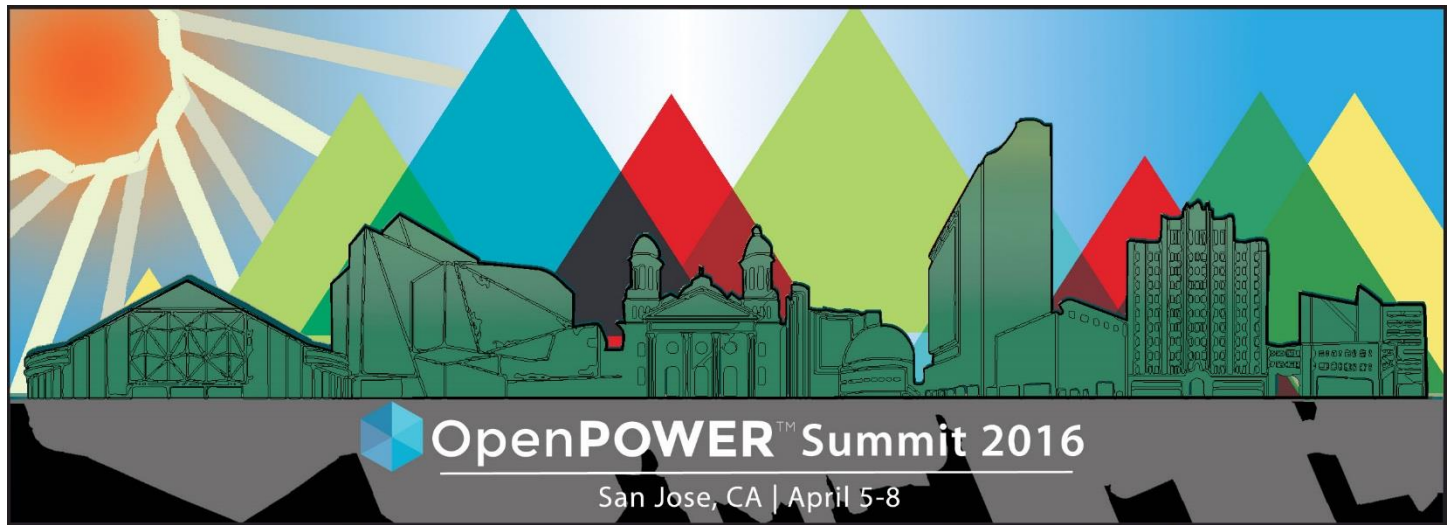# Apache Bigtop, Bringing Big Data to Power8

Nate D'Amico – Apache BigTop PMC member
Amir Sanjar - OpenPOWER foundation member
Konstantin Boudnik – Apache BigTop PMC member

Revolutionizing the Datacenter

OpenPOWER™ Summit 2016

San Jose, CA | April 5-8

Join the Conversation #OpenPOWERSummit

Agenda

- Apache Bigtop Overview
- Integration of OpenPOWER with Apache BigTop
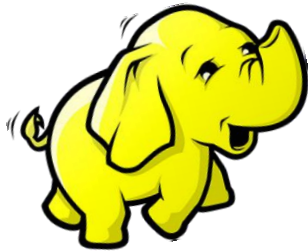 - Port big data stack to OpenPOWER
  - Build big data stack for OpenPOWER

What is Bigtop?

▪Apache Bigtop is setting the standard for testing, packaging and integration of leading big/fast data components

Components as Building Blocks

And many others…

Solving Dependency Hell

	ubuntu	suse	redhat		debian
hdfs					
zookeeper					
hbase					
kafka					
spark					
.					
.					
.					
mapred					
oozie					
hive					
etc					

Bigtop Early Mission Accomplished
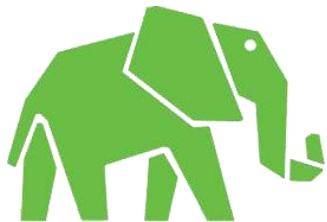
All major Hadoop distros leverage Bigtop to build its foundation

Looking at Road Ahead

▪Data engineers vs
distro builders

▪Enhance Operations &
Deployment

▪Reference
implementation &
tutorials

Port to OpenPOWER

- Porting to POWER has become effortless
  - Advent of OpenJDK for POWER
    - No porting required
    - %100 compatible with Oracle JAVA
  - POWER with Little endian
    - No porting, just recompile native (C/C++) libraries
    - Examples, Canonical port and tested 40,000 packages in 4 months
    - Ported 22 out of 24 Apache Big Top stacks to POWER in two week.

Build for OpenPOWER

- Build of Apache projects are not for the faint-hearted.

- Many build dependencies

- Various development tools

- Many Linux distribution to support

Preparing for Build Hadoop for OpenPOWER & x86

On all systems	Also on RPM-based systems	Also on DEB-based systems
• Java JDK 1.6	• lzo-devel	• libxslt1-dev
• Apache Ant	• zlib-devel	• libkrb5-dev
• Apache Maven	• fuse-devel	• libldap2-dev
• wget	• openssl-devel	• libmysqlclient-dev
• tar	• python-devel	• libsasl2-dev
• git	• libxml2-devel	• libsqlite3-dev
• subversion	• libxslt-devel	• libxml2-dev
• gcc	• cyrus-sasl-devel	• python-dev
• gcc-c++	• sqlite-devel	• python-setuptools
• make	• mysql-devel	• liblzo2-dev
• fuse	• openldap-devel	• libzip-dev
• protobuf-compiler	• rpm-build	• libfuse-dev
• autoconf	• createrepo	• libssl-dev
• automake	• redhat-rpm-config (RedHat/CentOS only)	• build-essential
• libtool		• dh-make
• sharutils		• debhelper
• xmlto		• devscripts
		• reprepro

What Apache BigTop offers?

▪Apache Bigtop dockerized the entire build environment.

  –Ship the entire build environment for OpenPOWER in Docker Images

  ▪Available on Docker Hub

  ▪Made possible *only by five* patches

  ▪https://github.com/apache/bigtop/commit/ac04aaf6705db42f092fe0194c82ee8d0f58cc15

  ▪https://github.com/apache/bigtop/blob/master/docker/bigtop-slaves/ubuntu-15.04-ppc64le/Dockerfile

  –

One click to build packages

```
$ 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 to install packages

$ sudo apt-get install spark-master

Apache BigTop 1.1.0 for OpenPower

bigtop-groovy	**2.4.4**
bigtop-jsvc	**1.0.15**
bigtop-tomcat	**6.0.36**
bigtop-utils	**1.1.0-SNAPSHOT**
crunch	**0.12.0**
datafu	**1.0.0**
flume	**1.6.0**
giraph	**1.1.0**
hadoop	**2.7.1**
hama	**0.7.0**
hbase	**0.98.12**
hive	**1.2.1**
hue	**3.9.0**
ignite-hadoop	**1.5.0.final**
kafka	**0.8.1.1**
kite	**1.1.0**
mahout	**0.11.0**
oozie	**4.2.0**
phoenix	**4.6.0-HBase-0.98**
pig	**0.15.0**
solr	**4.9.0**
spark	**1.5.1**
sqoop	**1.4.5**
sqoop2	**1.99.4**
tachyon	**0.6.0**
tez	**0.6.2**
ycsb	**0.4.0**
zeppelin	**0.5.5**
zookeeper	**3.4.6**

Flexible build & test

Linux

Apache Bigtop CI

CentOS

Fedora

Ubuntu

Debian

OpenSuSE

Jenkins