

工信部教考中心《大数据技术应用工程师》简介

一、项目简介

大数据技术应用工程师是由工业和信息化部教育与考试中心推出一套专业化，科学化，系统化的人才考核标准，涉及在互联网、零售、金融、电信、医学、旅游、新闻媒体等行业专门从事数据采集、数据分析、机器学习、人工智能并能制作业务报告、提供决策的新型数据分析人才所需要的技能。

“大数据技术应用工程师”是对在数据分析、机器学习、深度学习、人工智能领域中使用大数据技术作为主要方法从业者的全面考核，分为初级，中级，高级三个等级。

二、能力标准

1、大数据技术应用工程师（初级）

需要掌握 Python 编程基础、Python 常用数据分析及处理工具 numpy、pandas、scikit-learn，能以 Python 作为工具，解决基本的数据分析问题。能够理解业务目标，并能将业务目标初步转化为大数据技术与应用问题，能结合具体技术进行初步目标实现。适合政府、金融、电信、零售等行业前端业务及从事市场、管理、财务、供应、咨询等职位的相关人员。

2、大数据技术应用工程师（中级）

一年以上大数据分析应用工作经验，或已获得大数据技术应用工程师（初级）证书。掌握 Python 编程基础；常用数据分析及处理工具 numpy、pandas、matplotlib、scikit-learn；Python 数据可视化、机器学习、常规案例应用，能够将业务目标准确转化为大数据技术与应用目标，能将业务目标拆解成不同任务并找到对应技术实现方法，提升工作价值。适合政府、金融、电信、零售、互联网、电商、医学等行业专门从事大数据分析应用的人员。

3、大数据技术应用工程师（高级）

三年以上大数据分析应用岗位工作经验，或已获得大数据技术应用工程师（中级）证书。掌握 Python 编程基础；Python 常用数据分析及处理工具 numpy、pandas、matplotlib、scikit-learn；数据可视化、自然语言处理、网络爬虫、人工智能应用、机器学习、多行业多领域大数据技术应用、深度学习、典型人工智能应用，能充分理解业务目标，并将业务目标精准转换、拆分为具体技术任务；具有超强数据规划能力，不仅能自主完成各环节任务还能带领数据分析团队完成大型项目，为企业生产赋能。适合政府、金融、电信、零售、互联网、电商、医学等行业数据分析应用资深人员。

三、课程内容

初级：

阶段	课程	知识模块
必备技能	Linux操作系统基础	Linux概述
		Linux系统安装
		Linux基本命令
		Linux Vi编辑器
		Linux Shell编程
	MySQL数据分析基础	MySQL 安装
		可视化工具Navicat安装
		数据库基础
		基础命令
		授权管理
		Navincat工具使用介绍
		数据库的创建和删除
		INSERT语句插入数据
		SELECT查询语句
		WHERE条件
		GROUP BY分组查询
	Scala编程基础	Scala简介
		Scala安装配置
		Scala基础语法

		函数
案例实战	基于MySQL的优惠券特征处理	背景&需求说明
		数据字段说明
		标签列构建
		标签列构建
		表合并&小结
职业技术考核	大数据技术应用工程师（初级） 职业技术考试	

中级：

阶段	课程	知识模块
大数据基础	大数据技术导论	认识大数据
		建立大数据思维
		机器学习简介
		数据可视化概述
大数据采集与存储技术	数据采集	爬虫环境简介
		网页前端基础
		简单静态网页爬取
		常规动态网页爬取
		模拟登陆
	数据存储	Hadoop相关概述
		安装虚拟机与Java
		Hadoop集群部署模式与固定IP
		Hadoop集群配置与启动
		Hadoop安全模式与HDFS操作
		MapReduce操作
		Hive安装配置
		工程创建与数据导入
		数据统计分析
		Hive自定义函数
大数据处理与分析技术	分布式计算	安装配置
		架构及原理
		RDD创建与transformation算子
		键值对RDD算子与action算子
		文件读取与时长统计
		搭建开发环境与SQL配置
		创建DataFrame
		DataFrame API操作
		数据保存与DataSet操作

