

# 無人值守 Apache BigTop 安裝光碟

Unattended Apache BigTop installer CD using preseed

Jazz Yao-Tsung Wang 王耀聰 <jazz@nchc.org.tw>

Co-Founder of Hadoop.TW

Associate Researcher,

National Center for High-performance Computing

2013/11/10

# On 11 Feb 2011, 4\$ shared about preseed !

感謝 4\$ 大大分享 Debian 6.0 自動化安裝

The screenshot shows a blog post on the 'FourDollars Blog' website. The page has a blue header with the blog name and a navigation menu with '首頁', 'Software Development', and 'Debian Package'. The main content area is white with a light blue background. The post title is '\$4 的 Debian 6.0 自動化安裝'. The author is '4\$'. The date is '2011年2月11日星期五'. The post content includes a sub-header '製作可開機安裝 Debian 6.0 的 USB Stick', a paragraph of text, a code block for mounting the ISO, another paragraph, and a sub-header '準備好要安裝的電腦跟可以透過有線網路 DHCP 上網的網路環境'. The footer of the post says '現在這隻 USB Stick 可以接上任何可以使用 USB Stick 開機的電腦上開機就會看到:'. On the right side, there is a '訂閱' (Subscribe) section with '發表文章' (Post Article) and '留言' (Comment) buttons, and a '企鵝星球' (Penguin Planet) logo at the bottom.

## FourDollars Blog

GNU/Linux, Debian/Ubuntu, Mac OS X, Free Software/Open Source Software, Freeware, and Programming.

首頁 Software Development Debian Package

2011年2月11日星期五

### \$4 的 Debian 6.0 自動化安裝

製作可開機安裝 Debian 6.0 的 USB Stick

首先去下載 `debian-6.0.2.1-i386-businesscard.iso` (46M) 回來  
然後再準備一個 USB Stick 接上電腦，如果 mount 在 Linux 系統上的話記得先 umount 掉  
以 root 權限執行 (這裡假設 USB Stick 的 Device Node 位於 `/dev/sdc`)

```
# cat debian-6.0.2.1-i386-businesscard.iso > /dev/sdc
```

這樣就可以製作出可開機安裝 Debian 6.0 的 USB Stick 沒錯就是這樣製作的  
這是 Debian 6.0 以後的新功能 hybrid CD/DVD image 不過只有在 i386/amd64 上面有作用

### 準備好要安裝的電腦跟可以透過有線網路 DHCP 上網的網路環境

現在這隻 USB Stick 可以接上任何可以使用 USB Stick 開機的電腦上開機就會看到:

訂閱

發表文章

留言

企鵝星球

Source: <http://fourdollars.blogspot.tw/2011/02/4-debian-60.html>

# Giveaway (1) :

**NARLabs**

## My tiny work on Apache BigTop installer CD

The screenshot displays the GitHub interface for the repository `jazzwang/haduzilla`. At the top, the GitHub logo and navigation links are visible. The repository is public and has 22 commits, 2 branches, 0 releases, and 1 contributor. The current branch is `master`. The file list shows a recent commit by `jazzwang` 6 months ago, with files `isolinux`, `preseed`, and `Makefile`. A dialog box titled "Switch branches/tags" is open, showing a search bar and a list of branches: `debian` (checked) and `master` (checked).

<https://github.com/jazzwang/haduzilla>

- master branch: for Ubuntu 12.04
- debian branch: for Debian 7.0.2

# ISO files for BigTop 0.6.0~0.7.0

Home / Browse / Home & Education / Administration / drbl-hadoop / Files



**drbl-hadoop** Prealpha

Brought to you by: jazzwang

Summary | **Files** | Reviews | Support | Wiki | Donate | Code

Looking for the latest version? [Download bigtop-0.6.0.iso \(744.5 MB\)](#)

Home

Name ↕	Modified ↕	Size ↕	Downloads ↕
<b>0.7.0</b>	< 10 mins ago		
<b>0.6.0</b>	2013-08-11		
<b>0.5.0</b>	2013-06-03		
<b>0.1.0</b>	2010-07-19		

**Totals: 4 Items**

<http://sourceforge.net/projects/drbl-hadoop/files/>

# Screenshot of BigTop installer CD

BigTop Installer (0.7.0\_131110-0103)

Single Node BigTop (DHCP)

Boot From Hard Disk



Apache BigTop

Free Software Labs  
自由軟體實驗室

NCHC, Taiwan

國家高速網路與計算中心

# Basic Info. of BigTop installer CD

```
Ubuntu 12.04.3 LTS bigtop tty1

bigtop login: user
Password:
Welcome to Ubuntu 12.04.3 LTS (GNU/Linux 3.8.0-29-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

sudo juser@bigtop:~$ sudo jps
[sudo] password for user:
789 DataNode
855 NameNode
920 SecondaryNameNode
1161 ResourceManager
1092 NodeManager
1570 Jps
user@bigtop:~$
```

- ID:

**user**

- Password:

**hadoop.TW**

- Default installed

**Hadoop 2.0**

**YARN**

- Suggest:

**Install Hue**

- Note:

**amd64 only**

**How to deal with Big Data ?**

---

**Let's talk about  
“Big Data  
Architecture”.**

# Current Status of Big Data .....

**Big data** is like **teenage sex**:  
everyone **talks about** it,  
nobody really knows **how to do** it,  
everyone **thinks** everyone else is doing it,  
so everyone **claims** they are doing it ..

– Dan Ariely, Professor at Duke University  
and Professor at Center for Advanced Hindsight



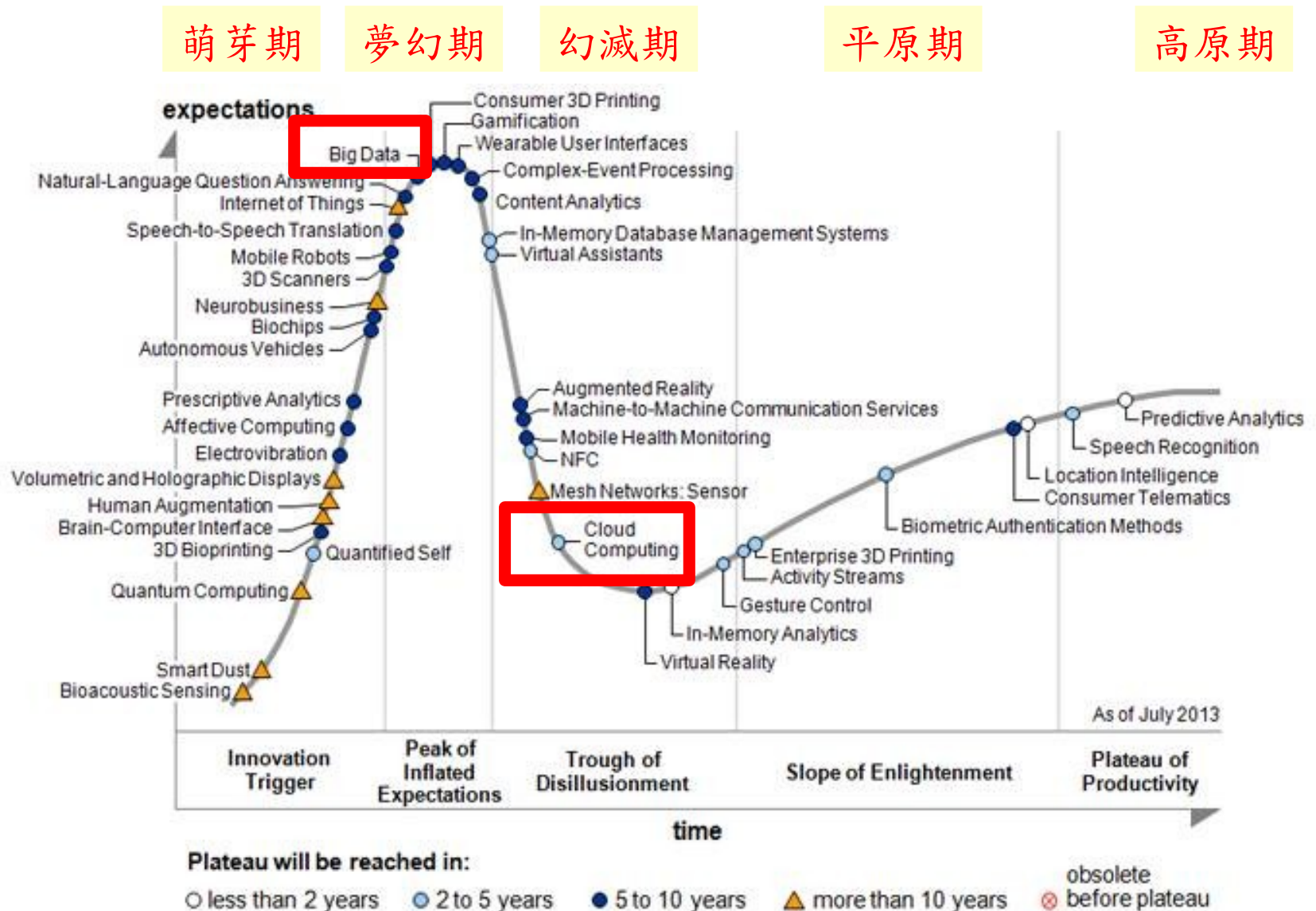
**Dan Ariely** · 92,283 個追蹤者  
1月6日 8:02 · 🌐

