

APACHE  CON

DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

Apache Bigtop: a crash course in deploying a Hadoop platform

Presented For The Apache Foundation By
 **LINUX FOUNDATION**

APACHE CON

DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

“3 + 7 + 9 + 1”

The complexity of the stack

Presented For The Apache Foundation By
LINUX FOUNDATION

Hadoop ecosystem

- Many components
- Gazillions of versions
- Lot of patches if you like it hot and dirty



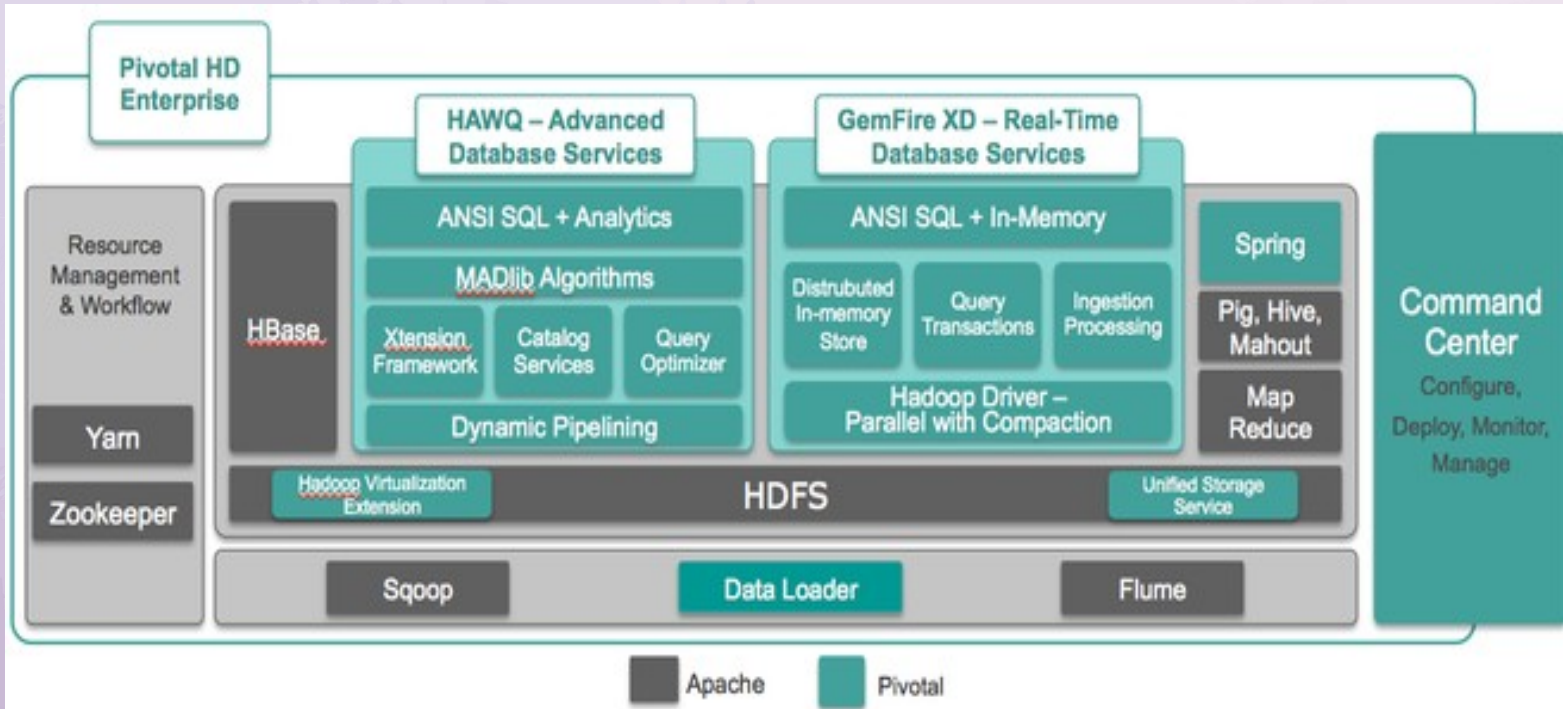
Presenters

- Dr. Konstantin Boudnik, Roman Shaposhnik
- Initial co-inventors of Apache Bigtop
- Active contributors, committers, and PMCs on multiple Apache TLPs in Hadoop ecosystem
- Expertise in
 - Compilers
 - Operating systems, Virtual machines, JVM
 - Distributed systems
 - Complex software stacks

