

钢铁行业数据仓库系统应用与创新

韩勇¹ 葛海洋²

(1.首钢京唐公司信息计量部 河北 唐山 063000 ;

2.首钢京唐公司信息计量部 河北 唐山 063000)

摘要: 现在,企业信息化已经成为一股势不可挡地潮流,数据仓库工具在大型企业中发挥着越来越重要的作用,它们给企业带来管理优势、价值优势的同时,在实施、应用过程中也会不可避免地出现一些难点与问题,特别是对于钢铁行业这种工序复杂的生产企业,如何结合企业实际,对复杂的信息流做到有效地多维处理与分析,是每一个实施、应用数据仓库的企业都会遇到的问题。本文以首钢京唐公司业务数据仓库平台的搭建为背景,对钢铁行业实施数据仓库系统中可能遇到的挑战、应对及创新进行了分析。

关键词: 钢铁行业 数据仓库 应用 创新

Application and Innovation of Data Warehouse System in Steel Industry

Han Yong¹, Ge Haiyang²

(1. Department of Information Computation of Shougang Jingtang Iron and Steel Company, Hebei,

Tangshan, 063000; 2. Department of Information Computation of Shougang Jingtang Iron and Steel Company, Hebei, Tangshan, 063000)

Abstract: At present, IT application has become an inevitable trend of the time. Data warehouse is playing an increasingly important role in big companies. While bringing management advantages and advantages in value, they will inevitably encounter difficulties and problems in the process of implementation and application. In particular, as for production companies in steel industry with complex procedures, how to effectively deal with and analyze the complex information flow from multiple perspectives in combination of the reality of the company is a problem every company encounters. This paper analyzes the possible challenges, responses and innovation in implementing the data warehousing system against the backdrop of the building of the platform for business data warehouse in Shougang Jingtang Iron and Steel Company.

Key Words: Steel Industry, Data Warehouse, Application, Innovation

0 引言

随着近年来信息技术的日益成熟,ERP 系统已逐渐成为不少成功企业提高经营管理效益的关键事务处理系统。ERP 的成功实施和应用,在使企业流程规范化的同时,也使得企业管理系统的数据量呈指数性增长。这就需要一个有效的信息分析处理工具来对数据信息进行存储、集成、处理、分析从而实现信息价值,为企业增效节支。SAP BW 是 SAP 公司的 mySAP 商务智能(BI)组件。它可以为 SAP 数据和非 SAP 数据的采集、存储、分析和管理提供一个集成的,面向商务的平台。

钢铁企业生产过程复杂,生产过程包括冶炼工序、轧制工序、成材工序。生产过程的复杂性以及产品种类的多样性使得数据仓库系统的应用与其它行业相比面临更多的挑战。本文以首钢京唐公司业务数据仓库(BW)项目的实施案例,分析钢铁企业业务数据仓库平台搭建的过程中面临的难点及创新。

1 首钢京唐公司业务数据仓库平台项目简介

1. 项目时间

首钢京唐公司业务数据仓库平台项目的实施共经历两个阶段:

(1) 2009.7-2009.11 公司第一次实施 BW,经过需求调研、系统设计、开发实现、上线准备等阶段工作,于 2009 年 11 月 18 日正式上线。

(2) 2010.7-2010.10 随着公司一期二步建设工作的进行,新设备、新产线的投产,公司 BW 系统也进行了相应的扩充和完善,于 2010 年 10 月 31 日正式上线。

2. 项目成果

经过两个阶段的工作,公司搭建了完善的数据仓库平台,形成了数据分析体系,建立京唐公司统一的数据中心,与企业运营有关的各专业数据都集中管理。并按照各专业间交互需求及保密需要分层次满足数据共享。

