

云计算大数据实战班

六大真实企业大数据项目，助你成为大型项目（数据量大，并发高）的顶级架构师！

课程内容涵盖：高并发架构，hadoop,storm,spark,openstack 等热门技术！

官网 [HTTP://WWW.BJSXT.COM/HTML/CLOUD/](http://www.bjsxt.com/html/cloud/)

目录

目录	1
课程说明（必读）	3
课程线路	3
Linux 大纲	4
大型网站高并发处理	5
Hadoop 离线计算大纲	6
分布式数据库 Hbase	8
数据仓库 Hive	9
Lucene 课程	10
Solr/solrcloud 课程	11
elasticsearch 分布式搜索	12
CM+CDH 集群管理课程	13
Impala 课程	13
Oozie 课程	14
数据迁移工具 Sqoop	14
Flume 分布式日志框架	15
Zookeeper 开发	16
R 语言机器学习	16
Mahout 机器学习	17
Redis 缓存数据库	19

Kafka 分布式队列系统课程	20
Storm 实时数据处理	20
Python 课程	22
Scala 课程	23
Spark core 大数据编程	24
Spark SQL	26
Spark Streaming 实时计算	26
Spark MLlib 机器学习	27
Sparkt GraphX 图计算	28
Tachyon 内存分布式文件系统	29
Docker 课程	29
虚拟化 KVM	30
云平台 OpenStack	31
大型企业项目实战	31
项目一某大型电商日志分析和订单管理	31
项目二某运营商基站保障平台	33
项目三微博营销数据挖掘项目	33
项目四智慧城市大数据分析项目	34
项目五手机软件推荐系统项目	34
项目六网络流量异常检测项目	35
毕业学员就业薪资	35
就业前景调查	40
当前教室一览	49
联系方式	52

课程说明（必读）

本系列课程适合有一点编程基础的人员学习（比如 java,python,c/c++），最好是 java 编程人员，特别是从事过 j2ee 开发的人员。因为本套课程属于高端课程，所以报名过程中我们会有对报名人员进行简单技术评估，希望大家不要嫌麻烦。因为这样是对您最好的负责，也能更好保障您能学会并消化我们的课程（有一定难度和深度）内容，另外我们也可以给你一些比较好的复习和学习建议。学习完本套课程，可以帮助你成为大型项目架构师，特别是数据量大，并发量高的大型项目架构师，当然也能很大程度上提高你的薪资待遇。

课程线路

六大真实企业大数据项目，助你成为大型项目（数据量大，并发高）的顶级架构师！

课程内容涵盖：

高并发架构，hadoop,storm,spark,机器学习，docker,openstack 等热门技术

云计算极限班课程大纲



Linux 大纲

这章是基础课程，帮大家进入大数据领域打好 Linux 基础，以便更好地学习

Hadoop，NoSQL，Oracle，MySQL，Spark，Storm 等众多课程。因为企业中无一例外的是使用 Linux 来搭建或部署项目。

- 1) Linux 的介绍，Linux 的安装：VMware Workstation 虚拟软件安装过程、CentOS 虚拟机安装过程
- 2) 了解机架服务器，采用真实机架服务器部署 linux
- 3) Linux 的常用命令：常用命令的介绍、常用命令的使用和练习
- 4) Linux 系统进程管理基本原理及相关管理工具如 ps、pkill、top、htop 等的使用；
- 5) Linux 启动流程，运行级别详解，chkconfig 详解

- 6) VI、VIM 编辑器：VI、VIM 编辑器的介绍、VI、VIM 的使用和常用快捷键
- 7) Linux 用户和组账户管理：用户的管理、组管理
- 8) Linux 磁盘管理，lvm 逻辑卷，nfs 详解
- 9) Linux 系统文件权限管理：文件权限介绍、文件权限的操作
- 10) Linux 的 RPM 软件包管理：RPM 包的介绍、RPM 安装、卸载等操作
- 11) yum 命令，yum 源搭建
- 12) Linux 网络：Linux 网络的介绍、Linux 网络的配置和维护
- 13) Shell 编程：Shell 的介绍、Shell 脚本的编写
- 14) Linux 上常见软件的安装：安装 JDK、安装 Tomcat、安装 mysql、web 项目部署

大型网站高并发处理

通过本章的学习大家将会了解大数据的源头，数据从何而来，继而更好的了解大数据。

并且通过学习如果处理大型网站高并发问题反向更深入的学习了 Linux，同时站在了更高的角度去触探了架构。

- 1) 第四层负载均衡
 - a) Lvs 负载均衡
 - i. 负载均衡算法，NAT 模式，直接路由模式（DR），隧道模式（TUN）
 - b) Haproxy
- 2) 第七层负载均衡
 - a) Nginx/tengine
- 3) Tomcat、jvm 优化提高并发量

- 4) 缓存优化
 - a) Java 缓存框架
 - i. OScache , ehcache
 - b) 缓存数据库
 - i. Redis , Memcached
- 5) Lvs+nginx+tomcat+redis|memcache 构建二层负载均衡千万并发处理
- 6) Fastdfs 小文件独立存储管理

Hadoop 离线计算大纲

一、 初识 hadoop

听过大数据，必听过 hadoop，此部分带领大家了解 hadoop 的用途，在大数据中的用途，以及快速搭建一个 hadoop 的实验环境，在本过程中不仅将用到前面的 Linux 知识，而且会对 hadoop 的架构有深入的理解，并未你以后架构大数据项目打下坚实基础。

- 1) Hadoop 生态环境介绍
- 2) Hadoop 云计算中的位置和关系
- 3) 国内外 Hadoop 应用案例介绍
- 4) Hadoop 概念、版本、历史
- 5) Hadoop 核心组成介绍及 hdfs、mapreduce 体系结构
- 6) Hadoop 的集群结构
- 7) Hadoop 伪分布的详细安装步骤
- 8) 通过命令行和浏览器观察 hadoop

二、 HDFS 体系结构和 shell 以及 java 操作

详细剖析 HDFS，从知晓原理到开发网盘的项目让大家打好学习大数据的基础，大数据之于分布式，分布式学习从学习分布式文件系统（HDFS）开始。

- 1) HDFS 底层工作原理
- 2) HDFS datanode,namenode 详解
- 3) Hdfs shell
- 4) Hdfs java api

三、 详细讲解 Mapreduce

Mapreduce 可以说是任何一家大数据公司都会用到的计算框架，也是每个大数据工程师应该熟练掌握的，此处的学习除了老师详细的讲解理论外，会通过大量的案例让大家彻底掌握。

- 1) Mapreduce 四个阶段介绍
- 2) Writable
- 3) InputSplit 和 OutputSplit
- 4) Maptask
- 5) Shuffle : Sort , Partitioner , Group,Combiner
- 6) Reducer

四、 Mapreduce 案例案例

- 1) 二次排序
- 2) 倒排序索引
- 3) 最优路径
- 4) 电信数据挖掘之-----移动轨迹预测分析（中国棱镜计划）

- 5) 社交好友推荐算法
- 6) 互联网精准广告推送 算法
- 7) 阿里巴巴天池大数据竞赛 《天猫推荐算法》案例
- 8) Mapreduce 实战 pagerank 算法

五、 Hadoop2.x 集群搭建

前面带领大家开发了大量的 MapReduce 程序，此部分将带领大家让开发的程序运行在分布式集群中，并且运行在健壮高可用的集群中。

- 1) Hadoop2.x 集群结构体系介绍
- 2) Hadoop2.x 集群搭建
- 3) NameNode 的高可用性 (HA)
- 4) HDFS Federation
- 5) ResourceManager 的高可用性 (HA)
- 6) Hadoop 集群常见问题和解决方法
- 7) Hadoop 集群管理

分布式数据库 Hbase

大数据中使用 Hbase 的案例多的举不胜举，也可凸显大家学习的必要性。及时工作多年的大数据工程师 Hbase 的优化也是需要好好学习的重点。

- 1) HBase 定义
- 2) HBase 与 RDBMS 的对比
- 3) 数据模型
- 4) 系统架构

- 5) HBase 上的 MapReduce
- 6) 表的设计
- 7) 集群的搭建过程讲解
- 8) 集群的监控
- 9) 集群的管理
- 10) HBase Shell 以及演示
- 11) Hbase 树形表设计
- 12) Hbase 一对多 和 多对多 表设计
- 13) Hbase 微博 案例
- 14) Hbase 订单案例
- 15) Hbase 表级优化
- 16) Hbase 写数据优化
- 17) Hbase 读数据优化

