

文章编号: 1000-5862(2018)06-0634-05

# 基于 DAMT 的企业级数据仓库建设关键路径研究

赵宏斌<sup>1</sup>, 白开峰<sup>2</sup>, 崔丙锋<sup>3\*</sup>, 韩丽芳<sup>4</sup>, 马 勇<sup>5</sup>

(1. 国网陕西省电力公司 陕西 西安 710032; 2. 国网陕西省电力公司西安供电公司 陕西 西安 710032;

3. 国家电网公司 北京 100031; 4. 中国电力科学研究院有限公司 北京 100089;

5. 江西师范大学计算机信息工程学院 江西 南昌 330022)

**摘要:** 针对企业级数据仓库的关键问题, 该文基于 DAMT 构建了一种关键路径。这主要包括 4 个模块: 决策支持设计、数据来源分析、数据仓库建模、技术实现。利用所给的关键路径模式能够使企业管理决策者快速掌握信息, 从而制定更加适宜的竞争战略, 增强自身的竞争力。

**关键词:** 决策支持; 数据仓库建设; DAMT 路径

中图分类号: TP 319 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2018.06.15

## 0 引言

20 世纪 70 年代, 企业将业务处理系统和分析处理单独运行, 逐步形成了数据抽取、转换、验证和加载等基础流程。1991 年 Bill Inmon<sup>[1]</sup> 指出数据仓库应该具有面向主题、集成、不可更新、包含最明细的业务数据, 通过该基础流程生成面向全企业的决策支持的“分析处理数据”。企业信息化建设的初期, 各个业务纷纷单独建立自己的信息系统, 形成了多个信息孤岛。随着企业经营的要求, 大型企业于 21 世纪纷纷实施了 ERP 系统, 构建了面向企业全业务流程, 对数据交换、集成和融合提出了新的要求。业务发展和管理需求与相对落后的数据存储与分析技术构成了企业信息化的主要矛盾。尤其是大数据技术的突破, 更加亟需建立科学高效的构建数据仓库。

国内外学者在该领域积累了丰富的实证研究。R. Kimbal 等<sup>[2-3]</sup> 对 DW 的设计、实现、运行、维护和管理等方面的问题进行深入探讨, 提出了数据仓库业务维度生命周期开发方法和数据仓库的总线结构, 目前已被广泛应用到制造业、零售业、医疗保健、金融业和政府决策等领域, 取得了良好的经济和社会效益。在国内, 借助产学研平台, 各大知名高校(清华大学<sup>[4-6]</sup>、北京大学<sup>[7-8]</sup>、复旦大学<sup>[9-10]</sup>、浙江大学<sup>[11-12]</sup> 等) 竞相开展数据仓库的理论探索研究, 开

发出了相应的工具平台, 取得了较满意的效果。

以上研究较好地解决了数据仓库技术层面实现, 但均比较繁琐。本文提出构建企业级数据仓库建设的关键路径: DAMT。通过简明扼要、通俗易懂的语言让企业管理层从宏观上快速了解数据仓库建设的方法。希望越来越多的企业, 借助数据仓库制定出更加适宜的竞争战略, 赢得竞争优势。

## 1 企业级数据仓库建设关键路径

企业级数据仓库建设的关键路径: DAMT, 即决策支持设计( Design)、数据来源分析( Approach)、数据仓库建模( Modeling)、技术实现( Technic) 4 大步骤, 如图 1 所示。

### 1.1 决策支持设计( Design)

企业首先要明确决策支持是构建数据仓库的真正目的。信息化规范了企业经营, 提升了运转效率, 对决策层和管理层来说充分挖掘数据资产价值更为重要。然而, 决策支持不能囿于现有信息系统能够提供的数据, 而是要站在企业经营全局的高度“设计信息”。决策信息需要从企业管控模式、价值链、企业生命周期等方面综合考虑设计。

1.1.1 集团管控模式 自 20 世纪以来, 国内经济的高速发展和经济全球化趋势, 赋予了中国企业前所未有的成长环境, 孕育了一大批大型企业及企业

收稿日期: 2018-06-17

基金项目: 国家自然科学基金( 61462045) 资助项目。

通信作者: 崔丙锋( 1976-), 男, 安徽濉溪人, 博士, 主要从事能源互联网的 ICT 关键技术研究。E-mail: bfcui@sgcc.cn

集团. 企业集团开始作为一种组织形式逐步发展壮大, 在激烈的市场竞争中其优越性日益突出. 在企业经营决策中, 决策层和管理层首先需要明晰企业的管控模式, 方能抓住重点, 从而设计出具有针对性的决策信息. 按照企业集团业务特征, 管控模式主要分为金融控股型<sup>[13]</sup>、战略管理型<sup>[14]</sup>和业务管理型<sup>[15]</sup> 3 种模式, 如图 2 所示.

