

大数据技术应用--数据归集员

职业能力等级评价标准

(试行稿)

1 项目概况

1.1 项目名称

数据归集员

1.2 项目定义

指从事数据调研、数据采集、数据归类、数据存储等大数据相关基础服务工作的从业人员。

1.3 能力等级

本项目共设三个等级，分别为：初级、中级、高级。

1.4 能力特征

具有一定的学习、表达、计算能力，手指、手臂灵活，动作协调。

1.5 职业能力等级评价要求

1.5.1 申报条件

具备以下条件之一者，可申报初级：

(1) 累计从事相关职业工作1年（含）以上。

(2) 相关专业在校学生。

具备以下条件之一者，可申报中级：

(1) 取得本项目或相关职业初级评价证书（含职业资格证书、职业技能等级证书等）后，累计从事相关职业工作2年（含）以上。

(2) 累计从事相关职业工作4年（含）以上。

(3) 取得相关专业毕业证书。

具备以下条件之一者，可申报高级：

(1) 取得本项目或相关职业中级评价证书（含职业资格证书、职业技能等级证书等）后，累计从事相关职业工作3年（含）以上。

(2) 累计从事相关职业工作6年（含）以上。

(3) 具有高等职业学校、高级技工学校、技师学院相关专业毕业证书，并取得本项目或相关职业中级评价证书（含职业资格证书、职业技能等级证书等）。

(4) 具有大专及以上学历相关专业毕业证书，并取得本项目或相关职业中级评价证书（含职业资格证书、职业技能等级证书等）后，累计从事相关职业工作 1 年（含）以上。

1.5.2 申报条件注释

(1) 满足本项目高级别申报条件可申报本项目低级别。

(2) 相关职业：大数据技术应用相关职业。

(3) 相关专业(根据《普通高等学校高等职业教育专业目录（2015 年）》归类)：

计算机类 6102；

通信类 6103；

经济贸易类 6305（限经济信息管理 630505）；

财务会计类 6303（限会计信息管理 630304）。

1.5.3 评价方式

职业能力等级评价考试包括理论知识、技能操作两个科目，较高等级必要时可增加综合评审。

理论知识考试以笔试为主，可以机考，条件成熟时试点开展网络考试，主要考核从业人员从事本职业应掌握的基本要求和相关知识要求。技能操作考核主要采用现场操作、模拟操作、面试答辩等方式进行，主要考核从业人员从事本职业应具备的技能水平。综合评审通常采取审阅申报材料、技术答辩等方式进行全面评议和审查。理论知识考试和技能操作考核均采用百分制，成绩达到 60 分以上者为合格。

