**附件1：课程安排（上课时间：上午9:00-12:00；下午14:00-17:00）**

|  |
| --- |
| **大数据** |
| **阶段1** | Hadoop基础实战 | **周期** | 7天 |
| **综合项目** | 1、基于Ubuntu的Hadoop集群安装、配置及管理 (高可用HA)2、Eclipse连接HDFS进行文件操作3、HBase分布式数据库的安装、配置及应用开发4、MapReduce编程模型实践（数据排序、数据去重、数据统计）5、分布式文件系统HDFS命令操作。6、Hive数据仓库应用实践 |
| **阶段目标** | 1. 掌握Hadoop原理机制，熟悉Hadoop集群体系结构、核心技术。
2. 掌握基本的Hadoop管理方式方法
3. 掌握HDFS API开发技能。
4. 掌握非关系数据库HBase的基本安装管理管理。
5. 掌握MapReduce并行计算模型的原理与实践。
6. 掌握Hive数据仓库工作原理及HiveQL应用
7. 掌握Spark生态系统及Spark Streaming、Spark MLlib的原理与实践
 |
| **日期** | **内容** | **时间** | **目标/知识点** | **实战演示** |
| **第一天** | Hadoop入门 | 9：00-12:00AM | 1、Hadoop概述2、Hadoop的功能与作用3、Hadoop应用现状及发展趋势4、Hadoop项目及其结构5、Hadoop体系结构6、Hadoop与分布式开发7、Hadoop计算模型 |  |
| 14:00-17:00PM | 8、HDFS的数据管理9、HBase的数据管理10、Hive的数据管理11、Hadoop集群安全策略12、本章小结 |
| **第二天** | Hadoop的安装与配置 | 9：00-12:00AM | 13、Hadoop平台搭建技术依赖14、JDK安装与配置15、tzseleck的功能介绍16、ntp时间同步服务器的描述17、SSH安全外壳协议18、SSH免密登陆验证19、Hadoop集群配置文件20、Hadoop集群配置参数21、Hadoop集群搭建帮助文档22、Hadoop分布式集群搭建的运行方式23、搭建Hadoop集群的基本流程 |  |
| 14:00-17:00PM | 24、集群网络拓扑25、定义集群拓扑26、基于Ubuntu环境分布式集群设计27、Ubuntu权限用户root登陆设置28、集群服务器网络设置29、集群服务器主机名设置30、集群服务器IP地址解析设置31、Ubuntu下JDK的安装32、Ubuntu下SSH免密登陆验证安装设置33、zseleck的时区时间设置34、集群ntp时间同步服务器的安装与设置35、Hadoop基本文件的配置36、Hadoop环境变量的配置37、Hadoop基本配置的分发38、Hadoop集群启动39、Hadoop集群测试40、Hadoopfs帮助命令41、Hadoopdfsadmin帮助命令42、Hadoop集群状态查看管理帮助命令43、本章小结 |
| **第三天** | **实验** | 9：00-12:00AM | 44、基于Ubuntu的Hadoop集群安装与配置 |
| Hbase分布式数据存储原理 | 14:00-17:00PM | 45、大数据数据存储的类型46、HBase概念47、HBase体系结构48、HBase数据存储原理49、HBase数据模型50、HBase与Hadoop关系51、HBase与HDFS关系52、Hregion概念与HRegion服务器53、HMaster概念与HMaster服务器54、ROOT表与METE表55、Hbase的表结构56、行、列、列族、时间戳57、逻辑视图与物理视图 |  |
| **第四天** | Hbase分布式数据存储实践 | 9：00-12:00AM | 58、Zookeeper概念59、Zookeeper在Hadoop中的工作原理60、Zookeeper与Hbase的关系61、Zookeeper的安装与配置62、Hbase的搭建与部署63、Hbase Shell简介64、Hbase Shell的基本用法65、create创建表命令66、list查询表命令67、describe描述表结构命令68、put增加记录命令69、get查询记录命令70、scan遍历表命令71、其他命令72、Hbase API简介73、HBaseAdmin、HBaseConfiguration类的使用74、HTable、HTableDescriptor、HColumnDescriptor类的使用75、Put、Get、Scanner类的使用 |