Hadoop \neq HDFS + MR

- Today's Hadoop is more than storage + MR
- Baby elephant outgrew his cradle
 - Hbase
 - SQL frontends
 - In-memory processing
 - Storage caching
 - Connectors
 - DSL languages
 - <your name is here>

Complexity in the extreme



APACHE CON

DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

“One to bring them all”
Bigtop is a simple answer

Presented For The Apache Foundation By
LINUX FOUNDATION

Bigtop stack: take it & go

- Modify a stack BOM
 - Build
 - Deploy
 - Configure w/ Puppet (included)
 - Test (scenarios are provided / easy to add)
 - Grab an appliance if short of hardware

Rinse and repeat



APACHE  CON
DENVER
WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

Presented For The Apache Foundation By
 **LINUX FOUNDATION**

APACHE CON

DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

Field case studies: Pivotal WANdisco

Presented For The Apache Foundation By
LINUX FOUNDATION

APACHE CON

DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

From 0 to full stack in 28 days: WANdisco case study

Presented For The Apache Foundation By
LINUX FOUNDATION

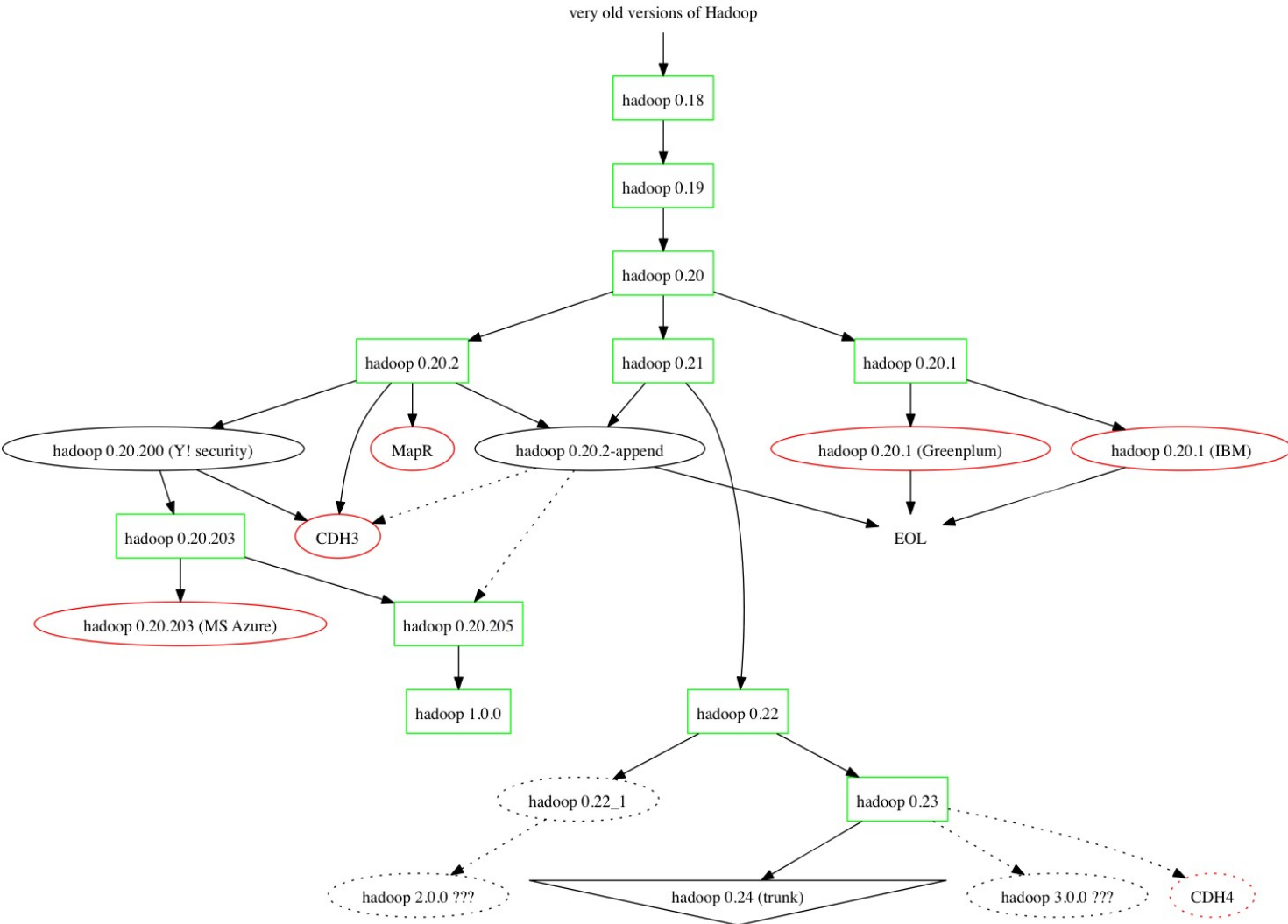
New distro in 4 weeks

- Major challenges for a new player:
 - Need to have a stable development platform
 - Offering Apache Hadoop certified binaries
 - Team with no prior expertise in the field
 - Very complex Hadoop ecosystem landscape

