Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Linux Containers (LXC)

Linux User Group Meeting #2

Wednesday, 16th June 2014 17:00 – 18:00

BFH – Room HALL 217 (Hallerstrasse, Bern)

Daniel Baumann <daniel.baumann@bfh.ch>

IT System Engineer, Infrastructure Team

Overview

Introduction

- BFH News

Linux Containers (LXC)

- Theory
- Practice

Aperitif

BFH News

BFH News

https://otrs.bfh.ch

- Sub-Queue: Infrastructure Services / Linux

https://lists.bfh.ch

- Use our mailinglists, specifically bfh-linux-users@lists.bfh.ch

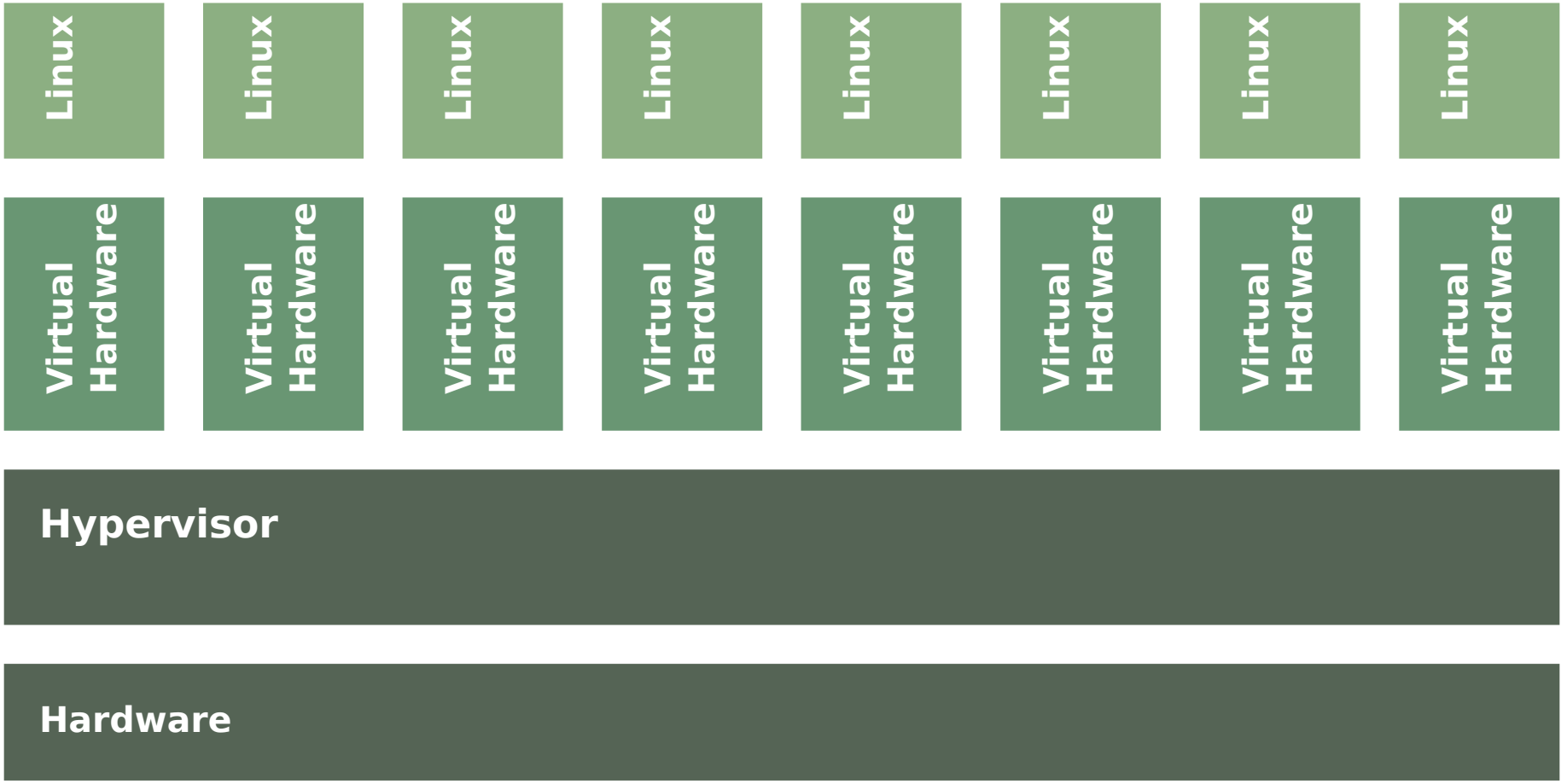
https://linux.bfh.ch/services/linux/server

