How to build and run Bigtop Sandbox (Experimental)

What is Bigtop Sandbox?

A handy tool to run big data pseudo clusters on Docker.

How to run

Make sure you have Docker installed. We've tested this using Docker for Mac

Currently supported OS list:

- debian-8
- ubuntu-16.04

Run Hadoop HDFS

```
docker run -d -p 50070:50070 bigtop/sandbox:1.2.1-ubuntu-16.04-hdfs
```

For HDFS, it takes around 30 secs. You can use docker logs to see whether it has been provisioned:

```
BIGTOP=$(docker run -d -p 50070:50070 bigtop/sandbox:1.2.1-ubuntu-16.04-hdfs)
```

```
docker logs -f $BIGTOP
```

```
Warning: This method is deprecated, please use the stdlib validate_legacy function, with Stdlib::Compat::Hash.
There is further documentation for validate_legacy function in the README.
   (at /etc/puppet/modules/stdlib/lib/puppet/functions/deprecation.rb:25:in `deprecation')
Warning: This method is deprecated, please use match expressions with Stdlib::Compat::Bool instead. They are
described at https://docs.puppet.com/puppet/latest/reference/lang_data_type.html#match-expressions.
   (at /etc/puppet/modules/stdlib/lib/puppet/functions/deprecation.rb:25:in `deprecation')
Warning: This method is deprecated, please use match expressions with Stdlib::Compat::Array instead. They are
described at https://docs.puppet.com/puppet/latest/reference/lang_data_type.html#match-expressions.
   (at /etc/puppet/modules/stdlib/lib/puppet/functions/deprecation.rb:25:in `deprecation')
Notice: Scope(Class[Node_with_components]): Roles to deploy: [namenode, datanode]
Warning: This method is deprecated, please use the stdlib validate_legacy function, with Pattern[]. There is
further documentation for validate_legacy function in the README.
   (at /etc/puppet/modules/stdlib/lib/puppet/functions/deprecation.rb:25:in `deprecation')
Warning: This method is deprecated, please use the stdlib validate_legacy function, with Stdlib::Compat::Bool.
There is further documentation for validate_legacy function in the README.
   (at /etc/puppet/modules/stdlib/lib/puppet/functions/deprecation.rb:25:in `deprecation')
Warning: This method is deprecated, please use the stdlib validate_legacy function, with Stdlib::Compat::
String. There is further documentation for validate_legacy function in the README.
   (at /etc/puppet/modules/stdlib/lib/puppet/functions/deprecation.rb:25:in `deprecation')
Warning: This method is deprecated, please use match expressions with Stdlib::Compat::Numeric instead. They are
described at https://docs.puppet.com/puppet/latest/reference/lang_data_type.html#match-expressions.
   (at /etc/puppet/modules/stdlib/lib/puppet/functions/deprecation.rb:25:in `deprecation')
Notice: Compiled catalog for 9c26fcceafad.local in environment production in 1.45 seconds
Notice: Baseurl: http://repos.bigtop.apache.org/releases/1.2.1/ubuntu/16.04/x86_64
Notice: /Stage[main]/Bigtop_repo/Notify[Baseurl: http://repos.bigtop.apache.org/releases/1.2.1/ubuntu/16.04
/x86_64]/message: defined 'message' as 'Baseurl: http://repos.bigtop.apache.org/releases/1.2.1/ubuntu/16.04
/x86_64'
Notice: /Stage[main]/Bigtop_repo/Exec[bigtop-apt-update]/returns: executed successfully
Notice: /Stage[main]/Hadoop::Common_hdfs/File[/etc/hadoop/conf/core-site.xml]/content: content changed '{md5}
71506958747641d1a5def83b021e7f75' to '{md5}ce32af59eb015a3bb3774d375be10f11'
Notice: /Stage[main]/Hadoop::Common_hdfs/File[/etc/hadoop/conf/hdfs-site.xml]/content: content changed '{md5}
784883dd654527ae577de19ecdec0992' to '{md5}ddc0a621878650832f30eb9690aa7565'
Notice: /Stage[main]/Hadoop::Namenode/Service[hadoop-hdfs-namenode]/ensure: ensure changed 'stopped' to
'running'
Notice: /Stage[main]/Hadoop::Datanode/File[/data/1/hdfs]/mode: mode changed '0700' to '0755'
Notice: /Stage[main]/Hadoop::Datanode/File[/data/2/hdfs]/mode: mode changed '0700' to '0755'
Notice: /Stage[main]/Hadoop::Datanode/Service[hadoop-hdfs-datanode]/ensure: ensure changed 'stopped' to
'running'
Notice: /Stage[main]/Hadoop::Init_hdfs/Exec[init hdfs]/returns: executed successfully
Notice: Finished catalog run in 29.46 seconds
```

