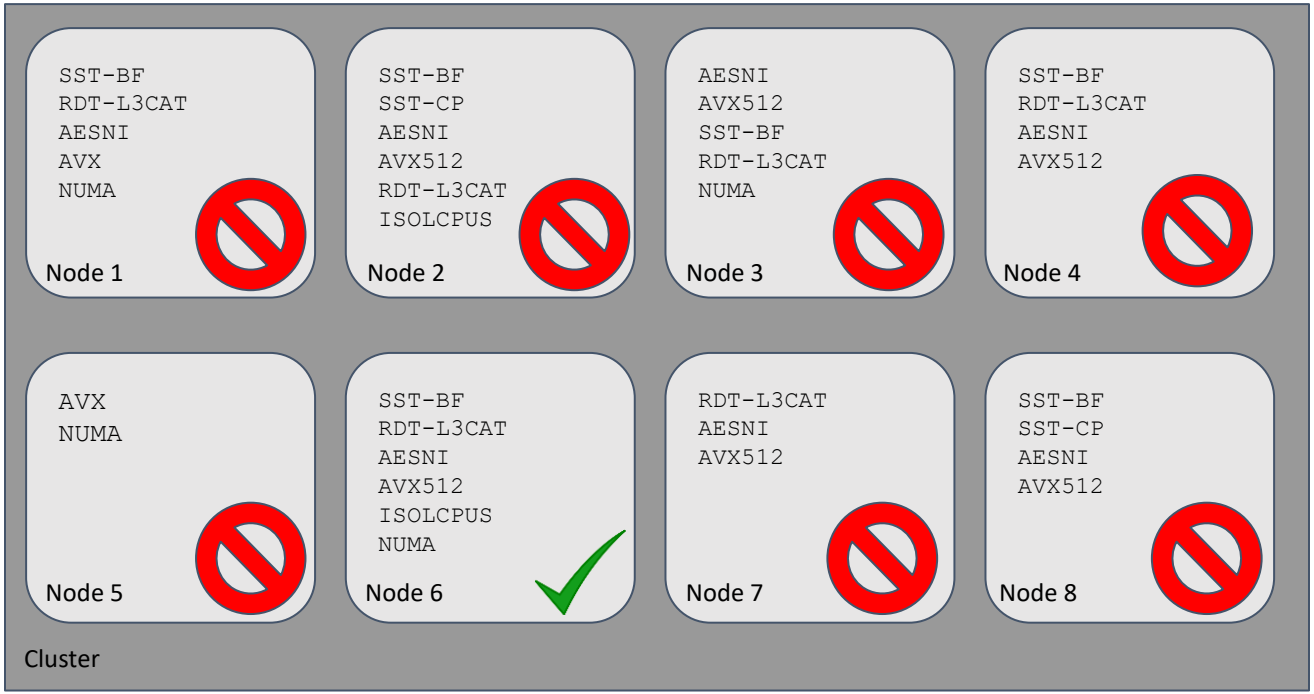
- **Feature Labels**
  - Pod spec bloat
    - Individual feature awareness - tight coupling to platform capabilities
    - Configuration nightmare - extra complexity
    - Scheduling headaches
- **Feature Discovery**
  - Multiple ways to label a node - NFD/node-labeller etc
  - Laundry lists of individual features
  - Number of features will continue to grow - existing pattern
- **Too much individual feature focus**
  - Does not adhere to abstraction layers

We need to pivot towards a ***platform centric*** perspective for the node to be perceived as an optimized accelerator for your specific workload

