

"产业大脑"数据仓建设指南

Construction guidelines for the data warehouse of the "industry brain"

2025 - 07 - 29 发布

2025 - 08 - 29 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 基本原则 1

 4.1 需求牵引 1

 4.2 场景赋能 1

 4.3 安全合规 2

5 总体架构 2

6 数据来源 3

 6.1 公共数据 3

 6.2 企业数据 3

 6.3 其他数据 3

7 数据能力层 3

 7.1 数据采集 3

 7.2 数据资源编目 3

 7.3 数据治理 4

 7.4 数据融合分析 4

 7.5 数据存储 4

 7.6 数据共享交换 5

 7.7 数据质量管理 5

8 数据服务层 6

 8.1 基础库 6

 8.2 行业库 7

 8.3 场景库 7

 8.4 数据服务 8

9 数据仓对接 8

10 安全保障 8

 10.1 安全管理 8

 10.2 安全技术 8

 10.3 安全运营 9

附录 A（资料性） 场景库建设示例（企业服务信息） 10

 A.1 企业服务信息 10

 A.2 基础数据表 10

参考文献 12

前 言

本文件按照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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由山东省工业和信息化厅提出、归口并组织实施。

"产业大脑"数据仓建设指南

1 范围

本文件确立了“产业大脑”数据仓建设的基本原则，提供了总体架构、数据来源、数据能力层、数据服务层、数据仓对接、安全保障等方面的指导。

本文件适用于指导各行业“产业大脑”数据仓的规划和建设。

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文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB/T 22239 信息安全技术 网络安全等级保护基本要求
- GB/T 35273 信息安全技术 个人信息安全规范
- GB/T 39786 信息安全技术 信息系统密码应用基本要求
- GB/T 43697 数据安全技术 数据分类分级规则
- DB37/T 4646.2 公共数据 数据治理规范 第2部分：数据清洗比对

3 术语和定义

下列术语和定义适用于本文件。

3.1

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]

3.2

数据标准 data standard

数据的命名、定义、结构和取值范围方面的规则和基准。

[来源：GB/T 36344—2018，2.8]

4 基本原则

4.1 需求牵引

“产业大脑”数据仓建设宜以产业发展需求为牵引，推动公共数据、企业数据、其他数据等多方数据汇聚。

4.2 场景赋能

宜通过规范“产业大脑”数据仓数据采集、编目、治理、融合分析、存储、共享交换和质量管理，以知识、规则、模型、算法等方式赋能产业服务和场景。

4.3 安全合规

宜采取措施保护“产业大脑”数据仓中数据不被非法访问、修改或删除。

5 总体架构

“产业大脑”数据仓宜整合产业相关数据，实现产业数据的采集、编目、治理、融合分析、存储、共享交换和质量管理，是“产业大脑”的核心组成部分和数据底座。“产业大脑”数据仓宜包括数据能力层、数据服务层、安全保障，“产业大脑”数据仓架构见图1。

- a) 数据能力层：作为支撑产业数据处理的核心层，宜具备数据资源采集、编目、治理、融合分析、存储、共享交换和质量管理等能力。
- b) 数据服务层：宜通过对数据仓汇聚的原始数据进行数据治理、融合，构建形成基础库、行业库和场景库，同时汇聚各类知识、规则、模型、算法，为各类场景应用提供服务支撑。
- c) 安全保障：宜为数据仓提供安全管理、安全技术和安全运营等支撑。

