

选择 Apache Hadoop 版本

目录

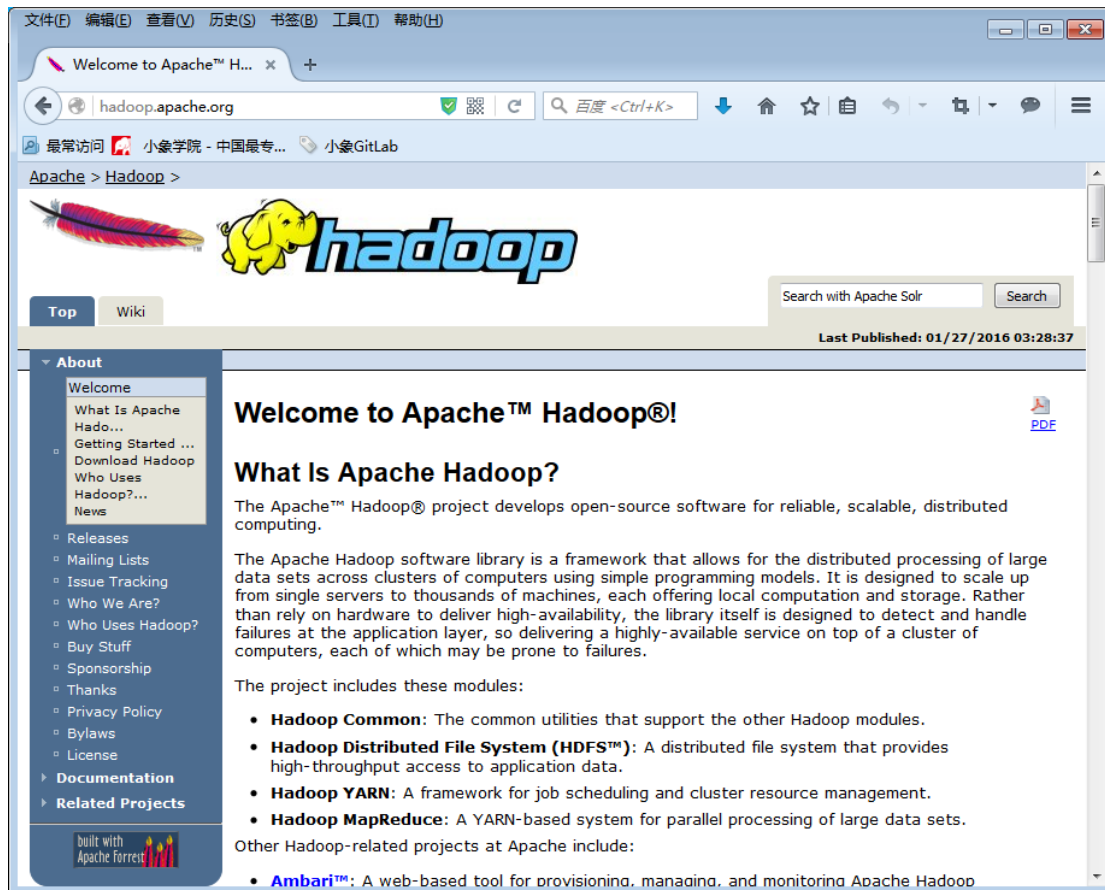
选择 Apache Hadoop 版本.....	1
1. 官网首页.....	2
1.1. Apache Hadoop.....	2
1.2. Apache HBase	2
1.3. Apache HIVE.....	3
2. 官网上有那些 Releases 版本.....	4
2.1. Apache Hadoop Releases 版本.....	4
2.2. Apache Hive Releases 版本.....	5
2.3. Apache HBase Releases 版本.....	6
3. 各模块版本兼容性.....	7
3.1. 查看官方文档.....	7
3.1.1. HBase 上手指南.....	7
3.1.2. HIVE 上手指南	10
3.1.3. HIVE 和 HBase 整合	11
3.2. 查看源码包中的 pom.xml 文件.....	14
3.2.1. 下载并解压源码包.....	14
3.2.2. 查看 HBase 中的 pom.xml	14
3.2.3. 查看 HIVE 中的 pom.xml.....	14
3.2.4. 总结.....	14
4. 选择 Apache Hadoop 版本.....	14
5. 补充.....	15



1. 官网首页

1.1. Apache Hadoop

地址: <http://hadoop.apache.org/>



Apache Hadoop 项目包括四个模块:

Hadoop Common: 公用程序。

Hadoop Distributed File System (HDFS): 分布式文件系统。

Hadoop YARN: 作业调度和集群资源管理框架。

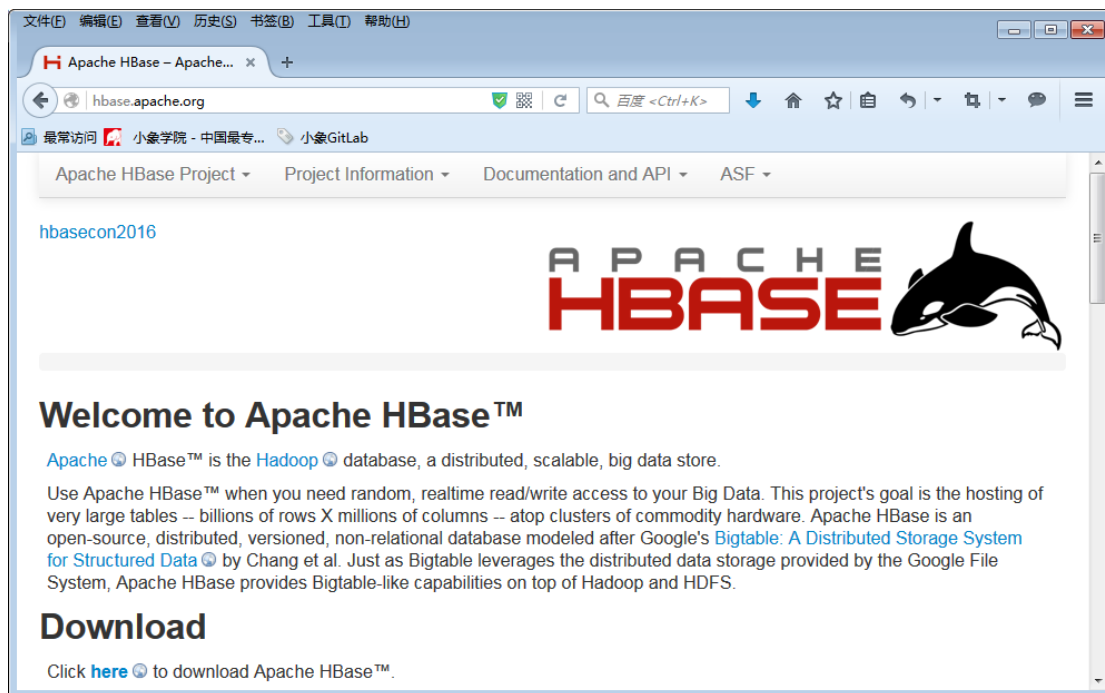
Hadoop MapReduce: 并行处理系统。

Apache Hadoop 相关的项目有: HBase, Hive, Spark 和 ZooKeeper 等等。

1.2. Apache HBase

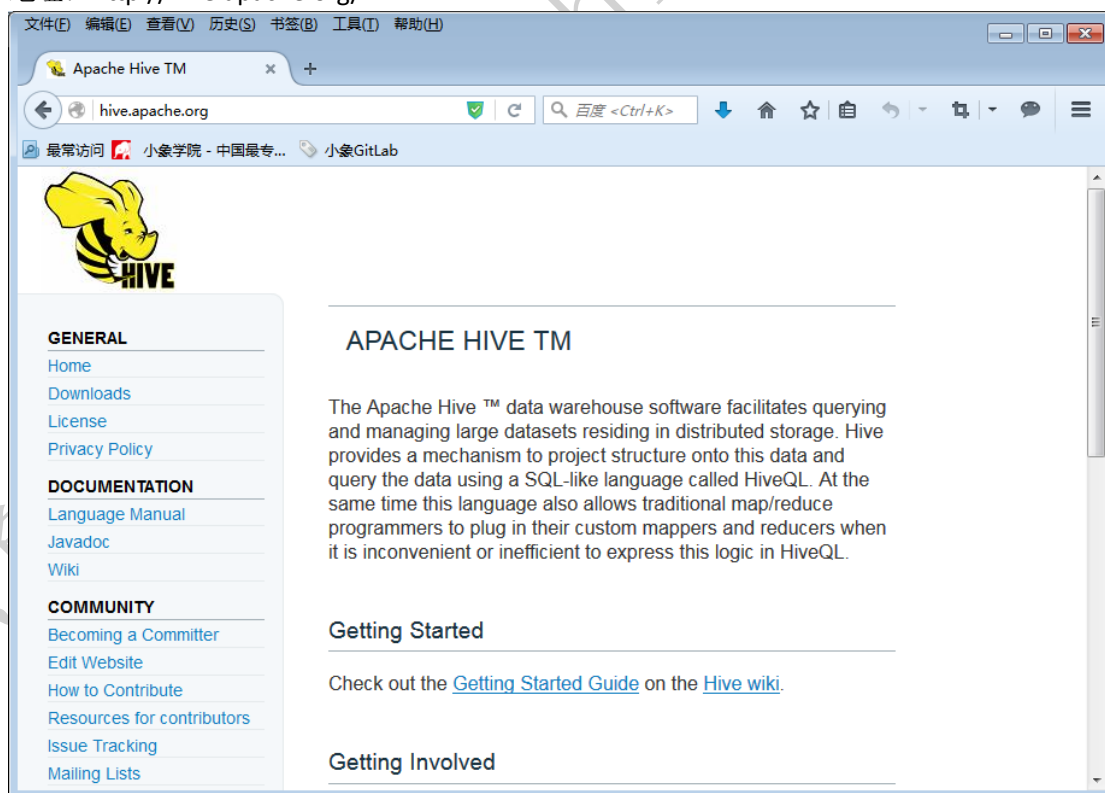
地址: <http://hbase.apache.org/>





1.3. Apache HIVE

地址: <http://hive.apache.org/>

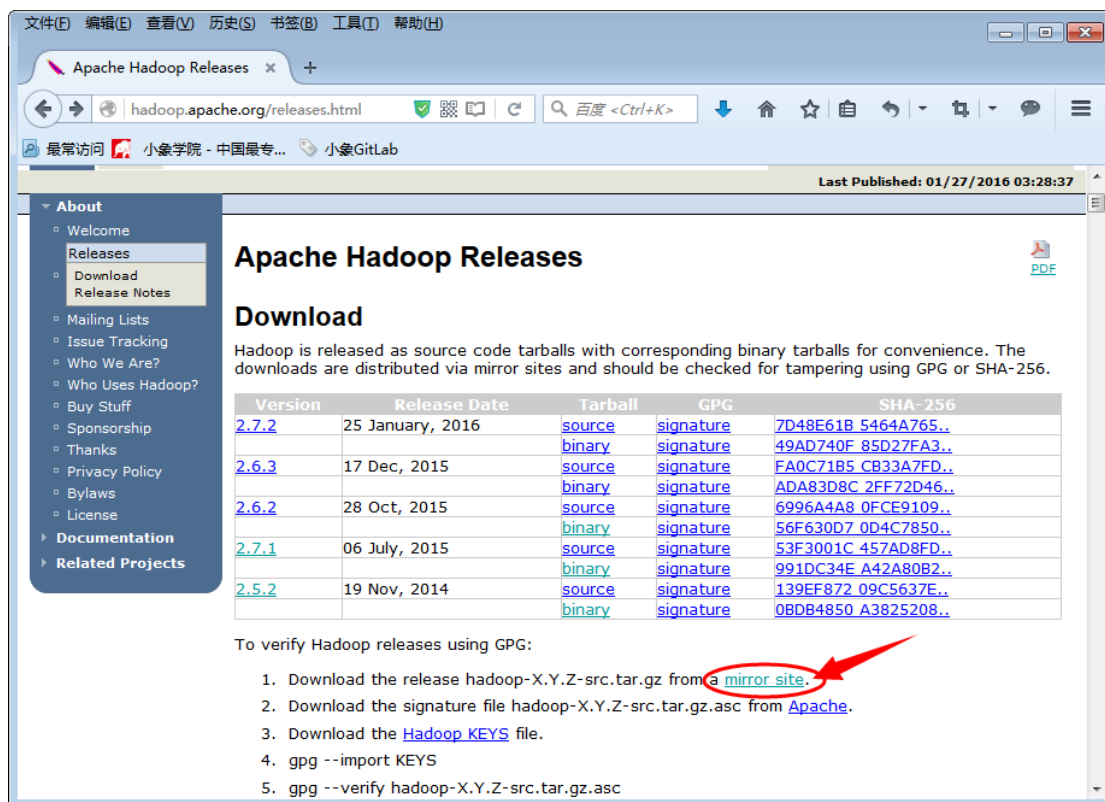


2. 官网上有那些 Releases 版本

2.1. Apache Hadoop Releases 版本

点击 Releases 进入页面 <http://hadoop.apache.org/releases.html>

点击镜像站 (mirror site)



The screenshot shows the Apache Hadoop Releases page in a web browser. The page title is "Apache Hadoop Releases". On the left, there is a navigation menu with links like "About", "Releases", "Download", "Release Notes", "Mailing Lists", "Issue Tracking", "Who We Are?", "Who Uses Hadoop?", "Buy Stuff", "Sponsorship", "Thanks", "Privacy Policy", "Bylaws", "License", "Documentation", and "Related Projects". The main content area has a "Download" section with a table of releases. Below the table, there are instructions on how to verify releases using GPG, with a red arrow pointing to the "mirror site" link in the first step.

