

浙江省地方标准

DB33/T 1354.1—2024

产业数据仓
第1部分：总体框架和要求

Industry data warehouse—
Part1:General framework and technical requirements

2024-01-04 发布

2024-02-04 实施

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 总体框架 2

 5.1 逻辑架构 2

 5.2 技术架构 2

6 能力要求 3

 6.1 数据资源编目 3

 6.2 数据采集 4

 6.3 数据交换共享 4

 6.4 数据治理 4

 6.5 数据融合 5

 6.6 数据存储 5

 6.7 数据管理 5

 6.8 数据质量评价 6

7 服务要求 6

 7.1 产业主题库 6

 7.2 产业专题库 8

8 对接要求 8

 8.1 数据目录对接 8

 8.2 用户体系对接 9

 8.3 系统日志对接 9

 8.4 功能模块对接 9

9 安全体系建设 9

 9.1 总体要求 9

 9.2 数据分类分级要求 9

 9.3 系统管理 9

 9.4 数据运维安全要求 10

附录 A（资料性） 各级各类数仓建设内容参考 11

附录 B（资料性） 产业数据仓数据分层架构参考 12

附录 C（资料性） 业务专题库建设示例（企业诉求专题库） 13

附录 D（资料性） 行业专题库建设示例（化工行业专题库） 15

前 言

本部分按照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。

本部分是DB33/T 1354的第1部分。DB33/T 1354《产业数据仓》已经发布了以下部分：

——第1部分：总体框架和要求。

——第2部分：数据资源编目规范。

请注意本部分的某些内容可能涉及专利。本部分的发布机构不承担识别专利的责任。

本部分由浙江省经济和信息化厅提出、归口并组织实施。

本部分起草单位：浙江省数字经济发展中心、浙江省电子信息产品检验研究院、浙江经信信息技术中心有限公司、杭州数梦工场科技有限公司、杭州数政科技有限公司、浙江省智慧城市促进会。

本部分主要起草人：张君、田燕翔、陆澄澹、汪越、李晓波、季永炜、徐强、念灿华、王宁、胡光剑、廖文睿、黄益超。

产业数据仓

第 1 部分：总体框架和要求

1 范围

本部分规定了产业数据仓的总体框架、能力要求、服务要求、对接要求、安全体系建设等内容。
本部分适用于行政产业数据仓、行业数据仓和企业数据仓的规划建设。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本部分必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本部分；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本部分。

GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求

3 术语和定义

下列术语和定义适用于本部分。

3.1

数据仓 data warehouse

业务过程中收集、产生的，随时间动态更新，信息类型、格式等相对稳定的各类公共数据资源集合。
[来源：DB33/T 2350—2022，3.2.3.9]

3.2

产业数据仓 industry data warehouse

有效开展各种经济活动的实体在组织生产、经营或业务活动的过程中收集、产生的，随时间动态更新，信息类型、格式等相对稳定的各类数据资源集合。

3.3

数据治理 data governance

对数据进行处置、格式化和规范化的过程。

注1：数据治理是数据和数据系统管理的基本要素。

注2：数据治理涉及数据全生存周期管理，无论数据是处于静态、动态、未完成状态还是交易状态。

[来源：GB/T 35295—2017，2.1.43]

3.4

数据融合 data fusion

基于一组或多组数据，通过一定的处理过程以获得新的或更高质量信息的过程。

[来源：GB/T 33745—2017，2.5.6]

4 缩略语

下列缩略语适用于本部分。

API：应用程序编程接口（Application Programming Interface）

ETL：业务系统的数据经过抽取、清洗转换之后加载到数据仓库的过程（Extract-Transform-Load）

5 总体框架

5.1 逻辑架构

产业数据仓是由行政产业数据仓（以下简称“行政数仓”）、行业数据仓（以下简称“行业数仓”）、企业数据仓（以下简称“企业数仓”）构成的逻辑集合。政企两侧数据仓之间宜以应用需求为导向，通过数据安全传输通道开展数据治理与融合加工，使数据资源发挥更大价值。产业数据仓逻辑架构见图1。具体如下：

- a) 行政数仓：依托政务云资源，实现政府侧产业数据的统一管理，可分省市两级节点或省市县三级节点；汇聚政府侧相关应用系统共享的产业数据，并按需采集行业和企业自愿授权公开的产业相关数据；
- b) 行业数仓：以特定组织为主体建设运营，实现本行业的数据采集、治理和融合应用；汇聚本行业范围内企业基本信息数据、产业链节点数据、产品库数据、供应链数据、政府部门授权开放的公共数据、第三方数据服务机构数据等；
- c) 企业数仓：以企业为主体建设运营，实现企业内的数据采集治理、融合分析、挖掘应用；汇聚企业自身采购、生产、经营等相关数据及行业产业链上下游业务数据，并可按需向行业数仓申请可共享数据。