数据仓库 Hive

Hive 是使用 sql 进行计算的 hadoop 框架，工作中最常用到的部分，也是面试的重点，此部分大家将从方方面面来学习 Hive 的应用，任何细节都将给大家涉及到。

- 1) 数据仓库基础知识
- 2) Hive 定义
- 3) Hive 体系结构简介
- 4) Hive 集群
- 5) 客户端简介

- 6) HiveQL 定义
- 7) HiveQL 与 SQL 的比较
- 8) 数据类型
- 9) 外部表和分区表
- 10) ddl 与 CLI 客户端演示
- 11) dml 与 CLI 客户端演示
- 12) select 与 CLI 客户端演示
- 13) Operators 和 functions 与 CLI 客户端演示
- 14) Hive server2 与 jdbc
- 15) 用户自定义函数 (UDF 和 UDAF) 的开发与演示
- 16) Hive 优化

Lucene 课程

在大数据里面文本数据的搜索是很重要的一块，特别是里面的分词技术，是后面机器学习里面文本挖掘的基石，我们需要深入学习 java 领域里面的搜索核心技术 lucene，同时也可以了解到百度 google 这样的搜索系统是怎么架构实现的。

- 1) Lucene 介绍
- 2) Lucene 倒排索引原理
- 3) 建索引 IndexWriter
- 4) 搜索 IndexSearcher
- 5) Query
- 6) Sort 和 过滤 (filter)

7) 索引优化和高亮

Solr/solrcloud 课程

接着前面 lucene 技术搜索，如果把 lucene 技术比如为发动机，那 solr 就是一两成型的汽车了。学习完 solr 可以帮助你在企业里面快速的架构搜索系统。首先 Solr 是基于 Lucene 做的，Lucene 是一套信息检索工具包，但并不包含搜索引擎系统，它包含了索引结构、读写索引工具、相关性工具、排序等功能，因此在使用 Lucene 时你仍需要关注搜索引擎系统，例如数据获取、解析、分词等方面的东西。而 Solr 的目标是打造一款企业级的搜索引擎系统，因此它更接近于我们认识到的搜索引擎系统，它是一个搜索引擎服务，通过各种 API 可以让你的应用使用搜索服务，而不需要将搜索逻辑耦合在应用中。而且 Solr 可以根据配置文件定义数据解析的方式，更像是一个搜索框架，它也支持主从、热换库等操作。还添加了飘红、facet 等搜索引擎常见功能的支持。

- 1) 什么是 solr
- 2) 为什么工程中要使用 solr
- 3) Solr 的原理
- 4) 如何在 tomcat 中运行 solr
- 5) 如何利用 solr 进行索引与搜索
- 6) solr 的各种查询
- 7) solr 的 Filter
- 8) solr 的排序
- 9) solr 的高亮
- 10) solr 的某个域统计

- 11) solr 的范围统计
- 12) solrcloud 集群搜索详解
- 13) solr+hbase 大型搜索系统架构

elasticsearch 分布式搜索

Elastic Search 是一个基于 Lucene 构建的开源, 分布式, RESTful 搜索引擎。设计用于云计算中, 能够达到实时搜索, 稳定, 可靠, 快速, 安装使用方便。支持通过 HTTP 使用 JSON 进行数据索引。

- 1) elasticsearch 简介
- 2) elasticsearch 和 solr 的对比
- 3) elasticsearch 安装部署
- 4) elasticsearch service wrapper 启动插件
- 5) 使用 curl 操作 elasticsearch 索引库
- 6) elasticsearch DSL 查询
- 7) elasticsearch 批量查询 meet
- 8) elasticsearch 批量操作 bulk
- 9) elasticsearch 插件介绍
- 10) elasticsearch 配置文件详解
- 11) java 操作 elasticsearch
- 12) elasticsearch 的分页查询
- 13) elasticsearch 中文分词工具的集成
- 14) elasticsearch 优化

15) elasticsearch 集群部署

16) elasticsearch+hbase 大型搜索系统架构

CM+CDH 集群管理课程

Cloudera 的定位在于 Bringing Big Data to the Enterprise with Hadoop

Cloudera 为了让 Hadoop 的配置标准化，可以帮助企业安装，配置，运行 hadoop 以达到大规模企业数据的处理和分析。由 cloudera 公司开发的集群 web 管理工具 cloudera manager(简称 CM)和 CDH 目前在企业中使用的比重很大，掌握 CM+CDH 集群管理和使用 不仅简化了集群安装、配置、调优等工作，而且对任务监控、集群预警、快速定位问题都有很大的帮助。

- 1) CM + CDH 集群的安装
- 2) 基于 CM 主机及各种服务组件的管理
- 3) CDH 集群的配置和参数调优
- 4) CDH 集群 HA 配置及集群升级
- 5) CM 的监控管理
- 6) 集群管理的注意事项
- 7) HUE 实战详解

Impala 课程

Impala 是 Cloudera 公司参照 Google Drem 系统进行设计并主导开发的新型查询系统，它提供复杂 SQL 语义，能查询存储在 Hadoop 的 HDFS 和 HBase 中的 PB 级大数据。

网址：<http://www.bjsxt.com/html/cloud/>

- 1) Impala 介绍和架构
- 2) Impala 实战安装, 架构, 外部 shell
- 3) Impala 内部 shell, 存储分区, SQL
- 4) Impala SQL、hbase 整合, JDBC、性能优化
- 5) Impala 配置及其调优
- 6) Impala 项目应用

Oozie 课程

oozie 是 Hadoop 生态系统中比较重要的框架, 在系统中扮演的角色是工作流引擎服务器, 用于运行 Hadoop Map/Reduce 任务工作流 (包括 MR、Hive、Pig、Sqoop 等), 企业中多由于整个项目流程任务调度。

- 1) Oozie 入门介绍
- 2) Oozie 安装配置及其简单操作
- 3) hPDL 语言学习及流程定义
- 4) oozie 工作流配置及元数据库定义
- 5) oozie 定时任务调度和 oozie API 操作

数据迁移工具 Sqoop

sqoop 适用于关系型数据库和 HDFS 分布式数据系统之间进行数据转换, 在企业中, 是构建数据仓库的一大工具。

- 1) 介绍 和 配置 Sqoop
- 2) Sqoop shell 使用

- 3) Sqoop-import
 - a) DBMS → hdfs
 - b) DBMS → hive
 - c) DBMS → hbase
- 4) Sqoop-export

Flume 分布式日志框架

Flume 最早是 Cloudera 提供的日志收集系统，目前是 Apache 下的一个孵化项目，Flume 支持在日志系统中定制各类数据发送方，用于收集数据。大家学习完此节后不但可以掌握 Flume 的使用，而且可以进行对于 Flume 的开发。

- 1) flume 简介-基础知识
- 2) flume 安装与测试
- 3) flume 部署方式
- 4) flume source 相关配置及测试
- 5) flume sink 相关配置及测试
- 6) flume selector 相关配置与案例分析
- 7) flume Sink Processors 相关配置和案例分析
- 8) flume Interceptors 相关配置和案例分析
- 9) flume AVRO Client 开发
- 10) flume 和kafka 的整合

Zookeeper 开发

Zookeeper 在分布式集群（Hadoop 生态圈）中的地位越来越突出，对分布式应用的开发也提供了极大便利，这也是这里我们带领大家深入学习 Zookeeper 的原因。本课程主要内容包括 Zookeeper 深入、客户端开发（Java 编程，案例开发）、日常运维、Web 界面监控。大家这里学好 Zookeeper，对后面学习其他技术至关重要。

- 1) zookeeper 架构
- 2) zookeeper 实战环境
- 3) zookeeper 内部算法详解
- 4) Zookeeper java api 开发
- 5) Zookeeper rmi 高可用分布式集群开发
- 6) Zookeeper 实现 SOA 高可用架构框架
- 7) Netty 异步 io 通信框架
- 8) Zookeeper 实现 netty 分布式架构的高可用
- 9) Zookeeper 分布式锁实现