- Linux as a Service (PaaS/SaaS) based on LXC

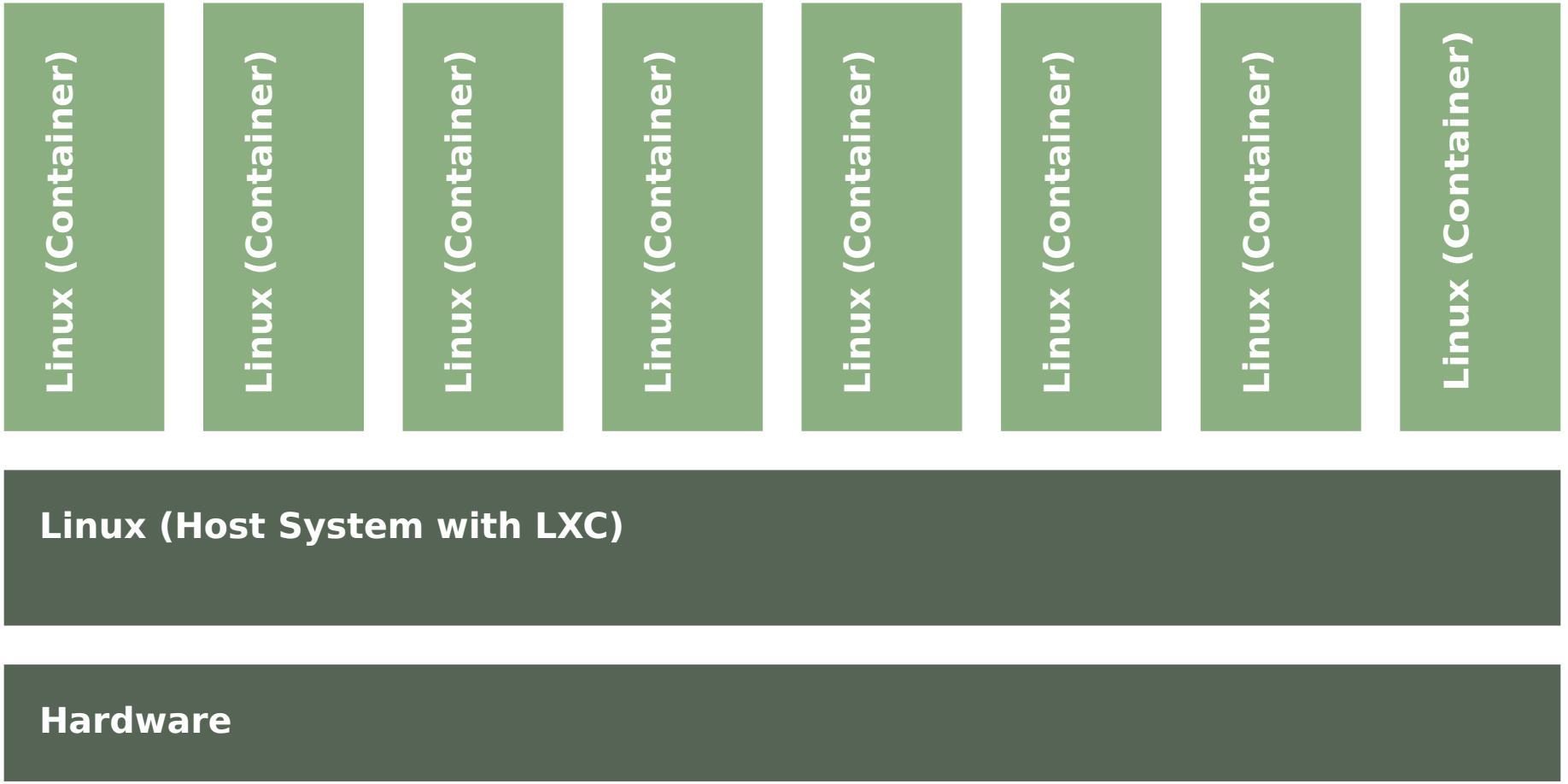
Topics Survey and Date Polls for future meetings coming soon...