Node Spec Example

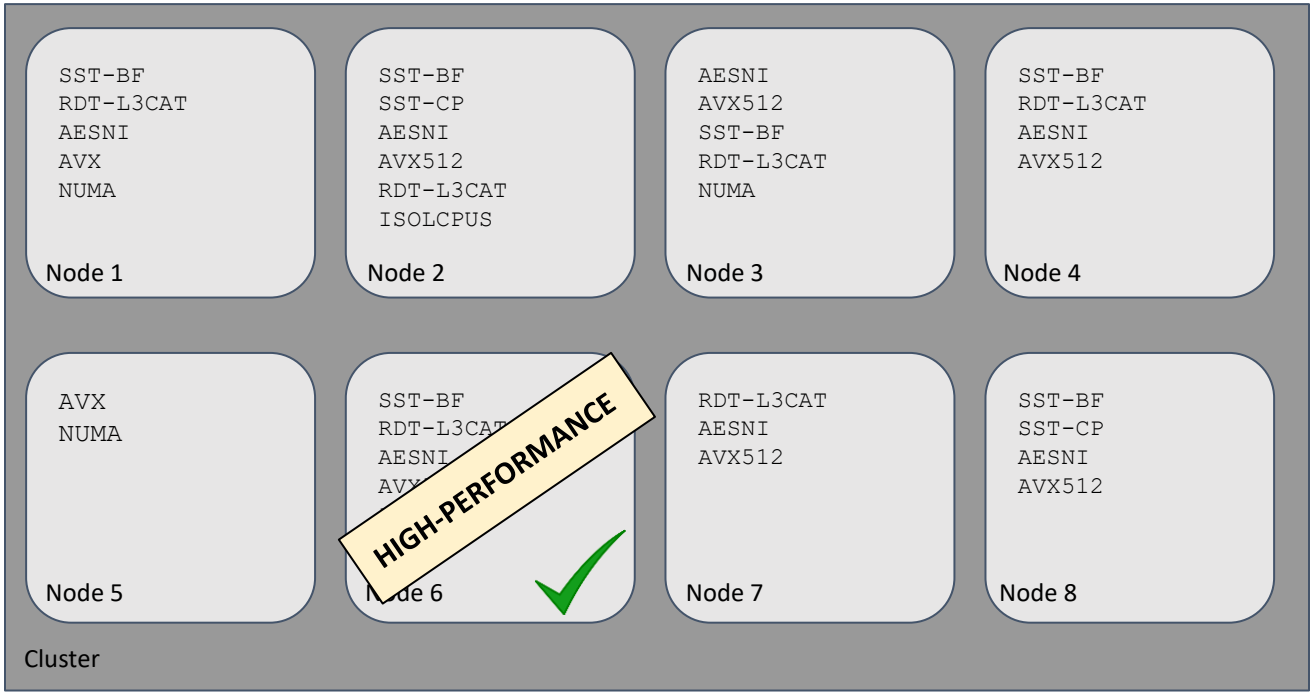

```
node-labeller-feature.node.kubernetes.io/cpu-model-Cascadelake-Server: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Cascadelake-Server-noTSX: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Haswell: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Haswell-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Haswell-noTSX: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Haswell-noTSX-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-IvyBridge: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-IvyBridge-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Nehalem: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Nehalem-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Opteron_G1: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Opteron_G2: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Penryn: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-SandyBridge: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-SandyBridge-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Skylake-Client: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Skylake-Client-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Skylake-Client-noTSX-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Skylake-Server: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Skylake-Server-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Skylake-Server-noTSX-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Westmere: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-Westmere-IBRS: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-kvm32: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-kvm64: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-qemu32: "true"
node-labeller-feature.node.kubernetes.io/cpu-model-qemu64: "true"
node-labeller-feature.node.kubernetes.io/kvm-info-cap-hyperv-base: "true"
node-labeller-feature.node.kubernetes.io/kvm-info-cap-hyperv-frequencies: "true"
node-labeller-feature.node.kubernetes.io/kvm-info-cap-hyperv-ipi: "true"
node-labeller-feature.node.kubernetes.io/kvm-info-cap-hyperv-reenlightenment: "true"
node-labeller-feature.node.kubernetes.io/kvm-info-cap-hyperv-reset: "true"
node-labeller-feature.node.kubernetes.io/kvm-info-cap-hyperv-runtime: "true"
```

