

数据采集

职业技能等级标准

标准代码：510007

(2021年2.0版)

浪潮集团有限公司 制定

2021年12月 发布

目 次

| | |
|-------------|----|
| 前言 | 1 |
| 1 范围 | 2 |
| 2 规范性引用文件 | 2 |
| 3 术语和定义 | 2 |
| 4 适用院校专业 | 4 |
| 5 面向职业岗位（群） | 5 |
| 6 职业技能要求 | 6 |
| 参考文献 | 20 |

前 言

本标准按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本标准起草单位：浪潮集团有限公司、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：穆建平、刘何秀、陈天真、徐宏伟、单震、商程、崔瑞娟、刘洋、高敬萍、宁玉富、李明、徐翠娟、宋磊、许文宪、刘洪海、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

[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 适用院校专业

4.1 参照原版专业目录

中等职业学校：网络信息安全、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、移动应用技术与服务、电子商务。

高等职业学校：信息网络安全监察、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、商务数据分析与应用、人工智能技术服务、会计、会计信息管理、信息统计与分析、司法信息技术。

高等职业教育本科学校：计算机应用工程、网络工程、软件工程、大数据技术与应用、信息安全与管理。

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、信息管理与信息系统、大数据管理与应用、智能科学与技术。

4.2 参照新版职业教育专业目录

中等职业学校：网络信息安全、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、移动应用技术与服务、大数据技术应用、电子商务。

高等职业学校：网络安全与执法、云计算技术应用、大数据技术、计算机应用技术、计算机网络技术、软件技术、移动应用开发、信息安全技术应用、商务数据分析与应用、人工智能技术应用、大数据与会计、会计信息管理、统计与大数据分析、司法信息技术。

高等职业教育本科学校：计算机应用工程、网络工程技术、软件工程技术、大数据工程技术、信息安全与管理、人工智能工程技术、工业互联网技术。

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、信息管理与信息系统、大数据管理与应用、智能科学与技术。

5 面向职业岗位（群）

【数据采集】（初级）：主要面向政府及企事业单位、向大数据转型的互联网企业的客户服务、数据运营等相关职能部门及岗位，主要完成数据库数据、互联网应用数据、业务系统应用数据的数据采集、数据清洗、数据预处理、数据存储等大数据采集相关工作。

【数据采集】（中级）：主要面向政府及企事业单位、向大数据转型的互联网企业的系统开发、数据分析、客户服务、数据运营等相关职能部门及岗位，主要完成数据库数据、互联网应用数据、业务系统应用数据、容器日志数据、操作系统日志数据的数据采集、数据清洗、数据预处理、数据存储等大数据采集相关工作。

【数据采集】（高级）：主要面向政府及企事业单位、向大数据转型的互联网企业的系统开发、数据分析、客户服务、数据运营等相关职能部门及岗位，主要完成数据库数据、互联网应用数据、业务系统应用数据、容器日志数据、操作

系统日志数据、设备日志数据的数据采集、数据清洗、数据预处理、数据存储等大数据采集相关工作。

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、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 法规制度规范合规 | 1.1.1 能够熟悉并严格遵守关于数据库操作和数据采集的相关法律规范。 1.1.2 能够在合法的情况下进行数据库数据采集。 |

| | | |
|---------------|----------------|--|
| | | <p>1.1.3 能够参考关于数据库操作的国家标准，确保数据采集过程合规。</p> <p>1.1.4 能够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。</p> |
| | 1.2 数据库数据的收集 | <p>1.2.1 能够使用常规访问数据库的方式方法获取到目标数据库的相关信息。</p> <p>1.2.2 能够使用常用数据库客户端工具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，查询、筛选的准确率应能达到 100%。</p> <p>1.2.3 能实时完成数据的安全备份。能够对采集到的原始数据进行完整文件备份。</p> <p>1.2.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> <p>1.2.5 能够通过正规渠道从专业信息采集机构获取合法公开的信息数据，并且准确的解读出其中的数据内容。</p> |
| | 1.3 数据库数据的清洗 | <p>1.3.1 能够编写基础的数据验证规则，并结合数据清洗工具进行数据的合法性验证。</p> <p>1.3.2 能够使用正则表达式等数据拆分规则运用拆分方法完成数据分解。</p> <p>1.3.3 能够保证拆解的高准确度和精确度，能够对全部的输入数据进行成功拆解。</p> <p>1.3.4 对拆分的字段要有完整的文档进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 1.4 数据库采集数据的存储 | <p>1.4.1 能够完成常用关系型数据库的日常管理、运维任务。</p> <p>1.4.2 可以合理的规划、设计数据表结构。</p> <p>1.4.3 可以使用 SQL 语言准确完成数据存储工作。</p> <p>1.4.4 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。</p> <p>1.4.5 能够保证数据的完整性。存储的数据结构能够准确反映采集的数据含义。</p> |
| 2. 业务系统日志数据采集 | 2.1 法规制度规范合规 | <p>2.1.1 能够严格遵守关于业务系统日志数据采集的法律规范。</p> <p>2.1.2 能够在合法的情况下进行业务系统日志数据采集。</p> <p>2.1.3 掌握并参考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。</p> <p>2.1.4 能够根据行业技术标准和规范准确获取业务系统日志数据信息。</p> |

