

## 【进阶】大数据云计算架构师专家课程

第一阶段 分布式架构技术	
讲解课程	知识要点
Linux	Vmware 安装配置、安装过程中典型问题及解决方案、Linux 文件系统、Linux 常用命令和工具、Linux 网络配置、linux 文件和目录操作、liunx 系统管理、文件权限操作、win7 和 centous 联网设置、克隆虚拟操作系统、Linux 免密码原理和配置、免密码登录、Linux Shell 编程、运行 Shell 脚本有两种方法、Shell 变量、Shell 数组、Shell 流程控制、Shell 函数、Shell 传递参数、Shell 文件操作、Shell 输入/输出重定向
JVM 调优&Java 并发	Jvm 内存管理、jvm 内存分析、内存分配原则、垃圾回收算法、堆的遍历、MarkCompact、内存泄漏、Jps、实时垃圾回收、Jstack、Jstat、Jmap、Jhat、Jconsole、收集器原理、堆大小设置、年老代年轻代大小划分  VisualVM、Java 并发包、Lock、线程池、阻塞队列、并发容器、Fork/join
Redis	Redis 安装配置、Redis 命令、Redis 数据类型、Redis 键(key)、Redis 字符串(String)、Redis 哈希(Hash)、Redis 集合(Set)、Redis 发布订阅、Redis 事务、Redis 脚本、Redis 安全、Redis 性能测试、Redis 分区、Jedis Api 开发、Redis 管道技术、Redis 集群、Sentinel、

MongoDB	MongoDB 安装配置、MongoDB - 连接、MongoDB 创建集合、MongoDB 删除集合、MongoDB 插入文档、MongoDB 更新文档、MongoDB 条件操作符、MongoDB Limit 与 Skip 方法、MongoDB 排序、MongoDB 索引、MongoDB 聚合、MongoDB 备份 (mongodump)与恢复(mongorestore)、MongoDB 监控、数据类型、数据库操作、集合、文档、映射、索引、复制、分片、Java API 编程、查询分析、高级索引、全文检索、Rockmongo
MySql 集群	mysql 的架构、mysql 事务、mysql 索引、数据备份/恢复、Mysql 性能优化、主从复制、mysql proxy、mysql 的 MMM 架构、MySQL 同步粒度及引擎、MySQL Replication 工作机制、mysql 的源码分析、Mysql Cluster、Mycat 配置、读写分离、分库分表、mysql 的负载均衡
Docker	Docker 介绍、镜像、容器、仓库、Docker 安装配置、Docker Hello World、Docker 应用、Docker 镜像使用、Docker 容器连接、网络端口映射、Docker 安装 Nginx、Docker 安装 MySQL、Docker 安装 Tomcat、Docker 安装 Redis、Docker 安装 MongoDB、Docker 安装 Apache
Zookeeper	Zookeeper 架构、Zookeeper 安装、Zookeeper 命令、工作流、Leader 选举、Zookeeper API 编程、Zookeeper 工作流、ZooKeeper 集合中的节点、Zookeeper leader 选举、Zookeeper CLI、连接到 ZooKeeper 集合

Dubbo	<p>Dubbo 的总体架构、Dubbo 架构解析、解析服务、服务注册、服务监控、服务监控、远程通信与信息交换、服务调用、暴露服务、dubbo 的底层原理、Dubbo Activate、Dubbo+ookeeper、dubbo-admin、Dubbo+zk+spring 整合案例</p>
Spring Cloud	<p>Spring boot、Spring cloud config、Spring Cloud 服务的注册与发现 ( Eureka )、Spring Cloud 服务提供者 Eureka + 服务消费者 ( rest + Ribbon )、Spring Cloud 断路器监控(Hystrix Dashboard)、Spring Cloud 高可用的分布式配置中心 Spring Cloud Config、Spring Cloud 高可用的分布式配置中心 Spring Cloud Config 集成 Eureka 服务、断路器、智能路由、客户端负载均衡、Docker 部署 spring cloud 项目、</p>
分布式爬虫	<p>爬虫搜索原理、网络爬虫系统的工作原理、抓取策略、宽度优先遍历策略、反向链接数策略、网络爬虫系统的基本结构、分布式网络爬虫的工作原理、基于 Web 数据采集的搜索引擎系统-基本架构、爬虫基本流程设计、爬虫开发实践、反爬虫技术、WebMagic 框架、Nutch 框架、爬虫项目实战、《网易新闻数据爬取解析》</p>
区块链	<p>区块链技术原理、区块链特性、区块结构、区块链如何工作、区块链发展趋势、Merkle Tree、数字指纹、区块链的 Java 实现、HyperLedger Fabric 环境搭建、HyperLedger Fabric 集群、加密货币、智能合约、主要代币、CouchDB、Fabric-SDK-Java 开发实例、</p>

