Shifter at NERSC

Lisa Gerhardt
Shane Canon
NERSC Data and Analytics Services

March 21, 2016
Deploying containers to HPC

• **Docker: open source, automated container deployment service**
  
  – Docker containers wrap up a piece of software in a complete filesystem that contains everything it needs to run (code, runtime, system tools and libraries)
  
  – Guaranteed to operate the same, regardless of the environment in which it is running

• **NERSC has enabled Docker-like container technology on its systems through a new software package known as Shifter**
Shifter at NERSC

- Secure and scalable way to deliver containers to HPC
- Implemented on Cori and Edison
- Supports Docker images and other images (vmware, ext4, squashfs, etc.)
- Basic Idea
  - Convert from native image format to common format
  - Chroot using common image on compute nodes

https://www.nersc.gov/research-and-development/user-defined-images/
Why Use Shifter?

• **Shifter allows you to fully customize your operating environment**
  – Want SL 6 with 32 bit libraries? Use Shifter
  – Have a very complicated software stack with lots of dependencies? Use Shifter

• **Shifter allows you to share images across platforms**
  – Images are NERSC-independent, can be run anywhere
Shifter is Fast

Another benefit: Improved shared library loading times compared to shared cluster file system
Shifter: How to Use It

module load shifter

shifterimg images

cori11> shifterimg images
cori  docker  READY  ce20c473cd  2015-12-11T09:53:47  centos:7
  
  cori  docker  READY  17583c7dd0  2015-12-04T08:32:58  busybox:latest
  
  cori  docker  READY  92b9f2dbf7  2015-12-04T10:48:12  scanon/shanetest:latest
  
  cori  docker  READY  d120eb68e8  2015-12-04T15:36:13  jcorrea/spot2:v1
  
  cori  docker  READY  621a496b0d  2015-12-08T06:26:58  registry.services.nersc.gov/scanon/shanetest:latest
  
  cori  docker  READY  d55e68e6cc  2015-12-22T09:16:46  ubuntu:14.04
  
  cori  docker  READY  86314f27ce  2015-12-09T00:33:46  registry.services.nersc.gov/patrick/psana:latest
  
  cori  docker  READY  d6bf89bd0b  2016-02-08T09:32:43  registry.services.nersc.gov/marius311/centos:7
  
  cori  docker  READY  b345ea6cf6  2015-12-09T07:57:41  registry.services.nersc.gov/marius311/centos:7
  
  cori  docker  READY  33200b4db5  2015-12-12T00:00:06  kbase/kbase_base:develop
  
  cori  docker  READY  e5ff6dcec0  2015-12-14T11:03:53  registry.services.nersc.gov/registry.services.nersc.gov
  
  cori  docker  READY  109b72e23c  2015-12-15T05:50:38  fedora:23
  
  cori  docker  READY  2bc6cdd62f  2015-12-15T06:06:25  miguelgila/wlcg_wn:20151223
  
  cori  docker  READY  6378f30b51  2015-12-16T17:13:54  atlas_cvmfs:latest
  
  cori  docker  READY  9ff944a24c  2015-12-18T03:08:09  marius311/run_sim:latest
  
  cori  docker  READY  56e5a8a6e0  2016-01-07T15:24:40  marius311/mpitest:latest
  
  cori  docker  READY  df9675993b  2015-12-21T14:47:19  paterno/centos67-art_v1_14
  
  cori  docker  READY  c0f009e667  2015-12-22T09:37:10  fenicsproject:dev:latest
  
  cori  docker  READY  0f73fcfb8a  2015-12-22T13:08:00  fenicsproject/stable:latest
  
  cori  docker  READY  f1c24227b7  2015-12-22T10:54:20  jbkowalkowski/art_test:latest
  
  cori  docker  READY  3f2d105bb2  2015-12-23T17:06:47  cms_cvmfs:latest
  
  cori  docker  READY  e51ffe812d  2016-01-07T11:46:19  paterno/centos-uboone_v04
  
  cori  docker  READY  9e6c507dae  2016-01-11T22:40:36  cmfs_test:latest
  
  cori  docker  READY  8e24aaaa895  2016-01-12T16:45:12  cmfs_mptest:latest
  
  cori  docker  READY  63e5da39eb02  2016-01-13T17:37:11  alice_cvmfs:latest
Pulling Down an Image from Dockerhub

```
shifterimg -vvv pull docker:ubuntu:16.04
```

Format is source:image_name: tag

repeat command to see status

```
Message: {
  "ENTRY": "MISSING",
  "ENV": "MISSING",
  "WORKDIR": "MISSING",
  "groupAcl": [],
  "id": "MISSING",
  "itype": "docker",
  "last_pull": "MISSING",
  "status": "INIT",
  "system": "cori",
  "tag": [],
  "userAcl": []
}
```

```
Message: {
  "ENTRY": "MISSING",
  "ENV": "MISSING",
  "WORKDIR": "MISSING",
  "groupAcl": [],
  "id": "MISSING",
  "itype": "docker",
  "last_pull": 1458578536.426084,
  "status": "PULLING",
  "system": "cori",
  "tag": [],
  "userAcl": []
}
```

```
Message: {
  "ENTRY": null,
  "ENV": [],
  "WORKDIR": "MISSING",
  "groupAcl": [],
  "id": "e0e03236f2d978612dae8da250eceffdf6f7057f",
  "itype": "docker",
  "last_pull": 1458578649.793376,
  "status": "READY",
  "system": "cori",
  "tag": [
    "ubuntu:16.04"
  ],
  "userAcl": []
}
```
#!/bin/bash
#SBATCH --image=docker:image_name:latest
#SBATCH --nodes=1
#SBATCH --partition=regular

module load shifter
srun -n 32 shifter python myPythonScript.py args

Use “sbatch” to submit as normal

Many more commands at
https://www.nersc.gov/research-and-development/user-defined-images/
Shifter Functionality

• Can bind mount file systems to specific directories:
  – #SBATCH --volume="/global/cscratch1/sd/lgerhard:/srv"

• Can also create a writeable temporary xfs file to each node to mimic local disk
  – #SBATCH --volume=""/global/cscratch1/sd/lgerhard/spark_tmp_dir/pca/pca:/tmp:perNodeCache=size=200G"

• Can load multiple images in a single job
• Images can use MPI via ssh out of the box
• Images can access Burst Buffer
Coming Soon

• Working on being able to pull images from private repositories

• Working on building a framework for using MPI w/full Cray interface in Shifter

• If you have any questions about using Shifter, please contact consult@nersc.gov
How Shifter Works

User loads image “docker pull <image_name>”

Image is stored by gateway service

Batch system prolog stages image on nodes

Uses chroot to produce transparent environment for user’s job