



你可能学了一个“假的”大数据

2018最新课程大纲

什么是大数据？

- 1、大数据是对海量数据存储、计算、统计、分析等一系列处理手段
- 2、处理的数据量是TB级，甚至是PB或EB级的数据，是传统数据处理手段无法完成的，例如关系型数据库oracle，mysql等，或者数据分析系统SAS、SPSS等。
- 3、大数据涉及分布式计算、高并发处理、高可用处理、集群、实时性计算等等，汇集的是IT最热门，最流行的IT技术
- 4、大数据是机器学习、深度学习、AI等尖端可以领域的基础架构，没有大数据，这些都无从谈起
- 5、大数据是企业财富的源动力，不是企业的成本，是企业的利润和财富，没有哪家企业会拒绝利润。

为什么要学习大数据？

- 1、大数据是最具有发展前景的IT技术
- 2、大数据行业人才急缺，尤其对技术过硬的大数据人才更是千金难求
- 3、大数据从业人员的平均薪资相较其他行业高出0.5~2倍左右
- 4、大数据行业发展势头迅猛，国家政府层面大力扶持，技术人员年薪百万者屡见不鲜
- 5、大数据行业上升空间广阔，分析师、架构师、算法工程师、AI工程师、机器学习工程师、深度学习工程师等等
- 6、大数据行业年龄有压倒性优势，年龄越大，身价越高，企业越青睐，其中年薪50万的就业记录就是由34岁的学员创造的，且年龄较长的学员月平均薪资都在25K左右。
- 7、马云在数博会上声称，未来30年是大数据时代，大数据和水电煤一样，是公司发展的必需品
- 8、很多IT人员都会面临一个35岁的尴尬年龄，届时年龄会成为硬伤，职业生涯遭遇天花板，跳槽困难。
- 9、党的十八届五中全会公报提出要实施“国家大数据战略”
- 10、习近平主席再次强调实施国家大数据战略，加快建设数字中国(万字长文解读)

大数据有哪些行业应用？

- 1、淘宝采用的是Hadoop集群，实现云梯模型。在云梯上，每天会对数据产品进行不同的MapReduce计算。
- 2、在存储层面上淘宝采用基于hbase集群的存储技术，满足高吞吐量下的实现随定定位与实时读写
- 3、亚马逊使用大数据对S3为数据元提供 99.999999999% 的耐久性和 99.99% 的可用性，并能够承受两个设施中的数据同时丢失。
- 4、亚马逊、淘宝、京东使用大数据技术为用户做商品推荐，其中亚马逊可获得15%的交易量增长
- 5、微信、搜狗等软件使用大数据做语音识别
- 6、中国银行、太保利用大数据做金融风控和防诈骗分析
- 7、迈宁数据、数尊等公司运用大数据做贷前贷后的风控处理
- 8、百分点公司运用大数据做精准营销
- 9、妙计旅行、滴滴和优步等公司运用大数据做线路出行优化
- 10、Facebook运用大数据做广告精准投放，09-10年实现市场份额翻番
- 11、英国BT运用大数据实现直邮回应率的100%增长
- 12、……

大数据是数据重要，还是技术重要，我们更注重哪一点？

- 1、数据重要，技术更重要。
- 2、没有数据，技术没有意义。
- 3、有数据，没有处理技术，数据就是死的，发挥不了价值，传统的处理手段也是技术，但是无法应对海量数据，数据自然也就发挥不了价值。
- 4、相对而言，数据较容易获取，采用埋点、定时等手段即可轻松获得海量数据，甚至可以从运营商或渠道商购买数据。
- 5、大数据如何优化存储，如何做数据结构优化选型，如何提升计算效率，如何避免内存泄漏，如何提升数据传输效率，计算资源如何配置，如何做可靠性保证，如何做准确性保证等问题都需要较高的技术能力。
- 6、不论是做风控、精准营销、推荐还是用户行为分析，都是大数据技术的具体应用，业务层面会有越来越多的产品和商业模型，自然会有专业的产品经理和运营团队来完成。
- 7、我们要做的是要把技术做到极致，上层的业务模型再复杂，进入底层之后，数据的处理方式的手段都会归并为以上所列出的问题，所以作为大数据工程师，掌握大数据核心技术，才是掌握真正的价值所在。

大数据未来的发展趋势如何？

- 1、越来越多的设备加入互联网，手环、手表、平衡车、电动汽车、车联网等等，物联网背景下，人们的行为数据会以指数方式进行增长。
- 2、在巨量数据的快速增长的情况下，会催生对大数据人才更大的需求量。
- 3、大数据技术在当下还是初级阶段，从我们刚刚完成对广州电信的企业培训了解到，公司正致力于大数据业务的开展，但没有相关技术团队，对大数据人才需求非常迫切。
- 4、机器学习继续成为大数据智能分析的核心技术。
- 5、人工智能和脑科学相结合，成为大数据分析领域的热点。
- 6、数据科学带动多学科融合。
- 9、大数据预测和决策支持仍然是应用的主要形式。
- 10、数据的语义化和知识化是数据价值的基础问题。
- 11、基于海量知识的智能是主流智能模式。
- 12、基于知识图谱的大数据应用成为热门应用场景等。

大数据很难，能学会吗？“我读书少，你不要骗我”

- 1、大数据相比传统IT技术，要难一些，但完全可以学得会！
- 2、难与不难完全取决于是否掌握了其中的原理，很多人觉得难，是根本没有掌握要领，完全靠记忆，那是不可能学会的，即使学会，也无法胜任工作要求。
- 3、大数据对讲师的要求很高，不论是技术体系的广度还是深度，都需要很深的功底。很多有着多年开发经验的人在无人指点的情况下，学习起来也是非常吃力的。
- 4、人一旦形成思维定势是可怕的，甚至有研究生问我们能不能学会大数据，这也可以看成是教育的悲哀了吧！
- 5、只有高中学历的学员毕业后也能拿到30K的月薪，甚至超过了很多高学历的人，根本原因是没有恐惧感，而是热衷于大数据技术。现在已经成成为了团队的骨干。

为什么说你可能学了一个“假”大数据？

- 1、页面设计高大上，什么？还有动图，简直不要美翻！……
- 2、项目描述狠牛掰，动辄上亿项目，又是某宝某东某车……
- 3、是不是眼花缭乱、真假难辨，办法很简单，那就是课程体系
- 4、如果你发现有大量的不相关课程，如下某著名机构的课程体系

阶段划分	主要内容	技术点
第一阶段： 电商基础架构	数据库	Mysql安装配置、SQL语法入门、数据导入导出
		数据建模、SQL高级查询
		JDBC入门、SQL注入、事务管理、批处理
	静态网站开发与部署	连接池 (Druid)、DBUtils框架
		HTML
		CSS
		JavaScript
		Bootstrap
		Linux常用操作
	动态网站开发与部署	Nginx服务器 网页配置
		Tomcat服务器、Servlet、Request、Response
		HTTP协议、Request、Response
		JSP入门、JSTL(forEach)、EL、MVC模式
		JQuery基础、选择器、对象、闭包调用
		JQuery DOM 操作、JQuery 事件、Ajax
JQuery Ajax、JSONP		
分页综合案例 (列表)、三层模式、DAO		
网站开发三大框架	Cookie、JQuery读取cookie、Session (登陆)	
	JavaMail注册、Redis入门、激活码	
	log4j、JUL、JCL、Slf4j、Tomcat日志分析	
	Filter过滤器	
	Maven	
	MyBatis	
	Spring	
	Spring MVC	

“你确定不是在逗我？”

- 5、还有java基础部分至关重要，在大数据技术面试中，能够占到50%的笔记，底层要求和一些原理性，尤其是多线程、数据结构、Hash原理、JVM内存结构、NIO、并发库等技术对大数据的掌握非常关键。下图是某著名机构放出的java部分体系，那么恭喜你中奖了！