<p><b>综合项目</b></p>	<p><b>项目讲解：</b>分布式电商架构（淘宝网架构搭建）</p> <p><b>功能讲解</b> Linux 下 Nginx+Tomcat 负载均衡和动静分离配置要点、Tomcat 负载均衡配置、静态资源分离配置、Echarts、支付系统实现、秒杀系统实现、图片系统实现、分布式缓存、数据库主备集群、分布式会话、单点登录系统实现 SSO、分布式消息队列、分布式系统分层架构、消息中间件 ActiveMQ 的安装和使用、Spring 整合 ActiveMQ 实践</p> <p><b>配套教材：</b>项目需求说明书、项目案例使用素材、教员 1 对 1 答辩答疑、远 程案例辅导、重要代码参考、sql 脚本</p>
<p><b>第二阶段 大数据离线处理</b></p>	
<p><b>讲解课程</b></p>	<p><b>知识要点</b></p>
<p>Hadoop</p>	<p>大数据技术生态体系、Hadoop 系统框架原理分析、hadoop 系统功能和结构功能介绍、Hadoop 集群搭建、secureCRT 和 FileZilla 的使用介绍、hadoop 集群搭建--配置文件、Hadoop 集群中各节点功能分析、hadoop 集群数据观察、向集群提交任务演示、向集群提交任务、Hdfs 命令使用、hdfs 指令实践、Hdfs 编程实践、java 操作 hdfs、MapReduce 框架介绍、mapreduce 编程框架和 wordcount 程序思路分析、编程实践—单词统计、wordcount 代码编写、wordcount 集群运行和观察、编程实践—单词去重、win7 下调试运行单词去重程序、编程实践—流量排名、流量统计、编程实践—流量倒排、流量倒排思路分析、流量倒排索引实现、编程实践-手机号分区、手机号分区解决思路、手机号分区代码实现、编程实践-表的 join 操作、Combiner 实现思路分析、jion 思路分析、编程实践-combiner、编程实践-qq</p>

	<p>好友、qq 共同好友、QQ 好友推荐实现、MapReduce 框架剖析、mapreduce 计算框架、mapreduce 计算框架总结、mapreduce 项目--商品推荐系统</p>
Hive	<p>Hive 原理和框架搭建、数据仓库是什么、hive 和 hadoop 的关系、hive 中的数据存储、linux 中安装 mysql、安装和配置 hive、hive 初步体验、在 mysql 中查看元数据信息、hive 中的 ddl、hive 中 dml、Hive 视图和索引、Hive 分区、Hive 内置运算符、Hive 内置函数、hive 中 join 语句、hive 常用函数、java 操作 hive</p>
Hbase	<p>hbase 概念、hbase 中的表结构、hbase 系统结构和存储、Hbase 几个特点介绍、列式存储、海量存储、zookeeper 介绍、hbase 的安装和配置、hbase shell 指令介绍、Hbase 的 Region、hbase shell 指令实践、java 操作 hbase、Hbase 写入逻辑、MemStore 刷盘、Hlog 结构、Hlog 的生命周期、RegionServer 的故障恢复、Region 的拆分、Region 的合并、mapreduce 中使用 hbase、</p>

Flume	<p>Flume 的概念、Flume 的结构、Flume 的应用、Flume 搭建_案例一：单个 Flume、Flume 日志数据采集、聚合、传输、Flume 搭建_案例二：两个 Flume 做集群、Flume 搭建_案例三：如何监控一个文件的变化？、Flume 搭建_案例四：如何监控一个文件：目录的变化？、Flume 搭建_案例五：如何定义一个 HDFS 类型的 Sink?、案例六_Spooling Directory Source：监听一个指定的目录</p>
Oozie& Azkaban	<p>Oozie 安装、workflow 工作流、coordinator 讲解、bundle 介绍、Oozie 测试和使用、Oozie shell 编写、workflow.xml 解析、WordCount 例子、MapReduce 的 old API 跟 new API、azkaban 介绍、工作流调度实现方式、Azkaban 与 Oozie 对比、azkaban 安装部署、azkaban web 服务器安装、azkaban 执行服务器安装、创建 SSL 配置、azkaban 脚本导入、command job 示例、Command 类型多 job 工作流 flow、HDFS 操作任务、HIVE 脚本任务、MAPREDUCE 任务、</p>
综合项目	<p><b>项目讲解：</b>电商平台（天猫）点击日志大数据分析</p> <p><b>功能讲解：</b>mapreduce 数据处理、Flume 日志数据采集、聚合、传输、Flume 监控一个文件的变化 azkaban 工具使用、Command 类型多 job 工作流 flow、HDFS 操作任务、HIVE 脚本任务、MAPREDUCE 任务、商品推荐的实现思路训练、大型 web 电商项目前后台系统实现</p> <p><b>配套教材：</b>项目需求说明书、项目案例使用素材、教员 1 对 1 答辩答疑、远程案例辅导、重要代码参考</p>
第三阶段 实时处理与内存计算	

