

国家职业技能标准

职业编码：2-02-10-11

大数据工程技术人员

(2021 年版)

中华人民共和国人力资源和社会保障部
中华人民共和国工业和信息化部

制定

说 明

为贯彻落实《关于深化人才发展体制机制改革的意见》，推动实施人才强国战略，促进专业技术人员提升职业素养、补充新知识新技能，实现人力资源深度开发，推动经济社会全面发展，根据《中华人民共和国劳动法》有关规定，人力资源社会保障部联合工业和信息化部组织有关专家，制定了《大数据工程技术人员国家职业技术技能标准》（以下简称《标准》）。

一、本《标准》以《人力资源社会保障部办公厅 市场监管总局办公厅 统计局办公室关于发布人工智能工程技术人员等职业信息的通知》（人社厅发〔2019〕48号）为依据，按照《国家职业技术技能标准编制技术规程》有关要求，坚持“以职业活动为导向、以专业能力为核心”的指导思想，在充分考虑科技进步、社会经济发展和产业结构变化对大数据工程技术人员专业要求的基础上，以客观反映大数据技术发展水平及其对从业人员的专业能力要求为目标，对大数据工程技术人员的专业活动内容进行规范细致描述，明确了各等级专业技术人员的工作领域、工作内容以及知识水平、专业能力和实践要求。

二、本《标准》为首次制定，依据有关规定将本职业分为初级、中级、高级三个等级，包括职业概况、基本要求、工作要求、权重表和附录等五个方面内容。

三、本《标准》的编制工作在人力资源社会保障部专业技术人员管理司、工业和信息化部人事教育司、中国就业培训技术指导中心的指导下，由工业和信息化部教育与考试中心具体组织实施。

四、本《标准》主要起草单位有：工业和信息化部教育与考试中心、新大陆科技集团技术研究院、北京大学、中国电子技术标准化研究院、西安电子科技大学、杭州数梦工场科技有限公司、腾讯科技（深圳）有限责任公司、新华三技术有限公司。主要起草人有：谭志彬、陈钟、张正球、鲍亮、邓立、吴子颖、吴东亚、陈海、龚玉涵、姚明、张耀华、李冰、念灿华、陈旭、胡国华、张融、刘朝晖、罗盛、肖李晨。

五、本《标准》主要审定人员有：朱小燕、虞文明、穆勇、朱敏、张群。

六、本标准的开发得到了左仁贵、邵昕、刘鹏、张良均、张淑平、宋胜利、李祥林的支持和帮助，在此表示感谢。

职业编码：2-02-10-11

七、本《标准》经人力资源社会保障部、工业和信息化部批准，自颁布之日起施行。

大数据工程技术人员 国家职业技术技能标准

(2021 年版)

1 职业概况

1.1 职业名称

大数据工程技术人员

1.2 职业编码

2-02-10-11

1.3 职业定义

从事大数据采集、清洗、分析、治理、挖掘等技术研究，并加以利用、管理、维护和服务的工程技术人员。

1.4 专业技术等级

本职业共设三个等级，分别为初级、中级、高级。

初级、中级分为三个职业方向：大数据处理、大数据分析、大数据管理。

高级不分职业方向。

1.5 职业环境条件

室内，常温。

1.6 职业能力特征

具有较强的学习能力、计算能力、表达能力及分析、推理和判断能力。

1.7 普通受教育程度

大学专科学历（或高等职业学校毕业）。

1.8 职业培训要求

1.8.1 培训期限

大数据工程技术人员需按照本《标准》的职业要求参加有关课程培训，完成规定学时，取得学时证明。初级 128 标准学时，中级 128 标准学时，高级 160 标准学时。

1.8.2 培训教师

承担初级、中级理论知识或专业能力培训任务的人员，应具有相关职业中级及以上专业技术等级或相关专业中级及以上职称。

承担高级理论知识或专业能力培训任务的人员，应具有相关职业高级专业技术等级或相关专业高级职称。

1.8.3 培训场所设备

理论知识培训在标准教室或线上平台进行；专业能力培训在具有相应软、硬件条件的培训场所进行。

1.9 专业技术考核要求

1.9.1 申报条件

——取得初级培训学时证明，并具备以下条件之一者，可申报初级专业技术等级：

- (1) 取得技术员职称。
- (2) 具备相关专业大学本科及以上学历（含在读的应届毕业生）。
- (3) 具备相关专业大学专科学历，从事本专业技术工作满1年。
- (4) 技工院校毕业生按国家有关规定申报。

——取得中级培训学时证明，并具备以下条件之一者，可申报中级专业技术等级：

- (1) 取得助理工程师职称后，从事本专业技术工作满2年。
- (2) 具备大学本科学历，或学士学位，或大学专科学历，取得初级专业技术等级后，从事本专业技术工作满3年。
- (3) 具备硕士学位或第二学士学位，取得初级专业技术等级后，从事本专业技术工作满1年。
- (4) 具备相关专业博士学位。
- (5) 技工院校毕业生按国家有关规定申报。

——取得高级培训学时证明，并具备以下条件之一者，可申报高级专业技术等级：

- (1) 取得工程师职称后，从事本专业技术工作满3年。
- (2) 具备硕士学位，或第二学士学位，或大学本科学历，或学士学位，取得中级专业技术等级后，从事本专业技术工作满4年。