| | | |
|--------------|-----------------|---|
| | 2.2 业务系统日志数据的收集 | <p>2.2.1 能够明确业务系统的行为数据含义,根据数据采集常见流程使用常用的系统日志采集工具收集业务系统日志数据。</p> <p>2.2.2 能够根据业务系统部署信息找到业务系统日志所在的路径和文件。</p> <p>2.2.3 能实时完成数据的安全备份。能够对采集到的原始数据进行完整文件备份。</p> <p>2.2.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> <p>2.2.5 能够通过正规渠道从专业信息采集机构获取合法公开的信息数据,并且准确的解读出其中的数据内容。</p> |
| | 2.3 业务系统日志数据的清理 | <p>2.3.1 能够根据业务系统日志的组成规则快速获取、分析业务系统日志中可用数据信息。</p> <p>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。</p> <p>2.3.3 能够运用正则表达式等数据拆分规则完成数据分解。并保证拆解的高准确度和精确度,且对全部的输入数据进行成功拆解。</p> <p>2.3.4 对拆分的字段要有完整的文档进行描述,保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 2.4 业务系统日志数据的存储 | <p>2.4.1 能够根据常用的数据结构和业务系统的日志数据特点合理选择合适的数据结构,存储的数据结构需能够准确反映数据含义。</p> <p>2.4.2 能够各类文件存储格式特点将数据保存成不同类型文件。</p> <p>2.4.3 具备关系型数据库的管理、运维能力。</p> <p>2.4.4 具备将清理后的数据存储到目标数据库或数据表的能力。</p> <p>2.4.5 能够保证数据的完整性,不遗漏数据。</p> |
| 3. 互联网应用数据采集 | 3.1 法规制度规范合规 | <p>3.1.1 严格遵守国家《数据安全管理办法》及相关法律法规。</p> <p>3.1.2 能够在合法的情况下进行互联网应用数据采集。</p> <p>3.1.3 能够参考关于互联网数据采集的国家标准,确保数据采集过程合规。</p> <p>3.1.4 熟悉行业内的技术标准和规范,能够准确提取互联网应用的数据信息。</p> |
| | 3.2 互联网应用数据的采集 | <p>3.2.1 能够运用常见的网页分析工具(如浏览器里的调试工具等)来分析网页内容。</p> <p>3.2.2 能够根据网站型互联网应用的数据静态表现形式以及常见解网络协议,通过网络</p> |

| | | |
|--|----------------|--|
| | | <p>准确获取并解析互联网应用的数据。</p> <p>3.2.3 能够使用工具或编写程序获取网站等互联网应用的静态数据并进行数据抽取。</p> <p>3.2.4 能够通过正规渠道从专业信息采集机构获取合法公开的互联网应用数据，并且准确的解读出其中的数据内容。</p> |
| | 3.3 互联网应用数据的清洗 | <p>3.3.1 能够根据网页数据类型，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。</p> <p>3.3.2 具备脏数据的初步清洗能力，去除常见的无用数据。</p> <p>3.3.3 能够使用常用数据拆分规则完成数据分解。能够对网站等互联网应用的静态数据进行准确拆解。</p> <p>3.3.4 能够编写完整的文档对拆分的字段进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 3.4 互联网应用数据的存储 | <p>3.4.1 能够根据互联网应用数据的特点以及常用的数据结构的特点合理选择适当的数据结构。</p> <p>3.4.2 能够根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>3.4.3 能够将清理后的数据存储到目标数据库或数据表，并完成相关关系型数据库的管理和运维等日常工作。</p> <p>3.4.4 能够保证数据完整性，不遗漏数据。</p> |