讲解课程	知识要点
Storm	Storm 核心概念、流分组、 Storm 工作流程、Storm 应用场景、Storm 安装配置、Storm 开发初步、Storm 分布式消息系统、Thrift 协议、Storm 工作实例、Trident 的工作实例、Storm 在 Twitter 上的应用、Hashtag 阅读器 spout、Storm 集群搭建、Storm 进阶案例、Storm 在雅虎财经上的应用、天气频道使用 Storm 拓扑来获取天气数据、Storm 项目实践
Kafka	Kafka 介绍、Kafka 基础、Kafka 环境搭建、Kafka 工作流程、订阅消息的工作流程、队列消息/用户组的工作流、ZooKeeper 的作用、Kafka 生产者示例、Kafka 消费者组示例、Kafka 整合 Storm、Kafka 与 Spark 的集成、KafkaConfig API、Kafka Java 编程实践、Kafka 集群搭建、Kafka 工具 Kafka+Storm+flume 集成、Kafka 实时应用程序(Twitter)
storm 项目	车辆实时数据采集 , storm 实时处理 web 页面展示 对外提供短信查询服务
Scala	scala 介绍及环境安装、Finagle 介绍、第一个 Scala 程序、Searchbird 数值类型、if 和 match、块表达式和循环、方法和函数的定义和使用、集合之数组、集合之 list、集合之 Map、集合之 Set、集合之元组、模式匹配与函数组合、类型和多态基础、参数化多态性、高级类型、简单构建工具 SBT、序列 Seq、层次结构、可变集合、使用 specs 测试、Scala 并发编程、线程安全问题、不安全的搜索引擎、Java 与 Scala、

Akka	<p>akk 介绍、使用 akka 带来的好处、Akka 环境搭建、Actors、actor 生命周期、消息投递、inbox 消息收件箱、路由器 Router、不可变对象、容错机制、消息通道、事务、内置状态转换 Procedure、Actor 中的 Future-询问模式、软件事务内存、agent 的安全处理、akka 持久化、Routing、IO、Spring 与 Akka 的集成、akka 分布式、分布式 Akka 应用程序开发</p>
Spark	<p>Spark 应用场景、Spark 安装、Spark SQL 的特性、弹性分布式数据集、Spark SQL 架构、Spark SQL 数据帧、DataFrame 的特性、Spark SQL 数据源、MapReduce 中的数据共享缓慢、Hadoop+Spark 集群搭建、idea 工具上编写程序任务并提交集群、Spark 交互式命令、利用 spark shell 实现 wordcount、rdd 算子简单应用、高级算子、Spark 实践-统计停留时间最长的基站、判断位置-背景讲述、统计停留时间最长的基站、Spark 实践-统计家庭住址和工作地点、Spark 实践-统计网站访问排名前三的网页、Spark 实践-网站前三名独立文件-自定义分区、Spark 实践-自定义排序、Spark 实践-ip 归属地查询、Spark 实践-mysql 数据库、统计结果写入 mysql 数据库、从传统数据库中读取数据、Spark Streaming 概述、wordcount 代码、wordcount 之累加统计、kafka 集群搭建、从 kafka 中实时拉取并处理数据、spark sql 了解、dataframe 操作的两种方式、spark sql 编程、Spark 机器学习库 MLlib 介绍、Spark 机器学习算法-回归分析原理与实践、Spark 机器学习算法-Bayes 原理与实践、Spark 机器学习算法-Knn 原理与实践、Spark 机器学习算法-svm 原理与实践、Spark 机器学习算法-决策树原理与实践、Spark 机器学习算法-KMeans 原理与实践、Spark GraphX 介绍、Graphx 算法-社区发现原理与实践</p>



## 综合项目

**项目讲解：**《基于 Spark 的大数据房源画像》《Spark 云日志分析》

**功能讲解：**MapReduce 中的数据共享缓慢、Hadoop+Spark 集群搭建、Spark 机器学习库 MLlib 介绍、Spark 机器学习算法运用、Spark 统计家庭住址和工作地点、wordcount 之累加统计、Spark GraphX 实战

