

# Tachyon (20:30准时开课)

--讲师: Yasaka

3月12日周末班 2月26日全日制班 3月19日 在线班 欢迎您的到来!

需要代码、PPT、视频等资料请加以下几位老师QQ:

贾老师: 1786418286

何老师: 1926106490

詹老师: 2805048645

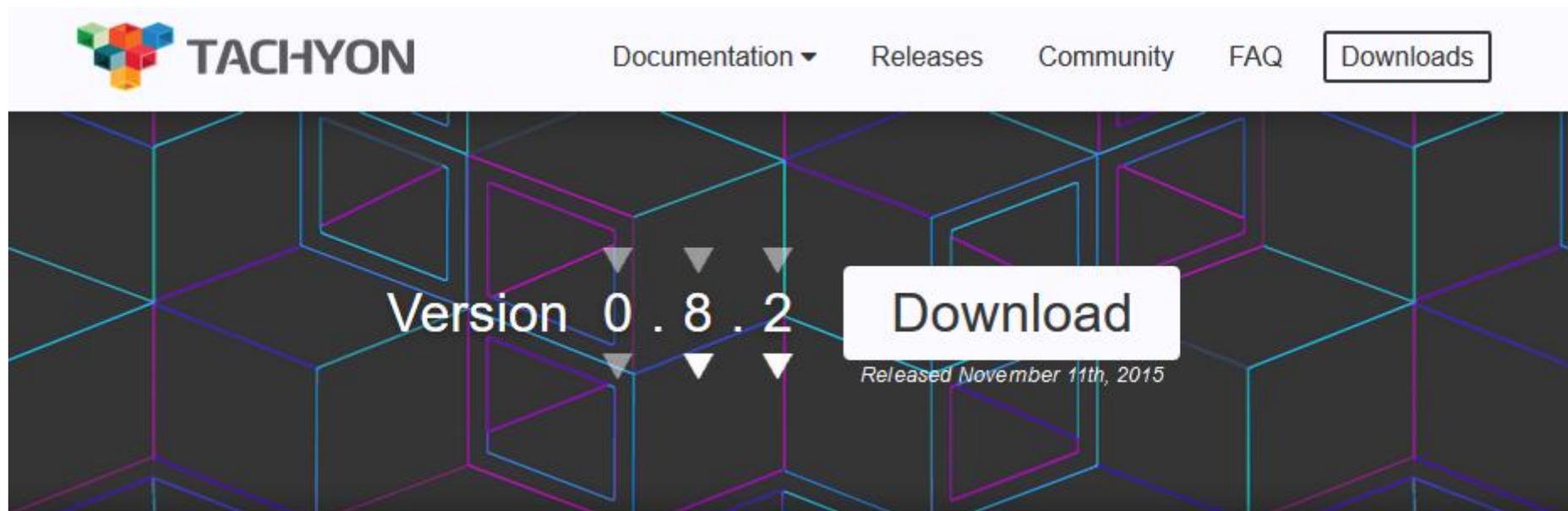
讨论技术可以加入以下QQ群: 172599077 , 156927834

热烈庆祝1221全日制班以及1226周末班爆满开班!!!

16.1.1日之后学费上调, 考虑春节后培训的学生可以提前报名, 预订座位, 不管以后何时过来学习, 费用都是以报名进的费用为主!



- 可以直接下载使用<http://tachyon-project.org/downloads/>



- <https://github.com/amplab/tachyon/releases>

Latest release

v0.8.2

bc59cc3

## Tachyon 0.8.2

 **apc999** released this on 12 Nov · **1923 commits** to master since this release

Release Date: November 11st, 2015

Please see <http://tachyon-project.org/releases> for detailed release notes.

For downloads, please checkout <http://tachyon-project.org/downloads/>

### Downloads

 [Source code \(zip\)](#)

 [Source code \(tar.gz\)](#)



- 为了更好地契合用户的本地环境，如java版本、hadoop版本或其他一些软件包的版本，可以下载Tachyon源码自行编译
- Tachyon项目采用Maven进行管理，因此可以采用 *mvn package* 命令进行编译打包
- 默认依赖的java版本为1.6，默认依赖的hadoop版本为1.0.4，如果要更改这些依赖的版本号可以在编译时加入选项
- `mvn clean package -Djava.version=1.7 -Dhadoop.version=2.5.0 -DskipTests`
- 编译好后我们可以在./assembly/target/中找到
- `tachyon-assemblies-0.8.2-jar-with-dependencies.jar`



