



数据中台 SG-CIM 模型应用方法

万齐鸣, 王思宁, 何鑫

(北京中电普华信息技术有限公司, 北京 100192)

摘要: 分析了国网电力物联网总体架构、数据中台分层架构, 研究了 SG-CIM 在数据中台定位与应用方式、模型对于数据分析应用的支撑。对于数据中台的物理模型设计, 在采用基线版管控的前提下, 建议采取创建平行拓展表和联合视图的方式, 进行个性化模型拓展设计; 对于数据中台模型数据的整合实施, 建议结合数据库选型, 采用批量数据装载方式, 以提高数据整合实施效率; 对于数据中台模型的升级与更新, 建议选择创建中间表方式, 以尽量减少对于数据分析应用的影响, 实现对于基于数据中台的数据分析应用的良好支撑。

关键词: SG-CIM; 数据模型; 数据中台; 模型应用

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020066

SG-CIM model application method in data middle platform

WAN Qiming, WANG Sining, HE Xin

Beijing China-Power Information Technology Co., Ltd., Beijing 100192, China

Abstract: The overall framework of State Grid ubiquitous power IoT and the layered architecture of data middle platform were analyzed. The positioning and application mode of SG-CIM in data middle platform, and the support of model for data analysis application were studied. The physical model design of data middle platform was put forward, on the premise of adopting the baseline version control, parallel expansion tables and joint views was created for personalized model expansion design. As to the integrated implementation of model data for data middle platform, combining database selection and adopt bulk data loading was proposed, so as to improve efficiency of data integration. As for the upgrading and updating of data middle platform model, creating a middle table to minimize the impact on the data analysis application was recommended, and a good support was achieved for the data analysis application based on data middle platform.

Key words: SG-CIM, data model, data middle platform, model application

1 引言

统一数据模型(SG-CIM)是国家电网公司参考国际标准(IEC 61970/61968/62325)和行业最

佳实践(SAP/ERP), 结合公司核心业务需求、在运系统数据字典等, 采用“业务需求驱动自顶向下”和“基于现状驱动自下向上”相结合的模式, 基于面向对象建模技术而构建的企业数据模型。

2019 年,国家电网公司提出建设“坚强智能电网、电力物联网”,打造“枢纽型、平台型、共享型”企业,即“三型两网”发展战略。SG-CIM 是国家电网公司电力物联网建设的重要基础,是电力物联网业务中台、数据中台(data middle platform)进行业务协同和数据共享的重要保障。

2 电力物联网总体架构

电力物联网是围绕电力系统各环节,充分应用移动互联、人工智能等现代信息技术、先进通信技术,实现电力系统各环节万物互联、人机交互,具有状态全面感知、信息高效处理、应用便捷灵活特征的智慧服务系统。

电力物联网建设通过升级信息通信基础架构、推动电网业务优化、促进新兴业务拓展,从而实现全时空通信覆盖、全环节物物互联、全业务云上运行、全数据统一管理、全方位数据应用、全过程可信互动,全面提升全息感知能力、泛在连接能力以及融合创新能力。

电力物联网主要包括感知层、网络层、平台层、应用层 4 个层次。感知层重点是统一终端标准,加强跨专业统筹,推动设备智能化升级,实现业务终端按需接入,提高边缘智能水平;网络

层重点是构建“空天地”协同一体化电力泛在通信网,增强网络带宽,提升网络资源调配能力,实现网络深度全覆盖;平台层重点是落实“一平台、一系统、多场景、微应用”信息化核心理念,构建企业中台,实现超大规模物联统一管理和高效处理能力,促进数据融通、支持应用灵活构建;应用层重点是全面提升核心业务智慧化运营能力,积极打造能源互联网生态,促进管理提升和业务转型。

