**大数据实训环境安装**

**Spark**

|  |  |  |
| --- | --- | --- |
|  | 文件标识： |  |
| 当前版本： | 1.0 |
| 作 者： | 马卫花 |
| 完成日期： | 2022-10 |

**目 录**

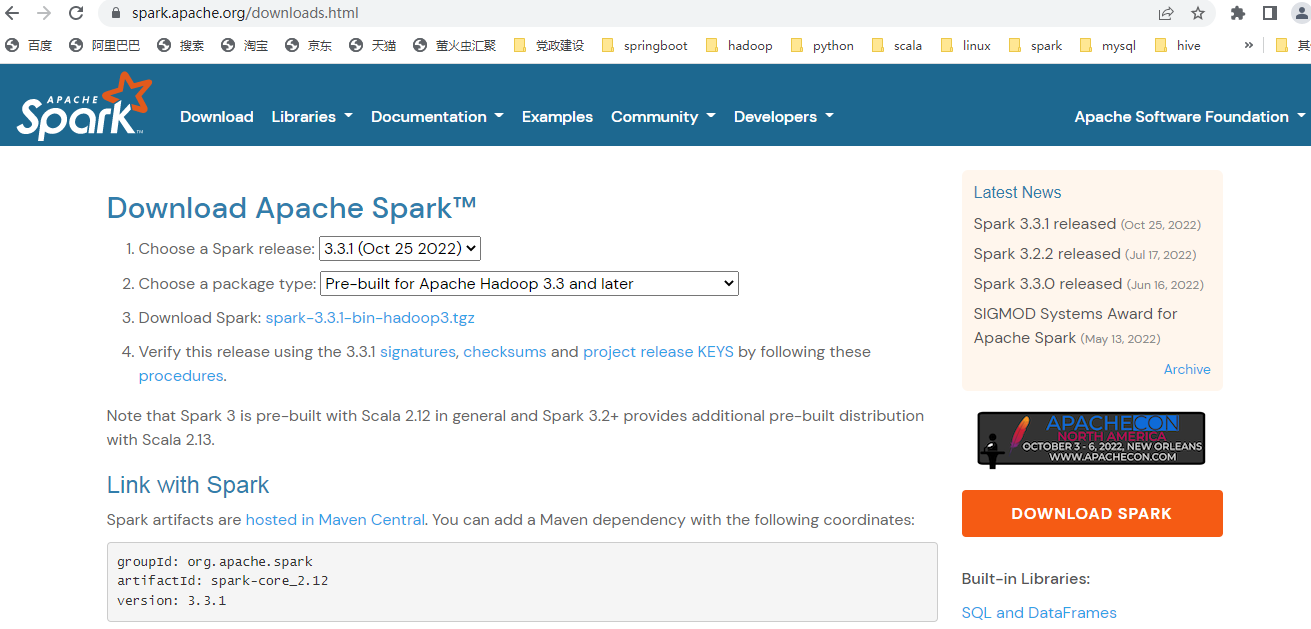
[1 Spark Local模式安装 3](#_Toc5758)

[2 Standalone环境配置 7](#_Toc2408)

[3 Spark运行示例 10](#_Toc24552)

# 1 Spark Local模式安装

1.从Spark官网下载Spark的安装文件，界面如下所示,下载spark-2.3.0-bin-hadoop2.7.tgz



2.下载后解压下载的文件如下图所示:

tar zxvf spark-2.3.0-bin-hadoop2.7.tgz

解压后的目录路径可按如下路径设置。

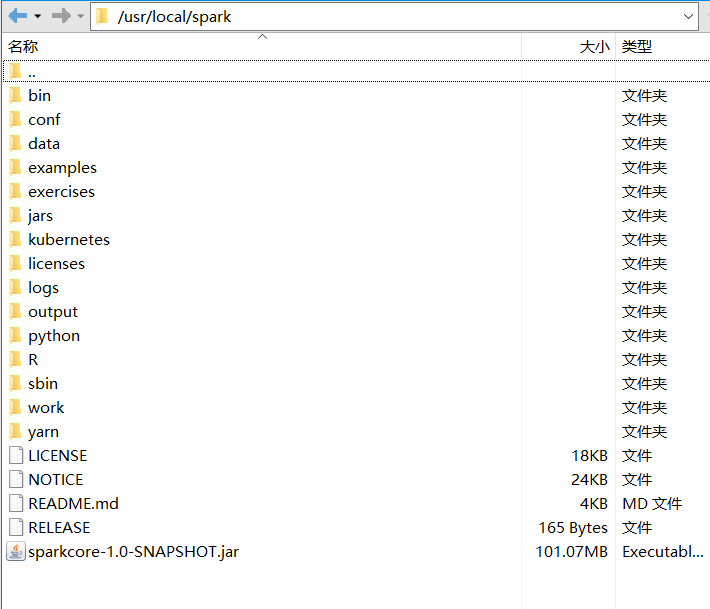
[root@master local]# cd spark

[root@master spark]# ls

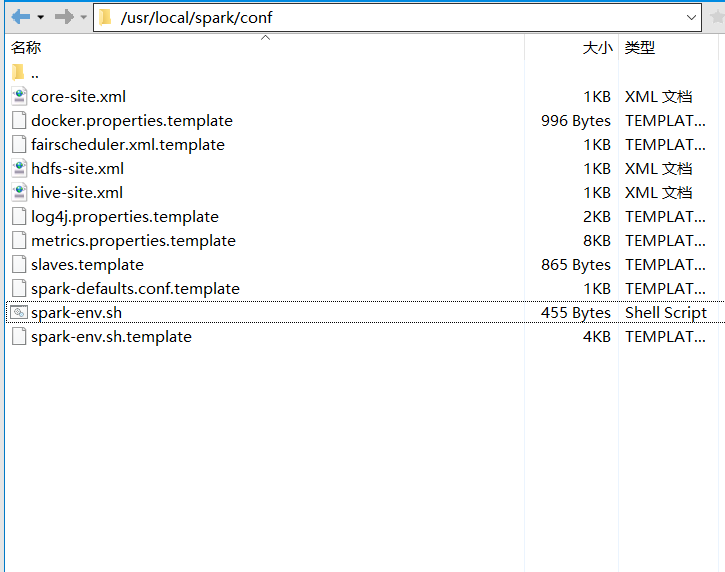
[root@master spark]# pwd

/usr/local/spark

Spark解压后的目录如下所示:



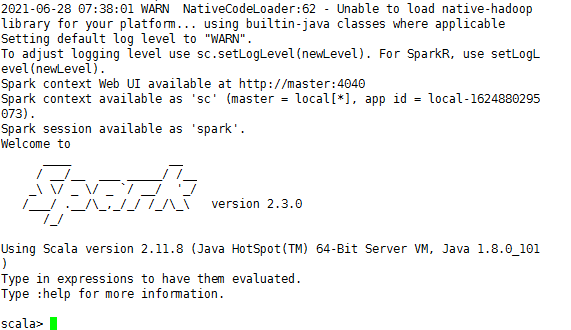
Spark目录中conf文件夹中的目录以及文件如下图所示：



3.配置环境变量如下所示



4.运行测试



# 2 Standalone环境配置

1.配置spark-env.sh

常见配置说明

|  |  |
| --- | --- |
| SPARK\_MASTER\_HOST=node01 | Spark的master运行在哪个机器上 |
| SPARK\_MASTER\_PORT=7077 | Spark master的通讯端口 |
| **SPARK\_MASTER\_WEBUI\_PORT**=8080 | Spark master的WebUI端口，默认端口是8080 |
| SPARK\_WORKER\_CORES=1 | Worker cpu可用内核数 |
| SPARK\_WORKER\_MEMORY=4G | 当前worker可用的总内存,默认值为1G |
| SPARK\_WORKER\_PORT=7078 | Worker的工作通讯地址 |
| SPARK\_WORKER\_WEBUI\_PORT=8081 | Worker的WebUI地址 |
| SPARK\_HISTORY\_OPTS=”  -Dspark.history.ui.port=18080  -Dspark.history.retainedApplications=30  -Dspark.history.fs.logDirectory=hdfs://node01:9000/spark-directory  -Dspark.history.fs.cleaner.enabled=true  ” | 设置历史服务器，配置的意思是将spark程序运行的历史日志存到hdfs的sparklog文件夹中 |

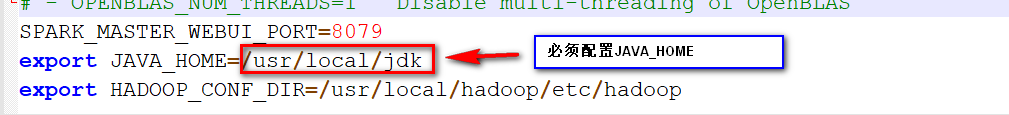
常见的配置说明:

Standalone集群常见配置https://spark.apache.org/docs/3.2.1/spark-standalone.html#cluster-launch-scripts

历史服务配置

<https://spark.apache.org/docs/3.2.1/monitoring.html#spark-history-server-configuration-options>