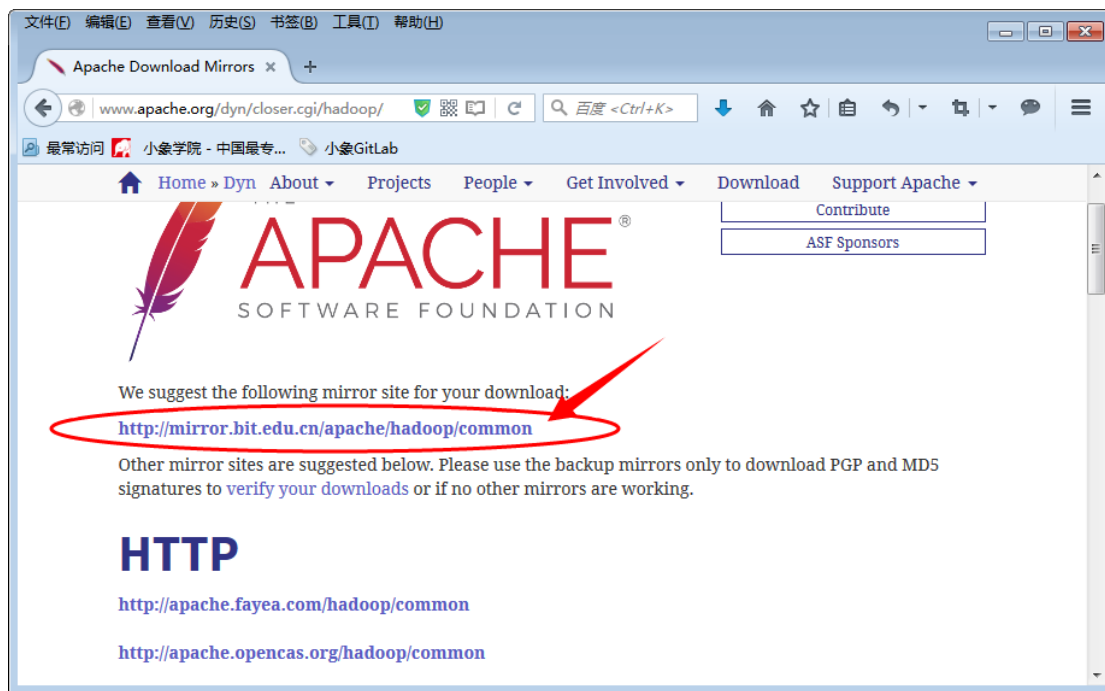
Version	Release Date	Tarball	GPG	SHA-256
2.7.2	25 January, 2016	source	signature	7D48E61B 5464A765..
		binary	signature	49AD740F 85D27FA3..
2.6.3	17 Dec, 2015	source	signature	FA0C71B5 CB33A7FD..
		binary	signature	ADA83D8C 2FF72D46..
2.6.2	28 Oct, 2015	source	signature	6996A4A8 0FCE9109..
		binary	signature	56F630D7 0D4C7850..
2.7.1	06 July, 2015	source	signature	53F3001C 457AD8FD..
		binary	signature	991DC34E A42A80B2..
2.5.2	19 Nov, 2014	source	signature	139EF872 09C5637E..
		binary	signature	0BDB4850 A3825208..

To verify Hadoop releases using GPG:

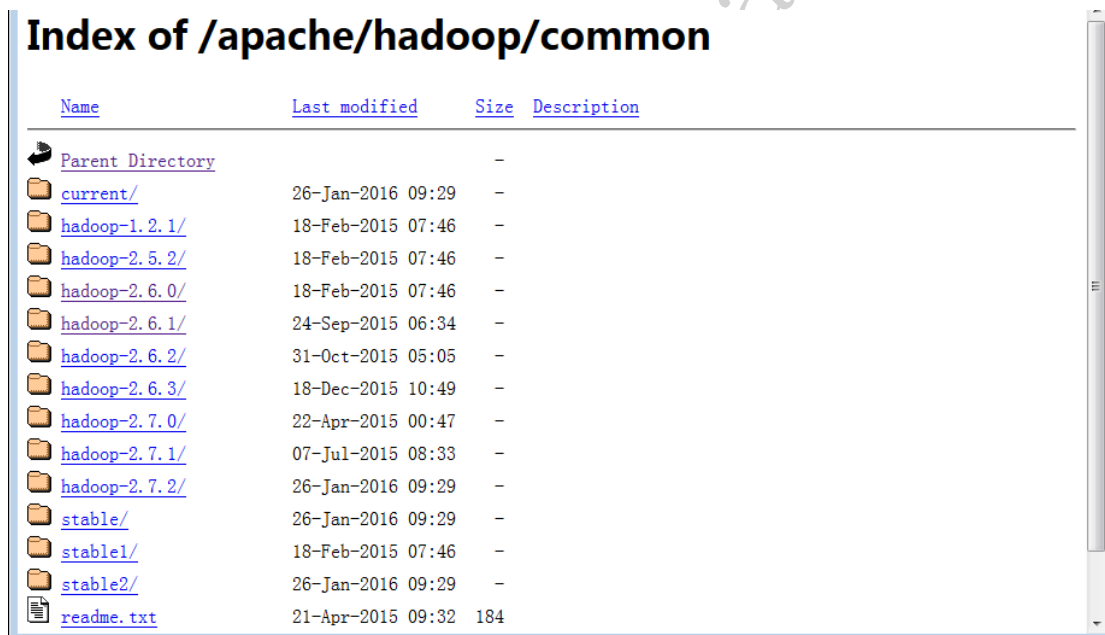
1. Download the release `hadoop-X.Y.Z-src.tar.gz` from [a mirror site](#).
2. Download the signature file `hadoop-X.Y.Z-src.tar.gz.asc` from [Apache](#).
3. Download the [Hadoop KEYS](#) file.
4. `gpg --import KEYS`
5. `gpg --verify hadoop-X.Y.Z-src.tar.gz.asc`