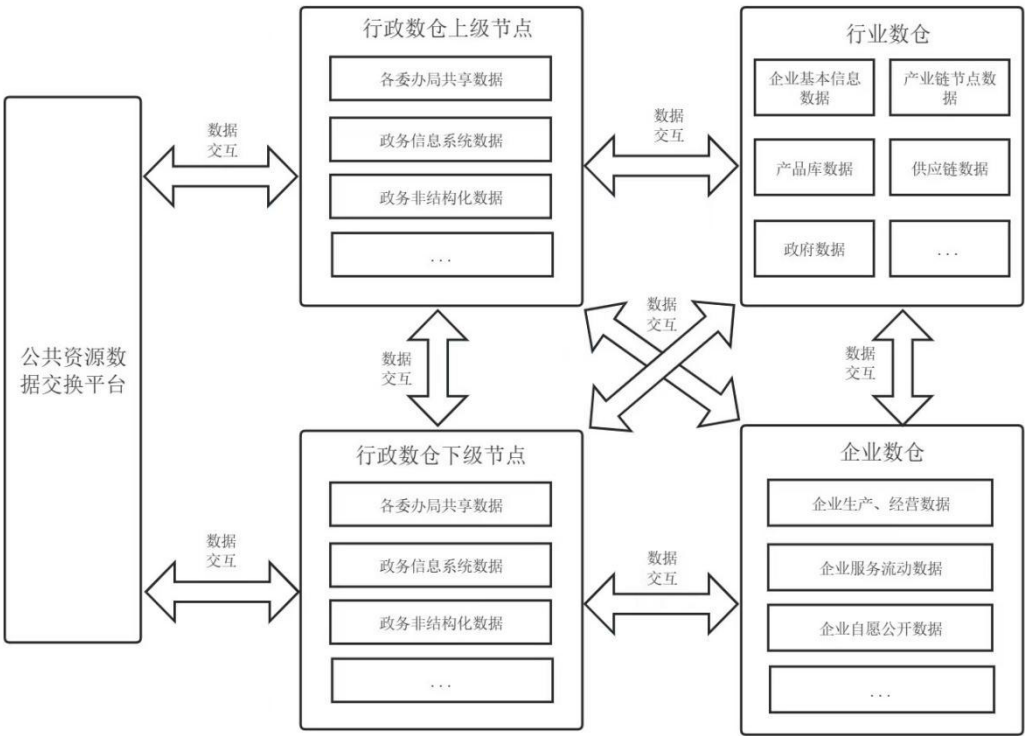


图 1 产业数据仓逻辑架构图

5.2 技术架构

产业数据仓基于网络资源、算力资源和感知资源等基础设施服务，构建数据支撑层、数据能力层、数据服务层，为各类应用场景提供服务支撑。产业数据仓技术架构见图2，具体如下：

- a) 数据支撑层：数据支撑层由各种途径采集汇聚形成的不同结构的数据源组成，按产业结构分类分为工业数据、农业数据、服务业数据，按数据来源分为政府侧数据、行业侧数据、企业数据、第三方数据服务机构数据等；
- b) 数据能力层：数据能力层是支撑产业数据处理的核心层，可具备数据采集、传输、存储、处理、交换、销毁等数据生命周期各阶段的支撑能力，包括数据资源编目、数据采集、数据交换共享、数据治理、数据融合、数据存储、数据管理、数据质量评价等功能；
- c) 数据服务层：数据服务层是基于数据仓汇聚的原始数据，通过数据治理、融合，构建形成主题库和专题库等，同时汇聚各类知识、规则、模型、算法，为各类应用场景提供服务支撑。

