

大数据工程化处理与应用

职业技能等级标准

标准代码：510062

（2021 年 2.0 版）

北京新奥时代科技有限责任公司 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.....	3
5 面向职业岗位（群）.....	4
6 职业技能要求.....	4
参考文献.....	14

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：北京新奥时代科技有限责任公司、工业和信息化部教育与考试中心、北京新大陆时代教育科技有限公司、中南大学、福建师范大学、北京信息职业技术学院、天津电子信息职业技术学院、福建信息职业技术学院、山西职业技术学院、贵州轻工职业技术学院、常州信息职业技术学院、重庆电子工程职业学院、烟台职业学院、北京市经济管理学校、浙江交通职业技术学院、新大陆数字技术股份有限公司、上海四道计算机科技有限公司、悦锦软件系统(上海)有限公司、创络(上海)数据科技有限公司、福建省国腾信息科技有限公司、中国人民人寿保险股份有限公司福建省分公司。

本标准主要起草人：谭志彬、邓立、王欣欣、张正球、聂强、武斌儒、秦华伟、睦碧霞、陈力捷、刘松、张海建、郭荣文、王作鹏、胡锦涛、姚志强、杨善仕、廖琼城、杨辰韵、张吉林、龚玉涵、龙九清、洪顺利。

声明：本标准的知识产权归属于北京新奥时代科技有限责任公司，未经北京新奥时代科技有限责任公司同意，不得印刷、销售。

1 范围

本标准规定了大数据工程化处理与应用职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于大数据工程化处理与应用职业技能培训、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 5271.6-2000 信息技术 词汇 第6部分：数据的准备与处理

GB/T 5271.9-2001 信息技术 词汇 第9部分：数据通信

GB/T 28821-2012 关系数据管理系统技术要求

3 术语与定义

国家、行业标准界定的以及下列术语适用于本标准。

3.1

结构化数据 structured data

一种数据表示形式，按此种形式，由数据元素汇集而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。

[GB/T 35295-2017，定义2.2.13]

3.2

非结构化数据 unstructured data

不具有预定义模型或未以预定义方式组织的数据。

[GB/T 35295-2017，定义2.1.25]

3.3

数据仓库 data warehouse

在数据准备之后用于永久性存储数据的数据库。

[GB/T 35295-2017，定义2.1.35]

3.4

抽取 to extract

从一组项中选择与取出符合某些特定准则的项。

[GB/T 5271.6-2000, 定义06.06.02]

3.5

数据源 data source

为传输提供数据的功能单元。

[GB/T 5271.9-2001, 定义09.01.03]

3.6

联机分析处理 online analytical processing; OLAP

对数据仓库中海量数据进行复杂、有效的分析处理,从而实现对决策的支持。

[GB/T 28821-2012, 定义3.5]

4 适用院校专业

4.1 参照原版专业目录

中等职业学校: 计算机应用、软件与信息服务、移动应用技术与服务、计算机网络技术、网站建设与管理、客户信息服务、通信运营服务、电子商务。

高等职业学校: 大数据技术与应用、云计算技术与应用、计算机应用技术、计算机网络技术、计算机信息管理、软件技术、软件与信息服务、信息统计与分析、人工智能技术服务。

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

应用型本科学校: 数据科学与大数据技术、大数据管理与应用、数据计算及应用、计算机科学与技术、信息管理与信息系统、软件工程、应用统计学、信息资源管理。

4.2 参照新版职业教育专业目录

中等职业学校: 大数据技术应用、计算机应用、软件与信息服务、移动应用技术与服务、计算机网络技术、网站建设与管理、客户信息服务、通信运营服务、电子商务专业。

高等职业学校: 大数据技术、云计算技术应用、计算机应用技术、计算机网络技术、软件技术、统计与大数据分析、人工智能技术应用专业。

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

应用型本科学校: 数据科学与大数据技术、大数据管理与应用、数据计算及应用、计算机科学与技术、信息管理与信息系统、软件工程、应用统计学、信息

资源管理。

5 面向职业岗位（群）

【大数据工程化处理与应用】（初级）：主要面向大数据行业、大数据处理与应用领域企事业单位数据采集标注工程师、大数据 ETL 工程师、大数据开发工程师等职业岗位，主要完成数据系统内项目文档及大数据处理、数据采集、数据抽取、数据转换、数据加载、数据系统应用等工作，从事数据采集与处理等基础技术工作。