(3) 具备博士学位，取得中级专业技术等级后，从事本专业技术工作满 1 年。

(4) 技工院校毕业生按国家有关规定申报。

1.9.2 考核方式

分为理论知识考试以及专业能力考核。理论知识考试、专业能力考核均实行百分制，成绩皆达 60 分（含）以上者为合格，考核合格者获得相应专业技术等级证书。

理论知识考试以闭卷笔试、机考等方式为主，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；专业能力考核以开卷实操考试、上机实践等方式为主，主要考核从业人员从事本职业应具备的技术水平。

1.9.3 监考人员、考评人员与考生配比

理论知识考试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；专业能力考核中的考评人员与考生配比不低于 1:5，且考评人员为 3 人（含）以上单数。

1.9.4 考核时间

理论知识考试时间不少于 90min，专业能力考核时间不少于 150min。

1.9.5 考核场所设备

理论知识考试在标准教室进行；专业能力考核在具有相应软、硬件条件的考核场所进行。

2 基本要求

2.1 职业道德

2.1.1 职业道德基本知识

2.1.2 职业守则

- (1) 遵纪守法，爱岗敬业。
- (2) 精益求精，勇于创新。
- (3) 爱护设备，安全操作。
- (4) 遵守规程，执行工艺。
- (5) 认真严谨，忠于职守。

2.2 基础知识

2.2.1 基础理论知识

- (1) 操作系统知识
- (2) 计算机网络知识
- (3) 编程基础知识
- (4) 数据结构与算法知识
- (5) 数据库知识
- (6) 软件工程知识
- (7) 云计算知识
- (8) 大数据知识

2.2.2 技术基础知识

- (1) 大数据系统环境安装、配置和调试知识
- (2) 大数据平台架构知识
- (3) 软件应用开发知识
- (4) 接口开发与功能模块设计知识
- (5) 数据采集与数据预处理知识
- (6) 数据计算与数据指标知识
- (7) 常用数据分析与挖掘方法
- (8) 常用数据报表与可视化技术方法
- (9) 数据管理知识
- (10) 数据运营及技术指导知识

2.2.3 安全知识

- (1) 大数据应用、设备与外部服务组件安全管理知识
- (2) 大数据服务用户身份鉴别与访问控制管理相关知识
- (3) 大数据服务数据活动安全管理知识
- (4) 大数据服务基础设施安全管理知识
- (5) 大数据系统应急响应管理知识

2.2.4 其他相关知识

- (1) 环境保护知识
- (2) 文明生产知识
- (3) 劳动保护知识
- (4) 资料保管保密知识

2.2.5 相关法律、法规知识

- (1) 《中华人民共和国劳动法》相关知识
- (2) 《中华人民共和国安全生产法》相关知识
- (3) 《中华人民共和国网络安全法》相关知识
- (4) 《关于加强网络信息保护的決定》相关知识
- (5) 《关键信息基础设施安全保护条例》相关知识
- (6) 《网络安全等级保护条例》相关知识
- (7) 《数据安全管理办法》相关知识
- (8) 《电信和互联网用户个人信息保护规定》相关知识

3 工作要求

本标准对初级、中级、高级的专业能力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

大数据处理方向的职业功能包括大数据系统搭建、大数据平台管理与运维、大数据技术服务、大数据处理与应用，大数据分析方向的职业功能包括大数据系统搭建、大数据平台管理与运维、大数据技术服务、大数据分析与挖掘，大数据管理方向的职业功能包括大数据平台管理与运维、大数据技术服务、大数据管理。

职业功能	工作内容	专业能力要求	相关知识要求
1. 大数据系统搭建	1.1 硬件系统搭建	1.1.1 能根据施工方案，进行需求沟通并确认设备参数 1.1.2 能参照施工方案，对大数据机架及大型设备进行机房空间规划并部署服务器 1.1.3 能根据组网规划方案，对各服务器或需联通网络设备进行组网布置 1.1.4 能根据现场设施及电力系统，对设备进行上电测试及点亮测试	1.1.1 硬件设备功能知识 1.1.2 服务器组网知识 1.1.3 服务器配置知识
	1.2 软件系统部署	1.2.1 能根据系统部署方案，安装集群环境、硬件环境、虚拟化环境所需的各类系统 1.2.2 能根据软件部署方案使用脚本部署产品或用原生方法安装各类大数据功能组件 1.2.3 能根据节点连接信息配置大数据集群 1.2.4 能根据集群功能对组件进行启动调试	1.2.1 操作系统安装及操作知识 1.2.2 云计算及虚拟化部署知识 1.2.3 大数据组件安装知识 1.2.4 大数据集群配置知识 1.2.5 大数据组件基础操作知识
2. 大数据平台管理与运维	2.1 平台管理	2.1.1 能对现有大数据集群的各类组件进行应用变更或版本更迭 2.1.2 能根据上线计划，按时完成功能上线 2.1.3 能对提交代码的版本进行管理	2.1.1 应用变更管理知识 2.1.2 代码仓库托管知识 2.1.3 功能持续集成知识 2.1.4 代码版本控制知识

