KubeCon

North America 2017

SIG Schedule + Resource Management Working Group Deep Dive

Wednesday, December 6 • 3:40pm - 5:00pm

Agenda

- Introductions
- Roadmap Walkthrough
- 2017 RMWG Themes
- Graduating features … Beta -> GA
- Resource Classes and Resource API
- Device Plugins
- Dynamic Resource Binding
- Group Discussion Topics

RMWG 2017 Themes

Avoid performance regressions when running on Containers vs VMs

- Better CPU Isolation - Static CPU Pinning: Design
- ~~Better CPU Isolation - Automated CPU allocation~~
- Support for HugePages: Design & Tracker
- NUMA: Tracker

RMWG 2017 Themes

Support for Hardware Accelerators

- [Device Plugin](#)
- [Add monitoring support for hardware accelerators](#)
- [Add ExtendedResourceToleration admission controller](#)

Graduating Features

- Device Plugin to Beta to GA
- Hugepages to Beta (on track v1.10), to GA
- CPU Manager to Beta (on track v1.10) (alpha to beta [Tracker](#))
  - CPU assignments are durable across Kubelet restarts.
  - We run the CPU manager node e2e tests periodically in CI ([testgrid link](#)). Next step is to make it PR-gating.

Resource API

- Overview and justification
- [Resource Classes Design Proposal Draft](#)
- How to bring a new resource type (logical/real)
- Improved compute resource APIs to handle device metadata
- Handling Resource Quota, LimitRange
- Handling cluster-level resources (Software Licenses/TPU)
- Discovery … what resources are available in this cluster?

Device Plugin Status

- Alpha in 1.8
- Only one implementation (NVIDIA)
- Ongoing work on Solarflare
- Intel/FPGA prototype?
- Infiniband? (Talk with Mellanox)

Device Plugin Roadmap

- [Device Plugin Work Tracker](#) (Google)
- [Device Plugin 1.9 Roadmap](#) (NVIDIA)
- [FPGAs and Kubernetes Device Plugins](#) (Intel)

Dynamic Resource Binding

- Persistent volumes and GPUs are resources that are being bound dynamically (after finding a node for a pod).
- More such resources may come in the future.
- Current implementations are not ideal.
- We should discuss possible ways of address dynamic resource binding and its integration with Scheduler.
- JE: scheduler extensions, node extended fitness
  - Logging daemon, GPU plugins, kernel cmdline, tuned

Group Discussion Topics

- Areas of overlap between SIG/wg
- How is the two-level scheduling working out?

Group Discussion Topics

- What use-cases are we trying to cover in the coming year that would generate features from each other?

Group Discussion Topics

- What's good and not good about Resource Management in Kubernetes? What are the most common pain points?

Group Discussion Topics

- How replaceable should the default scheduler be?
- Seeking testers for Priority and Preemption

Group Discussion Topics

- How does anyone keep track of anything on Github?

Example

- Example