

The **Apache**
Software Foundation

<http://www.apache.org/>

Apache Bigtop: 100% Apache Bigdata management distribution (and so much more!)

Roman Shaposhnik
rvs@apache.org, Cloudera Inc.

And did we mention:

first ever?

One way of using Apache software

```
$ wget http://apache.org/httpd.tar.gz
```

```
$ tar xzvf httpd.tar.gz
```

```
$ cd httpd
```

```
$ ./configure ; make
```

```
$ make install
```

ERROR: can't write to /usr/local/bin

```
$ sudo make install
```

A different way

```
$ sudo apt-get install httpd
```

Would you like to also upgrade your conf?

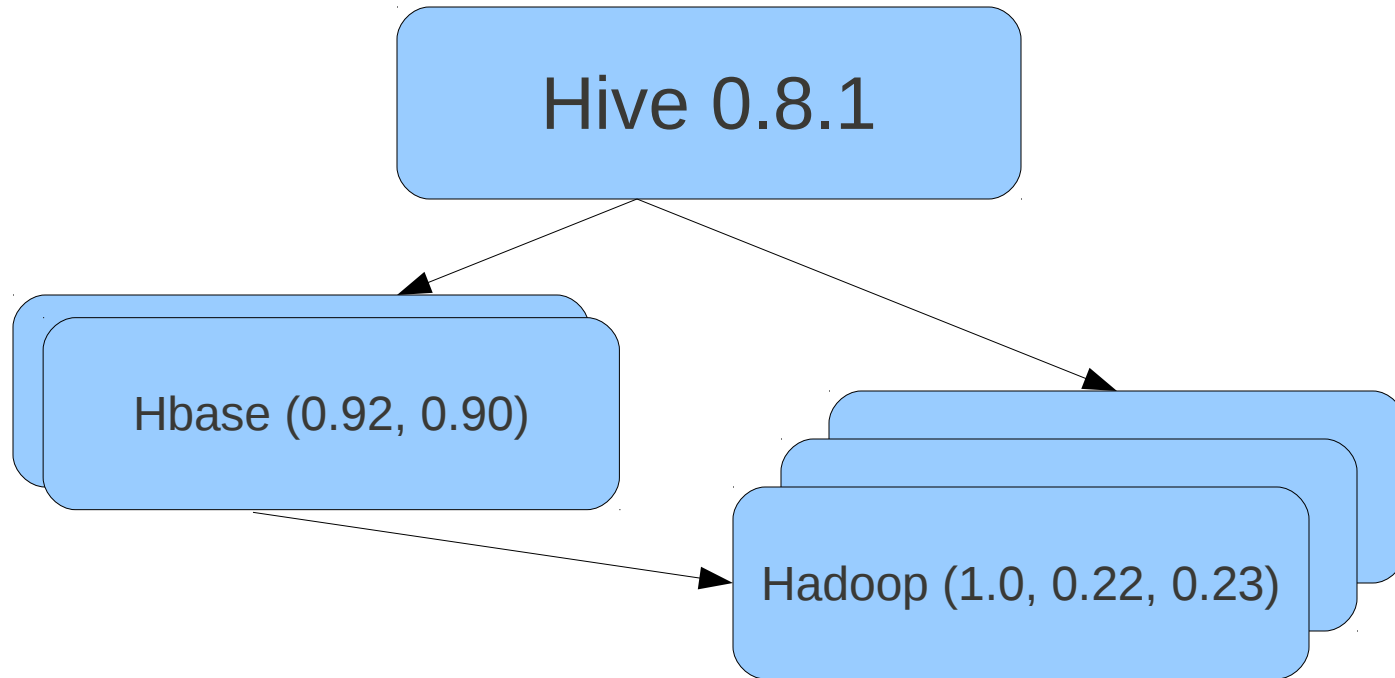
Is there apt-get install hadoop ?