	2.2 系统运维	<p>2.2.1 能使用工具对集群的软硬件运行状态进行监控管理</p> <p>2.2.2 能使用工具对大数据集群的各类组件、服务的运行状态进行监控管理</p> <p>2.2.3 能使用工具对作业运行情况和资源占用情况进行监控管理</p> <p>2.2.4 能根据故障报告，参与故障排查，处理故障问题</p> <p>2.2.5 能根据容灾计划，定期备份和迁移关键数据</p>	<p>2.2.1 管理平台操作知识</p> <p>2.2.2 系统环境监控知识</p> <p>2.2.3 常见故障排查知识</p> <p>2.2.4 容灾备份知识</p>
	2.3 安全维护	<p>2.3.1 能根据权限规范，使用工具配置和管理用户权限</p> <p>2.3.2 能针对各类突发的外部攻击或异常事件进行应急处理</p> <p>2.3.3 能对安全系统进行升级和维护</p>	<p>2.3.1 权限管理知识</p> <p>2.3.2 常见异常处理知识</p> <p>2.3.3 网络攻防知识</p>
3. 大数据技术服务	3.1 技术咨询	<p>3.1.1 能根据团队既定计划，收集市场目标信息</p> <p>3.1.2 能配合销售团队制作宣讲材料及解决方案展示材料</p> <p>3.1.3 能配合工程师解决客户技术咨询问题并提供参考信息</p>	<p>3.1.1 大数据行业应用知识</p> <p>3.1.2 大数据技术体系知识</p>
	3.2 解决方案设计	<p>3.2.1 能根据已有的产品解决方案调整输出具体的项目解决方案</p> <p>3.2.2 能进行产品演示和讲解产品特性</p> <p>3.2.3 能根据客户沟通反馈整理需求文档</p> <p>3.2.4 能根据客户需求提供产品咨询及软件架构</p>	<p>3.2.1 大数据基础理论知识</p> <p>3.2.2 大数据行业发展知识</p> <p>3.2.3 大数据软件架构知识</p>
4. 大数据处理与应用	4.1 数据采集	<p>4.1.1 能根据业务需求进行在线、离线数据采集</p> <p>4.1.2 能根据调度策略使用框架设置调度作业</p> <p>4.1.3 能根据存储策略进行数据存储</p>	<p>4.1.1 网络爬虫知识</p> <p>4.1.2 离线数据采集知识</p> <p>4.1.3 实时数据采集知识</p> <p>4.1.4 作业调度知识</p> <p>4.1.5 文件系统数据存储知识</p> <p>4.1.6 关系型数据库知识</p> <p>4.1.7 非关系型数据库知识</p>

	4.2 数据预处理	<p>4.2.1 能根据业务需求对遗漏数据、噪音数据、不一致数据等进行清洗</p> <p>4.2.2 能根据业务需求对多源数据进行整合</p> <p>4.2.3 能根据业务规则对数据格式进行转换</p> <p>4.2.4 能根据数据归一性原则对数据进行单位、数值规约</p>	<p>4.2.1 数据清洗知识</p> <p>4.2.2 数据 ETL 知识</p> <p>4.2.3 数据库基础操作知识</p> <p>4.2.4 SQL 函数知识</p> <p>4.2.5 结构化、半结构化与非结构化知识</p>
	4.3 数据计算	<p>4.3.1 能根据业务需求编写批量、实时数据计算作业</p> <p>4.3.2 能根据数据特征计算数据标签并进行汇总</p> <p>4.3.3 能根据数据指标规则计算关键业务指标</p>	<p>4.3.1 分布式计算知识</p> <p>4.3.2 内存计算知识</p> <p>4.3.3 数据结构封装知识</p> <p>4.3.4 关键业务指标知识</p>
	4.4 数据查询	<p>4.4.1 能根据数据平台构建联机事务分析系统并进行即席查询</p> <p>4.4.2 能根据检索引擎创建索引库并进行数据检索</p> <p>4.4.3 能使用交互式查询工具创建数据接口并提供对外服务接口</p> <p>4.4.4 能使用交互式查询平台制作报表及展示图表</p>	<p>4.4.1 OLAP 系统应用知识</p> <p>4.4.2 数据检索知识</p> <p>4.4.3 交互式计算知识</p> <p>4.4.4 报表制作知识</p>
5. 大数据分析 & 挖掘	5.1 数据预处理	<p>5.1.1 能根据业务需求对遗漏数据、噪音数据、不一致数据等进行清洗</p> <p>5.1.2 能根据业务需求对多源数据进行整合</p> <p>5.1.3 能根据业务规则对数据格式进行转换</p> <p>5.1.4 能根据数据归一性原则对数据进行单位、数值规约</p> <p>5.1.5 能根据数据特征及规律, 选择合适方法对数据进行采样</p>	<p>5.1.1 数据清洗知识</p> <p>5.1.2 数据 ETL 知识</p> <p>5.1.3 数据库基础操作知识</p> <p>5.1.4 SQL 函数知识</p> <p>5.1.5 结构化、半结构化与非结构化知识</p>
	5.2 数据分析	<p>5.2.1 能结合业务场景使用工具对数据集进行概要、描述性统计分析</p> <p>5.2.2 能在描述结果基础上, 对数据进行特征和规律的分析与推测</p> <p>5.2.3 能结合业务场景编写数据统计报告</p>	<p>5.2.1 描述性统计知识</p> <p>5.2.2 统计工具使用知识</p> <p>5.2.3 线性相关及回归相关知识</p>
	5.3 数据挖掘	<p>5.3.1 能评估挖掘需求并使用工具对数据进行特征工程处理</p> <p>5.3.2 能调用常规模型进行模型训练</p>	<p>5.3.1 特征工程知识</p> <p>5.3.2 机器学习基础知识</p> <p>5.3.3 数据挖掘类库应</p>