专业管理人员可通过该系统完成各项查询、整理、汇总等数据处理功能,达到动态、多维、可追溯的信息分析。决策者可通过该系统查询企业的实时信息,为企业决策提供依据。

2 首钢京唐公司业务数据仓库平台搭建中的难点与创新

钢铁企业的生产过程工序繁多,流程复杂,业务流的复杂性,使对信息流、数据流的处理同样面临许多难点:由于钢铁企业信息化系统的多级架构模式,数据来源复杂多样,信息整合难度高;钢铁企业兼具离散生产与连续生产的特性,系统数据表结构的固定性与数据分析的多态性需求使数据分析更加困难;传统数据仓库的设计模式不能满足企业复杂业务需求的现实;由于信息需求具体多样,BW 提供的标准套件不能完全满足企业的定制化需求。针对这些难点,我们使用了许多在 BW 新版本中提供的新技术并在标准系统的基础上对模型或使用方法进行创新,下面将对上述几点分别论述。

1. 多提供者数据模型的应用

首钢京唐公司的信息化系统有四级架构，其中为 BW 提供数据的有四级 SAP ECC6 系统，三级 MES 系统，计量系统等，运行的操作系统平台既有 Windows 又有 Unix，数据库平台既有庞大复杂的 Oracle 也有相对简单的 MS-SQL Server，对于一部分调整数据，还需要从文件接口获取。多数据源多系统平台的集成数据分析，成为首钢京唐公司业务数据仓库平台生产模型搭建中的难点。

经过细致的调研与分析后，我们制定了如下技术方案：

(1)在数据源层对各系统分别建立基于 S-API（SAP BW 的标准数据源接口）、DB 连接、文件适配器的相同数据粒度的数据源。

(2)通过 DSO、CUBE 层进行数据处理。

(3)上层通过多提供者（生产模型）进行数据汇总。多信息提供者合并来自不同的信息提供者的数据。多信息提供者可以集成其他任何信息提供者的数据，包括信息立方体、数据存储对象二信息对象、信息集和汇总级别，它本身并没存储物理数据。

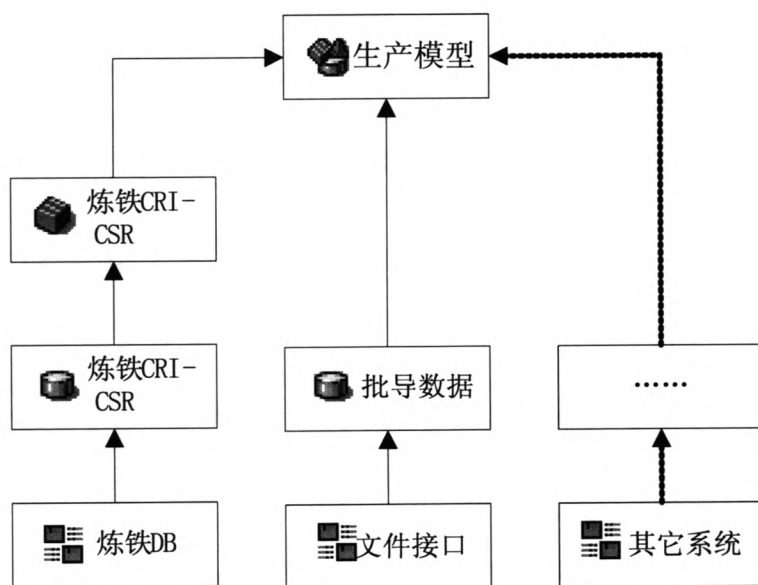


图 1 多提供者模型使用示意图

多提供者模型的使用，不但解决跨平台跨系统的数据集成问题，而且由于多提供者的数据查询在不同的底层模型间是并行进行的，所以同时大大地提高了查询效率，充分利用了系统资源，减少了查询等待时间。

2. 信息集模型的应用

在钢轧产品的数据分析需求中，需要对整个钢后产品链做综合分析。以热轧

为例，热轧产品统计人员不但需要知道热轧每个批次的产量、规格、钢种，而且需要追溯至投入板坯的投入量、钢种、订单等。冷轧产品的分析链则更长。而由于 ECC6 系统中物流数据存储结构为普通二维表形式，所有钢轧产品的物流数据都以表项目的形式记录在一个物料数据库中，包括报产、改挂、改判、转储等业务都无区别地记录在里面，而分析时需将业务分类整理。钢轧产品分

