

团 体 标 准

T/BDIA 001—2020
T/CA 008—2020

大数据工程师职业能力要求

Requirements for professional ability of Big Data Engineer

2020 - 09 - 01 发布

2020 - 09 - 01 实施

中关村大数据产业联盟 发布
中国通信工业协会

目 次

前言	III
1 范围	1
2 职业概况	1
2.1 职业名称	1
2.2 职业定义	1
3 职业技能等级	1
4 基本条件	1
4.1 职业环境条件	1
4.2 职业能力特征	1
4.3 普通受教育程度	1
5 职业技能鉴定要求	1
5.1 申报条件	1
5.2 鉴定方式	2
5.3 监考人员、考评人员与考生配比	2
5.4 鉴定时间	2
5.5 鉴定场所设备	2
6 基本要求	3
6.1 职业道德	3
6.2 基础知识	3
7 能力要求	3
7.1 大数据初级工程师	3
7.2 大数据中级工程师	4
7.3 大数据高级工程师	6

T/BDIA 001—2020
T/CA 008—2020

全国团体标准信息平台

前 言

本标准按照GB/T 1.1-2009给出的规则起草。

本标准由中关村大数据产业联盟、中国通信工业协会提出并归口。

本标准起草单位：中国人事科学研究院、中关村大数据产业联盟标准委员会、北京大学、北京邮电大学、北京航空航天大学、深圳大学、南京理工大学、东北大学、西安邮电大学、北京信息科技大学、北京化工大学、北大方正软件学院、北京腾云天下科技有限公司、用友软件股份有限公司、北京京东尚科信息技术有限公司、贝壳找房（北京）科技有限公司、北京字节跳动科技有限公司、泛华产业投资有限公司、方正国际软件有限公司、深圳数询教育科技有限责任公司、天津中德应用技术大学、中国联合网络通信集团有限公司、中国普天信息产业股份有限公司、北京先科众诚科技有限公司、北京易华录信息技术股份有限公司、软通动力信息技术（集团）有限公司、北京亚信数据有限公司、美林数据技术股份有限公司、北京久其软件股份有限公司、成都数联铭品科技有限公司、天德爱民（北京）科技有限公司、北京明略软件系统有限公司、上海海略管理信息咨询有限公司、朗新科技股份有限公司、北京数资科技有限公司、北京网智天元科技股份有限公司、北京启元资本市场发展服务有限公司、北京优炫软件股份有限公司、北京百分点信息科技有限公司、睿智大数据科技有限公司、山西中交高速数据中心有限公司、中软国际有限公司、中科软科技股份有限公司、北京北大千方科技有限公司、北京江民新科技有限公司、北京银江智慧城市技术有限公司、泰豪科技股份有限公司、四川千行众联控股集团有限公司、江西高创安邦技术有限公司、厦门精位文化地理信息有限公司、北京龙坤软件科技有限公司、百思网络科技有限公司、智绘九州（北京）科技有限公司、上海和鲸科技有限公司。

本标准起草人：佟娅丽、赵国栋、孙科、高宇、乐以长、张成文、李薇薇、杨青、刘恒农、孔涛、辛亚平、刘庆振、杨天杰、姚菲、杨慧、韩伟、胡加艳、刘磊、王晨光、杨天举、杨洋、徐建立、李静怡、马波、苏民、雷磊、孙科、张记高、张雪峰、宋伟、蒋怀超、李妍、王哲鹏、王维、郭远洋、石坚、游涛、王颖、李志勇、曾庆凯、王现增、万涛、郝文军、聂永春、李冰、贺俊、赵晶、胡卫彪、杜佳、徐志强、何英锋、范向伟。

本标准由中关村大数据产业联盟标准委员会负责解释。

本标准为首次制定。

大数据工程师职业能力要求

1 范围

本标准规定了大数据工程师的定义、职业技能等级、职业基本要求和工作要求。
本标准适用于从事大数据工程师的培训与认定。

2 职业概况

2.1 职业名称

大数据工程师。

2.2 职业定义

本标准所涉及的大数据工程师是指从事大数据相关工作的技术人员。包括但不限于从事大数据采集、清洗、存储、挖掘、建模、分析等，并加以开发、利用、管理、维护和服务的相关工作的技术人员。