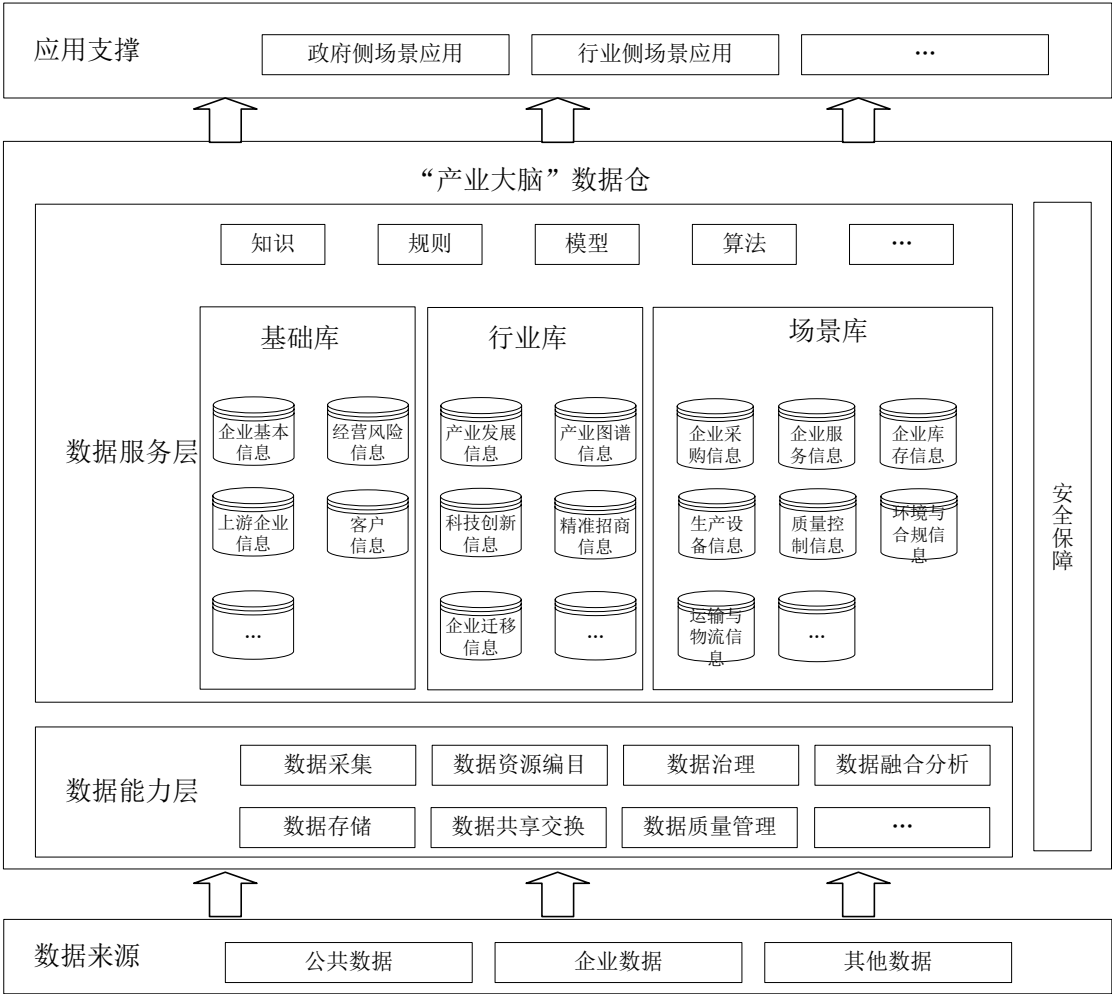


图1 “产业大脑”数据仓架构

6 数据来源

6.1 公共数据

6.1.1 宜通过省一体化大数据平台，根据业务需求申请政府侧相关产业基础数据，包括但不限于企业基础数据、公共资源交易数据、科技创新数据、金融综合服务数据、政策人才数据。

