

## 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：吴双	
上机实践名称：Yarn 框架下的系统部署	学号：10164102141	上机实践日期：
上机实践编号：#9	组号：23	上机实践时间：

### 一、实验目的

为 hadoop 和 spark 部署 yarn，并开启历史服务器

### 二、实验任务

使用 Yarn 本地部署 Hadoop；使用 Yarn 本地部署 Spark。

### 三、使用环境

Ubuntu LTS 18.04

Hadoop 2.7.3

### 四、实验过程

1. 修改 hadoop 的配置文件 mapred-site.xml，运行 jps 查看 yarn 是否配置成功：

```
$ start-dfs.sh
$ start-yarn.sh
$ jps
```

修改前：

```
hadoop@Master ~$ start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hadoop-na
menode-Master.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hadoop-da
tanode-Master.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-ha
doop-secondarynamenode-Master.out
hadoop@Master ~$ jps
5971 QuorumPeerMain
9211 SecondaryNameNode
8782 NameNode
8974 DataNode
9406 Jps
```

修改后：

```
hadoop@Master ~$ jps
11872 DataNode
14817 Jps
11681 NameNode
5971 QuorumPeerMain
12345 ResourceManager
12106 SecondaryNameNode
12686 NodeManager
```

2. 修改 spark-env.sh 文件, 解决虚存不够问题, 运行 spark-shell 不再报错, 可以使用

```

Warning: Master yarn-client is deprecated since 2.0. Please use master 'yarn' with specified deploy mode instead.
2018-11-19 15:38:26 WARN NativeCodeLoader:02 - 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 Spark2, use setLogLevel(newLevel).
2018-11-19 15:38:26 WARN Client:00 - Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.
Spark context Web UI available at http://PC-honwee:4040
Spark context available as 'sc' (master = yarn, app id = application_1542613023422_0001).
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_101)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val a = Array(1,2,3).parallelize()
=>res0: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <console>:24
scala>

```

3. 通过 yarn 提交 spark 作业(太长, 只截取部分)  
运行结果:

```

2018-11-19 16:14:28 INFO TaskSetManager:54 - Finished task 96.0 in stage 0.0 (TID 96) in 41 ms on PC-honwee (executor 2) (98/100)
2018-11-19 16:14:28 INFO TaskSetManager:54 - Finished task 98.0 in stage 0.0 (TID 98) in 23 ms on PC-honwee (executor 1) (99/100)
2018-11-19 16:14:28 INFO TaskSetManager:54 - Finished task 99.0 in stage 0.0 (TID 99) in 27 ms on PC-honwee (executor 2) (100/100)
2018-11-19 16:14:28 INFO YarnScheduler:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2018-11-19 16:14:28 INFO DAGScheduler:54 - ResultStage 0 (reduce at SparkPl.scala:38) finished in 2.409 s
2018-11-19 16:14:28 INFO DAGScheduler:54 - Job 0 finished: reduce at SparkPl.scala:38, took 2.513864 s
2018-11-19 16:14:28 INFO AbstractConnector:318 - Stopped Spark@5d00411a(HTTP/1.1,[http://1.1.1.1:4040])
2018-11-19 16:14:28 INFO SparkUI:54 - Stopped Spark web UI at http://PC-honwee:4040
2018-11-19 16:14:28 INFO YarnClientSchedulerBackend:54 - Interrupting monitor thread
2018-11-19 16:14:28 INFO YarnClientSchedulerBackend:54 - Shutting down all executors
2018-11-19 16:14:28 INFO YarnSchedulerBackendYarnDriverEndpoint:54 - Asking each executor to shut down

```

浏览器访问 localhost:8088 查看运行历史

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Black
application_1542613023422_0007	hadoop	Spark Pi	SPARK	default	Mon Nov 19 16:14:18 +0800 2018	Mon Nov 19 16:14:28 +0800 2018	FINISHED	SUCCEEDED		History	N/A
application_1542613023422_0006	hadoop	Simple	GDW.D	default	Mon Nov 19 16:14:18 +0800 2018	Mon Nov 19 16:14:28 +0800 2018	FINISHED	SUCCEEDED		History	N/A

## 五、总结

图形界面并没有比 terminal 界面好到哪里去, 差不多直观的感觉。