

商业智能系统建设若干问题的探讨

Research on Some Key Issues About The Implement of BI System

许建¹ 罗永强²

¹(清远职业技术学院网络信息中心 广东 清远 511510)

²(Sybase(软件) 中国有限公司广州分公司 广东 广州 510620)

摘 要 : 尽管商业智能到目前来说尚没有一个统一而明确的定义 , 但商业智能却在一些业务运营系统相对完善而成熟的企业得到了有效的应用。商业智能系统的建设是涉猎面广泛而又循序渐进的过程 , 涉及到数据仓库、数据挖掘、联机分析处理和面向应用的决策支持等多个领域与技术。结合作者多年从事企业商业智能建设的经验 , 本文在简要介绍企业商业智能与企业级数据仓库及其相互关系的基础上 , 重点对企业商业智能应用系统建设中若干关键问题进行了探讨。

关键词 : 商业智能 数据仓库 数据挖掘 联机分析处理 决策支持

随着市场竞争的加剧和计算机应用的普及 , 企业对计算机应用的需求也日益扩展 , 他们已不再满足于计算机能帮助他们迅速地处理具体业务 , 而要求计算机能帮助他们从积累的大量业务数据中探索业务活动的规律性 , 市场的运作趋势 , 并从中为他们参与市场竞争做出重要的决策。概括地说 , 即要求计算机系统为他们实现商务智能 (Business Intelligence)。完备的商务智能应包含五个层次 , 由低到高依次是 :

- ①查询 : 对企业范围数据的各种形式的访问 ;
- ②报表 : 查询结果的格式化展现 ;
- ③在线分析 (OLAP) : 多视角、多层次的数据分析 ;
- ④知识挖掘 : 探索、揭示企业数据之间的内在联系和规律 ;

⑤预测 : 对企业未来发展趋势的推测、模拟和仿真。

这五个层次互相关联 , 每一层次都满足企业的部分需求。查询可使企业了解自己的过去和现在 ; OLAP 分析本质上是一种复杂的查询 , 它所提供的多维视图使企业对自己的了解更全面、更深入 ; 查询结果和 OLAP 视图都可以以报表的形式展现和保存下来 , 供以后或上级部门使用 ; 知识挖掘运用各种高级分析方法和模型分析数据之间的内在关联 , 从而使企业准确把握自己的运作规律 ; 而预测则在知识挖掘的基础上推

测企业的未来状况 , 使企业可以根据预测提前规划未来 , 保证自己在竞争中的优势地位。

尽管商业智能到目前来说尚没有一个统一而明确的定义 , 但商业智能却在一些业务运营系统相对完善而成熟的企业得到了有效的应用 , 如电信、金融、税务等行业。从技术上来说 , 商业智能系统的建设涉及到数据仓库、数据挖掘、联机分析处理和面向应用的决策支持等多个领域与技术 ; 从应用与实现的角度来看 , 商业智能系统的建设涉及到企业资源的整合、业务流程重组、客户关系管理的优化等 , 其应用是一个逐步成熟的过程。结合作者多年从事企业商业智能建设的经验 , 本文在简要介绍企业商业智能与企业级数据仓库及其相互关系的基础上 , 重点对企业商业智能应用系统的建设中若干关键问题进行了探讨。