6.1.2 企业基础数据宜包括但不限于企业的登记信息、存续状态数据。

6.1.3 公共资源交易数据宜包括但不限于公共资源交易平台上的交易数据，如政府采购、土地拍卖、工程招投标数据。

6.1.4 科技创新数据宜包括但不限于科技创新项目申报、科研成果、专利申请和授权数据。

6.1.5 金融综合服务数据宜包括但不限于信贷记录、融资情况、保险业务数据。

6.1.6 政策人才数据宜包括但不限于税收政策、环保政策、贸易政策、鼓励性政策、行业规范、限制性措施、人才基本信息数据。

6.2 企业数据

6.2.1 宜通过产业链上下游企业许可范围内的合规数据共享交换获得企业运营、供应链等数据。

6.2.2 企业运营数据宜包括但不限于以下内容。

- a) 企业资源计划数据：物料需求计划、库存状态、生产计划、财务信息、人力资源数据。
- b) 客户关系管理数据：企业的客户信息、销售记录、售后服务数据。
- c) 制造执行数据：生产计划调度、生产进度跟踪、质量控制数据。

6.2.3 供应链数据宜包括但不限于以下内容。

- a) 供应链上下游企业数据：供应商的生产能力、质量水平、交货周期数据。
- b) 商品贸易和流通环节数据：进出口贸易数据、物流运输数据。
- c) 链主企业的招投标数据。

6.2.4 企业的其他数据宜包括但不限于知识产权、标准、企业认证和资质数据。

6.3 其他数据

宜通过互联网及其他资源渠道搜集获取电信运营商、行业协会、研究机构等第三方行业数据，包括但不限于互联网的新闻资讯、社交媒体信息、产品市场信息、行业从业机构信息，以及行业研究机构、咨询公司等提供的市场分析报告、行业研究报告。

7 数据能力层

7.1 数据采集

7.1.1 “产业大脑”数据仓宜具备数据采集能力，支持多源数据采集，包括物联网设备、社交媒体、交易系统等，并实施自动化数据采集流程，支撑数据的高效集成。

7.1.2 数据采集能力宜包括数据源管理、数据抽取、数据转换、采集任务调度、元数据管理、采集监控等。

7.2 数据资源编目

7.2.1 “产业大脑”数据仓宜具备数据资源编目能力，实现数据资源的统一编目和管理，为“产业大脑”、各业务系统提供数据资源清单。

7.2.2 数据资源编目能力宜包括数据资源目录编制、数据资源审核发布、数据资源挂载、数据资源目

录查询、数据资源目录变更、数据资源停用等。

7.2.3 数据资源目录内容宜包括但不限于数据资源名称、数据资源代码、数据资源提供方、数据资源分类、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、数据安全级别、开放属性、数据资源格式和类型、更新频率、发布日期、数据资源摘要。

7.3 数据治理

7.3.1 “产业大脑”数据仓宜具备数据治理能力，建立各行业数据治理规范，包括元数据标准、数据清洗比对标准等，对行业数据进行标准化和清洗比对。

7.3.2 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、标准质量检查、标准版本管理、标准模板管理、标准变更订阅等能力；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。

7.3.3 数据清洗比对宜具备清洗比对规则自定义与配置、数据过滤、数据比对、数据去重、数据补全、数据转化、数据降维、清洗任务管理等能力；清洗比对规则包括但不限于特定字符转换、大小写转换、汉语转拼音、同义词替换、规则校验，数据清洗比对宜参照 DB37/T 4646.2。

7.4 数据融合分析

7.4.1 “产业大脑”数据仓宜具备数据融合分析能力，通过数据模型管理、数据血缘分析、数据标签管理、数据融合分析，开发逻辑模型、物理模型，形成基础库、行业库和场景库，提供数据服务。

7.4.2 数据模型管理宜提供以逻辑模型创建物理模型能力，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；物理模型宜包含模型物化和模型加工两部分能力，模型物化宜创建物理模型定义的表结构，模型加工宜在线定义模型的加工逻辑，并支持垂直拆分、水平拆分、多表联合、多表连接或自定义映射等方式。

7.4.3 数据血缘分析宜提供数据血缘分析和展示能力，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能力；支持数据流动趋势、数据库间依赖关系和数量的展示。

7.4.4 数据标签管理宜提供数据标签体系建设和管理能力，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，提供数据分类标注和分级标注等服务，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，支持指标统计分析、标签的开发与运维、标签使用情况监测。