	大数据实时处理	Storm分布式安装
		架构及相关概念
		开发环境配置
		集群运行与并行度
		Storm集成Hadoop
	可视化分析	seaborn绘图基础
		seaborn图表类型
		交互式绘图
大数据应用	机器学习基础	构建回归模型
		构建决策树模型
		构建KNN与非线性支持向量机
		神经网络与集成算法
		Kmeans聚类算法
		密度聚类
		层次聚类与关联规则
		基于用户的协同过滤推荐
		基于流行度推荐
		时间序列
	深度学习基础	构建线性模型
		手写数字识别
项目实战	电影用户性别预测	案例介绍
		评分数据与用户数据
		统计用户对不同电影类型的访问次数
		数据拆分流程分析与实战
		模型构建与评估
职业技术考核	大数据技术应用工程师（中级） 职业技术考试	

高级：

阶段	课程	知识模块
大数据基础	大数据技术导论	大数据时代下的存储技术
		大数据认知
		机器学习概述
		模型评估
	大数据数学基础	回归、判别与聚类分析
		主成分、因子与典型相关分析
大数据采集与存储技术	数据采集技术	网络爬虫框架
		简介与快速入门
		Agent组件类型
		拦截器、Channel选择器和

		Sink 处理器
		常见的采集配置
		多 Agent 数据采集
		广告系统日志数据采集分析
		广告曝光日志数据采集分析
		用户行为日志数据采集分析
		农产品信息采集与分析
	数据存储技术	数据仓库
		数据库管理系统与数据存储
		查询事务管理与系统结构
		分布式文件系统概述
		环境配置
		Hadoop 集群配置与启动
		Hadoop 安全模式与 HDFS 操作
		非关系型数据库概述
		MongoDB 部署模式与数据分片
		系统部署单节点 MongoDB
		Linux 系统部署 MongoDB 副本集与分片集群
		数据库管理工具
大数据处理与分析技术	数据处理技术	创建数据库用户与 MongoDB 监控
		MongoDB 复制、副本集管理与分片
		聚合框架与聚合管道操作符
		MapReduce 与 PyMongo 简单操作
		DStream 转换与窗口操作
		图创建方式与图缓存管理
		图顶点与边转换
		Flink 概述
		IDEA 配置与运行示例
		商品实时推荐系统
	数据可视化技术	功能介绍与文本数据组件
		图表、多媒体与输入组件
		多页应用与社区云部署
		基于大数据技术的产品良率分析
		京东智能家居市场综合分析
大数据应用实战	机器学习技术与应用	MLlib 机器学习算法库与数据标准化
		回归、分类与聚类模型

		关联规则、协同过滤算法与模型评价
		电影智能推荐
		信用卡高风险客户识别
		网络入侵用户自动识别
	深度学习框架与实践	基于注意力机制的评论者满意度分析
		语音识别中的HMM声学模型
		食品图像数据分类识别
职业技术考核	大数据技术应用工程师（高级） 职业技术考试	

四、报考条件

初级：无要求，皆可报考。

中级：（满足其中一个条件即可）

1.获得初级证书。

2.年满18周岁，具备高中以上学历，工作年限满1年

高级：（满足其中一个条件即可）

1.获得中级证书。

2.年满20周岁以上，工作年限满2年

3.年满20周岁以上，具备大专及以上学历

五、考试方式

考试方式分为线下考试站点或线上考试系统统考两种方式，考试形式为上机答题，闭卷。

考试题型：

初级：120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

中级：120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

高级：120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

六、成绩评分

考试最终成绩满分为100分，成绩80-100分为优秀；成绩60-79分为合格；成绩60分以下为不合格。

七、证书样本

学员经考核合格，由工业和信息化部教育与考试中心颁发大数据技术应用工程师职业技术证书，证书可登录国家工信部教育与考试中心官网查询。

附：证书样本

