应用系统体系架构 — 作业12

# 学号： 522031910213 姓名： 朱涵 得分：

A.正确配置并运行统计功能

答：

## 使用docker部署spark集群，yml如下（worker类似）：

## 将书籍简介输入3个txt文件中，并编写python程序来进行mr作业，如下图：

## 

## 注意txt文件需要挂载到两个节点，因为是在本地容器运行而没有使用分布式文件系统。代码逻辑很简单，用flatmap方法将每一行文本都展开为关键词与出现次数的键值对，最后reduceByKey即可汇总成结果。最终结果如下：

## 

## 如下图是web页面中此任务的完成情况，可以看到确实被分发给了两个节点进行并发工作。

## 

## B.构建满足要求的RNN和LSTM模型

## 答：

## 仿照给出的示例即可，超参数基本都不变，除了记忆长度改为30天，最终训练结果如下：

## 

## 可以看到训练效果都很好，其中LSTM略低于RNN，可能是因为记忆长度较短或数据集较简单。