文章编号：1002-6673（2024）06-134-04

企业智能数据中台的建设探索

吕依濛[1]，金风明[2]，周立博[2]
（1.中煤科工开采研究院有限公司，北京 100013；2.北京东方国信科技股份有限公司，北京 100102）

摘 要：数据是企业和组织最宝贵的资产之一，本文结合企业数据中台建设的落地实践，探索了企业数据中台建设架构、建设步骤和价值，研究了智能数据中台功能体系，为企业数据价值发掘提供建设思路。

关键词：智能数据中台；平台功能架构；数据价值
中图分类号：TP39 **文献标识码**：A doi:10.3969/j.issn.1002-6673.2024.06.036

Exploration of Building an Intelligent Data Platform for Enterprises

LV Yi-meng[1], *Jin Feng-ming*[2], *Zhou Li-bo*[2]
（1.Coal Mining Research Institute Co.,Ltd.,CCTEG,Beijing 100013,China;
2.Business-intelligence of Oriental Nations Corporation Ltd.,Beijing 100102,China）

Abstract: Data is one of the most valuable assets for enterprises and organizations. This article explores the architecture, construction steps, and value of enterprise data platform construction, and studies the functional system of intelligent data platforms, providing construction ideas for exploring the value of enterprise data.

Keywords: Intelligent data platform; Platform functional architecture; Data value

0 引言

2022 年，国家发布了《关于构建数据基础制度更好发挥数据要素作用的意见》，加快构建数据基础制度，充分发挥我国海量数据规模和丰富应用场景优势，激活数据要素潜能。智能数据中台聚焦数据汇聚、处理、流通、应用、运营、销毁等技术，为数据生命周期提供技术支撑。平台旨在协助解决企业数据域管理面临的诸多问题，帮助建立企业级数据中心，自上而下开展大数据治理；利用数据管理平台提供的基础能力，将企业数据集成和整合，实现数据的统一管控和精细化管理，保障数据资产质量；支持海量数据应用，推动企业大数据分析，挖掘数据潜在价值，为企业战略决策、业务协同、风险管控等提供有力支持[1]。

1 智能数据中台总体架构

智能数据中台实现数据仓库数据模型标准化、数据关系脉络化、数据加工可视化、数据质量度量化、数据资产显像化、数据服务自动化、数据安全可控化等要求的一体化平台[2]。平台的总体功能架构如表 1 所示。

表 1 平台总体功能架构
Tab.1 Overall functional architecture of the platform

功能名称	功能模块
数据建模	包括逻辑建模、物理建模等
数据集成	包括离散数据和实时数据的采集与开发等
数据质量	包括稽核规则、稽核任务、质量问题、质量评估等
数据资产	包括资产管理、资产目录、资产价值评估、资产配置管理等
数据服务	包括服务管理、服务目录、服务订阅等
统一调度	包括调度流程管理、流程设计、流程监控、任务监控等
数据安全	包括数据安全分类分级、数据安全处理、数据安全策略、审计等
元数据	包括元数据模型、元数据采集、元数据目录、元数据应用等

1.1 数据建模

数据建模提供规范化、显象化的数据建模管理模式，模型设计继承已定义标准规范，提供贯穿数据的开发、部署、治理等各个阶段的模型管理功能，记录数据仓库模型建设的全过程。

1.2 数据集成

数据集成实现离散数据和实时数据的采集与开发，

修稿日期：2024-10-22
作者简介：吕依濛（19[redacted]-），男，北京人，硕士，主要研究方向：容器技术、大数据分析、人工智能。

解决数仓由原始数据向报表数据逐步转化加工的问题。

1.3 数据质量

数据质量解决数仓中数据是否可用问题,实现数据全生命周期的质量监控与质量稽核,生成完整的数据质量综合报告,保障数据的完整性、准确性、一致性、及时性等。

1.4 数据资产

数据资产生命周期管理,实现从资产盘点、评估、治理、共享、到资产监控的一体化服务,全方位展示数据中心所有数据间的脉络关系,数据特性及质量情况。全景视图体现数据中心所有对象的关系,通过分层及数据流向展示数据中心整体情况,实现对大数据中心数据资产的全方位展示以及数据的溯源和去向分析。