各级数仓建设内容参考相关示例见附录 A，产业数据仓数据分层架构相关示例见附录 B。

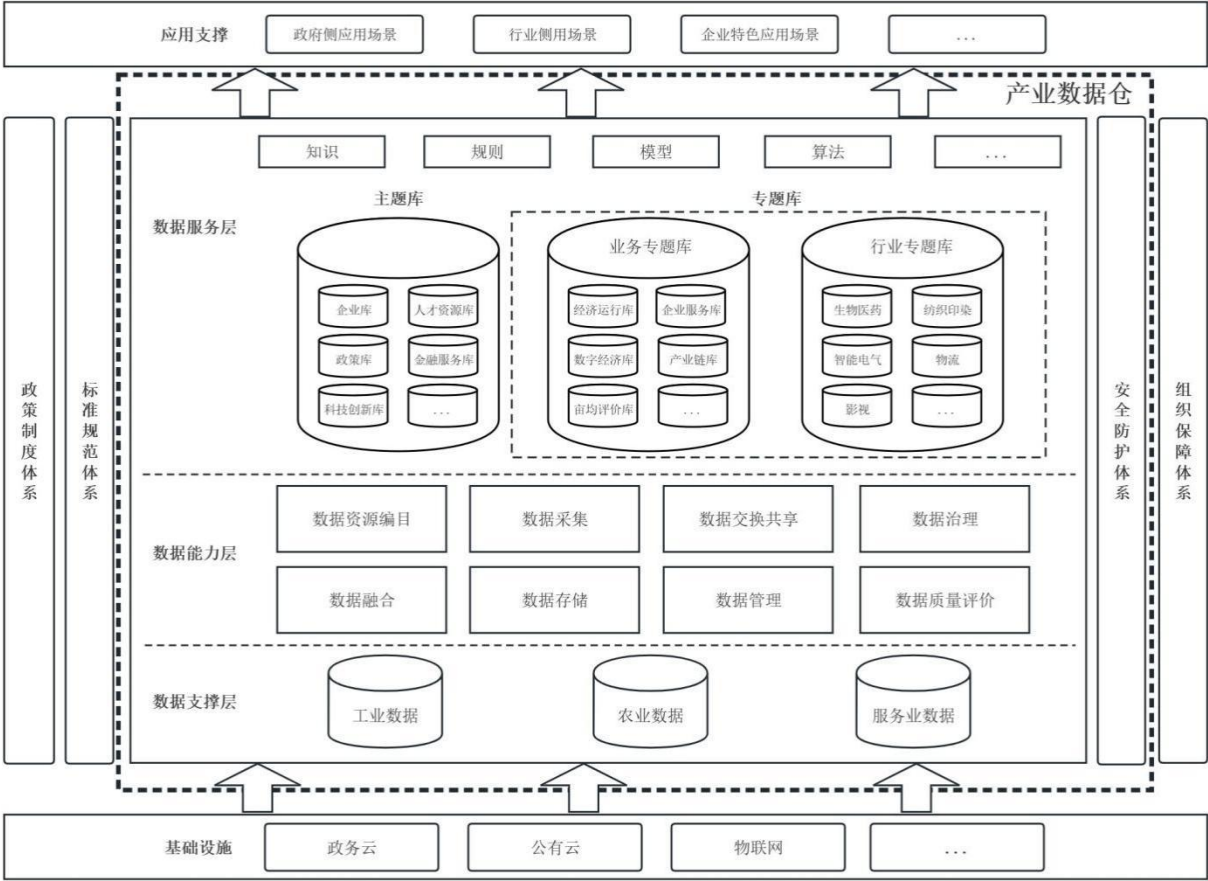


图 2 产业数据仓技术架构图

6 能力要求

6.1 数据资源编目

产业数据仓可具备数据资源编目功能，实现数据资源目录编制与管理，构建数据仓内的数据表与目录的对应关系，实现信息资源一本总账，为产业大脑、数字经济门户及各业务系统等提供信息资源清单。

数据资源编目可具备应用系统注册、目录编制、目录审核与发布、资源挂载、目录查询、目录变更与停用等功能，并构建产业数据资源标识符编码规则，进行统一编码。信息资源目录的内容包括但不限于信息资源的名称、数源单位、数据分类、数据格式、更新频率、重点领域分类、行业分类、行业代码、行业内企业共享属性、跨行业企业共享属性、政府侧共享属性、数据安全级别、开放属性及数据项等元数据信息。

6.2 数据采集

产业数据仓应具备数据采集功能，从不同结构的数据源中抽取数据，经过数据转换，加载到目标数据库，满足完整性、有效性、及时性、稳定性的要求。

数据采集应支持ETL的全过程管理，具备数据源管理、数据抽取、数据转换、采集任务调度、元数据管理、采集监控、数据稽查、日志审计等功能；可支持关系型数据库、半结构化数据、分布式文件数据、对象存储、流式数据、时序数据库、分析型数据库等多种类型数据源的采集。

6.3 数据交换共享

6.3.1 概述

产业数据仓应具备数据交换共享功能，通过构建覆盖多级政务部门间、企业侧各主体间的数据交换体系，实现异构应用系统间松耦合的数据交换共享，满足政企两侧的数据融合需要；涉及跨网的数据交互，应满足第9章安全体系建设要求。数据交换共享应支持但不限于以下方式：库表数据交换、接口数据交换、文件数据交换等。

6.3.2 库表数据交换

库表数据交换应由交换桥接、前置交换、交换传输、交换管理等组成，通过交换桥接将需要交换的信息交换到前置数据库，在交换管理的流程控制下，通过交换传输、前置交换，把需要交换的信息定向传输到接收单位。

6.3.3 接口数据交换

接口数据交换应具备数据服务化的封装与管理、统一的数据接口服务，以供应用系统调用数据。接口数据交换应具备接口注册发布、接口申请审批、接口管理、接口服务调用、接口鉴权、接口运行监测、日志审计、接口文档管理等功能，实现对接口全生命周期进行管理；支撑常用的接口协议和网络协议。

6.3.4 文件数据交换

文件数据交换应具备数据统一管理、数据备份、文件实时同步、版本管理、文件下载等功能。

6.4 数据治理

6.4.1 概述

产业数据仓宜具备数据治理功能，建立各行业的数据质量标准规范，实现对海量数据的清洗、转换和校核，提升产业数据仓的整体数据质量。

数据治理宜具备数据标准化和数据清洗等功能。

6.4.2 数据标准化

数据标准化宜具备标准创建、标准修改、标准删除、标准查询、标准详情查看、标准导入导出等功能；并实现标准质量检查、标准版本管理、标准模板管理、标准变更订阅。数据标准信息包括但不限于数据标准代码、数据标准名称、数据标准类型、长度、精度、数据标准描述等。