- Hadoop is still in a very active development
- Hadoop is Java based
- Hadoop is a distributed application
- Hadoop is way more than HDFS + MR

Project-by-project approach

- “Passively” maintained code
 - Packaging, OS-level (init.d)
- Developer-centric view
 - Edit-compile-debug cycle vs. deployment
 - Lack of integration testing
- Differences in distributions/packaging:
 - Where is this valid: /usr/libexec ?
- Combinatoric explosion of dependencies

Dependencies Inferno:

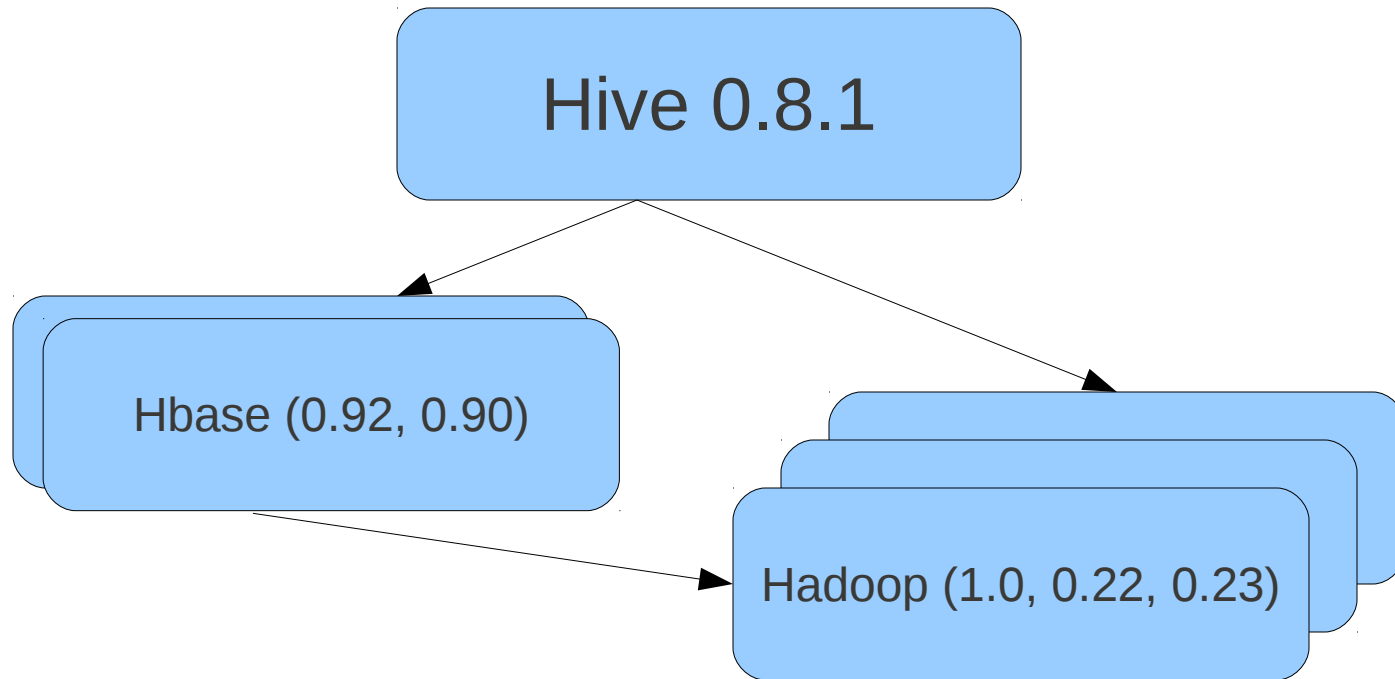


A million dollar question:

```
$ tar xzvf hive-0.8.1.tar.gz
```

```
$ ls hive-0.8.1/lib
```

Dependencies Inferno:



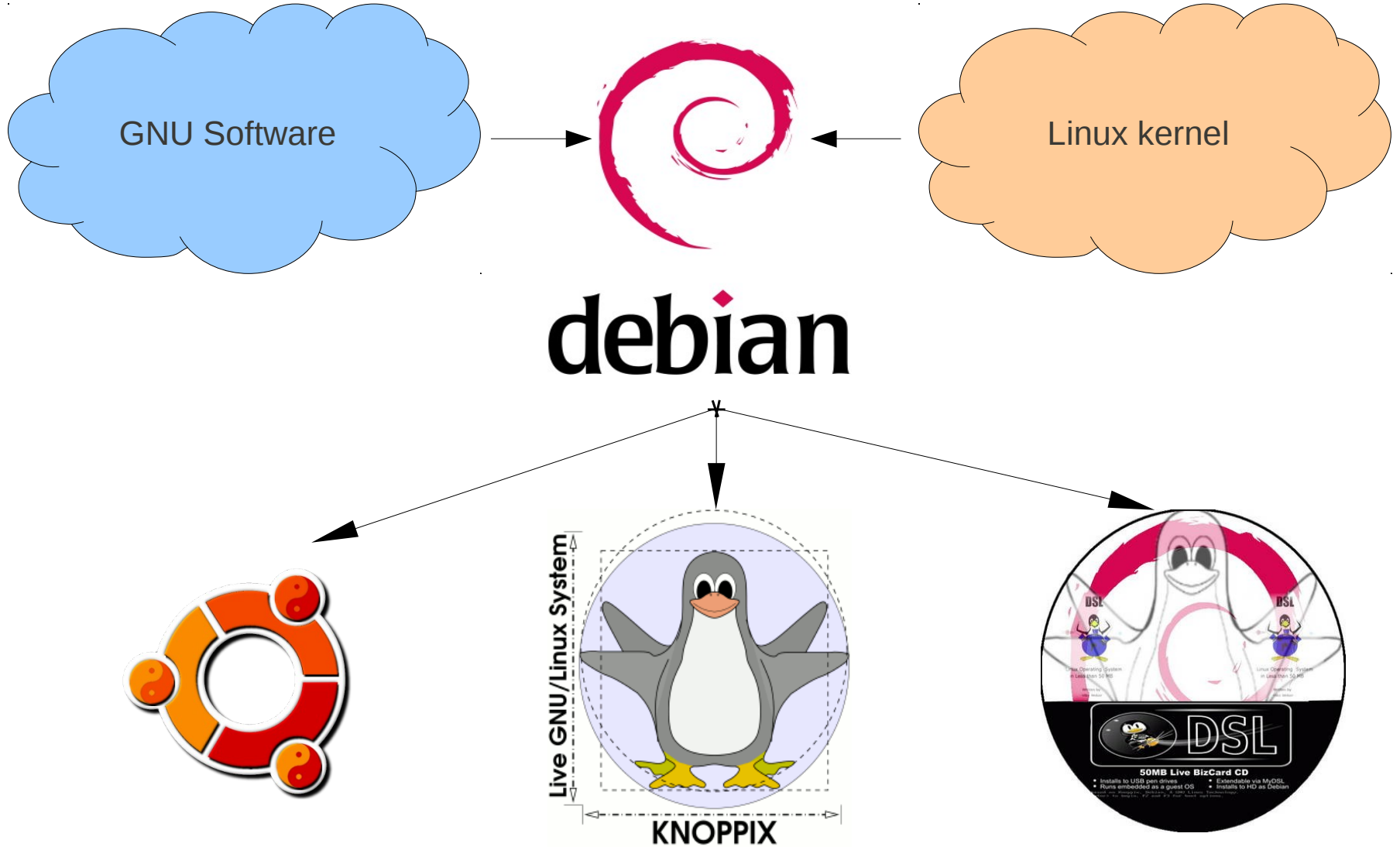
A million dollar question:

```
$ tar xzvf hive-0.8.1.tar.gz
```