		<p>5.3.3 能根据合适评价指标对模型进行验证和调参</p> <p>5.3.4 能根据合适评价指标对模型进行测试并输出最终模型的性能评估分数</p>	用知识
	5.4 数据可视化	<p>5.4.1 能选择关键指标抽取数据并进行图表展示</p> <p>5.4.2 能使用可视化组件库进行可视化页面开发并配置交互模式</p> <p>5.4.3 能根据产品反馈对可视化页面及图表进行调整和美化</p>	<p>5.4.1 BI 工具使用知识</p> <p>5.4.2 前端页面开发知识</p> <p>5.4.3 可视化平台配置及使用知识</p>
6. 大数据管理	6.1 数据管理	<p>6.1.1 能对大数据全生命周期进行监控，定义、管理元数据，并提供访问元数据接口服务</p> <p>6.1.2 能对数据质量进行评估及有效管控，校正异常数据和缺失数据</p> <p>6.1.3 能根据安全审计要求，对数据活动的主题、操作及对象等数据相关属性进行审核，确保数据活动过程和相关操作符合安全要求</p>	<p>6.1.1 数据标准知识</p> <p>6.1.2 数据血缘知识</p> <p>6.1.3 数据质量知识</p> <p>6.1.4 数据审计知识</p>
	6.2 数据管理评估	<p>6.2.1 能够编写或者受理评估申请</p> <p>6.2.2 能依据 DCMM 规则和组织需求确定评估范围</p> <p>6.2.3 能协助企业实施 DCMM 成熟度自评</p>	6.2.1 数据管理能力成熟度评估模型知识

3.2 中级

大数据处理方向的职业功能包括大数据应用开发、大数据系统搭建、大数据平台管理与运维、大数据技术服务、大数据处理与应用，大数据分析方向的职业功能包括大数据应用开发、大数据系统搭建、大数据平台管理与运维、大数据技术服务、大数据分析与管理，大数据管理方向的职业功能包括大数据平台管理与运维、大数据技术服务、大数据管理。

职业功能	工作内容	专业能力要求	相关知识要求
1. 大数据应用开发	1.1 应用服务开发	1.1.1 能根据系统所使用的组件接口，开发相应的数据访问层业务代码 1.1.2 能根据大数据存储系统结构，设计对接业务库表结构 1.1.3 能根据产品业务需求，开发相应数据或计算接口 1.1.4 能根据流程图梳理代码逻辑，优化接口及功能模块	1.1.1 大数据组件应用程序接口知识 1.1.2 模型层接口开发知识 1.1.3 服务层接口开发知识
	1.2 系统测试	1.2.1 能根据测试用例，对系统进行接口、功能、压力等黑盒测试并输出缺陷、测试报告 1.2.2 能根据测试用例，对代码进行逻辑、分支等白盒测试并输出缺陷、测试报告 1.2.3 能根据相应测试需求，开发自动化测试脚本	1.2.1 测试技术知识 1.2.2 测试用例设计知识 1.2.3 测试脚本开发知识
2. 大数据系统搭建	2.1 硬件系统搭建	2.1.1 能根据配置需求，规划及选型硬件配置设施 2.1.2 能根据机房环境和配置清单，制定工程实施方案 2.1.3 能根据物理硬件特性，制定组网规划方案 2.1.4 能根据硬件设备条件，进行底层及驱动配置 2.1.5 能根据现场施工情况进行故障处理指导	2.1.1 网络架构和规划知识 2.1.2 服务器底层配置知识
	2.2 软件系统部署	2.2.1 能根据应用需求，规划系统部署方案 2.2.2 能根据性能需求，对各运行系统进行配置和调优 2.2.3 能根据软件部署方案，编写自动化部署脚本，并完成部署 2.2.4 能根据集群组件进行高可用及容灾配置 2.2.5 能根据集群功能对各组件进	2.2.1 自动化脚本开发知识 2.2.2 集群配置知识 2.2.3 集群高可用及容灾知识