Overview




```
apiVersion: v1
Kind: Pod
spec:
 nodeSelector:
  “feature/cpu-aesni”: “true”
  “feature/cpu-avx512”: “true”
  “feature/cpu-sst-bf”: “true”
  “feature/cpu-rdtl3cat”:“true”
  “numa-alignment”: “true”
  “isolcpus”: “true”
```

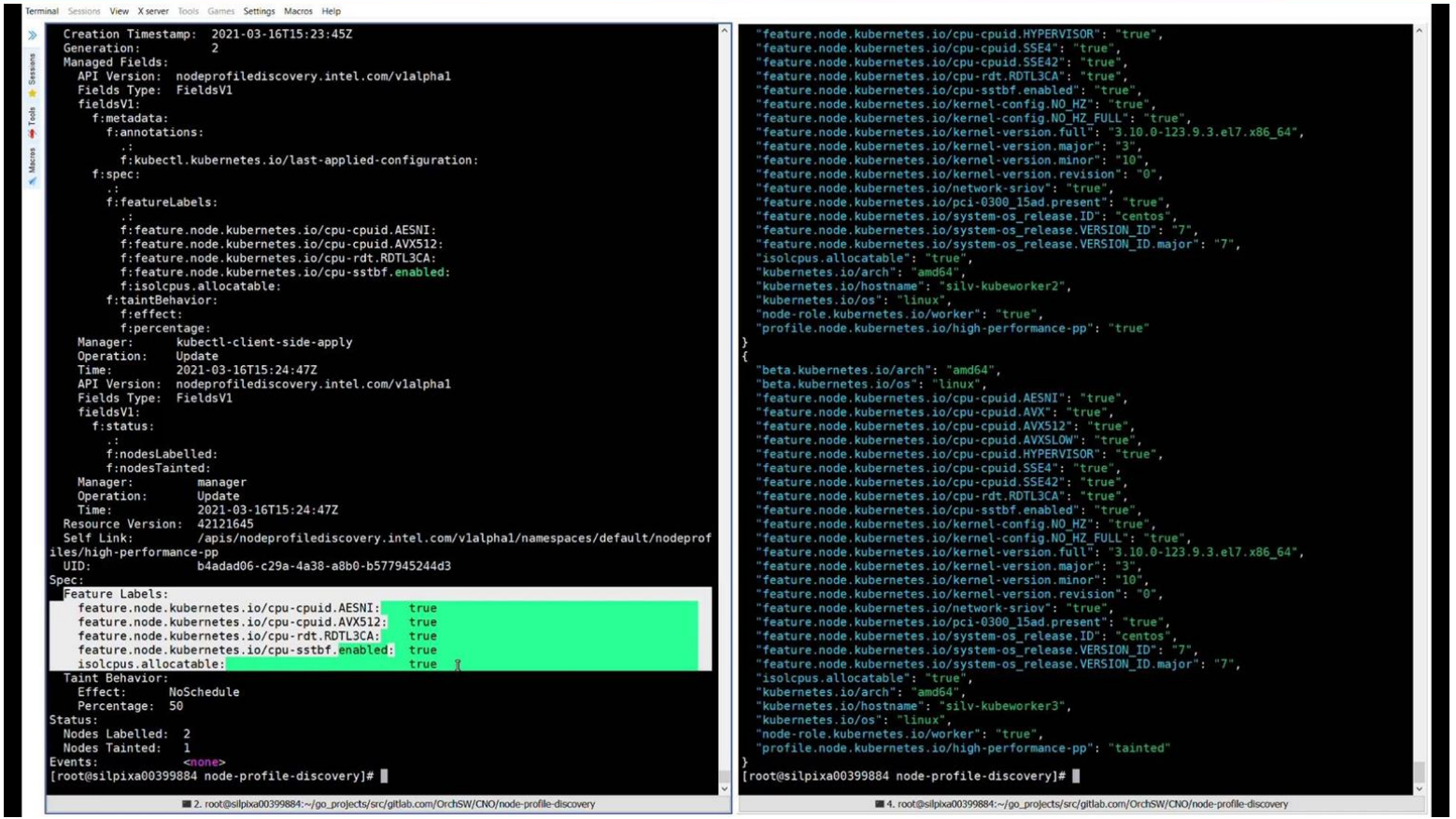
Overview




```
apiVersion: corev1
Kind: Pod
spec:
 nodeSelector:
  “profile/high-priority”: “true”
```

```
apiVersion: v1
Kind: NodeProfile
Name: high-performance
spec:
 featureLabel:
  “feature/cpu-aesni”: “true”
  “feature/cpu-avx512”: “true”
  “feature/cpu-sst-bf”: “true”
  “feature/cpu-rdtl3cat”:“true”
  “numa-alignment”: “true”
  “isolcpus”: “true”
```

