

包钢财务域数据治理及应用

王宏¹, 芦建文², 王晓云², 耿殿威²

- 包头钢铁(集团)有限责任公司信息服务中心, 内蒙古 包头 014010;
- 内蒙古新联信息产业有限公司, 内蒙古 包头 014010)

摘要: 针对包钢股份财务域存在的数据共享不及时、风险预警不及时等问题, 结合包钢股份财务部门的业务情况与未来规划, 以数据的“采、存、管、用”为主线, 规划设计包钢数据治理体系建设、数据平台搭建、数据应用设计等内容, 通过数据仓库设计和数据平台应用设计满足复杂多变的数据应用场景。在内部有效利用数据支撑业务决策, 改善经营指标; 对外部构建客户画像、剖析客户交易, 从而辅助营销, 提升企业经营管理效率。

关键词: 数据管理; 数据治理; 数据仓库设计; 数据平台

中图分类号: F406.72

文献标识码: B

文章编号: 1009-5438(2022)06-0072-04

Data Governance and Applications in Financial Domain of Baotou Steel

Wang Hong¹, Lu Jian-wen², Wang Xiao-yun², Geng Dian-wei²

- Information Service Center of Baotou Iron & Steel (Group) Co., Ltd., Baotou 014010, Inner Mongolia Autonomous Region, China;
- Inner Mongolia Xinlian Information Industry Co., Ltd., Baotou 014010, Inner Mongolia Autonomous Region, China)

Abstract: Such aspects as the establishment of data governance system, construction of data platform and data application design are planned and designed by taking “acquisition, storage, management and application” of data as the main line, aiming at such existing problems that the data sharing and risk early warning in financial domain of Baotou Steel are not timely as well as combining with the business situations and future plan of financial department of Baotou Steel. Moreover, the complex and changeable data application scenarios are satisfied through data warehouse design and data platform application design. The business indicators are improved and business decisions are supported through effectively using data internally; the management efficiency of enterprise is improved and marketing is assisted through constructing client portrait and dissecting customer transactions externally.

Key words: data management; data governance; data warehouse design; data platform

近年来,随着大数据、云计算、区块链、AI等技术迅速发展,数字化转型成为国民经济和社会发展的重要风向标,企业数字化转型升级既是顺应时代发展的必然要求,也是企业信息化发展的必经阶段。

数据已经成为继土地、劳动力、资本、技术之后的第五大生产要素,数据作为一种新型的生产要素,已成为推动国民经济高质量发展的重要动能,也是推动企业数字化转型的重要基础。只有做好数据治

理,充分挖掘数据价值,才能更快、更好地推进数字化转型。

面向“十四五”,包钢将全面开展信息化、数字化建设,依托多业态、全产业链的发展优势,开展新一代信息技术与制造装备融合的集成创新和工程应用。

以包钢发展战略为纲领,以包钢股份财务域为切入点,开展数据治理工作,构建包钢数据治理框架。针对数据管理组织、制度、流程、工具以及落地实施的举措等方面进行细化和分解,输出可信、安全的数据,最终能够以数据的深化治理和资产运营赋能业务发展,助力包钢数字化转型。

1 包钢财务域数据管理存在的问题

包钢股份财务部门日常进行数据管理时面临的问题主要分为两个方面,一是财务业务方面问题,二是财务数据管理方面问题。

1.1 财务业务问题

包钢财务方面存在对业务响应不及时的问题、对业务需求缺少提前洞察的问题、对风险缺少提前预警的问题,具体如下。

(1)财务对业务响应不及时。具体体现在财务需要花费大量时间在手工编制各类财务报表、生成财务报告上,无法投入更多的时间在数据分析与业务指导上,财务数据也缺乏时效性,不利于决策支持。

(2)财务对业务运作缺少洞察。具体体现在缺少业务过程完整清晰的记录,缺少数据联动机制,同时在面对维度多重的经营决策时,仅能“靠要”、“靠表”,只知其果,不知其因,无法深度支持决策。

(3)财务对风险缺少监控预警机制。具体体现在财务无法对采购价格、销售价格、资金管控等项目开展事前或事中预警。

1.2 财务数据管理问题

海量数据分布在不同系统和电脑中,缺少数据价值挖掘以及数据共享机制,也存在数据安全隐患,容易造成信息泄露,还存在物料编码一物多码的情况,降低管理效率。

2 包钢数据治理目标

2.1 数据治理概述

数据治理(Data Governance)是指对数据的全生命周期进行管理,包含数据采集、清洗、转换等传统

数据集成和存储环节的工作,同时还包含数据资产目录、数据标准、质量、安全、数据开发、数据价值、数据服务与应用等工作,是对整个数据生命周期而开展的业务、技术和管理活动的集合。