选择一个镜像站





Apache Hadoop 可以下载的版本



2.2. Apache Hive Releases 版本

Apache Hive 可以下载的 Releases 版本



Index of /apache/hive

Name	Last modified	Size	Description
 Parent Directory		-	
 hive-1.0.1/	21-May-2015 09:21	-	
 hive-1.1.1/	21-May-2015 09:22	-	
 hive-1.2.1/	27-Jun-2015 02:35	-	
 hive-parent-auth-hook/	28-Jan-2016 09:40	-	
 ldap-fix/	21-May-2015 21:43	-	
 stable/	27-Jun-2015 02:35	-	

Apache/2.2.22 (Ubuntu) Server at mirror.bit.edu.cn Port 81

2.3. Apache HBase Releases 版本

Apache HBase 可以下载的版本







[HBase Releases](#)

Please make sure you're downloading from [a nearby mirror site](#), not from [www.apache.org](#).

We suggest downloading the current [stable](#) release.

The 1.1.x series is the current stable release line, it supercedes 1.0.x, 0.98.x and 0.94.x (the 1.0.x, 0.98.x and 0.94.x lines are still seeing a monthly cadence of bug fix releases for those who are not easily able to update). Note that 0.96 was EOL'd September 1st, 2014.

For older versions, check the [apache archive](#).

Name	Last modified	Size	Description
 Parent Directory		-	
 0.98.17/	26-Jan-2016 05:17	-	
 1.1.3/	26-Jan-2016 13:57	-	
 hbase-0.94.27/	26-Mar-2015 08:21	-	
 hbase-1.0.3/	29-Jan-2016 05:49	-	
 HEADER.html	01-Sep-2015 06:11	679	

