**华东师范大学数据科学与工程学院实验报告**

|  |  |  |
| --- | --- | --- |
| **课程名称**：分布式编程模型与系统 | **年级**：2019 | **上机实践成绩**： |
| **指导教师**：徐辰 | **姓名**：赵煜硕 | **学号**：10195501415 |
| **上机实践名称**：Spark部署 **上机实践日期：** | |  |

很顺利无坑

**华东师范大学数据科学与工程学院实验报告**

|  |  |  |
| --- | --- | --- |
| **课程名称**：分布式编程模型与系统 | **年级**：2019 | **上机实践成绩**： |
| **指导教师**：徐辰 | **姓名**：赵煜硕 | **学号**：10195501415 |
| **上机实践名称**： Spark部署 **上机实践日期：** | |  |

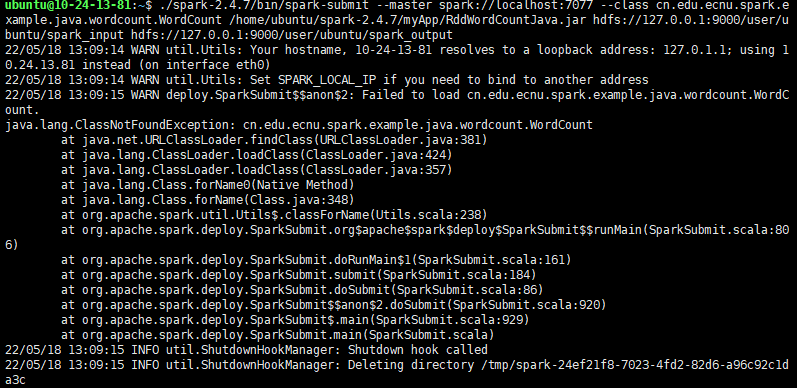
ScalaSDK死活装不上，解决方案：不用scala

执行

~/spark-2.4.7/bin/spark-submit

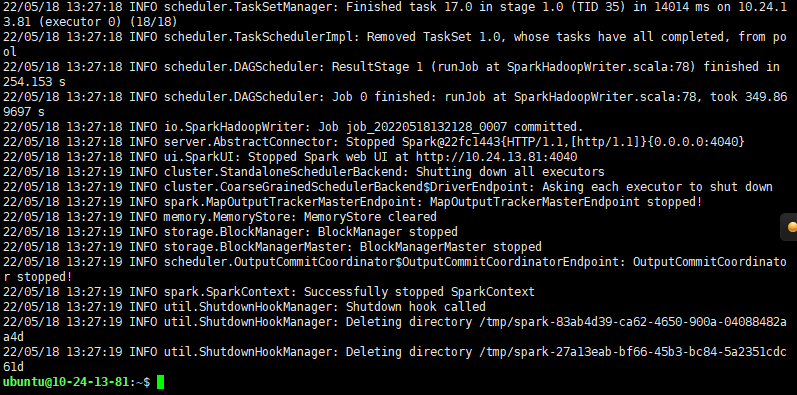
--master spark://localhost:7077 --class cn.edu.ecnu.spark.example.java.wordcount.WordCount

/home/ubuntu/spark-2.4.7/myApp/RddWordCountJava.jar hdfs://localhost:9000/user/ubuntu/spark\_input hdfs://localhost:9000/user/ubuntu/spark\_output



原因是在本地编写目录cn.edu.ecnu.spark.example.java.wordcount时出错

重新执行

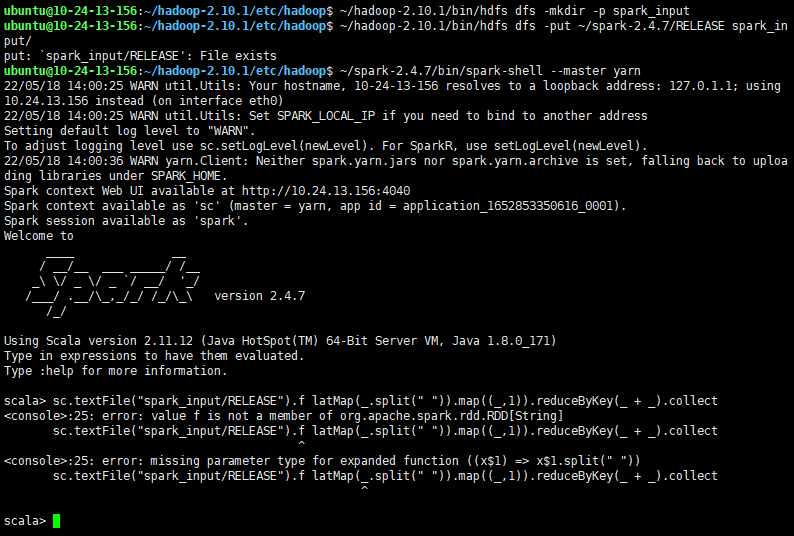


**华东师范大学数据科学与工程学院实验报告**

|  |  |  |
| --- | --- | --- |
| **课程名称**：分布式编程模型与系统 | **年级**：2019 | **上机实践成绩**： |
| **指导教师**：徐辰 | **姓名**：赵煜硕 | **学号**：10195501415 |
| **上机实践名称**：yarn部署 **上机实践日期：** | |  |

运行Spark程序

sc.textFile("spark\_input/RELEASE").flatMap(\_.split(" ")).map((\_,1)).reduceByKey(\_ + \_).collect



发现是书本印刷错误，faltMap连

重新执行