数据治理关注数据质量,保障数据稳定性、准确性,合理控制数据的生命周期,降低成本。数据治理的目标是理清有哪些数据、从哪来到哪去,最终用到什么地方。

数据治理就是进行数据架构、数据标准、数据质量、数据安全等领域的建设和管理的全流程。本质是盘点数据资产、提高数据质量,实施数据全生命周期的管理^[1]。

构筑一套企业级的数据综合治理体系,确保关键数据资产有清晰的业务管理责任,IT建设有稳定的原则和依据,作业人员有规范的流程和指导;有裁决机构和升级处理机制去解决争议问题;治理过程有所需的人才、组织、预算做基础保障。

2.2 数据治理目标

结合包钢财务业务实际情况与未来规划,以数据的“采、存、管、用”为主线规划设计包钢数据治理体系建设、数据平台搭建、数据应用设计等内容。

数据治理是所有数据应用的基础和根基,数据治理的好坏直接影响数据应用过程的价值达成,同时数据治理也是数据资产能够得到有效沉淀的基础。数据治理体系建设的目标就是帮助包钢能够及时、高效地发现数据问题,提升数据质量。助力包钢财务域业务应用场景落地,以及包钢数据管理体系建立。

包钢数据治理的目标具体为以下8个方面:

(1)数据管理方面。完善数据管理的组织职能,确认各业务部门和职能部门的数据所有者。发布并按照数据管理流程开展体系建设。

(2)数据质量方面。建立事前、事中、事后的数据质量控制体系,通过获取数据监控规则、进行数据校验、异常数据分析及改进,达到数据质量的监控管理,提高数据的质量。

(3)数据分布方面。建立数据模型,梳理数据分布,发布数据集成和共享的管理机制,统一业务领域和应用系统元数据,盘点包钢数据资产。

(4)数据标准方面。建立业务术语、参考数据和主数据、数据源、指标数据的数据标准和规范、管理流程。

(5)数据资产方面。梳理数据架构形成数据资

产,制定管控机制。成立数据管理组织,推动和管理按照数据架构机制和规范保障数据共享和质量。

(6)数据应用方面。规划业务运营场景及指标,业务数据统一入数据湖,建设数据底座,充分挖掘数据价值,满足业务可视、分析所需、决策可依的数据应用消费需求。

(7)数据安全方面。建立统一的数据安全标准、策略、管理流程,明确利益相关者的职责。定期开展数据安全风险分析。建立数据安全审计流程。

(8)数据生命周期管理方面。建立数据需求收集、验证和汇总流程、数据设计和开发流程、数据提供方管理流程,数据运维方案和流程,以及数据退役标准、数据保留和销毁策略,纳入数据管理体系日常工作。

3 包钢数据平台数据仓库设计

3.1 数据仓库分层作用和意义

数据分层是将杂乱无章、多数据源的数据按照业务主题进行分类,并有序地进行清洗、过滤,实现数据标准化的过程。在整个过程中,公共逻辑下沉将直接使用,以空间换时间,为业务提供更高效、安全隔离的应用数据。

在实际的工作中,开发人员和业务人员希望数据能够有序地流转,清晰地知道数据的整个生命周期,但是往往面临的数据状况是复杂且层级混乱的,导致我们可能会做出一套结构混乱,且出现循环依赖的数据模型。科学合理的设计数据分层,并且搭配一套行之有效的数据组织、数据管理和处理方法,使得数据体系更加有序^[2]。数据分层的好处有以下几点:

(1)清晰数据结构。每一个数据分层都有它的作用领域和职责,在使用表的时候能够更方便地定位和理解问题。

(2)减少重复开发、规范数据分层。通过开发一些通用的中间层数据,减少重复计算。

(3)统一数据口径。通过数据分层提供统一的数据出口,统一对外输出的数据口径。

(4)复杂问题简单化。将一个复杂的任务分解成多个步骤来完成,每一层解决特定的问题。

(5)数据血缘追踪。当业务使用的数据表出现问题时,可快速准确的定位到问题表,并清楚每张表的作用范围。

(6)减少业务的影响、应对多变的业务场景。

无需根据业务的改变而重新接入数据及修改应用加工逻辑。

3.2 包钢数据仓库设计思路

包钢数据平台中数据仓库的设计需要考虑三方面内容,一是要结合包钢财务域数据治理及数据应用场景的特点(典型的计算型数据应用场景,即以数据仓库批处理、实时计算、交互分析为主);二是要考虑钢铁行业的数据特点是数据源头多,数据质量不齐,数据访问量大;三是要能满足复杂多变的数据应用场景建设,要能高效的开放数据服务。

结合以上内容,包钢在构建数据仓库过程中要尽可能地将公共模型与自有模型区分,做到模型与数据的解耦,便于高效的支撑数据服务应用。同时,统一数据口径、减少重复建设、实现清洗的数据结构,并减少对原业务系统的影响。