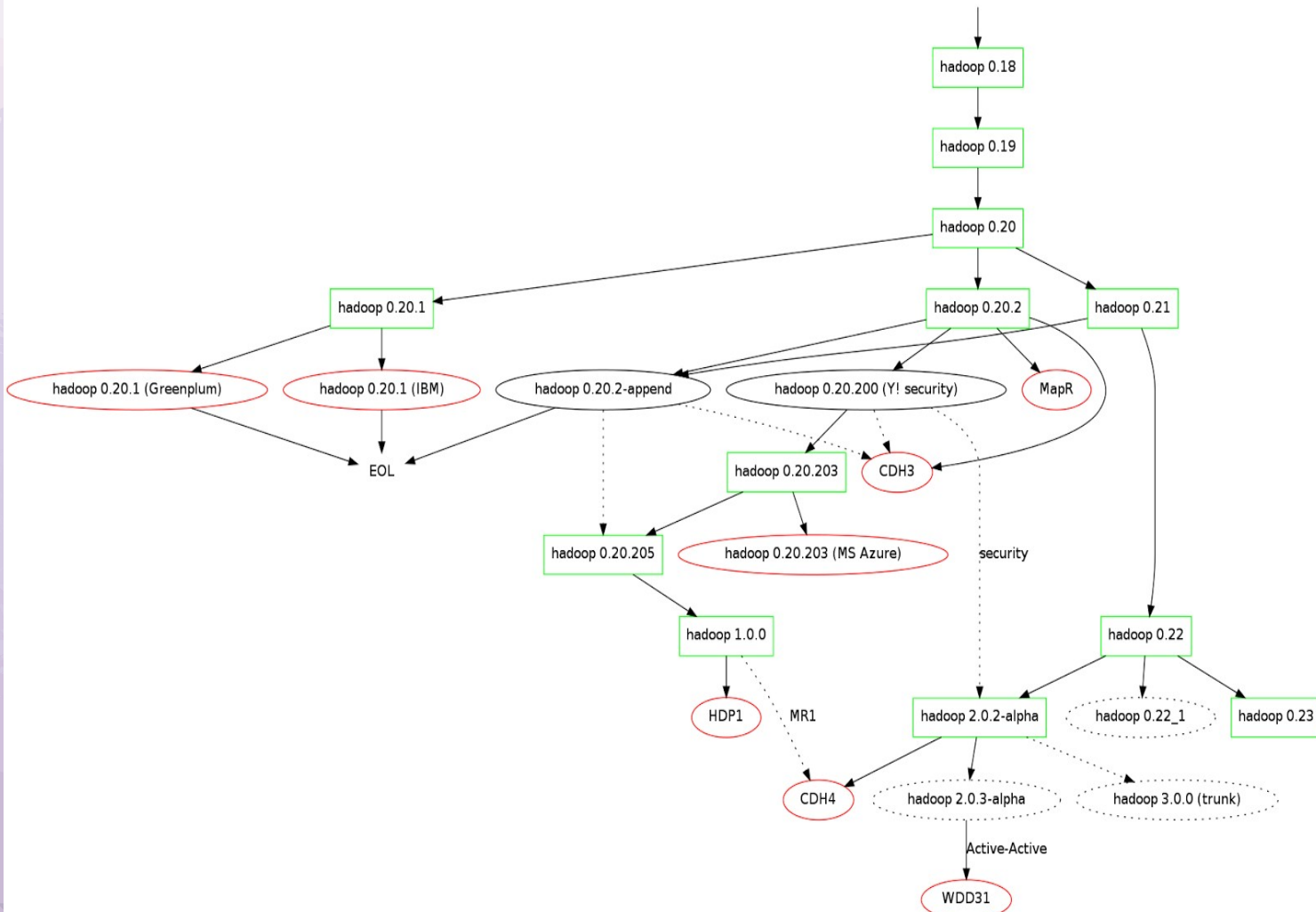


Bigtop to help

- Define component versions in the BOM
- Make component changes if needed
 - Run Bigtop build
 - Deploy the cluster w/ provided puppet
 - Test the cluster with integration suite
 - Rinse and repeat as needed
- Seamless integration into CI
- Easy provisioning and incremental updates



very old versions of Hadoop



APACHE  CON

DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

Hadoop in the cloud Pivotal case study

Presented For The Apache Foundation By
 **LINUX FOUNDATION**

Hadoop @Pivotal: history



- Started at Greenplum (GPHD)
 - A stand-alone Hadoop distribution
 - Based on a fork of Bigtop 0.3-incubating
 - Hadoop 1.0.1 based ecosystem
 - No community interaction
- Graduated as a PHD at Pivotal
 - A PivotalONE vision
 - Based on a fork of Bigtop 0.4
 - Hadoop 2.x based ecosystem
 - Integrated with HAWQ

Lessons for Pivotal

- Think of Bigtop as “Fedora”
- Become part of the community
- Say “no” to forking
- Work on custom requirements upstream
- Participate in release planning
- Leverage Bigtop's infrastructure internally



Lessons for Bigtop

- Promote common build and RE infrastructure
 - “Codifying” it
 - Avoid broken windows syndrome
- Bigtop releases don't patch, but vendors do
- Make tests easier to use
- Invest in documentation
 - Whitepapers, demos, etc.



APACHE  CON

DENVER

WESTIN DENVER DOWNTOWN

APRIL 7-9, 2014

Road ahead

Presented For The Apache Foundation By
 **LINUX FOUNDATION**

Packaging/deployment

- Deployment environments:
 - vanilla servers == packages ?
 - vanilla Vms/containers == JeOS/baking ?
 - specializes VMs == Osv
- Baking vs frying
- Rolling upgrades
 - side-by-side install



Validation

- Investing in iTest
 - Easier test management and execution
 - Cluster discovery and management framework
 - More test cases
- Lowering entry-level barriers
 - Less rigid user-facing interfaces: Gradle
 - Mix-n-match built-in integration tests
- Becoming a TCK for Hadoop ecosystem
- Engaging more into trunk testing

Growing the ecosystem

- 1st Hadoop distribution including Apache Spark
- But there's more:
 - GridGain
 - HBase indexer
 - Lipstick (not for Pig)
 - Ambrose
 - Launch it all to the Stratosphere?