1.5.4 监考人员、考评人员与考生配比

理论知识考试和技能操作考核中的监考人员与考生配比不低于 1：15，且每个考场不少于 2 名监考人员。技能操作考核中考评人员为 3 人以上单数。

1.5.5 评价时间

理论知识考试时间：不少于 90 分钟；技能操作考核时间：初级不少于 120 分钟，中级/高级不少于 180 分钟。

1.5.6 评价场所设备

理论知识考试：在标准教室或标准联网多媒体计算机教室进行。

技能操作考核：在标准联网多媒体计算机教室进行，考生计算机需要按照考核要求安装考试系统客户端及相关应用软件，考试结束后能完成环境的还原。



工业和信息化部教育与考试中心
EDUCATION & EXAMINATION CENTER OF MINISTRY OF INDUSTRY AND INFORMATION TECHNOLOGY

2 基本要求

2.1 职业道德

- (1) 遵纪守法，爱岗敬业。
- (2) 认真严谨，忠于职守。
- (3) 钻研业务，勇于创新。
- (4) 精益求精，工匠精神。

2.2 基础知识

2.2.1 计算机及网络知识

- (1) 计算机组成知识。
- (2) 计算机基础操作知识。
- (3) 计算机常用应用软件的安装及使用方法。
- (4) 计算机网络基础知识。

2.2.2 数据库知识

- (1) 关系型数据库的基本概念。
- (2) SQL 基础。
- (3) Oracle、MySQL、NoSQL 等数据库基础。

2.2.3 云计算、大数据知识

- (1) 云/大数据基本概念。
- (2) 云产品的操作使用。
- (3) 典型大数据平台的操作使用。

2.2.4 主流操作系统基础知识

- (1) Linux 基本操作知识。
- (2) Windows 基本操作知识。
- (3) Unix 基本操作知识。

2.2.5 信息安全基本知识

- (1) 信息安全基础知识。

2.2.6 数据管理相关国家标准

- (1) 数据管理知识体系。

(2) 数据管理能力成熟度评估模型（DCMM）。

(3) 数据管理宏观体系及关键职能操作方法。



工业和信息化部教育与考试中心
EDUCATION & EXAMINATION CENTER OF MINISTRY OF INDUSTRY AND INFORMATION TECHNOLOGY

3 工作要求

本标准初级、中级、高级的技能要求和相关知识要求依次递进，高级别涵盖低级别的要求。

3.1 初级

职业功能	工作内容	技能要求	相关知识要求
1. 数据调研	1.1 关系型数据库种类调研	1.1.1 能查看并记录所有生产系统数据库种类及版本 1.1.2 能记录数据库连接方式的所有信息 1.1.3 能使用数据库客户端工具连接数据库	1.1.1 关系型数据库基本知识 1.1.2 数据库种类及版本查看方法 1.1.3 数据库连接方法 1.1.4 常用数据库客户端管理工具使用
	1.2 元数据调研	1.2.1 能登录数据库查看并记录数据库表结构、字段信息 1.2.2 能通过 SQL 命令查询数据库元数据 1.2.3 能收集字段与字典表的映射关系	1.2.1 数据库表结构查看方法 1.2.2 查询类 SQL 命令的使用 1.2.3 元数据知识 1.2.4 数据库表结构查看方法 1.2.5 数据字典的概念
	1.3 数据字典调研	1.3.1 能导入导出数据库字典表 1.3.2 能核对数据字典	1.3.1 导入导出工具使用方法 1.3.2 导入导出命令使用方法
	1.4 数据规模调研	1.4.1 能获取数据库、表的数量 1.4.2 能统计数据库的记录数量规模、存储量规模	1.4.1 获取数据库表数量的方法 1.4.2 数据库基本信息的统计方法
2. 数据采集准备	2.1 网络环境测试确认	2.1.1 能识别网络拓扑结构图 2.1.2 能使用命令或工具验证网络状态	2.1.1 网络拓扑图的基础知识 2.1.2 常用网络命令的使用方法
	2.2 软硬环境准备	2.2.1 能使用远程连接工具检查服务器配置信息及软件服务安装情况 2.2.2 能安装部署数据采集工具	2.2.1 windows 远程桌面工具使用 2.2.2 SSH 工具使用 2.2.3 数据采集工具软件安装部署方法

	2.3 数据源连接设置	2.3.1 能通过采集工具配置数据源	2.3.1 采集工具的数据源配置方法
3. 数据归类存储	3.1 数据归类	3.1.1 能按照数据系统来源、部门来源方式进行数据分类管理 3.1.2 能按照数据类型（结构化、非结构化、半结构化）方式进行数据分类管理 3.1.3 能按照数据时间维度进行数据分类管理 3.1.4 能根据业务维度方式进行数据分类管理	3.1.1 数据分类管理方法
	3.2 存储管理	3.2.1 能创建维护数据库表 3.2.2 能创建维护分区表 3.2.3 能创建维护文件目录 3.2.4 能检查存储空间使用状况	3.2.1 数据库表创建方法 3.2.2 分区表创建方法 3.2.3 文件目录的维护方法 3.2.4 存储空间检查方法
4. 数据采集实施	4.1 数据源连接管理	4.1.1 能使用 ETL 工具创建维护数据库连接	4.1.1 ETL 工具数据源配置维护方法
	4.2 数据采集任务开发	4.2.1 能使用 ETL 工具进行数据采集转换的开发维护 4.2.2 能使用 ETL 工具进行数据采集作业的开发维护	4.2.1 ETL 软件采集转换组件使用方法 4.2.2 ETL 软件采集作业组件使用方法
	4.3 采集任务调度管理	4.3.1 能完成数据采集任务的调度管理 4.3.2 能解决任务调度失败问题处理	4.3.1 调度任务管理方法 4.3.2 采集任务调度问题分析处理方法
	4.4 数据导入、导出	4.4.1 能使用数据库客户端工具导入导出数据库数据 4.4.2 能使用数据泵命令脚本导入导出数据库表数据 4.4.3 能分析处理常见导入导出异常问题	4.4.1 数据库客户端工具数据导入导出方法 4.4.2 数据泵命令脚本工具数据导入导出方法 4.4.3 导入导出异常问题分析方法

