Dockerized Hadoop Platform and Recent Updates in Apache Bigtop

Yu-hsin Yeh (Evans Ye)
Apache Big Data NA 2016
Vancouver
Outline

• Bigtop Provisioner
• Integrating Docker Compose
• Integrating Docker Machine & Swarm
• Image Pre-build
• Project updates and PPC porting
Who am I

• Apache Bigtop PMC member
• Software Engineer @ Trend Micro
• Develop big data apps & infra
Bigtop Provisioner
Bigtop Provisioner

• A tool to demonstrate the full life cycle of Bigtop
Goal

- **Fast iterative development**
  - Test your code in the cluster, on your laptop, w/o human intervention

- **Flexibility**
  - Choose any combination of components you want

- **Responsive CI**
  - Integration tests that get you the result in mins

- **A Big Data Stack playground**
  - Spark + Tachyon, Spark + Ignite, Apex, etc
Vagrant + Automation Code + Bigtop Puppet = One-click Hadoop Provisioning
• We use Vagrant as an abstraction layer to support different kind of resource providers
One click Hadoop provisioning

./docker-hadoop.sh -c 3
One click Hadoop provisioning

bigtop/deploy image on Docker hub

VAGRANT

./docker-hadoop.sh -c 3
One click Hadoop provisioning

bigtop/deploy image on Docker hub

puppet apply

puppet apply

puppet apply

/docker-hadoop.sh -c 3
Bigtop/deploy Images

- **bigtop/deploy**
- Install Vagrant ssh key
- Install Puppet

Dockerhub official images
Bigtop Provisioner

- Supported providers in Bigtop 1.1.0 release
  - VirtualBox VM
  - Docker container
  - OpenStack
Use Cases

- For Hadoop app developers, cluster admins, users
  - Run a Hadoop cluster to test your code on
  - Try & test configurations before applying to Production
  - Play around with Bigtop Big Data Stack

- For contributors
  - Easy to test your packaging, deployment, testing code

- For vendors
  - CI out of the box —> patch upstream code made easier
Integrating Docker Compose
What’s the problem with Vagrant’s Docker Provider?

- Need to add vagrant public key into docker images
- Too many issues with auto-created boot2docker hosting VM
- A bug for docker provider keep opening for almost 2y
  - *Waiting for machine to boot’ hangs infinitely*
- Can not share same code for different providers anyway
- Not all the docker options supported in Vagrantfile
- Does not support Docker Swarm
- Slow
Docker Compose
Docker Compose

```
kafka:
  build: .
  ports:
    - "9092:9092"

spark:
  image: spark
  port:
    - "8080:8080"
```

up / stop / kill / rm
Integration Details

- Create docker containers:
  - `docker-compose scale bigtop=3`

- Volmes:
  - Bigtop Puppet configurations
  - Bigtop Puppet code
  - `/etc/hosts`

- privileged: true
Integrating with Docker Compose

./docker-hadoop.sh --create 3
$ docker-compose scale bigtop_local=3

Supported OS image on Docker hub

![Diagram showing supported OS image on Docker hub, followed by scaling with `docker-compose scale bigtop_local=3` command.](image)
$ docker inspect --format "{{.NetworkSettings.IPAddress}} {{.Config.Hostname}}.{{.Config.Domainname}}"
$ docker exec $node bash -c "./puppetize.sh"
$ puppet apply
bigtop-deploy/puppet/manifests/site.pp
Finished

Supported OS image on Docker hub
Advantages

• No need to create customised image beforehand
• Better compatibility with Docker’s native solutions
• Clear, simple yaml file for orchestration settings
• Has the **scale** option to easily scale up/down
• Native support to Docker Swarm
• Supports new features such as overlay network and named volume
• Fast —> better user experience
Integrating Docker Machine & Swarm
Docker Machine
Let you create Docker hosts

- On local machine
  - virtualbox, vmwarefusion
- On the cloud
  - amazonec2, azure, digitalocean, exoscale, google, rackspace, softlayer
- Inside your own datacenter
  - generic, openstack, vmwarevcloudair, vmwarevvsphere
Key Features

• The following are auto-configured:
  • Docker engine install/upgrade
  • TLS encryption & authentication
  • Certificate generation & key signing
  • Swarm
Docker Swarm
Docker Swarm
Resource Management