析的链特性与系统数据存储的集特性的不一致造成了数据分析的困难。

针对这些分析需求，我们决定使用信息集模型：

- (1)在物流模型上搭建上层 DSO 模型用来将业务按产出类和投入类进行分类。
- (2)中间层搭建 DSO 用来对数据进行处理，保证投入、产出的一一对应。
- (3)顶层用信息集模型通过钢卷号和物流记录时间实现投入、产出的对接。

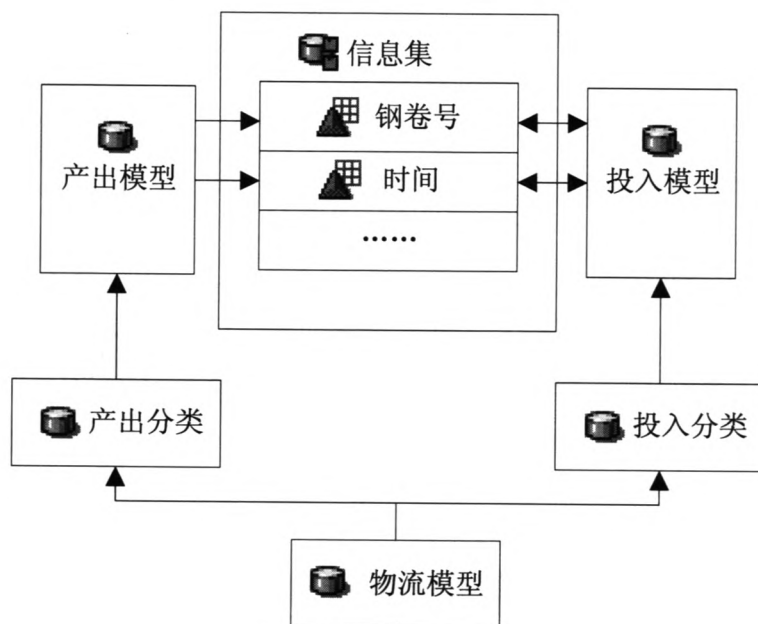


图2 信息集模型使用示意图

如果通过自开发方式实现同样的需求，需要大量的开发工作，操作不方便且容易出错，上线后的维护也相对困难。通过使用信息集，不但满足了钢后产品数据分析的需求，同时大大减少了开发工作量，缩短了项目周期，保证了项目进度，而且上线后的维护变更也简单易行。

3. 数据流与模型创新

首钢京唐公司的业务需求有着其特殊性，作为一个创立不久的企业，在生产工作日趋平稳的同时，基础建设工作也在紧张进行。由于财务分析的集成性需求，需要同时对工程项目控制数据和财务付款数据等进行综合分析。在 SAP ECC6 系统中，工程数据同付款数据一样是作为业务数据源存放的。而财务集成分析对工程数据的需求并没有工程管理部门那样精细，只需要各单项、单位工程的预算、投资控制目标等控制数据。

如果采用 SAP BW 的标准设计，就需要在系统额外建立工程数据的数据源，将工程数据全部存放在 BW 系统中。这样不但会造成系统存储的额外开销，而且由于冗余数据的影响，对财务集成分析的性能也会产生影响。

经过分析，我们并没有使用 SAP 的标准流程，而是对数据流向上进行了重新设计：

- (1)不使用 SAP BW 的标准工程数据源来进行集成查询。

- (2)创建一个独立的工程控制数据源，保存财务需要的工程数据。
- (3)将工程控制数据源内的控制数据导入工程的 WBS（工作分解结构）主数据中。
- (4)财务分析时直接从工程 WBS 主数据中读取工程控制数据。