6.4.3 数据清洗

数据清洗宜具备清洗规则自定义与配置、数据过滤、数据去重、数据补全、数据转化、数据降维、清洗任务管理等功能。清洗规则包括但不限于特定字符、大小写转换、汉语转拼音、同义词替换、身份证校验等。

6.5 数据融合

6.5.1 概述

产业数据仓应具备数据融合与数据分析功能，通过数据建模、数据标签、数据血缘等工具，开发逻辑模型、物理模型，形成主题库、专题库，提供数据服务。

6.5.2 数据建模

数据建模宜提供以逻辑模型创建物理模型功能，使物理模型继承逻辑模型定义的数据结构、约束和数据关系等。逻辑模型宜支持按照业务拆分主题域，并进行主题域内的数据逻辑结构、约束和数据关系设计，实现业务模型定义；物理模型包含模型物化和模型加工两部分功能，模型物化指创建物理模型定义的表结构，模型加工指在线定义模型的加工逻辑，支持垂直拆分、水平拆分、多表联合、多表连接或自定义映射等方式。

6.5.3 数据标签

产业数据仓可构建数据标签管理体系，开展数据标签规则建立、标签开发、标签生命周期管理，形成数据标签库，提供数据分类打标和分级打标等服务。数据标签宜具备标签地图、标签申请、标签数据预览、标签元数据查询、标签应用等功能，支持指标统计分析、标签的开发与运维、标签使用情况监测。

6.5.4 数据血缘

数据血缘宜提供数据血缘分析和展示，支持数据源库到目标数据库之间流向查看，库级、表级和字段级血缘分析，数据血缘信息展示。

基于数据血缘信息，宜提供溯源分析、影响分析、关联程度分析和数据时效性分析等元数据应用分析功能，支持数据流动趋势、数据库间依赖关系和数量的展示。

6.6 数据存储

产业数据仓应具备数据存储功能，支持基于关系型数据库和非关系型数据库的数据存储。非关系型数据库可包括键值存储数据库、列存储数据库、图形数据库和文档数据库等。

数据存储宜具备数据存储节点和组件动态伸缩、数据冗余存储、数据备份恢复等功能，可根据实际需要，选择合适的数据存储方式构建。

6.7 数据管理

产业数据仓应具备数据管理功能，构建统一的数据资源授权体系，实现数据入仓到出仓的全流程管控，为信息资源合规、有序流转提供保障。数据管理应具备目录检索、数据申请、数据需求填报、流程审批、待办提醒等功能，支持系统的运行监测、事项统计、日志记录。

6.8 数据质量评价

数据质量评价应具备数据规则管理、数据质量评价、质量报告生成、问题数据工单管理、数据质量监控等功能，实现问题数据的发现、反馈和修正。

7 服务要求

7.1 产业主题库

7.1.1 概述

产业数据仓可建设产业主题库。通过数据融合对产业数据进行主题分析，形成产业主题库并提供相关主题库服务。

产业主题库的建设可构建概念体系，完成数据资源的加工、整理和增建，提供统一的主题库服务。产业主题库可包括但不限于企业库、人才资源库、政策库、科技创新库、金融服务库等。

产业主题库建设应满足但不限于以下要求：

- a) 产业主题库应具有合理的概念体系，能够正确反应该主题库领域内数据的相关知识及知识之间的关系，基于该体系归纳出结构合理、层次清晰的多级类目，并通过数据标签功能，对所形成的类目进行维护，将概念体系固化至数据仓中；
- b) 根据数据资源内容特征及内容特征之间的关系将业务库合理地组织成若干逻辑数据库，基于逻辑数据库的数据模型实现资源的加工、整合和增建；
- c) 将主题数据库进行数据资源目录编制，供产业数据仓用户申请调用，提供统一的主题库服务。

7.1.2 企业库

企业库是以企业为主体的产业主题库类型，围绕构建现代产业体系目标，归纳为符合数字经济产业特色的三级类目概念体系。

一级类目包括但不限于基本信息、资本资产、生产经营、荣誉资质、信用评价和行政执法等6个类别，涵盖企业全生命周期的相关活动所产生的信息。

一级类目下包含若干二级类目，内容包括但不限于：

- 基本信息：登记信息、存续状态、企业类型、股东信息、场所信息、其他基本信息等；
- 资本资产：企业规模、资本信息、知识产权、人才信息、其他资本资产等；
- 生产经营：经营范围、营收财务、能耗水平、项目信息、商品服务、研发情况、亩均生产、税费缴纳、其他生产经营等；
- 荣誉资质：资质信息、荣誉信息、其他荣誉资质等；
- 信用评价：信用监管、社会评价、其他信用评价等；
- 行政执法：安全监管、行政处罚、司法处罚、其他行政执法等。

三级类目可根据实际业务需求进行维护。

7.1.3 人才资源库

人才资源库是以产业人才为主体的产业主题库类型，主要汇集在重点产业及各类优势产业、战略性新兴产业、未来产业中取得重大突破、解决重大技术难题的产业人才信息，提升人才资源信息的服务能力。归纳为三级类目概念体系。