电力物联网总体架构如图 1 所示。其中,电力物联网平台层主要包括云平台、物联管理平台、业务中台以及数据中台。

3 电力物联网数据中台

数据中台以企业多类型大数据量的汇聚为基础,以云平台为基础支撑,以 SG-CIM 为标准,通过丰富的数据标签,为前端应用提供敏捷的统一数据服务,为企业数据分析应用提供高效数据服务支撑。

数据中台包含数据技术层、统一数据层、统一服务引擎、数据资产管理以及数据运营管理等,为数据分析应用提供数据服务支撑。数据中台总体架构如图 2 所示。

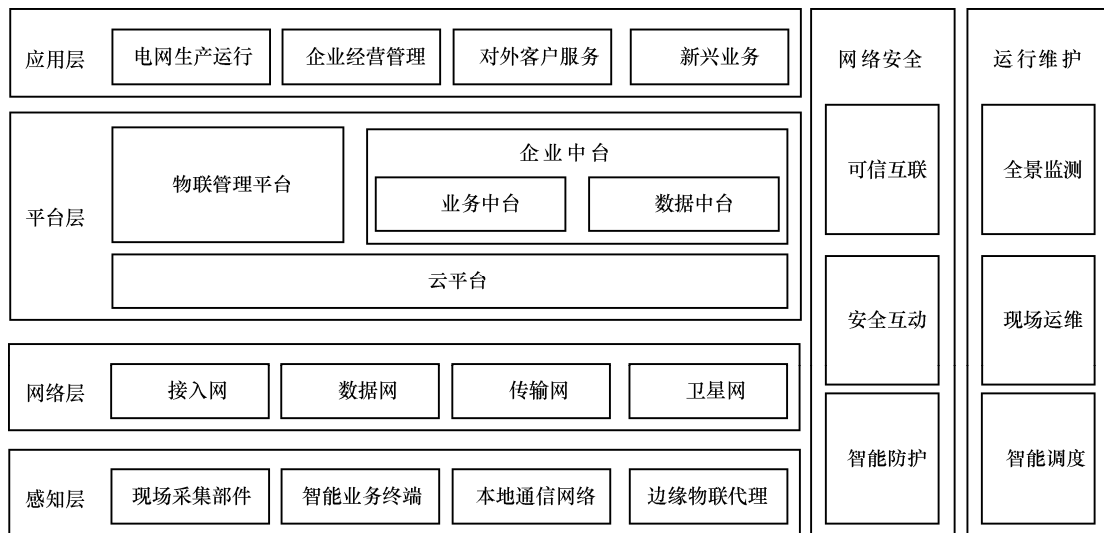


图 1 电力物联网总体架构

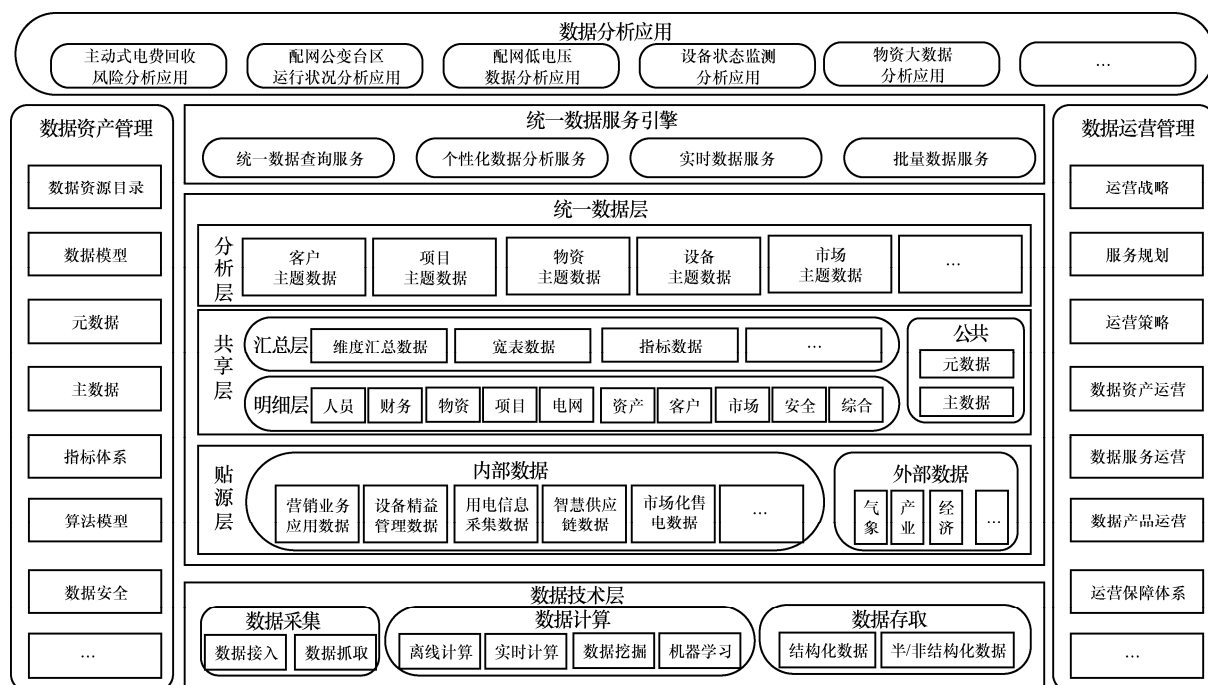


图2 数据中台总体架构

其中，统一数据层主要包含贴源层、共享层和分析层。贴源层是源端业务系统的数据复制；共享层是原始数据经过清洗整合后面向共享服务的数据，包含明细层和汇总层，明细层存储按照SG-CIM清洗整合后的业务明细数据，汇总层则存储维度汇总数据、宽表数据以及指标数据等；分析层则存储面向主题的数据集市数据，主要面向数据分析应用。

4 SG-CIM 在数据中台中的定位与应用

