



## 烟草行业数据仓库解决方案

苏宏亮

(赛贝斯软件中国有限公司)

中国烟草行业自1981年开始实行统一领导，垂直管理和专卖专营的专卖管理体制，长期以来保持着良好的发展势头，烟草行业每年向国家上缴的税收占税收总额的10%，是关系国民经济发展的支柱产业。

烟草行业信息化工作也取得较快发展，计算机应用水平有较大提高。到2000年，有四家列入国家“863”计划的烟草重点企业的CIMS系统通过国家验收，17家大型烟草企业已初步建立和运行了企业MIS系统。“十五”末通过实施“金叶信息系统工程”建设，建成全行业上下左右互通互联的卫星通信网和信息管理网。

烟草行业信息化建设经过十几年的努力已初具规模，由单机应用、局部系统应用发展到了网络化、集成化、数字化，基本完成了管理信息系统的应用阶段，在行业经济发展中已发挥了重要的作用；同时烟草业对信息化建设又提出了更高要求，希望通过信息化建设进一步降低成本、提高管理水平，优化资源配置和利用，进一步提高其竞争力，辅助生产和经营决策。

我们可以看到，在经济发达地区的烟草企业已经率先开始新一轮的信息化建设，它们充分利用企业内部积累的大量业务数据，通过数据仓库、联机分析处理（OLAP）等当前先进的信息处理技术，建立分析模型，挖掘数据内部规律，指导企业生产决策，取得了显著的效果。

### 1 烟草行业目前所面临的问题？

烟草生产与销售是个特殊的行业，由于我国实行的是烟草专卖制度，整个管理体制自上而下设置，从国家烟草专卖局到省局、再到地市局、再到县局，最后到农村一级的销售网点，烟草销售体现的是“垂直管理、专营专卖”特点。

多年的计划经济和专卖体制，以及信息化程度不高的管理手段，使得烟草的生产和销售面临着如下问题：

a. 卷烟销售信息反馈慢、信息失真现象严重。

b. 超产时有发生，限产压库难度大，造成商业环节卷烟库存增加，销售难度加大，企业在市场竞争中难以实现优胜劣汰和资源的优化配置。

c. 烟草专卖局对烟草销售的调控能力需要加强，同时需要更准确地把握市场变化的规律。

而中国加入WTO后，烟草市场的开放进入日程。国外卷烟的进入会加剧竞争局面。因此，更新观念，采用行政与市场相结合的管理模式和相应的信息化技术手段，是烟草公司发展的必由之路。

虽然目前大多数烟草企业信息化建设已打下了一定的基础，但当前的信息系统只是达到数据处理的初级水平，许多信息系统往往只限于对数据的统计分析，而真正涉及到预测、决策及优化方面的工作还比较少，结果是只能代替一部分繁重的手工劳动，并没有充分发挥信息系统的效益。

### 2 如何在激烈的商业竞争中保持竞争优势？

科学的经济活动分析依赖于强有力的数据分析能力，而当前的实际情况是：多数烟草企业的数据分析能力远远不

适应企业经济活动分析和经营决策的需要，大量宝贵的数据资源没有得到充分的开发和利用，“数据丰富、信息贫乏”的现象十分普遍，由此严重制约了企业经济活动分析效果和企业决策。迅速提高数据分析能力，为经营决策提供更有力的支持是众多企业管理者的呼声。为此有必要在原有的生产业务系统基础上，构建一个“烟草业数据仓库系统”，以及构筑在其上的商务智能，来管理和合理利用信息，使之成为战胜竞争对手的有力手段。

数据仓库技术是综合了客户服务与业务处理的信息，采用分析挖掘技术，为企业管理人员制定策略、开发市场、分析市场、效益评估、公共关系与企业形象设计等管理行为，以及决策提供科学的依据。

数据仓库通过从企业内部的不同系统和外部数据源收集数据，经过抽取、整合和统计，形成一个中心的数据集，这样既保持了数据的一致性，又易于被用户访问。同时，这些数据按照业务概念来组织，例如按地区、品牌等进行分类，便于分析。