3.2 中级

职业功能	工作内容	技能要求	相关知识要求
	1.1 调研规划制定	1.1.1 能根据业务需求制定调研规划 1.1.2 能制定调研模板	1.1.1 数据调研的方法

1. 数据调研	1.2 调研结果整理分析	1.2.1 能整合调研结果 1.2.2 能统计分析调研结果 1.2.3 能解决调研过程中异常问题处理	1.2.1 数据统计分析知识 1.2.2 调研异常问题处理方法
	1.3 非关系数据库调研	1.3.1 能使用命令或工具连接非关系型数据库、库 1.3.2 能完成数据结构分析、数据统计	1.3.1 非关系型数据库连接工具的使用方法 1.3.2 非关系型数据库的基本操作方法
2. 数据采集准备	2.1 采集工具部署	2.1.1 能检查系统软硬件环境 2.1.2 能安装部署 ETL 工具软件 2.1.3 能使用 ETL 工具软件验证开发采集任务	2.1.1 常用命令使用、网络配置、环境变量配置方法 2.1.2 安装软件的使用方法 2.1.3 ETL 工具软件的使用方法
	2.2 采集脚本开发准备	2.2.1 能开发 Shell 脚本完成任务调度 2.2.2 能开发 SQL 脚本完成数据处理	2.2.1 shell 脚本开发 2.2.2 基础 SQL 脚本开发
3. 数据归类存储	3.1 归类数据审核及问题处理	3.1.1 能够对数据归类进行技术核查	3.1.1 常见数据库客户端工具使用方法 3.1.2 SQL、java、JavaScript 语言运用
	3.2 数据存储平台检查	3.2.1 能检查存储平台的服务是否正常 3.2.2 能评估存储平台是否满足需求	3.2.1 评估存储平台服务能力的方法
4. 数据采集实施	4.1 采集实施计划制定	4.1.1 能制定可行的采集实施计划 4.1.2 能识别实施工作项，估算工作量，分析工作项依赖关系，安排工作计划	4.1.1 数据采集的工作流程及关键工作项内容
	4.2 采集问题分析及处理	4.2.1 能对采集中出现的异常问题进行分析、定位并处理 4.2.2 能总结问题处理方法，形成通用数据采集解决方案	4.2.1 异常问题处理分析方法
	4.3 采集策略制定	4.3.1 能通过数据探查获取数据整体规模、更新频率情况 4.3.2 能评估影响采集策略的关键因素。	4.3.1 数据探查方法 4.3.2 数据采集策略知识

		4.3.3 能够分析评估业务数据采集规模需求、及时性需求 4.3.4 能制订多种采集策略满足业务的需求	
	4.4 采集数据对账检查	4.4.1 能统计数据记录总数，检查字段值的完整性 4.4.2 能使用工具做数据库表核对校验 4.4.3 能对数据对账问题提出整改方案	4.4.1 数据对账检查方法 4.4.2 库表校验方法
	4.5 采集报告编写	4.5.1 能编写采集报告	4.5.1 采集报告的编写方法
5. 数据归集培训	5.1 培训组织管理	5.1.1 能通过 PPT 或图形工具制作培训宣传文稿 5.1.2 能完成现场培训环境的准备 5.1.3 能搭建现场培训演示环境 5.1.4 能对初、中级工现场培训授课和演示操作	5.1.1 常用办公软件使用方法 5.1.2 演示环境的搭建方法 5.1.3 培训规范和流程
	5.2 课件开发	5.2.1 能开发培训课件 5.2.2 能开发考试试题	5.2.1 PPT 制作方法 5.2.2 试题开发规范常识
	5.3 培训质量评估	5.3.1 能制作培训评估模板 5.3.2 能根据评估反馈改进评估模板 5.3.3 能运用统计分析类方法对评估进行分析	5.3.1 评估系统知识或统计分析知识
	5.4 讲师培养	5.4.1 能对初、中级工进行培训指导 5.4.2 能组织新晋讲师的评审	5.4.1 TTT 培训方法