数据中台为企业各种数据分析应用提供统一的数据服务。然而，由于在企业信息化建设过程中，大都采用业务部门为主导的分散模式建设，与各专业部门业务高度耦合，信息化系统整体呈“竖井”孤岛分布，造成各专业数据定义不一致、跨专业协同困难、数据无法融通共享等问题。因此，直接使用从源端系统复制的贴源层数据，可能存在数据语义不一致、标准代码不一致、数据多源、数据无法关联等诸多问题，导致难以支撑

企业各类跨专业数据分析应用。

为解决横向业务协同困难、数据无法融通共享的历史难题，国家电网公司设计了SG-CIM统一数据模型，作为企业公共信息模型，包含概念模型以及逻辑模型，同时为推动SG-CIM模型落地应用，提供了面向数据中台应用的物理模型。

SG-CIM物理模型是数据中台面向数据分析应用提供统一数据服务的基础，主要应用于数据中台统一数据层中的共享层，作为明细数据物理模型，存储从贴源层数据按照SG-CIM模型清洗整合后的业务明细数据，并为数据中台汇总层、分析层构建提供支撑。

5 数据中台模型对于数据分析应用的支撑

数据中台的贴源层、共享层（明细层、汇总层）、分析层均需对应的物理模型支撑。其中，贴源层物理模型与源系统物理模型基本一致；明细层物理模型则包含两部分：一是SG-CIM统一提供的物理模型基线版，二是省市公司结合本地数

据分析应用需求自行拓展设计的个性化拓展物理模型；汇总层物理模型则是省市公司结合本地数据分析应用需求设计的包含维度汇总数据、宽表数据、指标数据等的物理模型；分析层物理模型也是根据本地数据分析应用需求而设计的面向主题数据集市物理模型。通过数据中台各个层次物理模型的安装部署，实现各层数据的存储、管理与服务。