数据仓库有这样一些特点，例如：其中的数据是历史数据或统计数据；按照业务概念来组织；存储了大量的统计信息；数据易于分析；数据基本不作更改，仅用于查询；数据一般只存储一次，很少更改。此外，数据仓库还提供了联机分析（OLAP）功能，用户可以联机访问经过分析运算的多条数据。

数据仓库还克服了直接建立在业务系统数据库基础之上的决策支持系统的局限，例如：数据按照统一格式存放，用户可联机快速地查询复杂的数据；查询不会影响其它正在进行的交易；数据来自不同的系统，并按照分类存放。

简而言之，数据仓库是一个数据源，可以提供全面的企业信息。数据仓库也是信息仓库、数据集市的主要数据来源，是面向某类业务主题的。

### 3 数据仓库能为企业带来什么？

#### 3.1 烟草行业业务主题及主题关系

虽然不同类型的烟草企业有其各自的特色和业务关注点，但其对客户组成，客户行为，客户忠诚度，销售行为的关注却是相同的，因此，我们围绕客户、销售、生产和库存这些企业普遍关心的问题展开详细的分析，并归纳出如下相互关联的主要分析主题。除此之外，企业还可根据本企业业务需求灵活设定财务、人事、定单、服务等其它主题，并进行相应分析。



其中：

客户主题——客户关系管理和分析。是企业分析中比较重要的主题。

销售主题——从促销、卷烟品牌、时间等多个维度对影响客户购买行为的各种因素进行分析。是决策支持系统中最重要的主题。

库存主题——对周期库存、库存上、下限的合理分析，是影响内部叫强管理，降低成本的重要部分。

生产主题——对卷烟生产厂每个卷烟品牌的生产分析。

#### 3.2 烟草企业主题分析

在对每个主题进行完备的数据准备后，可以围绕每个主题展开详细的分析和统计，在下面的章节中我们将详细介绍每个主题可提供的主题分析内容。

##### 3.2.1 客户关系管理

建立客户关系管理系统(CRM)，在客户档案基础上，进行客户识别、客户细分、客户分析、指导经营。

建立以客户为中心的企业市场营销模式，对客户档案、联系记录、相关订单、业务进程等进行查询和分析。客户购买行为和价值取向进行深入分析，为企业挖掘新的销售机会，并对未来发展方向提供科学、量化的指导依据，使企业在快速变化的市场环境中保持持续发展能力。提高客户忠诚度和保有率，实现缩短销售周期、降低销售成本、增加收入、扩展市场，从而全面提升企业的赢利能力和竞争力。

#### a. 客户分析。

按地域、级别、进货量、销售量、存货量、产品品牌等类别对客户进行细分，找出关键客户。

#### b. 客户进货特征分析。

对客户购买行为和价值取向进行深入分析，挖掘新的销售机会，并对未来发展方向提供科学、量化的指导依据。

此外可以对客户忠诚度分析、客户满意度分析，确定客户忠诚度和保有率，找出目标客户，为客户提供信息支持，从而创造销售机会。

### 3.2.2 销售分析

根据各省、市、县卷烟销售公司和烟厂的历史销售数据，建立销售分析系统，分析卷烟销售情况（包括卷烟销售时间（年、季、月）、区域（省、市、县和地区）、数量、结构、产地、牌号、规格）、人口、流动人口、社会经济发展水平、居民可支配性支出、卷烟市场发展趋势等因素，设定卷烟库存定额总量（包括数量、结构、产地）；对于市、县等销售公司还可以根据销售网点的数量、配送能力、销售主导方式（大访销、大配送、访送一体、访送分离）等要素进行分析，预测每种品牌的卷烟在各个地区的销售量和销售额等情况。

### 3.2.3 库存分析

库存分析数据主要来源于两方面，一方面是随进货单，出货单随时变化的库存数，另一方面是定期进行库存盘点获得的实际库存数，合理的库存量可最大限度地减少脱销，并减少存货维护的成本和资金占用。对于配送中心，可对库存卷烟从时间、地域、厂家、品牌、类别、批次等多个角度，