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 | <p>1.1.1 能够通过访问包含分布式数据库在内的各类常用数据库获取到目标数据库的相关信息。</p> <p>1.1.2 能够使用各类客户端工具成功登录包含分布式数据库在内的目标数据库并进行相关数据的查询、筛选等数据收集操作。查询、筛选的准确率应能达到 100%。</p> <p>1.1.3 能实时完成数据的安全备份。能够对采集到的原始数据进行完整文件备份。</p> <p>1.1.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> |

| | | |
|---------------|-----------------|--|
| | 1.2 数据库数据的清洗处理 | <p>1.2.1 能够通过简单批处理程序的编写完成数据进行批量清洗操作。</p> <p>1.2.2 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。</p> <p>1.2.3 能够准确提取中间数据,根据常用的数据结构完成数据分解。保证拆解的高准确度和精确度,能够对全部的输入数据进行成功拆解。</p> <p>1.2.4 能够编写完整的文档对拆分的字段进行描述,保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 1.3 数据库采集数据的存储 | <p>1.3.1 能够完成常见数据库系统的日常管理、运维任务。</p> <p>1.3.2 可以合理的规划、设计数据库表结构。</p> <p>1.3.3 能够根据目标数据库的 SQL 特点,进行针对性的 SQL 优化。</p> <p>1.3.4 能够将清理后的数据存储到目标数据库或数据表的,并能够保证数据完整性,不遗漏数据。</p> |
| 2. 业务系统日志数据采集 | 2.1 业务系统日志数据的收集 | <p>2.1.1 能够使用常用的系统日志采集工具,允许使用主流的技术框架行为记录的日志框架,完成收集业务系统日志数据的任务。</p> <p>2.1.2 能够根据业务系统部署信息找到业务系统日志所在的路径和文件。</p> <p>2.1.3 能实时完成数据的安全备份。能够对采集到的原始数据进行完整文件备份。</p> <p>2.1.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> |
| | 2.2 业务系统日志数据的清理 | <p>2.2.1 能够根据业务系统日志的组成规则,快速获取、分析业务系统日志中特定数据信息。</p> <p>2.2.2 能够编写基础的数据验证规则,结合数据清洗工具,进行数据的合法性验证。</p> <p>2.2.3 能够使用正则表达式等数据拆分规则准确完成数据分解。</p> <p>2.2.4 能够在保证拆解的高准确度和精确度的前提下,对全部的输入数据进行成功拆解。</p> <p>2.2.5 能够编写完整的文档对拆分的字段进行描述,保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 2.3 业务系统日志数据的存储 | <p>2.3.1 能够根据常用的数据结构和业务系统的日志数据特点合理选择合适的数据结构,存储的数据结构需能够准确反映数据含义。</p> <p>2.3.2 能够各类文件存储格式特点将数据保</p> |

| | | |
|--------------|----------------|--|
| | | <p>存成不同类型文件。</p> <p>2.3.3 具备关系型数据库的管理、运维能力。</p> <p>2.3.4 具备将清理后的数据存储到目标数据库或数据表的能力。</p> <p>2.3.5 能够保证数据的完整性，不遗漏数据。</p> |
| 3. 互联网应用数据采集 | 3.1 互联网应用数据的采集 | <p>3.1.1 能够运用常见的网页分析工具(如浏览器、移动终端里的调试工具等)来分析网页内容。</p> <p>3.1.2 能够根据网站型互联网应用的数据静态及动态表现形式以及常见解网络协议、技术框架等，完成互联网数据的采集。</p> <p>3.1.3 能够针对不同类型互联网应用(网页、移动端应用等)的数据动态、静态的表现形式，通过网络准确解读和分析网络流量中的互联网应用数据。</p> <p>3.1.4 能够使用工具或编写程序获取网站等互联网应用的静态数据并进行数据抽取。</p> |
| | 3.2 互联网应用数据清洗 | <p>3.2.1 能够根据网页等数据类型，编写数据验证规则，进行数据的合法性验证。</p> <p>3.2.2 具备脏数据的剔除能力，准确清除无效数据。</p> <p>3.2.3 能够使用数据拆分规则完成数据分解。能够对网站等互联网应用的静态、动态数据进行准确拆解。</p> <p>3.2.4 能够编写完整的文档对拆分的字段进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 3.3 互联网应用数据的存储 | <p>3.3.1 能够独立完成数据表结构的设计。</p> <p>3.3.2 能够根据互联网应用数据的特点以及常用的数据结构的特点合理选择适当的数据结构。</p> <p>3.3.3 能够根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>3.3.4 能够独立完成关系型数据库的管理、运维，具备将清理后的数据存储到目标数据库或数据表的能力。</p> <p>3.3.5 能够保证数据的完整性，不遗漏数据。</p> |
| 4. 中间件数据采集 | 4.1 中间件数据收集 | <p>4.1.1 根据 HTTP Web 服务容器、FTP 服务容器等常见的容器日志生成方式、内容、格式。能够根据中间件容器的配置文件准确找到存放日志数据的目录和文件。</p> <p>4.1.2 能够运用采集工具对容器日志进行采集。采集工具对中间件容器的性能影响占比应低于 0.1%。</p> |