After provisioned, goto http://localhost:50070, you'll see the web UI is ready there.

To destroy the container:

```
docker stop $BIGTOP

docker rm $BIGTOP
```

Run Hadoop HDFS + HBase

```
BIGTOP=$(docker run -d -p 50070:50070 -p 16010:16010 bigtop/sandbox:1.2.1-ubuntu-16.04-hdfs_hbase)

docker exec -ti $BIGTOP hbase shell
```

Run Hadoop HDFS + Spark Standalone

```
BIGTOP=$(docker run -d -p 50070:50070 -p 8080:8080 bigtop/sandbox:1.2.1-ubuntu-16.04-hdfs_spark-standalone)

docker exec -ti $BIGTOP spark-shell
```

Run Hadoop HDFS + YARN + Hive + Pig

```
BIGTOP=$(docker run -d -p 50070:50070 -p 8088:8088 bigtop/sandbox:1.2.1-ubuntu-16.04-hdfs_yarn_hive_pig)

docker exec -ti $BIGTOP hive

docker exec -ti $BIGTOP pig
```

How to build

Download Bigtop

Go to http://bigtop.apache.org/download.html#releases and download the latest bigtop release. After downloaded:

```
tar zxvf bigtop-1.2.1-project.tar.gz
cd bigtop-1.2.1/docker/sandbox
```

Build a Hadoop HDFS sandbox image

```
./build.sh -a bigtop -o ubuntu-16.04 -c hdfs
```

Build a Hadoop HDFS, Hadoop YARN, and Spark on YARN sandbox image

```
./build.sh -a bigtop -o ubuntu-16.04 -c "hdfs, yarn, spark"
```

Build a Hadoop HDFS and HBase sandbox image

```
./build.sh -a bigtop -o ubuntu-16.04 -c "hdfs, hbase"
```

Use --dryrun to skip the build and get Dockerfile and configuration

```
./build.sh -a bigtop -o ubuntu-16.04 -c "hdfs, hbase" --dryrun
```

Change the repository of packages

```
export REPO=http://repos.bigtop.apache.org/releases/1.2.1/debian/8/x86_64
./build.sh -a bigtop -o ubuntu-16.04 -c "hdfs, yarn, ignite"
```

Customize your Big Data Stack

```
vim site.yaml.template.debian-8_hadoop # Configure your own stack

./build.sh -a bigtop -o debian-8 -f site.yaml.template.debian-8_hadoop -t my_hadoop_stack
```

Known issues

Fail to start daemons using systemd

Since systemd requires CAP_SYS_ADMIN, currently any OS using systemd can not successfully started up daemons during image build time.

Daemons can be brought up only if --privileged specified using docker run command.

Reference

Available Sandboxes: https://hub.docker.com/r/bigtop/sandbox/tags/

Build status: https://ci.bigtop.apache.org/view/Docker/job/Docker-Sandbox/

DataWorks Summit 2017 slide: https://www.slideshare.net/saintya/leveraging-docker-for-hadoop-build-automation-and-big-data-stack-provisioning