

华东师范大学数据科学与工程学院实验报告

课程名称： 分布式模型与编程	年级： 2016 级	上机实践成绩：
指导教师： 徐辰	姓名： 张宏伟	
上机实践名称： Spark 部署与 Shell 使用	学号： 10165101180	上机实践日期： 20181023
上机实践编号： 06	组号：	上机实践时间：

一、实验目的

在本地部署 Spark 环境，在集群上使用 SparkShell

二、实验任务

本地 spark 成功运行 SparkPick，集群上使用 SparkShell 读取本地和 HDFS 文件

三、使用环境

Spark2.3.2

四、实验过程

1.本地部署 spark

成功启动 spark-shell

```
hadoop@PC-honwee:/usr/local/spark$ ./bin/spark-shell
2018-10-23 09:26:40 WARN NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://PC-honwee:4040
Spark context available as 'sc' (master = local[*], app id = local-1540258005214).
Spark session available as 'spark'.
Welcome to

  ____ _
 / ___ \|
| |   | |
| |___| |
|  __ \|
| |___| |
|  __ \|
|  ___/
|_|

 version 2.3.2

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_181)
Type in expressions to have them evaluated.
Type :help for more information.

scala> 
```

跑样例程序测试结果

```
hadoop@PC-honwee:/usr/local/spark$ ./bin/run-example SparkPi 2>&1 | grep "Pi is roughly"
Pi is roughly 3.1418757093785468
```

2. 集群上使用 spark

直接读取统一上传的文本文件，并输出一行内容

```
scala> val textFile = sc.textFile("hdfs://10.11.6.91:9000/README.md")
textFile: org.apache.spark.rdd.RDD[String] = hdfs://10.11.6.91:9000/README.md MapPartitionsRDD[3] at textFile at <console>:24

scala> textFile.count()
res1: Long = 103

scala> textFile.first()
res2: String = # Apache Spark
```

自己上传文本文件到 HDFS

```
hadoop24@ubuntu16g-1:~/usr/local/hadoop$ ./bin/hdfs dfs -ls /user/zhw/spark_txt
18/10/26 09:55:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
Found 1 items
-rw-r--r--  3 hadoop24 supergroup        3809 2018-10-23 09:05 /user/zhw/spark_txt/README.md
```

用 Spark-shell 读取 HDFS 上的文本文件，并输出第一行内容

```
scala> val textFile=sc.textFile("/user/zhw/spark_txt/README.md")
textFile: org.apache.spark.rdd.RDD[String] = /user/zhw/spark_txt/README.md MapPartitionsRDD[3] at textFile at <console>:24

scala> textFile.first()
res2: String = # Apache Spark
```

五、总结

本地实验中，java 环境 JDK 当时用的是 10 版本，报错，后面改成 8 版本运行正常。JDK9、10 都不是长期支持更新的稳定版本，对于之前的应用，还是统一使用 8 较好。