1 数据、信息与知识

数据 : 是关于事务属性的事实。数据与上下文无关 , 但可以处理成有意义的信息。

信息 : 关于信息的观点有多种 , 但对于企业这种组织形式来说 , 我们倾向于定义为信息是对事物的了解的不确定性的度量。

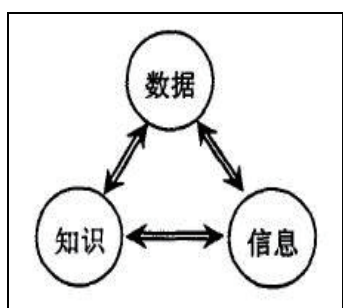


图1 数据、信息与知识的网状关系

知识：知识是信息之间的相互关系。在企业商业智能中，知识的概念通常是与数据和信息的概念相关联的，即知识是以多种方式把多个信息关联在一起的信息结构。

在这里我们之所以提起数据、信息与知识的基本概念，是因为在企业的 BI 解决方案中，基于数据、信息与知识的辅助决策支持是其核心的内容之一，而对知识和信息的理解和知识使用者的知识背景和思维方式有密切的关系，不仅有从数据中提取信息和知识的过程，还有外界的知识和信息作用到人，从而反过来作用数据的过程，因此，在企业的 BI 应用解决方案中，数据、信息与知识的关系也不仅仅是从数据 -> 信息 -> 知识的单链条关系，而是相互转化的网状关系，如图 1 所示。

2 商业智能

作为正式的概念，BI 是由美国 Gartner 公司的分析师 Howard Dresner 于 1989 年首次创造的这个名词，而早在 1985 年，美国宝洁公司就已经设计并开始利用类似商务智能的信息体系了。然而从概念到大规模的推广，还是有很长一段路要走。1996 年，Gartner 公司提出了信息民主的概念以后极大地推动了 BI 的应用，它认为商务智能是“使企业在竞争市场中保持领先地位的关键所在。正确的商务决策是以准确和及时的信息为基础的，而不是靠直觉。数据分析、报告及查询工具可帮助企业用户成功穿越数据海洋，并从中得到有价值的综合信息。”

自从关系数据库出现以后，OLTP 系统迅速发展，然而当人们初步设想从数据中提炼知识，却发现难以实现，而原因往往是找不到完整的数据，因为业务数据

存储在分散的、异构的数据库中。所以人们开始思考把数据格式统一并集中到一起的方法，这就是数据仓库（Data Warehouse）的由来。随着 DW 的海量数据的积累，就出现了数据挖掘（Data Mining）的概念。然而 DM 仅仅是产生了一堆统计报告，管理人员往往发现这些报告没有用，或不适用，无法对企业的经营决策带来帮助，于是数据库营销出现了。BI 开始与 CRM 相结合，对客户的行为进行有针对性的统计，并发现其中的规律，从而为企业的营销管理决策提供支持。为了更好地与企业的业务流程结合，报告展示层面的技术不断完善，并达到了可以定制和设计数据分析模式的高度，直接为企业提供上至战略规划，下至业务操作的支持。

综上所述，我们看到 BI 的产生和发展，正是数据分析与企业运营、业务流程紧密结合的过程，为企业创造持续的竞争优势。2004 年的 BI 峰会，确定了 BI 即是数据仓库之上的查询、报表和多维数据分析，正如本次峰会发言人所概括的，认为 BI 是一堆技术工具的集合。2007 年的 BI 峰会对 BI 的重定义：“我们开始将商业智能视为一个伞状的概念，它包括了分析应用、基础架构和平台和良好的实践”。可见，数据仓库、数据标准等平台已经涵盖在 BI 范畴里，BI 已不再是前端展现工具，BI 不再仅属于技术的范畴了。

3 企业级数据仓库

一个典型的、合理支持商业智能的企业数据仓库系统通常包含数据源、ETL、中央数据仓库与 ODS、数据集市以及数据访问与分析五个部分。

①数据源。通常包括企业内部信息和外部信息。内部信息包括存放于企业操作型数据库中（通常存放在 RDBMS 中）的各种业务数据和其他信息系统系统包含的各类数据及文档，及手工或半手工系统的数据及文档。外部信息包括各类法律法规、市场信息、竞争对手的信息以及各类外部统计数据及各类文档等。

②ETL。实际上，对企业级的数据仓库来说，ETL 在整个数据仓库系统的建设中起着十分重要的作用。我们可以 ETL 过程就是数据流动的过程，它包含有抽取、转换、装载三个环节，使得数据从不同的数据源流向不

同的目标数据,而这些数据源的数据可能是孤立的、不完整和不一致的,但到了目标数据区确是集成的、完整的、一致的。所以 ETL 事实上还是一个将企业数据源按企业信息模型进行数据集成、整合与映射的过程。

