Updates and Innovations with the Apptainer Platform

Forrest Burt - Solutions Architect, CIQ
February 3, 2024
FOSDEM 2024
Brussels, Belgium

Developments in Apptainer over the past few years

- Leveraging the User Namespace
- New recommendations for containerized MPI
- Increased adoption of ORAS (for Software Supply Chain)

Leveraging the User Namespace

Namespace types

The following table shows the namespace types available on Linux. The second column of the table show
ous APIs. The third column identifies the manual page that provides details on the namespace type. T
namespace type.

Namespace	Flag	Page	Isolates
Cgroup	CLONE_NEWCGROUP	cgroup_namespaces(7)	Cgroup root directory
IPC	CLONE_NEWIPC	ipc_namespaces(7)	System V IPC, POSIX message queues
Network	CLONE_NEWNET	network_namespaces(7)	Network devices, stacks, ports, etc.
Mount	CLONE_NEWNS	mount_namespaces(7)	Mount points
PID	CLONE_NEWPID	pid_namespaces(7)	Process IDs
Time	CLONE_NEWTIME	time_namespaces(7)	Boot and monotonic clocks
User	CLONE_NEWUSER	user_namespaces(7)	User and group IDs
UTS	CLONE_NEWUTS	uts_namespaces(7)	Hostname and NIS domain name

The namespaces API

As well as various /proc files described below, the namespaces API includes the following system calls

Leveraging the User Namespace (for `--fakeroot`)

```
[demouser@ciqbox ~]$ id
uid=1001(demouser) gid=1001(demouser) groups=1001(demouser) context=unconfined_u:unconfined_r:unconfi
ned_t:s0-s0:c0.c1023

[demouser@ciqbox ~]$ pwd
/home/demouser

[demouser@ciqbox ~]$ touch foo

[demouser@ciqbox ~]$ ll
total 0
-rw-r--r--. 1 demouser demouser 0 Jul 27 14:55 foo

[demouser@ciqbox ~]$
```

Leveraging the User Namespace (for `--fakeroot`)

```
[demouser@ciqbox ~]$ apptainer shell --fakeroot docker://alpine
Apptainer> id
uid=0(root) gid=0(root) groups=0(root)
Apptainer> pwd
/root
Apptainer> ls -l
total 0
-rw-r--r--    1 root     root             0 Jul 27 14:55 foo
Apptainer> touch bar
Apptainer> ls -l
total 0
-rw-r--r--    1 root     root             0 Jul 27 14:58 bar
-rw-r--r--    1 root     root             0 Jul 27 14:55 foo
Apptainer> exit

[demouser@ciqbox ~]$ ll
total 0
-rw-r--r--. 1 demouser demouser 0 Jul 27 14:58 bar
-rw-r--r--. 1 demouser demouser 0 Jul 27 14:55 foo

[demouser@ciqbox ~]$
```

Leveraging the User Namespace (for `build`)

```
[demouser@ciqbox ~]$ cat example.def
Bootstrap: docker
From: rockylinux:8

%post
        whoami
        dnf -y update
        dnf -y install vim
        # ^ note that these commands require privs

[demouser@ciqbox ~]$ apptainer build example.sif example.def
+ whoami
root
+ dnf -y update
Rocky Linux 8 - AppStream                                      594 kB/s |  10 MB     00:17
Rocky Linux 8 - BaseOS                                         460 kB/s | 4.9 MB     00:10
Rocky Linux 8 - Extras                                          40 kB/s |  13 kB     00:00
Dependencies resolved.
==========================================================================================
 Package                    Architecture   Version                        Repository        Size
==========================================================================================
Upgrading:
 krb5-libs                  x86_64         1.18.2-25.el8_8                baseos            841 k
 platform-python            x86_64         3.6.8-51.el8_8.1.rocky.0       baseos             86 k
 python3-libs               x86_64         3.6.8-51.el8_8.1.rocky.0       baseos            7.8 M
```

