

数据采集
职业技能等级标准

目 次

前言.....	01
1 范围.....	02
2 规范性引用文件.....	02
3 术语和定义.....	02
4 对应院校专业.....	04
5 面向工作岗位（群）.....	04
6 职业技能要求.....	04
参考文献.....	11

前 言

本标准按照GB/T 1.1-2009给出的规则起草。

本标准起草单位：浪潮集团有限公司、北京百分点信息科技有限公司、北京大唐高鸿数据网络技术有限公司、济南华信计算机技术有限公司、山东青年政治学院、山东中医药大学、哈尔滨职业技术学院、济南职业学院、莱芜职业技术学院、山东电子职业技术学院、济南信息工程学校、济南电子机械工程学校。

本标准主要起草人：刘何秀、穆建平、陈腾、商程、崔瑞娟、苏萌、刘洋、高敬萍、宁玉富、曹慧、李明、徐翠娟、宋磊、许文宪、徐健、刘学、赵素霞、赵庆华。

声明：本标准的知识产权归属于浪潮集团有限公司，未经浪潮集团有限公司同意，不得印刷、销售。

1 范围

本标准规定了数据采集职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于数据采集职业技能培训、考核与评价，相关用人单位的人员聘用、培训与考核可参照使用。

2 规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本标准。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 5271.1-2000 信息技术 词汇 第1部分:基本术语

GB/T 5271.17-2010 信息技术 词汇 第17部分: 数据库

GB/T 35295-2017 信息技术 大数据 术语

GB / T 33847-2017 信息技术 中间件 术语

GB/T 17143.6-1997 信息技术 开放系统互连 系统管理 第6部分:日志控制功能

GB/T 34840.3-2017 信息与文献 电子办公环境中文件管理原则与功能要求 第3部分:业务系统中文件管理指南与功能要求

3 术语和定义

GB/T 5271.1-2000 GB/T 35295-2017 GB/T 5271.17-2010 GB / T 33847-2017 GB/T 17143.6-1997 GB/T 34840.3-2017界定的以及下列术语和定义适用于本标准。

3.1 日志 log

以用作日志记录档案库的资源为模型的管理支持客体类。

[GB-T 17143.6-1997, 定义 3.1]

3.2 关系数据库 relational database

数据按关系模型来组织的数据库。

[GB/T 5271.17-2010, 定义 3.2]

3.3 非结构化数据 unstructured data

支持一个或多个应用领域，按概念结构组织的数据集合，其概念结构描述这些数据的特征及其对应实体间的联系。

[GB/T 35295-2017, 定义3.3]

3.4 中间件 middleware

位于系统软件之上，用于支持分布式应用软件，连接不同软件实体的支撑软件。

[GB / T 33847-2017, 定义 3.4]

3.5 硬件 hardware

信息处理系统物理组成部分的全部或部分。

[GB/T 5271.1-2000, 定义 3.5]

3.6 操作系统 operating system

控制程序*执行的软件，它能提供诸如资源分配、目录调度、输入输出控制及数据管理的服务。

[GB/T 5271.1-2000, 定义3.6]

3.7 业务系统 business system

生成或管理机构活动数据与证据的自动化系统。

注：主要包括旨在为机构部门与客户之间开展事务处理提供便利的应用程序。

例如，电子商务系统、客户关系管理系统、为特定目的而构建的或为客户定制的数据库、财务或人力资源系统等。

[GB/T 34840.3-2017，定义3.7]

4 对应院校专业

中等职业学校：计算机应用、计算机网络技术、网站建设与管理、软件与信息服务、网络信息安全、移动应用技术与服务等相关专业。

高等职业学校：云计算技术与应用、大数据技术与应用、计算机应用技术、计算机网络技术、计算机信息管理、软件与信息服务、软件技术、移动应用开发、电子商务技术、信息安全与管理等相关专业。