7.4.5 数据融合分析能力宜提供跨数据源的数据分析、融合和提取，形成面向业务和应用的数据集；支持基础统计分析、数据挖掘算法应用、多维分析等能力。

7.5 数据存储

7.5.1 “产业大脑”数据仓宜支持存储结构化数据、半结构化数据、非结构化数据等。

7.5.2 宜采用不同的存储方式存储数据仓不同类型的数据，存储方式包括但不限于以下方式。

- a) 传统关系型数据库：存储结构化数据，适用于少量存储数据，提供事务支持、数据一致性和完整性保障。
- b) 分布式关系型数据库：存储结构化数据，适用于处理大规模的数据集，并提供高性能的查询能力。
- c) 列式关系型数据库：存储结构化数据，适用于大规模数据的 OLAP 场景。

- d) 时序数据库：存储结构化数据，适用于存储物联网传感器数据、金融市场数据、应用性能监控等数据。
- e) 文档数据库：存储半结构化数据，适用于以文档作为数据存储和操作基本单元的场景。
- f) 分布式文件系统：存储非结构化数据，适用于以文件形式分散存储在多台独立服务器的场景；
- g) 对象存储：存储非结构化数据，适用于云计算环境。

7.5.3 宜制定数据备份策略，确保数据的备份频率、备份方式、备份介质等符合业务需求。备份策略包括但不限于全量备份、增量备份、差异备份。

7.5.4 宜制定数据恢复策略，确保在数据丢失或损坏的情况下能够快速恢复数据。恢复策略包括但不限于全量恢复、增量恢复、差异恢复。

7.6 数据共享交换

7.6.1 “产业大脑”数据仓宜根据数据的时效性、数据量等情况确定数据共享交换方式，数据共享交换方式包括但不限于以下方式。

- a) 数据库表方式：数据供给方将“产业大脑”数据仓中的数据推送到共享数据库中，数据需求方从共享数据库中获取数据，适用于数据时效性较低、共享交换的数据量较大的情况，主要能力包括数据源管理、数据字段映射、数据字段预处理、数据推送。
- b) 接口方式：数据供给方通过接口方式提供共享数据，适用于数据时效性较高、共享交换的数据量不大的情况，主要能力包括接口注册、接口代理、接口鉴权、接口流控。
- c) 文件方式：数据供给方将数据以文件为载体共享给数据需求方的方式，适用于数据时效性较低、数据对接双方隔离性较高的情况，主要能力包括文件源管理、文件清单管理、文件下载、文件上传。
- d) 标识解析方式：数据供给方将数据仓数据在工业互联网标识解析二级节点注册，通过标识映射数据仓 IP 地址，数据需求方通过标识解析实时获取数据。
- e) 区块链方式：利用区块链技术的分布式、不可篡改和加密特性，实现数据的安全共享交换，如交易信息、税务信息等，适用于数据时效性较低、共享交换的数据量较少、安全性较高的情况。

7.6.2 宜采用同步机制来确保“产业大脑”数据仓数据的一致性、完整性，同步机制包括但不限于以下内容。

- a) 全量同步：每次同步都会将所有数据都同步到目标端，适用于数据量不大、完整性较高、数据更新不可控的情况。
- b) 增量同步：只同步自上次同步以来发生变化的数据，适用于时效性较高、同步效率较高、资源占用较低的情况。
- c) 比对同步：比较源端和目标端之间的数据差异，然后将差异数据同步到目标端，适用于数据量较大、时效性较低、数据更新不可控的情况。

7.6.3 宜采用的数据同步方式包括但不限于实时同步、周期性同步、条件触发同步。

7.7 数据质量管理

7.7.1 “产业大脑”数据仓宜建设数据质量管理能力，包括但不限于数据规则管理、数据质量检查、数据质量分析评价、数据质量报告生成、问题数据工单管理、数据质量监控与改进能力。