		行联通调试	
3. 大数据平台管理与运维	3.1 平台管理	<p>3.1.1 能根据集群功能变更需求,制定组件升级及功能迁移方案</p> <p>3.1.2 能对上线功能进行测试,评估上线可行性,制定上线计划</p> <p>3.1.3 能对大数据平台中的各个组件使用权限进行管理</p>	<p>3.1.1 集群技术知识</p> <p>3.1.2 安全访问控制知识</p>
	3.2 系统运维	<p>3.2.1 能编写脚本对集群软硬件、组件与服务、作业运行情况进行监控及管理操作</p> <p>3.2.2 能对集群的运行性能、读写性能等指标进行调优</p> <p>3.2.3 能根据故障报告,排查故障原因,处理故障问题,并编写自动化运维脚本</p> <p>3.2.4 能制定容灾计划,对异常服务进行故障转移</p>	<p>3.2.1 性能调优知识</p> <p>3.2.2 故障排查知识</p> <p>3.2.3 容灾管理知识</p>
	3.3 安全维护	<p>3.3.1 能根据权限管理规范,编写日志监控脚本进行权限安全管理</p> <p>3.3.2 能根据漏洞报告和测试报告开发相应安全补丁</p> <p>3.3.3 能针对各类突发的外部攻击或异常事件制定应急处理方案</p> <p>3.3.4 能对安全系统进行开发、升级和维护工作</p>	<p>3.3.1 安全补丁开发知识</p> <p>3.3.2 异常处理知识</p> <p>3.3.3 安全工具产品知识</p>
4. 大数据技术服务	4.1 技术咨询	<p>4.1.1 能收集目标市场信息,分析行业需求</p> <p>4.1.2 能配合销售团队进行产品宣讲和解决方案展示</p> <p>4.1.3 能独立解决客户技术咨询问题并提供技术方案</p> <p>4.1.4 能参与项目架构设计并提出参考意见</p>	<p>4.1.1 大数据架构知识</p> <p>4.1.2 大数据技术趋势知识</p>
	4.2 解决方案设计	<p>4.2.1 能根据项目需求,在产品功能和技术架构相关技术文档基础上调整输出项目解决方案</p> <p>4.2.2 能进行产品调研、演示和产品特性讲解</p> <p>4.2.3 能结合业务情况主导或辅助原型项目交付</p> <p>4.2.4 能与业务部门合作挖掘客户需求并输出解决方案</p>	<p>4.2.1 大数据行业背景知识</p> <p>4.2.2 市场营销知识</p> <p>4.2.3 项目管理知识</p>
	4.3 指导与培训	<p>4.3.1 能制定技术员、助理工程师对应的人才培养计划</p>	<p>4.3.1 大数据技术知识</p> <p>4.3.2 技术教学知识</p>

		<p>4.3.2 能制作培训资源</p> <p>4.3.3 能使用培训材料开展对技术员、助理工程师的专业能力培训</p>	
5. 大数据处理与应用	5.1 数据采集	<p>5.1.1 能根据业务需求进行在线、离线数据采集，并配置数据缓存及消息队列</p> <p>5.1.2 能根据业务需求参与制定数据迁移方案</p> <p>5.1.3 能制定调度策略</p> <p>5.1.4 能根据业务特性，制定数据存储策略</p>	<p>5.1.1 信息系统配置知识</p> <p>5.1.2 数据监测与迁移知识</p> <p>5.1.3 数据存储策略知识</p>
	5.2 数据建模	<p>5.2.1 能根据数据建模规范设计数据模型</p> <p>5.2.2 能根据存储系统选型编写并优化数据模型实现脚本</p> <p>5.2.3 能根据业务需求对数据模型进行优化</p>	<p>5.2.1 数据仓库知识</p> <p>5.2.2 层次建模知识</p> <p>5.2.3 维度建模知识</p> <p>5.2.4 读写性能知识</p>
	5.3 数据预处理	<p>5.3.1 能根据数据质量要求制定数据清洗策略及评估方案</p> <p>5.3.2 能根据业务要求制定数据整合方案</p> <p>5.3.3 能根据业务需求及性能要求设计数据结构及格式调整方案</p> <p>5.3.4 能根据归一性需求制定数据规约方案</p> <p>5.3.5 能根据业务需求编写自定义数据预处理函数</p>	<p>5.3.1 信息技术文档编制知识</p> <p>5.3.2 数据序列化知识</p> <p>5.3.3 数据压缩知识</p>
	5.4 数据计算	<p>5.4.1 能根据业务需求编写批量、实时数据计算作业并优化作业参数</p> <p>5.4.2 能根据业务规则设计相应标签库并进行标签管理</p> <p>5.4.3 能根据业务规则设计相应数据指标计算算法</p>	<p>5.4.1 常用算法与数据结构知识</p> <p>5.4.2 数据画像知识</p> <p>5.4.3 数据倾斜知识</p>
	5.5 数据查询	<p>5.5.1 能使用大规模并行分析数据库优化联机事务分析系统性能</p> <p>5.5.2 能使用计算引擎优化数据查询效率</p> <p>5.5.3 能通过计算平台构建检索分析系统</p>	<p>5.5.1 大规模并行分析数据库知识</p> <p>5.5.2 数据立方知识</p> <p>5.5.3 查询引擎知识</p> <p>5.5.4 数据分词知识</p>
6. 大数据分析与应用	6.1 数据预处理	<p>6.1.1 能根据数据质量要求制定数据清洗策略及评估方案</p> <p>6.1.2 能根据业务要求制定数据整合方案</p> <p>6.1.3 能根据业务需求及性能要求</p>	<p>6.1.1 数据格式线性变换知识</p> <p>6.1.2 数据清洗需求分析方法</p> <p>6.1.3 数据清洗方案设计</p>