📡 追蹤

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

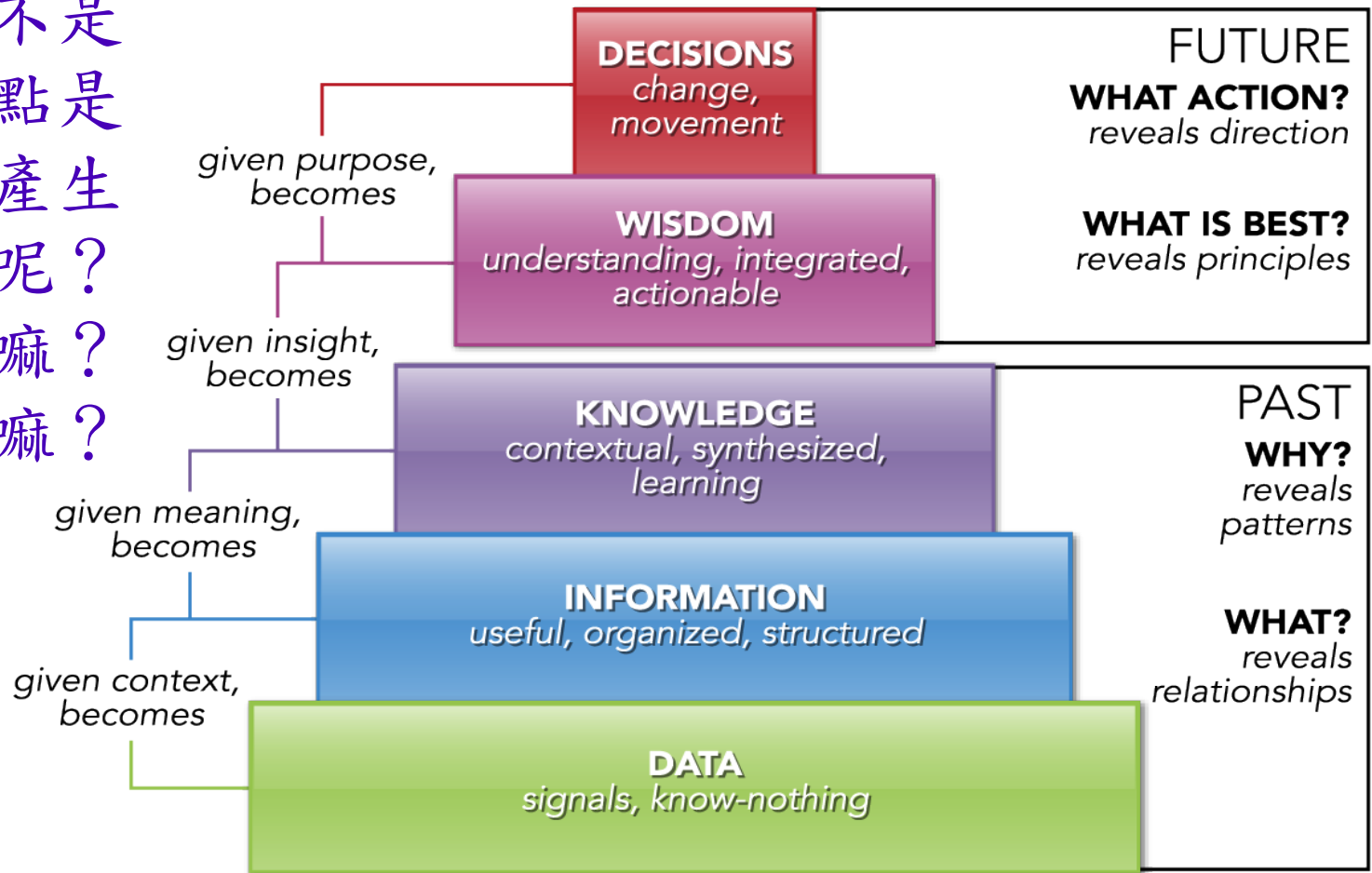


# Gartner Hype Cycle 2013



# 知識源自彙整過去， 智慧在能預測未來

資料多寡不是  
重點，重點是  
我們想要產生  
什麼價值呢？  
時效合理嘛？  
成本合理嘛？



<http://www.pursuantgroup.com/blog/tag/dikw-model/>

大家都說「資料是金礦」，  
那就讓我們拿採礦當類比吧！

國際金價

提供給客戶的價值

產品通路

開採成本

總擁有成本

軟硬體投資

提煉廠

分析平台與工具軟體

**SMAQ**

含金量

資料鑑價？

商業模式

開採權

分析資料的合法性