7.7.2 宜根据行业数据实际情况，制定数据质量管理流程，包括但不限于：

- a) 数据采集阶段，制定数据采集规范和验证规则；
- b) 数据清洗阶段，识别和处理数据中的错误、缺失值和异常值；
- c) 数据转换阶段，使用数据转换规则和映射表，确保数据在不同系统或格式之间的转换正确无误；

- d) 数据验证阶段，使用数据验证工具和技术对数据进行验证。

8 数据服务层

8.1 基础库

8.1.1 “产业大脑”数据仓宜建设基础库，将行业内企业基本信息、经营风险信息、上游企业信息、客户信息等基础信息进行归类和整合。

8.1.2 企业基本信息宜包括但不限于以下内容。

- a) 企业工商信息：企业名称、统一社会信用代码、企业法人、成立日期、企业官网、企业简介、经营范围、电话、电子邮箱、注册资金（万元）、实缴资本（万元）、纳税资质、企业类型、经营期限开始日期、经营期限结束日期、注册详细地址、登记机关、登记证号、参保人数、人员规模、登记状态。
- b) 企业专利信息：企业名称、统一社会信用代码、专利名称、专利摘要、专利类型、专利状态、专利申请号、专利申请人（机构）、申请日期、申请人地址、专利分类号、发明人、公开公告号、公开公告号日期、是否国际专利、代理机构、代理人。
- c) 企业商标信息：企业名称、统一社会信用代码、商标名称、商标图案（URL）、商标注册号、商标状态、商标或服务项目、申请人地址、申请人、申请日期、代理机构、国际分类编码、国际分类名称、注册公告日期。
- d) 软件著作权信息：企业名称、统一社会信用代码、软件全称、软件简称、版本号、登记号、开发完成日期、首次发布日期、登记日期、权利取得方式。
- e) 对外投资信息：企业名称、统一社会信用代码、被投资企业名称、企业状态、企业法人、注册资金（万元）、成立日期、持股比例、认缴出资额、认缴出资日期、最终受益股份、所属行业、投资数量、关联产品或机构。
- f) 创投融资信息：企业名称、统一社会信用代码、融资日期、融资轮次、融资金额、企业估值、投资方。
- g) 股权质押信息：企业名称、统一社会信用代码、质押人、质押人参股企业、质押权人、质押股份总数（股）、质押股份市值（元）、质押状态、公告日期。
- h) 网站备案信息：企业名称、统一社会信用代码、网站地址、网址、域名、网站备案或许可证号、审核日期、备案域名、公安备案号、备案地公安机关、备案日期、网站类型。
- i) 企业荣誉信息：企业名称、统一社会信用代码、荣誉名称、荣誉编号、荣誉类别、荣誉级别、发布单位、发布日期、有效期自、有效期至。
- j) 企业资质证书信息：企业名称、统一社会信用代码、证书名称、资质类别、资质等级或产品名称、证书编号、发证日期、发证机构、有效期至、证书状态。

8.1.3 经营风险信息宜包括但不限于以下内容。

- a) 司法案件信息：所属企业、统一社会信用代码、案件名称、案件类型、案件日期、案由、案件进程、案件身份、案号、法院、案件金额（元）。
- b) 限制高消费信息：所属企业、统一社会信用代码、案号、限消令对象、限制法定代表人、申请人、涉案金额（元）、执行法院、立案日期、发布日期。
- c) 行政处罚信息：所属企业、统一社会信用代码、决定文书号、违法事实、处罚结果、处罚金额、处罚单位、处罚日期。
- d) 被执行人信息：所属企业、统一社会信用代码、案号、被执行人、执行标的（元）、执行法院、立案日期。