• Filters
  • Memory, CPU, Network
    • docker run -m 1g -c 1 -p 80:80
  • Scheduling strategies
    • spread, binpack, random
High Availability

Replica

Replica

Forward

Primary

BOOM!
Integration Issues

• Swarm with overlay network has basic service discovery:

  • IP to container name mapping are auto-generated, maintained in each container’s /etc/hosts

• However, they’re invalid hostname with underscores:

<table>
<thead>
<tr>
<th>IP Address</th>
<th>Hostname</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0.0.2</td>
<td>20160507170202rdm10393_bigtop_swarm_3</td>
</tr>
<tr>
<td>10.0.0.2</td>
<td>20160507170202rdm10393_bigtop_swarm_3.bigtop</td>
</tr>
<tr>
<td>10.0.0.3</td>
<td>20160507170202rdm10393_bigtop_swarm_1</td>
</tr>
<tr>
<td>10.0.0.3</td>
<td>20160507170202rdm10393_bigtop_swarm_1.bigtop</td>
</tr>
<tr>
<td>10.0.0.4</td>
<td>20160507170202rdm10393_bigtop_swarm_2</td>
</tr>
<tr>
<td>10.0.0.4</td>
<td>20160507170202rdm10393_bigtop_swarm_2.bigtop</td>
</tr>
</tbody>
</table>

  • On the system, the hostname is still a hash

  / # hostname -f
  e5954c3e9b0c

• Hostname can be configured in config file, but it’s not working with **scale** —→ each scaled container gets same hostname
Integration details

- Use Docker Machine to create a Swarm cluster

- Instead of sharing volumes, use `docker cp` to copy files:
  - Bigtop Puppet configurations
  - Bigtop Puppet code
  - `/etc/hosts`
  - privileged: true

- Use overlay network
• docker-machine create -d virtualbox kvstore

• eval $(docker-machine env kvstore)

• docker run -d -p 8500:8500 --name=consul progrium/consul -server -bootstrap
Swarm on Virtualbox

- `docker-machine create -d virtualbox
  --swarm
  --swarm-master
  --swarm-discovery="consul://$(docker-machine ip kvstore):8500"
  --engine-opt="cluster-store=consul://$(docker-machine ip kvstore):8500"
  --engine-opt="cluster-advertise=eth1:2376"
swarm-master`
Swarm on Virtualbox

- docker-machine create -d virtualbox
  --swarm
  --swarm-discovery="consul://$(docker-machine ip kvstore):8500"
  --engine-opt="cluster-store=consul://$(docker-machine ip kvstore):8500"
  --engine-opt="cluster-advertise=eth1:2376"
  swarm-slave
Integrating with Docker Machine & Swarm

./docker-hadoop.sh --swarm --create 3
$ ./docker-hadoop.sh --swarm

Key-Value Store

Swarm Master

Swarm Slave
$ docker-compose scale bigtop_swarm=3

Supported OS image on Docker hub

Key-Value Store

Swarm Master

Swarm Slave
$ docker inspect --format "{{.NetworkSettings.Networks.$OVERLAY.IPAddress}}
{{.Config.Hostname}}.{{.Config.Domainname}}"
$ docker exec $node bash -c "./puppetize.sh"
$ puppet apply
bigtop-deploy/puppet/manifests/site.pp
$ puppet apply
bigtop-deploy/puppet/manifests/site.pp
Advantages

• Run Hadoop cluster on Docker anywhere

• Provision a fully distributed, multi-host Docker based Hadoop cluster
Not recommended for Production
Image Pre-build
5 mins provisioning time is too long!

You’re damned right.

An Apache guy

Bigtoper
Idea

• For users, they don’t want to build/test packages, they just want a cluster

• Docker images are immutable

• RPM/DEB packages are also immutable

• Let’s build an image that preload all the packages
Image Pre-build

bigtop/deploy:prebuild
Seriously?
Second Thought

- Each company has its own Big Data Stack

- TM Hadoop = Hadoop + HBase + Pig + Oozie
Second Thought

• Within same company, there might be multiple stacks serving different purpose

• Product Specific Platform = Spark + Docker + Akka + Cassandra + Kafka
Image Pre-build

components: [hadoop, yarn]

Dockerhub official images

bigtop/deploy:prebuild

Hadoop packages

Yarn packages
Provision a HDFS cluster took
2m43s → 0m50s
226% Faster
Better CI
New Components

• New in Bigtop 1.1 release Feb, 2016
  • Apache Hama 0.7.0
  • Apache Tajo 0.11.1
  • Apache Zeppelin 0.5.6

• New in Bigtop master branch
  • Apache Apex 3.3.0
  • QFS 1.1.4
  • Apache Flink - BIGTOP-1927, PR available
Amir Sanjar
OpenPower Foundation Member
Senior Software Engineer, IBM Power Systems Software & Solutions
What is OpenPower Foundation?