| | | |
|---------------|---------------|---|
| | | <p>4.1.3 能实时完成中间件容器的日志数据的安全备份，能够对采集到的原始数据进行完整文件备份。</p> <p>4.1.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> <p>4.1.5 能够通过正规渠道从专业信息采集机构获取合法公开的信息数据，并且准确的解读出其中的数据内容。</p> |
| | 4.2 中间件数据清洗处理 | <p>4.2.1 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。</p> <p>4.2.2 具备脏数据的清洗能力，去除常见的无用数据。</p> <p>4.2.3 能够使用正则表达式等数据拆分规则完成数据分解。对中间件日志数据的成功拆解率应能达到 99%。</p> <p>4.2.4 能够编写完整的文档对拆分的字段进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 4.3 中间件数据的存储 | <p>4.3.1 能够独立完成数据表结构的设计。</p> <p>4.3.2 能够根据互联网应用数据的特点以及常用的数据结构的特点合理选择适当的数据结构。</p> <p>4.3.3 能够根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>4.3.4 能够独立完成关系型数据库的管理、运维，具备将清理后的数据存储到目标数据库或数据表的能力。</p> <p>4.3.5 能够保证数据的完整性，不遗漏数据。</p> |
| 5. 操作系统日志数据采集 | 5.1 操作系统数据收集 | <p>5.1.1 能够根据 Windows/Linux 等常见的操作系统日志生成方式。识别用户在操作系统里进行的操作（如登录、系统级目录和内容修改）。</p> <p>5.1.2 能够使用采集工具对操作系统日志进行采集。</p> <p>5.1.3 能实时完成操作系统的日志数据的安全备份，能够对采集到的原始数据进行完整文件备份。</p> <p>5.1.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> <p>5.1.5 能够通过正规渠道从专业信息采集机构获取合法公开的信息数据，并且准确的解读出其中的数据内容。</p> |

| | | |
|--|----------------|---|
| | 5.2 操作系统数据清洗处理 | <p>5.2.1 能够根据操作系统的日志数据含义,对操作系统的操作日志数据进行初步提取。</p> <p>5.2.2 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。</p> <p>5.2.3 能够对数据进行过滤操作,根据采集目的从采集到的数据中准确选取所需数据。</p> <p>5.2.4 能够使用正则表达式等数据拆分规则,完成数据分解。对常见的操作系统日志数据成功拆解率应能达到90%。</p> <p>5.2.5 能够编写完整的文档对拆分的字段进行描述,保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 5.3 操作系统数据的存储 | <p>5.3.1 能够独立完成数据表结构的设计。</p> <p>5.3.2 能够根据互联网应用数据的特点以及常用的数据结构的特点合理选择适当的数据结构。</p> <p>5.3.3 能够根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>5.3.4 能够独立完成关系型数据库的管理、运维,具备将清理后的数据存储到目标数据库或数据表的能力。</p> <p>5.3.5 能够保证数据的完整性,不遗漏数据。</p> |