【大数据工程化处理与应用】（中级）：主要面向大数据行业、大数据处理与应用领域企事业单位的大数据爬虫工程师、大数据 ETL 工程师、大数据开发工程师、大数据应用开发工程师、报表开发工程师等职业岗位，主要完成脚本程序编写、离线数据采集、网络数据采集、数据处理、数据查询和检索工作，从事脚本程序编写、离线和网络数据采集与处理等工作。

【大数据工程化处理与应用】（高级）：主要面向大数据行业、大数据处理与应用领域企事业单位及科研机构的大数据爬虫工程师、大数据 ETL 工程师、实时计算工程师、大数据开发工程师、大数据应用开发工程师、报表开发工程师、大数据分析师等职业岗位，主要完成网络爬虫架构设计、离线或实时数据采集、处理计算、开发与优化和数据检索等工作，从事网络爬虫架构设计与开发、离线或实时数据采集与处理、优化等工作。

6 职业技能要求

6.1 职业技能等级划分

大数据工程化处理与应用的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。

【大数据工程化处理与应用】（初级）：主要面向大数据行业、大数据处理与应用领域企事业单位的数据采集标注专员、初级大数据 ETL 工程师、大数据开发工程师助理等岗位，利用正在运行的数据系统，根据项目文档及大数据处理

流程，能使用图形化或可视化工具完成工程化的数据采集、数据抽取、数据转换、数据加载、数据系统应用等。

【大数据工程化处理与应用】（中级）：主要面向大数据行业、大数据处理与应用领域企事业单位的大数据爬虫工程师、大数据 ETL 工程师、大数据开发工程师、大数据应用开发工程师、报表开发工程师等岗位，能根据业务需求，利用各种数据系统工具，编写脚本程序完成工程化的离线或网络数据采集、数据处理、数据清洗等。

【大数据工程化处理与应用】（高级）：主要面向大数据行业、大数据处理与应用领域企事业单位及科研机构的大数据爬虫工程师、大数据 ETL 工程师、实时计算工程师、大数据开发工程师、大数据应用开发工程师、报表开发工程师、大数据分析师等岗位，根据系统需求，能使用框架编写程序完成工程化的网络爬虫架构设计、离线或实时数据采集、处理计算、开发与优化和数据检索等。

6.2 职业技能等级要求描述

表 1 大数据工程化处理与应用职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1.大数据工程化采集	1.1 网络数据采集	<p>1.1.1 根据网页标题采集需求，能使用可视化爬虫工具，获取目标标题数据。</p> <p>1.1.2 根据网页文本采集需求，能使用可视化爬虫工具，智能解析文本数据内容。</p> <p>1.1.3 根据数据标签采集需求，能使用可视化爬虫工具，智能解析标签数据。</p> <p>1.1.4 根据循环遍历采集需求，能使用可视化爬虫工具，实现智能翻页采集。</p>
	1.2 离线数据采集	<p>1.2.1 根据离线数据采集要求，能使用图形化工具，查询数据源的存储路径信息，确定要抽取的离线数据数量，获取数据源准确的位置信息。</p> <p>1.2.2 根据数据格式，能使用图形化工具，创建转换对象并配置相应结构化数据输入模块。</p> <p>1.2.3 根据目标数据系统类型及原始数据集格式，能使用图形化工具，配置相应数据输出模块。</p> <p>1.2.4 根据配置完成后的模块，能使用图形化工具，获得离线采集数据。</p> <p>1.2.5 根据数据内容，能使用图形化工具创建存储</p>