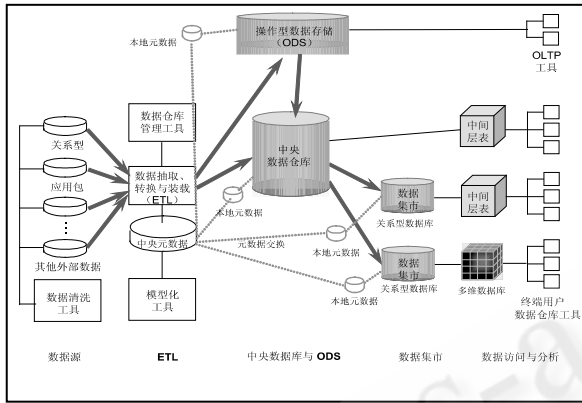


图 2 带 ODS 的企业数据仓库体系结构

③中央数据仓库与 ODS。在中央数据仓库与 ODS 中,存储的是抽取、清理,并有效集成后的,并按照主题进行重新组织的企业全局性的数据。同时也组织存储数据仓库元数据(具体包括数据仓库的数据字典、记录系统定义、数据转换规则、数据加载频率以及业务规则等信息)。不过中央数据仓库存储的以历史的以支持分析与决策的以历史为主的全局应用的数据集合、ODS 是能支持企业日常运营的以实时为主的全局应用的数据集合,是不同于 DB 的一种新的数据环境,是 DW 扩展后得到的一个混合形式。

④数据集市。那些小型的,面向部门或工作组级数据仓库就是数据集市(Data Mart)。在数据集市内数据的关联程度要高于集市与集市之间的关联程度。或者说,我们将某些关联程度较高的主题组织为主题域,一个或多个主题域也可以是数据集市。

⑤数据访问与分析平台。数据访问与分析平台主要包括前端工具与应用。它包含有包括各种数据分析工具、查询与展现工具、数据挖掘工具以及各种基于数据仓库或数据集市开发的应用。其中数据分析工具主要是指 OLAP 服务器,OLAP 的模式则又分为以关系数据库为基础的 ROLAP、以多维数据库为基础的 MOLAP 及前两种模式混合的 HOLAP。查询与展现工具、数据挖掘工具既针对数据仓库,同时也针对 OLAP 服务器。

4 基于数据仓库的商业智能

两个标志性的事件使得数据仓库与 BI 紧密联系在一起。1991 年,世界数据仓库之父 W. H. Inmon 出版了《建立数据仓库》一书;1993 年, E. F. Codd 等发表了一篇名为“为分析用户提供 OLAP(在线分析处理)”的论文,文章指出了多维概念模型的重要性,为企业的数据收集、管理、处理和表达提供一种多维的解决方案,以便进行分析和决策。应该说这些技术方案的成熟,为 BI 的真正有价值的实现奠定了坚实的基础。

对于基于数据仓库的商业智能(BI)来说,BI 可以理解为一种能力或方法:一种通过将企业中现有的运营数据(或政府的管理数据)转化为知识,进而帮助企业做出明智的经营决策的能力(或方法)。具体的说,就是通过对企业数据的收集、加工、管理(ETL)形成数据仓库(DW);再经由各式各样的查询分析工具(Query/Report Tools)、联机分析处理(OLAP)工具、数据挖掘(Data Mining)工具和决策规划人员的行业知识(Industry Knowledge);从而把复杂的业务数据转化为容易理解的信息,并将正确的信息在正确的时间内以正确的方式分配给正确的人,目的是使企业的各级决策者获得知识或者洞察力,促使他们更快地做出对企业有利的决策;帮助企业提高获利、增加生产力与竞争力。图 3. 展示的是基于企业数据仓库的闭环联合性 BI 的体系结构,所谓是指 BI 系统的分析与辅助决策建议反过来作用企业的 OLTP,包括人力资源、财务、客户关系管理及其他业务系统。闭环的 BI 是 BI 应用到一定阶段的必然需要。

