老师不会运行代码，重点看实验报告

核心代码必须使用JAVA，分割文档等预处理过程可以使用其他语言或软件工具

HDFS体系结构：为提高可靠性，可以设置备用nameNode

配置块副本以防止因DATANODE故障导致的数据丢失

实验报告的内容：（建议）

* 实验要求

实验内容概述一下

* 数据准备

数据预处理过程

* 环境的安装与配置

用的什么软件，用的哪个版本，抓重点，一个机器上多节点和多机器上多节点没什么区别

* 算法及实现
* 运行结果与分析
* 总结（心得、体会）