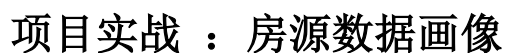
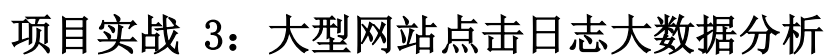
**配套教材：**项目需求说明书、项目案例使用素材、教员 1 对 1 答辩答疑、远程案例辅导、重要代码参考

注：学习过程中部分项目截图

## 项目实战：大型分布式电商《淘宝网》系统架构

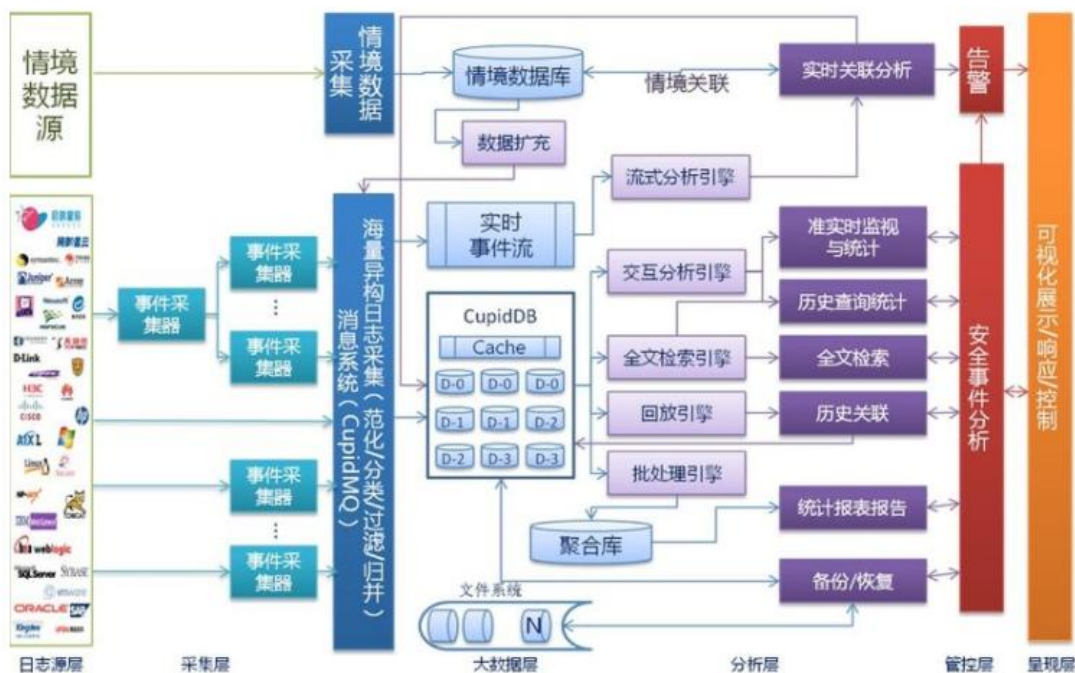


## 项目实战：互联网典型电商品牌信息分析





## 项目实战：情境日志分析系统



## 项目实战：地震大数据预警系统

## 1. 地震烈度速报



烈度速报就是在破坏性地震发生时能够快速给出不同地区的烈度分布情况。

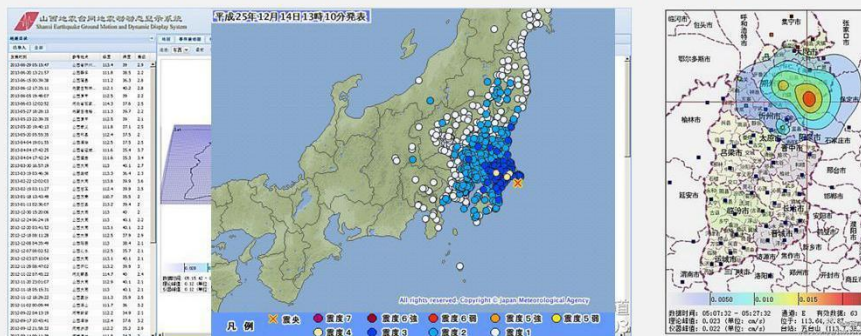


图10-1 山西地震台网的观测烈度速报

## 项目实战：金融投资大数据分析系统



## 项目实战：春运大数据分析系统



## 阿里云大数据云计算培训战略合作伙伴，权威资质，权威保障

证书图样

