

附件 1

课程内容及安排

1、网络学习课程

课程名称	课时	课程内容介绍
Linux 操作系统	10	Linux 操作系统介绍、文件与目录管理、用户以及用户组管理、磁盘管理、文本编辑工具 vim、文档的压缩与打包、软件安装、Shell 脚本等内容。
Hadoop 环境搭建及工作原理	10	讲解 Hadoop 生态系统，包括操作与开发；详细讲解 HDFS 和 Map-Reduce 的功能及作用；了解 MapReduce 原理、运行流程、压缩数据处理、作业调度、计算器等环节。
HDFS 程序开发实践	6	讲解 Hadoop 文件系统 HDFSJAVAAPI 的使用。如何使用 HDFSJavaAPI，读写文件、读写目录、以及对文件进行压缩处理等。
MapReduce 原理与开发实践	6	本课程针对 HadoopMapReduce 开发进行讲解。课程以案例为基础，以及如何使用 MapReduce 进行数据统计，去重，排序，Map 端 Join，Reduce 端 Join 等关联操作，掌握 MapReduce 处理过程。
Hive 数据仓库的应用	12	Hive 是基于 Hadoop 构建的数据仓库分析系统，通过学习掌握 Hive 的函数、Hive 数据的加载、Hive 的 DDL 操作、自定义函数 (UDF) 等内容，达到使用 Hive 进行查询、汇总、分析数据的能力。
分布式数据库 HBase 工作原理及应用	6	HBase 是 Hadoop 生态系统中的重要一员，主要用于海量结构化数据存储；通过学习对 HBase 表设计、表操作、数据操作、JavaAPI 等内容，掌握对 HBase 系统的开发及使用。
数据迁移工具 Sqoop 操作	4	Sqoop 是关系型数据库和 Hadoop 生态系统之间进行数据转换的主要工具；通过学习将 mysql 中的数据导入到 hdfs 中、将数据导入到 HBase 中、定义导入导出任务，掌握对数据迁移的能力。
分布式日志框架 Flume 配置与应用	4	Flume 对海量日志进行采集、聚合和传输的主流大数据工具；课程内容包含 Flume 应用场景、FlumeNG、FlumeOG、Flume 的核心组件、Flume 的架构、Flume 的 source、sink 配置说明等。

流式数据采集 Kafka 的部署与操作	4	Kafka 是分布式的消息队列，广泛应用于实时数据处理。学习内容包含 Kafka 的体系结构、安装模式及安装部署、Topic、Producer、Consumer、发布订阅消息以及 Kafka JAVA 开发等。
Spark 环境搭建及工作原理	6	Spark 是一款高性能的分布式计算框架，比 MapReduce 计算快百倍；本课程内容全面涵盖了 Spark 生态系统、Spark 与 Hadoop 对比、开发环境搭建、RDD、编程模型、Web 监控等内容。
SparkStreaming 原理与应用优化	4	SparkStreaming 是用户结合流式、批处理和交互式查询应用的实时计算框架；本课程内容详细讲解原理与特点、适用场景、容错、性能优化和内存优化等。
SparkSQL 的原理以及操作	4	SparkSQL 的出现，使得 SQL-on-Hadoop 的性能相对于 Hive 有了显著的提高。达到 Spark 兼容 Hive 的功能。本课程详细讲解特点、运行架构、数据源、数据缓存、DataFrame 等。

2、面授课程安排

数据科学与数据可视化课程（青海西宁）		
时间	课程模块	课程内容
第 1 天 8 月 14 日	数据科学核心理论与产品开发典型实践	<ul style="list-style-type: none"> 数据科学的核心理论（数据科学中的新理念、理论、方法、技术、工具、应用、机器学习以及如何跟踪数据科学的最新动态和如何继续学习数据科学） 数据科学的典型实践（最佳实践解读、典型算法及应用案例、基于 R 的数据科学实战、基于 SparkR 的数据科学实战） 数据产品开发（数据产品开发在数据科学中的重要地位；数据产品的特征；数据柔术；数据战略；数据能力；数据治理；开发技术；应用案例；如何继续学习
第 2 天 8 月 15 日		
第 3 天 8 月 16 日		
第 4 天 8 月 17 日	数据可视化	<ul style="list-style-type: none"> 数据可视化之美（数据与图形、数据可视化的运用、传统的数据分析图表、数据可视化工具） 数据可视化方法（整理数据源、直方图、圆饼图、散点图） 数据引导可视化设计（可视化对认知的帮助、掌握可视化设计组件、分析数据，指导视觉探索、空间数据的可视化） 数据可视化入门（Tableau、创建视图、仪
第 5 天 8 月 18 日		

		<p>表板与故事)</p> <ul style="list-style-type: none">• 数据管理与计算 (认识数据、数据架构、数据连接、数据加载、数据维护、高级数据操作)• 可视化分析 (条形图与直方图、饼图、压力图与突显表、树地图、气泡图与圆视图、标靶图、甘特图、地图分析)• 预测分析 (预测分析的可视化、预测、背景图像、趋势线)• 分享与发布 (导出和发布数据、导出图像文件、保存和发布)
--	--	--