2023-2024第1学期大数据系统开发考试说明

*本次考试的重点内容是基本原理和实验，所列出的内容要深入理解基本概念（没有列出的内容基本不考），这些基本概念包括实验中的很多内容，如：如何设置NameNode。除此之外，还要求掌握HDFS、HBase的基本操作命令。请大家在准备考试时按下面内容进行，特别是实验部分。*

1. **大数据基本概念**
2. 大数据特征
3. 关系数据库与NoSQL数据库的差异
4. 大数据库关键技术
5. 大数据处理流程
6. 大数据系统架构
7. **Hadoop**
8. 什么是Hadoop
9. Hadoop的核心架构
10. Hadoop生态圈
11. Hadoop的优点
12. Hadoop处理大数据的优势
13. **什么是HDFS**
14. HDFS的目标
15. HDFS体系结构
16. HDFS块：大小、副本
17. HDFS读写流程
18. HDFS创建目录、列表目录，传送文件的基本命令。
19. **什么是HBase**
20. HBase的特点，可扩展性、性能等。
21. HBase体系结构
22. HBase压缩
23. HBase日志
24. HBase读写流程
25. HBase创建表、增、删、改、查记录的命令。
26. **MapReduce**
27. MapReduce是什么？
28. MapReduce体系架构与处理流程
29. MapReduce如何在分布环境中处理词频统计、倒排索引
30. **实验部分**
31. Hadoop全分布部署模式
32. 各种配置文件core-site.xml、hdfs-site.xml、yarn-site.xml、zookeeper.xml、hbase-site.xml的主要作用、重点属性（少数几个，如配置hdfs端口的属性）。
33. Job、task与处理的数据之间的关系
34. MapReduce的主要接口
35. MapReduce实现倒排索引（原理及代码分析）
36. MapReduce移动计算、每个阶段的输入、输出。
37. HBase写入记录的过程（原理与分析代码）
38. 打包、运行实验程序(Java)的命令。
39. Hadoop进行开发搜索引擎的优势

**注：不要求写属性名、代码、安装、部署等相应的命令，给出属性、命令、代码能够识别与分析，例如，能够识别Mapper接口方法、该接口的作用及使用方法（例如，分析代码）。**