Apache/2.2.22 (Ubuntu) Server at mirror.bit.edu.cn Port 81

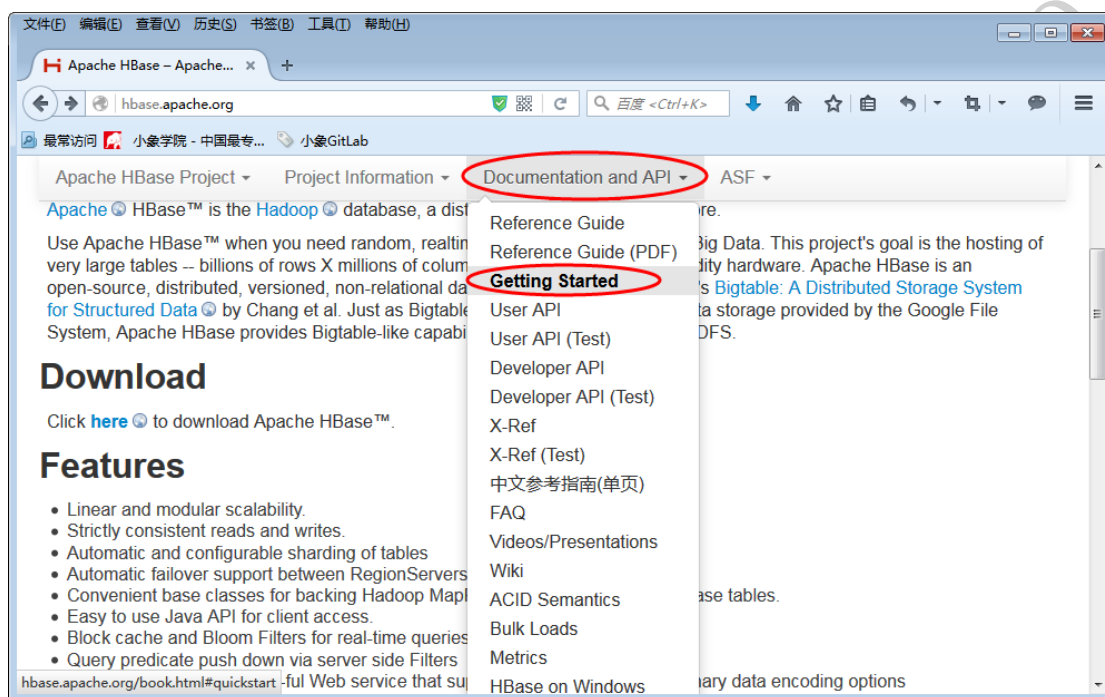


3. 各模块版本兼容性

3.1. 查看官方文档

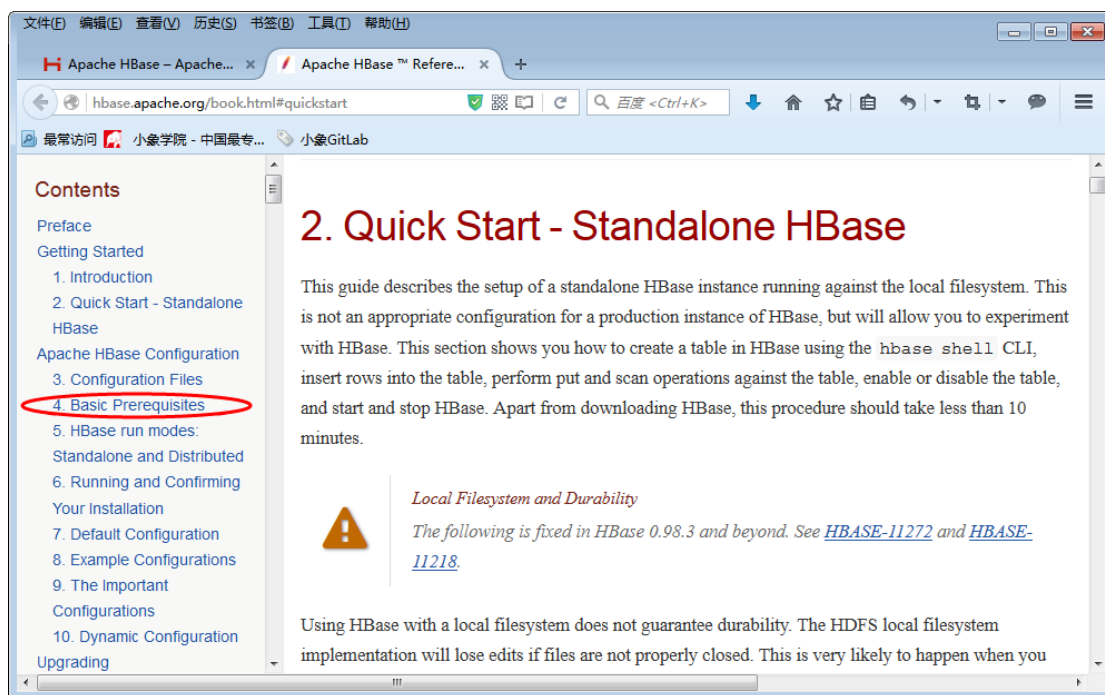
3.1.1. HBase 上手指南

在 Apache HBase 官网上找到 Getting Started（上手指南），点击它。

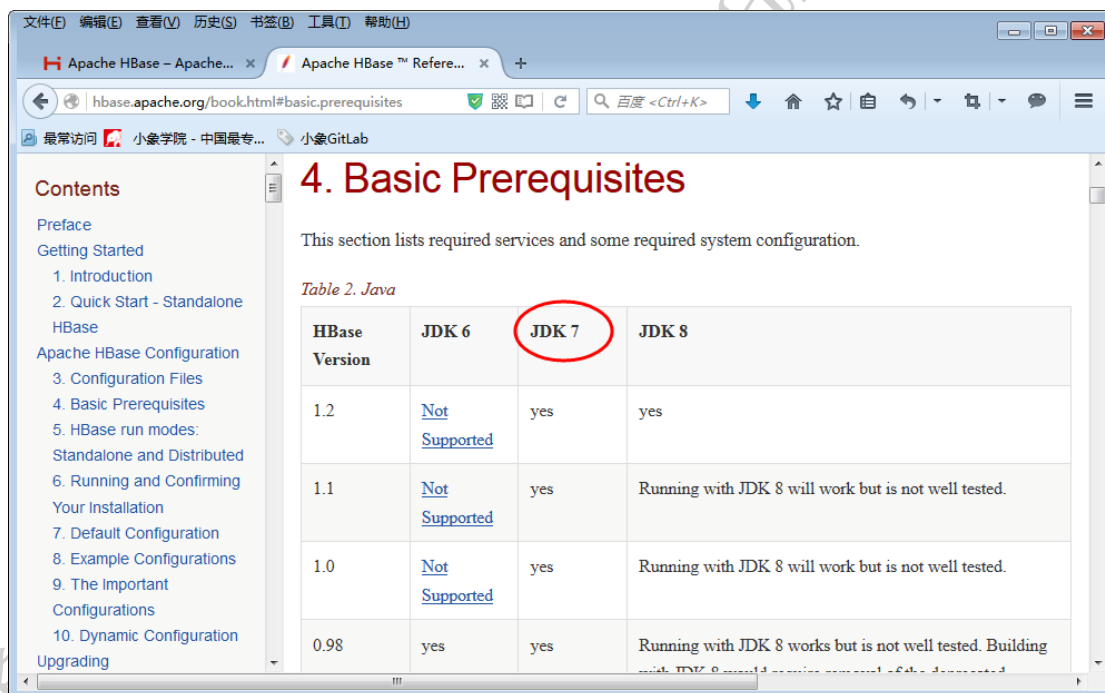


点击 Basic Prerequisites（基本前提）



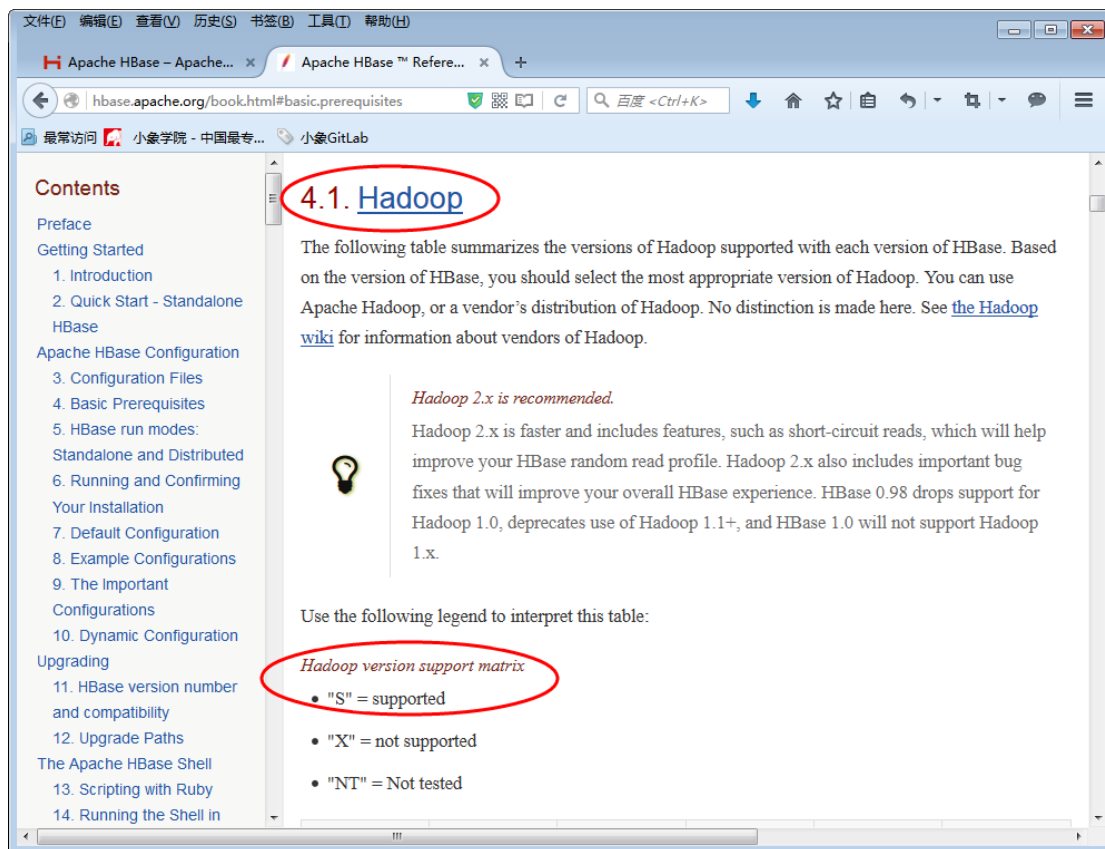


先看到 JDK 版本支持。



然后看到 Hadoop 版本支持。





下图是 Hadoop 修复了重要 bug 提高使用 HBase。

关于修复的重要 bug 后 HBase 各个版本和 Hadoop 各个版本对应关系表格如下：



Hadoop version support matrix

- "S" = supported
- "X" = not supported
- "NT" = Not tested

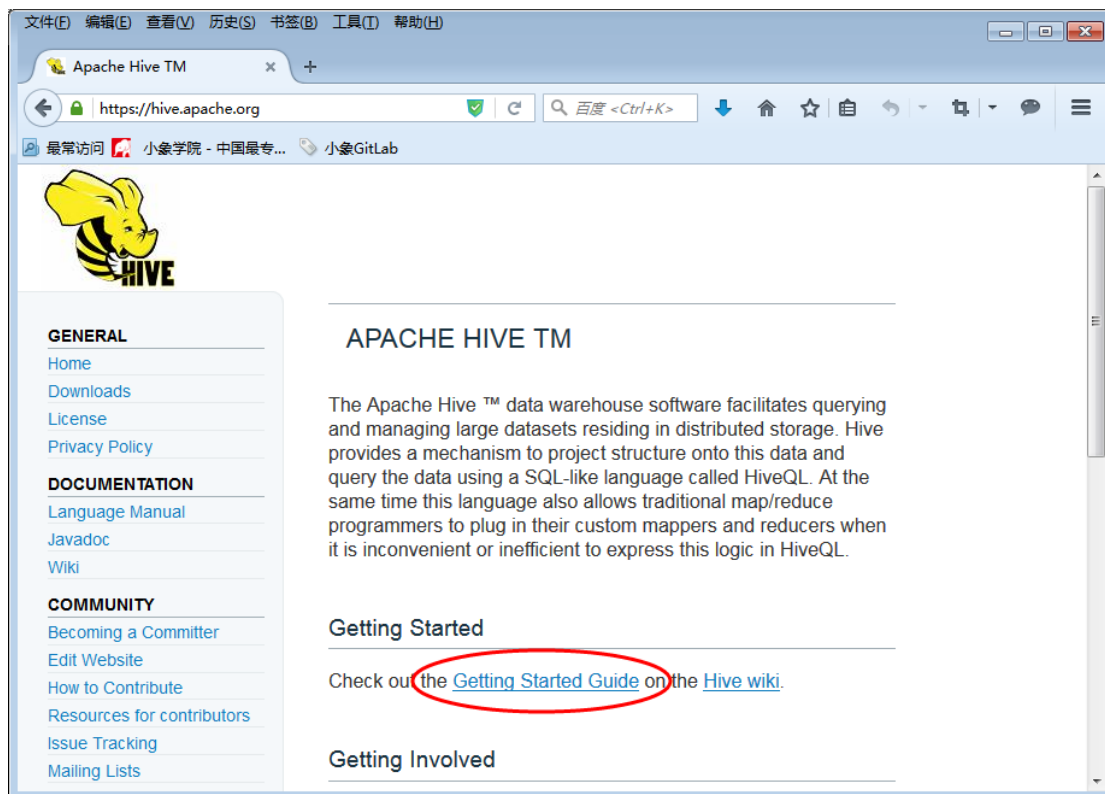