一级类目包括但不限于人才基本信息、人才评定、人才教育、人才工作、资质荣誉、人才信用、人才政策、人才技能评价等8个类别。

一级类目下包含若干二级类目，内容包括但不限于：

- 人才基本信息：基础身份信息、社保情况、家庭成员、住房情况、婚姻状况等；
- 人才评定：国内顶尖人才、国家级领军人才、省级领军人才、市级领军人才、高级人才等；
- 人才教育：学籍信息、学历登记信息、培训信息、考试信息、从业资质信息等；
- 人才工作：当前工作信息、历史履历信息、失业登记信息、工作技能信息、工作奖惩信息等；
- 资质荣誉：资质证书、荣誉证书、职称信息、专利信息、奖励信息等；
- 人才信用：处罚信息、失信信息、违法信息、欠税漏税信息、违纪信息等；
- 人才政策：人才激励政策、人才补贴政策、人才减免政策、职称评定政策、人才认定政策等；
- 人才技能评价：评价机构、人才证书、标准题库、课题研究、技能补贴等。

三级类目可根据实际业务需求进行维护。

7.1.4 政策库

政策库是以政策信息为主体的产业主题库类型，汇集政策文件中服务产业、行业、企业的政策信息，归纳为基于企业服务的二级类目概念体系，为政策精准推送、线上直兑等业务应用提供政策数据支撑。

一级类目包括但不限于税收优惠、财政支持、费金减免、降低企业制度性交易成本、降低企业人工成本、降低企业融资成本、降低企业用能用地成本等7个类别。

二级类目可根据实际业务需求进行维护。

各类惠企政策的字段包括但不限于发布层级、政策标题、责任部门、责任处室、政策原文、政策文号、政策要点、政策解读、支持对象、适用行业、开始时间、结束时间、印发日期、是否只适用出口企业等。

7.1.5 科技创新库

科技创新库是以科技创新资源为主体的产业主题库类型，汇集优质创新资源的科创要素数据，归纳为基于科创要素的二级类目概念体系，为匹配科创资源、核心技术攻关等业务应用提供科创数据支持。

一级类目包括但不限于科技创新项目、创新机构、人力资源、专利、文献、成果等6个类别。

二级类目可根据实际业务需求进行维护。

各类科技创新资源的字段包括但不限于名称、内容、技术领域、所属行业、所处阶段、项目金额、发布时间、投产时间、达产时间、成果质量、商用规模、应用成效等。

7.1.6 金融服务库

金融服务库是以涉企金融服务为主体的产业主题库类型，内容涵盖投资、储蓄、信贷、结算、证券买卖、商业保险和金融信息咨询等多个方面，包括但不限于银行、证券公司、保险公司、服务机构、投资产品、信贷产品、证券、商业保险、融资需求企业、融资需求情况、企业授信情况等数据。

一级类目包括但不限于投资、储蓄、信贷、结算、证券买卖、商业保险和金融信息咨询等7个类别。

一级类目下包含若干二级类目，内容包括但不限于：

- 投资数据：投资方、交易方、统计日期、投资国家（地区）、投资事件、投资金额、持股比例、投资方所属行业、交易方所属行业等；
- 储蓄数据：城乡居民定期储蓄额、城乡居民活期储蓄额、居民储蓄率等；
- 信贷数据：信贷申请数据、放款数据、还款数据、逾期数据等；
- 结算数据：结算方式、支持银行、结算周期、结算时间、结算费用、结算处理等；
- 证券买卖：委托比例、量比、内外盘数据、总成交量、换手率、总股本、流通股本、净资产、市盈率、收益、净流入额、大宗流入、所属板块概念等；
- 商业保险：投保数量、保费规模、保费收入、险种分布等；

——金融咨询信息：基础行情、量化指数、资讯舆情、研究报告、公告事件等。
三级类目可根据实际业务需求进行维护。

7.2 产业专题库

7.2.1 概述

产业数据仓宜建设产业专题库。通过数据融合对产业数据进行专题分析，根据产业数据仓数据分层架构，将数据划分成若干相对独立的业务域进行分阶段建设，形成满足特定场景需求的产业专题库。产业专题库包含业务专题库与行业专题库等。

产业专题库的建设宜按照实际业务需求，完成基础数据采集、专题明细表构建、应用数据汇总加工，为不同业务场景应用提供专题库服务。

专题数据库应满足但不限于以下要求：

- a) 实现对其面向的业务域产生和需要数据的收集、整理；
- b) 根据数据资源的内容特征及数据资源之间的关系，确定合理的数据模型，并将数据资源进行组织和管理；
- c) 根据专题业务逻辑或专题应用场景需求，对数据进行汇总分析，形成面向应用的专题数据汇总表；
- d) 将专题数据库进行数据资源目录编制，供产业数据仓用户申请调用，提供专题库服务；
- e) 实现不仅为数字经济领域的业务场景提供数据支撑，而且提供数据加工的环境和所需工具，使得数据能够直接在有关模型中进行分析和计算，实现数据和模型结合。