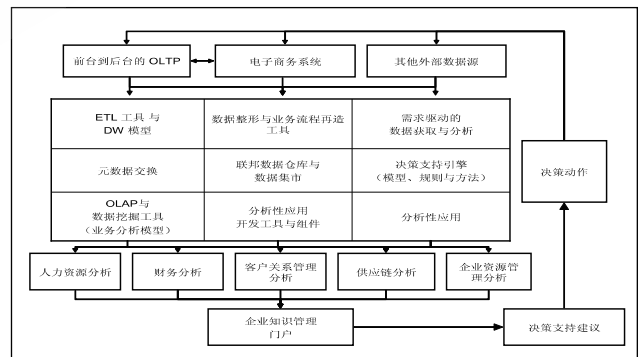


图 3 闭环的联合型 BI 体系结构

5 建设基于数据仓库的 BI 系统的若干关键问题

表 1. 简要概括了依据我们的经验,在基于企业数据仓库的 BI 系统建设过程中,各建设阶段或环节上的值得关注的键问题。事实上,如果考虑到 BI 闭环问题及实施全过程的项目管理,要关注的问题还远远不尽如此。以下,我们也仅就本表列出键问题的一些重点的键环节做进一步深入的探讨。

5.1 关于 ETL

ETL 是基于数据仓库的 BI 系统实施过程中的首要的重点也是难点问题之一。事实上,ETL 的实现与调度是一个极其复杂的过程,绝不是单一环节的数据映射,从空间跨度上,它涉及到从数据源到数据集市的整个环节;从时间跨度上,包含了数据仓库初始加载到以后

长期的数据更新。图 4. 简要展示了 ETL 过程及其数据流。

在我们的经验中,ETL 应该占整过 BI 应用系统建设 30% 以上的工作量,也是多数企业级数据仓库应用系统建设失败的常见环节之一。我们认为在 ETL 过程中,抽取效率问题,支持异种数据源问题,数据的抽取策略问题,数据的有效性、一致性、错误校验与回退等问题都非常关键。

更重要的是:对于 ETL,即使选择了良好的工具平台,二次开发或后续开发的工作量还不小;并且,抽取接口的后续维护的长期性是不可避免的。另外,灵活的数据接口设计与 ETL 全过程及有关数据仓库更新的后续过程的调度设计尤为重要。

表 1 BI 系统建设各阶段的键问题

系统建设的各阶段	键问题
系统运行与维护	业务流程重组、实用性、响应性能、需求升级
系统试运行	
系统修改、测试	数据仓库调优、展现的业务化
BI 门户构架	BI 展现及其业务逻辑、统一门户
功能与数据稽核	数据的一致性
数据仓库数据加载	ETL 性能
系统平台安装调试	系统集成方案、可扩展性、可靠性与稳定性
原型系统开发、测试	安全、效率、核心平台的生成
系统设计	数据仓库模型、与数据源的接口、系统的可扩展性,数据仓库粒度、对需求的修正
项目启动、需求分析	引导用户的需求、需求的合理性、对业务的分析与理解、对数据源的分析与理解

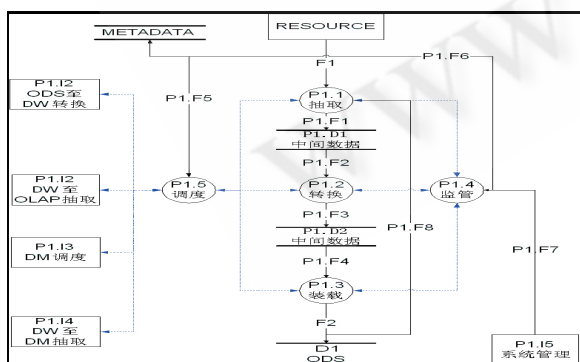


图 4 ETL 过程与数据流

5.2 主题及其扩展

当然,数据仓库是面向主题的,而主题又是来源于

业务需求的。虽然在数据仓库应用初期可能实现的分析主题域及其包含的分析主题只是有限的,但随着数据量的增加、应用经验的积累、应用需求的拓展,新增分析主题是不可避免的和长期的,这就要求应用系统具有这种可扩展的能力,而如果必须完全依赖于应用开发商来实现这种扩展,对用户来说将是一种沉重的负担。一般来说,这种新需求的扩展 50~80% 应该能让用户(高级数据分析师+数据仓库管理员)通过使用应用系统的模型维护、定制等应用功能来完成。也就是说,从数据仓库建设一开始,就应该考虑企业整体的数据模型,应用系统本身也是可扩展的。如果要对系统的整体设计做大幅度的修改才能实现如上所述的需求扩展,这样的系统就不能算是成功的。