	HBase-0.94.x	HBase-0.98.x (Support for Hadoop 1.1+ is deprecated.)	HBase-1.0.x (Hadoop 1.x is NOT supported)	HBase-1.1.x	HBase-1.2.x
Hadoop-1.0.x	X	X	X	X	X
Hadoop-1.1.x	S	NT	X	X	X
Hadoop-0.23.x	S	X	X	X	X
Hadoop-2.0.x-alpha	NT	X	X	X	X
Hadoop-2.1.0-beta	NT	X	X	X	X
Hadoop-2.2.0	NT	S	NT	NT	NT
Hadoop-2.3.x	NT	S	NT	NT	NT
Hadoop-2.4.x	NT	S	S	S	S
Hadoop-2.5.x	NT	S	S	S	S
Hadoop-2.6.0	X	X	X	X	X
Hadoop-2.6.1+	NT	NT	NT	NT	S
Hadoop-2.7.0	X	X	X	X	X

X 和 NT 也是可以用的，只不过是 hadoop 和 hbase 结合使用时的一些 bug 没有修复。那么这里选择已经修复 bug 的版本，选择 Hadoop-2.5.X 版本。

3.1.2. HIVE 上手指南

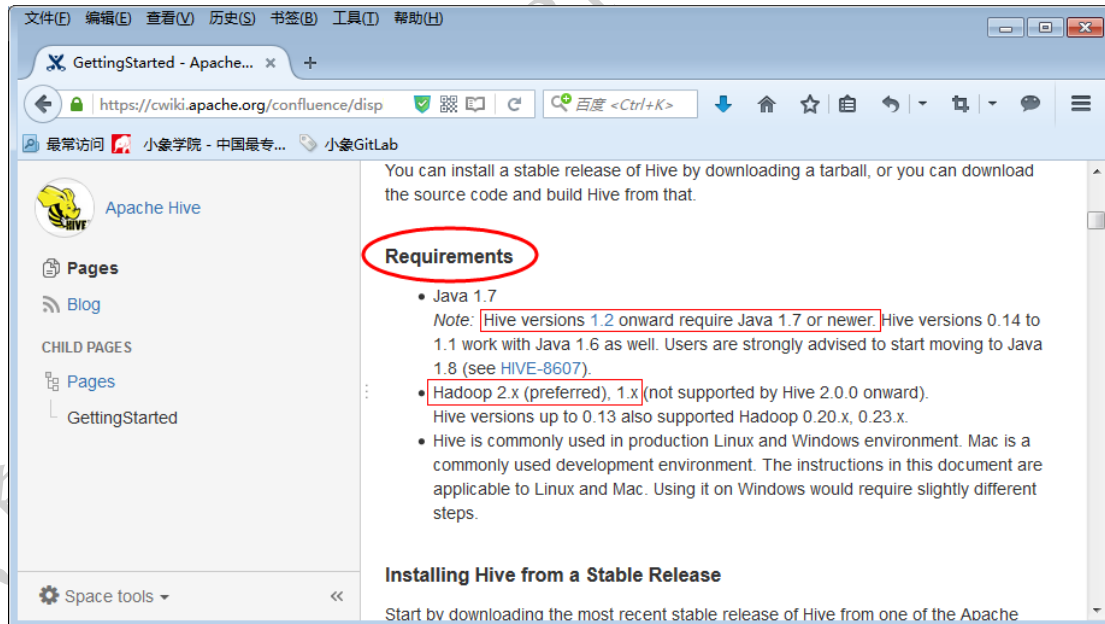
在 Apache HIVE 官网首页找到 Getting Started Guide，点击它。