個資法

金礦

資料集

**Open Data**

# 今天我們能談的是該怎麼蓋「提煉廠」

國際金價

提供給客戶的價值

產品通路

開採成本

總擁有成本

軟硬體投資

提煉廠

分析平台與工具軟體

**SMAQ**

含金度

資料鑑價？

商業模式

開採權

分析資料的合法性

個資法

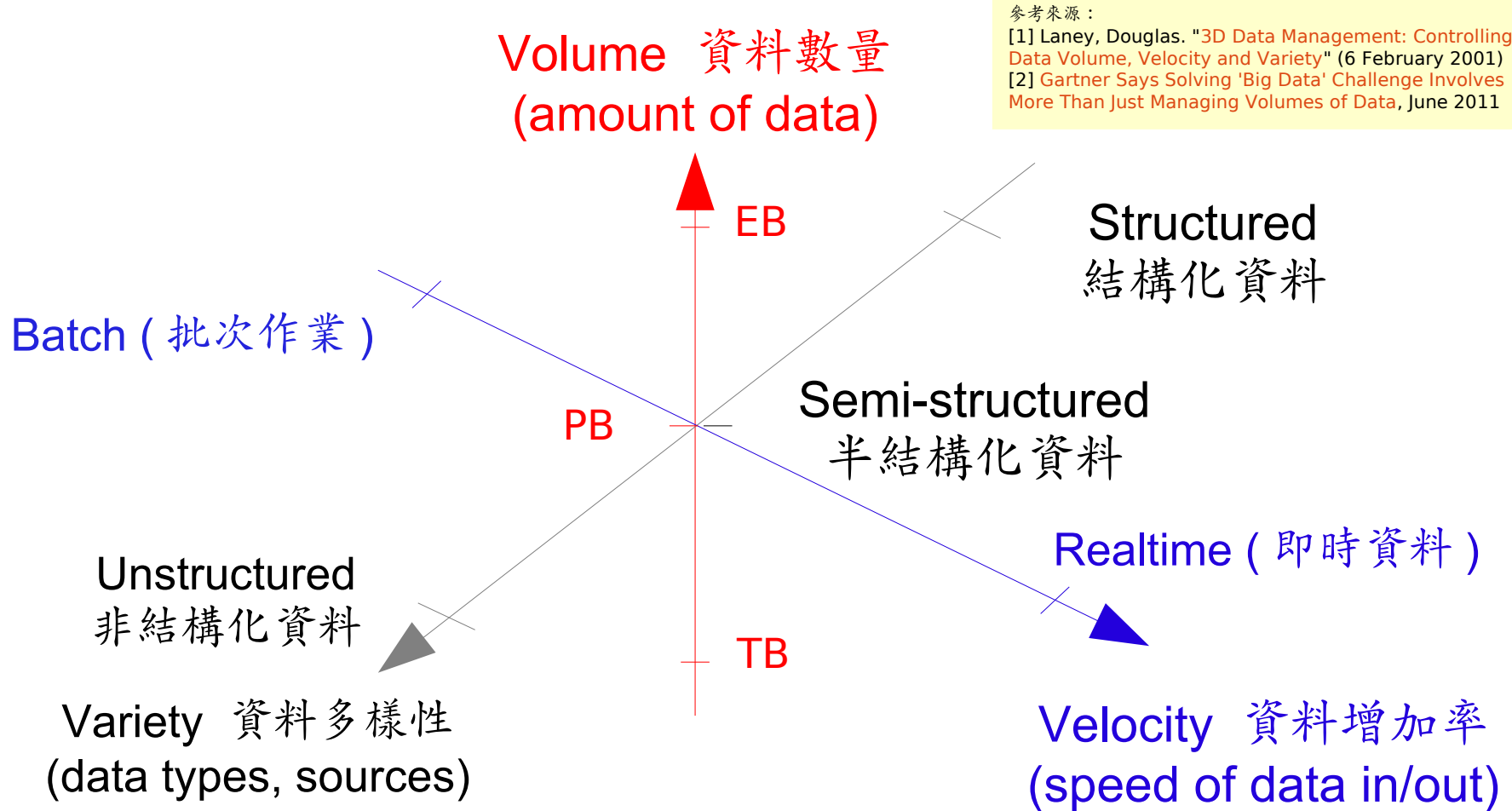
金礦

資料集

**Open Data**

# 巨量資料的標準定義

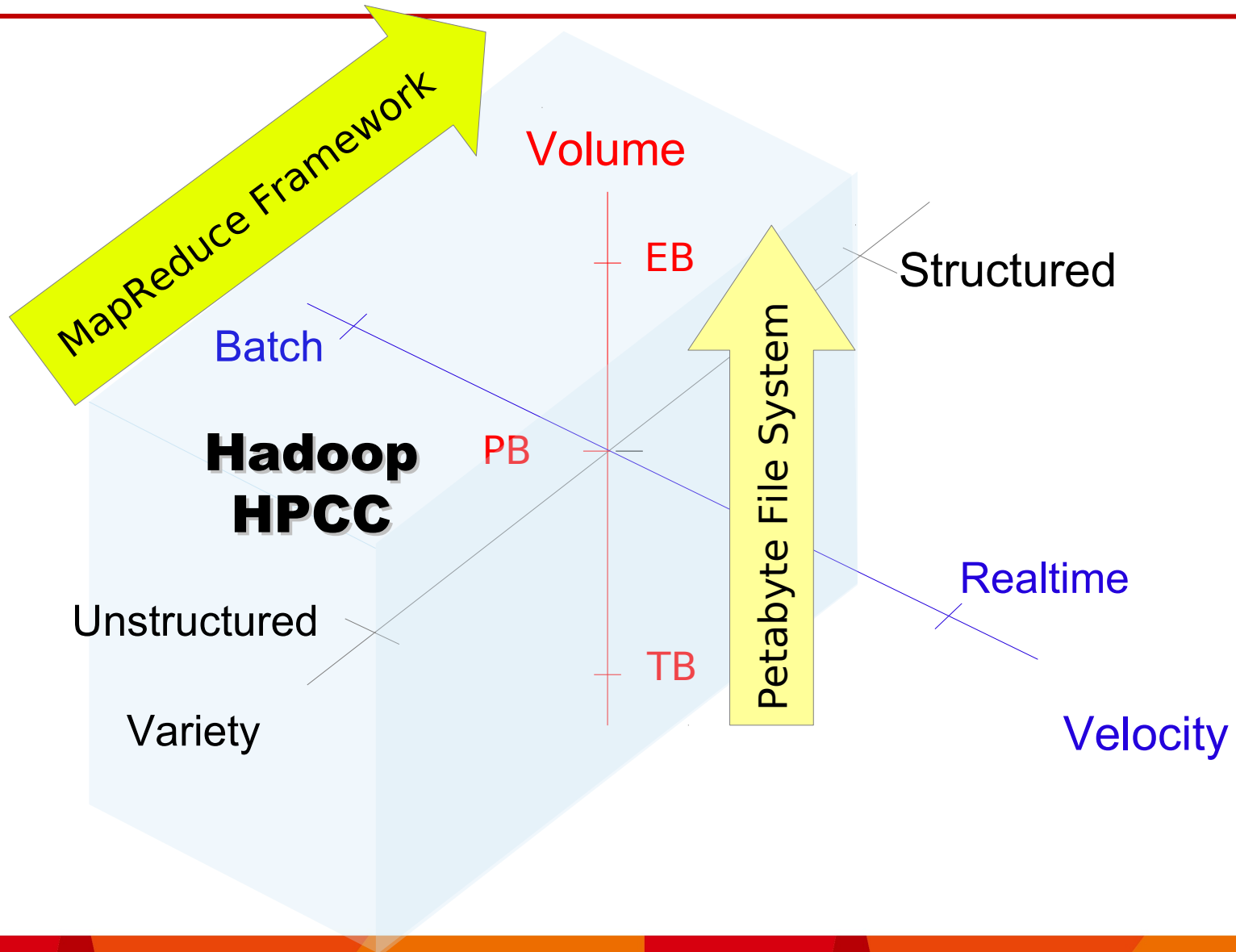
## 3 Vs of Big Data



巨量資料的挑戰在於如何管理「數量」、「增加率」與「多樣性」

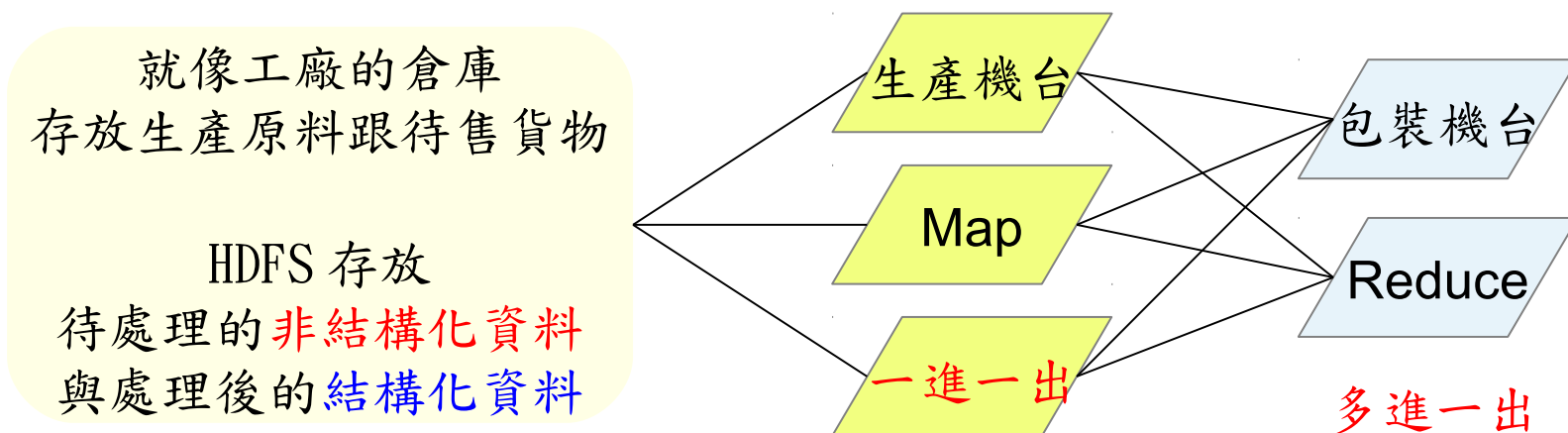
# 處理巨量資料的三類技術 (1)