R 语言机器学习

R 本身是一款十分优秀的数据分析和数据可视化软件，同时作为第一代机器学习的工具，其中包括大量用于机器学习的添加包。此部分带领大家学习 R 语言更是带领大家进入机器学习的领域，机器学习算法为主线的同时，通过案例学习将会让大家对内容脉络掌握的更加清晰。

- 1) R 语言介绍，基本函数，数据类型

- 2) 线性回归
- 3) 朴素贝叶斯分类
- 4) 决策树分类
- 5) k 均值聚类
 - a) 离群点检测
- 6) 关联规则探索
- 7) 神经网络

Mahout 机器学习

Mahout 提供一些可扩展的机器学习领域经典算法的实现，很多公司会使用 Mahout 方便快捷地创建智能应用程序。Mahout 包含许多实现，包括聚类、分类、推荐过滤、频繁子项挖掘。Mahout 通过使用 Apache Hadoop，可以有效地扩展到云中。被业界奉为第二代机器学习工具。此部分过后大家不仅会学习到 mahout 的组件而且会有项目让大家真正把它应用到工作中。

- 1) 介绍为什么使用它,它的前景
 - a) 简单介绍 Mahout
 - b) 简单介绍机器学习
 - c) 实例演示 Mahout 单机推荐程序
- 2) 配置安装 (hadoop2.x 版本的) 编译安装步骤说明
 - a) 命令行中测试运行协同过滤概念
- 3) 推荐

- a) 讲解基于用户的协同过滤
- b) 讲解基于物品的协同过滤
- 4) 分类
 - a) 分类概念
 - b) 分类的应用及 Mahout 分类优势
 - c) 分类和聚类、推荐的区别
 - d) 分类工作原理
 - e) 分类中概念术语
 - f) 分类项目工作流
 - g) 如何定义预测变量
 - h) 线性分类器的介绍，及贝叶斯分类器
 - i) 决策树分类器的介绍，及随机森林分类器
 - j) 如何使用贝叶斯分类器和随机森林分类器的代码展示
- 5) 聚类
 - a) 聚类概念
 - b) 聚类步骤流程
 - c) 聚类中的距离测度
 - d) 讲解 K-means 聚类
 - e) K-means 聚类算法展示
 - f) 聚类其他算法
 - g) 介绍 TF-IDF
 - h) 归一化

i) 微博聚类案例

Redis 缓存数据库

1. redis 特点、与其他数据库的比较
2. 如何安装 redis
3. 如何使用命令行客户端
4. redis 的字符串类型
5. redis 的散列类型
6. redis 的列表类型
7. redis 的集合类型
8. 如何使用 java 访问 redis 【a.python 访问 redis,scala 访问 redis】
9. redis 的事务(transaction)
10. redis 的管道(pipeline)
11. redis 持久化(AOF+RDB)
12. redis 优化
13. redis 的主从复制
14. redis 的 sentinel 高可用
15. twemproxy,codis 实战
16. redis3.x 集群安装配置

Kafka 分布式队列系统课程

Kafka 是当下流行的队列，可以说是从数据采集到大数据计算承上启下的重要环节，大家在此部分将会详细学习它的架构，kafka 在大家大数据的项目中几乎都会涉及到。

- 1) kafka 是什么
- 2) kafka 体系结构
- 3) kafka 配置详解
- 4) kafka 的安装
- 5) kafka 的存储策略
- 6) kafka 分区特点
- 7) kafka 的发布与订阅
- 8) zookeeper 协调管理
- 9) java 编程操作 kafka
- 10) scala 编程操作 kafka
- 11) flume 和kafka 的整合
- 12) Kafka 和 storm 的整合

Storm 实时数据处理

本部分学习过后，大家将全面掌握 Storm 内部机制和原理，通过大量项目实战，让大家拥有完整项目开发思路和架构设计，掌握从数据采集到实时计算到数据存储再到前台展

示，所有工作一个人搞定！譬如可以一个人搞定淘宝双 11 大屏幕项目！不光从项目的开发的层次去实现，并可以从架构的层次站在架构师的角度去完成一个项目。

- 1) Storm 的基本概念
- 2) Storm 的应用场景
- 3) Storm 和 Hadoop 的对比
- 4) Storm 集群的安装的 linux 环境准备
- 5) zookeeper 集群搭建
- 6) Storm 集群搭建
- 7) Storm 配置文件配置项讲解
- 8) 集群搭建常见问题解决
- 9) Storm 常用组件和编程 API : Topology、 Spout、 Bolt
- 10) Storm 分组策略(stream groupings)
- 11) 使用 Storm 开发一个 WordCount 例子
- 12) Storm 程序本地模式 debug、Storm 程序远程 debug
- 13) Storm 事物处理
- 14) Storm 消息可靠性及容错原理
- 15) Storm 结合消息队列 Kafka : 消息队列基本概念(Producer、Consumer、Topic、Broker 等)、消息队列 Kafka 使用场景、Storm 结合 Kafka 编程 API
- 16) Storm Trident 概念
- 17) Trident state 原理
- 18) Trident 开发实例
- 19) Storm DRPC(分布式远程调用)介绍

- 20) Storm DRPC 实战讲解
- 21) Storm 和 Hadoop 2.x 的整合 : Storm on Yarn
- 22) 淘宝核心架构套件
 - a) Jstorm+rocket 详解

Storm 开发实战： flume+Kafka+Storm+Hbase+redis 项目实战，以及多个案例

Python 课程

Python 语言的部分大家在学习后可以完全掌握 Python 的精髓，并通过这部分的学习给大家打好一个基础，在其他计算框架中多语言的使用上都会涉及到 Python 这门流行的语言。同时课程里会通过机器学习的案例让大家学习 Python 的同时去更好的理解机器学习。

- 1) 介绍 Python 以及特点
- 2) Python 的安装
- 3) Python 基本操作（注释、逻辑、字符串使用等）
- 4) Python 数据结构（元组、列表、字典）
- 5) 使用 Python 进行批量重命名小例子
- 6) Python 常见内建函数
- 7) 更多 Python 函数及使用常见技巧
- 8) 异常
- 9) Python 函数的参数讲解
- 10) Python 模块的导入
- 11) Python 中的类与继承
- 12) 网络爬虫案例

- 13) 数据库连接，以及 pip 安装模块
- 14) Mongodb 基础入门
- 15) 讲解如何连接 mongodb
- 16) Python 的机器学习案例

Scala 课程

在此部分内，将更注重 scala 的各种语言规则与简单直接的应用，而不在于其是如何具体实现，通过学习本课程能具备初步的 Scala 语言实际编程能力。本部分课程也可以视为大家下面学习 Spark 课程的铺垫，供大家扫盲熟悉 Scala，提前进行热身运动。

- 1) scala 解释器、变量、常用数据类型等
- 2) scala 的条件表达式、输入输出、循环等控制结构
- 3) scala 的函数、默认参数、变长参数等
- 4) scala 的数组、变长数组、多维数组等
- 5) scala 的映射、元组等操作
- 6) scala 的类，包括 bean 属性、辅助构造器、主构造器等
- 7) scala 的对象、单例对象、伴生对象、扩展类、apply 方法等
- 8) scala 的包、引入、继承等概念
- 9) scala 的特质
- 10) scala 的操作符
- 11) scala 的高阶函数
- 12) scala 的集合

13) scala 数据库连接

Spark core 大数据编程

本部分内容全面涵盖了 Spark 生态系统的概述及其编程模型，深入内核的研究，Spark on Yarn, Spark Streaming 流式计算原理与实践，Spark SQL, Spark 的多语言编程以及 SparkR 的原理和运行。不仅面向项目开发人员，甚至对于研究 Spark 的学员，此部分都是非常有学习指引意义的课程。

- 1) Spark 介绍
- 2) Spark 应用场景
- 3) Spark 和 Hadoop MR、Storm 的比较和优势
- 4) RDD
- 5) Transformation
- 6) Action
- 7) Spark 计算 PageRank
- 8) Lineage
- 9) Spark 模型简介
- 10) Spark 缓存策略和容错处理
- 11) 宽依赖与窄依赖
- 12) Spark 配置讲解
- 13) Spark 集群搭建
- 14) 集群搭建常见问题解决

