课程方向：大数据技术与实战

课程目标与主要内容：

1. 系统讲解大数据技术的主要理论、方法、技术、工具及业内典型实际应用；

2. 掌握基于容器的大数据平台技术架构、大数据分析的基本理论、大数据分析挖掘应用实战技能；

3. 利用大数据搜索挖掘与可视化工具，提升综合应用能力；

4. 提供大数据技术与实战的全套教学资料；

5. 分享大数据技术课程建设的经验；

6. 考评合格者，可获得大数据技术与实战研修班结业证书。

李楠

清华大学博士、北京信息科技大学信息系统研究所常务副所长、信管学院系主任、微租联盟（北京）汽车科技有限公司前 CIO/ 合伙人

钱兴会

思享会，楚门智能创始人。澳大利亚昆士兰理工数据科学硕士,前阿里巴巴、联想资深数据产品工程师，架构师，数据产品专家

课程大纲（共计 48 学时，6 天）

第 1 天 | 第一节：大数据概论（4 学时）

- 大数据技术相关课程建设经验介绍
- 大数据基础逻辑
- 大数据典型应用
- 大数据框架发展前沿
- 基于容器的大数据环境部署

代码与案例实践

- Docker 环境搭建
- 各类大数据框架的 docker-compose.yml 详解

第二节：大数据存储与计算模式（4 学时）

- 分布式文件系统基本原理
- Hadoop 框架与 HDFS
- S3、OSS、Blob、数据湖等公有云存储
- 分布式数据库 Hbase、Nosql 与 Newsql
- 批处理、流处理
- 图计算与实时计算的数据可视化

代码与案例实践

- 尝试公有云环境下分布式文件存储与访问检索

第 2 天 | 第三节：Flink 大数据处理分析平台（4 学时）

- 流批一体化大数据流程管理 Kafka 与 Flume
- Flink 基本原理与结构
- Flink 与 MapReduce、Spark、Storm、Samza 详细比较
- Flink 程序与数据流
- Flink 分布式运行
- Flink 应用实例

代码与案例实践

- 相关技术文档

第四节：流批一体化大数据实战（4 学时）

- 基于 Docker 搭建 kafka 与 Flink 容器集群
- 使用 Flink 进行某电商大数据处理分析
- 体验 kafka 处理音乐流数据

代码与案例实践

- 全部实验的源代码与文档

第 3 天 | 第五节：大数据项目实战：互联网电商系统实践（8 课时）

- 推荐算法原理
- Spark 协同过滤算法程序示例
- Item-based 协同过滤与推荐
- User-based 协同过滤与推荐
- 个性化推荐案例
- 互联网电商推荐系统架构设计
  - 电商推荐系统设计
  - 电商推荐系统架构
  - 数据集介绍
- 互联网电商推荐系统 – 召回模型开发
  - 基于协同过滤模型的推荐系统召回模型
  - 基于电商用户画像系统的推荐系统召回模型
  - 基于规则与预测算法的推荐系统召回模型开发
- 互联网电商推荐系统 – 排序模型开发
  - 基于逻辑回归实现推荐排序
  - 基于 GBDT 实现推荐排序
  - 基于 GBDT+LR 模型实现推荐排序
- 推荐系统模型部署
- 推荐系统模型优化与分析

第 4 天 | 第六节：大数据项目实战：某电信企业微博舆情系统开发实践（8 课时）

- 企业舆情系统背景介绍
- 企业舆情系统架构设计
- 企业舆情系统价值
- 企业舆情分析系统架构设计
  - 存储层设计
  - 数据层设计
  - 大数据技术层设计
- 企业舆情分析分析系统 HDFS 存储模块开发与设计
  - 需求分析
  - 模块开发
  - 模块定义
- 企业舆情分析系统 MapReduce 模块开发与设计
  - 需求分析
  - 模块开发
  - 模块定义
- 企业舆情分析系统自然语言处理模块开发与设计
  - 新闻文本分类模型实现
  - 模块开发
  - 模块定义

第 5 天 | 第七节：大数据项目实战：某金融企业用户交易行为分析系统（8 课时）

- 金融行业大数据产品及规划介绍
  - 银行大数据产品及战略
  - 保险大数据产品及战略
- 金融大数据行业特点介绍
  - 金融大数据都在哪里
  - 金融大数据与互联网公司的不同
- 金融大数据行业采用技术介绍
  - Hadoop 生态技术
  - Spark 生态技术
- 用户交易行为分析系统 Spark ETL 编程案例与实现
  - 需求分析
  - a. 用户交易地区分析
  - b. 热门交易区域分析
  - c. 热门 IP 分析
  - 模块开发
  - 模块定义
  - 流程实现

第 6 天 | 第八节：大数据项目实战：某互联网企业用户流量分析系统（8 课时）

- 互联网行业大数据产品及规划介绍
  - 百度大数据产品及战略
  - 阿里大数据产品及战略
- 互联网大数据行业特点介绍
  - 互联网大数据都在哪里
  - 互联网大数据与金融等公司的不同
- 互联网大数据行业采用技术介绍
  - Hadoop 生态技术
  - Spark 生态技术
- 用户流量行为分析系统 Spark ETL 编程案例与实现
  - 需求分析
  - a. 用户地区分析
  - b. 热门区域分析
  - 模块开发
  - 模块定义
  - 流程实现
- 用户流量分析系统算法模块实现 Spark MLlib
  - 使用 Spark Mllib 预测网站流量趋势
  - 培训总结与教师探讨企业大数据内容如何融合高校教育