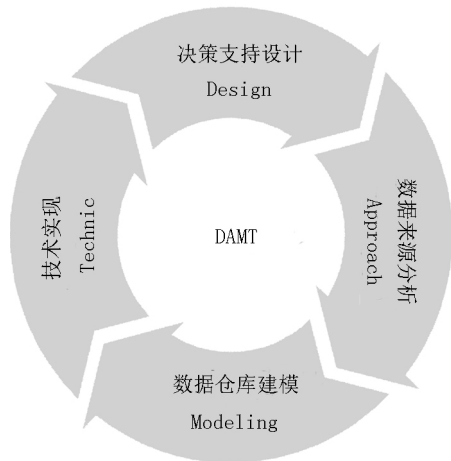


图 1 企业级数据仓库建设关键路径: DAMT

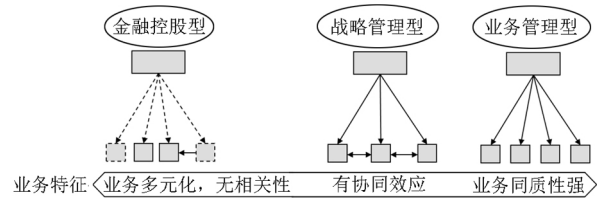


图 2 企业集团 3 种主要管控模式

金融控股型企业集团下属业务为多元不相关, 集团层面主要以财务投资为主营业务, 因此该类企业集团的决策层在设计决策信息时, 需重点关注投资回报类指标, 通过投资业务组合结构的优化, 追求公司价值最大化. 战略管理型企业集团下属业务为多元相关, 具有协同效应. 集团层面培育战略协同以保障公司整体有机成长. 因此, 该类企业集团既要关注财务指标, 又要关注各业务板块核心业务以及板块间协同指标. 业务管理型企业集团业务同质性强, 通常为单一产业领域内的运作, 往往采取集中控制与管理. 该类企业集团作为决策、管理、制度、标准输入平台, 管控层级深入下属企业的日常经营中, 以期提升标准化、规模化水平. 在决策信息设计时, 可参考价值链分析.

1.1.2 价值链 在某单一业务领域中, 为便于从宏观上更加清晰把握企业经营情况, 决策层和管理层可以从价值链的角度开展决策信息设计. 企业价值创造的基本活动包括购买原材料、产品生产、发货、商品销售和售后服务, 支持性活动包括人力资源、研

发设计、物资采购、财务管理等( 如果企业提供的为无形产品, 也可据此开展设计). 无论有形还是无形产品, 往往可以通过国内外行业对标方式明确指标体系, 并通过监控各分子公司及各部门执行情况, 来提升管理水平, 达到提升规模, 降低成本, 提高市场占有率, 提升与竞争对手的竞争力. 当然, 各部分的指标权重并不是一成不变, 而是随着企业的经营周期动态调整, 如图 3 所示.



图 3 价值链模型

1.1.3 企业生命周期 企业需要清楚不同生命周期的工作重点, 整合有限的资源, 最大限度地迎合客户的价值主张, 来最大化价值. 因此, 在不同阶段, 需由不同的职能来支撑公司发展, 这样才能延长企业生命线. 因此决策信息设计要根据企业所处阶段有所侧重, 如图 4 所示.

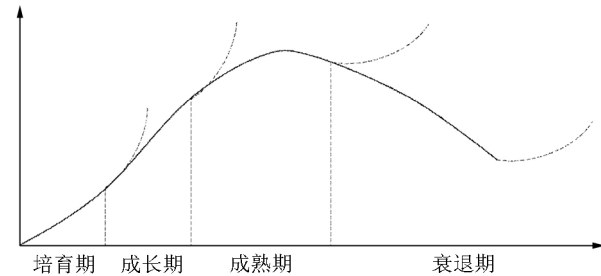


图 4 企业生命周期模型

处在培育期的企业, 需要验证新产品或服务、商业模式能否被市场接受, 开发职能主要关注新产品或新模式的设计、发现可能遇到的障碍并提出解决方案, 测试未来量产的可行性; 质量职能需要保障新产品的功能、品质、稳定, 为上市做好准备; 新产品具备条件后, 销售职能需要启动市场, 迅速进入销售状态. 因此, 在培育期决策信息设计时, 需要重点关注研发、质量、销售 3 个方面.