- 15) Spark 原理核心组件和常用 RDD
- 16) 数据本地性
- 17) 任务调度
- 18) DAGScheduler
- 19) TaskScheduler
- 20) Spark 源码解读
- 21) 性能调优
- 22) Spark 和 Hadoop2.x 整合 : Spark on Yarn 原理
- 23) Spark Core 核心编程
- 24) RDD 内核架构概览
- 25) RDD 的不同数据来源的创建方式详解
- 26) RDD 的操作算子综述与本质分析 (转换算子、行动算子)
- 27) 常用操作算子的案例实战
- 28) RDD 持久化实战以及 Checkpoint
- 29) RDD 共享变量以及累加器的使用实战
- 30) RDD 简单排序功能 (优化之前 WordCount 程序) 以及二次排序的实战
- 31) Spark 实战 Top N 功能详解
- 32) Spark 任务调度流程整体架构分析详解
- 33) Spark 任务划分流程整体架构分析详解 (宽依赖与窄依赖、DAGScheduler 源码分析)
- 34) Spark 执行任务相关原理以及源码分析 (TaskScheduler、Executor、Task、Shuffle)
- 35) Spark 实战之 PageRank
- 36) 性能优化与调优的分析

Spark SQL

1. Spark RDD 应用 SQL 实战
2. RDD 转化为 DataFrame 数据框的方式详解
3. Spark DataFrame 数据框操作实战
4. 加载和保存数据操作 (load 与 save)
5. JSON 数据源实战案例
6. JDBC 数据源实战案例
7. Hive 数据源实战案例
8. Parquets 数据源实战加载数据、自动分区推断、合并元数据
9. 内置函数的实战案例
10. 开窗函数的实战案例
11. Spark SQL UDF 自定义函数实战
12. Spark SQL UDAF 自定义聚合函数实战
13. Spark SQL 工作原理详解以及 Spark SQL 的源码分析
14. Hive on Spark

Spark Streaming 实时计算

- 1) Spark Streaming 和 Storm 对比讲解

- 2) Spark Streaming 本质原理分析
- 3) Wordcount 程序的实时版本开发
- 4) Spark Streaming 和 Spark Core 里面 context 的不同
- 5) 输入 DStream 和 Receiver 的讲解
- 6) 不同输入源 (Kafka、HDFS) 的 DStream 操作实战
- 7) 基于 DStream 的 window 滑动窗口实战案例
- 8) 基于 DStream 的 updateStateByKey 实战案例
- 9) 基于 DStream 的 transform 实战案例
- 10) DStream 的输出存储操作以及核心函数 foreachRDD 实战
- 11) Spark Streaming 的持久化实战以及 Checkpoint
- 12) 与 Spark SQL 结合使用实战案例
- 13) 架构原理分析与性能优化

Spark MLlib 机器学习

前面课程大家已经掌握第一代机器学习工具 R，而后又学习了第二代机器学习工具 Mahout，这里大家将会学习第三代机器学习工具 MLlib，大家不仅将会了解 MLlib 的组件及其调用，而且会通过 Spark 的项目深入了解 MLlib 的现实使用。通过此部分大家也可以看出课程不仅着眼于现在，更是着眼于大家的未来在行业中的发展。

- 1) 介绍
 - a) Spark MLlib 组件介绍
 - b) 基本数据类型

- 2) 回归算法
 - c) 广义线性模型
 - d) 逻辑回归
- 3) 分类算法
 - e) 朴素贝叶斯
 - f) 决策树
 - g) 随机森林
- 4) 第四章 推荐系统
 - a) 协同过滤
- 5) 第五章 聚类
 - h) Kmeans
 - i) Sparse kmeans
 - j) Kmeans++
 - k) Kmeans II
 - l) Streaming kmeans
 - m) Gaussian Mixture Model

Sparkt GraphX 图计算

- n) 二分图
- o) 概述
- p) 构造图
- q) 属性图

r) PageRank

Tachyon 内存分布式文件系统

- 1) Tachyon 带来的好处以及特性详解
- 2) Tachyon 架构原理分析
- 3) Tachyon 的安装部署实战
- 4) Tachyon 命令行操作实战
- 5) 整合 Spark 以 Tachyon 为输入输出源的实战
- 6) 整合 Spark 以 Tachyon 作为持久化 RDD 的实战

Docker 课程

Docker 是一个开源的应用容器引擎,让开发者可以打包他们的应用以及依赖包到一个可移植的容器中,然后发布到任何流行的 Linux 机器上,也可以实现虚拟化。容器是完全使用沙箱机制,相互之间不会有任何接口(类似 iPhone 的 app)。几乎没有性能开销,可以很容易地在机器和数据中心中运行。最重要的是,他们不依赖于任何语言、框架包括系统。

- 1) 基本介绍
- 2) vm docker 对比
- 3) docker 基本架构介绍
- 4) unfs cgroup namespace
- 5) 进程虚拟化 轻量级虚拟化
- 6) docker 安装

- 7) docker 镜像制作
- 8) docker 常用命令
- 9) docker 镜像迁移
- 10) docker pipework 【i.openvswitch】
- 11) docker weave

虚拟化 KVM

云计算平台，比如 openstack , cloudstack 底层采用的技术都是虚拟化方案，现在以 kvm 市场占有率最高，我们要深入的去学习这些原生的虚拟化，才能深入了解和架构 openstack 这样的云计算的平台，也才更有能力去开发自己的云计算平台。

- 1) 虚拟化介绍，虚拟化适用场景等等
- 2) Qemu Libvirt & KVM
- 3) 安装 KVM, Qemu, Libvirt
- 4) QEMU-KVM: 安装第一个能上网的虚拟机
- 5) Kvm 虚拟机 nat,网桥基本原理
- 6) kvm 虚拟机克隆
- 7) kvm 虚拟机 vnc 配置
- 8) kvm 虚拟机扩展磁盘空间
- 9) Kvm 快照
- 10) Kvm 迁移
- 11) Java,python,c 语言编程控制 kvm
- 12) 构建自己的虚拟云平台

云平台 OpenStack

在实战中学习，课程绝不纸上谈兵，通过搭建和调试一个真实的 openstack 平台，深入浅出，详细讲解 openstack 的各个组成模块：keystone, glance, nova, cinder, neutron, horizon。课程中遇到的各种实际问题，不仅演示了如何解决，更是要教会大家学会去找到解决问题的方法。难点问题全面讲解。在云计算的各种技术当中，网络部分是最难，也是最复杂多样的。课程中针对虚拟网络进行了详细的讲解，包括基本原理，以及实际环境搭建，问题的跟踪和解决。讲师拥有丰富的移动集团工作经验，负责云平台的各方面工作，讲课内容完全贴近企业需求，绝不纸上谈兵。

- 1) openstack 介绍和模块基本原理分析
- 2) openstack 多节点安装部署【a.采用 centos6.x 系统】
- 3) Keystone 基本原理
- 4) glance
- 5) Cinder
- 6) Swift
- 7) Neutron
- 8) Openstack api 二次开发

大型企业项目实战

项目一某大型电商日志分析和订单管理

架构技术 :CDH,flume、sqoop、HDFS,Mapreduce ,hive ,hbase ,zookeeper ,mysql ,oozie 等等)。在实战中学习，技术点非常多，怎么样实际运用这些点是在自学过程中

体验不到的。电商日志分析包括：pv、uv，跳出率，二跳率、广告转化率等。订单模块有：商家排名，历史订单查询，订单报表统计等。

- a) Web 项目和云计算项目的整合
- b) Flume 通过 avro 实时收集 web 项目中的日志
- c) 数据的 ETL
- d) Hive 批量 sql 执行
- e) Hive 自定义函数
- f) Hive 和 hbase 整合。
- g) Hbase 数据支持 sql 查询分析
- h) Mapreduce 数据挖掘
- i) Hbase dao 处理
- j) Sqoop 在项目中的使用。
- k) Mapreduce 定时调用和监控
- l) Flume 通过 avro 实时收集 web 项目中的日志
- m) Hive 批量 sql 执行
- n) Hive 自定义函数
- o) HBase 数据支持 sql 查询分析
- p) MapReduce 数据挖掘
- q) Mapreduce 定时调用和监控
- r) IP 流量分析
- s) 会话分析
- t) 浏览器 pv 分析

