Containers micro-conference

9:35 - 9:50	LXC: project update
9:50 - 10:05	Clear Containers: project update
10:05 - 10:20	OpenVZ: project update
10:20 - 10:35	runc: project update
11:05 - 11:20	rkt: project update
11:20 - 11:35	CGroup V2
11:35 - 11:50	File capabilities in user namespaces
11:50 - 12:05	State of the kernel support
12:05 - 12:20	Record and vPlay

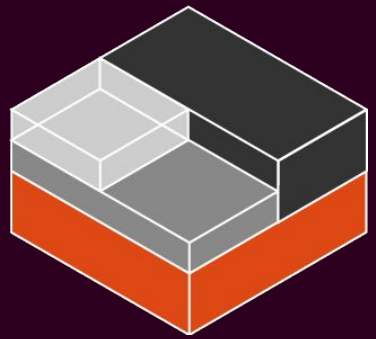
Each item is 10 minutes of presentation and 5 minutes of questions.

Linux Plumbers 2016
Santa Fe, New Mexico

ubuntu

LXC/LXCFS: project update

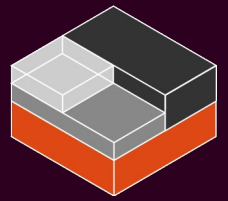
Linux Plumbers 2016
Santa Fe, New Mexico



Christian Brauner
Software engineer, Canonical Ltd.

christian.brauner@canonical.com @n06ab10

LXC/LXCFS: project update

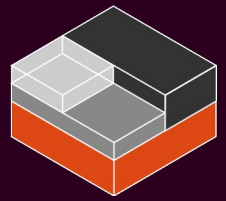


LXC

LXC 2.0 with completely backward compatible C API 1.2.0

- ➔ rewrote/refactored most low-level LXC tools
  - ◆ lxc-start-ephemeral and lxc-clone merged into lxc-copy
  - ◆ lxc-ls in C
  - ◆ unify behavior of tools wherever possible
  - ◆ lxc-attach: prevent tty pushback privilege escalation
- ➔ completely restructured and reworked the LXC storage backend
- ➔ cgfsng: Serge's new cgroup backend implementation
- ➔ correctly shut down systemd containers
- ➔ minimal unit tests

LXC/LXCFS: project update

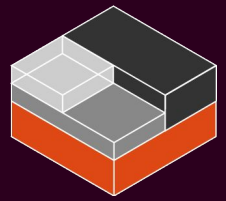


LXC

Added a bunch of new configuration items:

- ➔ lxc.ephemeral
- ➔ lxc.rebootsignal
- ➔ lxc.hook.destroy
- ➔ lxc.hook.stop
- ➔ lxc.monitor.unshare
- ➔ lxc.no_new_privs
- ➔ lxc.syslog

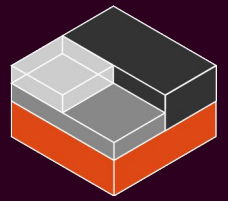
LXC/LXCFS: project update



LXC

```
chb@conventiont|~
> ls -al /var/lib/lxcfs/proc/
total 0
dr-xr-xr-x 2 root root 0 Oct 24 13:40 .
drwxr-xr-x 2 root root 0 Oct 24 13:40 ..
-r--r--r-- 1 root root 0 Oct 24 13:40 cpuinfo
-r--r--r-- 1 root root 0 Oct 24 13:40 diskstats
-r--r--r-- 1 root root 0 Oct 24 13:40 meminfo
-r--r--r-- 1 root root 0 Oct 24 13:40 stat
-r--r--r-- 1 root root 0 Oct 24 13:40 swaps
-r--r--r-- 1 root root 0 Oct 24 13:40 uptime

chb@conventiont|~
> ls -al /var/lib/lxcfs/cgroup/
total 0
drwxr-xr-x 2 root root 0 Oct 24 13:39 .
drwxr-xr-x 2 root root 0 Oct 24 13:39 ..
drwxr-xr-x 2 root root 0 Oct 24 13:39 blkio
drwxr-xr-x 2 root root 0 Oct 24 13:39 cpu,cpuacct
drwxr-xr-x 2 root root 0 Oct 24 13:39 cpuset
drwxr-xr-x 2 root root 0 Oct 24 13:39 devices
drwxr-xr-x 2 root root 0 Oct 24 13:39 freezer
drwxr-xr-x 2 root root 0 Oct 24 13:39 hugetlb
drwxr-xr-x 2 root root 0 Oct 24 13:39 memory
drwxr-xr-x 2 root root 0 Oct 24 13:39 name=systemd
drwxr-xr-x 2 root root 0 Oct 24 13:39 net_cls,net_prio
drwxr-xr-x 2 root root 0 Oct 24 13:39 perf_event
drwxr-xr-x 2 root root 0 Oct 24 13:39 pids
```