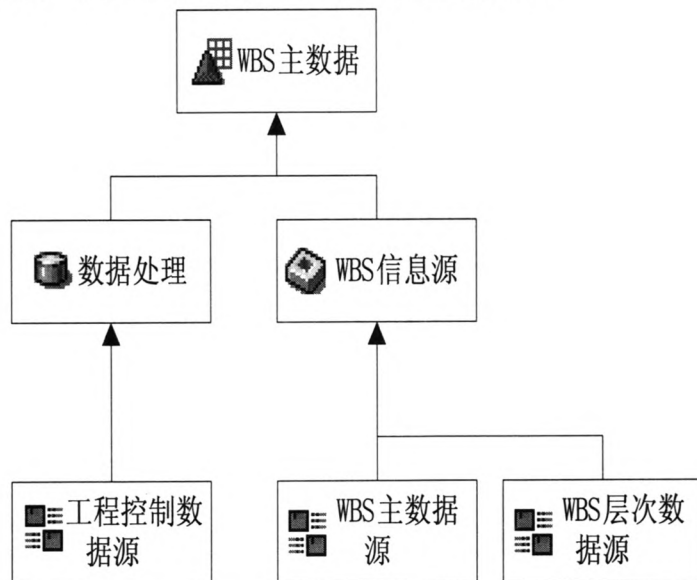


图 3 工程控制数据流

4. 定制与开发创新

大部分 ERP 软件公司提供的都是成形的标准套件，但是不同的企业的需求却是各不相同，需要对软件公司提供的标准套件进行本地化。对标准套件本地化的工作没有固定的模式，大多数本地化工作都是在行业经验的基础上结合公司实际业务需求进行的。在首钢京唐公司数据仓库平台搭建的过程中，我们以低开发周期和低维护成本为宗旨，结合行业经验同时不拘泥于经验，做到了可靠、创新的本地化。下面以物流数据源 2LIS_03_BF 和财务数据源 0FI_GL_4 为例说明。

作为物流数据的主要数据源，BW 提供的标准数据源 2LIS_03_BF 存储了几乎所有物流数据。首钢京唐公司 ECC6 系统中 2LIS_03_BF 数据源的特殊性在于，由于业务需要，在前期四级系统上线的过程中使用了一个一般项目中不会作为主要字段使用的字段 BKTXT 来存储钢后产品的钢卷号、坯号，且在后续业务调整中会对这个字段进行修改。但标准的 2LIS_03_BF 并不会自动记录 BKTXT 的更改，这就导致业务人员在对 BKTXT 字段变更后 BW 系统中的数据会与原始数据出现不一致。

如果按照行业经验来增强 2LIS_03_BF，需要修改 MCEX 的通讯结构，并做大量的判断更改代码来实现对数据变更的记录，为了保证项目周期，减少以后的维护工作的难度，我们使用了新的技术方案：

(1)对 2LIS_03_BF 使用的主要数据库表 MSEG 和 MKPF 建立数据库视图。

(2)创建基于视图的一般增量数据源 Z2LIS_03_BF，使用更改日期作为增量字段。

(3)Z2LIS_03_BF 的安全区间设为 2，即为了防止数据丢失，下一次数据加载与上一次数据加载间保持 2 天的数据重叠。

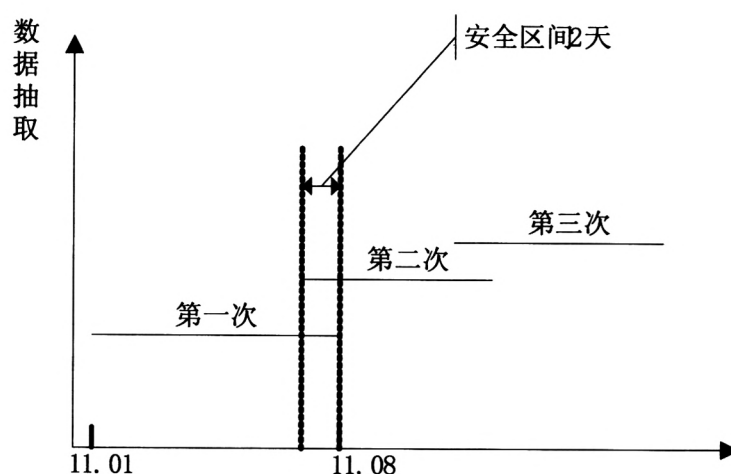


图 4 增强数据源增量抽取示意图

(4) 2LIS_03_BF 与 Z2LIS_03_BF 数据源在 DSO 层进行数据汇总。对于重复抽取的数据，在 DSO 层设置覆盖模式，保证不会有相同的数据在模型中保存两次。