Leveraging the User Namespace (for `build`)

```
Running transaction test
Transaction test succeeded.
Running transaction
  Preparing        :                                                        1/1
  Installing       : which-2.21-18.el8.x86_64                               1/5
  Installing       : vim-filesystem-2:8.0.1763-19.el8_6.4.noarch            2/5
  Installing       : vim-common-2:8.0.1763-19.el8_6.4.x86_64                3/5
  Installing       : gpm-libs-1.20.7-17.el8.x86_64                          4/5
  Running scriptlet: gpm-libs-1.20.7-17.el8.x86_64                          4/5
  Installing       : vim-enhanced-2:8.0.1763-19.el8_6.4.x86_64              5/5
  Running scriptlet: vim-enhanced-2:8.0.1763-19.el8_6.4.x86_64              5/5
  Running scriptlet: vim-common-2:8.0.1763-19.el8_6.4.x86_64                5/5
  Verifying        : gpm-libs-1.20.7-17.el8.x86_64                          1/5
  Verifying        : vim-common-2:8.0.1763-19.el8_6.4.x86_64                2/5
  Verifying        : vim-enhanced-2:8.0.1763-19.el8_6.4.x86_64              3/5
  Verifying        : vim-filesystem-2:8.0.1763-19.el8_6.4.noarch            4/5
  Verifying        : which-2.21-18.el8.x86_64                               5/5

Installed:
  gpm-libs-1.20.7-17.el8.x86_64                    vim-common-2:8.0.1763-19.el8_6.4.x86_64
  vim-enhanced-2:8.0.1763-19.el8_6.4.x86_64        vim-filesystem-2:8.0.1763-19.el8_6.4.noarch
  which-2.21-18.el8.x86_64

Complete!

[demouser@ciqbox ~]$
```

Leveraging the User Namespace (for installation)

- Default Apptainer installation is now unprivileged
- Squashfuse is used to mount the squashfs file system
- Apptainer enters a new User Namespace and then creates a new Mount Namespace to present the new root filesystem to processes

Leveraging the User Namespace (for installation)

apptainer/tools/install-un

github.com/apptainer/apptainer/blob/main/tools/install-unprivileged.sh

Product Solutions Open Source Pricing Search or jump to... / Sign in Sign up

apptainer / apptainer Public

Notifications Fork 92 Star 633

Code Issues 134 Pull requests 3 Actions Projects Security 2 Insights

main apptainer / tools / install-unprivileged.sh Go to file

DrDaveD update gocryptfs to 2.4.0 495f7cc · last month History

Code Blame Executable File · 430 lines (399 loc) · 12.2 KB Raw

```
#!/bin/bash
## Copyright (c) Contributors to the Apptainer project, established as
#   Apptainer a Series of LF Projects LLC.
#   For website terms of use, trademark policy, privacy policy and other
#   project policies see https://lfprojects.org/policies
#
# This installs relocatable released versions of unprivileged apptainer
# and all dependencies for a single architecture, from either Fedora or
# EPEL+CentOS7 or EPEL+RockyLinux (or AlmaLinux) distributions.  Supports
# the architectures that are available in those distributions.
```

Leveraging the User Namespace (for installation)

Important Note:

Because users can install in their own space they can manage their own configuration. This means that **things like ECL may be rendered invalid!**