数据中台物理模型分布与数据流向如图 3 所示。

贴源层从源端业务系统复制和接入数据；明细层首先按照 SG-CIM 物理模型基线版从贴源层进行数据溯源，再按照 SG-CIM 模型进行数据清洗、转换和数据接入，针对省市公司个性化需求，则按照个性化拓展物理模型，从贴源层进行数据溯源、数据清洗转换和数据接入；汇总层则基于省市公司汇总数据需求，按照汇总层物理模型，从明细层获取基础数据，进行轻度数据汇总并按照汇总层物理模型进行数据存储；分析层基于省市公司数据分析应用需求，从明细层、汇总层获取数据，构建面向主题的数据集市并按照分析层

主题数据集市物理模型进行数据存储。

数据中台基于明细层、汇总层、分析层物理模型进行数据存储管理，并通过统一数据服务引擎，为各类数据分析应用提供数据服务支撑。

6 数据中台个性化模型拓展设计

为推动模型落地应用，虽然 SG-CIM 提供了物理模型，但只是一个面向数据中台明细层的基线版本，该基线版本由国家电网公司总部统一设计和维护管理，省市公司不允许直接修改该物理模型基线版本。数据中台明细层数据是由汇总层、分析层构建的数据基础，同时也是数据分析应用的数据基础。然而省市公司数据分析应用需求存在差异，可能出现基线版物理模型难以满足本地全部数据分析应用需求的情况，因此，省市公司需要根据个性化需求进行个性化物理模型拓展设计。

省市公司在构建本地数据分析应用时，首先需要进行数据需求分析，并与总部下发的 SG-CIM 明细层物理模型基线版进行比对，分析 SG-CIM 基线版物理模型是否满足本地数据分析应用需求，并进行明细层物理模型个性化拓展设计。

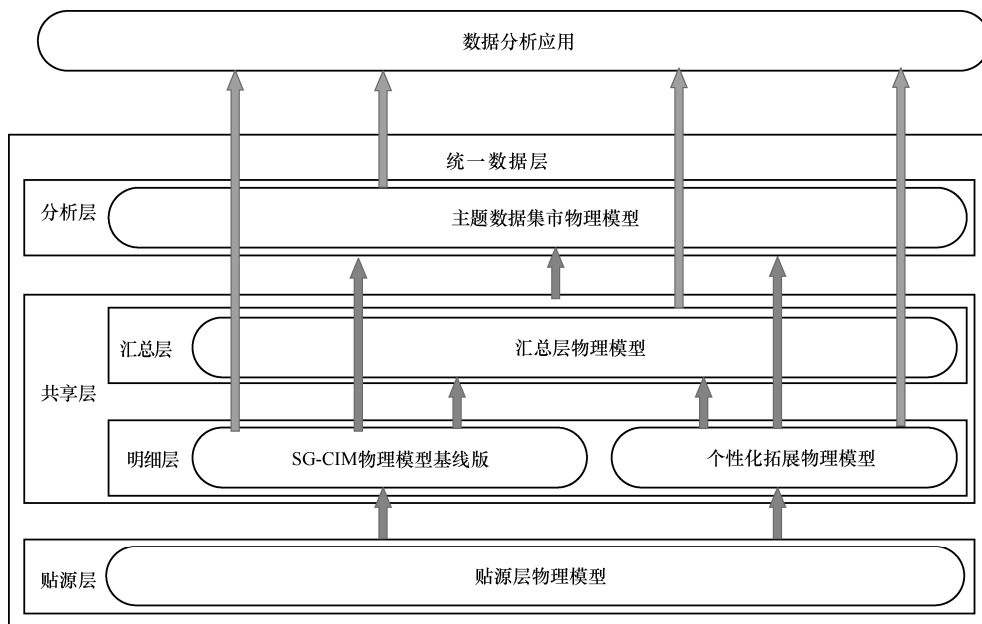


图 3 数据中台物理模型分布与数据流向



数据需求与 SG-CIM 物理模型基线版的对比分析结果分为如下 3 种情况。

(1) 完全匹配, 直接使用总部下发的物理模型基线版标准模型表, 满足本地数据需求。

(2) 部分匹配, 进行个性化物理模型拓展设计, 创建与基线版物理模型表平行的补充数据表, 包含标准模型表未覆盖的需求数据字段, 使用基线版物理模型表主键进行两表的关联, 创建包含标准模型表与平行补充表的联合视图, 通过联合视图满足本地数据需求。部分匹配时, 个性化物理模型拓展设计方法如图 4 所示。

