ICS 35.240.01
CCS L 70

DB32

江　苏　省　地　方　标　准

DB32/T 5059—2025

企业物流管理数据仓库建设指南

Data warehouse guide for the of construction of enterprise logistics management

2025-02-21发布　　2025-03-21实施

江苏省市场监督管理局　发布
中国标准出版社　出版

目　　次

前　言

本文件按照GB/T 1.1—2020《文件化工作导则　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。

本文件由江苏省软件和信息技术服务标准化技术委员会提出并归口。

本文件起草单位：江苏斯诺物联科技有限公司、诺得网络科技股份有限公司、上海大学。

本文件主要起草人：赵国荣、赵惠丹、武星、孙驰、吕斌、赵飓。

企业物流管理数据仓库建设指南

1 范围

本文件提供了企业物流管理数据仓库(下文简称“数据仓库”)设计的基本原则、设计指标、分层、数据构成、数据存储、数据建模、数据模型、数据采集、网络安全、数据备份与恢复、运行系统构成的内容,适用于企业物流管理数据仓库的规划、设计、开发和应用,支持物流行业与其他信息系统的互联互通。

本文件适用于企业物流管理数据仓库建设的过程。

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 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 11457—2006 信息技术 软件工程术语

GB/T 18768—2002 数码仓库应用系统规范

GB/T 20270—2006 信息安全技术 网络基础安全技术要求

GB/T 20271—2006 信息安全技术 信息系统通用安全技术要求

GB/T 20988—2007 信息安全技术 信息系统灾难恢复规范

GB/T 28452—2012 信息安全技术 应用软件系统通用安全技术要求

GB/T 29765—2021 信息安全技术 数据备份与恢复产品技术要求与测试评价方法

GB/T 33745—2017 物联网 术语

GB/T 35295—2017 信息技术 大数据 术语

GB/T 38667—2020 信息技术 大数据 数据分类指南

3 术语和定义

GB/T 5271.1—2000、GB/T 18768—2002、GB/T 11457—2006、GB/T 20270—2006、GB/T 20271—2006、GB/T 20988—2007、GB/T 5271.17—2010、GB/T 33745—2017、GB/T 35295—2017、GB/T 38667—2020界定的以及下列术语和定义适用于本文件。

3.1

数据仓库 data warehouse;DW

在数据准备之后用于永久性存储数据的数据库。

3.2

结构化数据 structural data

按次种形式,由数据元素汇集而成的每个记录的机构都是一致的并且可以使用关系模型予以有效描述的一种数据表示形式。

3.3

非结构化数据　unstructured data

不具有预定模型或以定义方式组织的数据。

3.4

元数据　metadata

关于数据或数据元素的数据(可能包括其数据描述),以及关于数据用有权、存取路径、访问权和数据易变性数据。

3.5

数据库　database

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。

4　缩略语

下列缩略语适用于本文件。

ADS:数据应用层(Application Data Store)

DIM:公共维表(Dimension Table)

DWD:基础数据层(Data Warehouse Detail)

DWER模型:实体联系模型(Entity-Relationship Model)

DWHS:基础标签层(Data Ware House Servce)

DWS:公共汇总粒度事实层(Data Warehouse Service)

ODS:数据接入层(Operational Data Store)

OLAP:联机分析处理(Online Analytical Processing)

RAID:磁盘阵列(Redundant Arrays of Independent Disks)