应用型本科学校：计算机科学与技术、数据科学与大数据技术、网络工程、软件工程、信息安全等相关专业。

5 面向工作岗位（群）

主要面向政府信息化监管机构相关技术岗位，面向企事业单位的数据库管理、互联网应用开发实施、系统平台设备运维管理、系统数据分析、互联网舆情监控等相关岗位。

6 职业技能要求

6.1 职业技能等级划分

数据采集职业技能等级分为三个等级：初级、中级、高级，三个级别依次递进，高级别涵盖低级别职业技能要求。

【数据采集】（初级）主要面向各企事业单位、政府等的信息化数字化、软件研发、软件测试、系统运维、系统审计部门，从事数据库数据、业务系统日志

数据、互联网应用数据等的采集、清洗与存储工作。

【数据采集】(中级): 主要面向各企事业单位、政府等的信息化数字化、软件研发、软件测试、系统运维、系统审计部门,从事数据库数据、业务系统日志数据、互联网应用数据、容器日志数据、操作系统日志数据等的采集、清洗与存储工作。并根据业务场景需求编制并实施解决方案。

【数据采集】(高级): 主要面向各企事业单位、政府等的信息化数字化、软件研发、软件测试、系统运维、系统审计部门,从事数据库数据、业务系统日志数据、互联网应用数据、容器日志数据、操作系统日志数据、设备日志数据等的采集、清洗与存储工作。并根据业务场景需求进行采集监控系统整体架构设计,编制并管理实施解决方案。

6.2 职业技能等级要求描述

表 1 数据采集职业技能等级要求(初级)

工作领域	工作任务	职业技能要求
1. 数据库数据采集	1.1 熟悉法规制度规范	1.1.1 熟悉关于数据库操作的法律规范,保证在合法的情况下进行数据库数据采集。 1.1.2 掌握关于数据库操作的国家标准,确保数据采集过程合规。 1.1.3 熟悉行业内的技术标准和规范,能够准确获取数据库数据信息。
	1.2 数据库数据的收集	1.2.1 能够掌握常规访问数据库的方式方法,能够获取到目标数据库的相关信息。 1.2.2 掌握常用数据库客户端工具的使用,能够成功登录目标数据库并进行相关数据表数据的查询、筛选等数据收集操作。
	1.3 数据库数据的清洗	1.3.1 能够利用数据清洗工具编写基础的数据验证规则进行数据的合法性验证。 1.3.2 熟练掌握数据拆分规则,能够完成数据分解。
	1.4 数据库采集数据的存储	1.4.1 掌握某种关系型数据库,具备数据库的管理、运维能力,可以合理的规划、设计数据库。 1.4.2 掌握 SQL 语言,熟悉数据库结构设计

		及优化，具备将清理的中间数据存储到另一目标数据库或数据表的能力。
2. 业务系统日志数据采集	2.1 熟悉法规制度规范	2.1.1 熟悉关于业务系统日志数据采集的法律规范，保证在合法的情况下进行业务系统日志数据采集。 2.1.2 掌握关于业务系统日志数据采集的国家标准，确保数据采集过程合规。 2.1.3 熟悉行业内的技术标准和规范，能够准确获取业务系统日志数据信息。
	2.2 业务系统日志数据的收集	2.2.1 熟悉目前常用的系统日志采集工具，具备可以收集业务系统日志数据的能力。
	2.3 业务系统日志数据的清理	2.3.1 熟悉业务系统日志的组成部分，具备快速获取、分析业务系统日志中特定数据信息的能力。 2.3.2 能够利用数据清洗工具编写基础的数据验证规则进行数据的合法性验证。 2.3.3 熟练掌握数据拆分规则，能够完成数据分解。
	2.4 业务系统日志数据的存储	2.4.1 掌握各类数据文件存储格式，并能使用相关技术将数据保存成不同类型文件。 2.4.2 掌握系统日志数据之间的关系及分类，能够按照其数据结构及时保存到数据库。
3. 互联网应用数据采集	3.1 熟悉互联网应用数据采集法律法规	3.1.1 遵守国家《数据安全管理办法》及相关法律法规。
	3.2 互联网应用数据的采集	3.2.1 熟悉不同互联网应用数据类型。 3.2.2 能够使用工具或编写程序获取不同类型互联网数据并进行数据抽取。
	3.3 互联网应用数据的存储	3.3.1 掌握各类数据文件存储格式，并能使用相关技术将数据保存成不同类型文件。 3.3.2 掌握数据之间的关系及分类，能够按照其数据结构保存到数据库。