- e) 失信被执行人信息：所属企业、统一社会信用代码、案号、失信被执行人、疑似申请执行人、涉案金额（元）、执行法院、立案日期、发布日期。

8.1.4 上游企业信息宜包括但不限于企业名称、统一社会信用代码、供应商、供应商状态、采购金额、报告期或公告日期。

8.1.5 客户信息宜包括但不限于企业名称、统一社会信用代码、客户、客户状态、销售金额、报告期或公告日期。

8.2 行业库

8.2.1 “产业大脑”数据仓宜建设行业库，行业库包括但不限于产业发展信息、产业图谱信息、科技创新信息、精准招商信息、企业迁移信息。

8.2.2 产业发展信息宜包括但不限于产业大脑代码、产业规模、产业聚集度、国家级奖项清单（集群名称、认定机构、认定时间）、省部级奖项清单（集群名称、认定机构、认定时间）、省十强“雁阵形”集群情况、省特色产业集群情况、省战略性新兴产业集群情况、其他省级荣誉清单（集群名称、认定机构、认定时间）。

8.2.3 产业图谱信息宜包括但不限于产业大脑代码、产业链名称、产业链节点名称（如微纳传感器产业链，上游铜原材料供应和芯片设计（一级节点名称）-传感器制造（二级节点名称）-传感器应用（三级节点名称））、产业链节点关联企业名称、产业链节点关联企业统一社会信用代码。

8.2.4 科技创新信息宜包括但不限于产业大脑代码、科创平台（平台名称、关联企业名称、关联企业统一社会信用代码）、产学研项目（项目名称、关联企业名称、关联企业统一社会信用代码）、高层次人才（人才姓名、学历层次、关联企业名称、关联企业统一社会信用代码）、专业技能人才（人才姓名、关联企业、关联企业统一社会信用代码）、晨星工厂（企业名称、企业统一社会信用代码）。

8.2.5 精准招商信息宜包括但不限于产业大脑代码、招商项目名称、关联产业链节点、关联企业、投资金额（元）、签约状态、项目进度。

8.2.6 企业迁移信息宜包括但不限于产业大脑代码、关联产业节点、迁移类型（入或出）、迁移概率。

8.3 场景库

8.3.1 “产业大脑”数据仓宜建设场景库，根据行业内企业的业务需求和场景需求，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场景库。

8.3.2 场景库宜包括但不限于企业采购信息、企业服务信息、企业库存信息、生产设备信息、质量控制信息、环境与合规信息、运输与物流信息，以企业服务信息为例，场景库建设示例见附录 A。

8.3.3 企业采购信息宜包括但不限于产业大脑代码、企业名称、统一社会信用代码、商品一级和二级分类名称和编码、商品编码、商品名称、采购时间、商品单位、采购数量、采购单价、原价金额、实付金额、采购状态、采购单位名称、供应商编码、供应商名称。

8.3.4 企业服务信息宜包括但不限于产业大脑代码、企业名称、统一社会信用代码、服务代码、服务名称、服务流程、服务资质要求、服务人才需求、服务交付成果、服务内容、服务报价。

8.3.5 企业库存信息宜包括但不限于产业大脑代码、企业名称、统一社会信用代码、物品编号、物品单位、物品数量、周转率、库房位置。

8.3.6 生产设备信息宜包括但不限于产业大脑代码、企业名称、统一社会信用代码、设备编号、设备名称、设备型号、月生产能力、月使用率、月空闲时段、使用状态、采购日期、报废年限。

8.3.7 质量控制信息宜包括但不限于产业大脑代码、企业名称、统一社会信用代码、产品编码、产品名称、质量标准、检验方法、检验时间、检验结果、检验记录、检验机构。

8.3.8 环境与合规信息宜包括但不限于产业大脑代码、企业名称、统一社会信用代码、记录日期、检查内容（检查项目、检查标准、检查方法、检查结果）、数据记录（监测数据、监测日期、监测人员、

监测设备)、判定结果、判定依据、判定日期、判定人员、整改要求、整改期限、整改结果、复查日期。

8.3.9 运输与物流信息宜包括但不限于产业大脑代码、企业名称、统一社会信用代码、所属企业、运输时间、运输费用、送达时间、发货地、目的地、物流公司名称、物流公司所在地、运输货物(货物名称、单位、数量)。