Growth in last 6 months

- Apache Spark: In-memory analytics
- Phoenix: Hbase SQL frontend
- Groovy
- Unification of user-facing interfaces
 - Gradle build system
- Received proposal in include GridGain in-memory platform



Community

- 23 committers
- Over 100 contributors
- Quarterly releases
- Hackathons and more



Demo (worth 1k words)



WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

```
ZOOKEEPER_DOWNLOAD_PATH=/zookeeper/zookeeper-$(ZOOKEEPER_BASE_VERSION)
ZOOKEEPER_SITE=$(APACHE_MIRROR)$(ZOOKEEPER_DOWNLOAD_PATH)
ZOOKEEPER_ARCHIVE=$(APACHE_ARCHIVE)$(ZOOKEEPER_DOWNLOAD_PATH)
$(eval $(call PACKAGE,zookeeper,ZOOKEEPER))

# HBase
HBASE_NAME=hbase
HBASE_RELNOTES_NAME=Apache HBase
HBASE_PKG_NAME=hbase
HBASE_BASE_VERSION=0.94.12
HBASE_PKG_VERSION=$(HBASE_BASE_VERSION)
HBASE_RELEASE_VERSION=1
HBASE_TARBALL_DST=hbase-$(HBASE_BASE_VERSION).tar.gz
HBASE_TARBALL_SRC=$(HBASE_TARBALL_DST)
HBASE_DOWNLOAD_PATH=/hbase/hbase-$(HBASE_BASE_VERSION)
HBASE_SITE=$(APACHE_MIRROR)$(HBASE_DOWNLOAD_PATH)
HBASE_ARCHIVE=$(APACHE_ARCHIVE)$(HBASE_DOWNLOAD_PATH)
$(eval $(call PACKAGE,hbase,HBASE))

56,1 13%
2 bash 11:06:13
-version, spark_vardefines]
phoenix
 [phoenix-clean, phoenix-deb, phoenix-download, phoenix-help, phoenix-info, phoenix-relnotes, phoenix-rpm, phoenix-sdeb, phoenix-srpm
, phoenix-tar, phoenix-version, phoenix_vardefines]
bigtop-groovy
 [bigtop-groovy-clean, bigtop-groovy-deb, bigtop-groovy-download, bigtop-groovy-help, bigtop-groovy-info, bigtop-groovy-relnotes, big
top-groovy-rpm, bigtop-groovy-sdeb, bigtop-groovy-srpm, bigtop-groovy-tar, bigtop-groovy-version, bigtop-groovy_vardefines]
bigtop-utils
 [bigtop-utils-clean, bigtop-utils-deb, bigtop-utils-download, bigtop-utils-help, bigtop-utils-info, bigtop-utils-relnotes, bigtop-ut
ils-rpm, bigtop-utils-sdeb, bigtop-utils-srpm, bigtop-utils-tar, bigtop-utils-version, bigtop-utils_vardefines]
bigtop-jsvc
 [bigtop-jsvc-clean, bigtop-jsvc-deb, bigtop-jsvc-download, bigtop-jsvc-help, bigtop-jsvc-info, bigtop-jsvc-relnotes, bigtop-jsvc-rpm
, bigtop-jsvc-sdeb, bigtop-jsvc-srpm, bigtop-jsvc-tar, bigtop-jsvc-version, bigtop-jsvc_vardefines]
bigtop-tomcat
 [bigtop-tomcat-clean, bigtop-tomcat-deb, bigtop-tomcat-download, bigtop-tomcat-help, bigtop-tomcat-info, bigtop-tomcat-relnotes, big
top-tomcat-rpm, bigtop-tomcat-sdeb, bigtop-tomcat-srpm, bigtop-tomcat-tar, bigtop-tomcat-version, bigtop-tomcat_vardefines]
BUILD SUCCESSFUL

Total time: 7.319 secs
<11.05:54 cos@tcos: bigtop/ BIGTOP-1201 (587)>
3 root@vmhost12-vm5:~/work/bigtop bash 11:06:13
0-!$ bash 1$ bash 2$ bash 3$* bash 4$ bash
```

Come & join us



<https://bigtop.apache.org/>

<https://blogs.apache.org/bigtop>

<https://cwiki.apache.org/confluence/display/BIGTOP>

dev@bigtop.apache.org

user@bigtop.apache.org

@ASFbigtop





APACHE CON
DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

Q & A

Presented For The Apache Foundation By
 **LINUX FOUNDATION**

APACHE CON

DENVER

WESTIN DENVER DOWNTOWN
APRIL 7-9, 2014

Thank you
@c0sin
@rhatr

Presented For The Apache Foundation By
LINUX FOUNDATION