



# **Introduction to OS Virtualization, Containers and OpenVZ**

Scott Dowdle <[dowdle@montanalinux.org](mailto:dowdle@montanalinux.org)>

# Who am I? Scott Dowdle

- I've been using Linux since 1995, Red Hat since 1996, and have been a sysadmin since 1998
- I'm NOT a kernel developer NOR programmer
- I've been active in the OpenVZ community ~4 years
  - #openvz channel on irc.freenode.net
- I helped staff the OpenVZ booth at LinuxWorld (07/08)
- I've done lots of OpenVZ presentations
- I'm active in Montana LUGs and run MontanaLinux.org
- One more time: [dowdle@montanalinux.org](mailto:dowdle@montanalinux.org)

# Preface

Linux native containers (LXC) is a part of the mainline kernel and is getting better with each kernel release...

...but it will be a few years before LXC reaches the functionality of much older, out-of-tree projects such as OpenVZ and Linux-VServer.

You DO NOT have to wait on LXC and can start using containers now. I've been using them for over 4 years now.

# Overview

- What is OS Virtualization and why are containers important?
- What virtualization use cases are containers especially well suited for?
- What virtualization use cases are containers not well suited for?
- A brief history of OpenVZ will be provided along with installation instructions and a demo of container creation, configuration and live migration.

# What is OS Virtualization?

- a.k.a. Containers
  - It is **NOT** hardware / machine virtualization
  - It is a **single kernel** on the host node
  - It is for **Linux on Linux** only virtualization
  - It offers **strongly isolated** environments
  - It very tunable with **dynamic resources**
  - It is a Linux distribution run as **grouped processes**
- Templates for a number of Linux distributions
- Implementations include OpenVZ and Linux-Vserver
- Coming to mainline kernel as LXC (Linux Native Containers)

# What are Containers NOT good at?

- Things that require a custom kernel
- Things that require multiple kernels
- Things that require custom kernel modules
- Things that require direct access to the hardware
- Huge Java server apps\*
- Running other virtualization products
- Multimedia / gaming

# What are Containers good at?

- Isolating server applications (better than chroot)
- Common server applications
- Offering higher density than hw virt
- Offering better scalability than hw virt
- IT Hosting and provisioning
- Educational environments
- Anything that doesn't fall into the NOT category

# OpenVZ History

- First Virtuozzo release by SWsoft (Moscow, Russia) in 2001
- Microsoft Windows version of Virtuozzo in 2005
- OpenVZ Project created in 2005
- Checkpointing and Migration added in Nov. 2006
- Parallels, Inc (Renton, WA) was an SWsoft company until January of 2008 when they merged. Known since as Parallels.

# OpenVZ Components

- Kernel
  - Container features added to various subsystems
- Userspace Utilities
  - vzctl, vzquota, vzmigrate
- OS Templates
  - Build System (vzpkg is suffering from bit rot)
  - Official pre-created
  - Community contributed pre-created
  - Manually created, Physical to Virtual

# OpenVZ Installation

- Install kernel with OpenVZ support
- Modify `/etc/sysctl.conf`
- Reboot to new kernel
- Install utility programs
  - `vzctl` and `vzquota`
- Start `vz` service

# OpenVZ Filesystem Layout

- Configuration

`/etc/vz/, /etc/vz/vz.conf, /etc/vz/conf/*,  
/etc/vz/dists/*`

- Containers

`/vz/private/*, /vz/root/*`

- OS Templates

`/vz/template/*`

- Quota

`/var/vzquota/*`

# Creating a Container

Make sure bash-completion is installed if you want vzctl tab completion

```
vzctl create {CTID} --ostemplate {filename} \  
  --conf {filename}
```

```
vzctl set {CTID} --name {name} \  
  --ipadd {ip address} \  
  --hostname {fqdn} \  
  --nameserver {first second} \  
  --diskspace {nG:nG} --save
```

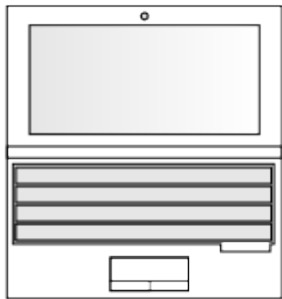
# vzctl Parameters

- **create:** `--config, --hostname, --ipadd, --ostemplate, --private, --root`
- **set:** `--applyconfig, --dgramrcvbuf, --ipdel, --netdev_add, --numiptent, --oomguarpages, --root, --userpasswd, --capability, --disabled, --iptables, --netdev_del, --numothersock, --othersockbuf, --save, --vmguarpages, --cpulimit, --diskinodes, --kmemsize, --netif_add, --numproc, --physpages, --searchdomain, --cpuunits, --diskspace, --lockedpages, --netif_del, --numpty, --private, --setmode, --dcachesize, --hostname, --meminfo, --noatime, --numsiginfo, --privvmpages, --shmpages, --devices, --ioprio, --name, --numfile, --numtcpsock, --quotatime, --tcprcvbuf, --devnodes, --ipadd, --nameserver, --numflock, --onboot, --quotauidlimit, --tcpsndbuf`

