

工信部教考中心《大数据采集工程师》简介

一、项目简介

大数据采集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。

“大数据采集工程师”是对在大数据采集、清洗、存储、处理及系统架构设计及大数据技术等前沿科技领域中，使用大数据采集技术作为主要工具和方法进行工作的专业人员的全面技能考核，分为初级，中级，高级三个等级。

二、能力标准

1、大数据采集工程师（初级）

需要掌握 Python 编程基础、用于编写数据采集脚本和进行数据处理，了解基本的数据结构和算法，以优化数据采集和处理过程，掌握SQL语言及MySQL等关系型数据库，便于从数据库中提取和整合数据、网络爬虫知识、离线数据采集和实时数据采集技术，以从不同渠道捕获数据，熟悉Linux系统命令和shell编程，为搭建大数据集群环境做准备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采集与应用问题，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、大数据采集工程师（中级）

一年以上大数据采集应用工作经验，或已获得大数据采集工程师（初级）证书。掌握 Python 编程基础、用于编写数据采集脚本和进行数据处理，了解基本的数据结构和算法，以优化数据采集和处理过程，掌握SQL语言及MySQL等关系型数据库，便于从数据库中提取和整合数据、网络爬虫知识、离线数据采

集和实时数据采集技术，以从不同渠道捕获数据，熟悉Linux系统命令和shell编程，搭建大数据集群环境做准备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采集与应用问题，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。

3、**大数据采集工程师（高级）**

三年以上大数据采集应用岗位工作经验，或已获得大数据采集工程师（中级）证书。掌握 Python 编程基础、用于编写数据采集脚本和进行数据处理，了解基本的数据结构和算法，以优化数据采集和处理过程，掌握SQL语言及MySQL等关系型数据库，便于从数据库中提取和整合数据、网络爬虫知识、离线数据采集和实时数据采集技术，以从不同渠道捕获数据，熟悉Linux系统命令和shell编程，搭建大数据集群环境做准备解决基本的数据分析问题，多行业多领域大数据技采集应用、深度学习、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；具有超强数据采集能力，不仅能自主完成各环节任务还能带领数据采集团队完成大型项目，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

阶段	课程	知识模块
基础知识	数据分析	数据分析概述
		Numpy数组
		Numpy矩阵与读写文件
	数据挖掘	数据挖掘概述
		编程基础
		常用库使用
	数据采集基础	Python爬虫环境与爬虫简介
		网页前端基础
		简单静态网页爬取-生成HTTP请求
		简单静态网页爬取-解析网页

		简单静态网页爬取-爬取并存储数据
大数据采集与存储技术	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
机器学习进阶知识	分类与回归	回归分析
		决策树算法介绍
		泰坦尼克号生还者预测
		人工神经网络介绍
		单样本网络训练
		全样本网络训练
		K 近邻算法
		朴素贝叶斯
		支持向量机算法介绍
		支持向量机代码实现
		市财政收入分析及预测
	聚类分析	聚类分析概述与实现
		密度与层次聚类
		根据刷卡经纬度信息分析乘客上车站点
	推荐算法	基于用户的协同过滤推荐

		基于物品的协同过滤推荐
		基于流行度推荐
		实操案例:新闻用户智能推荐
	时间序列	时间序列原理与实现
		脑卒中发病环境因素分析及干预
		金融服务机构资金流量预测
	集成学习与强化学习	集成学习概述与实现
		强化学习概述与实现
		网络入侵用户自动识别
职业技术考核	大数据采集工程师（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