		采集数据的数据表结构。
	1.3 作业调度	<p>1.3.1 根据任务调度要求，能使用调度工具，创建作业的图形视图，调度所需要的模块，获取起始到完结所需的任务模块。</p> <p>1.3.2 根据所有转换模块，能使用调度工具，配置各模块的关联信息，获得符合数据传输流程的模块调度链。</p> <p>1.3.3 根据各模块的关联关系，能使用调度工具，对所有模块进行参数配置，获得数据连通的作业图形视图。</p> <p>1.3.4 根据配置完成的作业属性及任务调度需求，能使用调度工具执行作业，检查目标数据系统数据的入库情况，并获得入库信息反馈。</p>
2. 大数据工程化处理	2.1 数据清洗	<p>2.1.1 根据采集的原始数据集，能使用图形化工具，删除离线数据集中的缺省数据，获得完整的数据集。</p> <p>2.1.2 根据完整数据集，能使用图形化工具，简单标记、删除内容重复的离线数据，获得无干扰的数据集。</p> <p>2.1.3 根据无干扰数据集，能使用图形化工具，归一处理离线数据中不符合标准单位要求的字段，获得标准化数据集。</p> <p>2.1.4 根据标准化数据集，能使用图形化工具，标记、删除离线数据中不符合数据质量要求的数据，获得可用数据集。</p>
	2.2 数据计算	<p>2.2.1 根据数据清洗后的数据集，使用图形化工具，连接各数据表并根据处理需求进行简单数值计算、字符切分等处理，获得正确处理后的多表数据集。</p> <p>2.2.2 根据多表数据集，能使用图形化工具，连接、关联处理多表数据，获得关联整合数据集。</p> <p>2.2.3 根据关联整合数据集，能使用图形化工具，进行数据聚合、分组，获得关联计算数据集。</p>
	2.3 数据派生	<p>2.3.1 根据给定数据标签规则，能使用图形化工具，依据给定数据标签，创建各主题的标签库。</p> <p>2.3.2 根据不同主题的标签，能使用图形化工具，对关联计算数据集中的数据，进行标签派生及验证，获得含标签的数据集。</p> <p>2.3.3 根据已标签数据集，能使用图形化工具，根据主题划分数据表，创建业务主题数据表。</p> <p>2.3.4 根据主题数据表的格式规范，能使用图形化工具，对同主题数据集进行全字段的数据组织，获得符合业务主题的宽表。</p>

3. 大数据工程化应用	3.1 OLAP 系统应用	<p>3.1.1 根据给定数据表结构，能使用图形化工具，在指定项目中选择数据的加载方式，配置数据系统的数据表，获得符合业务分析需求的模型。</p> <p>3.1.2 根据配置完成的模型，能使用图形化工具，以查看表格的形式进行检查验证，完成模型的调试。</p> <p>3.1.3 根据调试后的模型，能使用图形化工具，创建并使用数据立方，编写简单查询语句进行数据查询，正确查询目标数据。</p>
	3.2 搜索系统应用	<p>3.2.1 根据给定数据结构，能使用图形化工具，关联分布式搜索组件，并执行脚本完成索引结构创建。</p> <p>3.2.2 根据索引结构，能使用图形化工具，将数据完整地导入到分布式搜索组件中，获得可全文检索的数据系统。</p> <p>3.2.3 根据搜索业务需求以及搜索组件中数据，能使用图形化工具，进行数据查询，正确查询目标数据。</p>
	3.3 报表系统应用	<p>3.3.1 根据报表模板，能使用合适的可视化工具，配置所需展示数据源的连接信息，正确连通数据系统。</p> <p>3.3.2 根据报表数据及离线数据需求，能使用可视化工具，设置数据透视表、图表等，生成符合需求的静态数据报表。</p> <p>3.3.3 根据数据报表，能使用可视化工具，利用数据报表的筛选、自定义排序、统计等功能，以指定的格式处理并存储数据报表结果。</p>

表 2 大数据工程化处理与应用职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
1. 大数据工程化采集	1.1 网络数据采集	<p>1.1.1 根据网络数据采集需求，能使用编程或爬虫框架，查询网络数据源信息，确定网络地址及网页格式，获得网页代码准确的标签信息。</p> <p>1.1.2 根据网页信息，能使用编程或爬虫框架，采集网络数据标签中相应数据并配置存储方式，并输出网络采集脚本。</p> <p>1.1.3 根据网络采集脚本及数据过滤需求，能使用编程或爬虫框架，编写网络采集过滤脚本，并输出网络采集数据。</p>
	1.2 离线数据采集	<p>1.2.1 根据离线数据采集需求，能使用脚本方式，确定目标离线数据源格式，获取准确的格式信息。</p> <p>1.2.2 根据数据格式，能使用脚本方式，配置数据采集脚本进行结构化或半结构化离线数据采集，获得</p>