基础班课程大纲			
所处阶段	主讲内容	技术要点	学习目标
第1阶段： Java语言入门	1.计算机基础知识	1.计算机基础；2.DOS常用命令；3.Java概述；4.JDK环境安装配置；5.环境变量配置；6.Java程序入门	可掌握的核心能力： (1) 能够掌握DOS系统常用基本命令；
	2.编程基础	1.常量与变量；2.数据类型；3.运算符；4.流程控制语句；5.方法；6.数组	(2) 熟练使用eclipse编写java代码； (3) 熟练使用java语言的常用对象；
	3.面向对象	1.面向对象思想；2.类与对象；3.成员变量和局部变量；4.封装；5.this关键字；6.构造方法	(4) 使用java编写单机应用程序； (5) 掌握面向对象编程思想，为以后
	4.常用类	1.Object类；2.Scanner类；3.Random类；4.String；5.StringBuilder类	深入学习JavaEE就业课程打下坚实的基础。
	5.集合	1.集合概述；2.集合特点；3.ArrayList集合	
	6.IO	1.字符输入流；2.字符输出流；3.字符缓冲输入流；4.字符缓冲输出流；5.复制文件；6.集合与文件中数据相互读写	可解决的现实问题： 学生具备Java基础，为JavaEE就业班的学习打下基础。 市场价值： 具备Java语言编程基础，但是还不能达到企业的用人标准。

这些东西真的可以吗？

6、学习大数据，想要高薪，单单java基础部分就不能是初学者水平，你要掌握的东西应该是下面这些：

阶段	主题	描述
Java基础部分	计算机基础	<ol style="list-style-type: none"> 1. windows操作系统简介 2. 标题栏、状态栏、任务栏、工具栏、命令行、桌面、控制面板 3. 应用程序、进程、进程管理器、CPU、内存、硬盘、键盘 4. windows快捷键操作
	Windows常用命令	<ol style="list-style-type: none"> 1. 目录操作：cd、dir、mkdir、rmdir、del 2. 文件操作：cp、 3. 终端操作：echo、clear 4. 网络命令：ping、ipconfig 5. 进程操作：netstat、taskkill、taskmgr
	文本编辑	<ol style="list-style-type: none"> 1. 选择控制：Ctrl+A、Ctrl+X、Ctrl+C、Ctrl+V 2. 保存控制：Ctrl+S、Ctrl+shift+S 3. 光标控制：Home、End、Shift+Home、Shift+End、PageDown、PageUp
	二进制原理	<ol style="list-style-type: none"> 1. 原码、补码、反码、权 2. 二进制，八进制，十进制，十六进制，进制互转、bit、byte 3. -128存储形式与推导式，-1存储形式与推导式、正负零问题详解
	JAVA基础	<ol style="list-style-type: none"> 1. 基本概念：JDK、JVM、JRE、JSE、JEE、JME 2. 基本数据类型：byte、short、int、long、float、double、boolean、char 3. 数据类型转换：数据类型内存结构、数据升格与降格、强制类型转换 隐式类型转换 4. 循环语句：for循环、while循环、空心三角、实心三角、百钱买白鸡 5. 异常处理：异常声明、异常抛出、运行时异常、异常短路 6. 运算符：算术、赋值、比较、逻辑、三元 7. 位运算符：&、 、~、^、<<、>>、>>>
	数组	<ol style="list-style-type: none"> 1. 数组维度：一维数组、二维数组、多维数组、不规则数组、梯形数组实现 2. 数组访问：数组遍历、越界问题、空指针问题、0维度数组意义 3. 数组提高：数组拷贝、数组转置、数组排序、数组求极值
	排序算法	<ol style="list-style-type: none"> 1. 排序算法：冒泡、选择、快速、堆、基数、Timsort 2. 复杂度：时间复杂度、空间复杂度
	OOP	<ol style="list-style-type: none"> 1. 封装，继承，多态 2. private、protected、public、default可见性 3. override、overload、overwrite 4. this关键字详解，super关键字详解，栈、堆的使用过程分析 5. 字段与方法覆盖问题巧妙记忆与分析 6. 非法关键字组合与原理剖析 7. Java接口、抽象类、内部类、匿名对象详解、 8. 低耦合与高内聚的设计法则剖析
Java IO	<ol style="list-style-type: none"> 1. 字节流、字符流、输入流、输出流 2. 文件与文件夹的复制、驱动器扫描实现 3. 装饰模式与过滤器流、缓冲区流、压缩流， 4. 文件的归档与解档、归档协议设计与实现、归档文件名抽取 5. 内存流、对象流、对象序列化、对象图的深度复制、对象的反序列化、暂态数据 6. 序列化版本ID的作用与剖析、串行数据导致对内存的激增处理 7. 转换流与字符数据丢失处理、乱码处理、中英韩日波斯等多字符集混编与写 8. java中字符全量输出与编程技巧 9. ascii码、utf8码、iso8859-1码、unicode码、gb2312码、gbk码 10. 平台编码分析处理、中文字符的判断手段与实现原理 11. 正则表达式的中文提取策略、国际化实现原理与案例精讲 	

Java基础部分	Java集合	<ol style="list-style-type: none"> 1、java集合框架解析，常见集合使用与特点总结，List，Set，Map 2、链表与数组列表的底层原理实现与百万数据量的读写评测分析 3、java中HashMap中算法详解、hashCode方法详解、equals方法详解、 4、hash码设计原理与案例分析，特征值散列处理、位运算与hash算法的组合应用 5、hashmap中的分桶机制、优化技术手段、 6、hashset的本质原理解析，hashset在极端情况下的数据构成与原理剖析
	Eclipse 开发环境	<ol style="list-style-type: none"> 1、主题设置，字体，色彩，快捷键， 2、工程的导入导出，首选项的导入导出，自定义代码模板， 3、项目结构详解与常见故障处理，maven继承、单元测试、常见故障排查
	IntelliJ IDEA开发环境	<ol style="list-style-type: none"> 1. IDEA的安装与配置 2. 首选项设置、兼容eclipse快捷键设置、快速代码模板定制 3. 省电模式开关、debug时线程断点环境处理 4. 集成maven开发环境 5. Maven错误与依赖项无法解析的处理手段与技巧总结 6. 项目、模块、工件关系详解 7. jar、war等各种归档文件导出、实战中第三方类库或依赖的导出控制
	多线程技术	<ol style="list-style-type: none"> 1. 线程基本概念、线程进程内在关系、线程的内存访问模型 2. 线程常见方法：yield、sleep、join、daemon、线程优先级设定 3. 线程安全问题：安全问题之售票问题、生产消费之熊吃蜂蜜问题、最快消费问题与和尚吃馒头案例精讲 4. notify()、notifyAll()、wait()、wait(int n)、sleep()方法异同详解 5. 死锁场景设计再现与极端条件处理、监视器对象与等待队列IO模型揭秘 6. 线程的状态变换与触发条件以及UML下状态机图建模展示
	Java设计模式	<ol style="list-style-type: none"> 1. 单例模式（饿汉式 懒汉式）、单例与线程安全性 2. 装饰模式、池化模式、构建器模式、链式编程 3. 工厂模式、静态工厂、非静态工厂、适配器模式 4. 代理模式原理剖析、代理模式的编程要点，代理模式与AOP编程实现 5. 池化模式+单例模式+装饰模式+适配器模式实现数据库连接池
	Java NIO	<ol style="list-style-type: none"> 1. NIO内部原理、缓冲区、通道 2. 缓冲区的常见操作与相关概念，clip，clear，capacity，limit，mark，position 3. 虚拟内存映射，磁盘文件的高速IO实现 4. 零拷贝技术原理解析，零拷贝与常规IO速率评测，多平台下零拷贝技术实现原理 5. selector模型详解、事件注册模型分解、注册集、挑选集、撤销集异同分析 6. NIO与IO模式下的性能对比与原理分析 7. 高并发框架netty底层原理实现与JavaNIO源码剖析
	Socket网络技术	<ol style="list-style-type: none"> 1. 计算机网络、IP地址、端口以及之间对应关系 2. 网络协议与OSI开发系统互联 3. 七层协议：物理层、数据链路层、网络层、传输层、会话层、表示层、传输层 4. UDP协议传输方式与应用场景介绍、无固定路由与消息强顺序保证机制 5. TCP协议、三次握手原理、通信协议设计精讲、消息层结构设计精讲 6. Java socket编程技术实践、netstat进行端口状态检查