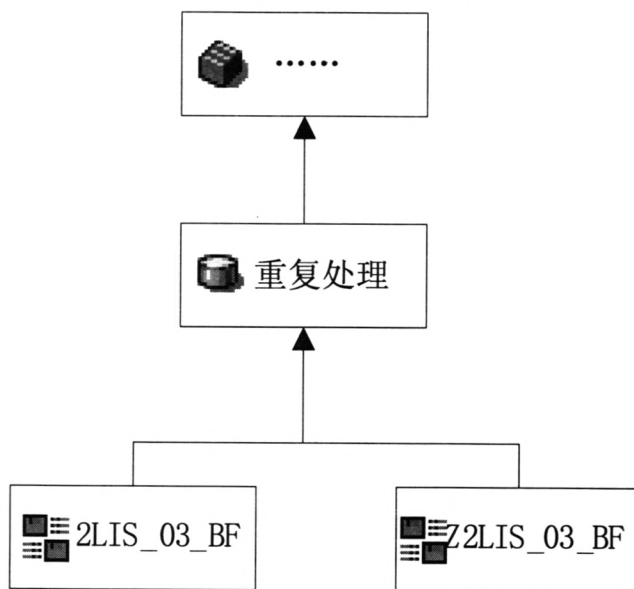


图 5 增强数据源数据流示意图

在 SAP ECC6 系统的财务部分架构中，财务数据分布在 6 个数据表里。在 SAP BW 中，用来存储财务行项目数据的数据源有 0FI_GL_4, 0FI_AP_4, 0FI_AR_4，分别对应总帐、应付、应收行项目。首钢京唐公司业务数据仓库项目中，财务部门需要对财务数据按进行多维度的收、付及一般总帐查询，基建财务部门需要按 WBS、网络等，生产财务部门需要明细到物料维度进行查询，由于复杂的查询维度需求，若分别使用三个数据源进行增强开发后汇总，不仅需要分别开发、测试增强代码而且数据的汇总、连接也是问题。最后我们确定了使用 0FI_GL_4 为基础，增强 AP、AR 数据及其它分析维度数据，所有的总帐、收付报表使用统一的数据基础的方案。

增强分析字段清单：

字段名	字段描述
ZAPLZL	订单的通用计数器
ZAUFPL	订单内工序的任务清单号
ZEBELN	采购凭证号
ZEBELP	采购凭证的项目编号
ZHZUON	特别总帐科目的分配号
ZKIDNO	付款参考
ZLIFNR	供应商或债权人的帐号
ZPOSID	工作分解结构元素 (WBS 元素)
ZPOST1	PS: 短描述 (第一行文本)
ZUONR	分配编号
ZWTGES	按业务货币计算的总计值
ZXRAGL	标识: 清帐已冲销
ZXREF2	业务伙伴参考码

