

附件

大数据分析师培训课程介绍

大数据分析初级课程（32 课时）			
时间	课程模块	内容	课时
第一天	大数据相关主流竞赛模式介绍	模式的讲解； 平台使用讲解； 知识点演示。	2
	Excel 数据分析	Excel 处理 基本设置 常用快捷键数据分列 筛选 Excel 函数统计分析函数 文本处理函数 数值运算函数 逻辑判断函数 日期计算函数 匹配查找函数 实现数据透视表 实现图表	6
第二天	SQL 高级查询	数据库介绍 创建和管理数据库 创建和管理表 操作表中数据 创建和管理索引 创建和管理视图 创建和使用存储过程和触发器 基于场景下的 SQL 语句查询	4
	数据分析与挖掘	SPSS 简介 数据读取 数据清洗 数据基本分析 统计图 Logistic 回归分析 关联分析 时间序列分析	4

时间	课程模块	内容	课时
第三天	BI 商业智能报表	Power BI 概述 数据整理 数据建模 常用的可视化图表	4
	python 数据分析	数据分析常用模块介绍 Numpy、 Pandas、 Matplotlib	2
	大数据分析师资格赛 样题串讲	使用专用的竞赛系统，结合样题进行讲解	2
第四天	经营管理中常用的业绩评估及考核方法	常用的考核指标统计维度 主要的业绩考核方法	1
	分析数据，形成初步结论	如何快速验证数据的准确性？ 如何避免“伪因果”推理导致的分析结论差错？ 如何避免因统计方法使用不当导致的分析结论差错？ 如何抓准业务发展的决定性因素？ 怎样速读报表，发现业务异动点？ 怎样精读报表，挖掘潜藏信息？	3
	经营分析报告的撰写与呈现	撰写数据分析报告的主要工序 常用的经营分析报告结构框架 用思维导图呈现数据分析报告的逻辑框架和数据要素 用金字塔结构规范分析报告的观点和论据 如何提炼报告段落文字，做到言简意赅，重点突出？ 如何用对数据图表，让图表展示	3
	大数据分析师职业技能证书考试题目串讲	知识点回顾 考试题型说明 模拟题讲解	1
培训后	模拟训练(竞赛实训)	提供 4 小时*6 次模拟练习	24

大数据分析中级课程（40 课时）			
时间	模块	内容	课时
第一天	Hadoop	大数据概论； Linux 服务系统； Hadoop 分布式集群搭建。	4
	分布式文件存储系统 HDFS	HDFS 文件系统； HDFS Shell 操作； HDFS Java API 操作。	
	分布式计算框架 MapReduce	初识 MapReduce； MapReduce 入门案例：WordCount。	
	数据仓库工具 Hive	Hive 基础入门、Hive 查询操作。	4
	分布式数据库 HBase	HBase 快速入门； HBase Shell 操作； HBase Java API 操作。	
第二天	开源的大规模数据处理分析引擎 Spark	Spark 框架概述； Spark Core； Spark SQL； Spark Streaming 大数据处理。	4
	python 基础及爬虫	Python 基本介绍； Python 常用数据类型和结构； 结构控制语句（if/while/for）； 函数（定义函数以及传参）； 文件处理 1 - I/O 操作。 Python 爬虫； 爬虫介绍； 获取网页； 解析网页； 数据爬取； 数据保存。	4

时间	课程模块	内容	课时
第三天	可视化工具 Echarts	Echarts 基本实例； Echarts 应用； 案例实战。	4
	可视化工具 D3.js	D3.js 概述； 选择集与数据； 比例尺和坐标轴； 绘图参数； 动画及交互效果； 图表； 案例实战；	
	企业项目实战	实时项目： 实时应用数据源 Kafka、Spark Streaming 编写流式应用、实时存储 HBase、数据可视化	4
第四天	了解数据建模	什么是数据建模？ 监督学习与非监督学习的异同点	1
	分类问题的经典算法及案例剖析	分类的基本概念 典型的分类问题 分类案例：客户满意度评估	6
	回归问题的经典算法及案例剖析	回归的基本概念 典型的回归问题 回归案例：用户流失预警	
	聚类问题的经典算法及案例剖析	聚类的基本概念 典型的聚类问题 聚类问题：精准广告	
	关联问题的经典算法及案例剖析	关联的基本概念 典型的关联问题 关联：个性化推荐系统	
	大数据分析师职业技能证书考试题目串讲	知识点回顾 考试题型说明 模拟题讲解	1
第五天	模拟训练（竞赛实训）	使用专用的竞赛系统，结合样题进行讲解，并提供 4 小时*6 次模拟练习。	8