Java基础部分	MySQL	<ol style="list-style-type: none"> 1. 关系型数据库原理、MySQL安装与配置 2. MySQL常用组件、数据库、表、触发器、存储过程DDL 3. MySQL下DML操作、CRUD数据 4. 数据库事务并发与脏读、幻读、不可重读并发风险介绍 5. 隔离级别精讲：read uncommitted、read committed、repeatable read、serializable各级别使用场景与并发现象的风险规避、隔离级别与性能评测 6. 锁原理精讲、共享读锁、独占写锁、表级锁、行级锁、悲观锁、乐观锁 7. MySQL下锁与索引的协同关系处理、MySQL百万数据量插入优化 8. 三种手段实现数据的读保护处理、读操作写锁定控制、插入操作串行控制保证 9. MySQL常见问题解决、BLOB字段缓存区调整、字符集乱码处理 10. MySQL数据彻底完全卸载处理、性能评测中高精度执行时间查看技巧 11. MySQL常见故障排查、远端拒绝连接、root密码重置
	JDBC	<ol style="list-style-type: none"> 1. JDBC API基本使用，完成insert、update、delete、select处理 2. statement、PreparedStatement、CallableStatement用法比较 3. SQL注入安全演示，百万数据量插入优化处理 4. 事务API编程控制、commit、rollback、savepoint控制 5. 并发状态下API的事务隔离级别控制与原理剖析 6. 第三方数据源配置与使用，c3p0，DBCP等等 7. 数据源容量控制、生产消费模式处理、资源回收控制
	Java反射技术	<ol style="list-style-type: none"> 1. java反射的技术原理与核心类Class，Method，Field，Constructor 2. 反射实现对象属性拷贝，基本数据类型与包装类判定 3. 类成员修饰符设计原理与代码解析，修饰符总和构成算法与判断技巧 4. 数组类型、多维数组类型、私有构造、静态方法在反射的使用技巧
	java内省	<ol style="list-style-type: none"> 1. Javabean的概念阐述、POJO、实体概念详解 2. 内省原理与实现、使用场景阐述、属性、字段、成员变量以及异同点比较 3. Introspector、BeanInfo、PropertyDescriptor用法展示 4. 编码实现异类对象之间的属性拷贝，继承属性复制处理 5. 属性复制工具类增强基本类与包装之间互通处理
	Java并发库	<ol style="list-style-type: none"> 1. 线程池原理精讲、swap锁、原子读写数据类 2. 可并发HashMap集合、双头队列、可阻塞队列 3. Executors类源码解析、FutureTask任务的跟踪处理、异步任务的同步执行控制 4. 线程池提交task的UML原理图剖析、源码分析精讲 5. 线程池元数据的一元数字化控制设计原理 6. 固定线程池、单线程守护线程池编程案例
	JVM	<ol style="list-style-type: none"> 1. JVM内存结构详解、方法区、java栈区、本地方法栈、程序计数器、堆区 2. 堆内存结构精讲、堆内内存、离堆内存、非堆内存的控制手段 3. 堆内内存结构精讲、年轻代，年老代、伊甸区、幸存区（一区，二区） 4. 内存碎片的整理原理、栈区空间调整、栈空间溢出、堆空间溢出 5. JVM调优参数控制、-Xmx -Xms -Xss -Xmn -XX:NewSize -XX:MaxNewSize -XX:NewRatio -XX:SurvivorRatio 6. Jvisualvm、Jconsole、Jmap等虚拟机监控工具的使用 7. GC垃圾回收原理，垃圾回收控制分析，堆内垃圾回收，major回收，minor回 8. 离堆内存垃圾回收处理手段、jdk保护shell穿透访问 9. Java三级ClassLoader原理剖析、Bootstrap ClassLoader、ExtClassLoader、AppClassLoader、自定义类加载实现代码热部署

Java基础部分	红蜘蛛电子课堂	<ol style="list-style-type: none"> 1. 系统概述、整体架构分析与设计、 2. UDP报文协议分析与设计、协议头设计、消息的顺序性保证设计思想 3. 教师端设计与实现：机器人的抓屏处理、报文的数据帧切割处理、数据的压缩处理、UDP数据报发送处理 4. 学生端设计与实现：UDP数据报接收处理、报文组装处理、解压缩处理、图像反串行化处理、画面帧单元数据丢失处理、图像的展示处理 5. Java程序的产品化处理、属性外部化配置处理、归档化处理、执行脚本编写、常见故障排查与处理
	QQ即时通信	<ol style="list-style-type: none"> 1. 系统概述、整体架构分析与设计 2. 消息报文协议设计与实现、服务器消息考量与处理，客户端消息考虑与处理 3. 基础类库设计与实现、工厂模式在消息解析时的经典运用 4. 消息模型的设计与思路、java基本类型的流中的串行与反串行处理 5. 服务器端NIO+线程池实现高并发消息处理 6. 客户端消息类型解析、消息报文的组装 7. 即时通信实现群聊、私聊、好友列表的手动刷新和自动刷新 8. 实现文件及文件夹传输 9. 使用UDP协议实现共享桌面
	迅雷下载软件	<ol style="list-style-type: none"> 1. 系统概述、整体架构分析与设计 2. 实现多线程资源下载、UI交互中各线程下载状况跟踪显式 3. 实现资源的断点续传处理 4. 实现资源的限速处理、暂停控制

我要怎么相信你们？

- 1、你不要相信我们，因为所有机构都会极尽所能的吹嘘自己是如何的牛掰，我们也一样！
- 2、记住，没有对比就没有伤害！你只需要拿到我们的课程视频和其他的视频对比一下就可以了，看一下是否适合自己。
- 3、什么？不知道如何判断好坏，那就找你周围懂点技术的人帮你参考下，千万不要找我们的咨询人员，他们都有业务压力，肯定都会说我们的好！那样对你太不负责了！
- 4、如果你连周围帮你参考的人都找不到，给你个建议，一是抛硬币决定，二是放弃学习，我们推荐第二种办法！

师资团队？

1、徐培成

首席大数据开发技术CTO亲自授课，曾任驿统金（上海）网络技术有限公司总经理，主导与光大银行总行重点合作的阳光驿站金融结算系统的研发，并指导团队在阿里公有云上规划、部署Hadoop大数据集群，以对用户的金融交易日志进行存储和分析，完成从传统RDMBS数据到大数据集群的移植和升级。



2、王建军

毕业于计算机专业，有多年开发和管理经验，团队骨干人员，曾参与中国银行总行股票基金交易大数据集群规划与建设，负责基金账户日常数据的维护和管理，参与股票投资客户群的用户投资行为数据ETL工作，对股票、基金以及期货交易的投资回报进行数据仓库分析，有丰富的金融行业经验，擅长hadoop运维管理经验，对常见故障排查有丰富的实践经验，熟练使用hive数据仓库进行业务分析与处理。



3、常庆琳

毕业于南开大学软件工程专业，先后在软件公司任软件工程师、系统分析师、项目经理，主持开发WIFI广告交易大数据拟合分析平台，互联网数据服务平台，擅长java高并发、高负载、高吞吐量下技术实现，对数据库技术、互联网技术、分布式计算集群等技术有着丰富的实战经验和深厚的技术功底。曾就职于北京亚信公司，担任PM，整体负责面向广告商和Portal商户提供广告投放及管理的综合业务平台，对机器学习、算法有者丰富项目和实战经验。



项目实战

项目一：基于人工智能和自然语言处理技术的自动客服系统

项目介绍、业务模型与行业术语阐述

核心算法库常用算法原理与使用场景分析、gensim、nltk，常用web框架flask使用说明，常见主题模型算法原理与使用技巧、SVD、LDA、LSI，基于gensim算法模型建立流程实现，分词算法原理与应用场景说明，推荐系统优化处理与效果展示。

项目二：南方省级电信公司千万级通话数据迁移改造

该项目包括App内嵌SDK核心模块、nginx反向代理与nginx web服务器集群构建模块、实现post静态资源处理以及json数据格式转换的nginx lua模块、lume日志收集模块、kafka消息存储模块、hive实现数据清洗和统计分析模块、azkaban调度模块、spark streaming实时计算模块、redis数据结果存储模块、基于SSM+echart实现数据可视化等众多模块。

该项目为东南某省电信公司的千万级通话数据的改造工程，主要完成从传统基于oracle的关系型存储方案改造成基于hbase数据库的分布式存储，利用hbase可伸缩性、可扩展、高吞吐量以及高可用的特点解决传统数据面临的查询性能低、扩展性差、伸缩性不强等瓶颈。