- `cp conf/tachyon-env.sh.template conf/tachyon-env.sh`
- 需要配置文件并且配置变量
- `JAVA_HOME` —— 系统中java的安装路径和编译版本保持一致
- `TACHYON_MASTER_ADDRESS` —— 启动TachyonMaster的地址，默认为localhost，所以在单机模式下不用更改
- `TACHYON_UNDERFS_ADDRESS` —— Tachyon使用的底层文件系统的路径，在单机模式下可以直接使用本地文件系统，如`"/tmp/tachyon"`
- `vi conf/workers` 里面内容不用修改，默认为localhost



- 格式化文件
- `./bin/tachyon format`
- 启动
- `./bin/tachyon-start.sh local`
- 去确认正常运行,可以浏览<http://spark001:19999>
  
- 停掉Tachyon
- 使用`./bin/tachyon-stop.sh`



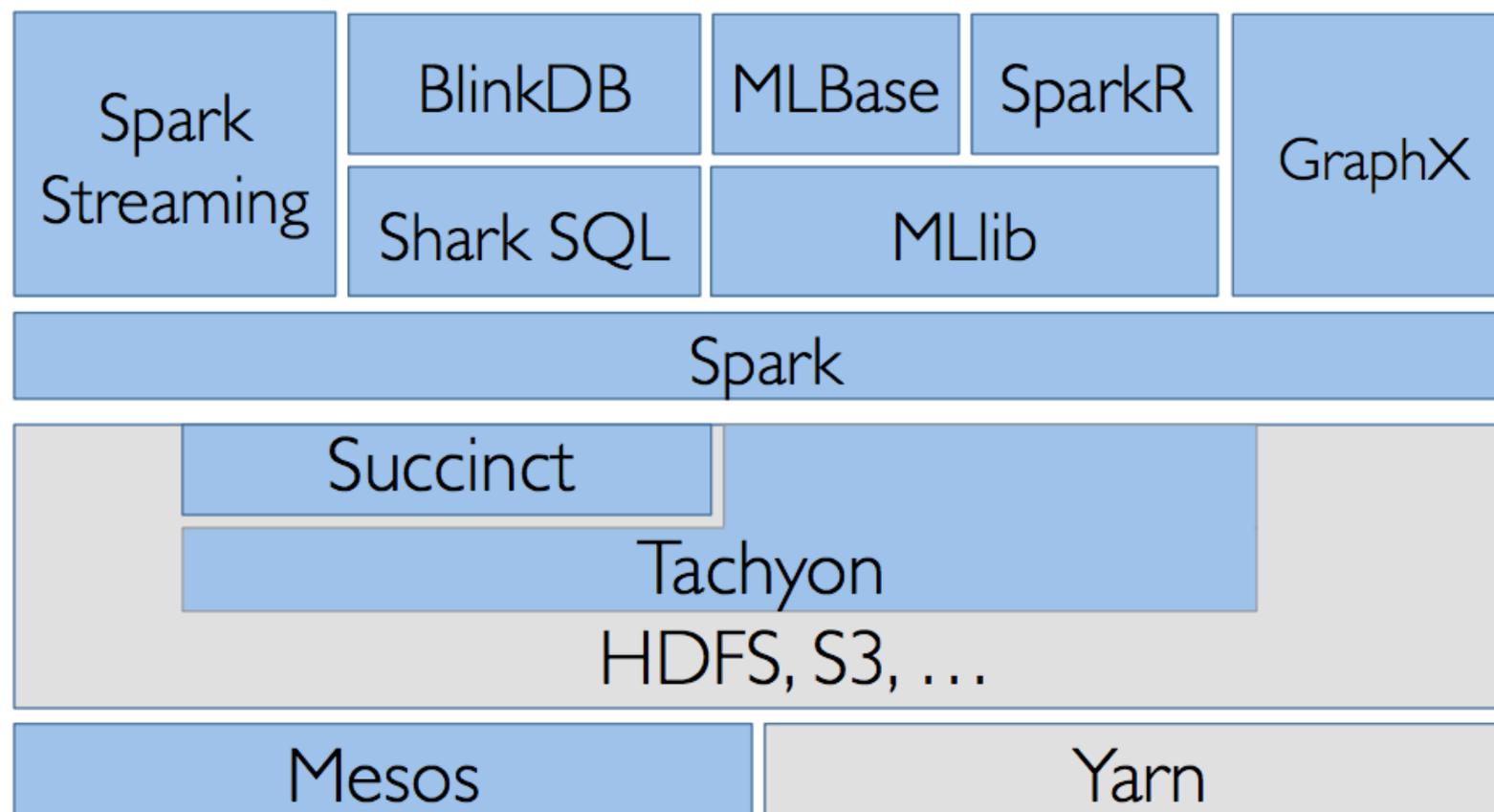
- `./bin/tachyon tfs`
- 从本地上传一个文件到Tachyon
- `./bin/tachyon tfs copyFromLocal LICENSE /LICENSE`
- 除了去页面上面确认，也可以通过命令行
- `./bin/tachyon tfs ls /`
- 也可以查看文件内容
- `./bin/tachyon tfs cat /LICENSE`
- <http://tachyon-project.org/documentation/Command-Line-Interface.html>

