## 大数据基础班课程大纲

[基础班学费价格详情（享受优惠价的条件是什么？）](http://www.itcast.cn/news/20170103/17414023367.shtml?kc" \t "http://www.itcast.cn/course/_blank)

|  |  |  |  |
| --- | --- | --- | --- |
| **大数据基础班课程大纲** | | | |
| **所处阶段** | **主讲内容** | **技术要点** | **学习目标** |
| **第一阶段： Java语言编程基础** | 计算机基础 | DOS常用命令、Java概述、JDK环境安装配置、环境变量配置、Java程序入门 | **核心能力培养：** 能够掌握DOS系统常用基本命令； 熟练使用eclipse编写java代码； 熟练使用java语言的常用对象； 使用java编写单机应用程序； 掌握面向对象编程思想，为以后深入学习JavaEE就业课程打下坚实的基础。  **具备能力及市场价值：** 能够完成B/S结构网站开发，具备了真实环境的项目部署能力，能够完成中小型企业管理系统等传统项目的开发。  **市场价值：** 继续努力学习。 |
| 编程基础 | 常量与变量、数据类型、运算符、流程控制语句、方法、数组 |
| 面向对象 | 面向对象思想、类与对象、成员变量和局部变量、封装、 this关键字、构造方法 |
| 常用类 | Object类、Scanner类、Random类、String、StringBuilder类 |
| 集合操作 | 集合概述、集合特点、ArrayList集合 |
| IO操作 | 字符输入流、字符输出流、字符缓冲输入流、字符缓冲输出流、 复制文件、集合与文件中数据相互读写。 |

## 大数据就业班课程大纲

|  |  |  |  |
| --- | --- | --- | --- |
| **大数据就业班课程大纲** | | | |
| **所处阶段** | **主讲内容** | **技术要点** | **学习目标** |
| **第二阶段： JavaWeb核心** | 前端 | HTML、CSS、JavaSript、BootStrap | **核心能力培养：** 运用常用的网页开发技术设计网页； 掌握WEB系统交互原理； 掌握JavaWeb开发核心技术； 运用JavaWeb核心技术完成简单功能实现； 掌握JavaWeb高级技术，创建更好的Web应用程序； 具备B/S结构软件开发能力。  **具备能力及市场价值：** 能够完成B/S结构网站开发，具备了真实环境的项目部署能力； 能够完成中小型企业管理系统等传统项目的开发。  **市场价值：** 继续努力学习。 |
| 数据库 | MySQL数据库、MySQL单表操作、MySQL多表操作、MySQL事物、 MySQL存储引擎、JDBC、JDBCDataSource |
| Web核心 | Tomcat、Http协议、servlet入门、Rquest、Response、JSP、MVC、 Cookie、Session、JSP、ETL、JSTL、Filter、listener |
| web增强 | Jquery、Ajax、ajax跨域、分页 |
| 基础增强 | 多线程入门、网络编程入门、反射、动态代理、注解 |
| **第三阶段： 网站开发三大框架** | 项目构建及管理 | Maven项目构建、管理、编译、仓库配置，SVN服务器部署、 SVN客户端、自动化部署 | **核心能力培养：** 掌握SSM框架，使用SSH框架开发出结构清晰、可复用性好、维护方便的Web应用程序； 掌握如何使用Maven管理项目工程； 掌握数据库的相关技术； 掌握系统开发中的性能、可扩展性及维护性的提升； 通过项目实战熟练掌握SSM框架的使用。  **具备能力及市场价值：** 能够使用SSM框架完成传统企业级项目开发，熟悉多种业务流程，丰富项目开发经验。  **市场价值：** 6000-8000元。 |
| 数据库操作框架 | mybatis框架原理、mybatis入门案例、mybatis开发DAO方式、 mybatis输入输出映射、动态sql、spring整合mybatis |
| Spring框架 | applicationContext、xml配置文件编写、IoC思想、DI依赖注入、 使用AspectJ切面编程、JdbcTemplate模板使用、声明式事务管理、SSH整合 |
