

工信部教考中心《大数据工程师》简介

一、项目简介

大数据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、金融领域，大数据工程用于信贷风控、精准营销、股价分析、智能投顾和反欺诈等行业专门从事大数据技术的系统搭建、数据存储处理与分析挖掘、应用解决方案设计，旨在通过处理和分析海量数据，挖掘数据价值、提供决策的新型数据应用人才所需要的技能。

“大数据工程师”是涉及大数据技术的系统搭建、数据存储处理与分析挖掘、应用解决方案设计，着重于构建完整大数据生态系统，涵盖操作系统与框架搭建，存储处理与深度分析挖掘关系型与非关系型数据，能针对特定场景制定大数据应用方案，实现数据价值最大化等从业者的全面考核，分为初级，中级，高级三个等级。

二、能力标准

1、大数据工程师（初级）

需要掌握 Linux、Hive、Python 的基本操作和工具使用。具备 Hadoop 的基础知识，包括 Hadoop 简介、集群安装与部署以及基础操作，能够搭建简单的 Hadoop 集群环境并进行基本的数据管理。具备一定的网页前端基础，能够理解网页结构，从而更好地通过工具进行网页数据的定位和抓取。

能够将所学知识应用到实际项目中，如针对市场营销应用，能够获取商品相关信息（链接、价格、详细信息）和评论数据，并进行预处理和分析，包括细分市场分析、竞争者分析、创新需求分析和营销策略分析等。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、大数据工程师（中级）

一年以上大数据工作经验，或已获得大数据工程师（初级）证书。掌握Linux、Hive、Python的基本操作和工具使用。深入理解Hadoop架构原理，能够优化Hadoop集群性能，包括调整集群参数、处理资源分配、解决常见的集群故障等，确保大规模数据处理的高效性和稳定性。

具备处理海量数据的能力，能够运用分布式计算技术和大数据存储技术，实现数据的高效存储、检索和更新。例如，对大规模日志数据进行实时收集、存储和分析，从中提取有价值的信息，为业务决策提供支持。

在大数据处理过程中，能够快速定位和解决各种技术问题，如系统故障、性能瓶颈、数据质量问题等。具备良好的问题排查能力和解决思路，能够运用系统监控工具、日志分析工具等手段进行问题诊断，并提出有效的解决方案。持续关注大数据技术发展动态，能够将新的技术和方法引入到实际项目中，对现有系统进行性能优化和功能升级。例如，探索新的数据存储格式（如Parquet、ORC）和计算引擎（如Flink）的应用，提高大数据处理效率和系统扩展性。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岗位的人员。

3、大数据工程师（高级）

三年以上大数据岗位工作经验，或已获得大数据工程师（中级）证书。掌握Hive、PySpark、Python、MongoDB的基本操作和工具使用。精通关系型数据（如MySQL）和非关系型数据（如MongoDB）的存储与处理。对于关系型数据库，熟练掌握数据库基础概念，能进行MySQL的基本操作（如数据插入、查询、更新、删除）、联表查询与复杂子查询，编写存储过程和函数以优化数据处理逻辑。对于非关系型数据库，熟练掌握MongoDB的入门实战操作，包括安全认证与管理、复制与分片技术以实现高可用和高性能的数据存储，熟

练进行文档操作、聚合操作，通过Python与MongoDB交互实现数据的灵活读写，合理运用视图与索引提升数据查询效率。具备扎实的数据挖掘基础，理解数据挖掘的流程和常用算法（如分类、聚类、回归等算法），能够根据业务需求选择合适的数据挖掘算法解决实际问题以及大数据分析框架配置能力，熟悉PySpark大数据分析概述，掌握PySpark安装配置，能够在大数据集群环境中部署和运行PySpark应用程序，实现高效的大数据分析任务。

在大数据项目实施过程中，能够迅速识别和分析遇到的各种问题，如数据质量问题、系统性能瓶颈、算法效果不佳等。运用所学知识和经验，提出有效的解决方案，确保项目顺利进行。具备创新思维，能够根据业务需求和技术发展趋势，提出创新性的大数据解决方案和应用思路，为企业创造更大的价值。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

阶段名称	课程名称	模块名称
大数据技术导论	大数据技术导论	大数据概述
		大数据与人工智能
		大数据采集技术
		大数据时代下的存储技术
		大数据分析技术
		机器学习
		模型评估
		数据可视化图形设计指南
大数据系统搭建	基础操作系统搭建	Linux概述
		Linux系统安装
		Linux基本命令
		Linux Vi编辑器
	系统基础架构搭建	Hadoop简介
		Hadoop集群安装与部署
		Hadoop基础操作
	数据存储配置	NoSQL与MongDB简介

		MongoDB单节点与高可用部署
		Hive简介
		Hive安装配置
	大数据分析框架配置	PySpark大数据分析概述
		PySpark安装配置
大数据采集技术	Python网络爬虫	Python爬虫环境与爬虫简介
		网页前端基础
		简单静态网页爬取
		常规动态网页爬取
大数据存储与处理	关系型数据存储与处理	数据库基础
		MySQL基本操作
		MySQL联表查询与子查询
		MySQL存储过程和函数
	非关系型数据存储与处理	MongoDB入门实战
		MongoDB安全认证与管理
		MongoDB复制与分片
		MongoDB文档操作
		MongoDB聚合操作
		Python与MongoDB交互
		MongoDB视图与索引
大数据分析挖掘技术	数据分析与挖掘	数据挖掘基础
		数据分析简介
		数据探索
		数据预处理
	大数据分析框架PySpark	DataFrame操作
		流式数据处理
		机器学习库
大数据应用解决方案	企业内推平台招聘信息采集与分析	页面数据获取
		数据处理与分析
	基于PySpark的招聘网站信息的职业类型划分	数据探索与处理
		模型构建与评估
职业技术考核	大数据工程师（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

