1. Linux基础知识
   1. 文件操作
   2. 文件属性
   3. 文件编辑
2. 大数据时代概念
   1. 大数据基本概念
   2. 大数据特征
   3. 大数据产生方式的3个阶段
   4. 大数据处理的基本流程
   5. 大数据关键技术分类
   6. 大数据处理的两大模式
   7. 典型的大数据应用
3. 大数据采集
   1. 大数据采集定义
   2. 大数据采集方法分类
   3. Flume功能、架构图、配置文件
   4. HTTP协议，报文，get和post请求方法区别
   5. 网络爬虫的工作原理、工作流程
   6. Scrapy简介、选择器
4. 大数据处理系统
   1. 集中式计算、分布式计算
   2. Google大数据处理架构的三大组件
   3. 什么是HADOOP
5. 大数据文件系统HDFS
   1. 什么是HDFS
   2. HDFS是如何解决简单分布式存储系统的负载不均衡问题
   3. HDFS是如何提高数据可靠性的
   4. HDFS NameNode和DataNode的主要作用
   5. HDFS架构图
   6. HDFS常用命令
6. NOSQL数据库HBASE
   1. 什么是NoSQL
   2. NoSQL四大类型
   3. 使用实例描述HBase数据模型
   4. HBase常用命令（建表，插入数据，查询数据）
7. 数据仓库Hive
   1. 数据仓库定义
   2. Hive概述
   3. Hive架构
   4. Hive常用命令：建表，加载数据，查询
8. 大数据批处理 Hadoop MapReduce
   1. Map和Reduce函数定义、作用
   2. MapReduce架构图
   3. MapReduce工作流程步骤
   4. MapReduce处理过程wordcount
9. 大数据实时处理SPARK
   1. SPARK简介
   2. 总体架构
   3. RDD基本概念
   4. RDD基本操作：构建、transform、action
   5. RDD宽窄依赖
10. 大数据的功能应用场景
    1. 应用场景简单介绍