Linux Container (LXC) Theory

Virtualization



Isolation


Linux (Container)
Linux (Container)
Linux (Container)
Linux (Container)
Linux (Container)
Linux (Container)
Linux (Container)
Linux (Container)
Linux (Host System with LXC)
Hardware

Linux Container (LXC) Theory: Leight Weight „Virtualization“ aka Isolation

- Nothing excatly new...
  - IBM LPAR
  - FreeBSD Jails
  - Solaris Zones
  - Linux chroot, Linux VServer, Linux OpenVZ, ...
- ...or not really the same...
  - Full/Para-Virtualization: KVM, Xen
  - Proprietary: VMware
- ...now, it is done right this time (aka Generic Ressource Management):
  - Linux Containers (LXC)
  - Docker
  - libvirt-lxc, lmctfy, linux-user-chroot, systemd-nspawn ...

Linux Container (LXC) Theory: Linux Ressource Management (1)

- Process Attributes
  - Process ID (PID)
  - Parent Process ID (PPID)
  - Nice number
  - Terminal (TTY)
  - Real and Effective User (RUID/EUID)
  - Real and effective Group owner (RGID/EGID)

Linux Container (LXC) Theory: Linux Ressource Management (2)

- Process Permission Checks (aka Capabilities)
  - Good Examples: CAP_MAC_OVERRIDE, CAP_MKNOD, CAP_SYS_TIME
  - Otherwise: CAP HELL :(
  - see capabilities(7)

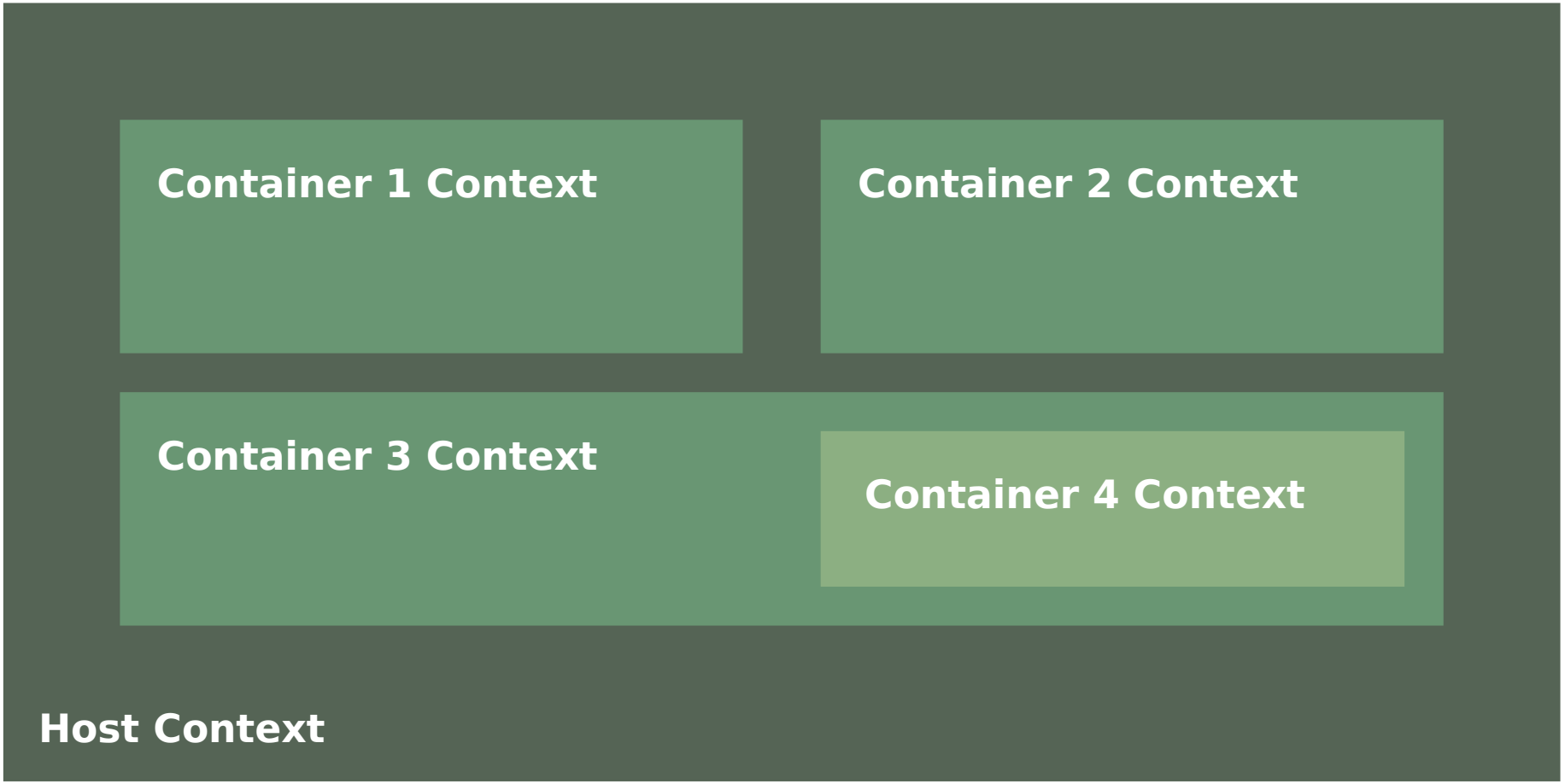
Linux Container (LXC) Theory: Linux Ressource Management (3)