3 职业技能等级

本职业共设三个等级，由低到高可分为：大数据初级工程师、大数据中级工程师、大数据高级工程师。

4 基本条件

4.1 职业环境条件

符合从业人员所处的客观工作环境。

4.2 职业能力特征

逻辑思维能力、跨界思维能力、编程能力、算法设计能力、数据分析能力、综合应用知识能力、沟通协调能力和团队合作能力、持续学习和创新能力。

4.3 普通受教育程度

高中毕业及以上（或同等学力）。

5 职业技能鉴定要求

5.1 申报条件

5.1.1 具备以下条件之一者，可申报初级工程师：

——累计从事本职业或相关职业工作1年(含)以上；

T/BDIA 001—2020

T/CA 008—2020

- 取得技工学校、职业学校本专业或相关专业毕业证书(含尚未取得毕业证书的在校应届毕业生);或取得经评估论证、以中级技能为培养目标的其他学校本专业或相关专业毕业证书(含尚未取得毕业证书的在校应届毕业生);
- 高等院校本专业或相关专业在校生。

5.1.2 具备以下条件之一者,可申报中级工程师:

- 取得本职业或相关职业初级工职业资格证书(技能等级证书)后,累计从事本职业或相关职业工作3年(含)以上;
- 取得本职业或相关职业初级工程师职业资格证书(技能等级证书)并具有高级技工学校、技师学院、职业学校毕业证书(含尚未取得毕业证书的在校应届毕业生);或取得本职业或相关职业初级工职业资格证书(技能等级证书)并具有经评估论证、以高级技能为培养目标的其他学校本专业或相关专业毕业证书(含尚未取得毕业证书的在校应届毕业生),累计从事本职业或相关职业工作3年(含)以上;
- 具有大学专科本专业或相关专业毕业证书,并取得本职业或相关职业三级/初级工职业资格证书(技能等级证书)后,累计从事本职业或相关职业工作3年(含)以上;
- 具有大学本科本专业或相关专业学历证书,并取得本职业或相关职业三级/初级工职业资格证书(技能等级证书)后,累计从事本职业或相关职业工作2年(含)以上;
- 具有硕士及以上本专业或相关专业学历证书(含尚未取得毕业证书的在校应届毕业生)。

5.1.3 具备以下条件之一者,可申报高级工程师:

- 取得本职业或相关职业中级工职业资格证书(技能等级证书)后,累计从事本职业或相关职业工作1年(含)以上;
- 取得本职业或相关职业中级工职业资格证书(技能等级证书)的高级技工学校、技师学院毕业生累计从事本职业或相关职业工作3年(含)以上或取得本职业或相关职业预备技师证书的技师学院毕业生累计从事本职业或相关职业工作2年(含)以上;
- 具有大学本科本专业或相关专业学历证书,并取得本职业或相关职业中级工职业资格证书(技能等级证书)后,累计从事本职业或相关职业工作3年(含)以上;
- 具有硕士本专业或相关专业学历证书,并取得本职业或相关职业中级工职业资格证书(技能等级证书)后,累计从事本职业或相关职业工作1年(含)以上;
- 具有博士本专业或相关专业学历证书。

5.2 鉴定方式

分为理论知识考试和技能操作考核,理论知识考试和技能操作考核均采用闭卷考试的方式,技能操作考核采用上机操作形式。高级工程师还须进行综合评审(能力展示与答辩)。

所有考试科目均实行百分制,成绩均达到60分(含)以上者为合格。

5.3 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,综合评审委员为3人(含)以上单数。

5.4 鉴定时间

理论知识考试时间150分钟,技能操作考核时间120分钟,综合评审时间30分钟。

5.5 鉴定场所设备

标准教室或具备相应条件的会议室，室内配备必要的照明设备、计算机、投影仪等多媒体设备，室内卫生，光线良好，通风条件良好。

6 基本要求

6.1 职业道德

大数据工程师的职业道德应包括但不限于以下要求：

- 诚实守信，遵纪守法；
- 严格遵守国家信息和数据安全各项法律法规及相关单位的各项制度，增强数据道德意识与警惕意识；
- 树立正确的“大数据主权观与责权观”，坚决抵制利用数据与大数据技术进行危害国家安全、经济环境与公民权益的各类非法活动；
- 自觉维护大数据产业各方的正当利益和合法权益。保守用户数据隐私与商业秘密；
- 不利用用户隐私数据从事任何与向用户作出的承诺无关的活动。

6.2 基础知识

数学、统计学、计算机编程语言、软件工程、数据库、计算机网络技术、算法分析与设计、数据安全、机器学习与数据挖掘。

7 能力要求

本标准对大数据初级工程师、大数据中级工程师和大数据高级工程师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。