- a. 分析卷烟的库存上、下限；
- b. 分析卷烟的最佳库存量；
- c. 分析卷烟的库存成本；
- d. 分析卷烟的库存周期；
- e. 分析卷烟的损坏和丢失等保管情况；
- f. 到期卷烟预警；
- g. 分析库房的使用率和成本；
- h. 分析最佳配送产品组合，时间，车次和路线；
- i. 分析合理的库房位置和卷烟存放；
- j. 分析库存卷烟的资金占用成本。

### 3.2.4 生产分析

根据销售和库存情况，以及销售趋势预测，制定每种品牌卷烟的生产量，保持安全、合理的库存。

对于卷烟生产企业，可以从企业ERP系统中提取相关的业务数据，对卷烟生产过程中涉及的每一个环节、工序和材料——烟叶、原辅材料、零配件等进行分析，分析卷烟的成本构成，从而降低生产成本，提高生产率。

### 3.2.5 分析决策

通过将CRM 分析系统、销售分析系统和生产分析系统的结合，将企业的各种业务信息统一存储在数据仓库之中，

可以方便的对客户、产品、进程、任务、预算、计划、费用等等信息进行分类统计，用以分析销售、市场和服务业务的运行情况，从而做出科学、正确的决策。利用分析决策功能可以产生一下报表：

- 1) 月度总销量表（作同比分析）；
- 2) 分烟厂销量表；
- 3) 骨干品牌销量分析；
- 4) 品牌利润贡献度分析；
- 5) 分批发部销量分析；
- 6) 分客户类型销量分析；
- 7) 客户利润贡献度分析；
- 8) 服务成本与效益增长率表；
- 9) 大客户销售预测（销售量与销售额）；
- 10) 品牌销售预测分析（销售量与销售额）；
- 11) 区域销售预测分析（销售量与销售额）。

### 3.3 企业从数据仓库获益

数据仓库可以帮助用户从繁杂的数据中获得有效的分析数据，并在此基础上做出正确的决策。引入数据仓库系统，可以使激烈竞争中的烟草企业进一步扩大利润，加大竞争实力，并在激烈的商业竞争中保持长久的竞争优势。

建立数据仓库系统能够带来的益处和收益：

- a. 实现烟草销售的快速供需链管理，打造烟草流通企业的核心竞争力；
- b. 及时有效的监控市场动态，对市场竞争做出快捷反应；
- c. 准确分析市场活动的回报，培育品牌挖掘市场卖点；
- d. 实现对烟草供应厂商完备的信息支持，实施合作伙伴关系管理；
- e. 建立完善的客户资料档案库，实施科学的客户分类评价指标体系；
- f. 提高服务质量降低服务成本，以最小的代价留住最大价值的客户；
- g. 与企业其它信息系统无缝连接，信息共享相互促进。

### 4 传统MIS系统能否替代数据仓库系统？

由于国内缺乏众多的数据仓库成熟案例，很多用户对于什么是数据仓库？数据仓库能提供什么了解不甚清楚。数据仓库本身分析的不确定性和灵活性，又给其蒙上了一层神秘的面纱，因此很多用户对数据仓库产生一定的误解，主要表现在三个方面：

- 1) 夸大企业需求，弱化数据仓库功能。

这类用户的典型现象是迫切需要数据仓库技术，而实际需求只是灵活的统计报表。实际上，报表分析只是数据仓库的一部分，是数据仓库的起步，数据仓库还可以为企业提供更深层次的专题分析和挖掘，并且层次越深，产生的价值越大。

- 2) 夸大OLTP系统需求，淡化数据仓库需求。

典型用户是认为现有的MIS系统可以支撑其全部数据仓库需求，所以不需要再建立单独的数据仓库系统，该类用户实际上也在一定层次上弱化了数据仓库的功能，实际上MIS系统可以替代的只是数据仓库中的统计报表功能，其它钻取挖掘等功能都不能实现。