处在成长期的企业, 需要快速打开市场, 这就要求销售职能快速组建队伍、搭建销售体系和管控系统以全力推进销售; 生产职能需要快速增加产能, 控制生产成本, 支持销售部门快速打开市场; 运营体系需要培养、研究老客户需求, 开发未来可能的运营模式、流程、标准, 为规模扩张打基础. 因此, 在成长期决策信息设计时, 需要重点关注销售、生产、运营 3 个方面.

处在成熟期的企业, 需要尽快转型到“运营”阶

段,形成竞争力.客户服务职能需要服务好大量老客户,持续关注并满足他们的需求;营销职能需要了解并预测市场和客户的新需求,保障企业对市场的灵敏感知;研发中心需要根据客户的新需求对产品及服务进行升级.因此在成熟期决策信息设计时,需要重点关注客户服务、市场营销、研发中心 3 个方面.

处在衰退期的企业,需要与客户建立牢固的关系,升级业务模式,确保业绩稳定.研发中心需要深入研究市场趋势,为客户创造新的消费体验;营销职能需要重构原有的商业关系,将客户或者供应商变为合作伙伴获得更多的机会;考虑建设研究院将企业沉淀多年的行业经验形成总结报告吸引更多的专业人士加盟以发掘更多的创意思路,为下一个培育期做好准备;因此在成熟期决策信息设计时,需要重点关注技术研发、市场营销、研究院 3 个方面.

以上是企业在设计决策信息时,根据不同生命周期需要重点考虑的指标,但是,财务、人力、物力等职能均要统筹设计,毕竟这些职能是企业经营的基础.另外,关于各职能指标如何设计(如财务的杜邦分析),鉴于理论界及实践界已开展多年研究,该部分理论相对成熟,本文不再论述.

### 1.2 数据来源分析( Approach)

在上一阶段,决策层和管理层依据企业管控模式、价值链、企业生命周期制定出供决策的主题和指标体系,接下来就需要开展指标数据来源分析.

对于尚未开展信息化建设的企业来说,需要将决策信息设计的结果作为信息系统建设的重要输入,以确保信息系统上线后能输出所需数据.由于受

到实施周期的限制,信息系统往往不能一次性涵盖所有的决策指标项,因此需要企业将指标进行优先级排序,依次纳入信息系统实施范围,同时在系统建设过程中要预留足够多的字段和存储空间,支持后续扩展.

对于正在开展信息化建设的企业来说,需要以决策信息设计为标准,分析将要上线的信息系统是否涵盖所有指标项,若不涵盖,则需及时增补.在此过程中,需要明确指标所有者,并制定数据维护标准和制度.若信息系统此时已经完成最终测试准备上线,则可待系统上线后再逐步将新增的指标项植入其中,以避免对业务产生过大影响.

对于信息化成熟的企业来说,需要梳理现有系统的数据是否涵盖决策支持指标.对于缺少的数据项,首先要明确应在哪个系统中实现,其次明确维护数据的岗位职责.对于质量不合格的数据,则需要开展数据治理保证合规性.

### 1.3 数据仓库建模( Modeling)

经过上述 2 个步骤,企业已明晰决策支持主题与指标,并明确了数据的来源.接下来需要将数据按照一定的逻辑转换成指标并展现出来,这就需要开展数据仓库建模.企业级数据仓库主要分为抽取层、转换层、分析层、展现层,如图 5 所示.

1.3.1 抽取层 抽取层的数据来自于数据源层,数据源可以是企业内部的数据,也可以是外部公开渠道获取的数据;可以是信息系统中的数据,亦可以通过 Excel、TXT 等媒介导入的数据.为保证数据的真实完整性,便于后期维护,原则上抽取层的数据颗