7.1 大数据初级工程师

7.1.1 大数据开发要求

7.1.1.1 具备开展大数据应用开发环境安装与部署的能力

- 能够安装主流操作系统；
- 掌握操作系统基础知识及常用操作命令。

7.1.1.2 具备数据采集与清洗的能力

- 能够安装部署并运行主流大数据平台；
- 能够使用数据采集工具对多源异构数据进行采集；
- 能够使用数据清洗工具对数据进行抽取、转换和加载。

7.1.1.3 具备数据存储、计算与交换的能力

- 能够将采集的数据按业务需求进行存储；
- 能够使用主流分布式工具对数据进行计算；
- 能够将大数据平台中的数据同多种数据库进行交换。

7.1.2 大数据运维要求

7.1.2.1 具备平台基础设施管理的能力

- 能够掌握分片、主从、管理节点、工作节点等集群基础知识；

T/BDIA 001—2020

T/CA 008—2020

- 能够搭建大数据集群的硬件环境，包括服务器、网络设备等；
- 能够安装大数据的基础软件，包括操作系统、大数据平台等；
- 能够对大数据集群中的网络环境、软硬件环境参数进行配置。

7.1.2.2 具备平台基础服务管理的能力

- 能够安装大数据自动化运维、监控、故障处理工具软件；
- 能够使用集群监控工具，监控大数据平台运行情况并进行日常巡检；
- 能够通过监控平台的警告及日志信息，进行故障识别及软硬件修复。

7.1.3 数据分析要求

7.1.3.1 具备需求分析及数据提取的能力

- 能够通过客户调研访谈等方式确定业务需求；
- 能够根据需求，利用数据提取工具获取适当的数据。

7.1.3.2 具备数据处理及可视化展示的能力

- 能够按照最终输出报告的要求对现有数据进行处理；
- 能够使用数据可视化工具，选择适当的统计分析方法对数据进行可视化展示，并形成报告；
- 能够准确清晰的在报告中体现出数据结果。

7.1.4 大数据治理要求

应具备数据安全治理的能力包括但不限于以下要求：

- 能够了解数据安全相关政策法规、标准规范；
- 能够了解数据加解密、4A 系统、DLP（数据防泄漏）、数据脱敏、数字水印等数据治理内容，以及数据安全审计技术：包括日志审计、网络流量分析；
- 能够了解大数据安全运营的主要内容，如数据分级管理、数据输出安全审核、数据平台权限管理及优化等；
- 能够了解数据安全的主要风险与防护手段以及密码学技术体系知识。

7.2 大数据中级工程师

7.2.1 大数据开发要求

7.2.1.1 具备项目规划与过程管理的能力

能够参与大数据项目的全生命周期，包括设计、开发、测试、部署和交付等。

7.2.1.2 具备数据采集与清洗的能力

- 能够梳理业务数据，并根据数据特征制定采集策略，参与采集系统的开发和维护；
- 能够根据数据清洗规范，设计数据接口及数据模型，并实现数据清洗流程；
- 能够掌握分布式文件系统知识，根据业务数据规模进行分布式存储结构选型及实现。

7.2.1.3 具备数据计算与应用开发的能力

- 能够掌握主流分布式算法知识，如 MapReduce；
- 能够掌握主流流式计算框架，如 Spark、Storm 等；
- 能够对海量数据的计算架构进行技术选型；
- 能够对数据的搜索、查询、分析应用进行设计与编码实现。

7.2.2 大数据挖掘要求

7.2.2.1 具备数据提取及预处理的能力

- 能够利用编程语言如 SQL（结构化查询语言）、R 语言等从数据平台中提取所需数据；
- 能够利用专用工具软件从数据平台中提取所需数据；
- 能够根据业务需求将数据分发到预置的计算平台中并进行预处理。

7.2.2.2 具备数据挖掘实现的能力

- 能够掌握主流的机器学习算法及其原理；
- 能够掌握主流的分布式计算框架并熟悉其使用；
- 能够掌握主流流式计算框架并熟悉其使用；
- 能够熟练使用至少一种开发语言进行代码编写；
- 能够根据业务需求进行建模，评估模型效果并进行针对性的优化。