5　建设原则

5.1　开放性原则

应基于业界开放文件,以确保系统能够与不同的数据源和工具兼容,便于未来的扩展和维护。

5.2　数据完整性原则

数据在生成、存储、传输和处理过程中保持其准确性、一致性和完整性

5.3　可扩展性原则

可支持体系结构的扩展,适应未来的业务发展和技术升级,通过添加新功能或修改现有功能来满足不断变化的需求。

5.4　灵活性原则

能适应多样化的源数据,以及不断变化的需求和业务环境的能力,并向目标系统提供多样化的数据支持。

5.5 安全性原则

建设过程中应采取一定的措施保护数据仓库中的数据不被非法访问、修改或删除。

5.6 兼容性原则

可支持多种数据源和数据库系统，包括关系型数据库和非关系型数据库。

6 设计指标要求

6.1 性能指标

6.1.1 响应时间：每一百并发数＜3 s。
6.1.2 吞吐量：100 TPS。
6.1.3 并发数：数百并发是基本要求，且需要具备扩展到数千的能力。
6.1.4 存储容量：约6 TB。
6.1.5 数据量的大小：每年的数据量在100 MB～1 GB之间。
6.1.6 数据的类型和结构：结构化数据（如客户关系数据、订单数据等），可以使用关系型数据库，在10 GB～20 GB的硬盘容量即可存储数年的数据；非结构化数据需要使用对象存储或分布式文件系统，存储方式通常需要较大的硬盘容量。
6.1.7 实时同步：实时同步是将数据仓库与源数据库实时保持一致，确保数据的即时更新，能够7×24 h运行高负载业务。
6.1.8 批量同步：批量同步是定期将源数据库的数据批量导入到数据仓库中。

6.2 可扩展性

6.2.1 采用分布式架构：通过将数据仓库部署在多个节点上，实现数据的分布式存储和处理，提高系统的处理能力和扩展性。
6.2.2 引入云计算技术：利用云计算资源，实现数据仓库的弹性伸缩，根据业务需求动态调整存储和计算资源，降低企业的研发成本。
6.2.3 优化数据模型：通过对数据进行建模和优化，减少冗余数据，提高数据的存储效率和查询速度。
6.2.4 引入大数据处理技术：利用Hadoop、Spark等大数据处理框架，实现数据的并行处理和高效分析，提高数据仓库的处理能力。
6.2.5 采用列式存储技术：通过采用列式存储技术，减少数据冗余，提高数据的压缩率和查询速度。

7 数据仓库分层

数据仓库建设是一个整体性工作，从数据产生到入库的整个环节应尽量遵循数据架构图进行搭建，各环节采用一套标准。

数据仓库搭建结构图如图1所示。

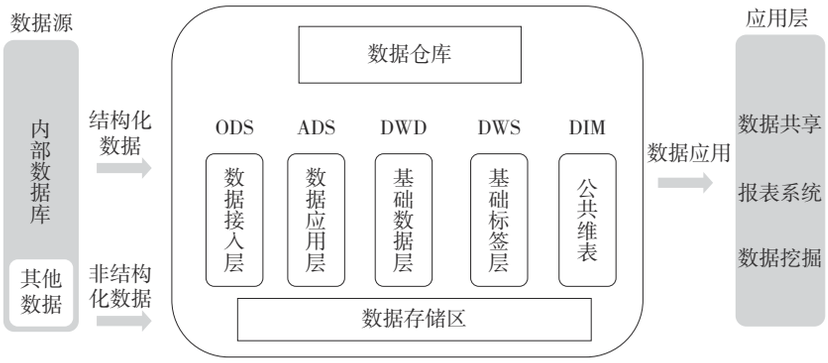

图1　数据仓库搭建结构图

数据仓库可分为数据接入层(ODS)、数据应用层(ADS)、基础数据层(DWD)、基础标签层(DWS)、公共维表(DIM)。

数据仓库分层结构表如表1所示。

表1　数据仓库分层结构表

结构名	英文全称	中文名	层次定义
ODS	Operational Data Store	数据接入层	实现功能:业务源系统数据接入到此层,此层数据不做任何加工,禁止重复进入。 数据来源范围:业务源系统。 数据存储时长:永久。
ADS	Application Data Store	数据应用层	实现功能:该层为数据应用层,根据业务需求组织数据,该层定期需要定期review,据层将公共指标沉淀到DWS中。 应用数据来源范围:DWS、DWD。 数据存储时长:根据业务需求状况保留。
DWD	Data Warehouse Detail	基础数据层	实现功能:该层为基础数据层,主要操作包括数据清洗、数据过滤、数据历史变更记录等。 数据来源范围:此层数据来源于0DS。 数据存储时长:根据业务需求状况保留。
DWS	Data Warehouse Summary	基础标签层	实现功能:该层为基础标签层,主要从DWD层的数据进行粗粒度聚合汇总;按不同维度进行统计,主要操作包括基于业务整合、关联计算得到的明细数据;着力公共指标、排序聚合得到的汇总数据 数据来源范围:DWD。 数据存储时长:根据业务需求状况保留。
DIM	Dimension Table	公共维表	实现功能:该层为公共维表层,该层独立于DWD、DWS、ADS,为DWD、DWS、ADS提供维度字段说明。 数据来源范围:0DS。 数据存储时长:根据业务需求状况保留。

