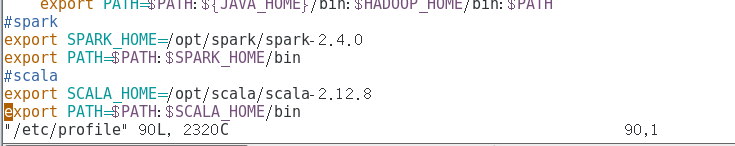
**作业9报告：**

**161278050 张旭**

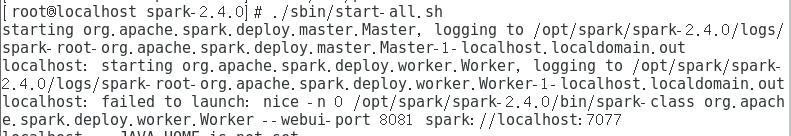
1.官网下载spark文件并解压

2.下载scala并解压

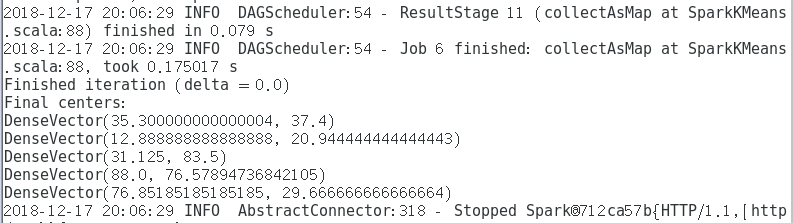
3.配置环境变量

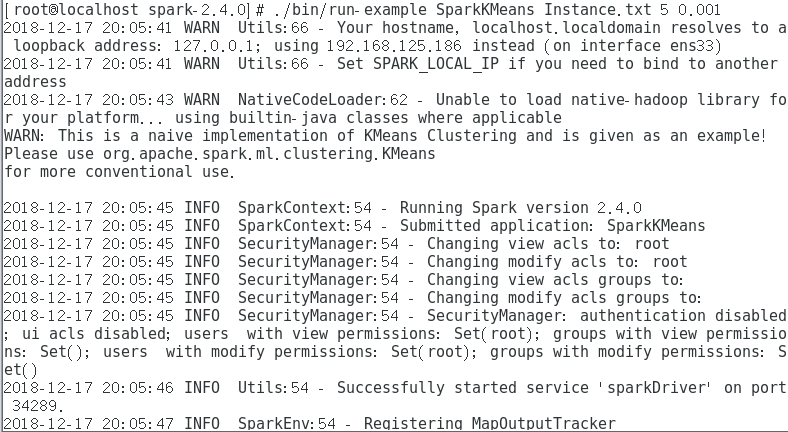


4.启动spark：



5.运行kmeans代码，观察结果和日志信息





6.与mapreduce的kmeans的性能比较：

与mapredce对比，spark的kmeans代码量更少，迭代次数更快，速度也快很多，效率更高。

|  |  |  |
| --- | --- | --- |
|  | Mapreduce | Spark |
| 中心节点数 | 5 | 5 |
| 迭代次数 | 10 | 5 |
| 耗时 | 6220523 | 175017 |