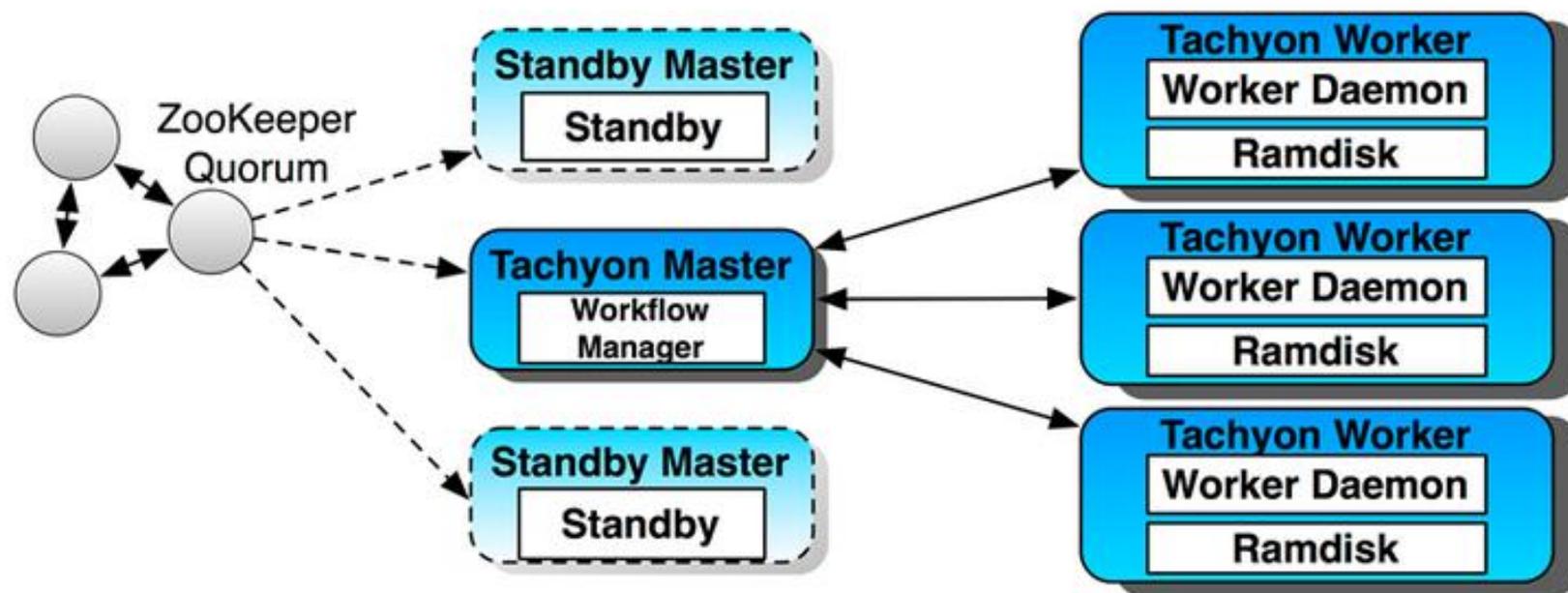


- cat : 将文件内容输出到控制台
- count : 输出符合路径前缀的文件总数
- ls : 输出目录中的文件信息
- lsr : 递归输出目录中的文件信息
- mkdir : 创建指定目录包括路径中的父目录, 如果目录已经存在则创建失败
- rm : 删除文件或者目录
- tail : 将文件的最末1KB输出到控制台
- touch : 在指定的位置创建空的文件
- mv : 将文件移动到指定位置
- copyFromLocal : 将文件从本地文件系统拷贝到Tachyon文件系统指定位置
- copyToLocal : 将文件从Tachyon文件系统拷贝到本地文件系统指定位置
- fileinfo : 打印指定文件的块信息
- pin : 将指定文件常驻内存
- unpin : 将常驻内存的文件撤销常驻状态



# Berkeley Data Analytics Stack





- vi conf/tachyon-env.sh
- TACHYON\_MASTER\_ADDRESS —— 启动TachyonMaster的地址，默认为localhost，所以在HDFS模式下修改为作为Tachyon集群里面Master的那台机器
- TACHYON\_UNDERFS\_ADDRESS —— Tachyon使用的底层文件系统的路径，在HDFS下配置为hdfs://master\_host:9000
- vi conf/workers 里面内容修改，默认为localhost，注释掉localhost之后，添加上Tachyon集群里面所有Works的那些机器