| SpringMVC框架 | springmvc框架原理、springmvc入门案例、springmvc整合mybatis、 参数绑定、json数据交互、拦截器 |
| CRM项目实战 | 使用springmvc+spring4+mybatis+svn来开发项目、 使用BootStrap进行布局 |
| **第四阶段： 大数据娱乐头条** | Linux服务器 | VMware虚拟机安装、linux常用命令、linux用户权限与网络安全 | **核心能力培养：** 掌握大数据爬虫技术开发； 掌握大数据搜索引擎技术开发； 掌握大数据消息队列使用及调优； 掌握Storm实时数据分析； 掌握分布式系统开发； 掌握互联网用户日志分析。  **具备能力及市场价值：** 能够独立开发爬虫系统； 能够独立开发搜索系统； 能够完成实时数据采集、存储、计算及商业应用。  **市场价值：** 8000-10000元。 |
| linux下应用tomcat、linux下应用MySQL、linux下应用nginx、 nginx负载均衡配置 |
| 分布式爬虫实战 | 大数据娱乐头条-项目整体介绍 |
| 大数据娱乐头条-爬虫基础、Http协议、HttpClient网络请求、 Jsoup网页解析、黑客行为之后台登录 |
| 大数据娱乐头条-Java并发、多线程、阻塞队列、 网易娱乐爬虫开发实战 |
| 大数据娱乐头条-Redis基础、Redis集群、Redis常用API、购物车、 排行榜、Redis持久化 |
| 大数据娱乐头条-分布式爬虫、代理IP、爬虫攻防技术、分布式爬虫开发实战 |
| 分布式搜索 | 大数据娱乐头条-lucene基础、搜索系统原理、Lucene创建索引、 Lucene查询索引、Lucene分页、Lucene高亮 |
| 大数据娱乐头条-solr基础、solr在Linux部署、solr管理界面、 solr创建索引、solr查询索引、solr高亮、solrj客户端管理 |
| 大数据娱乐头条-SSM搜索服务、搜索引擎界面、搜索分页、搜索高亮、 搜索热词联想 |
| 大数据娱乐头条-zookeeper集群、zookeeper原理、dubbox、 搜索服务开发 |
| 大数据娱乐头条-Kafka基础、Kafka集群、生产分发策略、消息不丢失、 存储机制、消费者负责均衡、Kafka配置文件详解 |
| 大数据娱乐头条-爬虫集成Kafka、爬虫创建索引、 FreeMarker热门搜索结果静态化 |
| 大数据娱乐头条-综合部署、nginx负载均衡、solrcloud集群、 solrcloud原理分析 |
| 搜索性能优化 | 大数据娱乐头条-nginx+lua基础、点击流日志收集系统部署 |
| 大数据娱乐头条-Storm基础、Storm架构、Storm编程模型、 Storm实时看板、Storm消息不丢失 |
| 大数据娱乐头条-Storm热词统计，提供实时热词静态化 |
| 大数据娱乐头条-Storm爬虫日志监控项目实战 |
| 大数据娱乐头条-Storm日志分析项目实战 |
| JVM与数据库优化 | JVM虚拟机基础与性能调优 |
| 数据库分析与优化 |
| **第五阶段： 大数据Hadoop实战** | 大数据环境准备 | linux基础、linux的shell编程、大数据环境准备、zookeeper、网络编程概述 | **核心能力培养：** 掌握离线数据收集、数据存储、数据计算、任务调度、数据导入导出、数据报表开发技术； 掌握用户日志分析系统（业务分析、编码实现、调度配置、数据导出、数据可视化）； 掌握数据仓库管理、元数据管理、数据稽查等常见处理技术掌握Hadoop高可用配置及管理。  **具备能力及市场价值：** 能够胜任离线相关工作，包括ETL工程师、任务调度工程师、Hive工程师、数据仓库工程师等。  **市场价值：** 12000-14000元。 |
| Hadoop集群部署 | Hadoop的发展简史、Hadoop的版本介绍、 三个公司对Hadoop版本的支持了解、 Hadoop1.x版本与2.x版本的架构比较、 Apache版本Hadoop三种环境构建、 CDH版本的Hadoop重新编译 |
| HDFS&MapReduce | HDFS的来源、HDFS设计目标、Hadoop的架构图、文件副本机制、 block块存储、HDFS的元数据信息、FSimage以及edits、 ScondaryNN的作用、HDFS的文件写入过程、HDFS的文件读取过程、 HDFS的API操作、HadoopMapReduce设计构思、 MapReduce框架结构、MapReduce编程规范及示例编写、 MapReduce程序运行模式 |