表 2 数据采集职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
1. 数据库数据采集	1.1 数据库数据的收集	1.2.1 能够掌握访问分布式数据库的方式方法，能够获取到目标数据库的相关信息。 1.2.2 掌握分布式数据库客户端工具的使用，能够成功登录目标数据库并进行相关数据的查询、筛选等数据收集操作。
	1.2 数据库数据的清洗处理	1.3.1 掌握简单批处理程序的编写，可以编写批处理程序对数据进行批量清洗操作。
2. 互联网应用数据采集	2.1 互联网应用数据的采集	2.1.1 能够编写采集客户端程序来进行互联网应用行为数据的动态采集。
	2.2 互联网应用数据清洗	2.2.1 能够按照行业要求，通过工具及编写数据校验规则提取有效数据，剔除脏数据。
3. 中间件数据采集	3.1 熟悉法规制度规范	3.1.1 熟悉关于信息数据采集的法律法规，保证在合法的情况下进行中间件容器日志采集。 3.1.2 掌握关于信息数据采集的国家标准，确保采集过程合规。
	3.2 中间件数据收集	3.2.1 熟悉 HTTP Web 服务容器、FTP 服务器等常见的容器日志生成方式、内容、格式。 3.2.2 能够运用采集工具对容器日志进行采集。
	3.3 中间件数据清洗处理	3.3.1 能够利用数据清洗工具编写基础的数据验证规则进行数据的合法性验证。 3.3.2 熟练掌握数据拆分规则，能够完成数据分解。
	3.4 中间件数据的存储	3.4.1 掌握数据表结构的设计能力。 3.4.2 掌握关系型数据库，将采集到的数据合理得存入数据库。
4. 操作系统日志数据采集	4.1 熟悉法规制度规范	4.1.1 熟悉关于信息数据采集的法律法规，保证在合法的情况下进行操作系统日志采集。 4.1.2 掌握关于信息数据采集的国家标准，确保采集过程合规性。

	4.2 操作系统数据收集	4.2.1 熟悉 Windows 等常见的操作系统日志生成方式。 4.2.2 能够识别操作系统中常见的操作日志种类（如登陆、系统级目录和内容修改）。 4.2.3 能够使用采集工具对操作系统日志进行采集。
	4.3 操作系统数据清洗处理	4.3.1 能够利用工具进行数据的合法性验证。 4.3.2 熟练掌握数据拆分规则，能够完成数据分解。 4.3.3 掌握操作系统的常见操作，能够对操作系统的操作日志数据进行初步提取。
	4.4 操作系统数据的存储	4.4.1 掌握数据表结构的设计能力。 4.4.2 熟练掌握关系型数据库，将采集到的数据合理得存入数据库。

表 3 数据采集职业技能等级要求（高级）

工作领域	工作任务	职业技能要求
1. 互联网应用数据采集	1.1 互联网应用数据的采集	1.1.1 能够在移动终端应用里编写数据采集插件来搜集应用数据。
2. 中间件数据采集	2.1 熟悉法规制度规范	2.1.1 熟悉行业内的技术标准和规范，能够准确得采集信息数据。
	2.2 中间件数据收集	2.2.1 熟悉代理服务容器、负载均衡服务等容器日志生成方式、内容、格式。 2.2.2 能够基于采集框架熟练运用编程语言对容器日志进行采集。
	2.3 中间件数据清洗处理	2.3.1 能够利用工具和编写数据验证规则进行数据的合法性验证。 2.3.2 熟练掌握数据拆分规则和方法，能够完成数据分解和归类。 2.3.3 掌握网络协议内容，能够对网络信息数据进行提取。
	2.4 中间件数据的存储	2.4.1 掌握数据模型关系设计能力，具备数据表结构的规划和设计能力。 2.4.2 具备非关系型数据库的使用能力，能够将采集到的海量数据存入到非关系型数据库。