8　数据构成

8.1　概述

企业物流数据仓库的数据构成包含但不限于物流运输中的业务数据、用户行为数据和爬虫数据等,

其中业务数据是核心。

8.2 业务数据

8.2.1 客户数据

基本信息:客户名称、客户代码、企业类型与规模、经营信息、联系人的姓名与联系方式等。

交互数据:购买记录、客服沟通记录等。

行为数据:积极(主动申请注册试用、主动沟通需求、主动提出建议等)、消极(拒接电话、减少预算等)。

8.2.2 物流公司数据

企业名称、企业代码、企业规模、经营方式、历年经营信息、联系人的姓名与联系方式、拥有车辆规模等。

8.2.3 承运商数据

车辆数、车辆具体信息、司机基本信息、证件信息等。

8.3 系统数据

系统监控日志、接口运行日志、用户系统操作日志、系统消息数据、冗余报表数据、系统对接中转数据等。

用户行为数据主要是指用户在使用过程中的行为记录,例如查询物流信息、投诉、评价等,此类数据对于分析用户行为和改进服务质量非常重要。

8.4 招投标数据

招标编号、招标企业、投标企业、招标文件、资审文件、投标文件、合同信息等。

8.5 订单流转数据

订单标题性资料:订单单号、订货日期、客户代号、客户名称、送货日期、送货地址等。

订单明细资料:货品代号、货品名称、货品单价、货品规格、订购数量、金额、折扣、交易类别等。

订单状态:是否完成、账目到位等。

8.6 过程跟踪数据

车辆形态、车辆位置、车辆轨迹、承运人信息等。

8.7 财务结算数据

订单金额、折扣信息、应收账款、资产抵押信息等。

9 数据存储

数据存储方式有集中式存储和分布式存储两种,按照数据的类别和特点进行选择,其特点如表2所示。

表2　存储方式表

项目	集中式存储	分布式存储
物理介质分布	物理介质集中布放	物理介质分布到不同的地理位置
数据上传	数据上传到中心	数据就近上传
对机房要求	对空间、承重、散热要求较高	要求较低，可采用多套低端的小容量的存储设备分布部署
存储设备	大型硬盘阵列、磁盘库和存储服务器	Hadoop、Ceph、GlusterFS等

数据仓库在存储数据时，在现有生产系统的基础上，对数据进行抽取、清理，并按照主题与类别有效地组织数据。在存储模式上，可参考Hdfs、Hbase及RDBMS相结合的模式。

10　数据建模

10.1　概述

数据建模分为：

——范式建模：依据数据仓库中的范式站在企业角度面向主题的抽象，而不是针对某个具体业务流程的实体对象关系抽象，它更多的是面向数据的整合和一致性治理；

——维度建模：是目前大数据场景下推荐使用的建模方法，面向分析场景而生，针对分析场景构建数仓模型；重点关注快速、灵活地解决分析需求，同时能够提供大规模数据的快速响应性能；

——数据值建模：一种中心辐射式模型其设计重点围绕着业务键的集成模式，这些业务键是存储在多个系统中的、针对各种信息，用于定位和唯一标识记录或数据。

10.2　核心步骤

10.2.1　选择业务过程

对业务全流程中的活动过程进行分析。

10.2.2　声明粒度

选择事实表的数据粒度。

10.2.3　维度设计

确定维度字段，确定维度表的信息。

10.2.4　事实设计

基于粒度和维度，将业务过程度量。

10.3　建模原则

10.3.1　易用性

冗余存储换性能，公共计算下沉，明细汇总并存。

10.3.2 高内聚低耦合

核心与扩展分离，业务过程合并，考虑产出时间。

10.3.3 数据隔离

业务与数据系统隔离，建设与使用隔离。

10.3.4 一致性

业务口径一致，主要实体一致，命名规范一致。

10.3.5 中性原则

弱业务属性，数据驱动。

11 数据模型

11.1 业务模型

主要解决业务层面的分解和程序化。按照业务部门的划分，进行各个部分之间业务工作的界定，理清各业务部门之间的关系、了解各业务部门的具体业务流程并将其程序化。界定数据建模的范围并划分整个数据仓库项目的目标和阶段。