		<p>离线采集数据并输出数据采集脚本。</p> <p>1.2.3 根据数据格式，能使用脚本方式，完成数据过滤脚本。</p> <p>1.2.4 根据编写的离线数据采集脚本，能使用脚本方式，测试离线数据采集脚本的采集、过滤、存储，获得目标数据并对脚本进行持久化操作。</p> <p>1.2.5 根据数据内容，能使用脚本创建存储采集数据的数据仓库结构并创建分区表及视图表。</p>
	1.3 作业调度	<p>1.3.1 根据任务调度需求，能使用调度工具，配置数据采集模块与数据采集流程，进行采集作业调度并检查采集情况，并获得采集结果信息。</p> <p>1.3.2 根据联合调度需求，能使用调度工具，上传多个脚本文件，获得可调度执行的作业文件。</p> <p>1.3.3 根据作业文件类型，能使用调度工具，编写作业调度脚本，配置定时、触发信息，获得作业调度脚本。</p> <p>1.3.4 根据编写的作业调度脚本，能使用调度工具，调试作业调度效果。</p>
2. 大数据工程化处理	2.1 数据清洗	<p>2.1.1 根据采集的原始数据集及完整性规则，能编写使用简单删除法或空值替代法处理离线数据集中缺省数据的脚本，获得完整的数据集。</p> <p>2.1.2 根据完整数据集及去重规则，能编写标记、删除离线数据集中重复字段的脚本，获得无干扰数据集。</p> <p>2.1.3 根据无干扰数据集及标准化规则，能编写归一处理离线数据集中不符合标准单位要求或给定结果集字段的脚本，获得标准化数据集。</p> <p>2.1.4 根据标准化数据集及可用性规则，能编写替换、标记或删除离线数据集中不符合数据质量要求数据的脚本，获得可用数据集。</p> <p>2.1.5 根据可用数据集，能编写去除离线数据中无关字段的脚本，获得有效数据集。</p>
	2.2 数据计算	<p>2.2.1 根据数据清洗后的数据集，能编写连接各数据表并根据处理需求进行数值计算、字符转换、时间计算等处理的脚本，获得正确处理后的多表数据集。</p> <p>2.2.2 根据多表数据集，能编写对多表数据进行字段合并、拆分等操作的脚本，获得字段对应的多表数据集。</p> <p>2.2.3 根据字段对应多表数据集，能编写连接、关联处理多表数据的脚本，获得关联整合数据集。</p> <p>2.2.4 根据关联整合数据集，能编写数据条件聚合、分组的脚本，获得关联计算数据集。</p>
	2.3 数据派生	<p>2.3.1 根据关联计算的数据及数据特征，能编写脚本进行数据聚合结果分析并设计数据标签，创建各主题标签库。</p>

		<p>2.3.2 根据不同主题的标签，能编写脚本，对关联计算数据集中的数据，编写标签计算脚本并进行标签派生，获得含属性标签的数据集。</p> <p>2.3.3 根据已标签数据集，能根据主题编写脚本划分数据表，创建业务主题数据表。</p> <p>2.3.4 根据主题数据表的业务需求，能使用脚本方式将同主题指标、维度、属性均关联的数据集进行数据组织，获得符合业务主题的宽表。</p>
3. 大数据工程化应用	3.1 OLAP 系统应用	<p>3.1.1 根据数据表结构及特征，能编写脚本，基于现有的事实表，创建维度表，获得符合分析需求的基础数据表结构。</p> <p>3.1.2 根据事实表以及维度表，在指定项目中选择数据的加载方式，配置数据仓库中的事实表以及维度表的字段信息，获得符合业务分析需求的模型。</p> <p>3.1.3 根据配置完成的模型，以查看表格的形式或可视化的方式进行模型验证，对模型进行修改和调试，获得修改后的模型。</p> <p>3.1.4 根据修改后的模型，创建并使用数据立方，编写条件查询语句进行数据查询，查询目标数据并存储。</p>
	3.2 搜索系统应用	<p>3.2.1 根据数据结构及组织方式，能编写脚本、REST 请求命令，编写相应索引、类型及文档，获得匹配的索引结构。</p> <p>3.2.2 根据索引结构，能使用图形化工具或 REST 请求命令，将数据完整导入到分布式搜索组件，获得可全文检索的数据系统。</p> <p>3.2.3 根据搜索业务需求以及搜索组件中数据，能使用图形化工具或 REST 请求命令，进行简单数据查询，正确查询目标数据并存储。</p>
	3.3 报表系统应用	<p>3.3.1 根据参考报表样式，能使用合适的可视化工具及脚本编程方法，配置所需展示离线数据源的连接信息，正确连通数据系统，并获得报表初始数据。</p> <p>3.3.2 根据报表数据及离线数据需求，能使用可视化工具及脚本编程方法，生成符合业务需求的可交互查询动态数据报表。</p> <p>3.3.3 根据数据报表，能使用可视化工具及脚本编程方法，能根据业务需求，完成可供决策支持的数据报表处理。</p>