7.2.2 业务专题库

业务专题库基于应用场景实际需求，采集业务专题基础数据，通过产业数据仓的数据融合功能，构建业务专题明细表、业务应用数据表，形成业务专题分析的专项数据集合，提供业务专题库服务，支撑业务应用场景。业务专题库包括但不限于经济运行库、资源要素库、企业服务库、数字贸易库、产业链库等。

业务专题库建设示例见附录C。

7.2.3 行业专题库

行业专题库围绕第一、二、三产业，采集行业专题基础数据，通过产业数据仓的数据融合功能，构建行业专题明细表、行业分析数据表，形成行业专题分析的专项数据集合，提供行业专题库服务，支撑行业应用场景。行业专题库聚焦赋能产业集群和细分行业产业大脑建设，领域包括但不限于化工、化纤、电机、畜牧业、渔业、种植业、物流业、影视业等。

行业专题库的数据主体宜包括但不限于行业的企业、园区、协会等，以及以数据主体为核心产生的业务数据，如企业基本信息、企业产品信息、企业经营信息、园区基本信息、园区产业标签、协会基本信息、协会统计信息等。

行业专题库建设示例见附录D。

8 对接要求

8.1 数据目录对接

产业数据仓可提供标准化的目录对接接口，实现与第三方数据平台的数据目录同步更新。

8.2 用户体系对接

产业数据仓应提供用户调用标准接口和接口字段规范要求，满足各级各类数仓之间、数仓与其他应用系统之间的用户统一身份认证和互联互通需求等。

8.3 系统日志对接

产业数据仓可提供规范化的接口，以日志同步方式实现与数据安全防护产品的对接，日志包括但不限于数据申请审批日志、库表数据交换日志、接口调用日志等。

8.4 功能模块对接

产业数据仓应支持与第三方通用软件的交互和集成。第三方通用软件包括但不限于审批流引擎、数据标准管理、数据质量管理、数据治理与融合等。

9 安全体系建设

9.1 总体要求

9.1.1 产业数据仓应围绕数据采集、存储、处理、交换、销毁等数据的全生命周期构建安全防护体系，符合 GB/T 22239—2019 相应等级的技术要求，核心数据的传输和存储应支持国密算法加密，防止数据的非法采集、非授权访问、篡改、泄漏和滥用等。

9.1.2 产业数据仓应采用权限管控、数据脱敏、数据加密、数字签名、数据溯源、态势感知、日志审计、区块链、隐私计算等安全技术手段，提升数据安全防护能力，实现数据全生命周期各阶段的技术安全风险防控。

9.1.3 在中华人民共和国境内收集和产生的产业数据，应依照法律、行政法规要求在境内存储。确需向境外提供的，应依法依规进行数据出境安全评估，并报数据管理机构批准。非经数据管理机构批准，数据处理者不得向境外机构提供存储于中华人民共和国境内的数据。

9.1.4 数据安全体系建设应根据产业数据的分类分级情况采取差异化管理，兼顾成本与安全效益。

9.2 数据分类分级要求

9.2.1 数据分类

数据分类宜采用多维度和线分类法相结合的方法，根据行业要求、管理方式、业务应用、数据来源和用途等因素进行分类。

9.2.2 数据分级

结合产业数据特点，根据数据破坏后对国家安全、社会秩序、公共利益以及对公民、法人和其他组织的合法权益（受侵害客体）的危害程度来确定数据的敏感级别，将产业数据共分为4级，由高至低分别为：敏感数据（L4级）、较敏感数据（L3级）、低敏感数据（L2级）、不敏感数据（L1级）。

9.2.3 分类分级安全防护体系

产业数据仓分类分级安全防护体系宜遵循以下原则：L1级数据可直接共享；L2级数据应脱敏脱密后进行共享；L3级数据应采用数字签名、区块链等安全技术手段进行共享，确保传输和存储安全性；L4级数据不宜进行共享，应采用多方安全计算技术，在数据不出域情况下进行逻辑加工，以发挥数据价值。

9.3 系统管理

产业数据仓应具备具备用户管理、角色管理、权限管理、日志管理、流程引擎等功能，构建统一的用户体系，对组织架构、人员角色、功能权限、数据访问权限等内容进行统一管控，支持对数据审批流程的可视化配置，以流程引擎驱动审批流程的执行。

9.4 数据运维安全要求

9.4.1 应按照“谁主管谁负责、谁运营谁负责”的原则，明确各级各部门数据安全管理的职责分工，落实数据安全主体责任。

9.4.2 应建立事前审批、事中留痕、事后追溯的全链路数据安全监管机制，明确协同运行的处置机构，及时发现处置各类数据安全风险，切实防范数据篡改、泄漏、滥用。

9.4.3 数据处理者应根据传输的数据类型、级别和应用场景，制定安全策略并采取保护措施；与数据提供方通过签署相关协议、承诺书等方式，明确双方法律责任。

附 录 A
(资料性)
各级各类数仓建设内容参考

各级各类数仓建设内容参考见表A. 1。

表 A. 1 各级各类数仓建设内容参考

项目	行政数仓		行业数仓	企业数仓
	省级数仓	市级数仓		
数据资源编目	●	——	●	——
数据采集	●	●	●	●
数据交换共享	●	●	●	●
数据治理	●	○	●	○
数据融合	●	○	●	○
数据存储	●	●	●	●
数据管理	●	——	●	——
数据质量评价	●	○	○	○
安全防护体系	●	●	●	●
产业主题库	●	——	——	——
产业专题库	●	●	●	○
注：“●”代表应建内容，“○”代表宜建内容，“——”代表不建内容。				