1.5 数据服务

数据服务解决数据中心对外数据流动的问题。提供基于数据仓库的全量数据,发布成标准化数据产品,统一对外提供数据服务。并提供数据产品申请、审核、发布、获取、监控、评估完整的闭环业务流程,形成完善的服务运营管理机制。

1.6 统一调度

统一调度解决数仓各种加工生产作业执行时序的问题。提供数据仓库调度流程的统筹管控,根据企业级数据仓库的加工层级,对于定期加工的源头流程采用计划调度,下游流程采用事件触发,临时性的流程采用手动启动,三类调度方式混合使用可以应对全部调度场景需求。

1.7 数据安全

数据安全解决数据中心数据安全性的问题。提供各种访问控制、数据加解密、数据脱敏、水印和数据审计等功能。同时提供了安全策略制定,用户安全态势监控和日志分析预警以及审计功能,打造事前防范,事中管控,事后评估全流程安全体系。

1.8 元数据

元数据管理能够对数据仓库的各类元数据进行统一管理,并且完整记录数据处理链路的血缘关系,形成企业全局数据地图,助力企业方便轻松的管理数据仓库的海量数据。

2 智能数据中台存储技术

数据中台采集并存储了大量结构化和非结构化数据,其中存在图片、视频等多媒体数据,存储空间大,读取速度慢。数据读取访问速度渐渐成为中台性能的瓶颈。为了解决现有数据存储性能的瓶颈问题,数据中台创建了一种数据校验及存储的方法,极大提升中系统数据的读取效率。

2.1 数据的存储格式

数据中台的数据格式分为五部分:数据校验等级(未校验)、数据本体信息、数据采集来源、读取量 n1、修改量 n2 、是否精密数据。

在可靠性要求较高的存储系统中,一般采用 RAID (Redundant Arrays of Independent Disks,磁盘阵列)对数据进行冗余处理,以实现当部分磁盘损坏或数据块时被损坏。由于数据中台采集数据量大,且并不是所有信息在系统中全部被使用。因此,系统使用的存储分为三个区。

(1)非校验区。连续以位或字节为单位分割数据,并行读写于多个磁盘上,因此具有很高的数据传输率,但是没有数据冗余,满足快速、大量、高性能读写存储的功能。

(2)镜像区。通过磁盘数据镜像实现数据冗余,在成对的独立磁盘上产生互为备份的数据。当原始数据失效时,系统可以自动切换到镜像磁盘上读写。

(3)高精度区。数据以位或字节为单位分割后存入各磁盘,并使用国产 SM2 算法进行加密校验码,SM2 具体算法可参考相关公开文献,使用 SM2 屋里加密机,利用校验冗余信息提供错误检查及恢复,安全性极高。

三种存储区的比较如表 2 所示。

表 2 三种存储区的性能对比

Tab.2 Performance comparison of three storage areas

序号	功能名称	非校验区	镜像区	高精度区
1	读写速度	高	中	低
2	安全性	低	中	高
3	读写性能	高	中	低

2.2 数据的校验步骤

进入数据中台的数据按照如下步骤进行校验:

(1)采集系统推入 NAS 存储的数据,默认进入非校验区,数据校验等级为未检验,4 和 5 初始值为 0。连续以位或字节为单位分割数据,并行读/写于多个磁盘上,有很高的数据传输率,能够实现高性能快速存储,没有数据冗余,一旦发生数据错误或缺失,无法靠磁盘冗余恢复。由于出错率较少,发生缺失时,需依靠数据采集来源字段,返回采集系统,进行重新采集。

(2)计算存储的信息数据读取量 n1,修改量 n2,利用计算公式

N=(lgn1+lgn2)/10,当 0<N<=0.8 时,将该数据转移到镜像区,当 0.8<N<1 时,将该数据转移到高精度区。

(3)当字段 6 标记为精密数据时,将该数据移动到高精度区。

通过独创此种架构及校验方式,可实现太极六合信创知识库系统百万级数据的高性能快速存储,在保证读写性能的同时,对于经常访问修改的新创知识信息进行冗余设置,保证了数据的准确性。对于保密数据及高精度数据,精密区极大程度保证了数据的安全性、准确性及可恢复性。