| MapReduce优化 | MapReduce的分区、ReduceTask的数量设置、 MapReduce排序以及序列化、MapReduce计数器、MapReducecombiner、MapReduce上网流量统计、 MapTask运行机制详解、Map任务的并行度、ReduceTask工作机制、reduceTask的并行度、MapReduceshuffle过程、shuffle阶段数据的压缩机制 |
| 自定义及资源调度 | MapReduce实现join、社交粉丝数据分析、mapreduce案例:倒排索引建立、 自定义inputFormat合并小文件、自定义outputFormat、 自定义GroupingComparator求取topN、mapreduce参数优化、 Yarn资源调度 |
| Hive数据仓库 | 数据仓库特征、数据仓库架构、Hive的概念、Hive架构、Hive部署及使用、 HiveDDL、HiveDML、Hive命令行、Hive参数配置、Hive内置函数、 HiveUDF开发、Hive的数据压缩、Hive的文件格式、Hive调优、 Hive语句综合练习 |
| 网站流量日志分析 | 网站流量日志采集、数据分析系统介绍、系统开发架构、Flume实现数据采集、 数据预处理、数据仓库设计、ETL、统计分析、Sqoop结果导出、 工作流调度azkaban、数据可视化 |
| Impala&Hue | impala的介绍、impala安装准备、制作本地yum源、impala的安装、 impala基本使用、impala的java开发、Hue的介绍、Hue的安装、 hue与HDFS集成、hue与yarn集成、配置hue与hive集成、 配置hue与impala的集成、配置hue与mysql的集成 |
| 大数据新技术 | Oozie介绍、Oozie架构、Oozie的执行流程、Oozie组件、Oozie的安装、 hue整合Oozie、hadoopHA、hadoopFederation、CDH整体架构、 CDH环境安装、Kylin、kettle、kudu |
| **第六阶段： 大数据Spark实战** | Scala语言基础 | Scala基础语法、Scala高级特性、Scala的Akka编程实战 | **核心能力培养：** 掌握Spark基础、SparkRDD、SparkSQL、SparkStreaming开发技术； 掌握互联网电商用户画像建模、开发、可视化（业务知识、技术开发、架构）； 掌握数据数据存储及存储（Hbase+Phoenix）。  **具备能力及市场价值：** 能够胜任Spark相关工作，包括ETL工程师、Spark工程师、Hbase工程师、用户画像系统工程师、大数据反欺诈工程师。 目前企业急缺Spark相关人才。  **市场价值：** 15000+元。 |
| Spark基础 | Spark概述、Spark集群安装部署、Spark运行架构、Spark编程模型 |
| SparkRDD | RDD概述、RDD特征、RDD算子操作、RDD依赖、RDD缓存、 Spark任务调度、checkpoint、RDD编程实战 |
| SparkSQL | SparkSQL概述、DataFrame、DataFrame常用操作、DataSet介绍、 SparkSQL整合JDBC、SparkonYarn |
| SparkStreaming | 概述、与Storm的对比、SparkStreaming原理、DStream操作实战、 开窗函数、整合Flume、整合Kafka |
| 项目实战阶段 | 互联网电商用户画像建模、开发、数据导入、 数据存储（Hbase+Phoenix）、可视化 |
| 项目实战阶段 | 大数据反欺诈系统、航空领域反爬虫项目实战、系统架构、防爬规则、 Nginx+lua+kafak整合、高频IP检测、SparkStreaming规则引擎开发、 数据可视化 |
| 新技术 | flink的简介、最值函数aggregate和minBy、maxBy、 去重distict和关联join函数、union合并和rebalance负载均衡、 3种分区方式（hash、range、sort）、source数据源、flink的sink操作、 本地执行和集群执行、广播变量的操作、分布式缓存的使用、 无重叠数据处理、 窗口划分、windowTime、同步hbase、flink的source源、 基于mysql的sink操作、flink的容错、flink对接kafka数据、案例实战 |