5.3 关于 BI 的用户层次

企业 BI 的应用应该是有用户层次的,并不是所有的用户都必须和都能够使用所有的功能的。BI 系统的使用角色与用户层次如图 5. 所示。

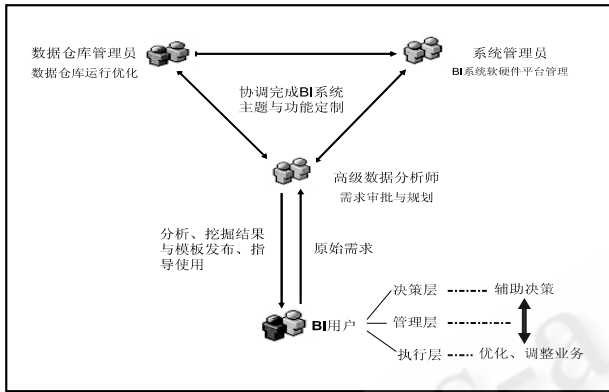


图 5 BI 系统的角色与用户层次

值得重点关注的是, BI 的应用功能从业务层面上应该面向决策层、管理层、业务操作层三个不同层面,不同层面的使用需求也是不一样的,事实上,我们不能要求所有的领导和所有的一般业务人员都能完成类似数据挖掘等这样复杂的操作,他们要的一般只是结果。这种结果,领导用它来辅助决策;一般业务人员参照它来调整、优化业务。大部分管理层的人员也是如此,当然他们既有部分决策职能,也有部分一般业务职能,他们的使用程度可能会深入一些。

而真正能够进行深入的数据分析或数据探索的人我们称他们为“高级数据分析师”,在整个用户单位内部,他们也许就那么几个人。他们既需要精通业务又需要熟悉技术,他们会不断地、深入地进行数据分析、数据挖掘与探索,并把他们的分析结果或分析模板分门别类发布给应该了解或使用这些结果的领导与业务人员。他们也会不断发现与收集新的需求,并配合数据仓库管理员扩展新的分析主题。

为了使 BI 系统真正应用成功,从一立项开始,企业就应该来定位和培养企业自己的高级数据分析师。

5.4 数据质量问题

数据质量问题是数据仓库应用上一个令人烦恼的问题,特别是对于企业信息化建设还没有达到一定程度的应用单位或企业来说。一般来说,应该在 ETL 这

个环节上,通过数据的有效性、一致性检查与完整性处理来着力解决这一问题,但据我们的经验,也不能全面解决问题。

一方面,通过实施数据仓库,建立企业或行业完整的业务模型和逻辑数据模型(LDM),来促使业务系统的数据进一步规范化,解决信息孤岛、数据一致性、数据完整性问题,但这样做是要一个相当长的时期的。所以在另一方面,我们也要在数据仓库应用上采取适当的办法来解决这一问题,如人为的数据完整性转化、多种方式的数据正确性稽核、或让所有的“问题数据”都带上“口径”的标记或烙印等办法来辅助解决这一问题。否则,系统的可用性是值得怀疑的。

5.5 关于数据仓库的逻辑模型与物理模型

对数据仓库建设来说,数据模型是非常重要的,数据仓库的逻辑模型与物理模型的设计对整个系统至关重要。从 ODS、底层数据仓库到数据集市,前二者涉及到企业数据的一致性、完整性问题的根本解决,后者涉及到分析主题、OLAP 的可扩展性。