8.4 数据服务

宜结合政府侧及企业侧应用场景需求,基于已有的基础库、行业库、场景库数据,采集场景应用所需的其他数据,对场景相关数据治理融合后进行封装,为应用场景提供数据产品或应用支撑能力,通过接口等方式支撑“产业大脑”和行业内各企业的业务决策和分析需求,数据产品或应用支撑能力包括但不限于知识、规则、模型、算法。

9 数据仓对接

“产业大脑”数据仓宜具备数据资源对接能力,统一规范数据接口类型、协议、文档等内容,并提供数据资源调用接口,满足数据仓之间、数据仓与其他应用系统之间的数据资源调用,数据接口其他内容可参照DB37/T 4225。

10 安全保障

10.1 安全管理

10.1.1 宜建立数据安全管理制度,制定数据安全制度规范,明确数据安全管理目标、原则、责任边界等。

10.1.2 宜建立统一身份认证机制,对对接双方的身份进行认证,确保对接双方身份的真实性,保证仅授权方可以进行对接活动,统一身份认证机制内容包括但不限于:

- a) 对数据仓对接双方进行双向身份鉴别,保证对接双方身份的真实性;
- b) 在安全周期范围内对对接双方定期重新认证;
- c) 涉及交换敏感数据时,双方采用密码技术(如数字证书)进行身份鉴别;
- d) 涉及多方对接时,对各接入方进行交叉认证。

10.1.3 宜建立数据使用过程的责任机制和评估机制,防止数据资源被用于不正当目的。

10.1.4 宜建立数据灾备系统和机制,制定数据灾备策略和应急计划,定期进行灾备的演练和测试。

10.1.5 宜建立数据销毁机制,采用数据销毁、存储媒体销毁等方式,实现数据的有效销毁。

10.1.6 宜建立数据安全审计机制,对重要行为和重要安全事件进行安全审计,定期对审计记录进行分析和结果记录,审计内容包括但不限于数据申请审批、库表数据交换、接口调用。

10.1.7 网络安全等级保护宜符合 GB/T 22239,宜通过商用密码应用安全性评估,个人信息安全宜符合 GB/T 35273。

10.2 安全技术

10.2.1 宜按照 GB/T 43697 制定数据分类分级规则,按照各行业数据需求,对数据进行分类分级管理,重点对重要数据、核心数据进行识别和分级保护。

10.2.2 宜采用符合 GB/T 39786 的密码技术,对数据进行加密保护;根据应用需求保留敏感数据的原格式、属性或关联,并对脱敏操作进行记录。

10.2.3 宜建立数据脱敏安全策略,对敏感数据进行匿名化、去标识化等脱敏处理,屏蔽敏感信息。

10.2.4 宜对“产业大脑”数据仓的数据采取加水印的方式实现数据溯源管理和版权确认。

10.2.5 数据分析过程宜采取适当的安全控制措施，如差分隐私保护、K 匿名等，防止数据挖掘、分析过程中有价值信息和个人隐私泄漏的安全风险。

10.2.6 对于涉及产业数据中的特定级别敏感数据的共享交换，宜采用数据可用不可见技术。

10.3 安全运营

10.3.1 宜开展数据安全风险监测，建立数据安全事件应急响应机制，制定应急响应流程。

10.3.2 宜开展数据安全巡检自查，采用手工巡检、脚本化巡检和平台化巡检相结合的方式，对数据仓的可用性、可靠性和性能进行综合评定。

10.3.3 宜采用数据仓事务与日志记录、数据仓审计、数据备份与还原、数据可视化与报表等方法，对数据采集、治理、处理、分析、共享交换和应用等数据全流程进行溯源管理提高数据质量。

10.3.4 宜定期开展数据仓安全漏洞评估，从数据、网络、应用程序等方面进行全面的安全检测，及时发现并修复潜在的安全隐患。

10.3.5 宜开展数据安全培训工作，培训内容包括数据仓安全基础知识、安全操作规范、应急响应流程等。

附 录 A
(资料性)
场景库建设示例 (企业服务信息)