## Data at Rest – MapReduce Framework



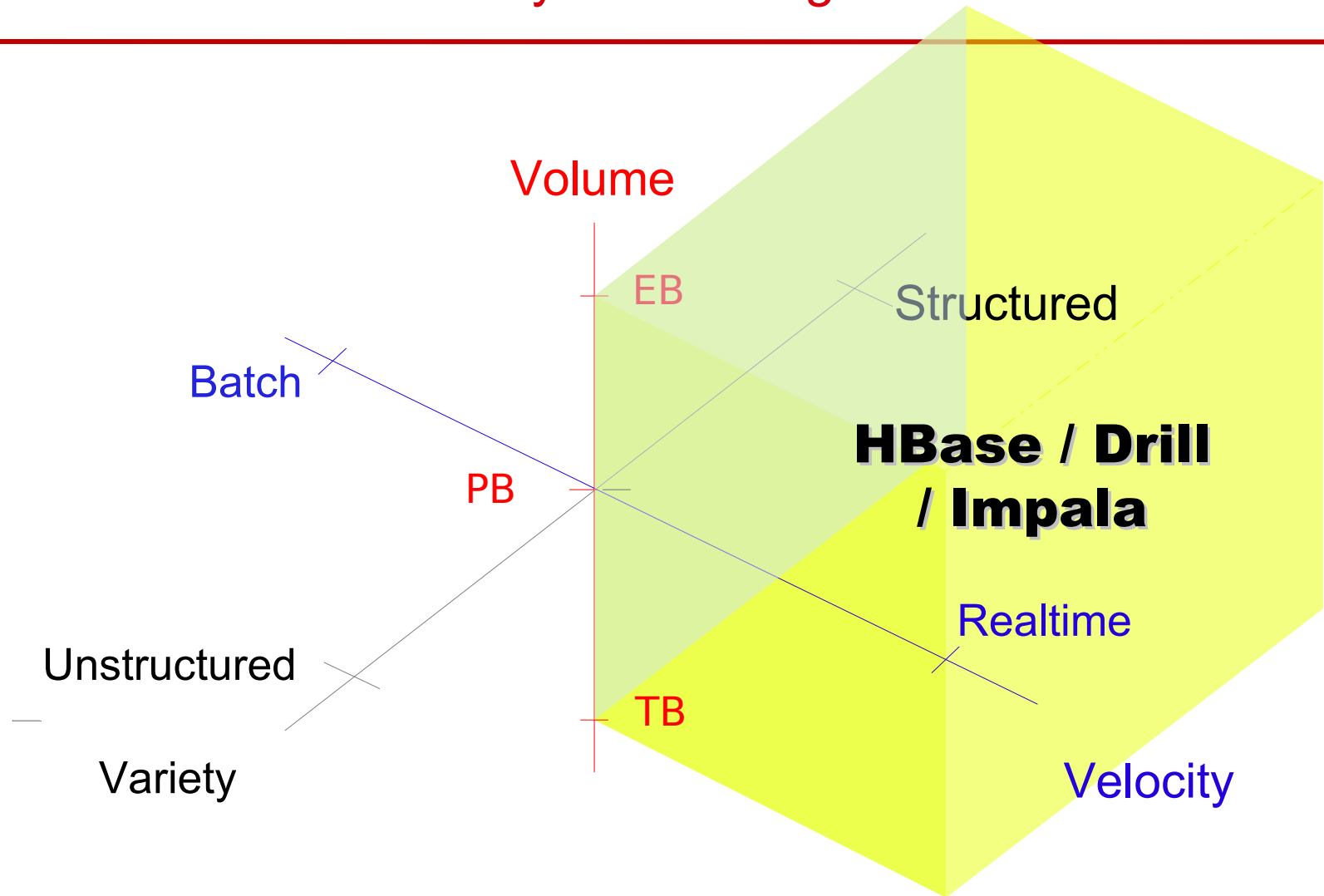
**Hadoop** 是一個讓使用者簡易撰寫並執行處理海量資料應用程式的軟體平台。

亦可以想像成一個處理海量資料的生產線，只須學會定義 **map** 跟 **reduce** 工作站該做哪些事情。



# 處理巨量資料的三類技術 (2)

## Data in Motion – In-Memory Processing





# Google 的技術演進 VS Apache 專案

Big Query  
(JSON, SQL-like)

Dremel  
(2010)

Apache Drill  
(2012)

Incremental Index Update  
(Caffeine)

Percolator  
(2010)

Graph Database

Pregel  
(2009)

Apache Giraph  
(2011)

Query

BigTable  
(2006)

Apache HBase  
(2007)

Map Reduce

MapReduce  
(2004)

Hadoop MapReduce  
(2006)

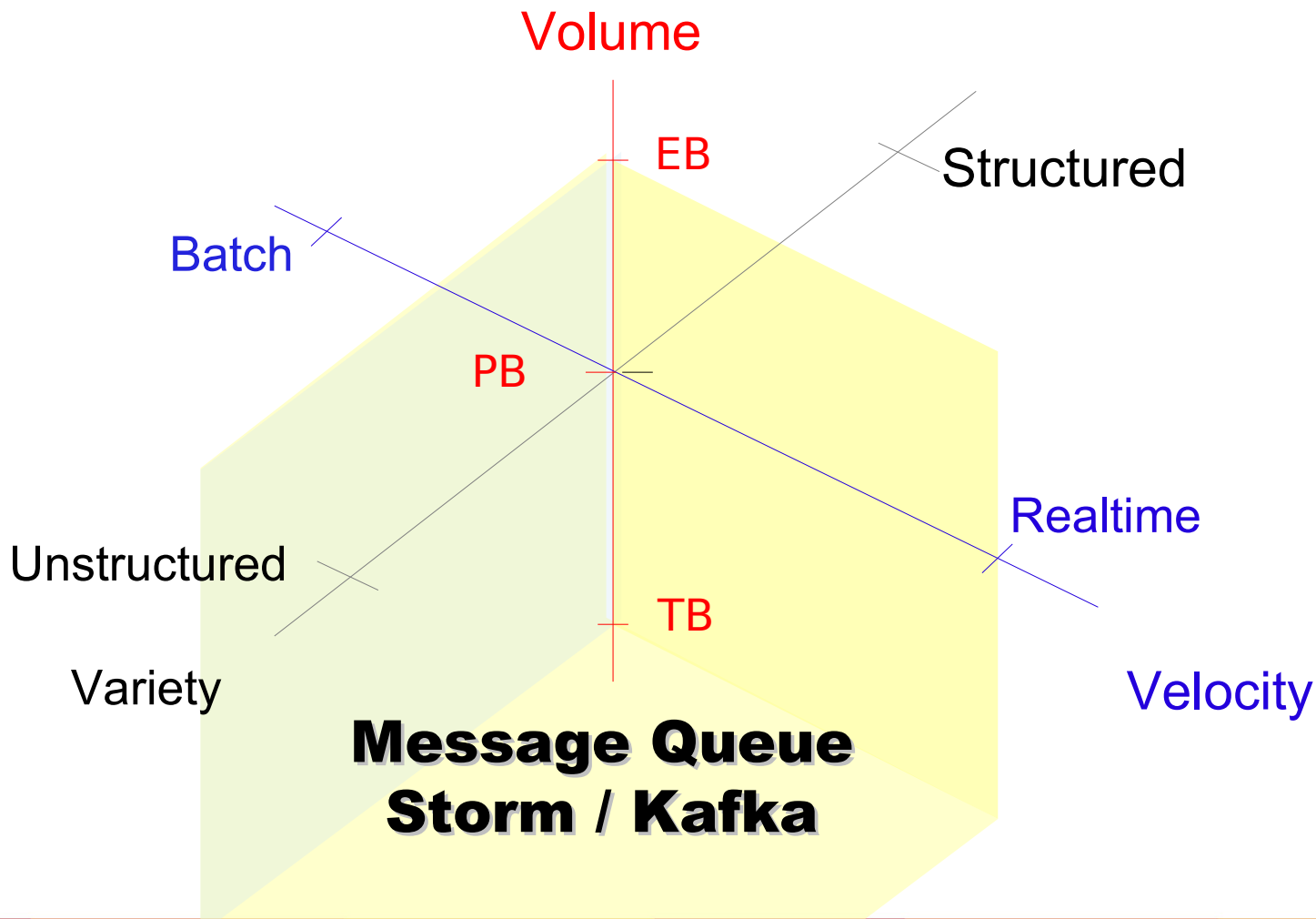
Storage

Google File System  
(2003)

HDFS  
(2006)

