Hadoop.next
Who are we?

- Hadoop downstream community
- Well, specifically:
  - Roman Shaposhnik
    - VP, Apache Bigtop, IPMC member at ASF
    - Software engineer, Cloudera inc.
  - Dr. Konstantin Boudnik,
    - ASF Hadoop committer, Bigtop PMC,
    - Director of Engineering, WANdisco
What are we dealing with?

- **Hadoop 1.x**
  - stable, but old
- **Hadoop 2.0.x**
  - modern, used to be alpha, now stabilizing
- **Hadoop 2.1.x**
  - modern, feature-driven
- **Hadoop 3.x**
  - perpetual trunk
What are the implications?

- YARN's appeal as IaaS
- Fragmentation
- Repeat of “UNIX vendor wars”
- Cutting off vital sources of feedback
- Jaded downstream
- Confused users
- Delayed world domination
What's downstream to do?

- `mvn help:all-profiles`
  - Profile Id: hadoop_0.20.203 (active)
  - Profile Id: hadoop_1.0
  - Profile Id: hadoop_non_secure
  - Profile Id: hadoop_facebook
  - Profile Id: hadoop_0.23
  - Profile Id: hadoop_yarn
  - Profile Id: hadoop_2.0.0
  - Profile Id: hadoop_2.0.1
  - Profile Id: hadoop_2.0.2
  - Profile Id: hadoop_2.0.3
  - Profile Id: hadoop_trunk
  - Profile Id: hadoop_cdh4.1.2
That active profile?

- http://mvnrepository.com/artifact/org.apache.hbase/hbase/0.94.3

```xml
<dependency>
  <groupId>org.apache.hbase</groupId>
  <artifactId>hbase</artifactId>
  <version>0.94.3</version>
</dependency>
```

- Try finding Apache Giraph artifacts!
We don't have the TCK, but...

- Zookeeper
- HBase
- Pig
- Hive
- Impala
- Sqoop
- Oozie
- Whirr
- Mahout
- Flume
- Giraph
- Hama
- Hue
- Solr
- Crunch
Apache Bigtop

“open-source software related to a system for integration, packaging, deployment and validation of a big data management software distribution based on Apache Hadoop”
Remember what Debian did to Linux?
Bigtop is trying to do it with Hadoop

Hadoop Ecosystem
(Pig, Hive, Mahout)

Hadoop
(HDFS + MR)
What does Bigtop offer:

• Community focused on all of the above

• Software for:
  – Integration
  – Build (make, Maven)
  – Packaging (RPM, DEB)
  – Deployment (Puppet)
  – Testing (iTest)

• A continuous integration Jenkins server
Embrace asynchronous nature

- Don't expect flag days
- Don't expect agreement on releases
- Do practice Last Known Good Builds
Who's on-board?

- Cloudera
  - CDH4 is 100% based on Bigtop (hadoop v2)
- WANdisco
- TrendMicro
- Hortonworks, EMC, EBay, Intel (partially)
- Canonical
  - Ubuntu Server: Hadoop and Bigdata blueprint
- Illumos (early stages of interest)
Who's on-board?

- Cloudera
  - CDH4 is 100% based on Bigtop (hadoop v2)
- WANdisco
- TrendMicro
- Hortonworks, EMC, EBay, Intel (partially)
- Canonical
  - Ubuntu Server: Hadoop and Bigdata blueprint
- Illumos (early stages of interest)
Who's on-board?

- Cloudera
  - CDH4 is 100% based on Bigtop (hadoop v2)
- WANdisco
- TrendMicro
- Hortonworks, EMC, EBay, Intel (partially)
- Canonical
  - Ubuntu Server: Hadoop and Bigdata blueprint
- Illumos (early stages of interest)
Who's on-board?

- Cloudera
  - CDH4 is 100% based on Bigtop (hadoop v2)
- WANdisco
- TrendMicro
- Hortonworks, EMC, EBay, Intel (partially)
- Canonical
  - Ubuntu Server: Hadoop and Bigdata blueprint
- Illumos (early stages of interest)
What's happening

- A special release: Bigtop 0.3.0-incubating
  - Hadoop 1.0.1

- Last stable release: Bigtop 0.5.0
  - Hadoop 2.0.2-alpha

- Next stable release: Bigtop 0.6.0
  - End of Mar 2013 release
  - Hadoop 2.0.4-alpha
  - Major focus on developers
A special note on 2.0.4-alpha

- It really will be 2.0.4.1
- First release to use Bigtop as release criteria
- A 100% community effort
- First non-vendor stabilization effort
- A stable base for 18 applications and counting!
<your idea here>