本实训案例的配置可参考如下的设置。



**提示:**

spark-env.sh文件必须配置JAVA\_HOME

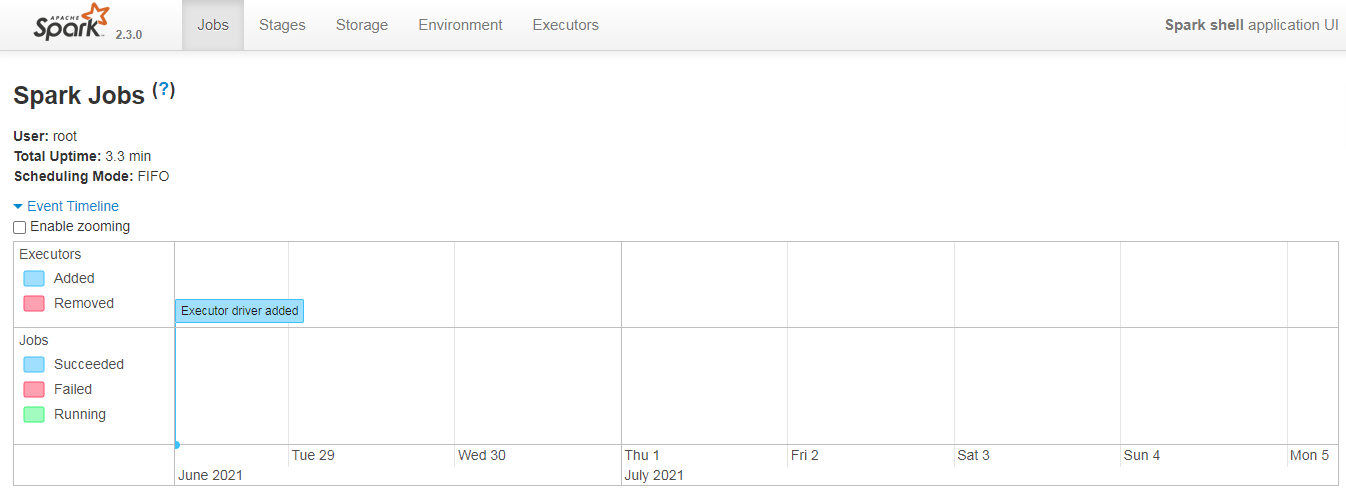
SPARK\_MASTER\_WEBUI：Spark Master的WEBUI端口(默认为8080)

**3.4 启动和停止命令**

启动spark，命令如下所示：

|  |
| --- |
| [root@master spark]# pwd  /usr/local/spark  [root@master spark]# cd sbin  [root@master sbin]# ./start-all.sh  [root@master bin]# ./spark-shell |

在浏览器的地址栏中输入http://192.168.174.15:4040/jobs/ ，显示如下的界面



关闭spark命令如下所示。

|  |
| --- |
| [root@master spark]# cd sbin  [root@master sbin]# ls  slaves.sh start-mesos-dispatcher.sh stop-master.sh  spark-config.sh start-mesos-shuffle-service.sh stop-mesos-dispatcher.sh  spark-daemon.sh start-shuffle-service.sh stop-mesos-shuffle-service.sh  spark-daemons.sh start-slave.sh stop-shuffle-service.sh  spark-warehouse start-slaves.sh stop-slave.sh  start-all.sh start-thriftserver.sh stop-slaves.sh  start-history-server.sh stop-all.sh stop-thriftserver.sh  start-master.sh stop-history-server.sh  [root@master sbin]# ./stop-all.sh |

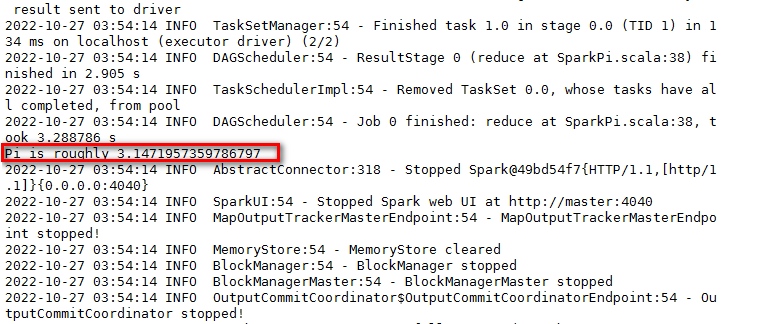
# 3 Spark运行示例

运行Spark自带的例子计算圆周率。

首先启动Spark服务，然后输入如下命令

**[root@master spark]# bin/spark-submit --class org.apache.spark.examples.SparkPi --master local --executor-memory 1G ./examples/jars/spark-examples\_2.11-2.3.0.jar**

运行后的结果如下所示：



【over】