Using Software Containers on Blue Waters

Maxim Belkin
Education and Training Coordinator
mbelkin@illinois.edu
What are software containers?
What are software containers?

Modern applications have multiple dependencies…
What are software containers?

and when we move them to a different place, they can lose their dependencies and stop working.
What are software containers?

Software containers offer a solution!
What are software containers?

Software containers are light-weight Virtual Machines

Virtual Machines

<table>
<thead>
<tr>
<th>App #1</th>
<th>App #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>bins/libs</td>
<td>bins/libs</td>
</tr>
<tr>
<td>Guest OS 1</td>
<td>Guest OS 2</td>
</tr>
<tr>
<td>Hypervisor</td>
<td>Hypervisor</td>
</tr>
<tr>
<td>Host OS</td>
<td>Host OS</td>
</tr>
<tr>
<td>Hardware</td>
<td>Hardware</td>
</tr>
</tbody>
</table>

Containers

<table>
<thead>
<tr>
<th>App #1</th>
<th>App #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>bins/libs</td>
<td>bins/libs</td>
</tr>
<tr>
<td>Container Daemon</td>
<td>Container Daemon</td>
</tr>
<tr>
<td>Host OS</td>
<td>Host OS</td>
</tr>
<tr>
<td>Hardware</td>
<td>Hardware</td>
</tr>
</tbody>
</table>
Why software containers?

Software containers…

… provide a light-way solution for packaging applications together with their dependencies without the overhead of Virtual Machines.

… allow users substitute entire software stacks (except for Linux kernel).
Why do we need Software Containers on Blue Waters?

- Solve "dependency nightmare" problems
- Enable use of "certified" software stacks
- Enable new (for HPC) workflows
- Let users get up and running on Blue Waters fast!
Software containers: many solutions

- Docker
- Charliecloud
- UDOCKER
- Singularity
- podman
- Sarus
- ENROOT
- Shifter
- FreeBSD jails
- LXC
Dockerfiles and Docker images

Dockerfiles are recipes for creating containerized applications, e.g.:

```bash
FROM ubuntu:16.04
RUN apt-get update && apt-get install patchelf
WORKDIR /tmp
ENTRYPOINT ["/usr/bin/patchelf"]
```

The above Dockerfiles can be used to build a Docker image:

```
docker build -t patchelf:ub1604 -f Dockerfile .
```

Created Docker image can be used on any system with Docker installed like so:

```
docker run -v "$PWD":/tmp patchelf:ub1604 --print-needed filename
```
Software containers: Shifter

- Designed for HPC (users don't need root privileges)
- Works with public and private repositories on DockerHub
- Provides users with fine-grained access control over private images (user- and group-level control)
- Works with local images (no need for DockerHub)
- Integrates well with the Blue Waters system
Shifter: Software containers for HPC

To use Shifter in a job, request **shifter** generic resource:

```
qsub -l gres=shifter
```

or in a script:

```
#PBS -l gres=shifter
```

Once the jobs starts, access Shifter by loading its **module**

```
module load shifter
```
Shifter Module on Blue Waters

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>shifter</td>
<td>launch an application in a container</td>
</tr>
<tr>
<td>shifterimg</td>
<td>return information about available images</td>
</tr>
<tr>
<td>images</td>
<td>return ID of an UDI, if available</td>
</tr>
<tr>
<td>lookup</td>
<td>download specified Docker image and convert it to UDI (squashfs) format</td>
</tr>
<tr>
<td>login</td>
<td>Authenticate to DockerHub (for private repositories)</td>
</tr>
</tbody>
</table>
Shifter: Software containers for HPC

Once loaded, using **Shifter** is as easy as:

```
aprun -b ... \n    shifter --image=centos:6.9 -- \n    ./path/to/application.exe
```

Can use any image available on DockerHub ([hub.docker.com](http://hub.docker.com))!

or a local one:

```
aprun -b ... \n    shifter --image=custom:centos:6.10 -- \n    ./path/to/application.exe
```
Shifter: MPI Jobs: impact on start-up time

![Graph a](ppn = 1, 36 MB, 1.7 GB, CLI prologue)

- Node count: 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096
- Job start-up time, s: 1.7 GB
- CLI prologue

![Graph b](80 nodes, 36 MB, 1.7 GB, CLI prologue)

- Processes per node (ppn): 1, 2, 4, 8, 16
- Job start-up time, s: 1.7 GB
- CLI prologue
Shifter: MPI Jobs: impact on performance

(a) Time, µs vs. Message size, bytes
- Shifter CLE
- MPI_Alltoall
- MPI_Alltoallv

(b) Time, ms vs. Message size, kilobytes
- Shifter CLE
- MPI_Bcast
- MPI_Reduce

IOR Read, GB/s vs. Message size, bytes
- Shifter
- Cray Linux Environment

IOR Write, GB/s vs. Message size, bytes
- Shifter
- Cray Linux Environment
Shifter: Software Containers for HPC

bluewaters.ncsa.illinois.edu/shifter

Blue Water Symposium presentations:

- 2017: https://youtu.be/nsBqo__UZg8
- 2018: https://youtu.be/Zlm95VNeDBk

help+bw@ncsa.illinois.edu
Questions?