从逻辑模型上来说,行业逻辑数据模型是关键。从物理模型上来说,粒度划分、Cube 定义或事实表、维表的定义及中间表的定义与基于性能考虑的优化是十分重要的。

5.6 展现问题

尽管目前许多前端展现工具厂家已有的展现系统确实也有较为完善与优秀的功能了,但我们还是认为, BI 的数据展现(数据仓库数据可视化)是一个极为重要的环节,绝对不仅仅是一些简单的直方图、饼状图、动态报表之类的展现方法所能解决问题的。在某种意义上,我们可以认为 BI 的展现门户是一个辅助决策支持的引导过程。所以,在 BI 的展现门户上,我们不得不考虑企业全局的指标体系与分析角度,基于指标体系的企业 KPI 定义、基于规则的预警、自动的信息与知识推送、可基于预警驱动的关联分析这样一个在更高层次上的展现体系或功能。

5.7 其他关键因素

首先,用户要有需求,也就是我们的用户要知道建立商业智能系统与企业数据仓库是为了做什么。反之也要知道什么是我们不能要求在商业智能系统与企业数据仓库上来作的,很多失败的案例就是无休止地在

数据仓库应用系统上增加本来应该在 OLTP 系统上实现的功能,这是不合适的。

也正是因为有需求,用户才会真正能够来使用系统,使用有初步成效了,又有新的需求产生,新的需求又推动应用系统的更新与扩展。所以从这个意义上来说,基于数据仓库应用系统决不是一个具体的产品,而是一个不断完善、不断更新的过程;系统长期的维护与升级对系统的成功应用至关重要。

虽然数据仓库应用的整体过程上不是一个产品,但对不同环节上的功能(如 ETL、展现、报表、即席查询、数据挖掘)在应用层次上产品化,对系统通用性的构架是有良好意义的。而即使如此,数据仓库的应用也决不是一个产品安装过程与集成过程,大量的二次开发是不可避免的。

最后,还要强调的是 BI 系统建设的统一规划与分步实施。所以既要在数据仓库一启动就考虑企业数据模型的整体规划与应用系统的可扩展性构架,但也不能和不必对第一阶段所实现的效果要求太高。

6 结束语

目前,我国对商业智能的研究与开发工作仍可说处于起步阶段,目前市场没有一个真正的领军者,突出的问题在于数据分析、知识发现能力效率低,或者缺乏知识发现,而更像一个操作型应用系统。当然,BI 的低端市场或某一义务点的商业智能应用也许可以用到一些套装化的产品来解决,但企业级的 BI 应用更应该说是一个整体的解决方案,是多个适用于不同环节的支撑平台或产品与业务及基于需求的应用开发的有机结合。希望本文关于若干问题的探讨能对着手进行 BI 系统建设的企业或 BI 解决方案提供与实施商有一定的帮助。

参考文献

- 1 Colin J W. Decision Processing :the Next Generation of Business Inteligence. Database Associates International Inc[EB\OL]. www. dmreview. com. 2001.
- 2 Hammer J ,et al. The Stanford Data Warehousing Projeet . http : //www - stanford. edu /warehousing /warehouse . htm1. 2001.
- 3 Inmon W H. What Is a Data Warehouse ? www. china3i. net. 2001.
- 4 Klauer P ,Brobst S. Building a Data Warehouse for Decision Support.(2nd Edition). Prentice Hall PIR Prentice - Hall , Inc. 1998.
- 5 谢榕. 基于数据仓库的决策支持系统框架,系统工程理论与实践 2000 4 27 - 30
- 6 余长慧,潘和平. 商业智能及其核心技术. 计算机应用研究 2002 19(9) :14—1
- 7 Michael. 数据仓库的数据稽核方法. www. ivpp. ac. cn 2003
- 8 A delman S , Mo ss L T. 数据仓库项目管理. 薛宇,王剑锋译. 北京:清华大学出版社,2003.
- 9 W H Inmon. Building the Data Warehouse. John Wiley&Sons 1997.
- 10 Rob Ericsson. Delve into Data with Business Intelligence. http : //www. flponline. com /wss/2003 _ 03/ magazine/features/erisson.
- 11 Jiawei Han ,et al. Intelligent Query Answering by Knowledge Discovery Techniques. IEEE Intelligent System ,1996 (8).
- 12 Freeman O. Competitor Intelligence Information or Intelligence. Business Information Review ,1999 16(2).