财务数据流示意如下，可以看到，改进后的数据流程更加紧凑有效，既节省了系统资源，又能满足复杂多维度查询的需要。

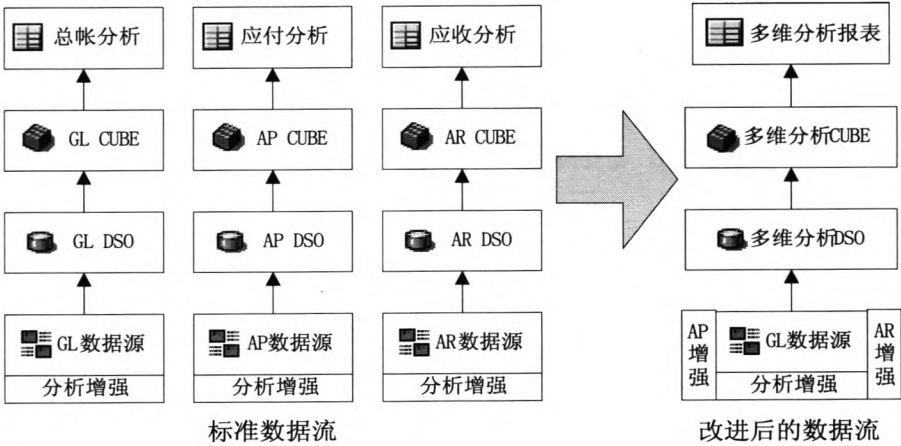
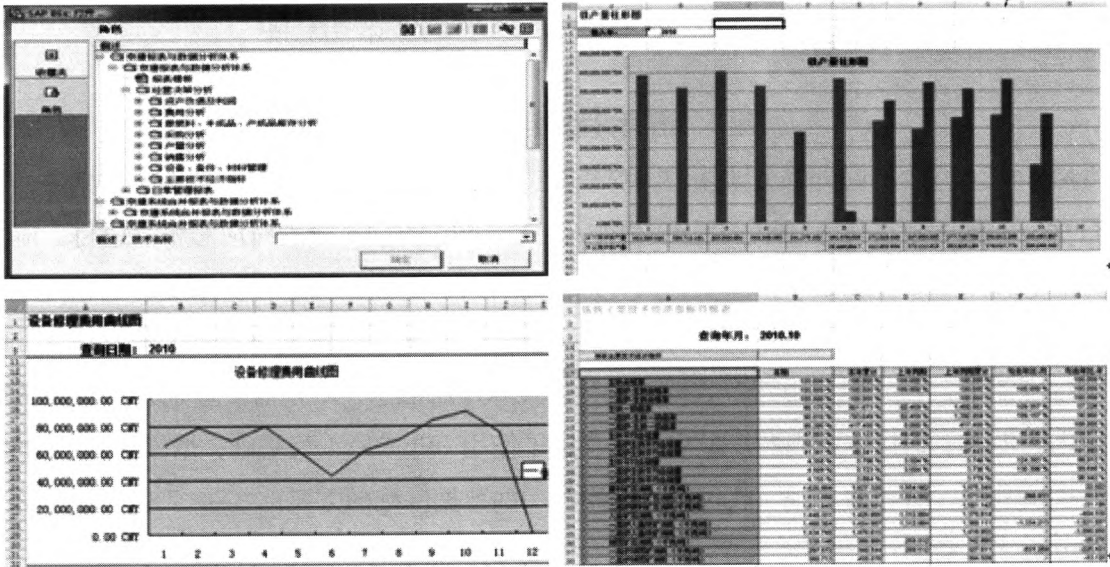


图 6 财务数据模型改进示意图

3 首钢京唐公司业务数据仓库平台应用示例



4 总结

信息化是现代企业发展的必然趋势，企业生产经营过程中积累的业务信息是企业的宝贵财富。而现代化的数据分析工具对企业的管理优化、节能增效、快速发展起着越来越重要的作用。但每一个信息化系统上线的过程都不会是一个一帆风顺的过程，特别是对于钢铁行业，业务流程、生产工艺的复杂性注定了信息化系统上线的过程中需要更多的探索、更多的创新。应该看到，数据仓库系统在大型钢铁企业中的应用尚处在摸索阶段，本文中论述到的内容相对于庞大复杂的数据仓库系统也只是冰山一角，谨希望能够抛砖引玉，使更多的钢铁企业在应用数据仓库系统的过程中找到一条适合企业自身实际的创新之路，为行业数据仓库系统的应用提供更多宝贵经验。

【参考文献】

- [1] (美) 萨蒙 (Inmon, W.H) 著, 王志海等译. 数据仓库 (原书第 4 版) [M]. 北京: 机械工业出版社, 2006.
- [2] (加) 韩家炜, 堪博著, 范明, 孟小峰译. 数据挖掘概念与技术 (原书第 2 版) [M]. 北京: 机械工业出版社, 2007.
- [3] 潘华, 项同德. 数据仓库与数据挖掘原理、工具、及应用 [M]. 北京: 中国电力出版社, 2007.
- [4] 陈京民. 数据仓库原理、设计与应用 [M]. 北京: 水利水电出版社, 2004.
- [5] 王丽珍. 数据仓库与数据挖掘原理及应用 (第二版) [M]. 北京: 科学出版社, 2009.
- [6] 姚家奕. 数据仓库与数据挖掘技术原理及应用 [M]. 北京: 电子工业出版社, 2009.
- [7] 夏火松. 数据仓库与数据挖掘技术 (第二版) [M]. 北京: 科学出版社, 2009.