表 3 大数据工程化处理与应用职业技能等级要求（高级）

工作领域	工作任务	职业技能要求
------	------	--------

1. 大数据工程化采集	1.1 网络数据采集	<p>1.1.1 根据网络数据采集需求，能使用编程或爬虫框架，查询网页提交方式及浏览器标识信息，提高网络采集脚本稳定性，获得网页代码准确的数据请求接口信息。</p> <p>1.1.2 根据数据格式，能使用编程或爬虫框架，进行网络数据采集，编辑网络采集脚本，测试数据爬虫脚本的采集、循环、过滤效果并迭代优化，并输出网络采集数据。</p> <p>1.1.3 根据数据格式，能使用编程或爬虫框架方式，遍历目标网站中的相应链接，使用合适算法过滤重复链接及规避身份验证，获得网络数据遍历脚本。</p> <p>1.1.4 根据网络遍历需求，能使用编程或爬虫框架方式，编辑完整的网络遍历脚本，并输出全站采集数据。</p>
	1.2 离线数据采集	<p>1.2.1 根据离线数据采集需求，能使用框架编程方式，确定要抽取的离线数据接口，获取数据源准确的连接信息。</p> <p>1.2.2 根据数据格式，能使用框架编程方式，配置数据采集模块进行各种离线数据采集（含非结构化离线数据），获得离线采集数据并输出数据采集程序类。</p> <p>1.2.3 根据数据格式，能使用框架编程方式，编写数据格式封装类，获得能够解析原始数据的方法。</p> <p>1.2.4 根据数据接口，能使用框架编程方式，进行数据过滤方法编写并配置其采集字段、过滤、存储规则，输出离线数据过滤方法。</p> <p>1.2.5 根据数据封装类，能使用框架编程方式，编写离线数据采集、过滤、存储方法，获得目标数据并输出数据采集程序。</p> <p>1.2.6 根据数据内容，能使用框架编程方式，创建存储采集数据的数据仓库结构并创建分桶、表约束及索引。</p>
	1.3 作业调度	<p>1.3.1 根据任务调度需求，能使用调度工具，通过数据采集作业并行调度，完成采集结果信息分析及进程日志监控。</p> <p>1.3.2 根据联合调度需求，能上传并调度执行作业文件、应用项目或自动化脚本。</p> <p>1.3.3 根据作业文件间依赖关系，能使用调度工具，配置作业流执行顺序，获得可执行作业调度流程。</p> <p>1.3.4 根据作业文件类型及作业混合调度需求，能使用调度工具，编写作业调度脚本，获得基于规则的作业流调度脚本。</p> <p>1.3.5 根据编写的作业流调度脚本，能使用调度工具，测试脚本并对作业执行性能进行优化。</p>
	1.4 实时数据采集	<p>1.4.1 根据实时数据采集需求，能使用图形化工具或脚本方式，确定日志数据的位置信息、来源方式等信息，获得实时数据的基本信息。</p>

		<p>1.4.2 根据数据采集需求，能使用实时数据采集组件，配置数据采集的来源、目标及采集方式，获取基础日志采集脚本。</p> <p>1.4.3 根据基础实时数据采集脚本，能使用图形化或脚本方式，编写实时数据拦截脚本拦截及过滤非需求之外的数据流，获得有效实时数据。</p> <p>1.4.4 根据实时数据采集脚本，能使用消息中间件，配置数据主题，以及生产者、消费者信息，获得数据缓存队列。</p>
2. 大数据工程化处理	2.1 数据清洗	<p>2.1.1 根据采集的原始数据集及完整性需求，能使用框架编程开发使用特殊值填充、均值填充离线或实时数据的缺省数据的程序，获得完整的数据集。</p> <p>2.1.2 根据完整数据集及去重需求，能使用框架编程方式开发替换、标记或删除离线或实时数据中特定字段或时间戳的重复数据的程序，获得无干扰数据集。</p> <p>2.1.3 根据无干扰数据集及标准化需求，能使用框架编程开发归一处理离线或实时数据中不符合标准单位要求或给定结果集字段的程序，获得标准化数据集。</p> <p>2.1.4 根据标准化数据集及可用性需求，能使用框架编程方式开发替换、标记或删除不符合数据质量要求的离线或实时数据的程序，获得可用数据集。</p> <p>2.1.5 根据可用数据集，能使用框架编程方式开发去除离线或实时数据中无关字段的程序，获得有效数据集。</p>
	2.2 数据计算	<p>2.2.1 根据数据聚合计算需求，能使用框架编程方式开发自定义数据处理函数，获得自定义数据处理方法。</p> <p>2.2.2 根据数据清洗后的数据集，能使用框架编程方式开发连接各数据表并根据处理需求进行数值计算、字符转换、时间计算、键值化、离散化等复杂处理的程序，获得正确处理后的多表数据集。</p> <p>2.2.3 根据多表数据集及多表关联需求，能使用框架编程方式开发对多表数据进行字段合并、拆分、归类等操作的程序，获得字段对应的多表数据集。</p> <p>2.2.4 根据字段对应多表数据集及多表关联需求，能使用框架编程方式开发连接、关联处理多表数据的程序，获得关联整合数据集。</p> <p>2.2.5 根据关联整合数据集，能使用框架编程方式开发条件聚合分组的程序，获得关联计算数据集。</p>
	2.3 数据派生	<p>2.3.1 根据关联计算的数据及数据标签需求，能使用框架编程方式开发程序进行数据聚合结果分析并设计数据标签，创建各主题标签库。</p> <p>2.3.2 根据不同主题的标签，能使用框架编程方式，对关联计算数据集中的数据，开发标签计算程序并对离线或实时数据进行标签派生，获得含属性标签的数据集。</p>