2.3 非校验区的结构

非校验区分为新推送区和旧推送区，存储大小比例为 2:8，当数据推送到非校验区时，新推入数据默认存储到新推送区，记录数据存入的时间，这里设定一个阀值 N，当数据的校验等级没有改变时，当时间达到 N 时，使用复制算法，将该数据转移到旧推送区，同时修改量 n2+1。

旧推送区的存储方式采用上文 1 中描述的传统磁盘存储，而新推送区采用 SSD 存储，同样不做冗余校验，以保证高速存储效率[3]。

3 智能数据中台建设与应用

3.1 企业数据管理问题

(1)数据分散割裂未整合：数据以域、系统为单位“烟囱”式分布，无横向关联，造成数据壁垒，无法从企业管理和整体视角，全面快速分析诊断痛点问题，发挥数据价值。

(2)数据标准字典未统一：数据治理体系建立不完整，未形成企业级数据治理体系和整合机制，没有定义企业统一标准数据字典，数据标准规范未及时更新，与现状不符。

(3)数据敏捷开发未具备：数据开发的采集、加工过程停留在需要大量有经验的数据开发人员写脚本阶段，成本高，复用低、上手难，无法实现自动化数据任务的统一编排和调度。

(4)数据质量安全未保障：在数据全生命周期过程中，没有对每个环节的数据进行识别、度量、分析、改进等一系列管理活动，保障数据的可用性和安全性[4]。

(5)数据智能服务未形成：数据共享服务停留在人工化、文档化的传统管理模式，无法将数据价值成果直接提供用户使用，实现数据开放共享运营分析能力。

(6)数据全局视图未建立：未对企业数据资产进行全面盘点和监控，大量企业数据散布在各业务和 IT 系统中，无法一点看全和及时掌握企业数据有什么、在哪里、发生了什么变化。

3.2 建设目的

为了解决上述企业问题，探索开展企业数据管理平台，进行企业统一的数据标准和规范顶层设计，完成数据元标准、数据模型标准、数据交换标准、数据安全标准、数据治理标准等系列标准规范的建设[3]，为后续企业信息化建设及数据应用提供标准规范和指导，促进数字化与业务提升的转型。建设企业数据仓库，实现企业数据资产的管理，以采购业务、人力资源业务、经营管理业务、科研管理业务、项目管理业务、投资管理业务为核心，以人员数据、项目数据、客户数据、供应商数据、物料数据、投资数据为业务主线，实现企业数据资产的标准化、规范化、高质量的管理。

3.3 平台建设步骤

智能数据中台建设整体内容如图 1 所示，包括咨询规划、平台建设、服务应用等三部分。

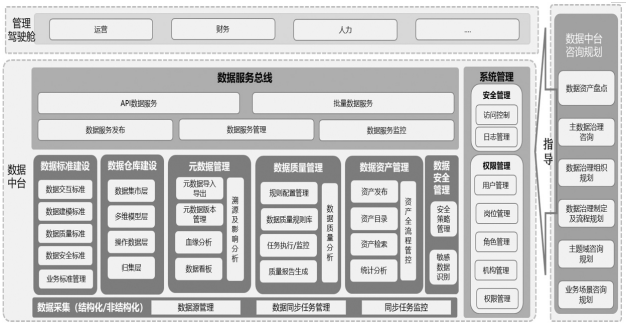

图 1 智能数据中台建设内容

Fig.1 Construction content of intelligent data platform

(1)咨询规划。咨询规划通过对现状进行深入调研，进行数据资产盘点与业务蓝图规划，完成数据盘点(盘点内容如图 2 所示)主数据治理实施、数据治理、业务应用场景等内容的咨询规划工作，为后续的数据中台实施提供管理框架及指导方向。