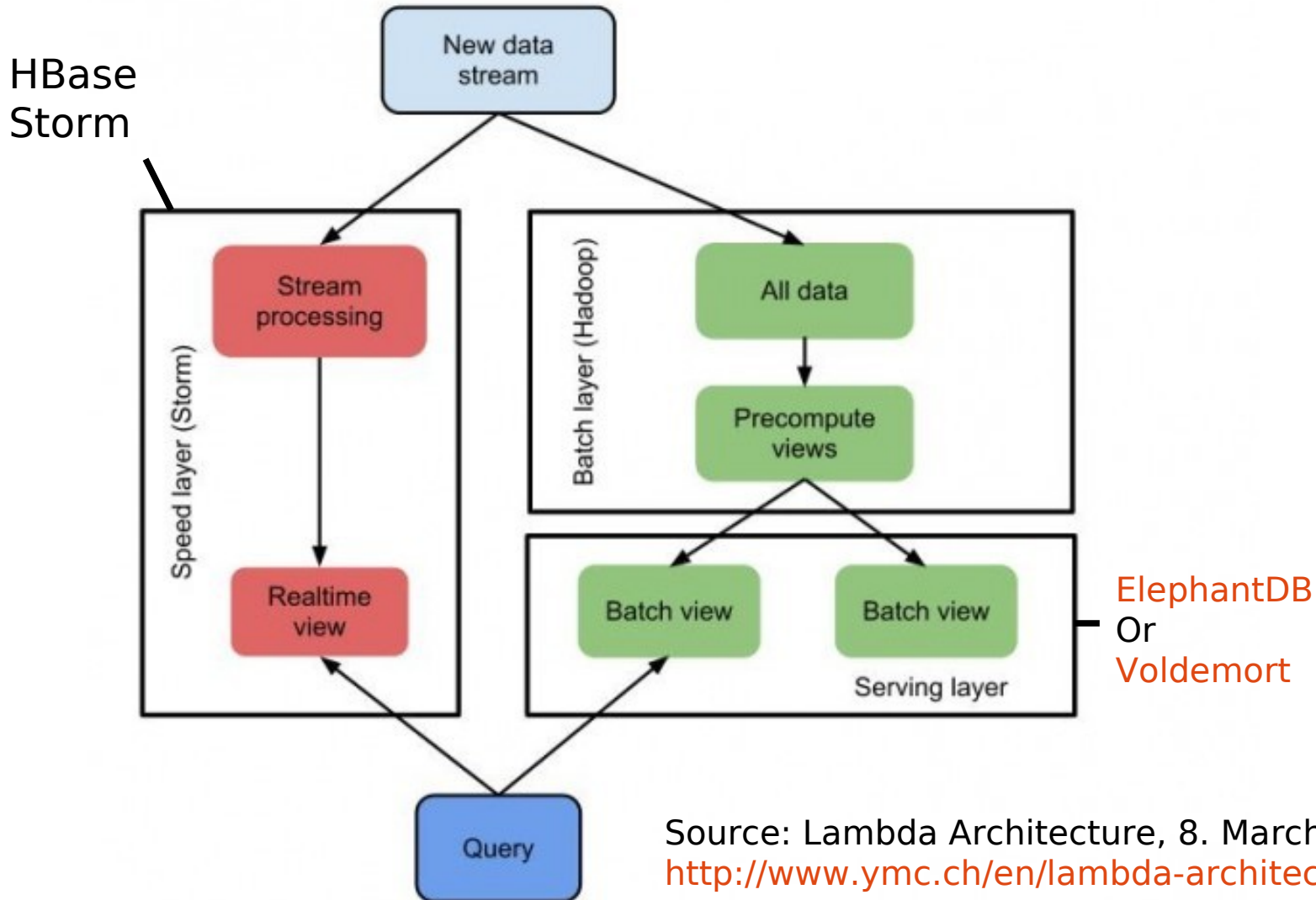
# 處理巨量資料的三類技術 (3)

## Streaming Data Collection



# 混合模式的巨量資料處理架構

## Lambda Architecture



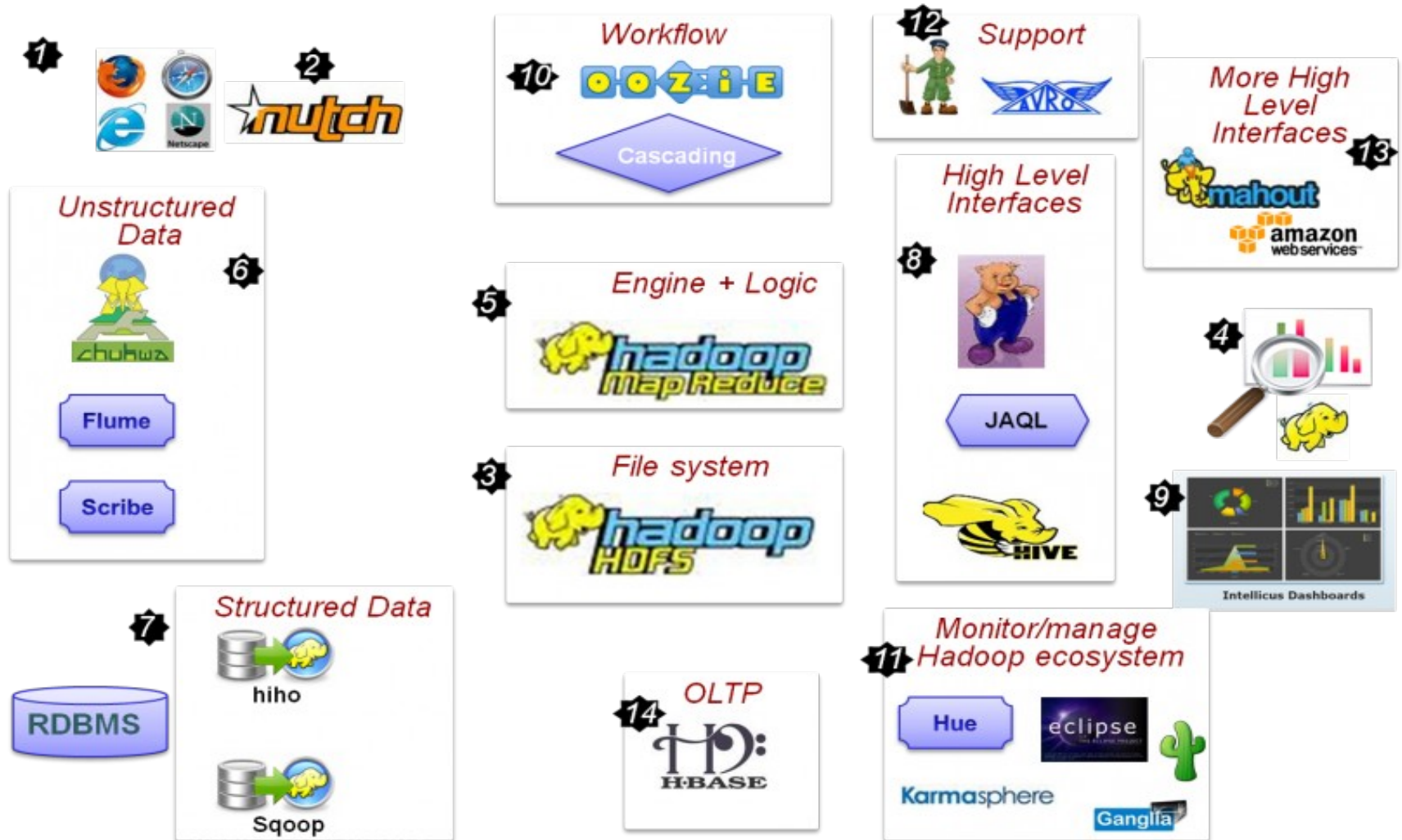
Source: Lambda Architecture, 8. March 2013  
<http://www.ymc.ch/en/lambda-architecture-part-1>