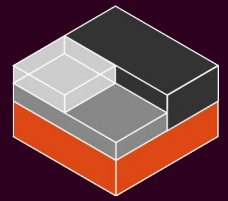
LXC/LXCFS: project update



LXCFS

- ➔ make LXCFS a versionless libtool module
- ➔ make LXCFS behave (mostly) like a standard filesystem
- ➔ move LXCFS filesystem to minimal namespaced chroot
  - ◆ Requirements
    - private cgroup mounts to not confuse Docker, Libvirt
    - make sure to not pin any host mounts in the new namespace
  - ◆ Solution
    - create minimal namespace
    - mount cgroups in there and open fd for each mounted controller

LXC/LXCFS: project update



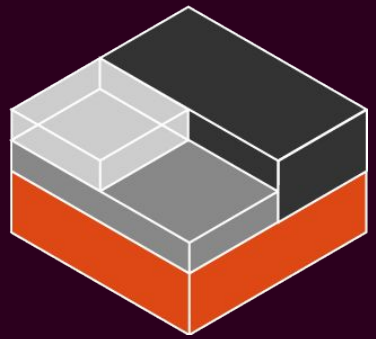
LXCFS

```
f = fopen("/proc/self/cgroup", "r"));

/* Parse cgroup mounts and store them in: */
static int num_hierarchies; // number of controllers
static char **hierarchies; // name of controller
static int *fd_hierarchies; // fd for each controller mounted in private mntns

/* Preserve initial namespace. */
init_ns = preserve_ns(getpid());

fd_hierarchies = malloc(sizeof(int *) * num_hierarchies);

for (i = 0; i < num_hierarchies; i++)
        fd_hierarchies[i] = -1;

/* Change to new mount namespace. */
unshare(CLONE_NEWNS) < 0);

/* Mount cgroups. */

/* Open fd in private namespace for each mounted controller. */
for (i = 0; i < num_hierarchies; i++)
        fd_hierarchies[i] = open(target, O_DIRECTORY);

/* If on ramfs chroot() else pivot_root() and umount2() everything we don't need. */

/* Switch back to initial mount namespace. */

setns(init_ns, 0);
```

ubuntu

LXD: project update

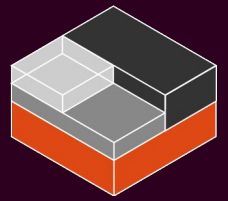
Linux Plumbers 2016
Santa Fe, New Mexico



Stéphane Graber
LXD project leader, Canonical Ltd.

stgraber@ubuntu.com @stgraber
https://www.stgraber.org

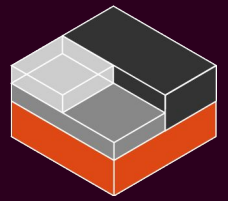
LXD: the container lighter-visor



A year of LXD

- ➔ First LTS release of LXD (2.0, supported for 5 years)
- ➔ Monthly stable releases
- ➔ Stable REST API to manage containers
- ➔ Support for live-migration and stateful snapshots through CRIU
- ➔ USB and GPU passthrough
- ➔ Resource limits (CPU, memory, block and network I/O and disk)
- ➔ Network management API
- ➔ Support for AppArmor namespacing and nesting
- ➔ Multiple storage backends (ZFS, btrfs, LVM+ext4, LVM+xfs, directory)

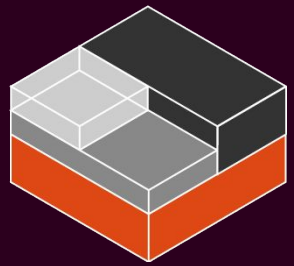
LXD: the container lighter-visor



LXD moving forward

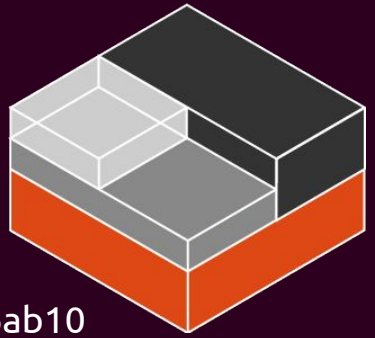
- ➔ Improved storage handling
- ➔ Better language bindings for the API
- ➔ Improved live migration
- ➔ More device passthrough and resource limits
- ➔ Better scripting user experience

Demo time!



Stéphane Graber
LXD project leader, Canonical Ltd.
stgraber@ubuntu.com @stgraber
https://www.stgraber.org

ubuntu



Christian Brauner
Software engineer, Canonical Ltd.
christian.brauner@canonical.com @n06ab10

Questions?

https://linuxcontainers.org/lxd
https://github.com/lxc/lxd

Try it yourself at: https://linuxcontainers.org/lxd/try-it
LXD stickers are available at the front!