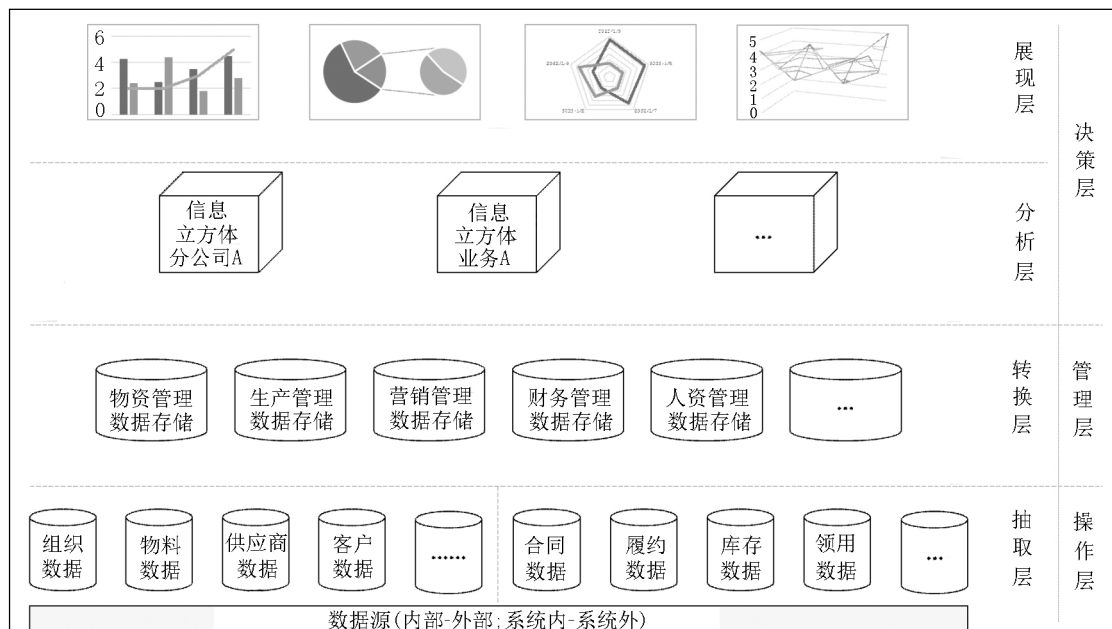


图 5 企业级数据仓库模型

粒度需要同数据源保持一致。抽取层的数据不进行编码转换、语义逻辑加工、数据清洗。抽取层的数据分为2类:主数据和业务数据。主数据主要是各种重要的分析特性的特征值和文本,如组织、物料、供应商、客户等。业务数据主要是各种业务单据,如采购订单、合同履行、库存记录等。

1.3.2 转换层 基于抽取层的数据进行一定的组织、关联、合并后,在业务转换层通常会实现基于一类主题的数据聚合,其数据结构和定义更贴近业务概念和业务需求。如按照企业管理职能(物资管理、生产管理、营销管理、财务管理等)形成相应的数据存储。基于转换层的数据,管理层即可获得大部分决策信息,这部分数据也可作为企业集团层面垂直管理的决策依据。值得强调的是,在形成转换层数据之前,需要对抽取层的数据完成编码统一、业务逻辑统一,只有基于此,才能保证数据统计口径在企业集团内上下一致。

1.3.3 分析层 如果说转换层的数据是弹药库中的子弹,那么分析层的数据就是战场上枪膛中的子弹。分析层的数据直接瞄准决策信息设计阶段的结果,需要按照决策设计对分析层的数据颗粒度展开设计(如可以按照某个分公司或某条业务线的角度组织转换层的数据)。因为数据粒度将严重影响数据存储的记录条数,记录数越多,付出的数据下载时间越长,查询性能越低,维护的成本也越高。考虑到当今的商业环境变化越来越快,企业需要据此灵活调整重点关注的决策指标,因此除必要处理外,从转换层到分析层禁止数据逻辑转换,否则将大大降低数据获取、信息提炼的时效性和准确性。

1.3.4 展现层 展现层是决策支持信息的最终体现,针对企业、业务线、职能部门等维度创建的分析层,其目的是为最终用户提供需要展现的信息。从格式上来说,分为格式化报表和自定义报表2类。格式化报表主要应用于统计性的日报、月报、季报和年报等需要采用固定格式并易于打印的报表。而自定义报表,是基于信息立方体的特性,开展多维分析、切片剖析、业务追溯等灵活查询(如查询华北地区、各业务线、各分子公司的销售情况)。从形式上来说,可以基于企业 Portal、IE 浏览器、Microsoft Excel,以及企业仓库专业软件呈现数据。值得一提的是,展现层也可以架构在转换层之上,为管理层提供决策支持。

综上,抽取层是面向企业的操作层,数据仓库是否可以发挥应有的价值,较大程度上取决于抽取层,即数据源的数据质量和全面性。转换层更多是面向管理层,便于其掌握职能领域的运作情况。分析层和展现层更多是面向决策层,使得企业高层既能从宏观上掌握企业集团的概况,又能透视到微观层面了解业务运转情况,为审慎合理地制定战略提供可能。