- u) 地域信息分析
- v) 订单分析
- w) 事件分析
- x) MR,hive,impala 任务 oozie 部署

项目二某运营商基站保障平台

一个市级运营商公司,每天的产生海量话务数据(一线城市更高),通过大数实时分析,监控每个基站的掉话率,基站通话总数,基站掉话总数,基站告警,3g/4g 上网流量实时监控。对以上维度进行实时分析以达到对基站工作情况的监控。

技术架构: Storm+hbase+kafka+flume+echarts

- a) flume 实时采集日志
- b) kafka 缓冲队列
- c) storm 实时处理
- d) Hbase dao 存储处理结果
- e) 前端 Web 实时展示报表

项目三微博营销数据挖掘项目

使用数据来自微博平台,项目目标通过机器学习所学知识挖掘目标客户群体,找到代言人进行微博营销广告投放。

- a) 分布式平台 Hadoop, MapReduce
- b) 数据采集 Flume

- c) 数据清洗 ETL
- d) 数据库 Hbase , Redis
- e) 机器学习 Mahout

项目四智慧城市大数据分析项目

城市中每时每刻都会产生海量数据，应用数据挖掘、机器学习和可视化技术，分析出的数据可以改进城市规划，缓解交通拥堵，抓捕罪犯。项目会使用真实的数据。

涉及到所学知识如下：

- a) 实时流处理 Kafka , Spark Streaming
- b) 分布式运算 Hadoop , Spark
- c) 数据库 Hbase , Redis
- d) 机器学习 Spark Mllib
- e) 前台 web 展示数据 Struts2 , echart

项目五手机软件推荐系统项目

使用数据来自某互联网平台手机助手，项目目标通过机器学习所学知识挖掘平台手机用户喜好，给用户准确推荐手机软件，类似 360 手机助手、华为手机助手、百度手机助手推荐功能。

- a) 分布式平台 Hadoop , Spark
- b) 数据清洗 Hive
- c) 数据分析 R RStudio
- d) 推荐服务 Dubbox

- e) 规则过滤 Drools
- f) 机器学习 MLlib

项目六网络流量异常检测项目

项目目标通过机器学习所学知识检测出异常，包括检测欺诈，网络攻击，服务器传感器故障灯（本项目用户现在热门的电商网站的流量分析检测，比如京东，天猫，淘宝等）。项目功能应用于各大互联网平台中，各大互联网平台均需要网络安全予以重视和预防以及检测。

- a) 数据存储 Hadoop
- b) 数据准备 Spark
- c) 数据分析 R RStudio
- d) 机器学习 MLlib
- e) 数据评价及调参

毕业学员就业薪资

P.S. 由于现在公司内部同事之间薪资待遇都是保密的，避免就业信息给已就业学员带来不必要的麻烦，公布的就业信息会有部分信息暂时隐藏。

薪资为北上广深一线城市薪资，非一线城市不计入薪资。

下列表格内待遇没特殊说明都有五险一金，这是现在公司基本福利。有多家公司是说明当时这位同学拿了好几家公司的 Offer。

2015 年部分学员就业薪资一览表格：

姓名	性别	入学付款方式	学历	拿到offer和相应待遇
庞	男	一次性	本科	北京通软件股份有限公司 16k 北京东科技股份有限公司-电信事业部 12k 北京东科技股份有限公司-联通事业部 15k 北京东科技股份有限公司 16k(入职)
李	男	一次性	本科	市场研究股份有限公司 15k+1300房补+7险1金
赵	男	一次性	本科	科技(北京)有限公司, 月薪30k
鲁	男	一次性	本科	北京华科技发展有限公司12k
王	男	一次性	本科	面了6家, 拿了两个offer, 现在在中 年薪18W
王	男	分期	本科	北京文库信息有限公司 offer 17K 北京大电子有限公司offer 14K offer 11k 还有一个也是 11k忘记叫什么了 面试的公司不多大概7家, 我最后选择的是的offer。
祁	男	一次性	专科	一家是文 给我12k, 一家是 给我14k。 现在入职的这家是 14k吧
朱	男	一次性	本科	科技14k
刘	男	一次性	本科	面了两家: 两个offer; 一个是做在线 的, 我没去税前14K; 一个是我现在的公司: 有限公司: 税后16K 现在入职的公司: 有限公司: 税后16K
吴	男	分期	本科	第一周拿到offer: 科技 工资12k 第二周拿到offer: 金融 工资15k
甘	男	分期	专科	待遇13k
魏	男	分期	本科	北京通软件股份有限公司 14k
谢	男	分期	专科	面试拿到2个offer 入职公司名: 北京有限公司 薪资: 17000
王	男	一次性	本科	有限公司17k
姓名	性别	入学付款方式	学历	拿到offer和相应待遇
张	男	分期	本科	老师我收到 科技股份有限公司(入职)和 北京信息技术有限公司的offer (第一家14K, 13薪, 年终奖1-3月工资; 第二家, 15K, 12薪, 没有年终奖)
姜	男	分期	专科	市场研究股份有限公司 基本薪资-13000
白	男	分期	本科	17k(大)
张	男	分期	专科	北京移动科技股份有限公司 20k
位	男	分期	本科	上海科技14k
张	男	分期	本科	14k
洪	男	分期	本科	游戏18k*13+200(车补)+午餐+晚餐
李	男	分期	本科	13k*13
张	男	一次性	本科	北京世纪信息技术有限公司 14k
李	男	分期	本科	软件12k
吴	男	一次性	专科	软件14k 17k
刘	男	一次性	本科	科技(入职公司) 17k 北京世纪信息技术有限公司 16.5k
于	男	一次性	博士	北京环保科技有限公司 年薪43万外加期权
匙	男	一次性	本科	中国云计算 薪水是17K
江	男	分期	本科	数据15k 智能 15k
刘	男	分期	专科	北京环保科技有限公司 这个公司给了12k
邹	男	分期	本科	上海云网14k
雷	男	分期	硕士	科技17k*13

姓名	性别	入学付款方式	学历	拿到offer和相应待遇
陈	男	一次性	本科	北京 资讯有限公司 14k
胡	男	分期	硕士	北京 有限公司 17k*13
王	男	分期	本科	上海 电子 18k
王	男	分期	专科	北京E 有限公司 13k
杨	男	分期	本科	街14k
李	男	分期	专科	14k
姚	女	分期	本科	北京 快运股份有限公司，给了14K
杨	男	分期	专科	北京 科技13k*13
肖	男	分期	本科	上海 信息有限公司 16K*14 上海 信息有限公司 18K*15 +期权 上海 技术有限公司 17k offer
申	男	一次性	本科	中国 网 18K 盛 21K 在线 20K
朱	男	分期	本科	教育网 17k
姓名	性别	入学付款方式	学历	拿到offer和相应待遇
刘	男	分期	本科	一共4个offer: 金融16k 15 16 17
刘	男	分期	专科	深圳 有限公司，税后16k
邢	男	分期	本科	成功收到两个offer: (14k)，北京 世纪信息技术有限公司 (15k) offer
王	男	分期	专科	公司: 思 薪资: 16k*13薪
杨	男	一次性	本科	我现在去的青 ，待遇是18k. 平时会加下班
严	男	分期	本科	去的是一个小公司 在望京 拿了三个offer 现在待遇15K 叫海
荣	男	分期	专科	金融 27K
郭	男	一次性	本科	老师，现在去 去了！就拿到两个一个 一个 总共面试了四家 待遇17k
宋	男	分期	本科	13k 2个offer 还有一个中铁 没去
房	男	一次性	专科	跟中 签的 15K 外包进的电信
纪	男	一次性	本科	最后是去的 中国 有限公司 15K 一共3个offer
孙	男	一次性	本科	月薪 1.4W， 每年13月薪水。 试用期有五险一金 就业单位: 中 信
石	男	分期	本科	中国 云计算公司 14K
刘	男	分期	专科	北京 世纪信息技术有限公司 22K