| **第七阶段： 大数据机器学习实战** | 机器学习概念入门 | 1.基本概念：属性、属性的度量、属性类型、数据集类型、数据集的特性、训练集、测试集、特征值、监督学习、非监督学习、半监督学习等概念 2.数据的预处理：聚集、抽样、维度规约、特征子集选择、特征创建、离散化和二元化、变量变换 3.模型的评估：模型的过分拟合(过拟合)，欠拟合，评估分类器的性能(交叉验证和自助法)，模型评估方法、损失函数和风险函数、参数优化等，模型复杂度(奥卡姆剃刀)4.机器学习处理的一般流程分析 | **核心能力培养：** 掌握机器学习算法理论基础； 熟悉Python语言基础及数据科学库； 熟悉机器学习应用场景； 掌握Spark机器学习框架、能使用scikit-learn机器学习库结合Python完成全栈机器学习建模； 熟悉深度基本概念、流程、常见算法、能够使用算法解决简单的业务问题（图像识别等）； 掌握用户标签预测全流程； 掌握互联网行业推荐业务开发； 掌握点击率预估使用场景及项目全流程开发。   **具备能力及市场价值：** 能够胜任机器学习、数据挖掘等相关工作，包括推荐算法工程师、数据挖掘工程师、机器学习工程师，填补人工智能领域人才急剧增长缺口。  **市场价值：** 18000-30000元。 |
| 机器学习数学基础 | 初等数学基础、函数求导以及链式求导法则、方向导数、梯度、泰勒级数、 拉格朗日乘子法、线性代数与矩阵、特征值与特征向量、概率分析、 极大似然估计、梯度下降法代码实践、牛顿法代码实战、 矩阵分解实战(SVD,PCA,QR) |
| 机器学习语言基础之Python语言 | 基础数据类型、list/tuple/dict/set、列表推导式、生成器推导式、 lambda函数、控制语句、文件读写、异常处理分析、面向对象编程、 GUI编程、Python基础项目实践 |
| Python数据分析库实战 | Numpy矩阵运算库基础及实战、Scipy数值运算库基础及实战、 Matplotlib绘图库基础及实战、Seaborn绘图库基础及实战、 Pandas数据分析库基础及实战 |
| Spark机器学习库实战 | SparkML和SparkMLLIB区别、Spark机器学习基础、Pipeline管道、 特征抽取(TF-IDF、Word2Vec、CountVectorizer）、特征转换(Tokenizer、 PCA、N-gram、DCT、one-hot、MinMaxScaler、Normalizer、SqlTransformer、VectorAssembler)、特征选择(VectorSlicer、RFormula、 ChiSqSelector) |
| 机器学习算法之用户标签预测项目实战 | 用户画像标签预测实战、KNN、KMeans、决策树算法模型(ID3、C4.5、 Cart树)、集成学习算法(Bagging、随机森林、Adaboost算法、GBDT算法、 XGBOOST算法、LightGBM算法模型)、人才流失模型项目实战 |
| 机器学习算法之推荐系统实战 | 基于记忆的CF实战(Surprise库实战)、基于模型的CF实战(SparkALS实战)、 基于Native-Bayes分类算法实战、基于内容推荐(jieba分词、提取词向量、 文本分类、特征聚类)、关联挖掘算法实战(基于Spark的FP-Growth算法实战)、推荐项目实战 |
| 机器学习算法之CTR点击率预估实战 | 特征工程实战、CTR点击率预估应用场景分析、 逻辑斯特回归算法理论基础推导及项目实战、推荐系统指标分析、 推荐系统架构分析、基于Wideanddeep模型理论及实战(学会读学术Paper) |
| 机器学习算法之深度学习基础及图片分类实战 | 神经网络和深度学习基础、MP神经元模型、感知机模型、BPNN模型实战、 CNN模型实战图像识别、Tensorflow基础、电影评论文本分析、 RNN文本情感分析实战 |
| 机器学习面试必备 | 机器学习算法串讲、机器学习面试题目详解，剖析结合人工智能实际场景、 机器学习或人工智能类岗位核心技能需求、所需知识和技能、主流机器学习工具和框架的使用方法、开放式问题和系统设计问题， 融汇贯通整个课程知识点、大数据和机器学习部分项目衔接 |