- 启动HDFS
- start-dfs.sh
- 格式化文件
- ./bin/tachyon format
- 启动
- ./bin/tachyon-start.sh all Mount
- 去确认正常运行,可以浏览<http://spark001:19999>
- ./bin/tachyon tfs copyFromLocal LICENSE /LICENSE



- Hadoop需要增加配置
- export  
HADOOP\_CLASSPATH=/pathToTachyon/clients/client/target/tachyon-client-0.8.2-jar-with-dependencies.jar
- 开启mapreduce
- start-yarn.sh
- hadoop jar /usr/hadoopsoft/hadoop-2.5.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.2.jar wordcount -libjars /usr/hadoopsoft/tachyon-0.8.2/clients/client/target/tachyon-client-0.8.2-jar-with-dependencies.jar tachyon://spark001:19998/LICENSE tachyon://spark001:19998/LICENSE-wc



- 查看结果
- `./bin/tachyon tfs ls /LICENSE-wc`
- `./bin/tachyon tfs cat /LICENSE-wc/part-r-00000`
- 除了可以配置tachyon-client-0.8.2-jar-with-dependencies.jar到HADOOP\_CLASSPATH中，另外一种方式就是可以将tachyon-client-0.8.2-jar-with-dependencies.jar拷贝到/usr/hadoopsoft/hadoop-2.5.2/share/hadoop/mapreduce/lib/
- `hadoop jar /usr/hadoopsoft/hadoop-2.5.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.2.jar wordcount tachyon://spark001:19998/LICENSE tachyon://spark001:19998/LICENSE-wc`



- 直接./bin/spark-shell --master local
- `val s = sc.textFile("tachyon://spark001:19998/LICENSE")`
- `s.count`
- 会报错 No FileSystem for scheme: tachyon
- `sc.hadoopConfiguration.set("fs.tachyon.impl", "tachyon.hadoop.TFS")`
- 加上上面一句不会报之前的错，但是会一直不停的输出ERROR：



- 版本问题，我们回看兼容性说明
- 立马发现我们使用spark-1.3.1是无法与tachyon-0.8.2兼容的
- 于是我们改用spark-1.5.0配合tachyon-0.7.1
- 另一个解决方案是修改pom.xml重新编译Spark

| Spark Version   | Tachyon Version |
|-----------------|-----------------|
| 1.0.x and Below | v0.4.1          |
| 1.1.x           | v0.5.0          |
| 1.2.x           | v0.5.0          |
| 1.3.x           | v0.5.0          |
| 1.4.x           | v0.6.4          |
| 1.5.x           | v0.7.1          |
| 1.6.x           | v0.8.2          |

```
<dependency>  
  <groupId>org.tachyonproject</groupId>  
  <artifactId>tachyon-client</artifactId>  
  <version>your_tachyon_version</version>  
  ...  
</dependency>
```



- 如何编译、配置、格式化tachyon-0.7.1这里不再赘述
- 启动tachyon
- `./bin/tachyon-start.sh local`
- 启动spark
- `./bin/spark-shell --master local`
- `val s = sc.textFile("tachyon://spark001:19998/testdata/people.txt")`
- `s.count`
- `s.saveAsTextFile("tachyon://spark001:19998/testdata/people2.txt")`



- 需要设置两个参数,
- spark.externalBlockStore.url
- spark.externalBlockStore.baseDir
- 1,可以配置到Spark的spark-defaults.conf文件中
- 2,可以通过代码设置到SparkConf中
- 3,也可以动态添加上,
- ./bin/spark-shell --master local --conf  
"spark.externalBlockStore.url=tachyon://spark001:19998"



- `sc.hadoopConfiguration.set("fs.tachyon.impl", "tachyon.hadoop.TFS")`
- `var file = sc.textFile("file:///usr/hadoopsoft/spark-1.5.0-bin-hadoop2.4/CHANGES.txt")`
- `val counts = file.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(_ + _)`
- `counts.persist(org.apache.spark.storage.StorageLevel.OFF_HEAP)`
- `counts.take(10)`
- `counts.take(10)`



```
[root@spark001 tachyon-0.7.1]# /usr/soft/jdk1.7.0_71/bin/java -jar tachyon-test.jar  
log4j:WARN No appenders could be found for logger ().  
log4j:WARN Please initialize the log4j system properly.  
0123456789
```

TACHYON Overview **Browse File System** System Configuration Workers In-Memory Files Log Files Enable Auto-Refresh

root tmp

| File Name  | Size    | Block Size | In-Memory  | Pin | Creation Time           | Modification Time       |
|--|---------|------------|--|-----|-------------------------|-------------------------|
|  test | 10.00 B | 512.00 MB  |  100% | NO  | 12-29-2015 21:59:33:310 | 12-29-2015 21:59:35:886 |

[View Settings](#)

Tachyon is an [open source](#) project developed at the UC Berkeley [AMPLab](#).