姓名	性别	入学付款方式	学历	拿到offer和相应待遇
师	男	分期	本科	信息股份有限公司 12k试用期
侯	男	分期	本科	北京恒 股份有限公司，15000
吴	男	分期	专科	科技 试用期11k 转正13k
李	男	分期	本科	科技(北京)有限公司 试用期10000 转正13k
王	男	分期	专科	信息股份有限公司 13k
顾	男	一次性	本科	资讯有限公司 16k
周	男	分期	本科	文 16k
李	男	分期	专科	科技 18k
谭	女	一次性	硕士	科技有限公司18k 13个月
秦	男	分期	硕士	中国 信息中心：16k
林	男	一次性	专科	深圳 13K 15薪，北京 15k 13薪
李	男	一次性	本科	仁 13k 13个月
刘	男	一次性	本科	招 科技有限公司 15k
李	男	分期	本科	在线 16k 13个月
代	男	一次性	本科	在线电子商务有限公司 34k
张	男	一次性	本科	华13k*15, 传媒 14*13
魏	男	分期	本科	北京 科技有限公司，15K绩效。 北京 科技信息公司，12K绩效各种福利补贴 北京仁 信息科技公司，17Kx14 (入职公司) 北京 互动信息科技有限公司，17K
费	男	分期	专科	贯资讯 12k
姓名	性别	入学付款方式	学历	拿到offer和相应待遇
李	男	分期	本科	中 海基本工资就12k
杨	男	分期	专科	中 税前18K

姓名	性别	入学付款方式	学历	拿到offer和相应待遇
张	男	一次性	本科	北京科技发展有限公司 试用期1.5万 转正1万8
李	男	分期	本科	北京软件科技有限公司 试用期13000 转正16000
张	男	分期	本科	北京科技发展有限公司 15500+850补助
范	男	分期	硕士	公司：通信有限公司 试用期待遇：15k；转正待遇：17k
蔡	男	一次性	本科	北京科技发展有限公司 试用期17000 转正21000
赵	男	分期	本科	大数据科技有限公司 试用期 12000
徐	男	一次性	本科	北京科技 试用期10000 转正12000
陈	男	一次性	本科	北京股份有限公司 16k
王	男	一次性	本科	科技股份有限公司 15k+1300房补+7险1金
胡	男	一次性	本科	科技（北京）有限公司，月薪25k
刘	男	一次性	本科	北京科技发展有限公司12k
王	男	一次性	本科	北京云游科技有限公司 年薪18W
于	男	分期	本科	北京信息科技有限公司 offer 17K
郭	男	一次性	专科	北京电力，给我14k。
朱	男	一次性	本科	科技14k
刘	男	一次性	本科	科技有限公司：税后16K
李	男	分期	本科	金融 工资15k
姓名	性别	入学付款方式	学历	拿到offer和相应待遇
李	女	一次性	本科	科技有限责任公司 试用期2月，转正后交五险一金，没有具体问工资是试用还是转正薪资，我要了14000，就给了14000，每月500电脑补贴，200交通补贴，300餐补，税后14K
刘	男	一次性	本科	深圳网络技术有限公司 试用期12000转正14000
侯	男	一次性	本科	中国云计算公司 试用期17000 转正20000
赵	男	分期	硕士	北京系统技术有限公司（北京市海淀区花园路时代大厦）试用期14000 转正17000
梁	男	一次性	专科	北京信息技术有限公司 15000 北京网 17000（入职公司）
刘	女	一次性	硕士	北京科技有限公司14000

就业前景调查

hadoop • 工资收入水平



<input type="checkbox"/> 高级hadoop工程师 (JAVA组)	安徽钰诚控股集团股份有限公司北京分公司	15001-20000	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> 大数据开发工程师 (Hadoop)	北京北森测评技术有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> Hadoop\大数据开发工程师	新宇联安信息科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Hadoop运维	北京汉朔科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> hadoop工程师(ltsy)	北京东方国信科技股份有限公司	8001-10000	北京	09-17	▼
<input type="checkbox"/> Hadoop开发工程师 (LTSY)	北京东方国信科技股份有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Hadoop专家OHA	北京联动天地科技有限公司	20001-30000	北京	09-17	▼
<input type="checkbox"/> Hadoop工程师	北京源信创想信息技术有限公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> 高级 Hadoop 研发工程师	北京智明星通科技有限公司	20001-30000	北京	09-17	▼
<input type="checkbox"/> Hadoop资深研发工程师	北京中科天玑科技有限公司	10001-15000	北京 - 海淀区	09-17	▼
<input type="checkbox"/> Hadoop开发工程师	黑龙江邮政易通信息网络有限公司	面议	北京 - 丰台区	09-17	▼
<input type="checkbox"/> Hadoop开发工程师	北京网德天下信息技术有限公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> Hadoop HBase 工程师	江苏中禾网络科技有限公司	10001-15000	北京 - 西城区	09-17	▼
<input type="checkbox"/> hadoop数据开发工程师	魔秀科技(北京)有限公司	15001-20000	北京	09-17	▼
<input type="checkbox"/> Hadoop研发工程师-大数据	北京百分点信息科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Hadoop工程师 (北京分公司办公)	上海朗趣软件科技有限公司	15001-20000	北京 - 东城区	09-17	▼
<input type="checkbox"/> Hadoop高级开发工程师	酷我音乐	面议	北京	09-17	▼
<input type="checkbox"/> 大数据平台开发工程师 (Hadoop)	北京筑巢新游网络技术有限公司	15001-20000	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> Hadoop开发工程师	马上消费金融股份有限公司	15001-20000	北京	09-17	▼

solr • 工资收入水平



<input type="checkbox"/> Solr搜索高级开发工程师	北京风行在线技术有限公司	15001-20000	北京 - 海淀区	09-17	▼
<input type="checkbox"/> 高级搜索开发工程师 (Java/Solr/Elasticsearch)	北京第三石科技发展有限公司	15001-20000	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> Java (Lucene/solr) 开发工程师	北京捷通无限科技有限公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> java软件工程师	中网互联(北京)资讯有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Java工程师	微课网	15001-20000	北京	09-17	▼
<input type="checkbox"/> 高级软件工程师	北京华赢凯来投资担保有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 高级系统开发工程师 (java)	好大夫在线	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> 高级java开发工程师	北京立思辰新技术有限公司	面议	北京 - 东城区	09-17	▼
<input type="checkbox"/> PHP技术主管	北京联云格科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 高级Java开发工程师	北京合享创新信息科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 平台开发部——爬虫搜索软件工程师	药渡经纬信息科技(北京)有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 高级JAVA工程师-SaaS平台	美团网	面议	北京	09-17	▼
<input type="checkbox"/> JAVA高级软件工程师	北京千里马网信科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 技术经理	上海艾融软件股份有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 技术经理	上海艾融软件股份有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 资深/高级PHP开发工程师	北京优叁企业管理顾问有限公司	15001-20000	北京	09-17	▲

地点：北京 公司性质：民营 公司规模：20-99人 经验：1-3年 学历：不限 职位月薪：15001-20000元/月

... solr等开源框架； 4、精通Java、C++语言，具有设计模式思想和代码开发经验的优先； 5、对OOP、MVC 有较为深入的认识、能熟练使用常用的设计模式的优先； 6、熟悉Javascript、XHTML、CSS、DIV 等前端技术，能熟练使用至少一种Javascr...

申请职位

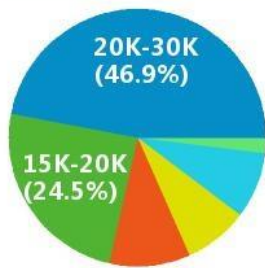
☆ 收藏职位

hbase • 工资收入水平

最新招聘

工资收入

就业形势

小组^{NEW}

区间	占比
2K-3K	2%
6K-8K	8.2%
10K-15K	10.2%
15K-20K	24.5%
20K-30K	46.9%
30K-50K	8.2%

¥ 18670

近1年 49 份样本 / 可信度：高

你觉得该统计数据准确吗？

偏低

靠谱

偏高

0

0

0

收起 >>

<input type="checkbox"/> Hadoop HBase 工程师	江苏中禾网络科技有限公司	10001-15000	北京 - 西城区	09-17	▾
<input type="checkbox"/> 大数据工程师 (Hbase)	北京亿玛在线科技有限公司	20001-30000	北京	09-17	▾
<input type="checkbox"/> 广告数据平台高级工程师 (Java、Hadoop、Hive、HBase)	视界环球(北京)文化传媒有限公司	面议	北京 - 朝阳区	09-17	▾
<input type="checkbox"/> Hbase工程师	全时云商务服务股份有限公司	面议	北京	09-17	▾
<input type="checkbox"/> hbase系统架构工程师	全时云商务服务股份有限公司	面议	北京	09-17	▾
<input type="checkbox"/> HBase工程师	北京陌陌信息技术有限公司	面议	北京	09-17	▾
<input type="checkbox"/> Hbase工程师	北京亿玛在线科技有限公司	20001-30000	北京	09-17	▾
<input type="checkbox"/> PHP开发工程师	北京市燕医堂医疗科技有限公司	10001-15000	北京	09-17	▾
<input type="checkbox"/> PHP开发工程师	北京市燕医堂医疗科技有限公司	10001-15000	北京 - 朝阳区	09-17	▾
<input type="checkbox"/> 平台开发工程师	生命奇点(北京)科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 大数据高级开发工程师	互爱(北京)科技有限公司	15001-20000	北京	09-17	▾
<input type="checkbox"/> 急招一爬虫工程师	北京中搜网络技术股份有限公司	15001-20000	北京 - 朝阳区	09-17	▾
<input type="checkbox"/> C++高级开发工程师 (云计算)	北京捷通华声语音技术有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 推荐工程师	聚美优品	15001-20000	北京	09-17	▴