看下 Requirements（要求）。

仔细阅读文档中，jdk 版本和 hadoop 版本的一些要求。

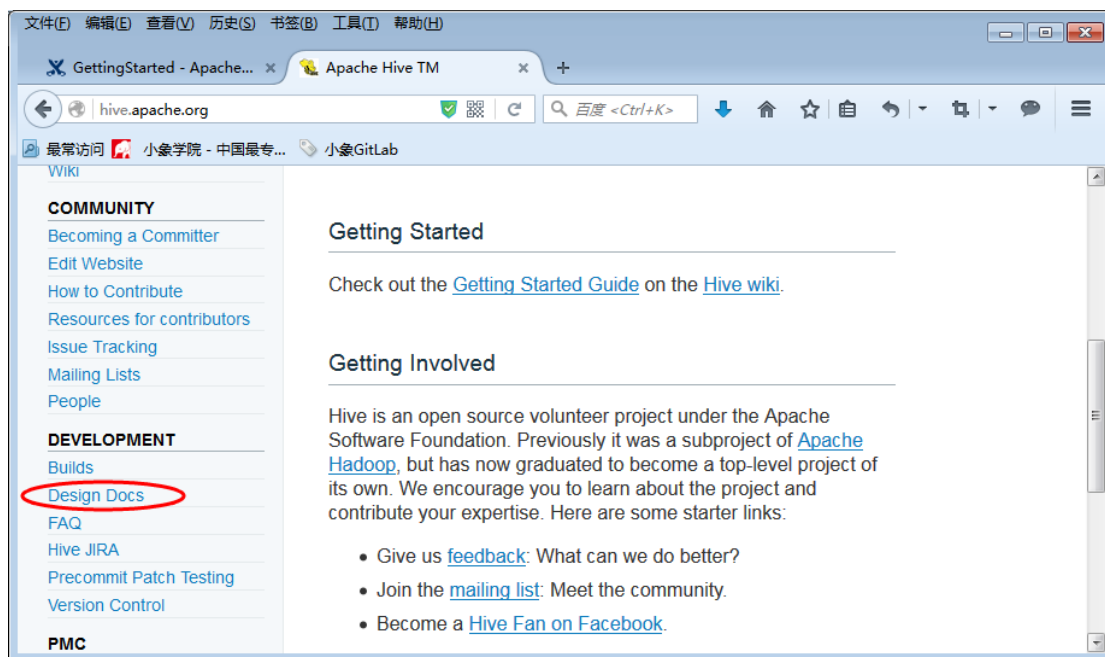


3.1.3. HIVE 和 HBase 整合

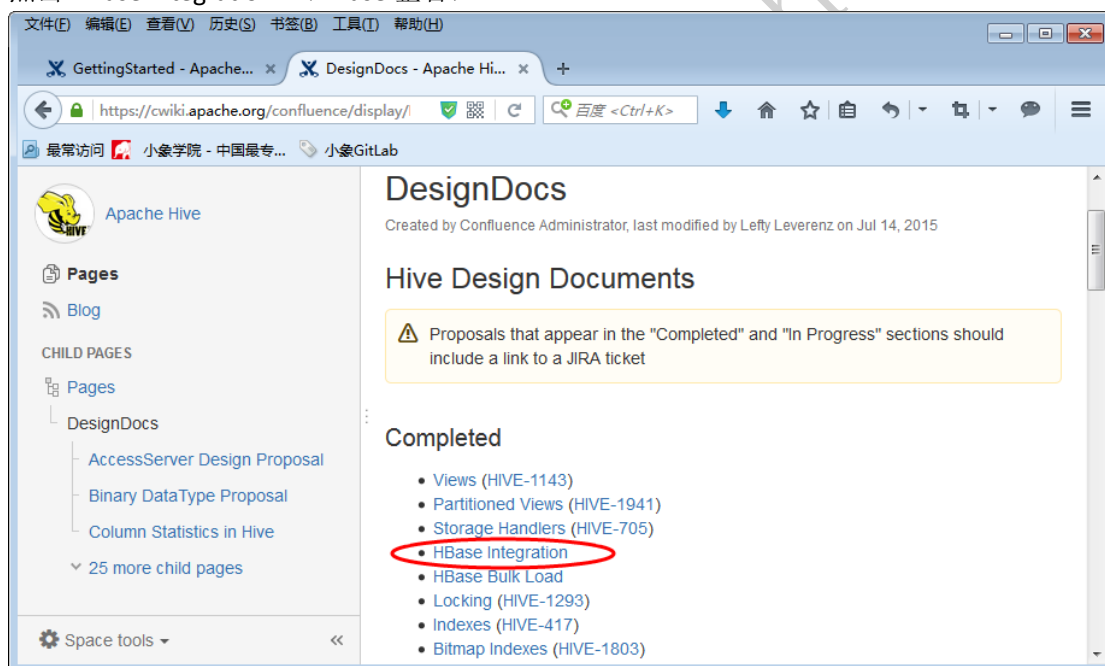
要 HIVE 和 HBase 整合，则还需要考虑他们的版本对应问题。

在 Apache HIVE 官网首页上找到 Design Docs（设计文档）点击它。



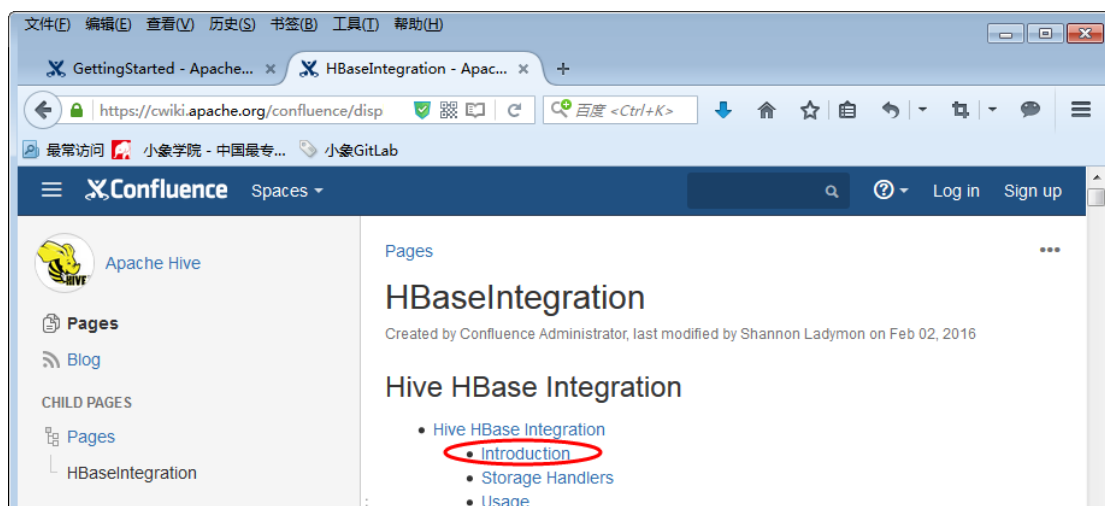


点击 HBase Integration （HBase 整合）



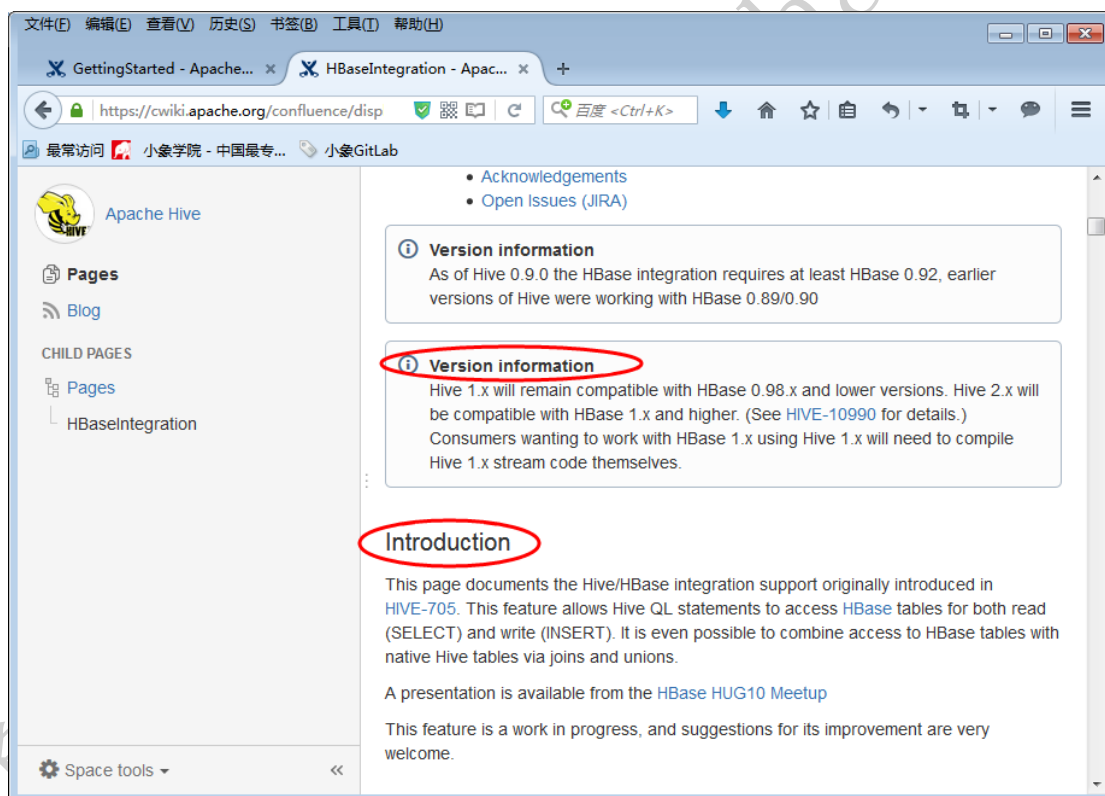
点击 Introduction （简介）





找到介绍和版本信息，仔细阅读下。文档中写到：

Consumers wanting to work with HBase 1.x using Hive 1.x will need to compile Hive 1.x stream code themselves.







大概意思是：用户需要编译 Hive 1.x stream code 才能使用 HBase 1.x 的版本。



3.2. 查看源码包中的 pom.xml 文件

3.2.1. 下载并解压源码包

	apache-hive-1.0.1-src	2016/2/5 18:39	文件夹	
	hbase-1.0.3	2016/2/5 18:39	文件夹	
	apache-hive-1.0.1-src.tar.gz	2016/2/5 18:34	好压 GZ 压缩文件	10,606 KB
	hbase-1.0.3-src.tar.gz	2016/2/5 18:39	好压 GZ 压缩文件	14,826 KB

3.2.2. 查看 HBase 中的 pom.xml

在 hbase-1.0.3 目录下的 pom.xml 文件中找到 hadoop-two.version

```

1125 <!-- Build dependencies -->
1126 <maven.min.version>3.0.3</maven.min.version>
1127 <java.min.version>${compileSource}</java.min.version>
1128 <!-- Dependencies -->
1129 <hadoop-two.version>2.5.1</hadoop-two.version>
1130 <hadoop-three.version>3.0.0-SNAPSHOT</hadoop-three.version>