		设计数据结构及格式调整方案 6.1.4 能根据归一性需求制定数据规约方案 6.1.5 能根据业务需求编写自定义数据预处理函数 6.1.6 能根据数据特征及规律, 制定数据采样方案	计知识
	6.2 数据分析	6.2.1 能根据分析需求进行数据准备 6.2.2 能根据业务需求构建合适的分析模型 6.2.3 能使用合适的算法实现分析模型并对拟合结果进行优化 6.2.4 能分析数据的主成分及因子等相关特征, 重构数据内容 6.2.5 能针对数据结果进行归纳并输出分析报告	6.2.1 多元统计分析知识 6.2.2 判别分析知识 6.2.3 聚类分析知识 6.2.4 主成分分析知识 6.2.5 因子分析知识 6.2.6 时间序列分析知识
	6.3 数据挖掘	6.3.1 能评估挖掘需求并选择合适方法对数据进行特征工程处理 6.3.2 能使用算法库及工具创建数据挖掘模型并进行模型训练 6.3.3 能选择合适评价指标对模型进行验证及调优 6.3.4 能选择合适评价指标对模型进行测试并输出最终模型的性能评估报告 6.3.5 能使用编程语言对模型进行部署和重构	6.3.1 模型训练知识 6.3.2 模型测试知识 6.3.3 模型部署知识
	6.4 数据可视化	6.4.1 能根据业务需求及分析结果, 制定数据展示方案 6.4.2 能设计数据可视化实现方式 6.4.3 能与产品、运营人员合作美化数据报表及数据展示页面 6.4.4 能开发并优化数据可视化组件库 6.4.5 能对数据可视化结果进行业务分析并输出分析报告	6.4.1 数据可视化设计知识 6.4.2 可视化组件库开发知识
7. 大数据管理	7.1 数据管理	7.1.1 能制定数据标准管理制度, 通过制度约束、系统控制等手段, 提高平台治理水平 7.1.2 能制定数据质量管理规范, 确保平台数据质量符合规范 7.1.3 能制定数据生命周期管理规范、数据血缘关系管理规范	7.1.1 数据标准管理知识 7.1.2 数据质量管理知识 7.1.3 数据生命周期管理知识 7.1.4 数据安全知识

		<p>7.1.4 能制定安全审计要求，确保数据活动过程和相关操作符合安全要求</p>	
	<p>7.2 数据管理评估</p>	<p>7.2.1 能独立开展 DCMM 调研访谈，收集、解读评估材料</p> <p>7.2.2 能运用评估表等工具进行 DCMM 评估</p> <p>7.2.3 能分析企业数据管理现状，识别数据管理问题及改进项，给出数据管理能力成熟度等级建议</p>	<p>7.2.1 DCMM 评估方法</p> <p>7.2.2 数据治理知识</p>

3.3 高级

职业功能	工作内容	专业能力要求	相关知识要求
1. 大数据应用开发	1.1 大数据组件技术研发	1.1.1 能根据相关论文、材料实现存储、计算功能的分布式并行算法 1.1.2 能根据算法构造存储、读写或处理工具的海量数据计算引擎 1.1.3 能根据应用需求开发基于计算引擎的算子、函数或方法 1.1.4 能根据算子、函数或方法，构造队列或流程，实现计算作业功能	1.1.1 分布式算法 1.1.2 计算引擎开发知识
	1.2 应用服务开发	1.2.1 能根据系统架构，规划各项组件接口规范 1.2.2 能根据业务功能，设计接口权限及参数规范 1.2.3 能对整体系统进行数据打通方案设计 1.2.4 能对整体系统进行库表结构设计及优化 1.2.5 能对整体系统的数据传输、缓存、推送设计方案	1.2.1 软件应用接口开发知识 1.2.2 数据通信知识 1.2.3 数据缓存知识 1.2.4 消息中间件知识
	1.3 系统测试	1.3.1 能根据产品说明文档，规划系统测试计划 1.3.2 能根据测试计划，协调人力、设备等测试资源 1.3.3 能根据测试需求，编制测试脚本 1.3.4 能根据性能需求，进行系统深度性能优化测试	1.3.1 自动化测试脚本技术 1.3.2 测试工具开发方法
2. 大数据系统搭建	2.1 硬件系统搭建	2.1.1 能根据安全施工规范，整体规划硬件设施安全方案 2.1.2 能根据应用需求，规划网络配置实施方案 2.1.3 能根据产品特性，制定统一施工标准 2.1.4 能根据系统部署方案，与产品开发部门整体规划硬件承载、配置及扩展方案 2.1.5 能根据不同硬件设施，制定故障处理规范及流程	2.1.1 安全施工规范 2.1.2 硬件产品知识 2.1.3 故障管理知识
	2.2 软件系统部署	2.2.1 能根据权限安全规范，制定软件权限安全方案 2.2.2 能根据系统组件关系，配置组件使用权限	2.2.1 权限安全规范 2.2.2 软件产品交付知识 2.2.3 联邦集群知识

