LXD
Live Migration of Linux Containers

Tycho Andersen, Canonical Ltd.

tycho.andersen@canonical.com
http://tycho.ws
Sorry, no demos today :(
Who is this guy?

➔ LXD
➔ LXC
➔ CRIU
➔ Kernel
The history of LXC

Over 6 years of Linux system containers
LXD: the container lighter-visor

What it **IS**

**Simple**

*Clean command line interface, simple REST API and clear terminology.*

**Fast**

*No virtualization overhead so as fast as bare metal.*

**Secure**

*Safe by default. Combines all available kernel security features.*

**Scalable**

*From a single container on a developer’s laptop to thousands of containers per host in a datacenter.*
LXD: the container lighter-visor

What it ISN’T

Another virtualization technology
LXD tries to offer as similar a user experience as that of a virtual machine but it doesn’t itself virtualize anything, you always get access to the real hardware and the real native performance.

A fork of LXC
LXD uses LXC’s API to manage the containers behind the scene.

Another application container manager
LXD only cares about full system containers and doesn’t care about what runs inside the container.
Hypervisor-y things
LXD: the container lighter-visor

- nova-compute-lxd
- lxc (command line tool)
- your own client/script?

LXD REST API

Host A
- LXD
- LXC
- Linux kernel

Host B
- LXD
- LXC
- Linux kernel

Host C
- LXD
- LXC
- Linux kernel

Host D
- LXD
- LXC
- Linux kernel

Host ...
- LXD
- LXC
- Linux kernel
Hypervisor-y things
lxc move host1:c1 host2:
lxc move host1:c1 host2:

➔ host1 negotiates three “channels” with host2
  ◆ control
  ◆ filesystem
  ◆ container process state

➔ Using a tool called CRIU for process state
The history of CRIU

Five years of checkpointing!

2006
Attempts to merge OpenVZ in-kernel upstream

... 2011  2012  2013  2014  2015

First kernel patches merged

CRIU 0.1

LXC gets support for C/R via CRIU

libcontainer gets support for C/R via CRIU

2011
First discussion on lkml about C/R from Userspace

C/R in OpenVZ Kernel

2012
Memory page tracking patches merged upstream
What’s the catch?

“A note on this: this is a project by various mad Russians to perform c/r mainly from userspace, with various oddball helper code added into the kernel where the need is demonstrated… However I'm less confident than the developers that it will all eventually work!”

- Linus Torvalds (kernel commit 09946950)
What’s the catch?

“This is not an enterprise feature. It's a promise one cannot keep. We will not add code to systemd that works often but not always, and CRIU is certainly of that kind.”

- Lennart Pottering (systemd-devel, 2015)
Always playing catch-up
Plug-ins needed for custom /dev
Not all kernel features supported
Security!?!111?1
Security

Secure

Migratable
Security

Secure & Migratable
Security

➔ cgroups
➔ apparmor, selinux, etc.
➔ seccomp (STRICT, FILTER)
➔ user namespaces
Correct and Fast

Pick two
Making Migration Fast

→ Three channels
  ◆ control
  ◆ filesystem specific
  ◆ memory state specific

→ Filesystems:
  ◆ btrfs, LVM, ZFS, (swift, nfs?), etc.
  ◆ rsync between incompatible hosts

→ Memory state:
  ◆ Stop the world
  ◆ Iterative incremental transfer (via p.haul)
Administrivia

➔ **LXD**
- Current stable release 0.16
- 1.0 targeted for February 2016
- Two week release cadence
- (at least) One more release before Wily
- [https://linuxcontainers.org](https://linuxcontainers.org)
- [https://github.com/lxc/lxd](https://github.com/lxc/lxd)

➔ **CRIU**
- Current stable release 1.6.2
- Three month release cycle
- [http://criu.org](http://criu.org)
- [https://github.com/xemul/criu](https://github.com/xemul/criu)
Questions?