1131 <commons-cli.version>1.2</commons-cli.version>
1132 <commons-codec.version>1.9</commons-codec.version>

```

说明当前下载的 hbase-1.0.3 版本是根据 hadoop-2.5.1 编译发布的。

3.2.3. 查看 HIVE 中的 pom.xml

在 apache-hive-1.0.1-src 目录下的 pom.xml 文件中找到 hadoop-23.version

```

115 <groovy.version>2.1.6</groovy.version>
116 <hadoop-20.version>0.20.2</hadoop-20.version>
117 <hadoop-20S.version>1.2.1</hadoop-20S.version>
118 <hadoop-23.version>2.5.0</hadoop-23.version>
119 <hadoop.bin.path>${basedir}/${hive.path.to.root}/testutils/hadoop</hadoop.bi
n.path>
120 <hbase.hadoop1.version>0.98.3-hadoop1</hbase.hadoop1.version>
121 <hbase.hadoop2.version>0.98.3-hadoop2</hbase.hadoop2.version>
122 <!-- httpcomponents are not always in version sync -->

```

说明当前下载的 hive-1.0.1 版本是根据 hadoop-2.5.0 编译发布的。

3.2.4. 总结

由此可见，在 Apache 官网上下下载到的各个模块，它们是基于不同版本的 Hadoop 编译发布的。

4. 选择 Apache Hadoop 版本

因为 hadoop2.5.0, hadoop2.5.1 和 hadoop2.5.2 它们的 API 变化不大，一些基本功能可



以运行，所以在练习搭建 `hadoop` 集群时，直接用它们也是可以运行的。

5. 补充

在生产环境搭建 `Apache Hadoop` 集群，需要重新编译源代码和版本兼容性测试。因为这些工作比较繁琐，所以企业会采用 `Cloudera` 的 `CDH` 或者 `Hortonworks` 的 `HDP`。

当然我们在练习时，可以用 `apache` 版本的。（若学会搭建集群后，建议用 `cdh` 版本的，小象训练营提供 `CDH` 版本的。）