		<p>2.2.3 能根据产品特性，制定部署及升级策略</p> <p>2.2.4 能根据集群组件特性，制定高可用及容灾策略</p> <p>2.2.5 能根据调试结果，制定部署优化方案</p>	2.2.4 异地多活知识
3. 大数据平台管理与运维	3.1 平台管理	<p>3.1.1 能评估应用变更风险，发布应用变更计划，管控变更流程，总结变更报告</p> <p>3.1.2 能根据软件部署方式，制定各类组件应用变更或版本更迭方案</p> <p>3.1.3 能制定代码管理规范并配置代码仓库管理系统</p> <p>3.1.4 能制定各部门平台功能使用权限规范</p>	<p>3.1.1 风险管理知识</p> <p>3.1.2 应用变更知识</p> <p>3.1.3 代码管理知识</p> <p>3.1.4 权限管理知识</p>
	3.2 系统运维	<p>3.2.1 能规划监控指标，制定监控管理规范</p> <p>3.2.2 能开发监控脚本</p> <p>3.2.3 能对系统性能进行调优</p> <p>3.2.4 能使用数据挖掘方法挖掘潜在故障</p> <p>3.2.5 能对故障事故进行复盘，编写故障预防规范</p> <p>3.2.6 能定期组织容灾备份演练</p>	<p>3.2.1 性能指标知识</p> <p>3.2.2 负载均衡知识</p> <p>3.2.3 故障分析方法</p> <p>3.2.4 容灾备份知识</p>
	3.3 安全维护	<p>3.3.1 能根据安全规范制定风险预警等级</p> <p>3.3.2 能明确安全需求，审核并制定权限管理规范和数据分类分级</p> <p>3.3.3 能制定应急管理策略并定期组织安全演练</p> <p>3.3.4 能根据漏洞测试报告和突发事件应对策略，评估系统潜在风险</p> <p>3.3.5 能构建系统安全机制并完成对安全系统的开发工作</p>	<p>3.3.1 安全规范知识</p> <p>3.3.2 应急管理知识</p>
4. 大数据技术服务	4.1 技术咨询	<p>4.1.1 能建立目标市场分析模型，分析行业需求</p> <p>4.1.2 能整体输出产品解决方案</p> <p>4.1.3 能独立解决客户技术咨询难题，并提供技术解决方案</p> <p>4.1.4 能参与项目架构设计与产品设计，并提出建设性意见</p>	<p>4.1.1 大数据架构分析知识</p> <p>4.1.2 大数据产品设计知识</p>

	4.2 解决方案设计	<p>4.2.1 能根据产品功能设计和技术架构，输出产品的配套文档，并根据项目需求针对性设计项目解决方案</p> <p>4.2.2 能与业务部门合作引导和挖掘客户需求，并输出解决方案</p> <p>4.2.3 能挖掘行业普遍需求，提炼产品价值特征，整理竞品分析报告</p> <p>4.2.4 能主动分析与挖掘市场情况，对市场策略制定提出建议</p>	<p>4.2.1 项目管理方法</p> <p>4.2.2 需求分析技术</p>
	4.3 指导与培训	<p>4.3.1 能结合技术发展方向，调研大数据先进技术并进行技术团队建设</p> <p>4.3.2 能分析现有大数据产品技术体系及可优化方向，并向技术团队培训</p> <p>4.3.3 能整理大数据产品操作手册，并指导技术或非技术人员产品使用方法</p>	<p>4.3.1 技术调研方法</p> <p>4.3.2 团队组建知识</p> <p>4.3.3 产品操作手册制作方法</p>
	4.4 流程优化与管理	<p>4.4.1 能建立业务需求收集业务指标数据，并根据实际数据建立业务分析模型</p> <p>4.4.2 能根据数据分析情况指导业务开展及流程优化</p> <p>4.4.3 能管理不同业务部门的开发生产活动</p>	<p>4.4.1 流程优化知识</p> <p>4.4.2 运营管理知识</p>
5. 大数据处理与应用	5.1 数据采集	<p>5.1.1 能根据业务场景制定数据采集策略并监控采集情况</p> <p>5.1.2 能根据业务场景制定数据迁移策略并监测迁移情况</p> <p>5.1.3 能根据业务及性能需求设计消息传输及推送方案</p> <p>5.1.4 能根据业务需求及依赖关系设计调度方案</p> <p>5.1.5 能根据业务需求及存储应用设计存储策略</p>	<p>5.1.1 数据采集与迁移策略</p> <p>5.1.2 依赖调度原理知识</p> <p>5.1.3 存储架构知识</p>
	5.2 数据建模	<p>5.2.1 能制定数据建模流程规范</p> <p>5.2.2 能根据业务需求，对模型进行优化</p> <p>5.2.3 能跨团队部门协作，系统性分析并解决各类数据平台相关的运行或数据问题</p> <p>5.2.4 能根据设计方法，构建面向服务或数据的数据建模体系架构</p>	<p>5.2.1 数据平台设计知识</p> <p>5.2.2 数据建模知识</p> <p>5.2.3 SOA/DOA 体系结构知识</p>