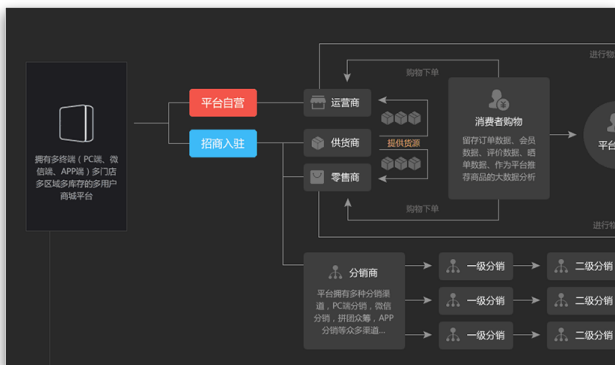
项目三：基于协同过滤的垂直电商实时推荐系统

项目介绍、业务模型与行业术语阐述、功能模块演示与讲解

评测模块中相关概念阐述与算法实现，算法与核心库实现、准确率、召回率、覆盖度、多样性、信息熵、基尼系数，实时推荐引擎模块数据结构设计与实现、产品选型与执行标准，用户行为数据清洗处理、行为类型分析与权重设计、推荐策略的设计与实现，推荐引擎中相似度算法实现、jaccard算法、余弦算法、john breese算法、基于商品协同过滤算法原理与实现，动态化调整推荐模型的设计与实现，推荐算法的设计与实现、基于邻域模型的协同过滤核心代码实现，基于隐语义模型算法原理原理与实现，实时计算模式下推荐模型的实时更新处理、周期性推荐模型校准处理，推荐策略中热门推荐与协同过滤的组合应用、用户与商品的冷启动处理。

选修项目

大型互联网电商数据项目



大型企业移动bug处理系统



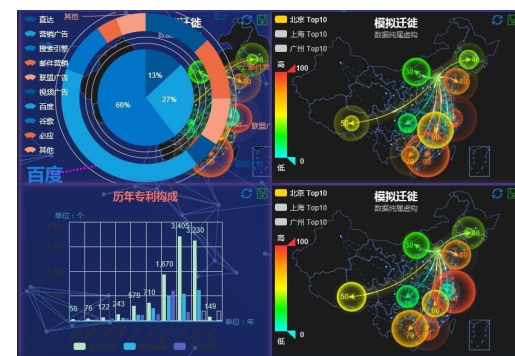
传统广告大数据分析项目（新增）



数亿级网站日志收集分析项目



大并发网站流量统计项目



部分学员就业薪资

姓名	工作单位	月薪	地点
于**	公司保密	税后18k	北京
廖*	用友软件	16k*16	北京
张*	上海新致软件有限公司	20k	北京
张*	公司保密	20k	北京
安**	亚信科技	15k	北京
范**	搜房网	15k	北京
张*	公司保密	20k	北京
常**	拓明科技	18k	北京
刘*	小牛学堂	30k	北京
李*	神州泰岳股份有限公司	10k	北京
贺**	上海圆通速递有限公司	19k	上海
雷*	北京智慧云行科技有限责任公司	22k	北京
李**	天曦网络科技(北京)有限公司	12k	北京
刘**	北京合力亿捷科技股份有限公司	14k	北京
陆*	江苏西格数据科技有限公司	联合创始人	北京
陶*	同鑫诚程科技	14.5k	北京
张**	浙江核新同花顺网络信息股份有限公司	14.5k	杭州
蔡**	保密	35k	北京
张*	北京惠福康信息咨询有限公司	13K	北京
张*	居理新房	19K	北京
唐**	广州，公司保密	13K	广州
史**	云和时空	18k	北京
张*	音悦台	15k	北京
梁*	保密	20K	北京

课程大纲

阶段	主题	描述
Java基础部分	计算机基础	<ol style="list-style-type: none"> 1. windows操作系统简介 2. 标题栏、状态栏、任务栏、工具栏、命令行、桌面、控制面板 3. 应用程序、进程、进程管理器、CPU、内存、硬盘、键盘 4. windows快捷键操作
	Windows常用命令	<ol style="list-style-type: none"> 1. 目录操作: cd、dir、mkdir、rmdir、del 2. 文件操作: cp、 3. 终端操作: echo、clear 4. 网络命令: ping、ipconfig 5. 进程操作: netstat、taskkill、taskmgr
	文本编辑	<ol style="list-style-type: none"> 1. 选择控制: Ctrl+A、Ctrl+X、Ctrl+C、Ctrl+V 2. 保存控制: Ctrl+S、Ctrl+shift+S 3. 光标控制: Home、End、Shift+Home、Shift+End、PageDown、PageUp
	二进制原理	<ol style="list-style-type: none"> 1. 原码、补码、反码、权 2. 二进制, 八进制, 十进制, 十六进制, 进制互转、bit、byte 3. -128存储形式与推导式, -1存储形式与推导式、正负零问题详解
	JAVA基础	<ol style="list-style-type: none"> 1. 基本概念 : JDK、JVM、JRE、JSE、JEE、JME 2. 基本数据类型: byte, short, int, long, float, double, boolean, char 3. 数据类型转换: 数据类型内存结构、数据升格与降格、强制类型转换 隐式类型转换 4. 循环语句 : for循环、while循环、空心三角、实心三角、百钱买白鸡 5. 异常处理 : 异常声明、异常抛出、运行时异常、异常短路 6. 运算符 : 算术、赋值、比较、逻辑、三元 7. 位运算符 : &、 、~、^、<<、>>、>>>
	数组	<ol style="list-style-type: none"> 1. 数组维度: 一维数组、二维数组、多维数组、不规则数组、梯形数组实现 2. 数组访问: 数组遍历、越界问题、空指针问题、0维度数组意义 3. 数组提高: 数组拷贝、数组转置、数组排序、数组求极值
	排序算法	<ol style="list-style-type: none"> 1. 排序算法: 冒泡、选择、快速、堆、基数、Timsort 2. 复杂度 : 时间复杂度、空间复杂度
	OOP	<ol style="list-style-type: none"> 1. 封装, 继承, 多态 2. private, protected, public, default可见性 3. override, overload, overwrite 4. this关键字详解, super关键字详解, 栈、堆的使用过程分析 5. 字段与方法覆盖问题巧妙记忆与分析 6. 非法关键字组合与原理剖析 7. Java接口、抽象类、内部类、匿名对象详解、 8. 低耦合与高内聚的设计法则剖析

Java基础部分	Java IO	<ol style="list-style-type: none"> 1. 字节流、字符流、输入流、输出流 2. 文件与文件夹的复制、驱动器扫描实现 3. 装饰模式与过滤器流、缓冲区流、压缩流， 4. 文件的归档与解档、归档协议设计与实现、归档文件名抽取 5. 内存流、对象流、对象串行化、对象图的深度复制、对象的反串行化处理、暂态数据 6. 串行化版本ID的作用与剖析、串行数据导致对内存的激增处理 7. 转换流与字符数据丢失处理、乱码处理、中英韩日波斯等多字符集混编读与写 8. java中字符全量输出与编程技巧 9. ascii码、utf8码、iso8859-1码、unicode码、gb2312码、gbk码 10. 平台编码分析处理、中文字符的判断手段与实现原理 11. 正则表达式的中文提取策略、国际化实现原理与案例精讲
	Java集合	<ol style="list-style-type: none"> 1. java集合框架解析，常见集合使用与特点总结，List，Set，Map 2. 链表与数组列表的底层原理实现与百万数据量的读写评测分析 3. java中HashMap中算法详解、hashCode方法详解、equals方法详解 4. hash码设计原理与案例分析，特征值散列处理、位运算与hash算法组合应用 5. hashmap中的分桶机制、优化技术手段 6. hashset的内部原理解析，hashset在极端情况下的数据构成与原理剖析
	Eclipse	<ol style="list-style-type: none"> 1. 主题设置、字体、色彩、快捷键 2. 工程的导入导出、首选项的导入导出、自定义代码模板 3. 项目结构详解与常见故障处理、maven继承、单元测试、常见故障排查
	Intellij IDEA	<ol style="list-style-type: none"> 1. IDEA的安装与配置 2. 首选项设置、兼容eclipse快捷键设置、快速代码模板定制 3. 省电模式开关、debug时线程断点环境处理 4. 集成maven开发环境 5. Maven错误与依赖项无法解析的处理手段与技巧总结 6. 项目、模块、工件关系详解 7. jar、war等各种归档文件导出、实战中第三方类库或依赖的导出控制
	多线程技术	<ol style="list-style-type: none"> 1. 线程基本概念、线程进程内在关系、线程的内存访问模型 2. 线程常见方法：yield、sleep、join、daemon、线程优先级设定 3. 线程安全问题：安全问题之售票问题、生产消费之熊吃蜂蜜问题、最快消费问题与和尚吃馒头案例精讲 4. notify()、notifyAll()、wait()、wait(int n)、sleep()方法异同详解 5. 死锁场景设计再现与极端条件处理、监视器对象与等待队列IO模型揭秘 6. 线程的状态变换与触发条件以及UML下状态机图建模展示
	Java设计模式	<ol style="list-style-type: none"> 1. 单例模式（饿汉式 懒汉式）、单例与线程安全性 2. 装饰模式、池化模式、构建器模式、链式编程 3. 工厂模式、静态工厂、非静态工厂、适配器模式 4. 代理模式原理剖析、代理模式的编程要点，代理模式与AOP编程实现 5. 池化模式+单例模式+装饰模式+适配器模式实现数据库连接池
	Java NIO	<ol style="list-style-type: none"> 1. NIO内部原理、缓冲区、通道 2. 缓冲区的常见操作与相关概念，clip，clear，capacity，limit，mark，position 3. 虚拟内存映射，磁盘文件的高速IO实现 4. 零拷贝技术原理解析，零拷贝与常规IO速率评测，多平台下零拷贝技术实现原理 5. selector模型详解、事件注册模型分解、注册集、挑选集、撤销集异同分析 6. NIO与IO模式下的性能对比与原理分析 7. 高并发框架netty底层原理实现与JavaNIO源码剖析