## What is Apache BigTop ?

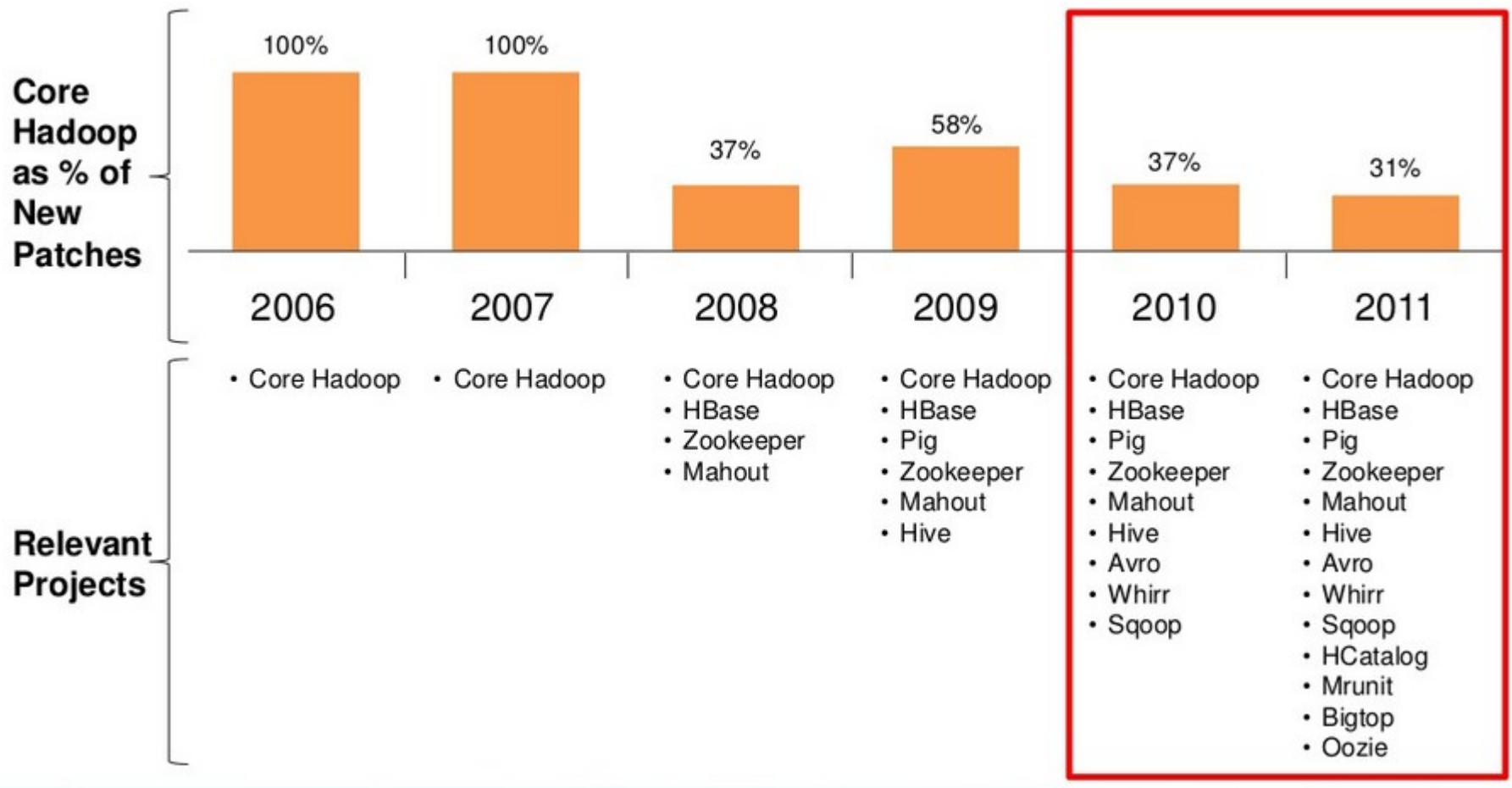
---

**Let's talk about**  
**“Apache BigTop”**  
**& Debian.**

# Apache Hadoop Ecosystem



# Apache Hadoop Ecosystem



Hadoop World 2011: The Hadoop Stack - Then, Now and in the Future

[http://www.slideshare.net/slideshow/embed\\_code/10110006](http://www.slideshare.net/slideshow/embed_code/10110006)

# Complexity of Apache Big Data Stack

Files	Integration	Web		Shell / CLI		Drivers				
		Languages		Libraries		Workflow		Scheduling		
		MapReduce		Stream		Graph		MPI		Other
RDBMS		Resource Management								
		Metadata storage								
		Time Series		ORM		OLAP		OLTP		
		Record storage								
Logs & events		File storage								
		Coordination								
		Formats		RPC		Authentication		Compression		

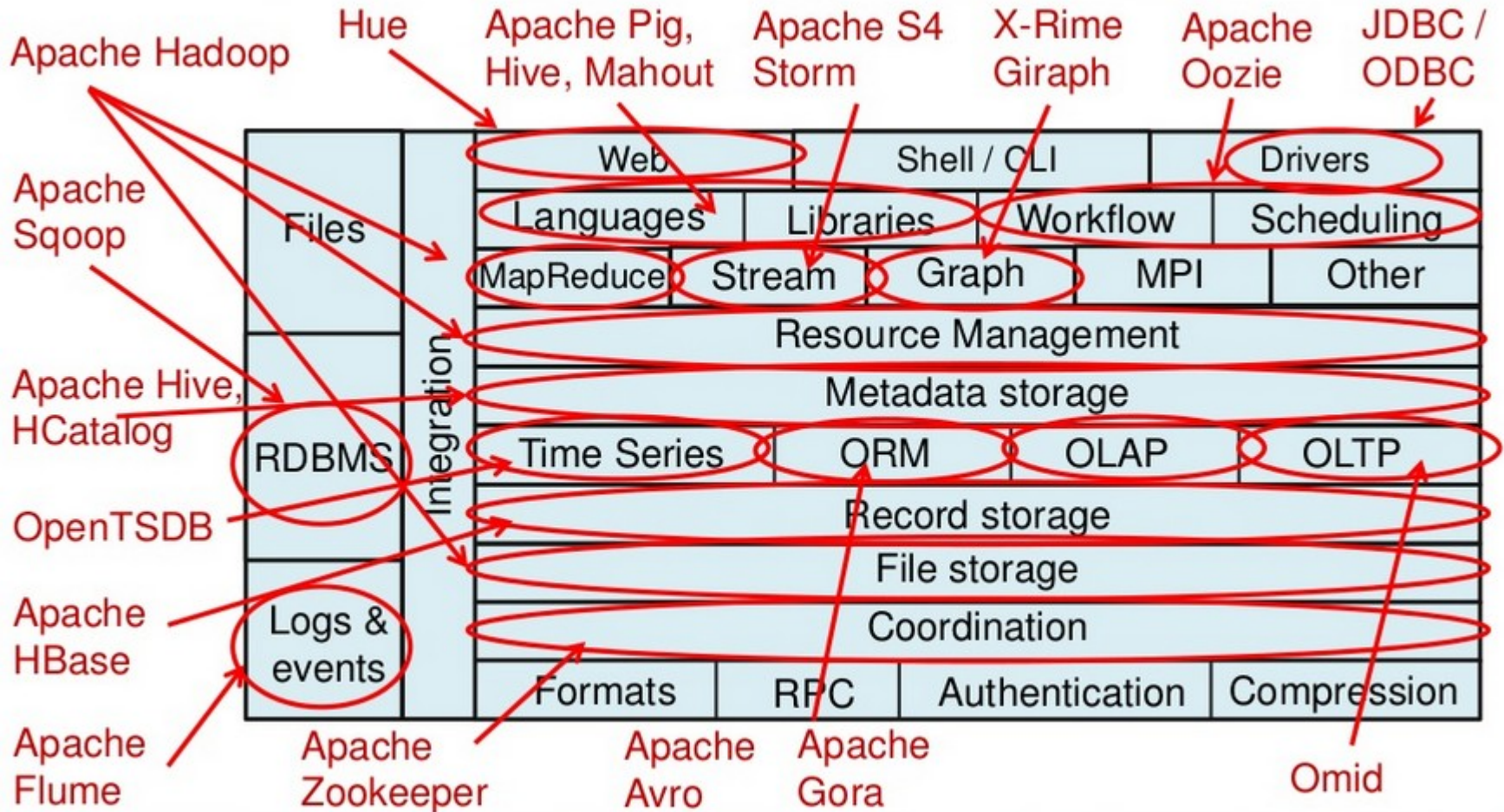
Hadoop World 2011: The Hadoop Stack - Then, Now and in the Future

[http://www.slideshare.net/slideshow/embed\\_code/10110006](http://www.slideshare.net/slideshow/embed_code/10110006)



