



# Spark编译

源码下载:

<https://spark.apache.org/downloads.html>

注意:

1. Choose a Spark release:3.2.3(Nov 28,2022)  
这里是3.2版本，之所以不选择3.3是因为3.3兼容性不好，各种支持，版本配置有许多问题，比如不适合Intelij
2. Choose a package type:Source Code  
既然要魔改，自然是要用源码来编程
3. Download Spark:spark-3.2.3.tgz  
下载在windows本地就行，编译是在windows上，运行考虑在linux上进行.
4. 安装InteliJ,有学生版免费，直接下ultimate,不用扣扣索索用社区版
5. 需要配置java 1.8.0,直接在IDEA里面自助下载环境就行,scala支持在插件里面下载,sbt那些不要管,maven命令会自己给你下好，scala要用最新的.

官方编译文档:

<https://spark.apache.org/docs/latest/building-spark.html#building-a-runnable-distribution>

## [1]Spark参考教程

请先按这个参考教程配置POM,否则编译必错无疑，其中最为可恶的就是Failed to execute goal net.alchim31.maven:scala-maven-plugin

## [2]net.alchim31.maven报错

不要按这个参考教程来，你乱插dependency没用，而且语法也不对，这里主要是要在pom.xml里找net.alchim31.maven这个字段，只有三个条目，找到有version的地方，改成3.2.2即可.

编译命令:

在Spark源码根目录下打开GitBash

```
./build/mvn -Phadoop-3.2 -Pyarn -Dhadoop.version=3.2.2 -Phive -Phive-thriftserver  
-DskipTests clean package
```

此处需要20min，这个命令会自动将各种依赖包安装好，所以会占网络资源，需要在maven的settings里面配置阿里云镜像，参照[3].

```
MINGW64/d/YuriRepo/spark-3.2.3
[WARNING] Usually this is not harmful and you can skip these warnings,
[WARNING] otherwise try to manually exclude artifacts based on
[WARNING] mvn dependency:tree -Ddetail=true and the above output,
[WARNING] See http://maven.apache.org/plugins/maven-shade-plugin/
[INFO] Replacing original artifact with shaded artifact.
[INFO] Replacing D:\YuriRepo\spark-3.2.3\external\avro\target\spark-avro-2.12-3.2.3.jar with
D:\YuriRepo\spark-3.2.3\external\avro\target\spark-avro-2.12-3.2.3-shaded.jar
[INFO] dependency-reduced POM written at: D:\YuriRepo\spark-3.2.3\external\avro\dependency-re
duced-pom.xml
[INFO] --- maven-source-plugin:3.1.0:jar-no-fork (create-source-jar) @ spark-avro-2.12 ---
[INFO] Building jar: D:\YuriRepo\spark-3.2.3\external\avro\target\spark-avro-2.12-3.2.3-sourc
es.jar
[INFO] --- maven-source-plugin:3.1.0:test-jar-no-fork (create-source-jar) @ spark-avro-2.12 -
-
[INFO] Building jar: D:\YuriRepo\spark-3.2.3\external\avro\target\spark-avro-2.12-3.2.3-test-
sources.jar
[INFO] -----
[INFO] Reactor Summary for Spark Project Parent POM 3.2.3:
[INFO]
[INFO] Spark Project Parent POM ..... SUCCESS [01:43 min]
[INFO] Spark Project Tags ..... SUCCESS [ 13.929 s]
[INFO] Spark Project Sketch ..... SUCCESS [ 6.259 s]
[INFO] Spark Project Local DB ..... SUCCESS [ 3.722 s]
[INFO] Spark Project Networking ..... SUCCESS [ 6.413 s]
[INFO] Spark Project Shuffle Streaming Service ..... SUCCESS [ 4.996 s]
[INFO] Spark Project Unsafe ..... SUCCESS [ 9.446 s]
[INFO] Spark Project Launcher ..... SUCCESS [01:09 min]
[INFO] Spark Project Core ..... SUCCESS [03:21 min]
[INFO] Spark Project ML Local Library ..... SUCCESS [ 25.317 s]
[INFO] Spark Project GraphX ..... SUCCESS [ 30.724 s]
[INFO] Spark Project Streaming ..... SUCCESS [ 57.930 s]
[INFO] Spark Project Catalyst ..... SUCCESS [02:02 min]
[INFO] Spark Project SQL ..... SUCCESS [03:21 min]
[INFO] Spark Project ML Library ..... SUCCESS [01:47 min]
[INFO] Spark Project Tools ..... SUCCESS [ 4.502 s]
[INFO] Spark Project Hive ..... SUCCESS [01:24 min]
[INFO] Spark Project Hive Thrift Server ..... SUCCESS [ 37.484 s]
[INFO] Spark Project REPL ..... SUCCESS [ 15.551 s]
[INFO] Spark Project YARN Shuffle Service ..... SUCCESS [ 18.143 s]
[INFO] Spark Project YARN ..... SUCCESS [ 57.786 s]
[INFO] Spark Project Assembly ..... SUCCESS [ 12.934 s]
[INFO] Kafka 0.10+ Token Provider for Streaming ..... SUCCESS [ 18.137 s]
[INFO] Spark Integration for Kafka 0.10 ..... SUCCESS [ 22.349 s]
[INFO] Kafka 0.10+ Source for Structured Streaming ..... SUCCESS [ 35.767 s]
[INFO] Spark Prct Examples ..... SUCCESS [ 29.259 s]
[INFO] Spark Integration For Kafka 0.10 Assembly ..... SUCCESS [ 6.266 s]
[INFO] Spark Avro ..... SUCCESS [ 30.182 s]
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 22:18 min
[INFO] Finished at: 2023-04-12T11:00:23:08:00
[INFO] -----
YuriYZXBLAPTOP-PEQM95V MINGW64 /d/YuriRepo/spark-3.2.3
$ |
```

```
./dev/make-distribution.sh
--name spark-3.2.3
--tgz -Phive
-Phive-thriftserver
-Pyarn -Phadoop-3.2
-Dhadoop.version=3.2.2
-Dscala.version=2.12.14
```