Java基础部分	Socket网络技术	<ol style="list-style-type: none"> 1. 计算机网络、IP地址、端口以及之间对应关系 2. 网络协议与OSI开发系统互联 3. 七层协议：物理层、数据链路层、网络层、传输层、会话层、表示层、传输层 4. UDP协议传输方式与应用场景介绍、无固定路由与消息顺序保证机制 5. TCP协议、三次握手原理、通信协议设计精讲、消息层结构设计精讲 6. Java socket编程技术实践、netstat进行端口状态检查
	MySQL	<ol style="list-style-type: none"> 1. 关系型数据库原理、MySQL安装与配置 2. MySQL常用组件、数据库、表、触发器、存储过程DDL 3. MySQL下DML操作、CRUD数据 4. 数据库事务并发与脏读、幻读、不可重读并发风险介绍 5. 隔离级别精讲：read uncommitted、read committed、repeatable read、serializable各级别使用场景与并发现象的风险规避、隔离级别与性能评测 6. 锁原理精讲、共享读锁、独占写锁、表级锁、行级锁、悲观锁、乐观锁 7. MySQL下锁与索引的协同关系处理、MySQL百万数据量插入优化 8. 三种手段实现数据的读保护处理、读操作写锁定控制、插入操作串行控制保证 9. MySQL常见问题解决、BLOB字段缓存区调整、字符集乱码处理 10. MySQL数据彻底完全卸载处理、性能评测中高精度执行时间查看技巧 11. MySQL常见故障排查、远端拒绝连接、root密码重置
	JDBC	<ol style="list-style-type: none"> 1. JDBC API基本使用，完成insert、update、delete、select处理 2. statement、PreparedStatement、CallableStatement用法比较 3. SQL注入安全演示，百万数据量插入优化处理 4. 事务API编程控制、commit、rollback、savepoint控制 5. 并发状态下API的事务隔离级别控制与原理剖析 6. 第三方数据源配置与使用，c3p0、DBCP等等 7. 数据源容量控制、生产消费模式处理、资源回收控制
	Java反射技术	<ol style="list-style-type: none"> 1. java反射的技术原理与核心类Class, Method, Field, Constructor 2. 反射实现对象属性拷贝，基本数据类型与包装类判定 3. 类成员修饰符设计原理与代码解析，修饰符总和构成算法与判断技巧 4. 数组类型、多维数组类型、私有构造、静态方法在反射的使用技巧
	java内省	<ol style="list-style-type: none"> 1. Javabean的概念阐述、POJO、实体概念详解 2. 内省原理与实现、使用场景阐述、属性、字段、成员变量以及异同点比较 3. Introspector、BeanInfo、PropertyDescriptor用法展示 4. 编码实现异类对象之间的属性拷贝，继承属性复制处理 5. 属性复制工具类增强基本类与包装之间互通处理
	Java并发库	<ol style="list-style-type: none"> 1. 线程池原理精讲、swap锁、原子读写数据类 2. 可并发HashMap集合、双头队列、可阻塞队列 3. Executors类源码解析、FutureTask任务的跟踪处理、异步任务的同步执行控制 4. 线程池提交task的UML原理图剖析、源码分析精讲 5. 线程池元数据的一元数字化控制设计原理 6. 固定线程池、单线程守护线程池编程案例
	JVM	<ol style="list-style-type: none"> 1. JVM内存结构详解、方法区、java栈区、本地方法栈、程序计数器、堆区 2. 堆内存结构精讲、堆内内存、离堆内存、非堆内存的控制手段 3. 堆内内存结构精讲、年轻代、年老代、伊甸区、幸存者（一区，二区） 4. 内存碎片的整理原理、栈区空间调整、栈空间溢出、堆空间溢出 5. JVM调优参数控制、-Xmx -Xms -Xss -Xmn -XX:NewSize -XX:MaxNewSize -XX:NewRatio -XX:SurvivorRatio 6. Jvisualvm、Jconsole、Jmap等虚拟机监控工具的使用 7. GC垃圾回收原理，垃圾回收控制分析，堆内垃圾回收，major回收，minor回 8. 离堆内存垃圾回收处理手段、jck保护shell穿透访问 9. Java三级ClassLoader原理剖析、Bootstrap ClassLoader、ExtClassLoader、AppClassLoader、自定义类加载实现代码热部署

Java基础部分	红蜘蛛电子课堂	<ol style="list-style-type: none"> 1. 系统概述、整体架构分析与设计、 2. UDP报文协议分析与设计、协议头设计、消息的顺序性保证设计思想 3. 教师端设计与实现：机器人的抓屏处理、报文的数据帧切割处理、数据的压缩处理、UDP数据报发送处理 4. 学生端设计与实现：UDP数据报接收处理、报文组装处理、解压缩处理、图像反串行化处理、画面帧单元数据丢失处理、图像的展示处理 5. Java程序的产品化处理、属性外部化配置处理、归档化处理、执行脚本编写、常见故障排查与处理
	QQ即时通信	<ol style="list-style-type: none"> 1. 系统概述、整体架构分析与设计 2. 消息报文协议设计与实现、服务器消息考量与处理，客户端消息考虑与处理 3. 基础类库设计与实现、工厂模式在消息解析时的经典运用 4. 消息模型的设计与思路、java基本类型的流中的串行与反串行处理 5. 服务器端NIO+线程池实现高并发消息处理 6. 客户端消息类型解析、消息报文的组装 7. 即时通信实现群聊、私聊、好友列表的手动刷新和自动刷新 8. 实现文件及文件夹传输 9. 使用UDP协议实现共享桌面
	迅雷下载软件	<ol style="list-style-type: none"> 1. 系统概述、整体架构分析与设计 2. 实现多线程资源下载、UI交互中各线程下载状况跟踪显式 3. 实现资源的断点续传处理 4. 实现资源的限速处理、暂停控制
Linux	VMware Workstation	<ol style="list-style-type: none"> 1、MVware软件安装与配置，网络连接模式，NAT、桥接、OnlyHost 2、宿主机、虚拟机、客户机异同，虚拟机完整克隆、虚拟机链接克隆 3、mac地址修改，虚拟网卡设置，虚拟网络编辑器，虚拟机移除
	CentOS	<ol style="list-style-type: none"> 1、Linux系统简介，CentOS, Ubuntu, Windows 2、CentOS下载、安装与配置
	目录结构	<ol style="list-style-type: none"> 1、目录结构：/bin /sbin /usr/bin /usr/sbin /usr/local/bin/usr/local/sbin /home /dev /tmp /boot /etc /mnt 2、文件类型：d：目录，-：文件，l：软连接，b：块设备 3、权限范围：User, group, other 4、权限类型：read, writer, execute, none 5、权限命令：chown, chmod u+x, chmod +x, chmod 777, chmod ugo+x 6、用户命令：adduser, useradd, userdel
	Linux命令	<ol style="list-style-type: none"> 1、目录操纵：ls, cd, rm, pwd, whoami, mkdir, mv, find, grep, , dirname, basename, 2、文件操纵：touch, cat, head, tail, cp, more, file, tar, gzip, gunzip, ln 3、终端操纵：clear, whoami, echo, redirect, help, reboot, cut, 4、网络命令：ifconfig, ping, hostname, firewalld服务, scp, rsync 5、进程控制：kill, nestat, jobs, ps, top, 6、管理控制：sudo password, uname, export, source, service, chkconfig 7、文本命令：nano, vi, wc, sed, awk 8、设备管理：mount, umount 9、帮助命令：help, info, man, 10、磁盘管理：fdisk, df, dd, 11、shell编程：if命令, for循环、while循环 12、控制命令：&&, , ;, (;), \$?, \$#,\$0,\$@, shift 13、瑞士军刀：nc, 监听端口, client连接, 聊天实现, 文件传输 14、日期控制：date, 时间滚动, 时间格式化, 时钟同步, 15、任务调度：crontab 16、软件安装：yum, make, make install, 17、提示符配置：export PS1='[\u@\h`pwd']\\$' 18、SSH配置：ssh-keygen, ssh-copy-id

