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： Yarn 使用	**学号：** 10165101180	**上机实践日期：**
上机实践编号： 09	**组号：**	**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

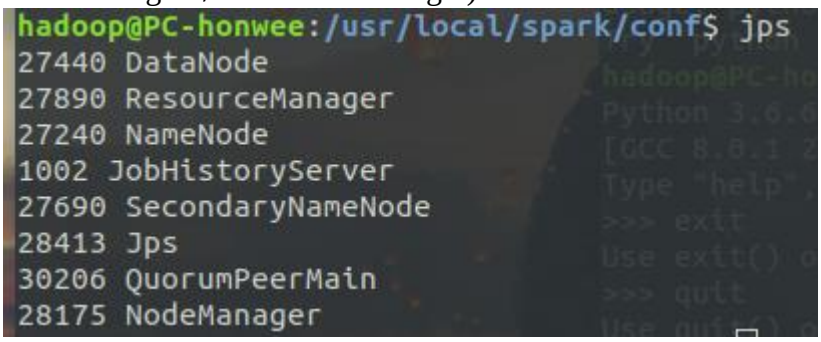
部署 yarn，并在配置 yarn 的 spark 下提交作业

三、使用环境

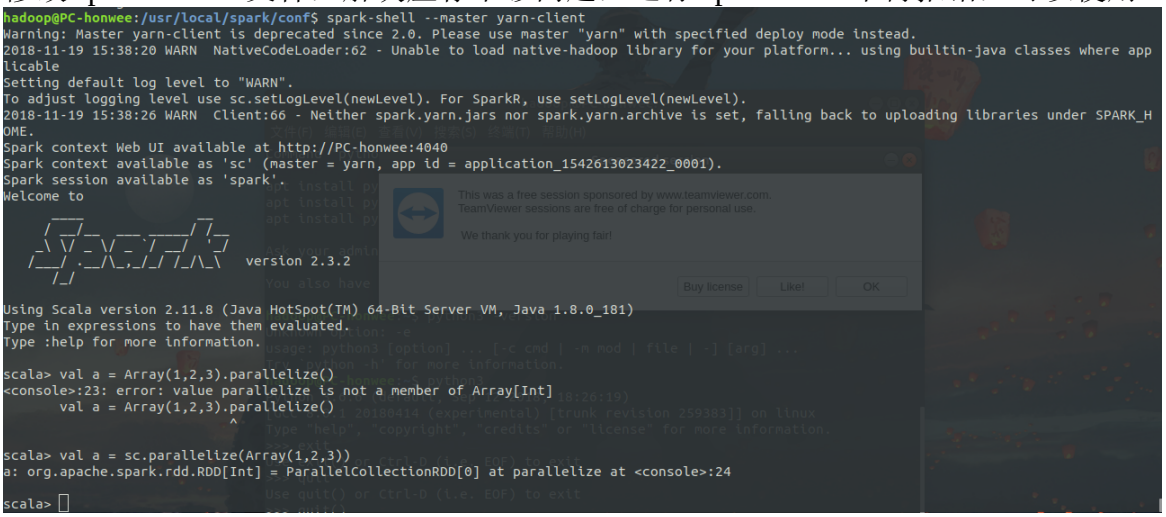
Hadoop2.7.7，spark2.3.2

四、实验过程

修改 hadoop 的配置文件 mapred-site.xml，运行 jps 查看 yarn 是否配置成功(多出 NodeManager 和 ResourceManager)



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



通过 yarn 提交 spark 作业(太长，只截取部分)

```
hadoop@PC-honwee:/usr/local/spark/examples/jars$ spark-submit --master yarn --class org.apache.spark.examples.SparkPi /usr/local/spark/e
xamples/jars/spark-examples_2.11-2.3.2.jar 100
2018-11-19 16:14:13 WARN  NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... using builtin-java classes whe
re applicable
2018-11-19 16:14:13 INFO  SparkContext:54 - Running Spark version 2.3.2
2018-11-19 16:14:13 INFO  SparkContext:54 - Submitted application: Spark Pi
2018-11-19 16:14:13 INFO  SecurityManager:54 - Changing view acls to: hadoop
2018-11-19 16:14:13 INFO  SecurityManager:54 - Changing modify acls to: hadoop
```

截取部分显示内容，其中 Pi is roughly 3.1415 为运行结果

```
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 SparkPi.scala:38) finished in 2.469 s
2018-11-19 16:14:28 INFO  DAGScheduler:54 - Job 0 finished: reduce at SparkPi.scala:38, took 2.513864 s
Pi is roughly 3.1415051141505113
2018-11-19 16:14:28 INFO  AbstractConnector:318 - Stopped Spark@5d00411a{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
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  YarnSchedulerBackend$YarnDriverEndpoint: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	Blackl
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	SPARK	default	Mon Nov 19	Mon Nov 19	FINISHED	SUCCEEDED		History	N/A

五、总结

解决虚存不够问题时，除了追加写入两个 property 外还要重启下 hadoop，不然还是会报错。