fedora

What’s new and what’s next in Fedora CoreOS

Dusty Mabe
CoreOS engineer at Red Hat

Timothée Ravier
CoreOS engineer at Red Hat

Agenda

- What is Fedora CoreOS?
- What’s new since last year?
- What’s coming soon?
- Becoming a better Fedora Project Citizen

What is Fedora CoreOS?

An emerging Fedora edition

- Came from the **merging** of two communities:
  - CoreOS Inc’s Container Linux
  - Project Atomic’s Atomic Host
- Incorporates **Container Linux**
  - Philosophy
  - Provisioning Stack
  - Cloud Native Expertise
- Incorporates **Atomic Host**
  - Fedora Foundation
  - Update Stack
  - SELinux Enhanced Security

Philosophy behind Fedora CoreOS

- **Automatic updates**
  - No interaction for administrators

- **Automated provisioning**
  - All nodes start from **~same starting point**
  - Use Ignition to provision a node on **first boot**

- **Immutable infrastructure**
  - **Automate** deployment and system configuration
  - Update configs and **re-provision** to apply changes

- Additional software runs in **containers**
  - Makes host updates more **reliable**

Supported platforms and architectures

- Available for a plethora of **cloud/virt platforms**
  - Alibaba, AWS, Azure, DigitalOcean, Exoscale, GCP, IBM Cloud, OpenStack, Vultr, VMware, QEMU/KVM
  - Directly launchable on AWS & GCP

- Several options for **Bare Metal**
  - Live ISO (automated or interactive installations)
  - PXE (network) boot
  - Raw and 4K native disk images

- Currently **x86_64** only (**aarch64** support coming soon)

What’s new in Fedora CoreOS?

(since August 2020)

cgroups v2 by default

- Switched to **v2 by default** since version 34.20210529.3.0
- **podman** & **Docker** support
- **No v1 to v2 auto update** (must re-create containers)
- Update existing systems with:

```
rpm-ostree kargs --delete=systemd.unified_cgroup_hierarchy --reboot
```

https://docs.fedoraproject.org/en-US/fedora-coreos/kernel-args/#_removing_existing_kernel_arguments

Reliable live changes to the system

- New options to change the system content **live** in a **safe**, **atomic** and optionally non-persistent way
- rpm-ostree usroverlay
  - Mounts a non persistent RW overlay on top of /usr
- rpm-ostree install --apply-live strace
  - **Install** a package into a new (offline) deployment
  - Atomically switch the running system to this deployment to **apply the changes live** (still RO)

rpm-ostree v2021.1 & rpm-ostree v2021.3 & https://coreos.github.io/rpm-ostree/apply-live/

Kernel arguments in Ignition

- Add, remove, replace kernel arguments **via Ignition**
- Applied on **first boot**, will trigger a reboot

```
# Disabling CPU
# vulnerability mitigations
variant: fcos
version: 1.4.0
kernel_arguments:
  should_exist:
    - mitigations=off
  should_not_exist:
    - mitigations=auto,nosmt
```

```
# Staying on cgroups v1
variant: fcos
version: 1.4.0
kernel_arguments:
  should_exist:
    - systemd.unified_cgroup_hierarchy=0
```

https://docs.fedoraproject.org/en-US/fedora-coreos/kernel-args/#_modifying_kernel_arguments_via_ignition

Introducing bootupd

- What?
  - Bootloader updater for rpm-ostree based systems
  - Currently **UEFI only** (BIOS planned)
- Why?
  - Transactional bootloader updates are really hard
  - Thus ostree/rpm-ostree do not update bootloaders
- How?
  - **Manually** triggered by users when **known to be safe**
  - bootupctl update

https://github.com/coreos/bootupd

/boot is now read-only

- Manually modifying content in /boot is **discouraged**
- Change **kernel arguments** with:
  - rpm-ostree kargs
- Change **boot order** with:
  - rpm-ostree rollback / update / deploy

https://docs.fedoraproject.org/en-US/fedora-coreos/storage/#_mounted_filesystems

Encrypted storage via LUKS in Ignition

- Unlock via a keyfile, TPM2 or a Tang server (via Clevis)
- Includes support for the **root partition**
  - Requires unlocking via a TPM2 or a Tang server

```
# LUKS for / using TPM2
variant: fcos
version: 1.4.0
boot_device:
  luks:
    tpm2: true
```

```
# LUKS for another device
variant: fcos
version: 1.4.0
storage:
  luks:
    - name: data
      device: /dev/vdb
      clevis:
        tpm2: true
  filesystems:
    - path: /var/lib/data
      device: /dev/mapper/data
      format: xfs
      label: DATA
      with_mount_unit: true
```

https://docs.fedoraproject.org/en-US/fedora-coreos/storage/#_encrypted_storage_luks