(3) 无模型支撑, 新建补充物理模型表, 包含需求数据字段、定义字段类型及字段长度、定义数据表类型、定义主外键等, 并使用补充物理模型表, 满足本地数据需求。

数据中台明细层个性化拓展模型设计完成后, 需要按照数据分析应用需求进行共享层汇总数据物理模型以及分析层主题集市物理模型设计; 模型设计和准备完成后, 需要在数据中台各层进行对应物理模型的部署和安装。

7 数据中台模型数据整合

数据中台各层次模型准备和安装部署完成

后, 需要按照模型及需求进行数据整合实施。模型数据整合实施包括从源端业务系统复制数据至贴源历史区、从贴源历史区抽取转换加载数据至明细数据区以及从明细数据区进行数据汇总计算并将结果存储至汇总数据区, 同时基于明细数据、汇总数据按照数据分析应用需求进行进一步汇总计算构建面向主题的数据集市, 并将结果存储至分析层主题数据集市区。数据抽取主要包括存量抽取、增量抽取两种方式。

数据中台模型数据整合实施过程如图 5 所示。

数据中台明细层数据库选型由于考虑海量数据大用户量并发应用需求, 多采用 MPP 并行数据库, MPP 数据库侧重于满足数据访问和分析应用性能, 但对于数据库的记录修改维护的效率并不高, 导致数据库在进行数据批量装载时性能远高于逐条记录维护的方式。如对于国网原全业务统一数据中心的明细层 GBase 8a 数据库, 在国网陕西电力全域综合应用示范测试, 同等配置下, 逐条加载为 300 条/s, 批量数据装载 18 万条/s。

因此, 对于数据中台模型数据整合实施, 存量数据的抽取、转换、加载应尽可能采用批量装载方式, 以提高数据加载效率; 对于增量数据的抽取, 也可考虑先将增量数据存入临时增量区,