地点：北京 公司性质：民营 公司规模：1000-9999人 经验：1-3年 学历：本科 职位月薪：15001-20000元/月

... HBASE开发经验者优先；

申请职位

☆ 收藏职位

hive • 工资收入水平

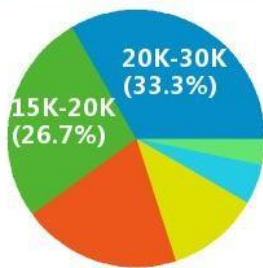
最新招聘

工资收入

就业形势

小组

NEW



区间	占比
2K-3K	3.3%
8K-10K	5%
10K-15K	20%
15K-20K	26.7%
20K-30K	33.3%
30K-50K	11.7%

收起 >>

¥ 18910

近1年 60 份样本 / 可信度：高

你觉得该统计数据准确吗？

偏低

0

靠谱

0

偏高

0

☐ 广告数据平台高级工程师 (Java、Hadoop、Hive、HBase)

视界环球(北京)文化传媒有限公司

面议

北京 - 朝阳区

09-17

☐ 大数据高级开发工程师

互爱(北京)科技有限公司

15001-20000

北京

09-17



地点：北京 公司性质：民营 公司规模：1000-9999人 学历：大专 职位月薪：15001-20000元/月

...Hive经验优先； 6). 熟悉Linux或Unix平台； 7). 熟悉JAVA编程； 8). 性格积极乐观，有良好的沟通能力，抗压能力，有强烈的学习/技术研究能力和良好的团队精神。

申请职位

☆ 收藏职位

☐ 急招一爬虫工程师

北京中搜网络技术股份有限公司

15001-20000

北京 - 朝阳区

09-17



地点：北京 - 朝阳区 公司性质：上市公司 公司规模：1000-9999人 经验：1-3年 学历：本科 职位月薪：15001-20000元/月

...HIVE 等 NoSQL 数据库，并至少精通大数据量的一种关系型数据库开发； 6、熟悉搜索引擎和个性化相关的机器学习算法相关技术，精通主流分词算法、分类、提取摘要、大规模网页聚类、索引、检索等相关开发经验者优先。

申请职位

☆ 收藏职位

☐ 高级hadoop工程师

北京海誉动想科技有限公司

面议

北京

09-17

☐ GIS架构师

北京智慧图科技发展有限公司

面议

北京 - 海淀区

09-17

☐ 高级软件工程师

全数软件科技(北京)有限公司

面议

北京

09-17

☐ 大数据开发工程师

全数软件科技(北京)有限公司

面议

北京

09-17

☐ 软件工程师

全数软件科技(北京)有限公司

面议

北京

09-17

☐ 大数据高级研发工程师

全数软件科技(北京)有限公司

面议

北京

09-17

☐ Hadoop开发工程师 (DXY-JSCP)

北京东方国信科技股份有限公司

面议

北京

09-17

☐ Hadoop / Spark开发工程师

北京百泉众合数据科技有限公司

面议

北京

09-17



地点：北京 公司性质：股份制企业 公司规模：20-99人 经验：1-3年 学历：本科

工作职责： 1、负责基于Spark的并行计算平台的维护、开发与优化工作 2、Spark全栈的研究和应用 职位要求： 1、非常熟悉Spark、Storm、Hadoop、Hive、Hbase、MapReduce这些常见的大数据框架，有至少1年的大数据经验； 2、熟悉Scala...

申请职位

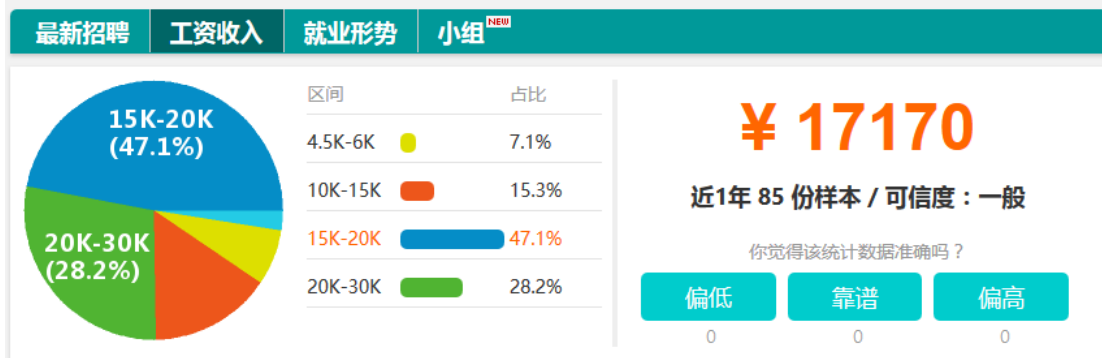
☆ 收藏职位

storm • 工资收入水平



- ☐ 大数据分析经理（数据抓取） 普惠融通科技(北京)有限公司 30001-50000 北京 - 朝阳区 09-17
- 地点：北京 - 朝阳区 公司性质：民营 公司规模：1000-9999人 经验：3-5年 学历：本科 职位月薪：30001-50000元/月
- ... Storm, Spark等大数据技术以及适应场景。有大规模数据处理和日志处理经验的优先，了解NoSQL技术，有实际使用者优先；6、熟练掌握多种统计和挖掘方法，熟练使用SPSS、SAS等相关数据分析软件；7、较强的数据敏感度，逻辑分析能力和文档写作能力。
- [申请职位](#) [收藏职位](#)
- ☐ 数据平台开发工程师 乐视网信息技术(北京)股份有限公司 面议 北京 - 朝阳区 09-17
- ☐ Java故障分析产品架构师 福建星网锐捷网络有限公司 20001-30000 北京 09-17
- 地点：北京 公司性质：合资 公司规模：1000-9999人 学历：大专 职位月薪：20001-30000元/月
- ... storm, zookeeper等分布式运算实践经验优先。素质要求：有志于走技术专家路线的人，对技术和架构设计有特别的热情；有阅读习惯，系统读过《JAVA编程思想》《设计模式》《EFFECTIVE JAVA》《重构》《敏捷开发》等经典书籍；有较新的新技术攻关能力...
- [申请职位](#) [收藏职位](#)
- ☐ 实时数据仓库工程师 乐视网信息技术(北京)股份有限公司 面议 北京 - 朝阳区 09-17
- ☐ 高级大数据开发工程师 乐视网信息技术(北京)股份有限公司 面议 北京 - 朝阳区 09-17
- ☐ 大数据分析系统资深研发工程师 北京中科天机科技有限公司 15001-20000 北京 - 海淀区 09-17
- 地点：北京 - 海淀区 公司性质：国企 公司规模：100-499人 经验：1-3年 学历：本科 职位月薪：15001-20000元/月
- ... Storm、GraphLab等 4. 了解并使用过常见大数据离线分析及流式分析框架 5. 了解MR、BSP、MPI等并行框架 6. 熟悉常用统计模型，有BI及数据仓库使用经验者优先 7. 具备SQL及自然语言处理经验者优先 8. 熟练掌握c/java/scala等开发语言 9...
- [申请职位](#) [收藏职位](#)
- ☐ 系统架构师/技术专家（Java）-新业务中心 国美电器电子商务公司 30001-50000 北京 09-17
- 地点：北京 公司性质：上市公司 公司规模：1000-9999人 学历：不限 职位月薪：30001-50000元/月
- ... storm大数据处理及高性能架构设计经验优先 10. 有支付交易、商品管理等电商系统研发经验者优先 11. 有丰富的性能调优经验，能快速定位性能瓶颈并给出解决方案优先 12. 有电商及其他知名互联网企业工作经历者优先
- [申请职位](#) [收藏职位](#)