RAID support in Ignition

- Setup any RAID level (0, 1, 5, etc.) on first boot **via Ignition**
- Mirrors EFI System Partition (ESP) & BIOS bootloader
- Side effect: ESP no longer mounted (empty /boot/efi)

```
# Mirror boot disk with RAID1
variant: fcos
version: 1.4.0
boot_device:
  mirror:
    devices:
      - /dev/sda
      - /dev/sdb
```

```
# Move / to RAID0
variant: fcos
version: 1.4.0
storage:
  raid:
    - name: myroot
      level: raid0
      devices:
        - /dev/disk/by-id/virtio-disk1
        - /dev/disk/by-id/virtio-disk2
  filesystems:
    - device: /dev/md/myroot
      format: xfs
      wipe_filesystem: true
      label: root
```

https://docs.fedoraproject.org/en-US/fedora-coreos/storage/#_reconfiguring_the_root_filesystem

More options for booting via (i)PXE

- Booting **transient systems** via (i)PXE
- Target system needs a kernel, initramfs and rootfs
- Final rootfs **used to be** included with the initramfs
- **Now split** to enable more flexibility:
    - Download from initramfs: coreos.live.rootfs_url= kargs
    - Use multiple initrd= for initramfs & rootfs in PXE config
    - Re-bundle: append rootfs to initramfs to use as initrd=

https://docs.fedoraproject.org/en-US/fedora-coreos/live-booting-ipxe/

What’s coming soon in Fedora CoreOS?

DNF Count Me support (Aug 2021)

- Enables **privacy preserving** and reliable system counting
- Only reports a **large approximation** of the age of a system
- Only reaches out to **official** Fedora repositories servers
- **No other information** sent or stored

https://fedoramagazine.org/getting-better-at-counting-rpm-ostree-based-systems/
https://github.com/coreos/fedora-coreos-tracker/issues/717

iptables using nftables by default

- **iptables** still using **legacy** backend instead of **nftables** one
- **Unintended** consequence of alternatives(8) 's behaviour
  - Configuration stored in a mix of **/var** and **/etc**
  - **Incompatibility** with **rpm-ostree** strict split between configuration and data
- Easy **workaround** available
- Full fix requires **adjustments** to alternatives(8) or an alternative(!)

https://docs.fedoraproject.org/en-US/fedora-coreos/alternatives/
https://github.com/coreos/fedora-coreos-tracker/issues/676
https://github.com/coreos/fedora-coreos-tracker/issues/677

systemd-resolved fully enabled

- Made the switch to systemd-resolved by default with **F34**
- Had to disable the stub listener due to **unexpected issues**
  - Reverse DNS lookups stopped working and caused system hostnames to not properly get set
- **Issue resolved** by augmenting NetworkManager to handle specific corner cases involving reverse DNS lookups
- Fix will be available in Fedora 35 and we'll fully enable systemd-resolved there

https://github.com/coreos/fedora-coreos-tracker/issues/834

ostree commits in container images

- New commands to export an ostree commit to a **container image**
- Enables **rebasing** to the content of a container image:
  - rpm-ostree rebase --experimental docker://quay.io/cgwalters/fcos:latest
- Enables **running** an ostree commit as a container for testing and debugging:
  - podman run --rm -ti quay.io/cgwalters/fcos:latest /bin/bash
  - **Not fit as a base** for application containers!

https://lists.fedoraproject.org/archives/list/devel@lists.fedoraproject.org/thread/B23FZILDI3J73OMION2IDEYMLKNKN5YE/

cliwrap: Helping with muscle memory

- CLI wrapper for **common** command line tools:
  - rpm, yum/dnf, grubby, etc.
- Easier to understand error messages and **hints**
- Help with the **transition** from classic dnf systems to rpm-ostree based ones
- **Optionally** enabled with:
  - rpm-ostree deploy --ex-cliwrap=true
  - Combine with: rpm-ostree ex apply-live

https://lists.fedoraproject.org/archives/list/devel@lists.fedoraproject.org/thread/7P5EYBYDG44LCTEGSMHHBFTFUCP4VN4R/

Becoming a Better Fedora Project Citizen

Background Context

- Fedora CoreOS…
    - is a merging and re-invention of:
        - Container Linux
        - Atomic Host
    - is the basis for upstream/downstream OKD/OCP
    - follows a different release model
        - stable/testing/next streams release every two weeks

Background Context

- Fedora CoreOS…
    - has a heavy reliance on CI and speed
        - releasing multiple streams every 2 weeks
        - OpenShift release cadence is much faster than RHEL
        - Automated tests++++++
    - needed custom release tooling
        - Build pipelines that can run many times a day
        - Containerized development environment
            - Quickly and easily build/run/test any FCOS artifact locally

Fedora Change Requests Reviews

- Actively reviewing Fedora Changes requests during the development release cycle
- The discussions/evaluations for Fedora 35 are in our issue tracker tagged with the F35-changes label

Building and Testing against Rawhide

- We are now building and testing a **rawhide** stream
  - Suite of automated tests now complement rawhide!
  - Helps identify unexpected breakage from new features.
  - Now participate closer upstream with developers and get general problems fixed.

