Hadoop大数据技术实践

# 1.实验目的

巩固前导课程Hadoop大数据技术的学习效果，根据前导课程学习的Hadoop安装步骤和配置事项，独立完成Hadoop平台及此次实践所需组件的部署。

掌握Hadoop系统的数据导入、预处理、统计和导出的全过程，能够运用Hadoop

平台的各种工具组件解决实际问题。

# 2.实验平台

Hadoop版本：3.1.3

# 实验内容

现有格式为docx的文档数份，请使用Hadoop相关技术实现以下功能：

1. 按从高到低的顺序进行词频排序，并显示出现频率最高的3个**词语**，例如：

大数据 1000

技术 999

实践 888

1. 输入词语后返回对应**词语**的词频搜索结果，并将结果导出到txt文档中。

输入：大数据

输出：1000

1. 输入词语后，根据对应**词语**在每篇文档的频率，按从高到低的顺序显示对应**词语**出现频率最高的3份文档名，并将结果导出到txt文档中。

输入：大数据

输出：文档a（100次） 文档b（99次） 文档c（98次）