OpenPOWER is an open development community, using the POWER Architecture to serve the evolving needs of customer
OpenPOWER Foundation reach

200+ members
60+ technologies revealed

24 countries
6 continents
100s innovations under way
Why OpenPower Foundation?
Moore’s law no longer satisfies performance gain
Growing workload demands more performance
Solution?
Accelerated Technology
## Accelerated Technology roadmap

<table>
<thead>
<tr>
<th>Year</th>
<th>IBM Systems</th>
<th>Mellanox</th>
<th>Nvidia</th>
<th>IBM CPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Connect-IB FDR Infiniband PCIe Gen3</td>
<td>ConnectX-4 EDR Infiniband CAPI over PCIe</td>
<td>Kepler PCIe Gen3</td>
<td>POWER8 OpenPower CAPI Interface</td>
</tr>
<tr>
<td>2016</td>
<td>ConnectX-5 Next-Gen Infiniband Enhanced CAPI over PCIe</td>
<td>Gen3 Pascal NVLink</td>
<td>ConnectX-4 EDR Infiniband PCIe Gen3</td>
<td>POWER8 with NVLink</td>
</tr>
<tr>
<td>2017</td>
<td>ConnectX-5 Next-Gen Infiniband Enhanced CAPI over PCIe</td>
<td>Gen3 Pascal NVLink</td>
<td>Volta Enhanced NVLink</td>
<td>POWER9 Enhanced CAPI &amp; NVLink</td>
</tr>
</tbody>
</table>
OpenPOWER Technology: 2.5x Faster CPU-GPU Connection via NVLink

GPUs Bottlenecked by PCIe Bandwidth From CPU-System Memory

NVLink Enables Fast Unified Memory Access between CPU & GPU Memories
System Performance of Apache Bigtop Spark 1.5.1 on POWER Spark Workloads

Relative System Performance

<table>
<thead>
<tr>
<th>Machine Learning</th>
<th>SQL</th>
<th>Graph</th>
<th>AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5-2620 v3</td>
<td></td>
<td></td>
<td>1.7X</td>
</tr>
<tr>
<td>100GB Mat. Fact.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100GB (in mem)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1TB (in mem)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1TB (50/50) LR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1TB SVM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10TB LR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1TB 5 query</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2TB 5 query</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13GB Page Rank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1TB Triangle Cnt.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1TB SVD++</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Price Performance of Apache Bigtop Spark on POWER

- Spend 33% less on infrastructure supporting the same amount of workload
- Spend the same on infrastructure but host 50% more workload

* - based on preliminary SoftLayer pricing targets - subject to change
Accelerated Spark Demo

*Adverse Drug Reaction Prediction*

*Bigtop Spark 1.5.1*

*NVidia GPU*
§ Apache BigTop distribution for Power

- Port Bigtop stack to Power
- Build Bigtop stack for Power
Porting to Power has become effortless.

• Advent of OpenJDK for POWER
  • No porting required
  • %100 compatible with Oracle JAVA.
• Power8 with Little endian
  • No porting, just recompile native (C/C++) libraries
• Ported 22 out of 24 Apache Big Top stacks to POWER in two week.
Build Bigtop Distribution for Power

§ Build of Apache big data projects are not for the faint-hearted.
§ Many build dependencies
§ Various development tools
§ Many Linux distribution to support
Preparing for Build Hadoop

