

ICS 35.240.01

CCS L 67

DB33

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

浙江省市场监督管理局 发布

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命周期管理，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：

- 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；
- 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；
- 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。

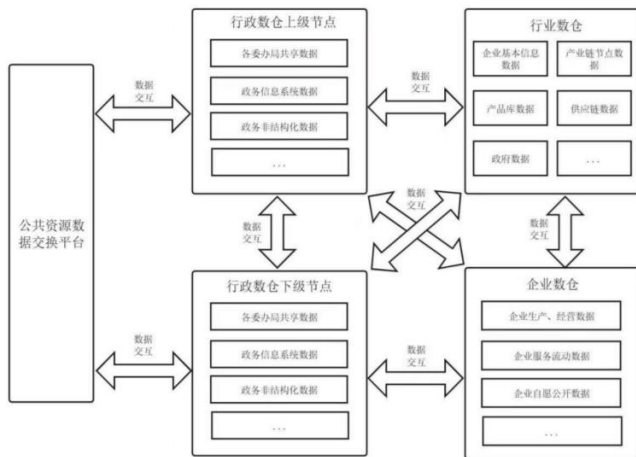


图 1 产业数据仓逻辑架构图

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：

- 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；
- 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；
- 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。

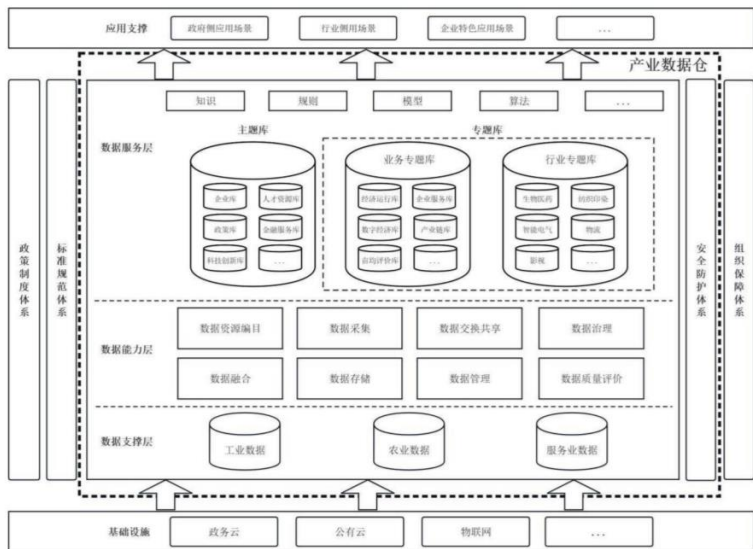


图 2 产业数据仓技术架构图

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。

数据资源编目应具备应用系统注册、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、数据源单位、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府部门间、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，建立各行业的数据库质量标准规范，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空间和组件动态伸缩、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。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/727155052012006145>

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：

- 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与知识之间的关系，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；
- 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；
- 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。