7.2.2.3 具备数据可视化的能力

能够利用可视化工具对挖掘结果进行展现及说明。

7.2.3 大数据运维要求

7.2.3.1 具备大数据平台运维规划与设计的能力

能够根据业务场景预测工作负载，设计集群环境的软硬件方案。

7.2.3.2 具备大数据平台监控管理的能力

- 能够维护或运维大型线上系统，设计及搭建监控系统；
- 能够对大数据平台软硬件资源的工作状态进行监控；
- 能够对大数据监控平台的预警、故障及问题进行定位、分析和解决；
- 能够将维护日志、故障、问题记录等形成运维报告。

7.2.3.3 具备大数据平台性能优化的能力

能够对运维报告进行分析，根据分析结果对软硬件环境及运维工作机制进行调优。

7.2.4 数据分析要求

7.2.4.1 具备需求分析及数据提取的能力

能够根据业务需求，指导并实现数据的提取及处理。

7.2.4.2 具备数据建模及结果展示的能力

- 能够根据业务需求进行建模；
- 能够选择适当的方法、工具或编程语言输出可视化结果，并形成报告；
- 能够准确清晰的体现分析结果，并对产生结果的原因进行分析。

7.2.5 大数据治理要求

7.2.5.1 具备数据安全治理的能力

- 能够熟知主流 Web 安全技术（SQL 注入、XSS、CSRF 等），能够实际动手对 Web 网站进行漏洞扫描、渗透测试、社会工程学等 Web 安全渗透测试和评估工作；
- 能够了解渗透测试、攻防对抗、应急响应等网络攻防等实际案例，能够熟练使用各类扫描工具、弱口令探测工具、嗅探工具、Web 系统渗透工具、后门检测工具；
- 能够参与数据安全的日常维护和管理的工作，包括数据安全事件的监控与应急响应、合规审计与调查分析与追踪；挖掘深层次潜藏的安全事件；

T/BDIA 001—2020

T/CA 008—2020

- 能够了解数据安全标准规范的制定和实施,包括需求识别、风险识别、分析、评级和安全管理;
- 能够掌握主流的数据库系统如MySQL、SQLite、Redis、Memcache等,Elasticsearch、Solr等常见搜索系统;熟悉基本的数据库操作语言;熟悉Hadoop大数据平台的安全防护措施;
- 能够了解互联网交易系统的数据流转、数据访问权限管理、日志审计以及动静态脱敏等的技术。

7.2.5.2 具备安全咨询的能力

- 能够熟练掌握数据安全相关法律法规;例如等保2.0数据合规、欧盟GDPR,《网络安全法》、《个人信息安全规范》、云安全、大数据安全等;
- 能够了解PIA、合规出境、DSMM、27001、ITIL、CISSP等相关安全与管理标准;
- 能够熟知数据安全咨询服务内容,包括数据安全治理、DSMM评估、个人信息安全评估、数据分类分级、数据安全规划、数据安全风险评估、数据中台安全评估等;
- 能够利用各类日志、流量数据或工具进行威胁猎捕,发现入侵事件、威胁活动、异常行为等安全问题,以帮助用户缓解安全风险;
- 能够针对数据的全生命周期流转,参与评估数据安全风险、设计数据安全规划;
- 能够基于大数据具体业务场景,分析、评估用户面临的数据安全风险和安全现状。

7.3 大数据高级工程师

7.3.1 架构设计要求

7.3.1.1 具备数据架构设计的能力

- 能根据业务需求及数据资源设计数据模型与接口,并完成总体架构设计;
- 能够开发核心技术框架,并根据业务及技术需求制定开发规范;
- 能够进行核心业务/系统架构的规划、设计、构建;对资源进行合理分配和利用,在质量、性能和效率方面达到平衡;
- 能够根据业务、技术、运维报告的结果调整优化现有系统及应用架构;
- 能够指导维护或运维大型线上系统,设计及搭建监控系统;
- 能够对系统稳定性、可用性进行评估并提出优化方案。

7.3.1.2 具备技术管理和培训的能力

- 能够对数据采集、清洗、存储等相关流程制定规范;
- 能够对大数据项目的全生命周期,包括设计、开发、测试、部署和交付等进行指导;
- 能够根据团队情况制定培训计划,从技术、业务、方法论等角度对初中级技术人员进行指导。

7.3.2 数据挖掘要求

7.3.2.1 具备模型构建与算法设计的能力

- 能够深入理解主流机器学习、深度学习等算法及模型原理,并应用于实际生产系统;
- 能够基于数据及业务需求设计高性能、可扩展的技术方案;
- 能够实现主流算法,选择最佳数据项进行模型构建及算法设计;
- 能够对所构建模型及算法进行评估,提出针对性改进方案;
- 能够根据挖掘结果总结数据价值,预测业务走向,提出模型应用、指导业务发展。