3. 操作系统数据采集	3.1 熟悉法规制度规范	3.1.1 熟悉行业内的技术标准和规范，能够准确得采集信息数据。
	3.2 操作系统数据收集	3.2.1 熟悉 Linux 等常见的操作系统日志生成方式。 3.2.2 能够识别操作系统中不同类别下的操作日志种类、内容、格式。 3.2.3 能够基于采集框架熟练运用编程语言对操作系统日志进行采集。
	3.3 操作系统数据清洗处理	3.3.1 能够利用工具和编写数据验证规则进行数据的合法性验证。 3.3.2 熟练掌握数据拆分规则和方法，能够完成数据分解和归类。 3.3.3 掌握操作系统的基本机制，能够对操作系统的操作日志数据进行提取。
	3.4 操作系统数据的存储	3.4.1 掌握数据模型关系设计能力，具备数据表结构的规划和设计能力。 3.4.2 具备非结构化数据库的使用能力，能够将采集到的海量数据存入到非结构化数据库。
4. 硬件设备数据采集	4.1 熟悉法规制度规范	4.1.1 熟悉关于信息数据采集的法律法规，保证在合法的情况下进行硬件设备日志采集。 4.1.2 掌握关于信息数据采集的国家标准，确保采集过程合规。 4.1.3 熟悉行业内的技术标准和规范，能够准确得采集信息数据。
	4.2 硬件设备数据收集	6.2.1 熟悉网络分层协议、TCP/IP 规范等网络协议。 6.2.2 熟悉交换机、路由器、VPN 等网络设备的配置和使用，熟悉设备操作日志的生成方式。 6.2.3 能够识别不同设备下的流量数据日志种类、内容、格式。 6.2.3 能够使用硬件设备协议和数据信息协议基于采集框架熟练运用编程语言对设备日志进行采集。
	6.3 硬件设备数据清洗处理	6.3.1 能够利用工具和编写数据验证规则进行数据的合法性验证。 6.3.2 熟练掌握数据拆分规则和方法，能够完成数据分解和归类。 6.3.3 掌握硬件设备的使用方式和使用目的，能够对硬件设备的操作和流量日志数据

		进行合理提取。
	6.4 硬件设备数据的存储	<p>6.4.1 掌握数据模型关系设计能力，具备数据表结构的规划和设计能力。</p> <p>6.4.2 熟练掌握关系型数据库，将采集到的数据合理得存入数据库。</p> <p>6.4.3 具备非结构化数据库的使用能力，能够将采集到的海量数据存入到非结构化数据库。</p>

参考文献

- [1]中等职业学校专业目录（征求意见稿）
- [2]普通高等学校高等职业教育（专科）专业目录及专业简介
- [3]普通高等学校本科专业目录
- [4]中等职业学校专业教学标准
- [5]高等职业学校专业教学标准（2018年）
- [6]本科专业类教学质量国家标准
- [7]《计算机科学技术名词》第三版
- [8]GBT 5271.1-2000 信息技术 词汇 第1部分基本术语
- [9]GB/T 5271.17-2010 信息技术 词汇 第17部分：数据库
- [10]GB/T 35295-2017 信息技术 大数据 术语
- [11]GB / T 33847-2017 信息技术 中间件术语
- [12]GB/T 17143.6-1997 信息技术 开放系统互连 系统管理 第6部分:日志控制功能
- [13]GB/T 34840.3-2017 信息与文献 电子办公环境中文件管理原则与功能要求 第3部分:业务系统中文件管理指南与功能要求
- [14]GB-T 17143.5-1997 信息技术 开放系统互连 系统管理 第5部分事件报告管理功能
- [15]GB-T 17143.6-1997 信息技术 开放系统互连 系统管理 第6部分日志控制功能
- [16]GB-T 17143.8-1997 信息技术 开放系统互连 系统管理 第8部分安全审计跟踪功能

[17]GB / T 36625.1-2018 智慧城市 数据融合 第1部分概念模型

[18]GB / T 36625.2-2018 智慧城市 数据融合 第2部分数据编码规范

[19]GB / T 34960.5-2018 信息技术服务 治理 第5部分数据治理规范

[20]SJ/T 11615.1-2016 网络数据采集分析软件规范 第1部分：框架

[21]GB/T 22240-2008 信息安全技术 信息系统安全等级保护定级指南

[22]GB/T 25058-2010 信息安全技术 信息系统安全等级保护实施指南

[23]数据安全管理办法