		<p>2.3.3 根据已标签数据集及规划设计需求,能根据主题使用框架编程方式开发程序对数据表进行划分,创建业务主题数据表。</p> <p>2.3.4 根据主题数据表的业务需求,能使用框架编程方式开发程序将同主题指标、维度、属性均关联的数据集进行数据组织或进行星型、雪花模型、宽表等表结构的创建,获得符合业务主题的模式格式。</p>
3. 大数据 工程化应用	3.1 OLAP 系统应用	<p>3.1.1 根据业务需求,能使用框架编程方式,创建对应的事实表以及维度表,获得符合分析需求的基础数据表结构。</p> <p>3.1.2 根据事实表以及维度表,在指定项目中选择数据的加载方式,配置数据仓库中的事实表和维度表以及关联条件、度量指标、分区字段,获得所需的数据模型。</p> <p>3.1.3 根据数据模型,以查看表格、可视化和 JSON 的方式进行模型验证,对模型进行修改及性能优化,获得优化后的模型。</p> <p>3.1.4 根据优化后的模型,创建指定参数的数据立方或二维表,编写复合查询语句进行即席查询,正确查询目标数据并存储。</p> <p>3.1.5 根据数据集成需求,能使用框架编程方式,编写 OLAP 系统的数据访问接口,获得可供其他数据系统访问的 API。</p>
	3.2 搜索系统应用	<p>3.2.1 根据业务需求,能使用框架编程方式,关联分布式搜索组件,创建索引、类型及文档,获得结构化或非结构化的索引结构。</p> <p>3.2.2 根据索引结构,能使用框架接口开发方式,将数据完整地导入及同步到分布式搜索组件中,获得可全文检索的数据系统。</p> <p>3.2.3 根据搜索业务需求以及搜索组件中数据,能使用框架接口开发方式,进行复合条件查询,正确查询目标数据并存储。</p> <p>3.2.4 根据数据集成需求,能使用框架编程方式,编写搜索引擎的数据检索接口,获得可供其他数据系统访问的 API。</p>
	3.3 报表系统应用	<p>3.3.1 根据报表展示需求,能使用框架编程方式,通过数据接口获得报表初始数据。</p> <p>3.3.2 根据报表数据及离线或实时数据需求,能使用框架编程方式,生成符合业务需求的可交互查询实时动态数据报表。</p> <p>3.3.3 根据离线或实时报表数据,能使用框架编程方式,编写前端查询和展示页,设置数据展示格式及功能,生成可交互查询、符合业务需求的网页数据报表。</p> <p>3.3.4 根据数据报表,能使用框架编程方式,获得供决策支持的数据报表并分析关键指标情况。</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》
- [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》
- [4] GB/T 35295-2017 信息技术 大数据 术语
- [5] GB/T 35589-2017 信息技术 大数据 技术参考模型
- [6] GB/T 37721-2019 信息技术 大数据分析系统功能要求
- [7] GB/T 37722-2019 信息技术 大数据存储与处理系统功能要求
- [8] GB/T 18714-2002 信息技术 开放分布式处理 参考模型
- [9] GB/T 32630-2016 非结构化数据管理系统技术要求