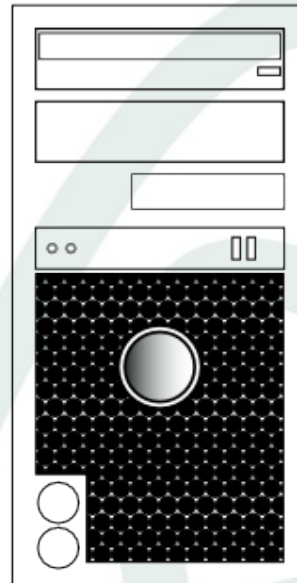
# Hands on Demo 1

- View of a container from the host node
  - Multiple init processes with children
- View from inside of the container
  - Same processes with different PIDs
- Live migration
  - No shared storage required
  - Container continues to run until ready to make the switch
  - Offline migration available

# OpenVZ Demo Setup



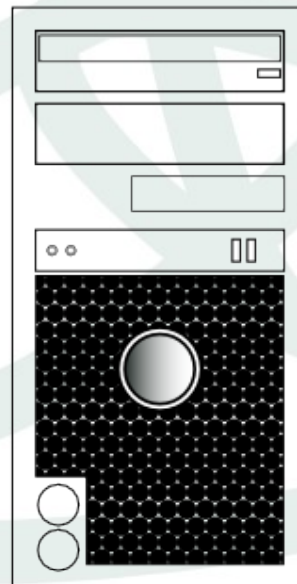
Laptop  
192.168.0.3



Hostnode1  
192.168.0.1  
CentOS 5.3 x86\_64  
OpenVZ kernel

Container1 - 192.168.0.11  
Fedora 11 i386 GUI

Container2 - 192.168.0.12  
CentOS 5.3 x86\_64



Hostnode2  
192.168.0.2  
CentOS 5.3 x86\_64  
Stock kernel

# Kernel

Mainline cycle is every 3 months with brief support period...  
want something long maintained and updated

- Stable Branches
  - RHEL4 2.6.9, RHEL5 2.6.18, and vanilla 2.6.18
  - ~~Lacks ppp in container support~~
- Development Branches
  - 2.6.24 (Ubuntu), 2.6.26 (Debian), 2.6.27 (long maintained)
  - Missing --cpulimits, use --cpuunits (it's better)
- Retired Branches
  - 2.6.8, 2.6.15, FC5, 2.6.16, SLES10 2.6.16, 2.6.20, 2.6.22
- Which kernel to use?
  - Next stable might be 2.6.27 or whatever RHEL6 ships with

# Userspace Utilities

- vzctl
  - create, destroy, start, stop, restart, enter, exec, set
  - chkpnt, restore
- Features coming in future vzctl releases
  - swappages UBC parameter, allows visible swap inside of a container and is mostly useful for installing Oracle
  - Disk image files?
- vzctl for mainline kernel?
  - Development was started but has mostly stalled
  - LXC thus far mostly useful for developers only
  - LXC namespace features of interest to HPC

# OS Templates

- **Build System**
  - vzpkg is broken and unmaintained
  - vzpkg2 & pkgcache by Robert Nelson
- **Official pre-created OS Templates**
  - Built using proprietary Virtuozzo system
  - Better quality
  - Built more frequently
- **Contributed pre-created OS Templates**
- **Wiki good source of “from scratch” and “physical to virtual” recipes**

# Kernel Contribution Stats

Also, here's the list of top10 contributors to the Linux 2.6.25.

Top changeset contributors by employer

(None)	1188	(9.3%)
Red Hat	1181	(9.3%)
Novell	817	(6.4%)
IBM	703	(5.5%)
Intel	472	(3.7%)
Bartłomiej Zolnierkiewicz	307	(2.4%)
Parallels	278	(2.2%) <---
Oracle	255	(2.0%)
bunk@kernel.org	227	(1.8%)
(Academia)	225	(1.8%)

Pavel Emelyanov has made it to top10 of developers.

Developers with the most changesets

Bartłomiej Zolnierkiewicz	307	(2.4%)
Adrian Bunk	234	(1.8%)
Patrick McHardy	225	(1.8%)
Ingo Molnar	213	(1.7%)
Paul Mundt	207	(1.6%)
Greg Kroah-Hartman	172	(1.4%)
Thomas Gleixner	166	(1.3%)
Jesper Nilsson	166	(1.3%)
Pavel Emelyanov	160	(1.3%) <---
Harvey Harrison	150	(1.2%)

# How many containers?

- In December I wanted to see how many containers I could create on a single host node
- Physical hardware
  - HP Proliant DL380 G5
  - Dual, quad-core Xeon with 32GB of RAM
- Outcome
  - 638 containers with reasonable resources
  - ~16,080 processes
  - Private IP addresses
  - Ran 600 containers for two weeks

<http://www.montanalinux.org/openvz-experiment.html>

# Resources

<http://www.openvz.org/>

<http://www.linux-vserver.org/>

<http://www.parallels.com/>

<http://www.montanalinux.org/>

[dowdle@montanalinux.org](mailto:dowdle@montanalinux.org)