大数据系统开发考试内容

1. **大数据基本概念**
2. 大数据特征
3. 大数据面临的挑战
4. 关系数据库与NoSQL数据库的差异
5. 大数据库关键技术
6. 大数据处理流程
7. 大数据系统架构
8. **Hadoop**
9. 什么是Hadoop
10. Hadoop的核心架构
11. Hadoop生态圈
12. Hadoop的优点
13. Hadoop处理大数据的优势
14. **什么是HDFS**
15. HDFS的目标
16. HDFS体系结构
17. HDFS块副本
18. HDFS元数据
19. HDFS读写流程
20. HDFS访问方式
21. **什么是HBase**
22. HBase的特点
23. HBase处理大数据
24. HBase体系结构
25. HBase数据模型
26. 行式存储与列式存储的区别
27. HBase日志
28. HBase读写流程
29. **MapReduce**
30. MapReduce是什么？
31. MapReduce体系架构与处理流程
32. MapReduce分区
33. MapReduce如何在分布环境中处理词频统计、倒排索引
34. **实验部分**
35. Hadoop本地、伪分布与全分布部署模式
36. HDFS上传文件（原理与分析命令）
37. 配置文件及具体的使用方法
38. MapReduce的主要接口
39. MapReduce实现倒排索引（原理及代码分析）
40. MapReduce如何实现移动计算？
41. HBase写入记录的过程（原理与分析代码）
42. Hadoop进行开发搜索引擎的优势

**注：不要求写代码、安装、部署等相应的命令，给出代码、命令能够识别与分析，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）。**