# Complexity of

# Apache Big Data Stack



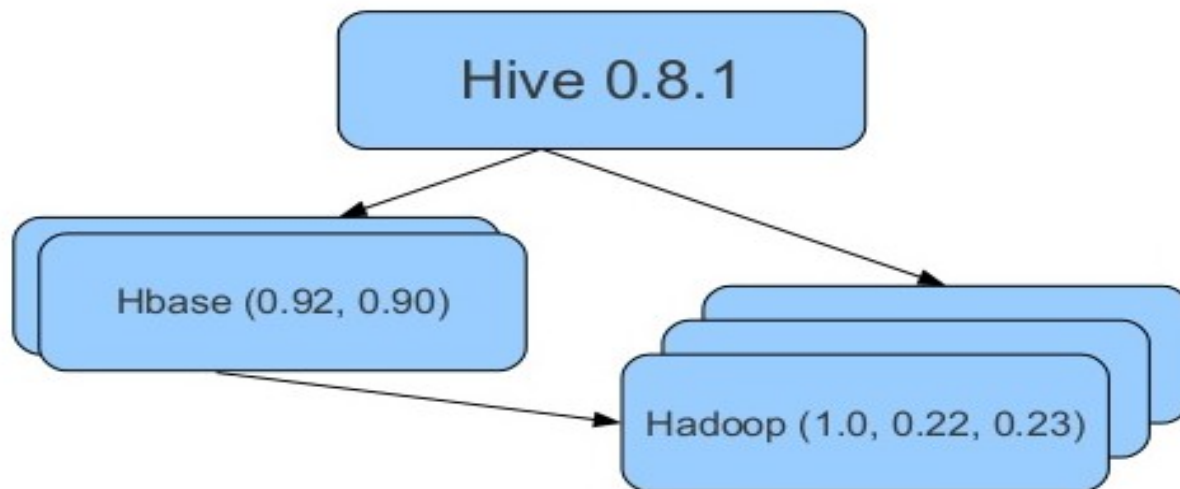
Hadoop World 2011: The Hadoop Stack - Then, Now and in the Future

[http://www.slideshare.net/slideshow/embed\\_code/10110006](http://www.slideshare.net/slideshow/embed_code/10110006)



# Big Problem: Package dependency

## Dependencies Inferno:



A million dollar question:

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

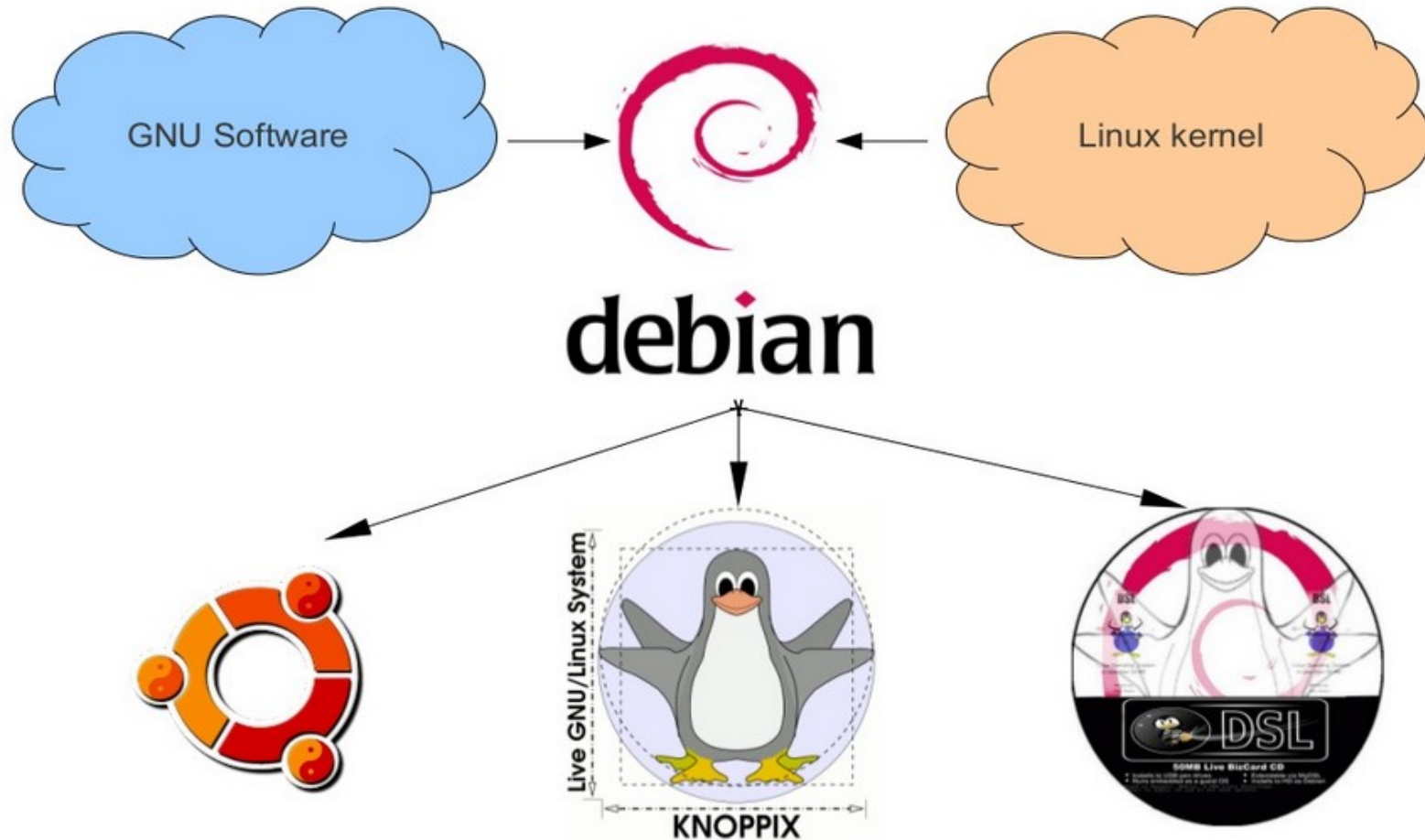
```
$ ls hive-0.8.1/lib
```

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

Deploying Hadoop-based Bigdata Environments,

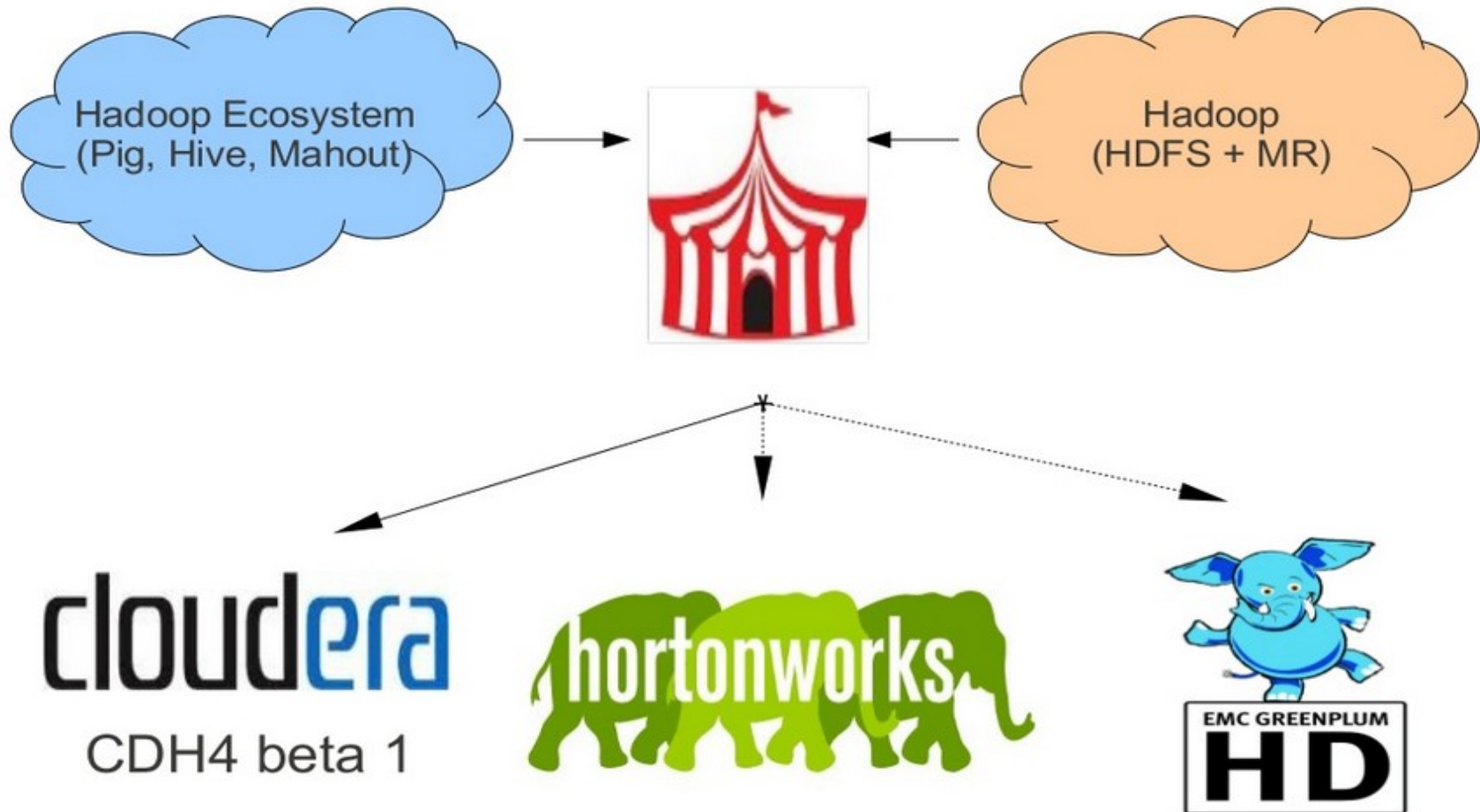
[http://www.slideshare.net/slideshow/embed\\_code/16370152](http://www.slideshare.net/slideshow/embed_code/16370152)