7.3.2.2 具备数据项目管理的能力

- 能够有效沟通并确定建模目标、建模需求;
- 能够根据数据生命周期梳理业务关键点,明确关键环节产出;根据资源、人员对大数据项目进度与质量进行管理;

- 能够独立或带领团队实现关键算法创新；
- 能够从数据和算法的角度下把握项目走向及资源分配。

7.3.2.3 技术管理和培训的能力

- 能够根据项目需求对模型选型、数据项选取、数据准备、特征工程及模型验证等进行实现及管理；
- 能够根据团队情况制定培训计划，从技术、业务、方法论等角度对初中级技术人员进行指导。

7.3.3 数据分析要求

7.3.3.1 具备问题转化、分解与分析的能力

- 能够从项目/业务需求中提取关键信息并转化为数据问题，形成可执行的目标；
- 能够根据对业务的理解识别和预测风险；
- 能够根据数据问题制定合适的分析思路并指导完成分析过程；
- 能够使用适当的挖掘工具提升分析效率。

7.3.3.2 具备数据可视化展示及业务指导的能力

- 能够根据产出结果选择合适的可视化形式，梳理可视化结果逻辑并选择合理的展示指标；
- 能够输出具有业内高水准的技术文献、规范、报告或行业方案；
- 能够根据业务需求指导建模；
- 能够通过分析结果产出行动指导建议。

7.3.3.3 具备技术管理和培训的能力

能够根据团队情况制定培训计划，从技术、业务、方法论等角度对初中级技术人员进行指导。

7.3.4 大数据治理要求

7.3.4.1 具备数据安全治理的能力

- 能够熟练掌握并灵活运用数据安全体系的技术方向、实现思路、产品架构、功能设计等，针对用户具体应用场景，能够建设数据安全管控和审计体系，包括数据安全架构、流程制度及技术工具，执行数据安全策略的落实，包括数据分类分级、元数据管理、数据加密、数据脱敏、数据调用审计等；
- 能够负责数据安全事件的风险预警、溯源、协同、跟踪、改进优化及事后评估；
- 能够负责数据安全咨询服务的项目管理和具体落地执行工作，编写项目用户的管理制度规范等交付文档，参与部分数据安全国标行标的编写；
- 能够识别用户的数据安全风险，分析风险影响，设计制定并实施基于风险的数据安全治理架构、方针策略、制度规范及标准流程等；并结合用户实际情况提供有效的数据安全管理解决方案，提升用户的数据安全管控能力及成熟度，具备数据安全风险管理能力；
- 能够结合国内与国际个人信息保护与隐私监管体系，行业最佳实践，例如等保 2.0 数据合规、欧盟 GDPR，《网络安全法》、《个人信息安全规范》、云安全、大数据安全等，进行对标解读，提出相应合规要求并跟踪落实，确保用户数据安全的合规性；
- 能够建设数据安全运营体系，包括数据安全使用指导、数据安全风险发现的整改落地推进、日常安全事件响应、问题咨询与方案支持、数据安全意识提升与培训等。

7.3.4.2 具备隐私保护治理的能力

- 能够结合国内与国际个人信息保护与隐私监管体系、行业最佳实践，例如等保 2.0 数据合规、欧盟 GDPR，《网络安全法》、《个人信息安全规范》等，健全隐私保护体系，包括隐私保护

T/BDIA 001—2020

T/CA 008—2020

技术方案、流程制度等；

- 能够跟踪、分析国内外隐私保护的法律法规、行业动态、典型案例等，并可面向不同用户开展对标评估，提出有效解决方案并推动实施；
- 能够对用户内涉及个人数据处理的业务实施隐私合规评审，把控业务的隐私合规风险。

7.3.4.3 具备合规与审计的能力

- 能够为用户设计并建设数据安全与隐私合规体系，组织实施用户的合规认证与运营，包括等保认证、ISO 认证、金融安全认证等。以满足监管要求，并符合行业实践标准；
 - 能够实施数据安全与隐私保护审计，通过季度、年度、不定期审计等，评估数据安全与隐私保护的落地成熟度，识别风险并推动整改；
 - 能够培训和推广数据安全与隐私保护策略、流程、方案，从技术、业务、方法论等角度对初中级技术人员进行指导。
-