3.3 高级

职业功能	工作内容	技能要求	相关知识要求
1. 数据调研	1.1 调研方案编写	1.1.1 能根据需求确定调研目标、调研范围、调研对象、调研方式 1.1.2 能根据需求制定调研内容的概要 1.1.3 能根据实际情况拟定调研步骤、调研计划	1.1.1 数据调研方法
	1.2 互联网数据调研	1.2.1 能根据需求确定互联网数据采集技术方法	1.2.1 互联网爬虫相关知识

		1.2.2 能制定互联网数据调研步骤及开展方法	1.2.2 调研步骤及开展方法
2. 数据采集准备	2.1 采集技术方案制订	2.1.1 能根据需求做采集技术工具技术选型 2.1.2 能评估软硬件需求、存储方式及存储容量 2.1.3 能制定技术实施方法步骤 2.1.4 能根据技术资源需求制定技术实施计划	2.1.1 常用数据采集技术工具知识 2.1.2 软硬件及存储容量的评估方法 2.1.3 采集实施的技术资源评估方法
	2.2 采集技术研究与开发	2.2.1 能选择可行的采集工具进行研究 2.2.2 能根据业务需求进行使用场景式技术验证 2.2.3 能整合采集工具	2.2.1 常用采集工具框架知识 2.2.2 系统集成开发基本知识
3. 数据归类存储	3.1 数据归类规则制定	3.1.1 能根据业务需求制定合理的归类规则 3.1.2 能制定数据归类规范 3.1.3 能确定数据归类规范技术实施方案	3.1.1 数据归类方法 3.1.2 数据资源管理知识
	3.2 存储策略制定	3.2.1 能根据业务需求进行存储技术选型 3.2.2 能根据资源情况选择合适的存储策略	3.2.1 数据存储平台技术知识 3.2.2 数据存储策略选择方法
4. 数据采集实施	4.1 采集实施方案制定	4.1.1 能确定采集实施内容及实施办法 4.1.2 能制定采集实施步骤及技术要求 4.1.3 能评估预测采集实施风险并制定风险管控办法 4.1.4 能制定采集实施计划	4.1.1 数据采集关键技术知识 4.1.2 数据采集任务评估办法 4.1.3 数据采集开发风险管控知识
	4.2 采集数据对账分析及处理	4.2.1 能对采集数据对账结果进行异常问题定位 4.2.2 能对采集对账分析发现问题并给出处理方案	4.2.1 对账异常问题分析方法 4.2.2 对账异常问题处理方法
	4.3 网络数据爬取	4.3.1 能部署网络爬虫工具软件 4.3.2 能使用网络爬虫工具爬取规律性网络数据 4.3.3 能开发静态网页、动态网页的爬取脚本程序	4.3.1 网络爬虫工具部署 4.3.2 网络爬虫工具使用 4.3.3 网页爬取的基本实现方法及 cookie 的追踪原理
	4.4 API 接口	4.4.1 能使用 ETL 工具组件调用 API 接	4.4.1 API 接口的概念及

	数据采集	口采集数据 4.4.2 能开发 API 接口调用脚本实现数据采集	调用方法 4.4.2 接口调用脚本开发知识
5. 数据归集培训	5.1 讲师培养及管理	5.1.1 能对中级及以下级别人员进行培训管理 5.1.2 能结合能力地图指导讲师的提升路线	5.1.1 教学组织管理技巧 5.1.2 岗位能力模型
	5.2 计划与规范制定	5.2.1 能制定培训管理规范 5.2.2 能根据技术发展路线修改培训管理规范 5.2.3 能制定周期性不同级别的详细培训计划 5.2.4 能持续根据技术发展及规则要求修订计划	5.2.1 培训体系管理知识



工业和信息化部教育与考试中心
 EDUCATION & EXAMINATION CENTER OF MINISTRY OF INDUSTRY AND INFORMATION TECHNOLOGY

4 权重表

4.1 理论知识权重表

项目		技能等级		
		初级	中级	高级
		(%)	(%)	(%)
基本要求	职业道德	5	5	5
	基础知识	25	20	15
相关知识	数据调研	20	25	15
	数据采集准备	20	10	20
	数据归类存储	10	10	15
	数据采集实施	20	25	20
	数据归集培训	—	5	10
合计		100	100	100

4.2 技能要求权重表

项目		技能等级		
		初级	中级	高级
		(%)	(%)	(%)
技能要求	数据调研	30	30	25
	数据采集准备	20	20	20
	数据归类存储	20	15	20
	数据采集实施	30	30	25
	数据归集培训	—	5	10
合计		100	100	100