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. 数据库数据采集 | 1.1 数据库数据的收集 | <p>1.1.1 能够通过访问各种数据库的方式方法,获取到目标数据库的相关信息。</p> <p>1.1.2 能够通过编写程序、选择合适的数据库客户端工具,进行相关表数据的查询、筛选等数据收集操作。查询、筛选的准确率应能达到100%。</p> <p>1.1.3 能够通过常见数据库的行为数据记录方式(如日志方式记录)、日志数据的格式和种类等信息。定位到数据库的日志数据文件,并从中读取到数据库日志内容。</p> <p>1.1.4 能实时完成数据的安全备份。能够对采集到的原始数据进行完整文件备份。</p> <p>1.1.5 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> |
| | 1.2 数据库数据的清洗处理 | <p>1.2.1 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。</p> <p>1.2.2 能够熟练掌握各种类型的数据结构,准</p> |

| | | |
|----------------------|------------------------|--|
| | | <p>确提取中间数据，完成数据分解。</p> <p>1.2.3 能准确的分解数据库日志数据，提取数据库的行为属性信息。具备脏数据的剔除能力，准确的过滤掉无关数据。保证拆解的高准确度和精确度，能够对全部的输入数据进行成功拆解。</p> <p>1.2.4 能够编写完整的文档对拆分的字段进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | <p>1.3 数据库采集数据的存储</p> | <p>1.3.1 能够独立完成关系型、非关系等多种类型数据库的管理、运维任务。可以合理的规划、设计数据库表结构。</p> <p>1.3.2 能够掌握常用的数据结构，能够根据被采集的目标数据的特点合理选择适当的数据结构。</p> <p>1.3.3 能够根据各类文件存储格式特点，将数据保存成不同类型文件。</p> <p>1.3.4 具备将清理后的数据存储到目标数据库或数据表的能力。能够保证数据的完整性，不遗漏数据。</p> <p>1.3.5 具备各类数据库（包含非关系型数据库）的使用能力，能够将采集到的海量数据存入到数据库中。</p> |
| <p>2. 业务系统日志数据采集</p> | <p>2.1 业务系统日志数据的收集</p> | <p>2.1.1 能够根据数据采集的完整过程、业务系统的行为数据含义，使用主流开发语言和技术框架的日志组件实现机制完成各种日志采集工具的使用，收集业务系统日志数据。数据采集插件的运行，对宿主终端应用的性能影响比率应低于 1%。</p> <p>2.1.2 能够根据各种业务系统的部署信息找到业务系统日志所在的路径和文件。</p> <p>2.1.3 能实时完成数据的安全备份。能够对采集到的原始数据进行完整文件备份。</p> <p>2.1.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> |
| | <p>2.2 业务系统日志数据的清理</p> | <p>2.2.1 能够根据业务系统日志的组成规则，快速获取、分析业务系统日志中特定数据信息。</p> <p>2.2.2 能够编写各种场景下的数据验证规则，结合数据清洗工具，进行数据的合法性验证。</p> <p>2.2.3 能够正则表达式等各种数据拆分规则，完成复杂的数据分解。</p> <p>2.2.4 能够在保证拆解的高准确度和精确度的前提下，对全部的输入数据进行成功拆解。</p> |

| | | |
|--------------|-----------------|--|
| | | <p>具备脏数据的剔除能力，能够准确的过滤掉无关数据。</p> <p>2.2.5 能够编写完整的文档对拆分的字段进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 2.3 业务系统日志数据的存储 | <p>2.3.1 能够根据各种数据结构以及业务系统的日志数据特点合理选择恰当的数据结构，并对数据结构进行综合利用。</p> <p>2.3.2 能够各类文件存储格式特点将数据保存成不同类型文件。</p> <p>2.3.3 具备各类数据库的管理、运维能力。具备将清理后的数据存储到目标数据库或数据表的能力。</p> <p>2.3.4 能够保证数据的完整性，不遗漏数据。</p> <p>2.3.5 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。</p> |
| 3. 互联网应用数据采集 | 3.1 互联网应用数据的采集 | <p>3.1.1 能够不同类型互联网应用（网页、移动端应用等）的数据动态、静态的产生和表现形式。根据互联网应用的网络数据发生的全过程，通过网络准确获取并解析互联网应用的数据。</p> <p>3.1.2 能够使用工具或编写程序从外部捕捉网站、移动端应用等互联网系统的动态和静态数据，并进行数据抽取。</p> <p>3.1.3 能够编写采集客户端程序，将程序植入宿主系统，以埋点式进行互联网应用行为数据的动态采集。</p> <p>3.1.4 采集客户端的运行对宿主系统的性能影响占比应低于 0.1%。</p> |
| | 3.2 互联网应用数据清洗处理 | <p>3.2.1 能够根据网页、移动端应用等各种应用的数据类型，编写数据验证规则，进行数据的合法性验证。</p> <p>3.2.2 具备脏数据的剔除能力，准确清除无效数据。</p> <p>3.2.3 熟练掌握数据拆分规则，能够完成数据分解。能够对网站等互联网应用的静态、动态数据进行准确拆解。</p> <p>2.2.4 掌握互联网应用的特征内容，能够对应应用数据和交互数据进行提取。对互联网应用的交互数据的成功拆解率应能达到 90%。</p> <p>3.2.5 对拆分的字段要有完整的文档进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |

| | | |
|------------|----------------|--|
| | 3.3 互联网应用数据的存储 | <p>3.3.1 能够根据数据模型关系设计方法对数据表结构进行系统性的规划和设计。</p> <p>3.3.2 能够根据互联网应用数据的特点以及常用的数据结构的特点合理选择适当的数据结构。并根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>3.3.3 具备将清理的中间数据存储到目标数据库的能力。并且能够保证存储数据的完整性。</p> <p>3.3.4 掌握各类（包含关系型、非关系型）数据库的使用方法，能够将采集到的数据合理的存入数据库。能够将采集到的海量数据存入到非关系型数据库。</p> |
| 4. 中间件数据采集 | 4.1 中间件数据收集 | <p>4.1.1 能够对 HTTP Web 服务容器、FTP 服务容器、代理服务容器、负载均衡服务容器等容器日志生成方式、内容、格式有完整的认识。能够根据中间件容器的配置文件准确找到存放日志数据的目录和文件。</p> <p>4.1.2 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。采集工具对中间件容器的性能影响占比应低于 0.1%。</p> <p>4.1.3 能实时完成中间件容器的日志数据的安全备份，能够对采集到的原始数据进行完整文件备份。</p> <p>4.1.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> |
| | 4.2 中间件数据清洗处理 | <p>4.2.1 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。</p> <p>4.2.2 具备脏数据的清洗能力，去除常见的无用数据。</p> <p>4.2.3 熟练掌握正则表达式等数据拆分规则，并能够运用拆分方法完成数据分解。对中间件日志数据的成功拆解率应能达到 90%。</p> <p>4.2.4 对拆分的字段要有完整的文档进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 4.3 中间件数据的存储 | <p>4.3.1 能够根据数据模型关系设计方法对数据表结构进行系统性的规划和设计。</p> <p>4.3.2 能够根据中间件应用数据的特点以及常用的数据结构的特点合理选择适当的数据结构。并根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>4.3.3 具备关系型数据库的管理、运维能力。</p> |

| | | |
|-------------|----------------|--|
| | | <p>能够将清理后的数据存储到目标数据库或数据表。</p> <p>4.3.4 能够根据各类(包含关系型、非关系型)数据库的使用方法,能够将采集到的海量数据存入到非关系型数据库。</p> |
| 5. 操作系统数据采集 | 5.1 操作系统数据收集 | <p>5.1.1 能够根据 Windows/Linux 等操作系统基本原理及核心组件使用方法。识别操作系统用户、应用程序在操作系统里的操作所产生的日志数据种类、内容、格式等信息。</p> <p>5.1.2 能够使用采集工具对操作系统日志进行采集。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。</p> <p>5.1.3 能实时完成操作系统的日志数据的安全备份,能够对采集到的原始数据进行完整文件备份。</p> <p>5.1.4 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> |
| | 5.2 操作系统数据清洗处理 | <p>5.2.1 能够根据操作系统的基本运行机制,对操作系统的操作日志数据进行提取。并能够编写复杂的数据验证规则,结合各类数据清洗工具进行数据的合法性验证。</p> <p>5.2.2 能够正则表达式等数据拆分规则运用拆分方法完成复杂数据分解和归类。对常见的操作系统日志数据的成功拆解率应能达到90%。</p> <p>5.2.3 能够对数据进行过滤操作,根据采集目的从采集到的数据中准确选取所需数据。</p> <p>4.3.4 能够编写完整的文档对拆分的字段进行描述,保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 5.3 操作系统数据的存储 | <p>5.3.1 能够根据数据模型关系设计方法对数据表结构进行系统性的规划和设计。</p> <p>5.3.2 能够根据操作系统日志数据的特点以及常用的数据结构合理选择适当的数据结构。并根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>5.3.3 具备关系型数据库的管理、运维能力。具备将清理后的数据存储到目标数据库或数据表的能力。能够保证数据的完整性,不遗漏数据。</p> <p>5.3.4 能够根据各类(包含关系型、非关系型)数据库的使用方法,能够将采集到的海量数据存入到非关系型数据库。</p> |