- Namespaces (NS)
  - providing isolation through new syscall flags (6 for clone) and new 3 syscalls (share/unshare/setns)
  - ips (System V IPC)
  - mnt (Mountpoints, Filesystems)
  - net (Networking)
  - pid (Processes)
  - user (UIDs/GIDs)
  - uts (Hostname)
  - In the Queue: devices, time

Linux Container (LXC) Theory: Linux Ressource Management (4)

- Control Groups (CGroups)
  - providing hierarchical ressource management through grouping of processes (ressource controllers, /sys/fs/cgroup)
  - blkio (block devices)
  - cpu (CPU Access), cpuacct (CPU Accounting), and cpuset (CPU Assignment)
  - devices
  - freezer
  - memory
  - swap
  - net_cls (Network Class identifier)
  - net_prio (Network Priority)
  - perf_event, hugetlb, debug

Control Groups



Linux Container (LXC) Practice

Linux Container (LXC) Practice: Installation

- Simple: sudo apt-get install lxc lxc-stuff debootstrap bridge-utils
- Make sure:
  - most recent Debian version of LXC on the host system
  - most recent version of live-debconfig in /usr/share/lxc/packages (unless building sid)
- Note:
  - Bumpy road ahead (live-debconfig migrates backend to puppet)
  - official Debian lxc-server images starting with jessie

Linux Container (LXC) Practice: Configuration (1)

Optionally add Boot parameter for Memory/Swap Ressource Controller

- cgroup_enable=memory
- swapaccount=1

Mount /sys/fs/cgroups (on non-systemd systems; #601757)

- mount cgroup -t cgroup /sys/fs/cgroup

Linux Container (LXC) Practice: Configuration (2)

- **Choose LXC Networking Types**
  - empty
  - veth (prefered; bridge, with or without NAT)
  - macvlan (private, vepa, or bridge)
  - vlan
  - phys

Enable IP Forward

- echo "net.ipv4.ip_forward = 1" > /etc/sysctl.d/ip_foward.conf
- sysctl -p

Linux Container (LXC) Practice: Terminology

- Contexts: Host System and Container
- LXC Commands: /usr/bin/lxc-*
- LXC Wrapper: /usr/bin/lxc
- LXC Container: /var/lib/lxc/$container
- Templates: /usr/share/lxc/templates
- Rootfs: /var/lib/lxc/$container/rootfs
- Config: /var/lib/lxc/$container/config
- Privileged and Unprivileged Containers

Linux Container (LXC) Practice: Usage (1)

Create a container

- Debian-specific: lxc-create -t debconfig -n example.org
- Upstream: lxc-create -t debian -n example.org

Start a container

- lxc-start -n example.org
- lxc-start -n examle.org -d

Stop a container

- lxc-stop -n example.org
- lxc-kill -n example.org

Linux Container (LXC) Practice: Usage (2)

Remove a container

- lxc-destroy -n example.org

Pause/Unpause a container

- lxc-freeze -n example.org
- lxc-unfreeze -n example.org

Attach a Console to container

- lxc-console -n example.org

List containers

- lxc-list
- lxc-ls –fancy

Linux Container (LXC) Practice: Preseeding

lxc-debconfig

- Uses debconf, therefore User and Admin friendly
- Two Stages: Bootstrapping and Postinst
- Preseed Files: /etc/lxc/debconfig
- Preseed Examples: /usr/share/lxc/preseed