11.2 领域模型

对业务模型进行抽象处理。本阶段主要工作为抽取关键业务概念，并将之抽象化。按照业务主线聚合类似的分组概念将业务概念分组；细化分组概念、理清并抽象化业务流程；理清分组概念间的关联关系，形成完整的领域概念模型。在设计物流数据的概念模型的时候可选择DWER模型进行建模。

11.3 逻辑模型

将领域模型的概念实体以及实体之间的关系进行数据库层次的逻辑化。通过逻辑建模，将概念模型完整串联成一个有机实体，表达业务间的关联性。

设计逻辑模型，可采用维度建模。事实表用来存储事实的度量及指向各个维的外键值。维度表用来保存该维的元数据，即维的描述信息，包括维的层次及成员类别等。在维度建模中可选择星型架构、雪花架构、星座架构等。

11.4 物理模型

解决数据的存储结构、索引策略、存储策略及存储优化等问题。根据数据仓库的逻辑模型，设计存储在数据仓库中表的结构，将领域概念模型中的实体映射为表格，表格中外键约束用来表示事实表和维度表之间的关系，实体的属性对应表格中列中的字段。在字段中主键约束用来唯一标识实体的实例。

由于数据仓库中的数据信息量比较大，可采用并行的存储结构，如RAID结构等。

在数据的索引策略上，为适应多维查询的环境，物流管理数据仓库可以采用Bit Map索引或Bit Wise索引等索引方式。

12 数据采集

数据采集层负责信息数据的汇集、转换与加载，提供多种数据采集方法，如ETL、Flume、Kafka等。数据仓库的接口由用户接口、业务量接口、账务接口等接口组成，并通过对相关表的设计具体实现。

数据仓库的数据主要来自企业自身使用的业务系统、标识码中的存储信息等。

13 网络安全

应按GB/T 20270—2006、GB/T 20271—2006、GB/T 28452—2012的规定执行。

14 数据备份与恢复

14.1 数据备份

应按GB/T 29765—2021的规定执行。

14.2 数据恢复

应按GB/T 20988—2007的规定执行。

15 运行系统的结构

15.1 数据元

数据元是数据仓库的基础，用于设计和定义数据仓库的元数据及数据来源，确定从数据源向数据仓库复制数据时的数据变换规则。可以包括内部系统的数据库、外部数据提供商的数据文件、Web上的数据源等。

15.2 数据提取

将数据从源系统中抽取出来并转换为数据仓库，采集产品包括编码发生器、归一化复制实用程序等。通过批处理、定时任务或实时流式传输等方式进行。

15.3 元数据管理

用于数据仓库管理工作，向其他部件提供使用和管理仓库数据集合的服务，以及对仓库数据和数据集合提供分配、安全、备份、归档、检测等处理服务。

15.4 数据清洗

数据清洗是数据仓库系统的重要功能之一，主要目的是对输入数据进行预处理和格式化。数据清洗的过程包括数据转换、数据规范化、数据匹配、数据删除等。通过数据清洗，可以确保进入数据仓库系统的数据质量，为后续的数据分析和决策支持提供可靠的基础。

15.5 查询与分析

数据访问和分析部件用于向最终用户提供存取、分析、研究仓库数据所需的工具，包括查询工具、数据分析多维产品等。

参 考 文 献

[1] GB/T 18354 物流术语
[2] GB/T 18768—2002 数码仓库应用系统规范
[3] GB/T 33745—2017 物联网术语
[4] 杨磊. 大数据的发展及数据仓库的融合应用[J]. 数字技术与应用,2019,37(06):62+64.