Demo


Terminal Sessions View X server Tools Games Settings Macros Help
Creation Timestamp: 2021-03-16T15:23:45Z
Generation: 2
Managed Fields:
API Version: nodeprofilediscovery.intel.com/v1alpha1
Fields Type: FieldsV1
fieldsV1:
f:metadata:
f:annotations:
.:
f:kubectl.kubernetes.io/last-applied-configuration:
f:spec:
.:
f:featureLabels:
.:
f:feature.node.kubernetes.io/cpu-cpuid.AESNI:
f:feature.node.kubernetes.io/cpu-cpuid.AVX512:
f:feature.node.kubernetes.io/cpu-rdt.RDTL3CA:
f:feature.node.kubernetes.io/cpu-sstbf.enabled:
f:isolcpus.allocatable:
f:taintBehavior:
f:effect:
f:percentage:
Manager: kubectl-client-side-apply
Operation: Update
Time: 2021-03-16T15:24:47Z
API Version: nodeprofilediscovery.intel.com/v1alpha1
Fields Type: FieldsV1
fieldsV1:
f:status:
.:
f:nodesLabelled:
f:nodesTainted:
Manager: manager
Operation: Update
Time: 2021-03-16T15:24:47Z
Resource Version: 42121645
Self Link: /apis/nodeprofilediscovery.intel.com/v1alpha1/namespaces/default/nodeprofiles/high-performance-pp
UID: b4adad06-c29a-4a38-a8b0-b577945244d3
Spec:
Feature Labels:
feature.node.kubernetes.io/cpu-cpuid.AESNI: true
feature.node.kubernetes.io/cpu-cpuid.AVX512: true
feature.node.kubernetes.io/cpu-rdt.RDTL3CA: true
feature.node.kubernetes.io/cpu-sstbf.enabled: true
isolcpus.allocatable: true
Taint Behavior:
Effect: NoSchedule
Percentage: 50
Status:
Nodes Labelled: 2
Nodes Tainted: 1
Events: <none>
[root@silpixa00399884 node-profile-discovery]#
2. root@silpixa00399884:~/go_projects/src/gitlab.com/OrchSW/CNO/node-profile-discovery
"feature.node.kubernetes.io/cpu-cpuid.HYPERVISOR": "true",
"feature.node.kubernetes.io/cpu-cpuid.SSE4": "true",
"feature.node.kubernetes.io/cpu-cpuid.SSE42": "true",
"feature.node.kubernetes.io/cpu-rdt.RDTL3CA": "true",
"feature.node.kubernetes.io/cpu-sstbf.enabled": "true",
"feature.node.kubernetes.io/kernel-config.NO_HZ": "true",
"feature.node.kubernetes.io/kernel-config.NO_HZ_FULL": "true",
"feature.node.kubernetes.io/kernel-version.full": "3.10.0-123.9.3.el7.x86_64",
"feature.node.kubernetes.io/kernel-version.major": "3",
"feature.node.kubernetes.io/kernel-version.minor": "10",
"feature.node.kubernetes.io/kernel-version.revision": "0",
"feature.node.kubernetes.io/network-sriov": "true",
"feature.node.kubernetes.io/pci-0300_15ad.present": "true",
"feature.node.kubernetes.io/system-os_release.ID": "centos",
"feature.node.kubernetes.io/system-os_release.VERSION_ID": "7",
"feature.node.kubernetes.io/system-os_release.VERSION_ID.major": "7",
"isolcpus.allocatable": "true",
"kubernetes.io/arch": "amd64",
"kubernetes.io/hostname": "silv-kubeworker2",
"kubernetes.io/os": "linux",
"node-role.kubernetes.io/worker": "true",
"profile.node.kubernetes.io/high-performance-pp": "true"
}
{
"beta.kubernetes.io/arch": "amd64",