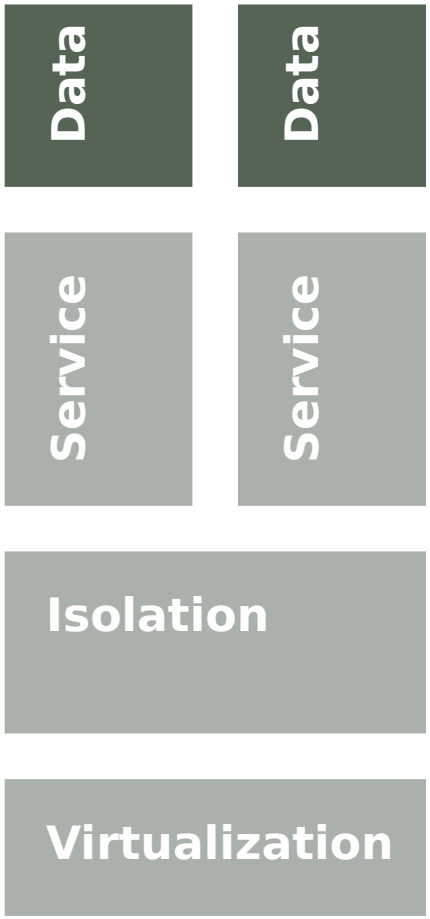
live-debconfig

- Uses debconf, therefore User and Admin friendly (dejavu :)
- Current backend shell, later puppet
- Documented, but also UTSL :(

Linux Container (LXC) Practice: Best Practice

- Always use FQDNs
- Data Separation
- Container Automation
- Use as many containers as reasonably needed/possible, only limit is IP address usage
- Combine LXC with your favourite tools (KISS)

Linux Container (1)



Host:
`/srv/lxc/data/$fqdn`

Container:
`/srv/$fqdn`

Host:
`/srv/lxc/containers/$fqdn`

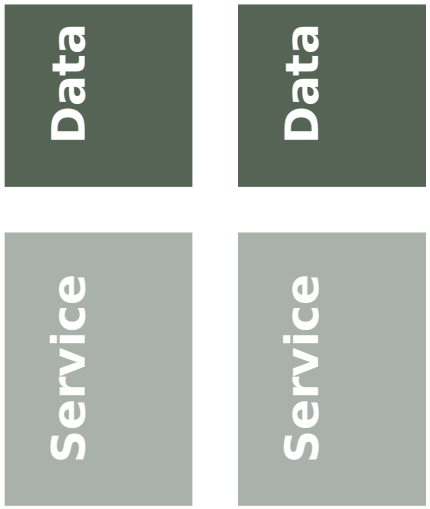
Container:
`/`

Host:
/

Container:
n/a

Host:
n/a

Linux Container (2)



Storage:
`bfhfilerbe01.bfh.ch:/vol/lx_data_01_v0_sata_enge`

Storage:
`bfhfilerbe02.bfh.ch:/vol/lx_containers_01_sas_enge`

Storage:
becenter

Storage:
becenter

Linux Container (LXC) Practice: Future Areas

Future

- Unprivileged containers
- Nested containers
- systemd/cgmanager
- criu
- puppet
- openvswitch
- LSM (SELinux)
- further finetuning and polishing Debian integration (config includes, image tarballs, etc.)

Rant

- libvirt-lxc
- Docker
- Security

Thank You for Your Attention.

Linux Container (LXC): Links

Namespaces/CGroups

- LWN Namespaces in Operation: http://lwn.net/Articles/531114/
- Kernel CGroups: Documentation/cgroups/cgroups.txt
- Pro-Linux CGroups Overview:
  http://www.pro-linux.de/artikel/2/1464/ressourcen-verwaltung-mit-control-groups-cgroups.html

LXC

- LXC Homepage: http://linuxcontainers.org
- LXC Networking Explained: http://containerops.org/2013/11/19/lxc-networking/
- LXC 1.0 Blog Post Series: http://www.stgraber.org/2013/12/20/lxc-1-0-blog-post-series/
- BFH: https://linux.bfh.ch/software/lxc/, https://linux.bfh.ch/services/lxc, and https://linux.bfh.ch/servers/

Further Reading...

- LWN The failure of operating systems and how we can fix it: http://lwn.net/Articles/524952/