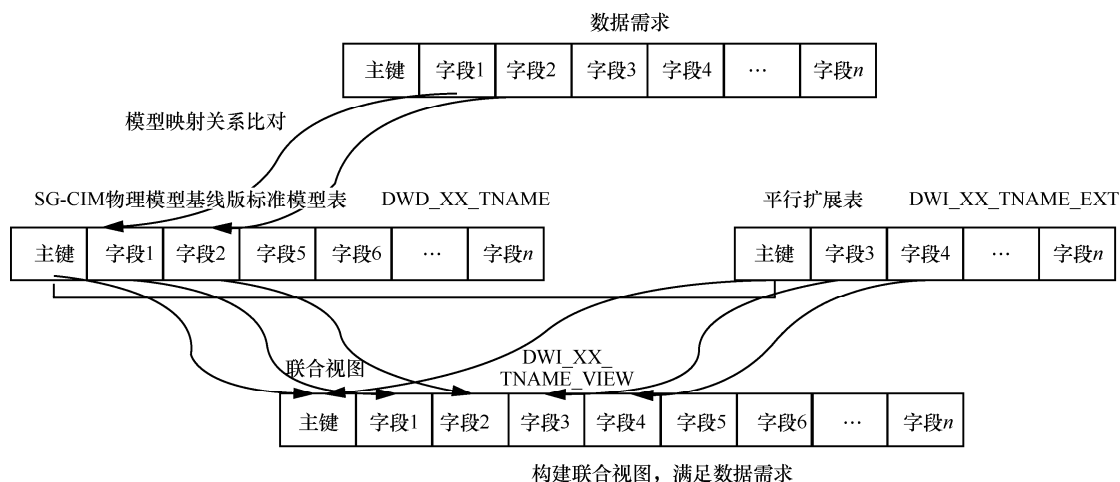


图 4 部分匹配个性化物理模型拓展设计方法

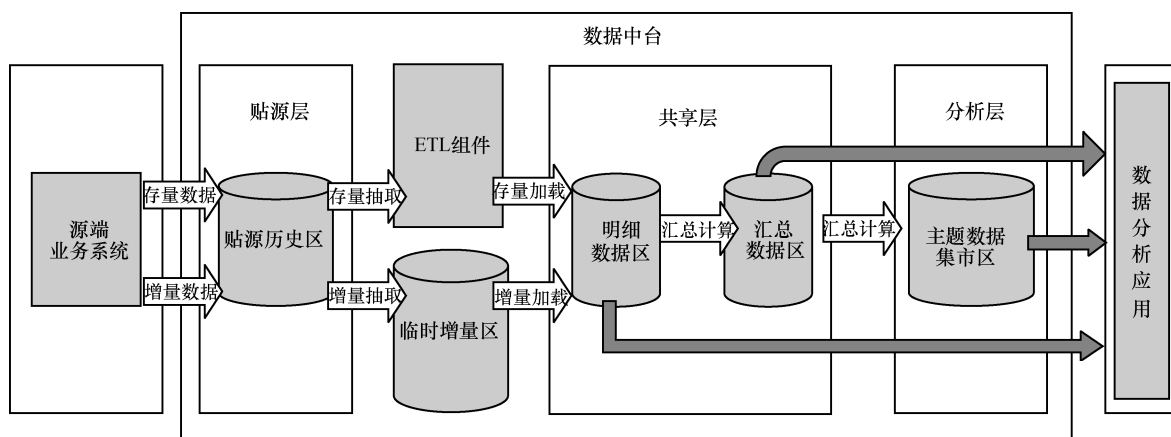


图 5 数据中台模型数据整合过程

再通过批量装载方式进行数据入库。

8 数据中台模型更新与升级

数据中台各层物理模型安装部署完成后，需要进行后续数据的抽取、转换、整合和数据加载，以提供统一数据服务，支撑各类数据分析应用。然而，由于 SG-CIM 仍然处于一个持续迭代和优化完善的阶段，随着其新版本模型的发布，必然会导致省市公司数据中台物理模型需要进行同步更新与升级，同时，省市公司本地数据分析应用需求的新增等变化，也可能导致数据中台各层模型变化，从而需要进行更新与升级。

数据中台模型更新与升级将会带来物理数据库结构的变化，需要重新进行数据抽取、转换和加载，将影响现有 ETL 的配置，需要进行 ETL 重构，同时在数据抽取、转换、加载的过程中，可能导致数据不可用，从而进一步地影响对数据分析应用的支撑。因此，需要审慎采用模型更新的升级方法，以尽可能地降低对于各类在运数据分析应用的影响。

一般情况下，数据中台模型更新与升级有 3 种方法。

方法 1 整表删除并重建。先删除数据中台中需要更新的数据表，然后按照模型更新，创建新的同名数据表，修改 ETL 配置，重新进行数据

抽取、转换和数据加载。

方法 2 创建中间表。先按照模型更新，创建一个中间表，重新配置 ETL 并指向中间表，进行数据抽取转换和数据加载，再在数据分析应用的空闲时点删除原数据表，修改中间表名称为原数据表名称，实现中间表的快速切换，并重新按原表名修改 ETL 配置。