Hadoop	Hadoop介绍	<ol style="list-style-type: none"> 1、Hadoop发展史与介绍, 2、模块说明, common, hdfs, mapreduce, yarn 3、Hadoop在国内国际上使用状况
	Hadoop集群	<ol style="list-style-type: none"> 1、SSH安装配置, 集群规划, 虚拟机环境准备 2、JDK安装与配置, 环境变量设置, JDK运行环境校验 3、Hadoop集群搭建, 本地模式, 伪分布式, 完全分布式 4、Hadoop临时目录修改
	Hadoop集群管理	<ol style="list-style-type: none"> 1. 集群管理命令: 启动, 停止, start-all.sh, start-dfs.sh, stop-all.sh, stop-dfs.sh 2. Hadoop架构原理详解、HDFS架构与进程集详解、YARN架构与进程集详解、NN、DN、2NN、RM、NM进程作用详解 3. 管理脚本分析、核心脚本源代码剖析、5大子进程的单独启动与停止 4. webui查看, 常用端口访问, 50070, 50090, 50075, 8088 5. Hadoop secondaryNameNode的周期性融合实现原理 6. Hadoop 2NN节点独立配置, 迁移元数据
	hdfs	<ol style="list-style-type: none"> 1、常用hdfs操作: 上传、下载、查看、移动、删除、创建目录 2、文件块大小设置、副本数设置、块物理切割与分布、物理块文件合成演示 3、oiv镜像文件考察, oev编辑日志考察, 元数据分析, 名字空间保存, 4、日志滚动, 配额管理, 快照管理及增量实现原理, 安全模式控制, 回收站管理, 5、黑名单和白名单, 节点的服役与退役, 6、HDFS API编程实现文件读写, HDFS文件写过程剖析, DFSPacket, truncck, 校验和处理 7、HDFS串行化writable原理详解, SequenceFile, MapFile文件, 同步点与可切割处理 8、HDFS压缩编解码器, default, deflate, gzip, bzip2, snappy, lzo, lz4以及性能评测
	MapReduce	<ol style="list-style-type: none"> 1. MapReduce编程模型原理, MR实现wordcount、Combiner过程控制 2. job本地模式运行, job导出与集群下执行 3. MR Job提交流程源代码跟踪解读, UML时序图分析 4. MRJob实现历年最高气温统计, combiner优化处理, 5. Map切片原理详解、reduce个数设置、数据倾斜处理手段 6. MR排序实现、部分排序、全排序、全排序的采样原理分析与常见错误处理、二次排序、自定义分区、自定义排序对比器、自定义分组对比器 7. MR的连接操作、map端连接、reduce连接异同与使用场景阐述 8. MR的多种inputFormat处理, 多种OutputFormat处理, MR与数据库交互 9. MR的Map的链式处理、reduce后的map链式处理, 多输入格式控制 10. 数据倾斜的处理手段、key的重新设计、自定义分区、二次job、多种手段的优缺点对比与优先使用场景 11. 使用MR实现团购网站App商家用户评论数统计双倒排实现。 12. 传统关系型数据库的MR集成
	YARN	<ol style="list-style-type: none"> 1. YARN资源调度框架, 事件二次分发架构原理 2. MR作业在yarn资源执行原理详解, Job, 切片原信息, 资源管理器 3. AppMaster, NodeManager, Yarn容器, Yarn子进程
	Hadoop高级管理	<ol style="list-style-type: none"> 1. Hadoop多模式共存, 编写脚本实现模式一键切换 2. Hadoop集群的HA架构原理与配置, NN的HA模式下容灾演练 3. Hadoop HA架构进程协同工作原理, NN, DN, JN, ZKFC, RM, NM 4. Hadoop Namenode联邦配置, 实现名称节点负载均衡 5. ResourceManger的HA配置与容灾演练 6. Hadoop NN的HA模式下整合Federation终极解决NN节点的负载与高可用问题

Hive	hive安装	<ol style="list-style-type: none"> 1、数据仓库，数据库，OLAT，OLTP概念阐述 2、hive简介、hive安装与配置 3、hive架构分析，hive元数据mysql配置
	hive应用	<ol style="list-style-type: none"> 1、hive元数据初始化，库，表，内部表，外部表，分区表，桶表，事务表 2、hive数据类型，简单类型，复杂类型，array，map，struct 3、hive表生成函数实现wordcount 4、hiveserver2+beeline以及jdbc客户端远程访问hive数据仓库 5、hive优化，map端连接，左半连接，本地模式，列格式转储，orc，rcfile，parquet，hive连接倾斜配置 6、hive sort by，order by，cluster by，distributed by
	hive高级	<ol style="list-style-type: none"> 1. hive高级聚合函数、grouping set、rollup、cube等运用技巧与优势分析 2. hive强大的分析函数、rank、percent rank、ntile、firstvalue、lastvalue、denserank 3. hive的开窗函数、行开窗、数据范围开窗、复杂查询下的窗口重用 4. hive自定义函数，UDF，UDAF，UDTF，分别实现周月星期起始时刻，日志聚合表的又分处理 5. hive执行计划分析与解释、hive优化手段与使用场景
Avro与Protobuf	Avro与Protobuf	<ol style="list-style-type: none"> 1. Java串行化与优缺点分析 2. avro简介、avro原理详解、avro下载与安装、avro串行化API读写实战、avro串行与反串行性能评测以及数据结构分析 3. PB串行化原理、pb读写API编程实战、pb串行与反串行性能评测
ZooKeeper	ZooKeeper基础	<ol style="list-style-type: none"> 1、ZooKeeper介绍 2、ZooKeeper集群搭建：本地模式，单机集群模式，分布式集群模式 3、ZooKeeper集群管理：启动，停止，状态查看 4、ZooKeeper常用命令：zkCli.sh，ls，get，set，rm，setACL，getACL，
	ZooKeeper编程	<ol style="list-style-type: none"> 1. ZooKeeper架构分析、ZooKeeper核心概念阐述、Leade、Follower 2. Zookeeper API编程，create，getChildren，setData，getData 3. Zookeeper节点类型详解，临时节点，序列节点，永久节点 4. Zookeeper观察者，一次性监听的重复监听实现手段
	ZooKeeper高级	<ol style="list-style-type: none"> 1、ZooKeeper工作原理剖析，分布式锁 2、ZooKeeper进程通信过程与端口详解，2181,2888,3888 3、ZooKeeper配置文件详解，日志文件，并发控制 4、ZooKeeper leader的快速选举原理剖析，zxid，myid双参优先排列控制 5、ZooKeeper HA下的容灾能力演练，容灾能力分析