| 14:00-17:00PM | 76、Eclipse开发HBase编程实例(学生成绩管理) |
| **第五天** | MapReduce核心技术 | 9：00-12:00AM | 1. MapReduce编程模型介绍
2. MapReduce的执行过程
3. MapReduce在数据流和控制流
4. MapReduce编程接口体系结构
5. MapReduce API基本概念
6. 序列化和回调机制
7. Reporter参数
8. Mapper、Reducer、Partition接口的使用
9. InputFormat、OutputFormat接口的使用
10. MapReduce内置数据类型
11. MapReduce的开发流程
12. MapReduce开发环境配置
13. MapReduce任务调度
14. 数据预处理与InputSplit的大小
 |  |
| 实验 | 14:00-17:00PM | 91、案例实战：MapReduce编程模型实践（数据排序、数据去重、数据统计） |
| **第六天** | Hive数据仓库实践 | 9：00-12:00AM | 1. 数据仓库的概念、特点
2. Hive数据仓库的概述
3. Hive体系结构
4. Hive的基本组成
5. Hive与MapReduce的关系简述
6. Hive工作原理
7. Hive的元数据存储
8. Hive数据仓库的搭建与部署
9. Hive访问方式
10. Hive的数据类型
11. Hive表的创建、数据的加载
12. Hive表的修改与删除
13. 内部表与外部表
14. 表的分区与桶的建立
15. Hive SQL简单查询语句
16. Hive SQL高级查询语句
17. HiveQL视图
18. Hive索引使用
19. Hive的存储类型和复合数据类型
20. Hive函数的使用
 |  |
| 14:00-17:00PM | 1. 案例实战：Hive数据仓库基本操作、高级查询。
2. 案例实战：基于Hive的Hadoop日志分析
 |
| **第七天** | 大数据处理挖掘技术 | 9：00-12:00AM | 1. Spark及其生态系统
2. Spark简介与原理
3. Spark Core、Spark SQL介绍
4. Spark Streaming、MLib、GraphX
5. Spark编译与部署
6. Spark编程模型（一）编程模型及SparkShell实战
7. Spark编程模型（二）IDEA搭建及实战
8. Spark运行架构分析
 |
| 14:00-17:00PM | 1. SparkSQL（一）SparkSQL简介
2. SparkSQL（二）深入理解运行计划及调优
3. SparkSQL（三）Spark实战应用
4. Spark Streaming实时流计算原理与实践
5. Spark MLlib（一）机器学习库SparkMLlib介绍
6. Spark MLlib（二）机器学习库SparkMLib实战
7. Spark图计算GraphX
 |

**附件2：大数据师资培训班**

报 名 回 执 表

|  |  |  |  |
| --- | --- | --- | --- |
| 学 校 |  | 院 系 |  |
| 地 址 |  | 邮 编 |  |
| 电 话 |  | 传 真 |  |
| 报名人员信 息 | 姓 名 | 性别 | 手 机 | 职务/职位 | 邮 箱 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| 发票信息 | 发票抬头： 发票内容：培训费 （还可提供 会务费、会议费，可在”培训费”处修改）发票金额： 元 ；发票数量： 张 （1人以上交费的，可开一张总额发票，或单笔开数张发票；根据自身情况自由改写。） |
| 住 宿 | 入住日期： 月 日； 退房日期： 月 日 房间数量： 间；(标间 或 大床) |

为使报名程序顺利完成，请完整提供以下信息并发送至：xiaoyao.fu@qi-tech.com.cn

如需要培训结业证书请提供身份证号码：

|  |  |
| --- | --- |
| 姓名 | 身份证号码 |
|  |  |
|  |  |