# Spark环境部署与Spark Shell的基本使用

1. 部署

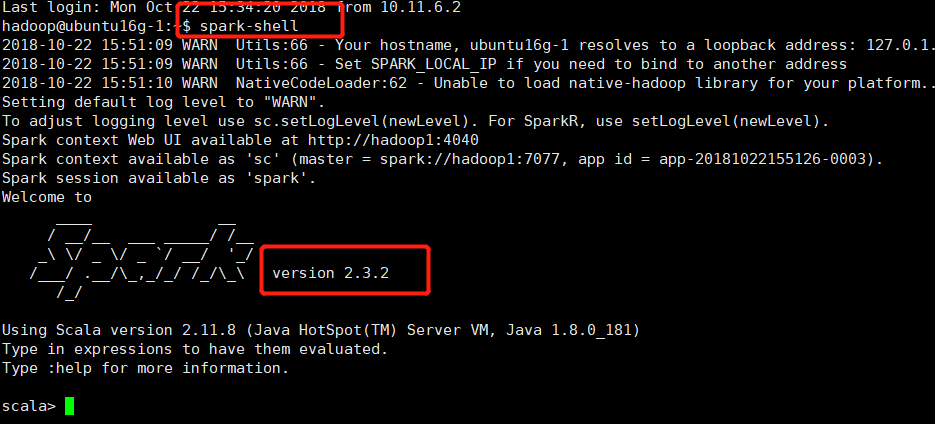
下载好安装文件后解压到/usr/local目录下

配置相关文件spark-env.sh，添加export SPARK\_DIST\_CLASSPATH=$(/usr/local/hadoop/bin/hadoop classpth)，这样spark就能把数据存储到hadoop分布式文件系统上，也可以从中读取数据，如果不配置这个，spark只能从本地读写数据

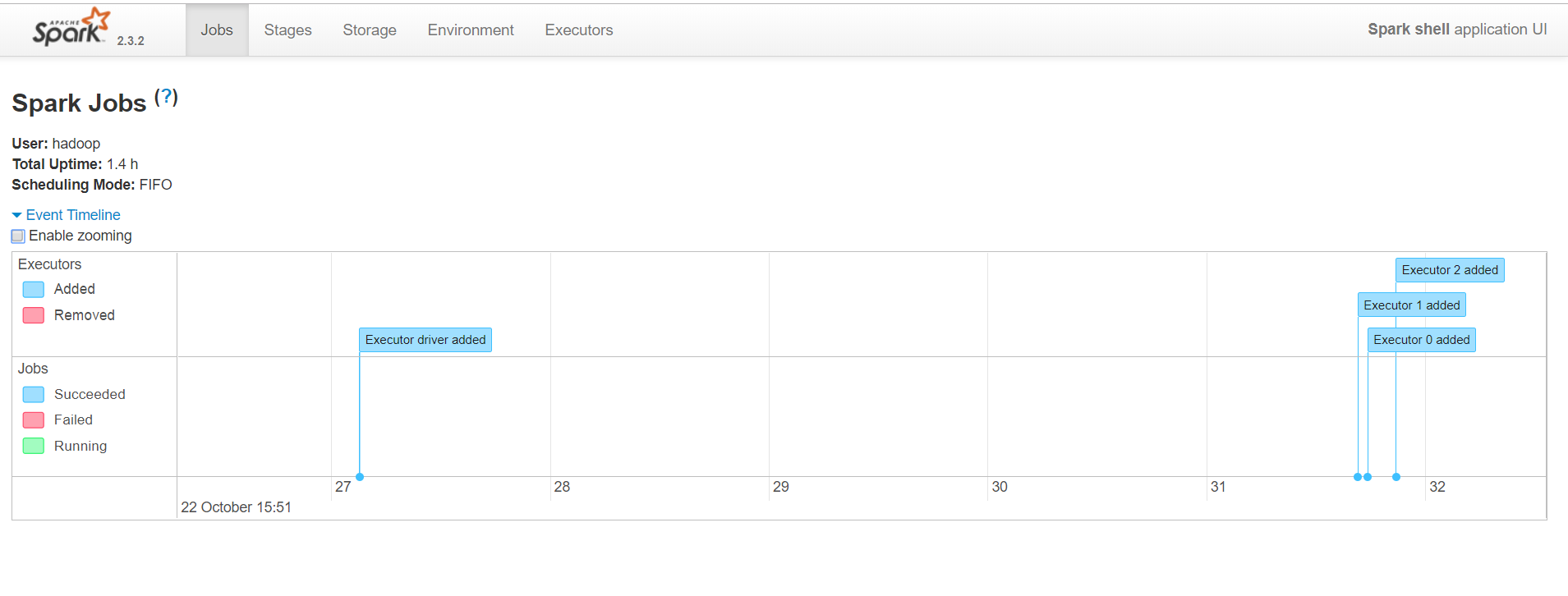
不需要像Hadoop那样进行启动命令就能运行Spark，执行run-example SparkPi

1. 使用Spark Shell

在Shell终端输入spark-shell，如果出现下面的就代表部署成功



spark-shell启动时会自动创建别名叫sc的spark context，并启动一个Spark的web ui，其默认端口号为4040，如果被占用，会自动+1



在集群上运行spark-shell

现在已经将/usr/local/spark下的README.md文件拷贝到HDFS的/spark\_txt目录下

在shell中执行spark-shell –master spark://10.11.6.91:7077

在spark-shell中输入下面的代码进行测试：(需要将hdfs://master:9000/README.md换成hdfs://10.11.6.91:9000ME.md)，可以在浏览器中输入10.11.6.91:8080以查看应用的运行情况



若记不得端口号可执行hdfs getconf -confKey fs.default.name