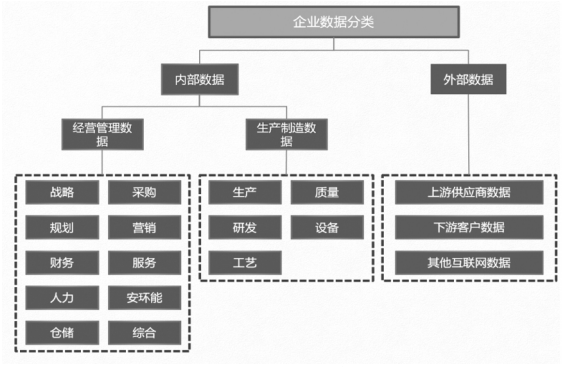

图 2 数据盘点

Fig.2 Data inventory

(2)平台建设。涵盖数据采集、数据标准建设、数据仓库建设、元数据管理、数据质量管理、数据资产管理、数据安全管理等数据全链路管控，通过大数据技术及工具支持，构建功能完善的数据中台(如图 3 所示)，夯实数据底座能力。

图 3 数据中台工作台

Fig.3 Data platform workbench

(3)数据服务应用。通过数据中台建设，形成规范化、及时化的数据服务体系和数据指标(如表 3 所示)，对外

提供统一数据服务，并以实际业务需求为牵引设计具体应用场景，构建管理驾驶舱，支撑公司决策应用。

表 3 数据指标体系

Tab.3 Data indicator system

一级指标主题	二级指标模块	指标名称	指标周期
运营管理主题	利润总额	利润总额目标值	年
	核准	核准目标值	年
		核准	天
	投产	投产目标值	年
	核准完成情况	核准完成值(集团)	年
		核准完成值(经营实体企业)	年
	投产完成情况	投产完成值(集团)	年
		投产完成值(经营实体企业)	年
	利润总额完成情况	利润总额完成值(集团)	年
		利润总额完成值(经营实体企业)	年
	企业财务分析指标对标	营业利润同比增长率	季度
		总资产周转率	季度
		总资产同比增长率	季度
		资产负债率	季度
		总资产净利率	季度
		净资产收益率	季度

4 智能数据中台建设价值

(1)解决了企业数据管理问题，规范了企业数据标准。通过统一数据标准制定，实现数据标准的统一管理。提供基于行业、主题、服务的统一标准规范制订，包括目录、数据元以及代码集的管理，促进数据标准规范的实施落地。

(2)制定了数据管理制度，提升了企业数据质量。由于数据质量可能出现从“数据产生”到“数据集成”再到“数据使用”在内的全过程，当出现数据质量问题，责任不清，数据质量问题的处理难以得到保障。数据质量管理集成流程管理功能，基于企业对数据质量问题处理的要求，可以灵活制定数据质量问题处理流程，指定流程处理的责任人，方便问题数据的处理和跟踪，形成企业数据质量管理流程制度。

(3)打通各业务线数据，进行精细化运营。目前，各系统彼此相互独立，同类型数据分散在不同业务系统中，仅能通过导出分析比对，才能看到数据差异。数据中台可实现统一的大数据分析，供给各个业务前端人员进行使用。

5 结束语

在当今数字经济时代，数据已成为企业不可或缺的重要资源。本文结合在企业数据中台建设的实际探索，分析了目前企业数据管理存在的问题，开展了数据中台建设的探索，研究了智能数据中台的功能架构，并开展了落地实践。通过平台，规范了企业数据标准，提升了企业数据质量，帮助企业实现了精细化运营。

参考文献

[1] 赵中良，梁建宾. 基于区块链技术配电网领域的发展应用前景[J]. 数字技术与应用期刊，2020，11.

[2] 赵明，林峰.有线电视网络大数据中心数据治理探析[J].广播与电视技术期刊，2019，12.

[3] 中煤科工开采研究院有限公司. 数据处理方法及装置：中国，2024103174206[P].2026-06-11.

[4] 蒋秀芳，于皓杰.大数据环境下电力企业智慧型运营监测管理研究[J].华北电力大学学报(社会科学版)期刊，2018，12.