spark • 工资收入水平



<input type="checkbox"/> Hadoop / Spark开发工程师	北京百泉众合数据科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> Hadoop/Spark大数据工程师	北京维盛网域科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> Elasticsearch/Spark开发工程师	北京翰思安信科技有限公司	15001-20000	北京	09-17	▾
<input type="checkbox"/> 大数据工程师(Spark)	北京腾云天下科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> Spark研发工程师	国美电器电子商务公司	面议	北京	09-17	▾
<input type="checkbox"/> Hadoop / Spark开发工程师	北京升鑫网络科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> Spark大数据架构师	北京新致君阳信息技术有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 大数据平台资深架构师-北京-00555	网宿科技股份有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 大数据算法工程师	北京海星融通软件技术有限公司	10001-15000	北京	09-17	▾
<input type="checkbox"/> 平台开发工程师	生命奇点(北京)科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 推荐工程师	聚美优品	15001-20000	北京	09-17	▾
<input type="checkbox"/> 运维总监	北京百泉众合数据科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 大数据技术架构师(DXSY)	北京东方国信科技股份有限公司	面议	北京	09-17	▾
<input type="checkbox"/> Hadoop开发工程师(DXSY-JSCP)	北京东方国信科技股份有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 软件开发工程师	北京中能博泰科技有限公司	面议	北京 - 海淀区	09-17	▾
<input type="checkbox"/> Linux系统运维经理	世界玖玖(北京)电子商务有限公司	面议	北京 - 朝阳区	09-17	▾
<input type="checkbox"/> 高级DBA(YDSY)	北京东方国信科技股份有限公司	10001-15000	北京	09-17	▾
<input type="checkbox"/> 大数据技术支持工程师D1	翱旗创业(北京)科技有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 即时通讯(IM)高级软件工程师	中律科技股份有限公司	面议	北京	09-17	▾
<input type="checkbox"/> 数据开发工程师	中关村在线	10001-15000	北京 - 海淀区	09-17	▾

python • 工资收入水平



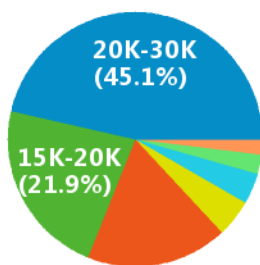
<input type="checkbox"/> python研发工程师	北京观止创想科技有限公司	10001-15000	北京 - 海淀区	09-17	▼
<input type="checkbox"/> Openstack开发工程师 (python)	中经云数据存储科技(北京)有限公司	15001-20000	北京	09-17	▼
<input type="checkbox"/> python开发工程师	北京卡来宝信息科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Python工程师	北京创信软通科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> python开发工程师	乐飞天下信息技术(北京)有限公司	8001-10000	北京	09-17	▼
<input type="checkbox"/> python研发工程师	北京娜迦科技发展有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Python开发工程师	东方合信(北京)投资有限公司	6001-8000	北京	09-17	▼
<input type="checkbox"/> python工程师	北京信善科技有限公司	8001-10000	北京	09-17	▼
<input type="checkbox"/> Python服务器工程师	Glu Mobile Limited	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> Python工程师	北京云信众成软件技术有限公司	10001-15000	北京	09-17	▼
<input type="checkbox"/> python后台开发XBA	北京联动天地科技有限公司	15001-20000	北京	09-17	▼
<input type="checkbox"/> Python开发工程师	北京四海通达网络科技有限公司	10001-15000	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> Python开发工程师	北京艾亿新融资本管理有限公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> python	北京网利科技有限公司	20001-30000	北京	09-17	▼
<input type="checkbox"/> 研发工程师 (C语言、python方向)	北京蓝讯通信科技有限责任公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> Python开发工程师	北京陌陌信息技术有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Python开发工程师	北京捷通华声语音技术有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> Python开发工程师	北京维旺明科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Java/Python工程师-大数据	北京百分点信息科技有限公司	面议	北京	09-17	▼

机器学习·工资收入水平

最新招聘

工资收入

就业形势

小组^{NEW}

区间	占比
2K以下	3.6%
6K-8K	1.8%
8K-10K	2.6%
10K-15K	4.6%
15K-20K	21.9%
20K-30K	45.1%
30K-50K	17.8%

收起 »

¥ 21560

近1年 388 份样本 / 可信度：高

你觉得该统计数据准确吗？

偏低

靠谱

偏高

0

0

0

<input type="checkbox"/> 机器学习工程师	北京光年无限科技有限公司	15001-20000	北京 - 海淀区	09-17	▼
<input type="checkbox"/> 机器学习算法工程师	北京中科天玑科技有限公司	10001-15000	北京 - 海淀区	09-17	▼
<input type="checkbox"/> 机器学习算法工程师	北京数想科技有限公司	20001-30000	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> 机器学习开发工程师	楚楚街	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> 机器学习/数据挖掘工程师	北京奇虎科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 机器学习工程师	北京奇虎科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 机器学习/算法工程师	北京昱美星翰科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 机器学习研究员	北京昱美星翰科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 机器学习算法工程师-qrs13	北京神舟航天软件技术有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 机器学习研究员	北京腾云天下科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 算法工程师/机器学习工程师	北京闪银奇昇科技有限公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> 机器学习(数据挖掘)架构师	风语者软件(北京)有限公司	15001-20000	北京 - 东城区	09-17	▼
<input type="checkbox"/> 研发实习生(机器学习方向)	信翰电子技术(北京)有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> 机器学习算法工程师	信翰电子技术(北京)有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> 机器学习算法工程师	北京金信网银金融信息服务有限公司	10001-15000	北京	09-17	▼
<input type="checkbox"/> 机器学习算法研发工程师	世纪佳缘上海VIP服务中心	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> 机器学习工程师	北京璧合科技股份有限公司	15001-20000	北京	09-17	▼
<input type="checkbox"/> 机器学习工程师	北京璧合科技股份有限公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> 机器学习与机器视觉算法工程师	深圳市智美达科技有限公司(北京)	面议	北京	09-17	▼

OpenStack • 工资收入水平



<input type="checkbox"/> Openstack开发工程师 (python)	中经云数据存储科技(北京)有限公司	15001-20000	北京	09-17	▼
<input type="checkbox"/> OpenStack工程师	天津书生云科技有限公司北京分公司	面议	北京 - 朝阳区	09-17	▼
<input type="checkbox"/> OpenStack软件工程师	北京百泉众合数据科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Openstack研发工程师	北京西普阳光教育科技有限公司	15001-20000	北京 - 海淀区	09-17	▼
<input type="checkbox"/> openstack工程师	北京中联润通信息技术有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> Openstack开发 (存储方向)	北京麓柏科技有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> openstack 开发工程师	北京快网科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Openstack测试实习生	北京奇虎科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> OpenStack测试工程师	北京奇虎科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> openstack开发工程师	北京奇虎科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Openstack云平台工程师	北京奇虎科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> linux web开 (openstack云平台)	北京盛华合创科技有限公司	面议	北京	09-17	▼
<input type="checkbox"/> Openstack研发工程师	TCL通讯科技控股有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> openstack高级开发	北京金山云网络技术有限公司	面议	北京	09-17	▼
<input type="checkbox"/> openstack开发工程师	北京金山云网络技术有限公司	面议	北京	09-17	▼
<input type="checkbox"/> openstack开发工程师	亿达信息技术有限公司	面议	北京	09-17	▼
<input type="checkbox"/> 系统工程师 (Openstack, Linux, VM)	甲骨文(中国)软件系统有限公司	面议	北京	09-17	▼
<input type="checkbox"/> OpenStack架构师	共致开源(北京)信息科技有限公司	面议	北京 - 海淀区	09-17	▼
<input type="checkbox"/> Openstack 开发工程师	北京众智运帷科技有限公司	面议	北京	09-17	▼

当前教室一览







联系方式

北京尚学堂官网: <http://www.bjsxt.com/html/cloud/>

QQ 讨论群: 172599077,156927834