<table>
<thead>
<tr>
<th>On all systems</th>
<th>Also on RPM-based systems</th>
<th>Also on DEB-based systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Java JDK 1.6</td>
<td>lzo-devel</td>
<td>libxlsx1-dev</td>
</tr>
<tr>
<td>Apache Ant</td>
<td>zlib-devel</td>
<td>libkrb5-dev</td>
</tr>
<tr>
<td>Apache Maven</td>
<td>fuse-devel</td>
<td>libldap2-dev</td>
</tr>
<tr>
<td>wget</td>
<td>openssl-devel</td>
<td>libmysqlclient-dev</td>
</tr>
<tr>
<td>tar</td>
<td>python-devel</td>
<td>libsqlite-dev</td>
</tr>
<tr>
<td>git</td>
<td>libxml2-devel</td>
<td>libxml2-dev</td>
</tr>
<tr>
<td>subversion</td>
<td>libxslt-devel</td>
<td>python-dev</td>
</tr>
<tr>
<td>gcc</td>
<td>cyrus-sasl-devel</td>
<td>python-setuptools</td>
</tr>
<tr>
<td>gcc-c++</td>
<td>sqlite-devel</td>
<td>libbz2-dev</td>
</tr>
<tr>
<td>make</td>
<td>mysql-devel</td>
<td>libfuse-dev</td>
</tr>
<tr>
<td>fuse</td>
<td>openldap-devel</td>
<td>libssl-dev</td>
</tr>
<tr>
<td>protobuf-compiler</td>
<td>rpm-build</td>
<td>build-essential</td>
</tr>
<tr>
<td>autoconf</td>
<td>createrepo</td>
<td>dh-make</td>
</tr>
<tr>
<td>automake</td>
<td></td>
<td>debhelper</td>
</tr>
<tr>
<td>libtool</td>
<td></td>
<td>devscripts</td>
</tr>
<tr>
<td>sharutils</td>
<td></td>
<td>reprepro</td>
</tr>
<tr>
<td>xmito</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What Apache BigTop offers?

Apache Bigtop has Dockerized the entire build environment.

The entire build environment in a Docker images
Available on Docker Hub

Porting Bigtop Docker images to Power was achieved without great effort.

Made possible with only five patches
Three Dockerfiles and three scripts

Dockerfile

Dockerfile

Dockerfile

Dockerfile

Dockerfile

Dockerfile

manifests/env.pp

manifests/protobuf.pp

puppetize.sh
Example

FROM ppc64le/ubuntu:15.04
MAINTAINER Amir Sanjar
COPY puppetize.sh /tmp/
puppetize.sh
RUN bash /tmp/puppetize.sh
One click build

$ git clone https://github.com/apache/bigtop.git
$ docker run -v `pwd`:/ws bigtop/slaves:trunk-ubuntu-15.04-ppc64le \
  bash -l -c 'cd /ws ; ./gradlew hadoop-deb'
One click install

$ sudo apt-get install spark-master
<table>
<thead>
<tr>
<th>Package</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigtop-groovy</td>
<td>2.4.4</td>
</tr>
<tr>
<td>bigtop-jsvc</td>
<td>1.0.15</td>
</tr>
<tr>
<td>bigtop-tomcat</td>
<td>6.0.36</td>
</tr>
<tr>
<td>bigtop-utils</td>
<td>1.1.0</td>
</tr>
<tr>
<td>crunch</td>
<td>0.12.0</td>
</tr>
<tr>
<td>datafu</td>
<td>1.0.0</td>
</tr>
<tr>
<td>flu</td>
<td>1.6.0</td>
</tr>
<tr>
<td>gir</td>
<td>1.1.0</td>
</tr>
<tr>
<td>hama</td>
<td>2.7.1</td>
</tr>
<tr>
<td>hbase</td>
<td>0.7.0</td>
</tr>
<tr>
<td>kafka</td>
<td>0.8.1.1</td>
</tr>
<tr>
<td>kite</td>
<td>1.1.0</td>
</tr>
<tr>
<td>mahout</td>
<td>0.11.0</td>
</tr>
<tr>
<td>oozie</td>
<td>4.2.0</td>
</tr>
<tr>
<td>phoenix</td>
<td>4.6.0</td>
</tr>
<tr>
<td>pig</td>
<td>0.15.0</td>
</tr>
<tr>
<td>solr</td>
<td>1.99.4</td>
</tr>
<tr>
<td>sqark</td>
<td>1.4.5</td>
</tr>
<tr>
<td>sqoop</td>
<td>1.99.4</td>
</tr>
<tr>
<td>sqoop2</td>
<td>1.99.4</td>
</tr>
<tr>
<td>tachyon</td>
<td>0.6.0</td>
</tr>
</tbody>
</table>

Apache BigTop Release 1.1.0 For OpenPower
Apache BigTop CI

- CentOS
- Fedora
- Ubuntu
- Debian
- OpenSuSE

Jenkins
Thank you.