附录 B
(资料性)

产业数据仓数据分层架构参考

产业数据仓根据实际使用需求对数据进行逻辑分层，确保仓内数据能够有序流转，数据结构更加清晰，便于数据的融合分析及数据价值挖掘。产业数据仓数据分层架构见图B. 1。

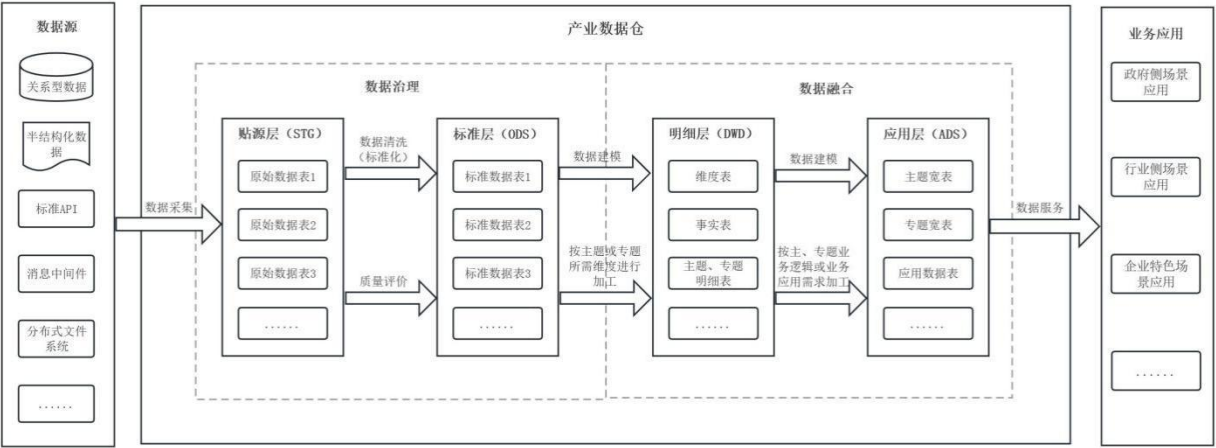


图 B. 1 产业数据仓数据分层架构

产业数据仓数据分层架构如下：

- a) 贴源层（STG）：异构数据源中的数据通过数据采集系统接入到此层，此层中数据与数据源保持完全一致，作为数据的快照存储，不做任何加工；
- b) 标准层（ODS）：贴源层中的数据通过标准化的数据清洗与质量提升后进入此层，数据结构与数据源保持一致，为数据仓后续数据加工提供标准数据；
- c) 明细层（DWD）：对标准层中的数据进行分类，依据数据的属性划分事实表和维度表，并基于维度建模思想，按主题或专题所需维度进行加工，构建最细粒度的主题、专题明细表；
- d) 应用层（ADS）：对明细层中的数据按主题、专题业务逻辑或业务应用需求进行汇总加工，形成主题、专题宽表或应用数据表，对外提供数据服务。

附 录 C
(资料性)
业务专题库建设示例 (企业诉求专题库)

C.1 企业诉求专题库

企业诉求专题库基于产业数据仓数据分层架构,由标准层的基础数据、明细层的业务专题明细数据、应用层的业务应用数据等三类数据组成:

- a) 基础数据: 企业诉求专题库基础数据表包括所有业务专题所需基础数据,如助企干部信息表、受助企业信息表、企业诉求信息表、诉求办理情况表、企业评价信息表等;
- b) 业务专题明细数据: 企业诉求专题库业务专题明细表包括按地市诉求信息明细表、按行业诉求信息明细表、按产业链诉求信息明细表、近 7 日诉求信息明细表、近 30 日诉求信息明细表等;
- c) 业务应用数据: 企业诉求专题库业务应用数据表包括按地市诉求信息汇总表、按行业诉求信息汇总表、按产业链诉求信息汇总表、近 7 日诉求信息汇总表、近 30 日诉求信息汇总表等。

C.2 基础数据

以企业诉求信息表为例,产业数据仓通过数据采集功能将数据从数据源抽取到数据仓中,经过数据清洗后,形成企业诉求信息表,相关信息见表C.1。

表 C.1 企业诉求信息表

序号	字段英文名称	字段中文名称	字段类型
1	zx_create_time	诉求提交时间	DATETIME
2	title	诉求标题	VARCHAR (500)
3	content	诉求描述	VARCHAR (4000)
4	cust_name	企业名称	VARCHAR (100)
5	org_code	企业组织机构代码	VARCHAR (100)
6	linkman	企业联系人	VARCHAR (100)
7	type_big_name	一级诉求分类名称	VARCHAR (30)
8	type_big_sub_name	二级诉求分类名称	VARCHAR (30)

