**（一）各章的重要知识点**

**第3章：**

HDFS中名称节点运行的基本原理以及只设置唯一的名称节点的局限性；数据节点的具体功能；HDFS常用的shell命令

**第4章：**

HBase各功能组件及其作用；HBase的三层结构中各层次的名称和作用；Hlog的工作原理；每个Region服务器维护一个Hlog 的优缺点；Hbase Shell命令的使用

**第5章：**

NoSQL四大类型数据库及每种数据库的典型应用和不适用情形；CAP的含义及满足CA、CP、AP的典型产品；BASE的基本含义； MongoDB数据库的Shell命令的使用；MongoDB的Java客户端编程；Redis哈希命令的使用

**第7章：**

JobTracker和TaskTracker的功能；MapReduce的工作流程；Map端和Reduce端的Shuffle过程

**第8章：**

HDFS HA的含义及架构组成；MapReduce 1.0体系结构中存在的问题；YARN架构中各组件的功能；Yarn和MapReduce 1.0框架的优劣对比分析

**第9章：**

Spark主要概念RDD、DAG、阶段、分区、窄依赖、宽依赖；Spark的运行架构； Spark运行的基本流程；RDD常用的转换操作和动作操作的含义和编程使用；常用的键值对转换操作及其用法；DataFrame的创建和保存方法；DataFrame的常用操作；DataFrame与RDD的相互转换；使用Spark SQL读写MySQL数据库的程序；编写Spark Streaming程序的基本步骤；掌握输入源为文件流和套接字流的流计算程序的编写方法；掌握把DStream写入MySQL数据库的程序的编写方法

**（二）实验与作业**

掌握实验1-4及作业1-3的内容