此处需要15min

```
cp /d/YuriRepo/spark-3.2.3/conf/fairscheduler.xml.template /d/YuriRepo/spark-3.2.3/conf/log4j.properties
cp /d/YuriRepo/spark-3.2.3/conf/defaults.conf.template /d/YuriRepo/spark-3.2.3/conf/spark-env.sh.template /d/YuriRepo/spark-3.2.3/conf/worker
cp /d/YuriRepo/spark-3.2.3/README.md /d/YuriRepo/spark-3.2.3/dist
cp -r /d/YuriRepo/spark-3.2.3/bin /d/YuriRepo/spark-3.2.3/dist
cp -r /d/YuriRepo/spark-3.2.3/python /d/YuriRepo/spark-3.2.3/dist
[' false == true ']
cp -r /d/YuriRepo/spark-3.2.3/sbin /d/YuriRepo/spark-3.2.3/dist
[' -d /d/YuriRepo/spark-3.2.3/R/lib/SparkR ']'
[' true == true ']
TARDIR_NAME=spark-3.2.3-bin-spark-3.2.3
TARDIR=/d/YuriRepo/spark-3.2.3/spark-3.2.3-bin-spark-3.2.3
rm -rf /d/YuriRepo/spark-3.2.3/spark-3.2.3-bin-spark-3.2.3
cp -r /d/YuriRepo/spark-3.2.3/dist /d/YuriRepo/spark-3.2.3/spark-3.2.3-bin-spark-3.2.3
tar czf spark-3.2.3-bin-spark-3.2.3.tgz -C /d/YuriRepo/spark-3.2.3 spark-3.2.3-bin-spark-3.2.3
rm -rf /d/YuriRepo/spark-3.2.3/spark-3.2.3-bin-spark-3.2.3
```

成功会在目录下产生spark-3.2.3-bin-spark-3.2.3.tgz，解压文件的目录下面就是编译成功的结果。

1. [https://blog.csdn.net/liweihope/article/details/122863127?ops\\_request\\_misc=%7B%22request%5Fid%22%3A%22168126273516800182150439%22%2C%22scm%22%3A%2220140713.130102334.pc%5Fall.%22%7D&request\\_id=168126273516800182150439&biz\\_id=0&utm\\_medium=distribute.pc\\_search\\_result.none-task-blog-2\\_allfirst\\_rank\\_ecpm\\_v1~times\\_rank-8-122863127-null-null.142<sup>v82</sup>insert\\_down1,201<sup>v4</sup>add\\_ask,239<sup>v2</sup>insert\\_chatgpt&utm\\_term=spark%E6%BA%90%E7%A0%81%E4%BF%AE%E6%94%B9%E9%A1%B9%E7%9B%AE&spm=1018.2226.3001.4187](https://blog.csdn.net/liweihope/article/details/122863127?ops_request_misc=%7B%22request%5Fid%22%3A%22168126273516800182150439%22%2C%22scm%22%3A%2220140713.130102334.pc%5Fall.%22%7D&request_id=168126273516800182150439&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2_allfirst_rank_ecpm_v1~times_rank-8-122863127-null-null.142<sup>v82</sup>insert_down1,201<sup>v4</sup>add_ask,239<sup>v2</sup>insert_chatgpt&utm_term=spark%E6%BA%90%E7%A0%81%E4%BF%AE%E6%94%B9%E9%A1%B9%E7%9B%AE&spm=1018.2226.3001.4187) ↵
2. <https://www.cnblogs.com/war3blog/p/7864804.html> ↵
3. [https://blog.csdn.net/ITzhangminpeng/article/details/79936598?ops\\_request\\_misc=%7B%22request%5Fid%22%3A%22168126114516800226566956%22%2C%22scm%22%3A%2220140713.130102334..%22%7D&request\\_id=168126114516800226566956&biz\\_id=0&utm\\_medium=distribute.pc\\_search\\_result.none-task-blog-2\\_allsobaiduend~default-1-79936598-null-null.142<sup>v82</sup>insert\\_down1,201<sup>v4</sup>add\\_ask,239<sup>v2</sup>insert\\_chatgpt&utm\\_term=maven%E4%BF%AE%E6%94%B9%E9%98%BF%E9%87%8C%E4%BA%91%E9%95%9C%E5%83%8F&spm=1018.2226.3001.4187](https://blog.csdn.net/ITzhangminpeng/article/details/79936598?ops_request_misc=%7B%22request%5Fid%22%3A%22168126114516800226566956%22%2C%22scm%22%3A%2220140713.130102334..%22%7D&request_id=168126114516800226566956&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2_allsobaiduend~default-1-79936598-null-null.142<sup>v82</sup>insert_down1,201<sup>v4</sup>add_ask,239<sup>v2</sup>insert_chatgpt&utm_term=maven%E4%BF%AE%E6%94%B9%E9%98%BF%E9%87%8C%E4%BA%91%E9%95%9C%E5%83%8F&spm=1018.2226.3001.4187) ↵