# What Debian did to Linux?



Deploying Hadoop-based Bigdata Environments,  
[http://www.slideshare.net/slideshow/embed\\_code/16370152](http://www.slideshare.net/slideshow/embed_code/16370152)

# Bigtop is trying to do it with Hadoop



Deploying Hadoop-based Bigdata Environments,  
[http://www.slideshare.net/slideshow/embed\\_code/16370152](http://www.slideshare.net/slideshow/embed_code/16370152)

# What's there in Bigtop

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

# Hot !! Apache BigTop 0.7.0 release on Nov. 6 2013

WEDNESDAY NOV 06, 2013

---

## Release of Apache Bigtop 0.7.0

Exciting times for the [Apache Bigtop](#) community and the [Apache Hadoop](#) ecosystem at large with the release of the brand new [Apache Bigtop 0.7.0](#)!

This new release brings tons of new features and fixes for the beloved 100% community and open source driven big data distribution.

Among the new features:

- Addition of [Phoenix](#), the Apache HBase SQL layer project
- Addition of [Apache Spark \(incubating\)](#), the general-purpose cluster computing system
- A brand new SolrCloud 4.5 integration with HDFS
- Improvement to our puppet recipes so setting up your Apache Bigtop toolchain will be easier than ever
- Addition of a new script `init-hdfs.sh` to initialize Apache HDFS filesystem structure so you can get up and running without manually setting up a complex directory structure
- Add flexibility in the way [Apache Oozie](#) manages its [Apache Tomcat](#) application
- Add a new standard location to install SQL connectors and other plug-ins

Also this releases fixes:

- Libsnappy is now included in [Apache Hadoop](#) packages
- Clean up some dependencies throughout our packages
- Update to our documentation

[https://blogs.apache.org/bigtop/entry/release\\_of\\_apache\\_bigtop\\_0](https://blogs.apache.org/bigtop/entry/release_of_apache_bigtop_0)

# Current Big Data Stack Packages

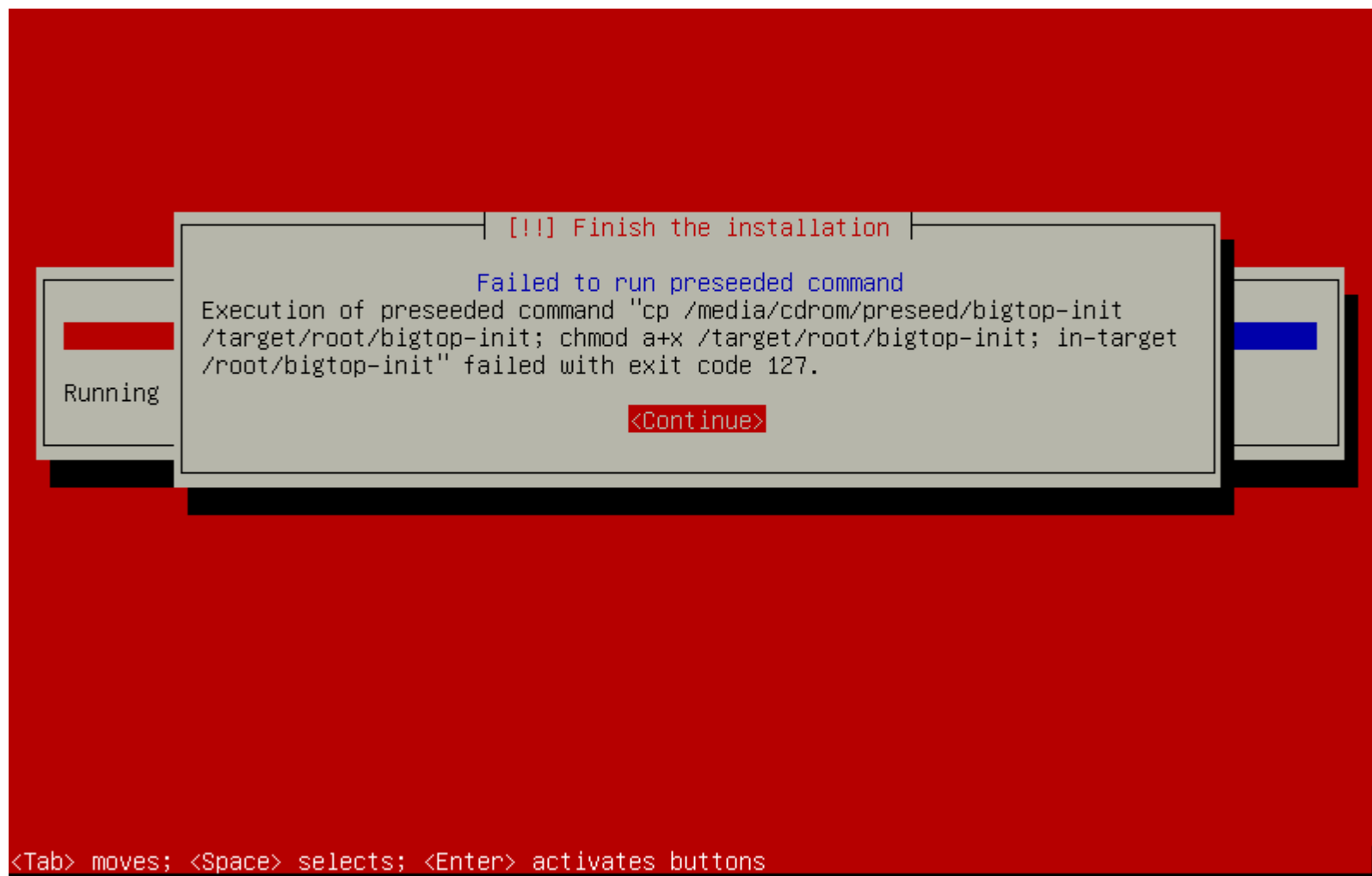
- Apache Hadoop 2.0.6
- Apache Zookeeper 3.4.5
- Apache Flume 1.4.0
- Apache HBase 0.94.12
- Apache Pig 0.11.1
- Apache Hive 0.11.0
- Apache Sqoop 2 (AKA 1.99.2)
- Apache Oozie 3.3.2
- Apache Whirr 0.8.2
- Apache Mahout 0.7.5
- Apache Solr (SolrCloud) 4.5.0
- Apache Crunch 0.7.0
- Apache HCatalog 0.5.0
- Apache Giraph 1.0.0
- LinkedIn DataFu 1.0.0
- Cloudera Hue 2.5.1

## **Current support GNU/Linux distribution**

---

- Centos 5 / 6
- Fedora 17 / 18
- Ubuntu Lucid / Precise / Quetzal
  - **Debian 7.0.2 (X)**
    - need 'testing' repo for **libc6 2.14**
- OpenSUSE 12
- SLES11

# Q: how to debug preseed errors ?





問題與討論  
Questions?