HBase	hbase基础	<ol style="list-style-type: none"> 1、Hbase简介，集群搭建，Zookeeper节点元素考察 2、Hbase集群管理，启动，停止，webui查看 3、Hbase核心概念详解，namespace，table，column family，column，version，rowkey
	Hbase应用	<ol style="list-style-type: none"> 1、Hbase shell常见操作：create_namesapce，list_namespace_tables，put，get，scan，delete 2、Hbase API编程，Admin，HTable
	Hbase架构	<ol style="list-style-type: none"> 1、HBase架构详解，HRegionServer内部构成原理 2、HMaster，HbaseRegion，WAL，Store，Memostore，HStore，HFile，HBase元数据存储 3、表切割，区域切割，表合并，区域合并，表预切割，区域移动， 4、Hbase千万数据批量写入优化处理，WAL关闭，缓存自动清理关闭，batch，表预切割，多线程，rowkey区域化 5、Hbase读查询优化，rowkey范围锁定，缓存设置，批处理设置，Hbase过滤器使用 6、Hbase计数器，点击流统计实现，
	HBase高级	<ol style="list-style-type: none"> 1. Hbase原生扫描、版本控制、最小版本控制、过期控制、删除数据保留控制 2. Hbase文件压紧优化、major_compact，内存缓冲区调整 3. Hbase切割风暴、热点问题、key加盐处理、协处理器、区域观察者、二级索引实现技巧 4. Hbase的大批量数据迁移、与Hive表映射、Hbase在MR中编程实战
phoenix	phoenix	<ol style="list-style-type: none"> 1. Phoenix介绍、phoenix安装、phoenix shell crud操纵 2. phoenix原理、phoenix clientAPI使用、phoenix二级索引原理
Redis	redis安装	<ol style="list-style-type: none"> 1、Redis简介，Redis安装与配置测试 2、Redis操作，hashmap，list，set，sorted set，get，put， 3、Redis linux下编译安装，Redis client访问API 4、Redis集群安装与配置，发布订阅模型，
Flume	Flume基础	<ol style="list-style-type: none"> 1. Flume介绍、Flume安装与配置 2. Flume源头配置：netcat、spooldir、exec、sequence序列源、stress压力源 3. Flume通道配置：内存通道、文件通道 4. Flume Sink配置：file sink，log4j sink，hdfs sink，hbase sink，hive sink 5. Flume拦截器：静态拦截器，时间戳拦截器 6. Flume通道选择器：replicating，multiplexing 7. Flume sink处理器的load_balance与failover配置 8. Flume的扇入扇出配置、avro实现hop跃点控制
	Flume高级	<ol style="list-style-type: none"> 1、自定义Flume拦截器实现限速处理 2、Flume通道选择器类型配置，replicating，multiplexing 3、Flume sink处理器的load_balance与failover配置 4、使用avro source，avro sink实现flume的hop跃点处理

SSM	Spring	<ol style="list-style-type: none"> 1、Spring原理, IOC, AOP思想, 配置实战 2、Spring API编程, Spring注解配置 3、Spring配置实现AOP编程, 注解AOP实现 4、IDEA下的Spring开发集成使用
	SpringMVC	<ol style="list-style-type: none"> 1、SpringMVC架构原理剖析, DispatcherServlet, 处理器映射, Controller, ViewResolver, View 2、IDEA下SpringMVC开发配置与项目实战 3、JavaEE编程规范, web开发中数据范围, page, request, session, application 4、SpringMVC中请求转发处理, 重定向处理, 静态资源配置
	Mybatis	<ol style="list-style-type: none"> 1、MyBatis原理, MyBatis使用 2、MyBatis client API编程实践, 一对多关联, 多对一, 一对一, 多对多
	SSM	<ol style="list-style-type: none"> 1、SSM整合原理剖析, 2、IDEA下SSM整合详解, 常见错误处理 3、中文乱码处理, json数据格式处理
Kafka	Kafka基础	<ol style="list-style-type: none"> 1、Kafka介绍, Kafka集群搭建, Kafka基本配置, Kafka服务器的启动与停止, 2、Kafka主题的建立、删除与显示, 控制台消息生产与消费, 3、Kafka API使用
	Kafka架构	<ol style="list-style-type: none"> 1、Kafka架构原理详解, 2、Kafka各组件详解, Broker, topic, partition, replication, partition offset, consumer offset, consumer group Zookeeper组件协同, leader 3、Kafka容灾能力详解, Kafka日志目录构成分析与分布,
	Kafka高级	<ol style="list-style-type: none"> 1、零拷贝原理, 分区内强顺序保证, 磁盘线性读写处理, 2、至少一次消费语义, 之多一次消费语义, 精准一次消费语义 3、Kafka新型API使用 4、Kafka消费者组、消费者、消费线程、分区、并发消费能力逻辑关系详解, 5、生产者配置与优化, 客户端缓冲区控制, 批处理控制, 失败重试控制, 请求超时控制, 逗留时间控制 6、确认回执控制, 异步消息机制下的同步消息保证, 定向分区消费, 7、二进制消息处理, Kafka实现屏幕广播

Scala	Scala基础	<ol style="list-style-type: none"> Scala简介, Scala解释器, Scala安装与配置 Scala shell使用、常量与变量、Scala REPL、 Scala常用类型, 运算符与操作符重载, 函数与方法, apply
	scala控制结构	<ol style="list-style-type: none"> 条件表达式, 语句终止, 块语句与赋值, 输出输入, 循环, for循环推导式, 函数, 默认参数, 命名参数, 变长参数, 过程, 延迟计算, 异常
	数据结构	<ol style="list-style-type: none"> 定长数组: 变长数组, 缓冲区, 数组转换与遍历 常用算法: 多维数组, 与java互操作 构造映射: 映射读写, 迭代映射, 排序映射, 同java互操作, 元组计算, 拉链计算
	OOP	<ol style="list-style-type: none"> 简单类、getter、setter、主构造、辅助构造 单例对象, 伴生对象, 特质, apply方法 中置操作符, 一元操作符, apply与update 高阶函数, 值表现, 参数表现, 匿名函数, 类型推断 闭包、SAM、柯里化、实战一元线性方程高阶函数处理
	集合	<ol style="list-style-type: none"> 可变集, 不可变集, Seq, List, 可变列表, 化简, 折叠, 扫描 拉链操作, 迭代器, java互操作,
	模式匹配	<ol style="list-style-type: none"> 守卫, 变量, 类型匹配, 数组匹配, 元组匹配, 列表匹配 提取器, for模式匹配, 样例类, 密封类, Option类, Some, None 偏函数, 尾递归
	类型参数与隐式转换	<ol style="list-style-type: none"> 泛型类, 泛型函数, 类型变量界定, 视图界定, 型变, 协变, 逆变 隐式转换, 引入隐式转换, 隐式转换规则, 隐式参数, 利用隐式参数进行隐式转换, 上下文界定

Spark	Spark基础	<ol style="list-style-type: none"> 1. Spark介绍, Spark安装配置, 本地模式实现wordcount 2. spark standalone搭建, spark集群启动与停止, spark脚本分析 3. spark webui, spark job提交 4. Spark shell下操作, Spark API Scala实现, Spark API java实现
	SparkJob	<ol style="list-style-type: none"> 1. Spark job提交流程详解与UML图绘制、调度架构源码分析、Dag调度、任务调度、后台调度 2. Spark核心组件: Job、RDD、Stage、Taskset、Task、Dependency 3. Spark依赖计算法则、宽依赖、窄依赖、串行化, 分区算法
	Spark RDD	<ol style="list-style-type: none"> 1. SparkRDD简介、RDD transform、RDD action 2. Spark RDD内部结构详解、Spark RDD执行原理剖析 3. Spark内存计算原理剖析 4. Spark并行度、分区数、并行能力、Executor、执行线程、slot、cpu内核、内存、task之间的对应关系详解 5. Spark shuffle优化, repartition, coalesce, 数据倾斜, 自定义key, 自定义分区类 6. Spark实现union操作, reduceByKey, aggregateByKey 7. Spark底层聚合函数原理详解
	spark job部署与资源分配	<ol style="list-style-type: none"> 1、部署模式详解, client, cluster 2、集群模式, local, standalone, yarn 3、driver-memory, driver-core, executor-memory, executor-core 4、spark的物理资源管理与配置, worker节点内存数, worker节点内核数
	Spark shuffle	<ol style="list-style-type: none"> 1. Shuffle简介、Shuffle涉及组件介绍、ShuffleManager、ShuffleHandle、ShuffleWriter、ShuffleReader 2. Shuffle依赖与Shuffle句柄的注册原理解析、shuffle句柄实现机制解析、串行shuffle、迂回shuffle、排序shuffle的异同点分析 3. Shuffle Writer配置策略分析、Shuffle Writer内部写执行原理分析、unsafe写入、Sort写入、迂回写入 4. Shuffle行为的控制手段、迂回阈值设置、map端合成控制、串行化器控制、聚合器定义控制、分区数阈值控制 5. 迂回Writer零拷贝写入控制、Unsafe Writer的写入过程分析
	Spark内存管理	<ol style="list-style-type: none"> 1. 内存管理器介绍、静态内存管理、统一内存管理与异同 2. Spark内存结构、用户内存、保留内存、spark内存、存储内存、执行内存 3. 内存空间控制处理、保留内存的开关控制、内存模型控制配置 4. Spark内存池实现原理、内存的分配与计量 5. 执行内存与存储内存的借调详解
	Spark广播变量	<ol style="list-style-type: none"> 1、Spark广播变量原理分析, Scala延迟计算在广播变量的应用, 2、Spark广播变量的使用技巧和场景, TorrentBroadcast中BT算法 3、BlockManager块管理器原理分析, 大型变量的广播化处理手段与优势分析 4、块管理器的内存储和磁盘存储, 存储级别控制与分析, 5、累加器原理与使用, 自定义累加器实现driver端双聚合极值,
	Spark SQL	<ol style="list-style-type: none"> 1. SparkSQL简介、SparkSQL使用与hive的无缝集成 2. Spark SQL常规操作、join、union、subselect、group by等 3. Spark SQL API编程实现、DataFrame使用、临时表、全局表、Shema定义、Row使用 4. Spark SQL多源操纵: parquet、json、jdbc 5. Spark SQL分布式查询引擎原理、配置、客户端API编程访问 6. Spark SQL调优处理、缓存表、DataFrame缓存、缓存撤销