## 1.4 技术实现(Technic)

经过多年的实践,市场上涌现出一批优秀的产品,如 SAP BI、Oracle-Oracle Database/Oracle Exadata、IBM-Red Brick、Sybase-IQ、Datastage、Sagent、Informatica 等。就其本质来说,都是围绕数据仓库主要技术:数据源开发、ETL、权限控制、展现设计、数据上载监控(处理链)、数据备份恢复、数据仓库加速(如 SAP 的 BIA)。鉴于数据仓库技术发展已较为成熟,本文不展开介绍。企业可以统筹考虑费效比、兼容性、安全性、可扩展性,选择适合自己的数据仓库产品。

## 2 结束语

本文提出构建企业级数据仓库的关键路径:DAMT。通过简明扼要、通俗易懂的语言使企业管理层从宏观上快速了解数据仓库建设的方法。希望越来越多的企业,能够充分发挥数据仓库的优势,更加了解企业经营状况,从而制定适宜的竞争战略,赢得竞争优势。下一步将研究国内外大型企业数据仓库应用实例,按照行业、企业集团管控模式、发展阶段,提出完备的决策指标体系。

## 3 参考文献

- [1] Willian H Inmon. Building the data warehouse [M]. 2nd ed. New York: John Wiley & Sons, Inc, 1996.
- [2] Kimball R, Laura Reeves, Margy Ross, et al. The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouse [M]. New York: John Wiley & Sons, Inc, 1998.
- [3] Kimball R, Margy Ross. The data warehouse lifecycle toolkit: the complete guide to dimensional modeling [M]. 2nd ed. New York: John Wiley & Sons, Inc, 2002.
- [4] 胡侃, 夏绍玮. 基于大型数据仓库的数据采掘: 研究综述 [J]. 软件学报, 1998, 9(1): 53-63.

- [5] 韩览山,邵贝恩.面向集团型企业的混合型分布式数据仓库构建[J].计算机集成制造系统,2003,9(1):80-84.
- [6] 蒋旭东,冯建华,周立柱.联机分析查询处理中的一种聚集算法[J].软件学报,2002,13(1):65-70.
- [7] 裴健,唐世渭,杨冬青,等.联机分析处理数据立方体代数[J].软件学报,1999,10(6):561-569.
- [8] 林子雨,杨冬青,宋国杰,等.实时主动数据仓库中多维数据实视图的选择[J].软件学报,2008,19(2):301-313.
- [9] 张旭峰.ETL若干关键技术研究[D].上海:复旦大学,2006.
- [10] 朱建秋,蔡伟杰,朱扬勇.CIAS:工一个客户智能分析数据挖掘平台[J].小型微型计算机系统,2003,24(12):2255-2259.
- [11] 陈义,宋执环,李平.基于Web的流程企业数据仓库体系研究[J].计算机集成制造系统,2003,9(6):493-495.
- [12] 曹蓟光.联机分析挖掘处理技术(OLAM)的研究[D].杭州:浙江大学,2001.
- [13] 殷红军.基于企业集团的金融产业发展研究[D].北京:中国社会科学院研究生院,2014.
- [14] 武常岐.中国战略管理学研究的发展述评[J].南开管理评论,2010,13(6):25-40.
- [15] 王秀华.利益相关者企业价值管理研究[D].青岛:中国海洋大学,2012.
- [16] 孙国强,韩强飞,陈俊. Kettle 在企业数据仓库建设中的应用与研究[J].信息系统工程,2017(2):28-28.
- [17] 邱菊,王岩,黄佩卓.大型电力企业基于GBase分布式数据仓库建设初探[J].计算机应用与软件,2018,35(5):184-189.

## The Study on the Key Path of Enterprise Data Warehouse Construction Based on DAMT

ZHAO Hongbin<sup>1</sup>, BAI Kaifeng<sup>2</sup>, CUI Bingfeng<sup>3\*</sup>, HAN Lifang<sup>4</sup>, MA Yong<sup>5</sup>

(1. State Grid Shanxi Electrical Company, Xi'an Shanxi 710032, China;

2. Xi'an Branch, State Grid Shanxi Electrical Company, Xi'an Shanxi 710032, China;

3. State Grid Cooperation, Beijing 100031, China;

4. China Electric Power Research Institute Company Ltd., Beijing 100089, China;

5. College of Computer and Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

**Abstract:** One critical path based on DAMT is built for key issues of enterprise-level data warehouses. The path consists of four modules that are decision-support design( Design) ,data source analysis( Approach) ,data warehouse modeling( Modeling) ,technology realization( Technic) . Enterprise policymaker by the path can quickly grasp information in order to develop more appropriate competition strategies and enhance its competition.

**Key words:** decision support; construction of data warehouse; DAMT path

(责任编辑:冉小晓)