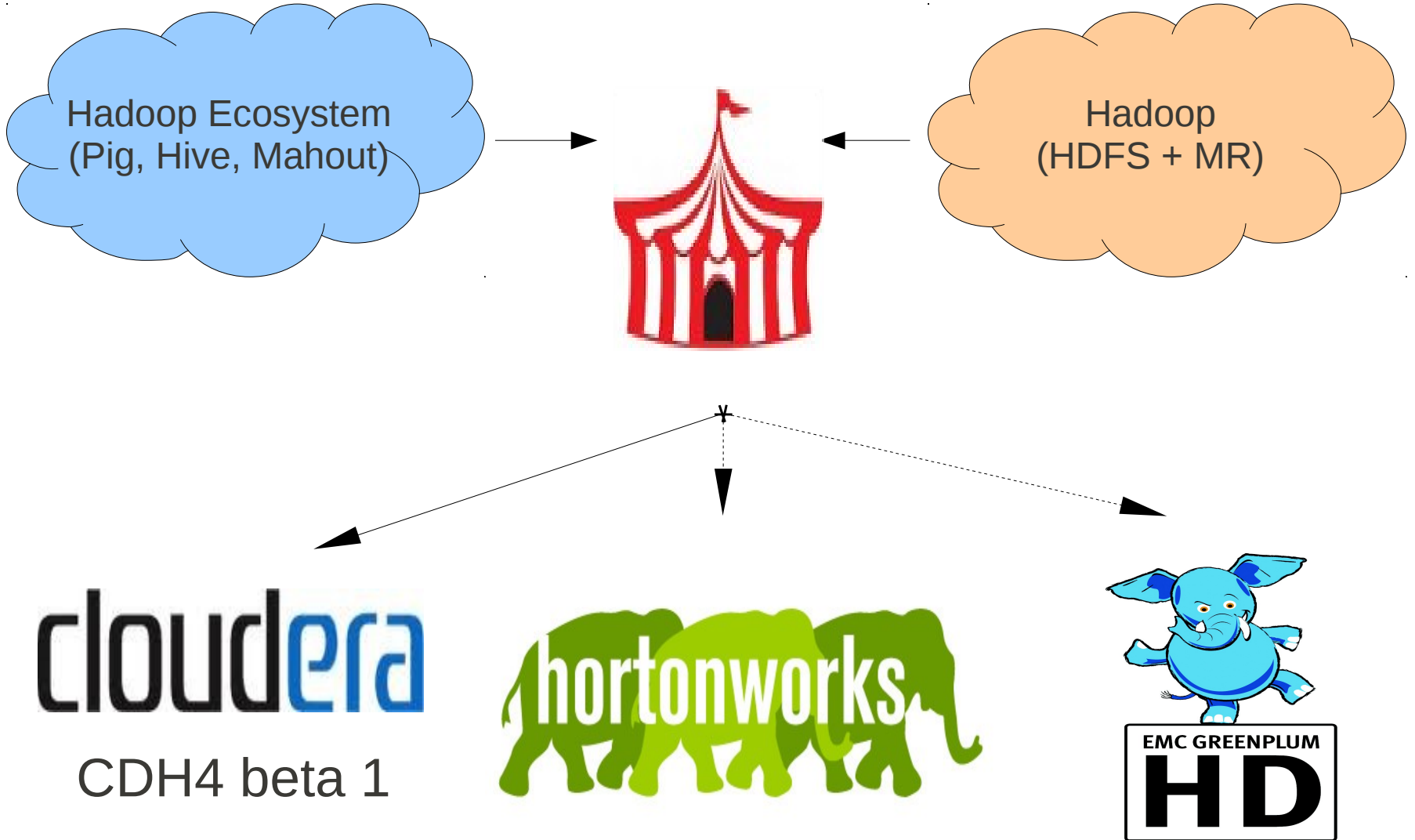
```
$ ls hive-0.8.1/lib
```

```
hbase-0.89.jar log4j-1.2.15.jar log4j-1.2.16.jar
```


Remember what Debian did to Linux?