Spark	Spark Streaming	<ol style="list-style-type: none"> 1. Spark流计算介绍、流计算编程体验、集群环境流计算应用部署 2. Spark流计算工作原理、receiver工作原理与常见故障、window操作 3. Spark流计算编程API实战、scala与java编程技巧辨析 4. Spark流计算updateStateByKey实现windows操作 5. Spark流计算优化、foreachPartition 6. Spark 流计算容灾处理、检查点原理与常见错误处理 7. Spark流计算与Kafka集成、定位策略详解，消费策略详解 8. Kafka流计算BlockRDD与KafkaRDD原理剖析、流计算分区控制手段
	Spark ML	<ol style="list-style-type: none"> 1. 数理统计：kmean、median、mode、range、variance、standard deviation、skewness、Kertosis 2. 监督学习、非监督学习、 3. 算法原理：逻辑回归、线性回归、神经网络、SVM、朴素贝叶斯、kmean、KNN、最小二乘法 4. 算法实战：酒类质量预测与分类、垃圾邮件分类、模型持久化 5. word分词处理、TF-IDF、HashingTF实现原理、特征值抽取、特征向量、Tokenizer、Pipeline、松散向量、密度向量，
azkaban	azkaban	<ol style="list-style-type: none"> 1. Azkaban介绍、azkaban安装与注意要点、Web server配置要点、executor配置要点、 2. Job编写、flow控制、flow计算法则、分叉控制、聚合控制、测试执行、调度控制、常见
Python 与 数据分析	Python 与 数据分析	<ol style="list-style-type: none"> 1. python介绍、python环境安装、python体验 2. python基础、语法、数据类型、分支、循环、判断、函数 3. python oop、多线程、io、socket、模块、包、导入控制 4. python正则表达式、python爬虫实现 5. 行列式基础、转置、矩阵定义、矩阵运算、逆矩阵、矩阵分解、矩阵变换、矩阵的秩、向 6. Python对常用矩阵算法实现 7. Python常用算法库原理与使用、numpy、pandas、sklearn 8. 数据加载、存储、格式处理 9. 数据规整化、绘图与可视化

项目	南方电信集团 通话数据迁移改造工程	<p>项目介绍、项目背景与客户需求 遗留系统数据模型分析 大数据产品方案选型与参考标准 Hbase数据模型设计与实现 Rowkey的分析与设计实现、协处理器实现主被叫信息的同步更新与一致性访问 大批量数据转移实现与mr编程优化处理 通话数据的实时查询与统计分析处理 集群Master HA保证与伸缩性处理</p>
项目	基于人工智能和自然语言处理技术的自动客服系统	<p>项目介绍、业务模型与行业术语阐述 核心算法库常用算法原理与使用场景分析、gensim、nltk 常用web框架flask使用说明 常见主题模型算法原理与使用技巧、SVD、LDA、LSI 基于gensim算法模型建立流程实现 分词算法原理与应用场景说明 推荐系统优化处理与效果展示</p>
项目	基于协同过滤的垂直电商实时推荐系统	<p>项目介绍、业务模型与行业术语阐述、功能模块演示与讲解 评测模块中相关概念阐述与算法实现 算法与核心库实现、准确率、召回率、覆盖度、多样性、信息熵、基尼系数 实时推荐引擎模块数据结构设计与实现、产品选型与执行标准 用户行为数据清洗处理、行为类型分析与权重设计、推荐策略的设计与实现 推荐引擎中相似度算法实现、jaccard算法、余弦算法、john breese算法、基于商品协同过滤算法原理与实现 动态化调整推荐模型的设计与实现 推荐算法的设计与实现、基于邻域模型的协同过滤核心代码实现 基于隐语义模型算法原理原理与实现 实时计算模式下推荐模型的实时更新处理、周期性推荐模型校准处理 推荐策略中热门推荐与协同过滤的组合应用、用户与商品的冷启动处理</p>
项目	大型外包公司项目风险评估系统	<p>项目介绍、业务模型与行业术语阐述、功能模块演示与讲解 人工智能背景与生产环境下应用阐述、自然语言处理背景与工业应用 核心python算法库介绍：numpy、pandas、sklearn、 工作流分解、数据标注、数据预处理、模型实现与模型评估 风险因子与维度划分、SVM算法理论与实战应用 k-folded交叉验证原理与实战应用、模型评估指标应用、准确、召回、F值、ROC曲线、AUS、K-S曲线、基尼系数 风险评估过程的过采样、欠采样控制 语料库使用分析、PFR人民日报语料库、宾州树库、华尔街日报、布朗语料库，ATIS和Switchboard语料库 语言模型处理、NN模型原理说明、语言模型评分处理、话题模型评分处理 n-gram语言模型处理、word2vec与Textrank算法比较、 开源NLP简介 核心功能模块算法及实现、词干化、词型标注、语法依存分析、命名实体识别、语义依存分析、指代消减、关系提取、最大熵模型检测评估、条件随机场中文精度处理、半-马尔科夫建模对实体识别干扰处理</p>

项目	移动物联网综合数据服务平台	<p>项目介绍、业务模型与行业术语阐述、功能模块演示与讲解 项目整体架构分析与设计、核心模块划分与设计思路 项目中大数据集群规划与设计、资源节点的配置与原则、大数据集群建设与核心脚本编写</p> <p>nginx web服务器集群规划与部署、nginx反向代理服务器安装与配置及后台服务器权重配给、日志格式的指定与参考依据、nginx非常规访问的故障处理、nginx多级代理情形的client真实ip提取处理、时钟对齐的消息头提取控制、lua插件的引入与安装、nginx日志的滚动处理与centos的crontab调度开发</p> <p>日志类别划分与准则、手机端程序日志格式分析与设计、数据模拟程序实现与设计技巧、设计模式的综合运用与内省模式对数据模型处理技巧、数据串行化处理与上报实现</p> <p>kafka集群规划与部署、kafka主题设计与实现、分区与副本的设定与实现原则 flume实时日志收集代理配置、收集日志的限速与动态调整实现、防丢失的处理、防重处理</p> <p>hdfs原生日志的存储设计与实现、hive数据仓库模型设计与实现、生产环境下的分区划分技巧、hive数据清洗处理、UDTF实现日志差分处理、时钟对齐处、地理位置处理、hive数据格式转储处理、自定义时间戳函数、特定时间段起始时刻提取、数据分析脚本实现、新增、活跃、留存、沉默等指标进行多维度分析</p> <p>使用spark流计算实现新增用户实时双向聚合、避免数据全表扫描、SSM实现数据可视化处理、使用echart完成数据展现</p>
----	---------------	--