Containers At Scale
At Google, the Google Cloud Platform and Beyond

Joe Beda – jbeda@google.com – @jbeda – google.com/+JoeBeda
Senior Staff Software Engineer, Google Cloud Platform
GlueCon - May 22, 2014
Google and Containers

**Everything** at Google runs in a container.

Internal usage:
- Resource isolation and predictability
- Quality of Services
  - batch vs. latency sensitive serving
- Overcommitment (not for GCE)
- Resource Accounting

We start over 2 billion containers per week.
Google and Containers

- Limited Isolation: 2004
- Released CGroups: 2006
- Released LMCTFY: 2013
- Using Namespaces: 2014
Let Me Contain That For You

github.com/google/lmctfy

- Replacement for LXC
- Integrating with Docker
  (https://github.com/dotcloud/docker/pull/4891)
- Separates policy from enforcement; buffers users from cgroups APIs
- Programmable API and CLI
The Managed Container Stack at Google

Managed Base OS

Node Container Manager
  • Common services: log rotation, watchdog restarting

Containers:
  • System container for shared daemons. Statically defined.
  • Dynamically scheduled containers

Cluster Scheduler
  • Schedules work (tasks) onto nodes
  • Work specified based on intents
  • Surfaces data about running tasks, restarts, etc.
Declarative Over Imperative

Imperative:
"Start this container on that server"

Declarative:
"Run 100 copies of this container with a target of <= 2 tasks down at any time"

Pros:
• Repeatable
• "Set it and forget it"
• Eventually consistent
• Easily updatable

Con:
• Tracing action/reaction can be difficult.
  "I made a change, is it done?"
Packaging Containers

Google:
- Host bind mounts
- Binary and deps built together
- Interfaces to Container Manager:
  Standard locations for logs, API

Docker Image and environment:
- More hermetic. Entire chroot is explicitly included.
- Less guaranteed file structure.
- Leverages OS distributions and package managers.
Warning
What follows is an early look at how we are integrating containers into the Google Cloud Platform.
Container Node Reference Architecture

Container Manifest
manifest.yaml

Open Source Node Container Manager

Docker

Start/Kill
Monitor
Container Manifest

Declarative description of a set of containers and required resources

A YAML File

"Scheduling unit": must be scheduled on a single node
  • Unit for sharing data, IPC, cpu/disk/ram limits, networking
  • Share fate. If the host machine goes down, all containers go down together.
Container Manifest Example

```yaml
version: v1beta1
containers:

- name: data-loader
  image: my-org/data-loader
  volumeMounts:
    - name: data
      path: /mnt/data

- name: server
  image: my-org/data-server
  ports:
    - name: www
      containerPort: 80
  volumeMounts:
    - name: data
      path: /mnt/data

volumes:
- name: data
```

**Diagram:**
- Data Loader
- Data Server
- Disk Volume
Reference Node Container Manager

Consumes a manifest and makes it happen. Layers on top of Docker.


Now:
- Starts containers when run start up
- Keeps containers running in face of failures

Soon:
- Dynamic update manifests
- Expose metrics, logs, history
Container VMs in Google Compute Engine

Container Manifest
manifest.yaml

Open Source Node Container Manager

Start/Kill
Monitor

Docker
Container VMs in Google Compute Engine

Cloud VMs optimized for Containers

Easiest way to use Container Manifests is on the Google Cloud Platform:
- Image preinstalled with: Docker, Node Container Manager
- Loads Container Manifest at start time
- [Soon] Integrate with UI, logging
- [Soon] Basic building block for dynamic systems

Also used by Managed VM driven by Google App Engine.
Using Container VMs

```
version: v1beta1
containers:
  - name: my-container
    image: my-org/my-server

$ gcloud compute instances create my-container \
  --image=project/google-containers/global/images/container-vm-v20140522 \
  --metadata-from-file google-container-manifest=my-containers.yaml
```
Next Steps

Launch a container VM:
developers.google.com/compute/docs/containers

Talk to Googlers:
Here at GlueCon
DockerCon June 9-10, Google I/O June 25-26

Send us comments/ideas:
Discussion group: groups.google.com/forum/#!forum/google-containers
IRC:
#google-containers on irc.freenode.net
Stack Overflow:
Use "google-compute-engine" and "docker" tags
Resources

LMCTFY:
  Feb 2014 SF Production Eng Meetup: http://goo.gl/6nbZsX
  Linux Plumbers Conference 2013: http://goo.gl/xqmDTp

Omega Cluster Management:

The Google Build system:
  DevOps talk from Cloud Platform Live 2014: http://goo.gl/jmzqwQ

MPM Package Management:
  Presentation from USENIX UCMS'13: http://goo.gl/aP9Rf6
Thanks!

Joe Beda
jbeda@google.com
google.com/+JoeBeda
@jbeda