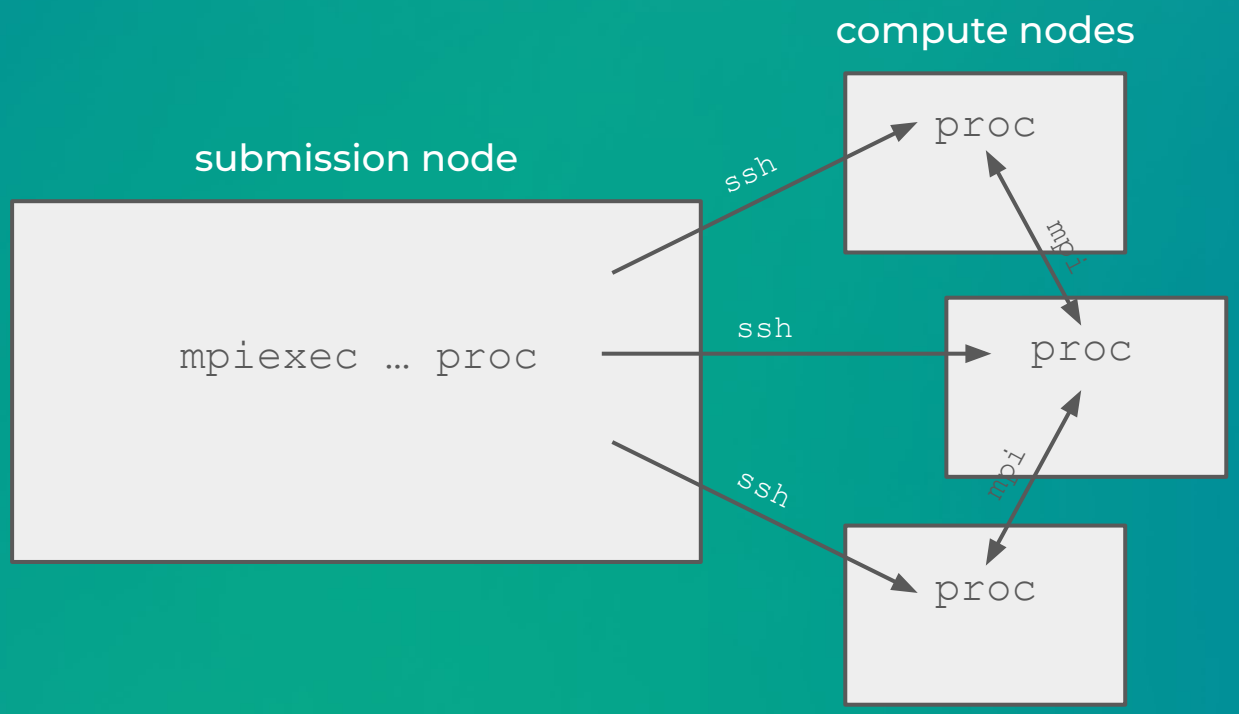
Must disable User Namespace in OS if this is a problem.

New recommendations for MPI jobs

- Containerized MPI jobs can present difficulties
  - wire-up
  - fabric adapter
- Historical approaches are difficult and lack portability
- PMI and libfabric are recommended to increase container portability
- Container-friendly package managers (like Spack) can reduce difficulty