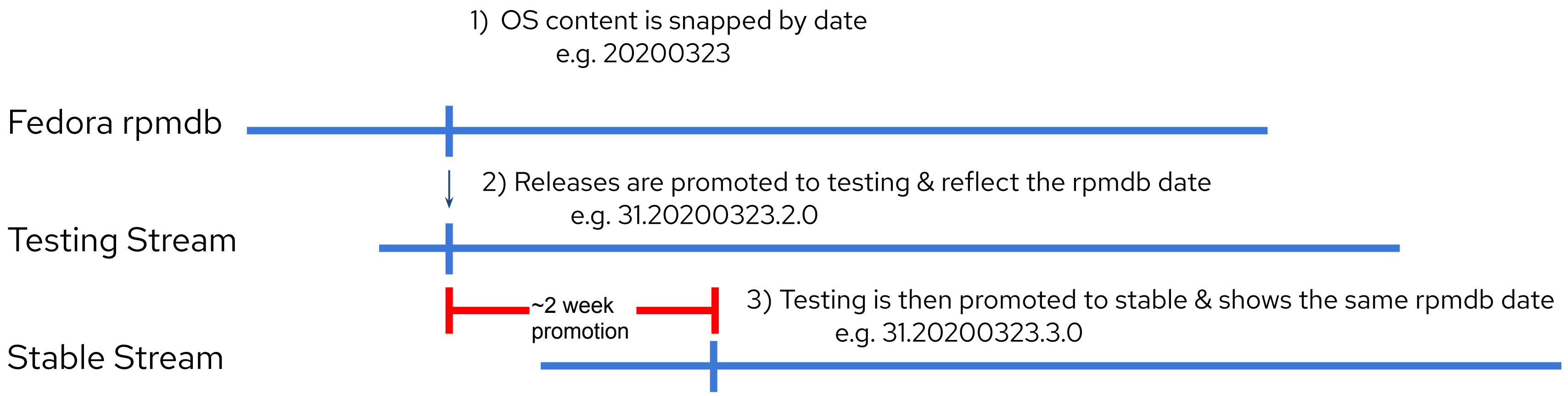
FESCO discussion/participation

- Participating in FESCO discussions
  - Allow the FCOS group to get advanced knowledge of future changes.
  - Allow us to help influence and add perspective on how changes affect Fedora CoreOS users.
- Potentially have FCOS representative run for FESCO

Default Settings Policy Changes

- Currently some friction between adopted Fedora Changes and Kubernetes required defaults.
- We’ve decided to adopt a policy that allows us to apply changes that aren’t reasonable defaults for K8s
  - https://github.com/coreos/fedora-coreos-tracker/issues/880
  - example: swap-on-zram, k8s doesn’t support swap
- For now, add documentation for kubernetes distributors
  - future: possibly gate changes with “feature flags”
  - https://github.com/coreos/fedora-coreos-tracker/issues/892

Closer Proximity to Fedora Releases



1) OS content is snapped by date
e.g. 20200323

Fedora rpmdb

2) Releases are promoted to testing & reflect the rpmdb date
e.g. 31.20200323.2.0

Testing Stream

~2 week
promotion

3) Testing is then promoted to stable & shows the same rpmdb date
e.g. 31.20200323.3.0

Stable Stream

Closer Proximity to Fedora Releases

- Fedora Beta Release
  - The **next** stream is switched over to the new Fedora release
- Fedora Final Freeze
  - The **next** stream ➡️ weekly releases to closely track GA content
- Fedora General Availability
  - Fedora CoreOS re-orients its release schedule:
    - Week 0 (GA release): **next** with latest Fedora N content
    - Week 1: **testing** release promoted from previous **next**
    - Week 3: **stable** release promoted from previous **testing**
      - now fully rebased to Fedora N.

https://github.com/coreos/fedora-coreos-tracker/blob/main/Design.md#major-fedora-version-rebases

Questions/Demo

Get involved!

- Web: https://getfedora.org/coreos
- Issues: https://github.com/coreos/fedora-coreos-tracker/issues
- Forum: https://discussion.fedoraproject.org/c/server/coreos
- Mailing list: coreos@lists.fedoraproject.org
- IRC: Libera.chat #fedora-coreos
- Other talks to get started:
    - Fedora CoreOS Introduction (Jul 13, 2020)
    - Getting Started with Fedora CoreOS (Mar 17, 2021)