A.1 企业服务信息

企业服务信息基础数据包括服务注册信息，提供服务的企业信息，服务交易信息等。企业服务数据可先由行业头部企业实际业务数据导入，后续可由行业内各企业主动录入或批量调取等方式补充并动态维护，主要服务形式包括：

- a) 按服务名称、服务内容等构建查询接口，为需要服务的客户提供主动公开的服务列表或服务检索；
- b) 对已有企业服务数据进行分析，统计提供不同服务的企业数量、服务交易量，对比服务价格、服务评价等信息，帮助客户择优选择服务企业；
- c) 根据服务资质、人才需求等信息，过滤选择不同等级的服务内容及服务企业，对行业企业及人才进行分类统计，研究行业发展规律，赋能产业高质量发展；
- d) 根据用户使用情况和数据采集量，搭建企业服务交易平台，为服务供需双方提供在线交易提供支持，并在交易业务中不断丰富数据，提供行业数据统计查询服务等。

A.2 基础数据表

以服务注册信息表、企业服务信息表、服务交易信息表为例，“产业大脑”数据仓通过数据采集能力将数据从数据源抽取到数据仓中，经过数据清洗、比对等数据治理工作后，形成服务注册信息表（见表A.1）、企业服务信息表（见表A.2）、服务交易信息表（见表A.3）。

表A.1 服务注册信息表

字段代码	字段名称	字段类型	备注
Service_ID	服务 ID 号	字符型	系统主键
MindCode	产业大脑代码	字符型	—
OrgName	企业名称	字符型	—
UnifiedIdentifier	统一社会信用代码	字符型	—
Service_code	服务代码	字符型	—
Service_name	服务名称	文本	—
Service_process	服务流程	文本	宜另附附件形式挂接具体流程图等
Service_qualifications	服务资质要求	文本	通过分隔符列示多个，资质名称宜另设计字典表或资质明细表
Service_talents	服务人才需求	文本	提供该服务的人才基本技能需求或资质认证需求
Service_deliverables	服务交付成果	文本	该服务最终交付成果（清单）
Service_content	服务内容	文本	该服务的简介、价值等
Service_quotation	服务报价	文本	该服务的成本组成、行业价格计算等

表A. 2 企业服务信息表

字段代码	字段名称	字段类型	备注
ComSer_ID	企业服务 ID 号	字符型	系统主键
MindCode	产业大脑代码	字符型	—
OrgName	企业名称	字符型	—
UnifiedIdentifier	统一社会信用代码	字符型	—
ComSer_Service_name	企业服务名称	文本	“服务注册信息表”的服务名称
ComSer_Service_URL	企业服务链接	字符型	企业服务的网络 URL 或其他业内可理解的服务提供渠道说明
ComSer_Customer_service	企业客服信息	字符型	企业服务的联系方式，如：联系人、电话、邮件等信息

表A. 3 服务交易信息表

字段代码	字段名称	字段类型	备注
Ser_Transaction_ID	服务交易 ID 号	字符型	系统主键
MindCode	产业大脑代码	字符型	—
Ser_Provider	服务供方	文本	提供服务单位的名称
UnifiedIdentifier	服务供方统一社会信用代码	字符型	—
Ser_Team	服务团队	文本	给出团队成员描述
Ser_Demander	服务需方	文本	被服务单位名称
UnifiedIdentifier	服务需方统一社会信用代码	字符型	—
Ser_Procurement time	服务采购时间	日期	服务采购或开始的时间
Ser_Continue	服务持续时间	文本	提供服务的时间区间
Ser_End time	服务结束时间	日期	服务结束或终止的时间

参 考 文 献

- [1] GB/T 35295—2017 信息技术 大数据 术语
 - [2] GB/T 36344—2018 信息技术 数据质量评价指标
 - [3] DB37/T 4225 政务信息资源 数据服务接口规范
-