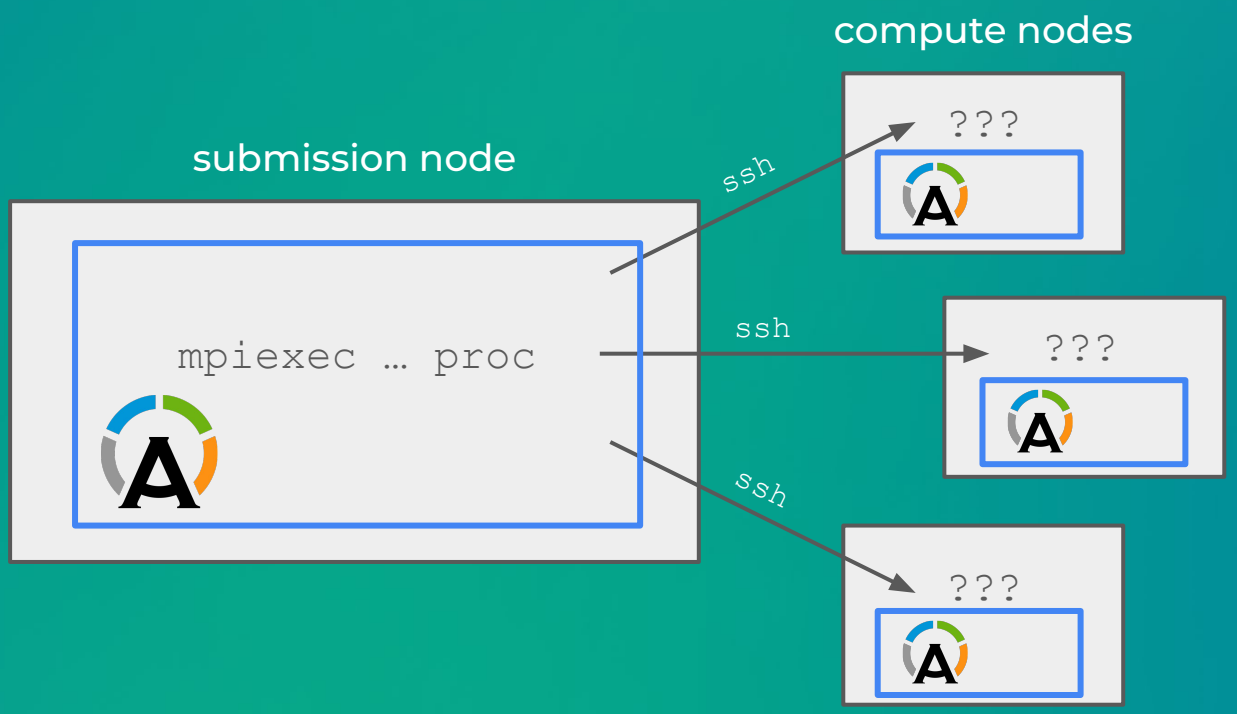
New recommendations for MPI jobs

Review of the wire-up problem



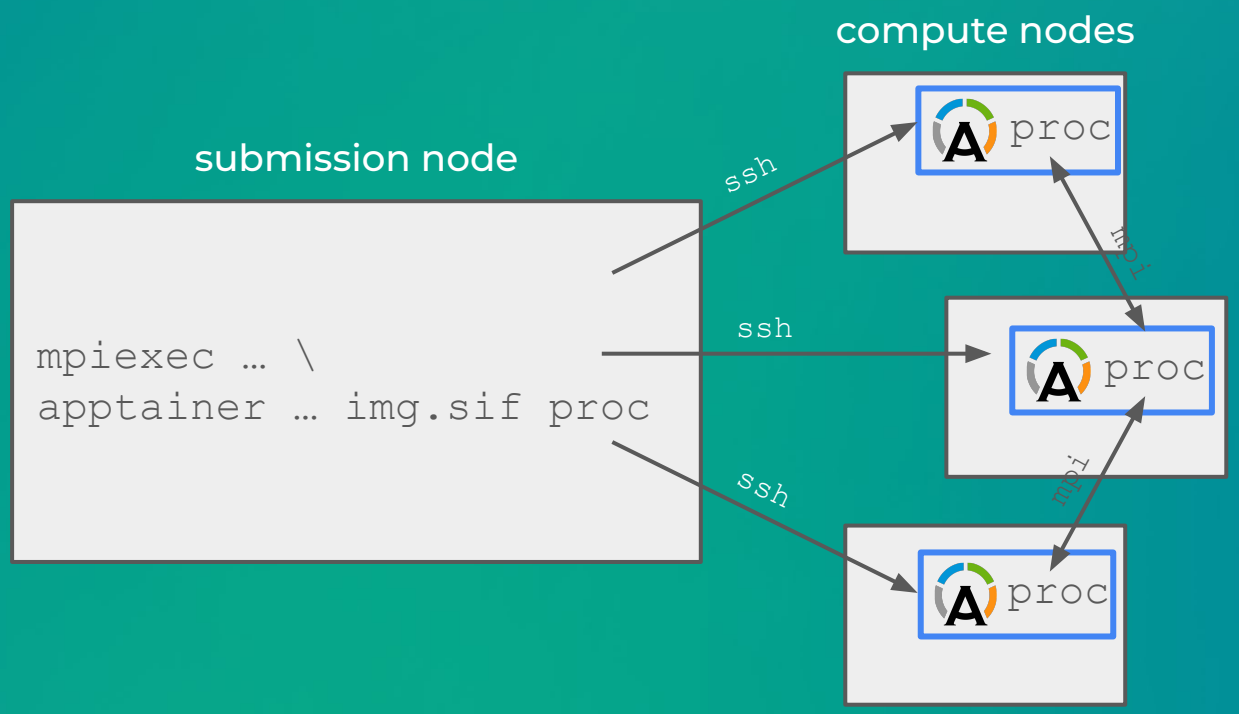
New recommendations for MPI jobs

Review of the wire-up problem



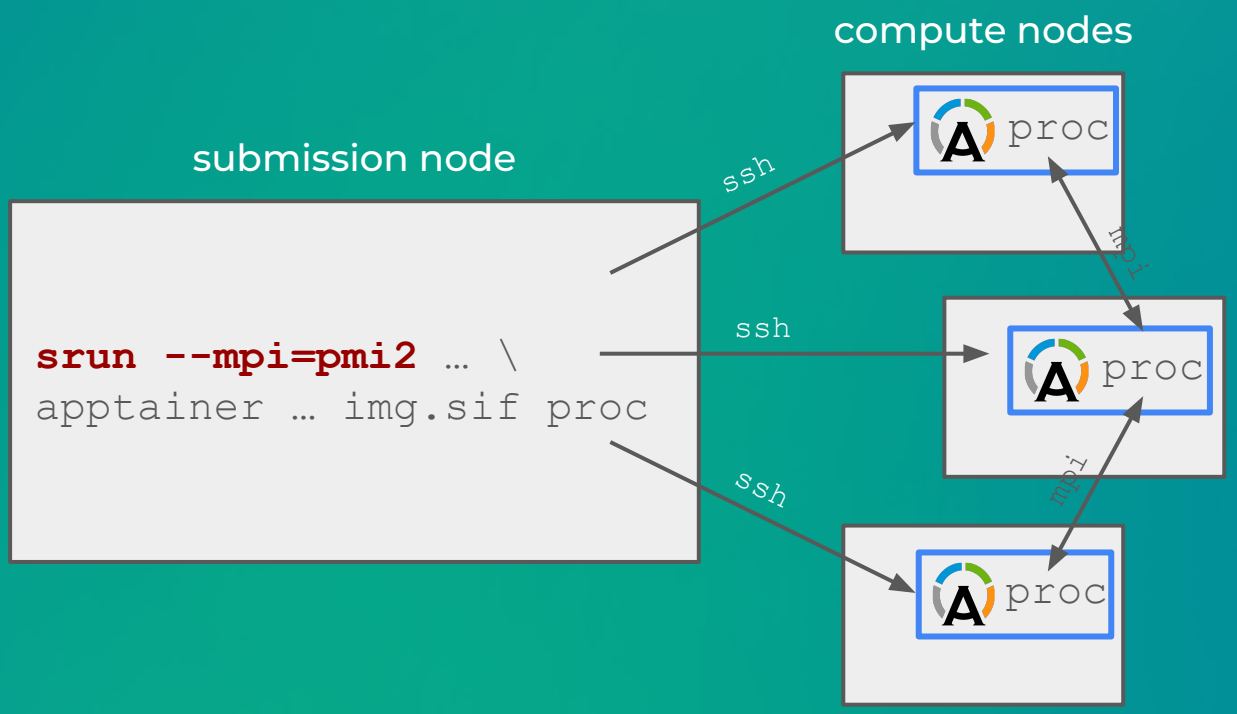
New recommendations for MPI jobs

Review of the wire-up problem



New recommendations for MPI jobs

Solving the wire-up problem with PMI