包钢数据平台的数据仓库划分为五层:ODS、DWD、DWM、DWS、ADS,满足未来丰富的数据应用建设。

(1)ODS 贴源层。贴源数据,跟源系统数据保持一致。

(2)DWD 数据明细层。对汇聚的数据进行数据标准化处理、数据清洗等操作。

(3)DWM 数据中间层。采用维度建模方法,设计通用原子性小宽表以适应业务频繁变化的场景。

(4)DWS 数据汇总层。基于数据中间层的数据,加工多维宽表、业务基础标签和业务基础宽表。

(5)ADS 数据应用层。负责生成宽表、报表和指标。

3.3 包钢数据仓库分层作用关系

根据包钢数据平台中数据仓库的分层逻辑,按照数据采集加工过程,其数据仓库五层作用关系如下。

3.3.1 贴源层 ODS 作用

数据接入:通过离线/实时抽取 ERP 系统数据到数据仓库,为上层数据加工提供原材料,还可以作为数据冗余备份,提供明细查询功能。

信息补充:在数据接入过程中,需要加入 ERP 数据接入时间、来源系统、主键信息、修改标记、删除标记等扩充信息。

大字段处理:交易表中有大对象的字段(如 clob, blob 等),如果数据仓库不使用的话,接入时会舍弃;如果后续要使用,则考虑使用对象存储单独处理,并与数据仓库表建立关联。

3.3.2 明细数据层 DWD 作用

主题域划分:逻辑实体、属性的划分及业务对象的识别。

数据建模:将业务过程相关的数据进行水平整合或垂直整合,输出最细原始粒度、完整属性的模型表。

一数之源:通过权威数据、数据优先级、一手数据、数据使用频率或使用人数等实现。

数据标准化:对原始层数据进行规范化编码转换、清洗、统一格式、脱敏等,保证数据质量。

3.3.3 数据中间层 DWM 作用

数据聚合:对标准化数据,以业务通用性原则生产中间表。

公共指标:对通用的核心维度进行聚合,计算出相应的统计指标,减少重复加工。

3.3.4 公共汇总层 DWS 作用

数据汇总:以分析的主题对象作为建模驱动,基于用户需求和钢种产品指标需求,提供面向主题、常用的、稳定的派生指标数据,以及轻度汇总预计算数据,按照一定的周期(比如天)、按照一定的维度进行聚合,采用一些维度退化手法,将维度退化至事实表中,采用更多的宽表化手段构建公共指标数据层,提高公共指标的复用性,减少重复加工。

3.3.5 应用层 ADS 作用

个性化指标:存放数据产品个性化的统计指标数据,直接为各种统计报表提供数据。

基于应用的数据组装:宽表集市、趋势指标、数据标签、算法等,不会全盘考虑整体的数据架构和应用,每个应用都有自己独立的空间。

4 包钢数据平台功能

以包钢数据平台财务域及未来数据汇集规划为前提,根据包钢数据实际需求选择相应的模块与组件进行数据平台的搭建。

包钢数据平台基于分布式架构,在包钢云服务器上进行部署且可以动态扩容。整个平台具备分布式存储引擎、分布式计算引擎、大数据消息中间件、数据处理引擎、DAG 计算引擎、作业调度系统、平台用户管控、集群管理监控系统、数据资产服务、透明加密、数据迁移的功能,以更好地支撑包钢大数据项目建设。

5 包钢数据平台应用设计

在数据应用视角下,信息化发展到数据驱动业

务价值达成的演变路径,历经数据产生与治理,再到数据业务使能。

数据平台投入使用后,借助平台解决线下收集数据难、数据不准、数据计算量大等问题,通过生产结构诊断应用场景的上线,节约线下超过 10 + Excel 文件处理,优化线下计算口径、计算逻辑不统一的问题,将数据处理过程复杂的标准成本核算通过数据平台完成,同时也解决数据计算过程耗时长等问题,生产结构调整预测效率有明显的提高。

数据应用在包钢财务域数据治理中,从五个方面发挥数据应用价值,一是以标准成本管理为基础,打通生产工序成本数据,构建端到端的生产管控链路,辅助优化生产结构诊断;二是以辅助经营预测为核心,打通关键业务,通过数据分析,赋能业务发展,辅助业务决策,有助于改善企业 ROE 综合指标;三是构建存货全链路检测,实现从库存采购、生产、销售全链路信息透明化、可视化、可追溯,支撑管理决策;四是建设综合财务监控平台,监控重点管理领域的风险,比如“两金”管理、资金监控、销售监控、库存监控和税务监控等,把控财务风险,洞察经营管理;五是构建客户画像,打造客户信息化标签,建立以客户为中心的统一视图,洞察客户交易,预判价格走势,指导精准营销。

6 结束语

数据治理为包钢财务领域的发展奠定了基础,可以弥补财务域在数据管理方面的不足。通过对包钢股份财务域的数据管理现状进行理解和分析,发现其中存在的问题,根据包钢股份财务域现状进行数据平台中的数据仓库设计,结合包钢数据源头多的特点,基于多元化的数据应用场景展开,将数据治理的效果体现在企业的多个生产经营场景中,为公司未来的进一步发展打下了坚实的数据基础。

参 考 文 献

- [1] 杨绍旭. 大数据及人工智能影响下的企业财务管理分析[J]. 中国物流与采购. 2022, (20): 90-91.
- [2] 梁婉容. 人工智能在企业财务管理中的应用及展望[J]. 国际商务财会. 2018, (6): 91-93.