创建中间表进行模型升级更新的过程如图 6 所示。

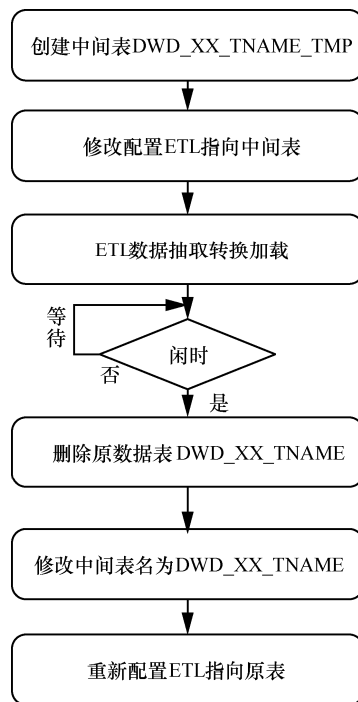


图 6 创建中间表进行模型更新升级的过程



方法 3 在线修改表结构。直接通过修改表结构的 SQL 语句进行数据表结构修改,如增加或者修改字段,然后开发脚本对于变化的字段内容进行逐条历史数据维护,保证数据的一致性和完整性。

由于数据中台数据库选型不同,采用何种方法进行数据中台模型更新与升级,对于数据分析应用的影响程度也各不相同。面向事务处理类数据库时,数据更新效率较高,可以承受增量逐条更新的方式;而面向分析应用数据库时,需要尽量避免逐条数据更新方式,采用批量数据装载方式以提高效率。

对于模型升级更新,方法 1 最为简单,由于采用批量数据装载方式,性能较高,对于数据分析应用的影响时间与数据装载时间直接相关,如数据量很大,也会造成数据分析应用的中断,此方法一般情况下适用于数据中台初建,数据分析应用尚未完全开展的情况;方法 2 较为复杂,需要创建中间表,需要修改两次 ETL 配置,但可以选择在凌晨或者闲时等不同时点进行模型更新和切换,对于数据分析应用的影响可以做到分钟级以内,甚至忽略不计;方法 3 较为复杂,虽然看似保留了原有数据,但数据表字段变化或更新,需要进行历史数据补齐,由于采用逐条维护方式,性能较低,一般仅适用于较小数据量的情况,限制较多。

由于数据中台模型更新升级会给 ETL 带来影响,需要重新进行数据抽取、转换和数据加载,同时数据加载又会影响数据中台对于数据分析应用的支撑,因此需要审慎考虑,结合具体情况再决定选择采用何种模型更新升级的方法。

9 结束语

基于数据中台支撑各类数据分析应用是一个较长的业务链条,首先需要进行数据需求分析;结合需求进行 SG-CIM 物理模型基线版比对,对于不能满足的需求,需要针对其进行模型拓展设

计,模型准备完成后进行模型安装部署;按照模型进行数据抽取、转换与数据加载,构建汇总层数据和面向主题的数据集市;模型更新与升级还需要重新进行模型部署和数据装载;数据准备完成后还需设计开发数据服务接口,进行数据分析应用相关的算法、功能以及可视化等方面的设计开发;最后进行测试上线。

对于数据中台物理模型设计,在采用基线版管控的前提下,建议采取创建平行拓展表和联合视图的方式,进行个性化模型拓展设计;对于数据中台模型数据整合实施,则建议结合数据库选型,采用批量数据装载的方式,以提高数据整合实施效率;对于数据中台模型升级与更新,则建议选择创建中间表方式,以尽量减少对数据分析应用的影响,以实现对于基于数据中台的数据分析应用的良好支撑。

参考文献:

- [1] 刘鸿宁,孔晓昀,黄文思,等.基于 SG-CIM 的企业数据库模型设计与应用[J].信息技术,2018(4).
LIU H N, KONG X Y, HUANG W S, et al. Design and application of SG-CIM based enterprise database model[J]. Information Technology, 2018(4).
- [2] 吴涛,刘韬,王斌.安徽联通企业级大数据平台构建及应用实践[J].电信科学,2018,34(1): 135-147.
WU T, LIU T, WANG B. Construction and application of Anhui Unicom enterprise big data platform[J]. Telecommunications Science, 2018, 34(1): 135-147.
- [3] 谢淑菁,王浩,张志伟.基于 CIM 的电网建模若干问题研究[J].电网与清洁能源,2011(2).
XIE S J, WANG H, ZHANG Z W. Discussions about power network modeling based on CIM[J]. Advances of Power System & Hydroelectric Engineering, 2011(2).
- [4] 李炳森,胡全贵,陈小峰,等.电网企业数据中台的研究与设计[J].电力信息与通信技术,2019(7).
LI B S, HU Q G, CHEN X F, et al. Research and design of data center of power grid enterprises[J]. Electric Power Information and Communication Technology, 2019(7).
- [5] 施传新,谢志林,楚黄维,等.基于数据挖掘的在线审计模型设计[J].电子技术与软件工程,2019(20).
SHI C X, XIE Z L, CHU H W, et al. Design of online audit model based on data mining[J]. Electronic Technology & Software Engineering, 2019(20).

- [6] 王彦明, 陈卫东, 黄恒, 等. 泛在电力物联网数据概念结构模型设计及数据抽象方法[C]//中国电机工程学会电力通信专业委员会. 电力通信技术研究及应用. 北京: 人民邮电出版社, 2019: 9.
- WANG Y M, CHEN W D, HUANG H, et al. Conceptual structure model design and data abstraction method of ubiquitous power Internet of things data[C]// Electrical communication Committee of Chinese Society of Electrical Engineering. Research and application of electric power communication technology. Beijing: Posts Telecom Press, 2019: 9.
- [7] 陈军. 泛在电力物联网功能架构初探[J]. 通信电源技术, 2019(7).
- CHEN J. Preliminary study on functional architecture of ubiquitous power internet of things[J]. Telecom Power Technology, 2019(7).
- [8] 李静, 马超, 黄镜宇, 等. 跨平台信息系统数据迁移技术研究与应用[J]. 电气技术, 2019(10).
- LI J, MA C, HUANG J Y, et al. Research and implementation of data migration technology of cross-platform information system[J]. Electrical Engineering, 2019(10).
- [9] 陈孝文, 陈文超, 刘振. ETL 技术在电网信息系统业务数据迁移中的应用[J]. 自动化与仪器仪表, 2019(8).
- CHEN X W, CHEN W C, LIU Z. Application of ETL technology in business data migration of power grid information system[J]. Automation & Instrumentation, 2019(8).
- [10] 于浩淼, 赵月芳, 陈盟, 等. 企业中台建设思路与实践方案[J]. 电信技术, 2019(8).
- YU H M, ZHAO Y F, CHEN M, et al. Idea and practice of construction in enterprise middle platform[J]. Telecommunications Technology, 2019(8).
- [11] 秦成. 制造业企业中台建设思考与实践[J]. 智能制造, 2019(2).
- QIN C. Thoughts and practice on the construction of manufacturing enterprises middle platform[J]. Intelligent Manufacturing, 2019(2).
- [12] 杨冬菊, 徐晨阳. 大数据环境下基于元模型控制的数据质量保障技术研究[J]. 计算机工程与科学, 2019(2).
- YANG D J, XU C Y. Data quality assurance based on metamodel control in big data environment[J]. Computer Engineering and Science, 2019(2).

[作者简介]



万齐鸣(1971—), 男, 北京中电普华信息技术有限公司高级工程师, 主要研究方向为大数据、智能计算、数据建模、知识管理等。

王思宁(1978—), 女, 北京中电普华信息技术有限公司电网运营监测事业部总经理、高级工程师, 主要研究方向为电力信息化项目管理及标准体系构建等。

何鑫(1988—), 男, 北京中电普华信息技术有限公司电网运营监测事业部副总经理, 主要研究方向为项目管理、数据模型设计等。