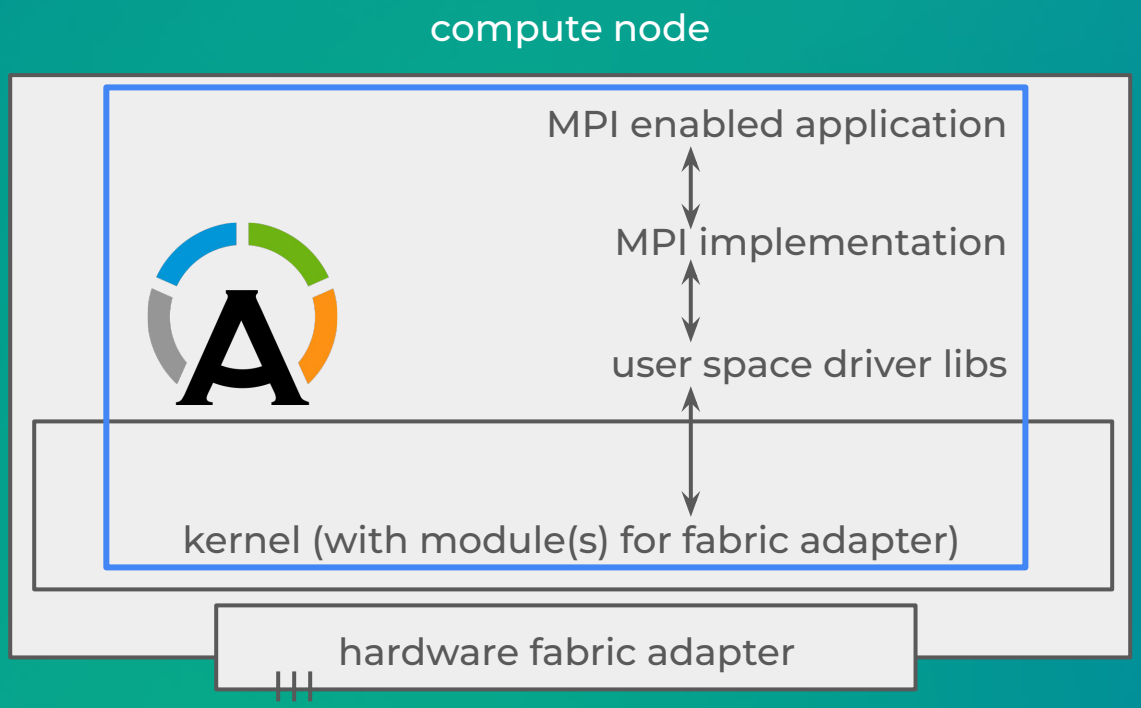


New recommendations for MPI jobs
Links to detailed resources

- Blog post with detailed (copy/paste-able) scripts
- Jonathon Anderson presenting this work at the 2023 HPC-AI Advisory council meeting at Stanford
- Dave Godlove and Jonathon Anderson discussing this work in a CIQ Webinar

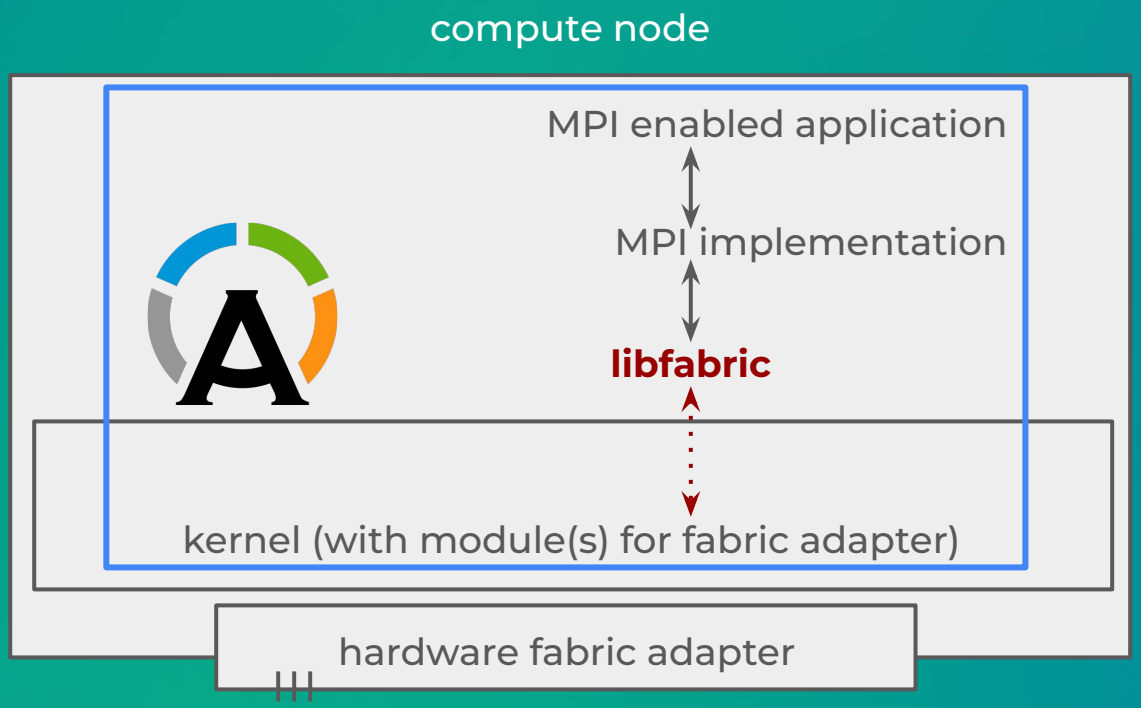
New recommendations for MPI jobs

Review of the fabric adapter problem



New recommendations for MPI jobs

Solving the fabric adapter problem with libfabric



ORAS for Software Supply Chain

- The OCI Registry As Storage (ORAS) protocol allows native SIF files to be stored on OCI registries (like Docker Hub)
- You can use advanced features like signing, verifying, and encryption without giving up the convenience of OCI registries!