Bigtop is trying to do it with Hadoop



Who's the target audience

- End users
 - YOU!
- ASF Projects/Bigdata developers
 - from Avro to Zookeeper
- Bigdata solutions vendors
 - Cloudera, EMC, Hortonworks, Karmasphere
- DevOPs
 - Ebay, Yahoo, Facebook, LinkedIn

End users:

- “Ubuntu of Bigdata”
 - One-click install experience for your platform of Choice
 - VMs for your virtualization pleasure
 - Community-driven, fully integrated and tested releases on a predictable schedule
- Curation of Apache Bigdata stack
 - User advocacy
 - Vulnerabilities management

ASF projects/developers:

- A stack view
 - Owning system-level functionality (packaging, init.d, user management, logging, monitoring)
 - Integration testing (iTest)
 - Harmonization of project dependencies
- A continuous integration view
 - Automation of build and release tasks
- Validation of release candidates
 - Combinatoric explosion of a dependency chain

Bigdata solutions vendors:

- A common, 100% Apache base for
 - Build systems
 - Custom patch management
 - Integration testing
 - Deployment infrastructure
- “Linux Test Project of Bigdata”
 - Integration tests and benchmarks

Who's on-board?

- Cloudera
 - CDH4 is 100% based on Bigtop (hadoop-0.23)
 - Available @cloudera.com
- Canonical
 - Ubuntu Server: Hadoop and Bigdata blueprint
<https://blueprints.launchpad.net/ubuntu/+spec/servercloud-p-hdp-hadoop>
- EMC, EBay
 - Early stages of prototyping
- *“Your name here”?*

What's happening?

- Last stable release: Bigtop 0.2.0-incubating
 - Hadoop 0.20.205
- Next stable release: Bigtop 0.3.0-incubating
 - End of Feb'12 release
 - Hadoop 1.0, HBase 0.92, Mahout 0.6, FlumeNG, Hive 0.8.1, Pig 0.9.2, Oozie 3.1.3, Sqoop 1.4.1
- Branch hadoop-0.23 (AKA Bigtop 0.4.0?)
 - Based on Hadoop 0.23.1
 - Basis of CDH4 beta 1

What's there in Bigtop

- Build/Packaging infrastructure
 - RPM, DEB
 - VirtualBox, VMWare and KVM VMs
 - Fedora, OpenSUSE, Mageia, CentOS, Ubuntu
- Deployment infrastrucutre
 - Puppet
- Integration test infrastrucutre (iTest)
- Bigtop Jenkins:
 - <http://bigtop01.cloudera.org:8080>

What Bigtop needs from you?

- More of you: we are still incubating!
 - Meetup: “Silicon Valley Hands-on Programming”
- More infrastructure for build/test
 - EC2, Supercell, EMC magic cluster
- More integration tests
 - Convince your bosses to commit to Bigtop
- Validate upstream release using Bigtop
 - HBase 0.92, Hadoop 0.22, etc.

Anatomy of Bigtop

- Build/Packaging
 - DEB, RPM, common (Solaris? Windows?)
- Deployment/Configuration Management
 - Puppet (Chef? cfEngine? smartFrog?)
- Integration testing framework
 - iTest
- Integration tests (based on iTest)

Anatomy of iTest

- Versioned, JVM-based test/data artifacts
- Dependency between test artifacts
- Matching stack of integration tests
- Implementation
 - Maven artifacts, pom files
 - JUnit test-execution entry point
 - Groovy for scripting

Contact

- § Bigtop home @Apache:
 - <http://incubator.apache.org/bigtop/>
- § Bigtop mailing lists:
 - bigtop-dev@incubator.apache.org
 - bigtop-user@incubator.apache.org
- § Roman Shaposhnik
 - rvs@apache.org

Presented by