| | | |
|-------------|----------------|--|
| 6. 网络设备数据采集 | 6.1 网络设备数据收集 | <p>6.1.1 能够基于采集框架熟练运用编程语言编写程序，完成对网络设备日志的采集。完成采集工作需能够掌握网络分层协议、TCP/IP 规范等网络协议、路由器等网络设备的配置和使用、设备操作日志的生成方式，并需要并能够对路由器等常见网络设备的流量数据种类、内容、格式进行解析。</p> <p>6.1.2 能实时完成网络设备的日志数据的安全备份，能够对采集到的原始数据进行完整文件备份。</p> <p>6.1.3 能保证数据的可追溯性。可以根据采集要求随时对备份的原始数据进行重新采集和核对。</p> <p>6.1.4 能够通过正规渠道从专业信息采集机构获取合法公开的信息数据，并且准确的解读出其中的数据内容。</p> |
| | 6.2 网络设备数据清洗处理 | <p>6.2.1 掌握网络设备流量数据的种类、格式等，理解数据的含义。</p> <p>6.2.2 能够根据网络设备流量数据的种类、格式等，对硬件设备的信息数据和流量日志数据进行合理提取。</p> <p>6.2.3 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。</p> <p>6.2.4 能够根据数据拆分规则和方法，完成数据分解和归类。对常见路由器日志数据的成功拆解率应能达到 90%。</p> <p>6.2.5 能够对数据进行过滤操作，根据采集目的从采集到的数据中准确选取所需数据。</p> <p>6.2.6 能够编写完整的文档对拆分的字段进行描述，保证拆分后的数据属性都有明确的意义和目的。</p> |
| | 6.3 网络设备数据的存储 | <p>6.3.1 能够根据数据模型关系设计方法对数据表结构进行系统性的规划和设计。</p> <p>6.3.2 能够根据互联网应用数据的特点以及各种数据结构合理选择适当的数据结构。并根据各类文件存储格式的特点将数据保存成不同类型文件。</p> <p>6.3.3 具备将清理的中间数据存储到目标数据库的能力。并且能够保证存储数据的完整性，不遗漏数据。</p> <p>6.3.4 能够独立完成各类数据库的管理、运维工作。能够将清理后的数据存储到目标数据库或数据表。能够将采集到的海量数据存入到非关系型数据库。</p> |

参考文献

- [1] 教育部关于印发《职业教育专业目录（2021年）》的通知（教职成〔2021〕2号）
- [2] 《教育部关于公布2019年度普通高等学校本科专业备案和审批结果的通知》（教高函〔2020〕2号）
- [3] 《教育部关于公布2020年度普通高等学校本科专业备案和审批结果的通知》（教高函〔2021〕1号）
- [4] 中等职业学校专业教学标准（2015年）
- [5] 高等职业学校专业教学标准（2018年）
- [6] 普通高等学校本科专业类教学质量国家标准（2018年）
- [7] 《计算机科学技术名词》第三版
- [8] GB/T 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 17143.8-1997 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

- [16] GB/T 36625.1-2018 智慧城市 数据融合 第1部分：概念模型
- [17] GB/T 36625.2-2018 智慧城市 数据融合 第2部分：数据编码规范
- [18] GB/T 34960.5-2018 信息技术服务 治理 第5部分：数据治理规范
- [19] SJ/T 11615.1-2016 网络数据采集分析软件规范 第1部分：框架
- [20] GB/T 22240-2008 信息安全技术 信息系统安全等级保护定级指南
- [21] GB/T 25058-2010 信息安全技术 信息系统安全等级保护实施指南
- [22] 中华人民共和国数据安全法(中华人民共和国主席令第八十四号)