ORAS for Software Supply Chain

godlovedc/rockylinux - Do

hub.docker.com/r/godlovedc/rockylinux

Search Docker Hub

Explore Pricing Sign In Register

Explore godlovedc/rockylinux

godlovedc/rockylinux

Pulls 48

By godlovedc • Updated 2 months ago

signed SIF images of Rocky Linux fingerprint is B7761495F83E6BF7686CA5F0C1A7D02200787921

Artifact

Overview Tags

These are signed SIF images of Rocky Linux built from the upstream mirrors (as opposed to building from existing containers). These can be pulled and verified as known good base containers to help with software supply chain concerns.

The fingerprint of my key (used for signature verification) is `B7761495F83E6BF7686CA5F0C1A7D02200787921`.

Docker Pull Command

```
docker pull godlovedc/r…
```

ORAS for Software Supply Chain

```
[demouser@ciqbox ~]$ apptainer pull oras://docker.io/godlovedc/rockylinux:8
INFO:    Downloading oras image

[demouser@ciqbox ~]$ apptainer verify rockylinux_8.sif
INFO:    Verifying image with PGP key material
[REMOTE]  Signing entity: David Godlove (production key) <davidgodlove@gmail.com>
[REMOTE]  Fingerprint: B7761495F83E6BF7686CA5F0C1A7D02200787921
Objects verified:
ID  |GROUP   |LINK    |TYPE
------------------------------------------------
1   |1       |NONE    |Def.FILE
2   |1       |NONE    |JSON.Generic
3   |1       |NONE    |FS
INFO:    Verified signature(s) from image 'rockylinux_8.sif'

[demouser@ciqbox ~]$
```

ORAS for Software Supply Chain

```
[demouser@ciqbox ~]$ apptainer inspect --deffile rockylinux_8.sif
BootStrap: yum
OSVersion: 8
MirrorURL: http://dl.rockylinux.org/pub/rocky/%{OSVERSION}/BaseOS/x86_64/os/
Include: dnf

%labels
    Author: davidgodlove@gmail.com

%post
    dnf -y update
    dnf install -y epel-release file

%environment
    LC_ALL=C

[demouser@ciqbox ~]$
```

ORAS for Software Supply Chain

```
Bootstrap: oras
From: docker.io/godlovedc/rockylinux:8
Fingerprints: B7761495F83E6BF7686CA5F0C1A7D02200787921

%post
        echo 'only install if signed and verified!'
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
-- INSERT --                                          6,45-52       All
```

ORAS for Software Supply Chain

```
[demouser@ciqbox ~]$ apptainer build sigexample.sif sigexample.def
WARNING: 'nodev' mount option set on /tmp, it could be a source of failure during build pr
ocess
INFO:    Starting build...
INFO:    Downloading oras image
INFO:    Checking bootstrap image verifies with fingerprint(s): [B7761495F83E6BF7686CA5F0C
1A7D02200787921]
INFO:    Running post scriptlet
+ echo 'only install if signed and verified!'
only install if signed and verified!
INFO:    Creating SIF file...
INFO:    Build complete: sigexample.sif

[demouser@ciqbox ~]$
```

ORAS for Software Supply Chain

```
[demouser@ciqbox ~]$ apptainer build sigexample.sif sigexample.def
WARNING: 'nodev' mount option set on /tmp, it could be a source of failure during build pr
ocess
INFO:    Starting build...
INFO:    Downloading oras image
INFO:    Checking bootstrap image verifies with fingerprint(s): [B7761495F83E6BF7686CA5F0C
1A7D02200787922]
FATAL:   While performing build: conveyor failed to get: while checking fingerprint: image
 not signed by required entities

[demouser@ciqbox ~]$
```

Developments in Apptainer over the past few years

- Leveraging the User Namespace
- New recommendations for containerized MPI
- Increased adoption of ORAS (for Software Supply Chain)

Get involved!

- Website: https://apptainer.org/
- Get Started: https://apptainer.org/get-started/
- GitHub: https://github.com/apptainer/apptainer
- Community Slack, mailing list, etc links: https://apptainer.org/support/
- User docs: https://apptainer.org/docs/user/main/
- Admin docs: https://apptainer.org/docs/admin/main/

THANK YOU