"beta.kubernetes.io/os": "linux",
"feature.node.kubernetes.io/cpu-cpuid.AESNI": "true",
"feature.node.kubernetes.io/cpu-cpuid.AVX": "true",
"feature.node.kubernetes.io/cpu-cpuid.AVX512": "true",
"feature.node.kubernetes.io/cpu-cpuid.AVXSLOW": "true",
"feature.node.kubernetes.io/cpu-cpuid.HYPERVISOR": "true",
"feature.node.kubernetes.io/cpu-cpuid.SSE4": "true",
"feature.node.kubernetes.io/cpu-cpuid.SSE42": "true",
"feature.node.kubernetes.io/cpu-rdt.RDTL3CA": "true",
"feature.node.kubernetes.io/cpu-sstbf.enabled": "true",
"feature.node.kubernetes.io/kernel-config.NO_HZ": "true",
"feature.node.kubernetes.io/kernel-config.NO_HZ_FULL": "true",
"feature.node.kubernetes.io/kernel-version.full": "3.10.0-123.9.3.el7.x86_64",
"feature.node.kubernetes.io/kernel-version.major": "3",
"feature.node.kubernetes.io/kernel-version.minor": "10",
"feature.node.kubernetes.io/kernel-version.revision": "0",
"feature.node.kubernetes.io/network-sriov": "true",
"feature.node.kubernetes.io/pci-0300_15ad.present": "true",
"feature.node.kubernetes.io/system-os_release.ID": "centos",
"feature.node.kubernetes.io/system-os_release.VERSION_ID": "7",
"feature.node.kubernetes.io/system-os_release.VERSION_ID.major": "7",
"isolcpus.allocatable": "true",
"kubernetes.io/arch": "amd64",
"kubernetes.io/hostname": "silv-kubeworker3",
"kubernetes.io/os": "linux",
"node-role.kubernetes.io/worker": "true",
"profile.node.kubernetes.io/high-performance-pp": "tainted"
}
[root@silpixa00399884 node-profile-discovery]#
4. root@silpixa00399884:~/go_projects/src/gitlab.com/OrchSW/CNO/node-profile-discovery

Advantages

- **Compound effect of multiple features on a platform vs single profile**
  - Less complex filtering process for the scheduler
  - Simplification of pod specs - easier for vendors to target configured platforms
  - Move towards top down perspective e.g. Power Efficient Packet Processing
  - Align with abstractions prevalent in today's deployments

- **Cluster slicing**
  - Designated nodes for acceleration of performance critical workloads
  - Lightweight and scalable

What's Next?

- **Schema based definition of Profiles**
  - JSON Schema for validation to promote consistency
  - Tie into automation pipeline to validate & create your profile
- **Future of NFD?**
  - Extend capabilities by incorporating direct support for profiles
- **Integration with NFD?**
  - Separate profile concerns with complimentary component that focuses on profile management
- **Integration with Policy Based Control systems**
  - Profile creation based on policies
  - Enforcement of profiles

Notices & Disclaimers

Intel technologies may require enabled hardware, software or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.


KubeCon
CloudNativeCon
Europe 2021
Virtual
Forward Together »