	5.3 数据预处理	<p>5.3.1 能根据质量要求，制定数据清洗流程规范</p> <p>5.3.2 能根据数据处理需求，制定统一数据预处理方案</p> <p>5.3.3 能根据作业存在风险，制定预处理异常处理机制</p> <p>5.3.4 能根据系统特性，优化预处理系统性能指标</p>	<p>5.3.1 数据清洗流程规范</p> <p>5.3.2 语法树解析知识</p>
	5.4 数据计算	<p>5.4.1 能根据业务需求设计离线或实时数据计算算法</p> <p>5.4.2 能制定数据标签库管理及规范</p> <p>5.4.3 能制定数据计算开发流程及规范</p> <p>5.4.4 能根据业务规则对关系对象进行图计算</p>	<p>5.4.1 数据算法设计方法</p> <p>5.4.2 数据分区及缓存知识</p> <p>5.4.3 外部程序管道知识</p> <p>5.4.4 图计算知识</p>
	5.5 数据查询	<p>5.5.1 能制定数据查询操作流程及规范</p> <p>5.5.2 能深入计算引擎对底层代码进行优化以提升查询性能</p> <p>5.5.3 能制定对外数据接口规范及权限</p>	<p>5.5.1 计算引擎优化知识</p> <p>5.5.2 索引优化知识</p> <p>5.5.3 驱动器与执行器知识</p>
6. 大数据分析与挖掘	6.1 数据分析	<p>6.1.1 能结合理论和业务实际，进行大数据分析相关算法研究</p> <p>6.1.2 能针对研究结果设计分析算法并指导算法模型实现</p> <p>6.1.3 能针对现有算法提出新的改进和优化方法，建立新的分析体系</p>	<p>6.1.1 大数据分析算法</p> <p>6.1.2 综合评价方法知识</p>
	6.2 数据挖掘	<p>6.2.1 能根据理论研究及数学原理，构建并行挖掘算法</p> <p>6.2.2 能根据挖掘性能及业务特征，优化挖掘算法</p> <p>6.2.3 能根据业务特性，制定合适的挖掘模型评价指标</p> <p>6.2.4 能对挖掘模型使用的多源异构数据源进行适配</p>	<p>6.2.1 大数据挖掘算法</p> <p>6.2.2 机器学习知识</p> <p>6.2.3 语义分析知识</p>
	6.3 数据可视化	<p>6.3.1 能根据业务分析需求及分析结果，指导数据展示方案制定</p> <p>6.3.2 能研发并设计前端图表展示功能代码</p> <p>6.3.3 能完成可视化组件开发、封装及优化</p> <p>6.3.4 能对数据可视化结果进行业务分析并输出分析报告</p>	<p>6.3.1 前端展示需求分析方法</p> <p>6.3.2 前端优化技术</p>

7. 大数据管理	7.1 数据规划	<p>7.1.1 能结合企业发展目标制定数据战略</p> <p>7.1.2 能根据主流的企业架构框架设计大数据框架</p> <p>7.1.3 能定制企业级数据管理解决方案，制定技术路线和标准规范</p> <p>7.1.4 能推动企业实施数据管理</p> <p>7.1.5 能制定元模型标准</p> <p>7.1.6 能制定数据资产管理规则，注册入库数据资产信息，并进行数据资产维护</p>	<p>7.1.1 整体规划知识</p> <p>7.1.2 经营管理知识</p> <p>7.1.3 产品优化知识</p> <p>7.1.4 系统架构知识</p> <p>7.1.5 数据资产管理知识</p>
	7.2 数据管理评估	<p>7.2.1 能编写、分析 DCMM 评估报告</p> <p>7.2.2 能解读 DCMM 模型并组织现场评估</p> <p>7.2.3 能建立企业数据管理组织和制度</p> <p>7.2.4 能指导企业实施 DCMM 改进</p>	<p>7.2.1 DCMM 评估报告的结构和内容</p> <p>7.2.2 企业业务流程优化和再造知识</p> <p>7.2.3 成熟度模型改进知识</p>

4 权重表

4.1 理论知识权重表

专业技术等级 项目		初级 (%)			中级 (%)			高级 (%)
		大数据处理方向	大数据分析方向	大数据管理方向	大数据处理方向	大数据分析方向	大数据管理方向	
基本要求	职业道德	5	5	5	5	5	5	5
	基础知识	20	20	25	15	15	25	10
相关知识要求	大数据应用开发	-	-	-	15	15	-	15
	大数据系统搭建	20	20	-	15	15	-	10
	大数据平台管理与运维	20	20	25	20	20	20	10
	大数据技术服务	10	10	20	10	10	20	15
	大数据处理与应用	25	-	-	20	-	-	10
	大数据分析与管理	-	25	-	-	20	-	15
	大数据管理	-	-	25	-	-	30	10
合计		100	100	100	100	100	100	100

4.2 专业能力要求权重表

专业技术等级 项目		初级 (%)			中级 (%)			高级 (%)
		大数据处理方向	大数据分析方向	大数据管理方向	大数据处理方向	大数据分析方向	大数据管理方向	
专业技术要求	大数据应用开发	-	-	-	15	15	-	15
	大数据系统搭建	30	30	-	15	15	-	10
	大数据平台管理与运维	20	20	20	25	25	20	20
	大数据技术服务	10	10	20	15	15	20	20
	大数据处理与应用	40	-	-	30	-	-	10
	大数据分析与管理	-	40	-	-	30	-	15
	大数据管理	-	-	60	-	-	60	10
合计		100	100	100	100	100	100	100

附录——参考文献

- [1] GB/T 35589-2017 《信息技术 大数据 技术参考模型》相关知识
- [2] GB/T 35295-2017 《信息技术 大数据 术语》相关知识
- [3] GB/T 38673-2020 《信息技术 大数据 大数据系统基本要求》相关知识
- [4] GB/T 37721-2019 《信息技术 大数据分析系统功能要求》相关知识
- [5] GB/T 37722-2019 《信息技术 大数据存储与处理系统功能要求》相关知识
- [6] GB/T 36073-2018 《数据管理能力成熟度评估模型》相关知识