C.3 业务专题明细数据

以按行业诉求信息明细表为例，以业务逻辑为导向，通过数据建模方式，将企业诉求信息表、诉求办理情况表与企业所属行业对应关系表进行数据融合加工，形成带有行业标签的按行业诉求信息明细表，相关信息见表C.2。

表 C.2 按行业诉求信息明细表

序号	字段英文名称	字段中文名称	字段类型
1	zx_create_time	诉求提交时间	DATETIME
2	title	诉求标题	VARCHAR (500)
3	content	诉求描述	VARCHAR (4000)
4	cust_name	企业名称	VARCHAR (100)
5	org_code	企业组织机构代码	VARCHAR (100)
6	linkman	企业联系人	VARCHAR (100)
7	type_big_name	一级诉求分类名称	VARCHAR (30)
8	type_big_sub_name	二级诉求分类名称	VARCHAR (30)
9	hydl	企业所属行业大类	VARCHAR (50)
10	hyzl	企业所属行业中类	VARCHAR (50)
11	hyxl	企业所属行业小类	VARCHAR (50)
12	status_name	诉求办理状态	VARCHAR (30)

C.4 业务应用数据

以按行业诉求信息汇总表为例，产业数据仓将按行业诉求信息明细表进行汇总分析，形成按行业诉求信息汇总表，相关信息见表C.3。

表 C.3 诉求信息汇总表

序号	字段英文名称	字段中文名称	字段类型
1	hydl	行业大类	VARCHAR (50)
2	hyzl	企业所属行业中类	VARCHAR (50)
3	hyxl	企业所属行业小类	VARCHAR (50)
4	count_hy	诉求总数量	VARCHAR (30)
5	count_processing	办理中诉求数量	VARCHAR (30)
6	count_finish	办结诉求数量	VARCHAR (30)

附 录 D
(资料性)
行业专题库建设示例（化工行业专题库）

D.1 化工行业专题库

化工行业专题库由基于产业数据仓数据分层架构，由标准层的基础数据、明细层的行业专题明细数据、应用层的行业分析数据等三类数据组成。

- a) 基础数据:化工行业专题库包括所有行业运行分析所需基础数据，如化工企业数据、化工园区数据、化工协会数据、化工行业指标信息等；
- b) 行业专题明细数据:化工行业专题明细表包括化工行业生产明细表、化工行业利润明细表、化工行业销售明细表等；
- c) 行业分析数据:化工行业专题库行业分析数据表包括化工行业生产指标汇总表、化工行业利润指标汇总表、化工行业销售指标汇总表等。

D.2 基础数据

以化工企业数据信息表及化工行业指标信息表为例，产业数据仓通过数据采集功能将数据从数据源抽取到数据仓中，经过数据清洗后，形成化工企业数据信息表，相关信息见表D. 1。

表 D. 1 化工企业信息表

序号	字段英文名称	字段中文名称	字段类型
1	Company_ID	企业 ID	VARCHAR (500)
2	Company_name	企业名称	VARCHAR (500)
3	unicode	统一社会信用代码	VARCHAR (100)
4	org_code	企业组织机构代码	VARCHAR (100)
5	Reg_address	注册地址	VARCHAR (100)
6	linkman	企业联系人	VARCHAR (100)
7	Label_name	产业标签	VARCHAR (30)
8	Found_date	成立日期	Date

D.3 行业专题明细数据

以化工行业生产明细表为例，以行业分析需求为导向，通过数据建模方式，将化工企业数据信息表与企业指标生产信息采集表进行数据融合加工，形成化工行业生产明细表，相关信息见表D. 2。

表 D.2 化工行业生产明细表

序号	字段英文名称	字段中文名称	字段类型
1	index_ID	生产指标 ID	VARCHAR (500)
2	index_product_name	生产指标名称	VARCHAR (500)
3	unit	单位	VARCHAR (100)
4	Company_ID	企业 ID	VARCHAR (500)
5	Company_name	企业名称	VARCHAR (500)
6	unicode	统一社会信用代码	VARCHAR (4000)
7	org_code	企业组织机构代码	VARCHAR (100)
8	Reg_address	注册地址	VARCHAR (100)
9	linkman	企业联系人	VARCHAR (100)
10	Label_name	产业标签	VARCHAR (30)
11	Found_date	成立日期	Date

D.4 行业分析数据

以化工行业生产指标汇总表为例，产业数据仓将化工行业生产明细表进行汇总分析，形成化工行业生产指标汇总表，相关信息见表D.3。

表 D.3 化工行业生产指标汇总表

序号	字段英文名称	字段中文名称	字段类型
1	index_ID	指标 ID	VARCHAR (500)
2	index_name	指标名称	VARCHAR (500)
3	Report_Date	报告日期	date
4	value	数值	Float8