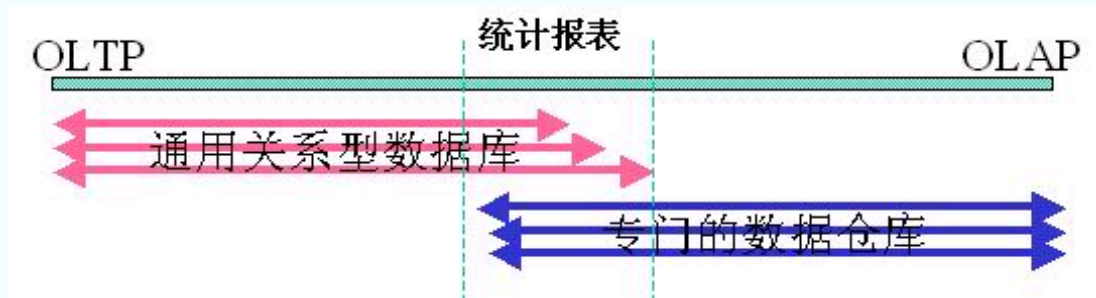
- 3) 希望上数据仓库，但提不出需求。

典型用户是知道数据仓库的部分功能，但不知道数据仓库具体能为本企业带来什么？

数据仓库与传统OLTP系统的区别是什么？为什么传统的MIS系统不能替代数据仓库？

a. 影响业务系统的正常运行。

实际上，数据仓库是从OLTP系统不断的抽取汇总需求发展出来的，因此统计汇总是两类系统的交界线，即不论是企业原有的MIS系统，还是新的数据仓库系统都提供了报表打印和统计汇总的功能，在这方面，两者的功能是交叉的，这也是为什么很多用户对数据仓库的功能产生误解的主要原因。



在数据量较少时，传统的业务系统的确可以满足企业统计报表的需求，但在数据量大时，如果将统计分析 with 日常业务处理混合放在一起，由于统计分析会占用大量的内存和CPU等资源，会极大地影响业务系统的正常运行。如在销售高峰期做一个全年销售数据统计排名，则会极大程度影响前端POS的正常销售业务。因此有必要也必须将统计分析转移到数据仓库中。

b. 不能支持灵活的即席查询需求。

在采用数据库技术的业务系统中，通常是采用索引等技术来提高查询性能，而这些索引的建立是建立在对既定查询优化的基础上。而在数据仓库项目中，数据仓库的查询需求是灵活的，未知的。这也是数据仓库区别于传统业务系统的一大优势和特点。但由于需求过于灵活，所以我们不能为每个查询请求进行即时优化，因此查询性能会受到极大的影响。而专门的数据仓库是专为查询分析而设计的，其对每个查询列的可能查询方式都提供了相应的优化技术，可支持灵活的查询需求，提高查询速度。

c. 不能容忍的查询性能。

通常情况下，业务系统只保存最近使用数据，而数据仓库为了更精确的分析挖掘系统规律，需要一定时间积累的历史数据，如3年，5年甚至10年，因此数据仓库中的数据远大于业务系统，如此大规模数据，不仅影响了业务系统的正常运行，而且也极大地降低了统计查询的速度，其速度几乎是不可忍受的。而专业的数据仓库由于是专门为查询分析设计的，提供了很多优化技术，因此可提高10~1000倍的性能。

d. 数据组织方式。

数据库主要是针对业务系统，一般采用实体关系模型，而数据仓库主要针对主题进行分析，所以较多采用多维模型，因此其数据的内部组织形式就完全不同，传统的实体关系模型组织形式不适合于企业灵活的挖掘，钻取等需求，因此也不能替代专业的数据仓库。

因此，如果是一个小业务系统的复杂报表统计，可以选择既有的OLTP业务系统；如果需要灵活的报表统计，请选择数据仓库；如果要对大量的历史数据进行分析，请选择数据仓库；如果在统计分析基础上，还要进行更深层次的钻取和挖掘的话，数据仓库则是您的必然之选。

## 5 如何成功实施烟草业数据仓库？

数据仓库是一个相对独立的项目。建立数据仓库为企业带来了一个独特的机会，其结果立竿见影，不会影响正常

的业务运营。

## 5.1 数据仓库实施过程

数据仓库决策分析的业务处理流程主要分为四个阶段，如图1所示。

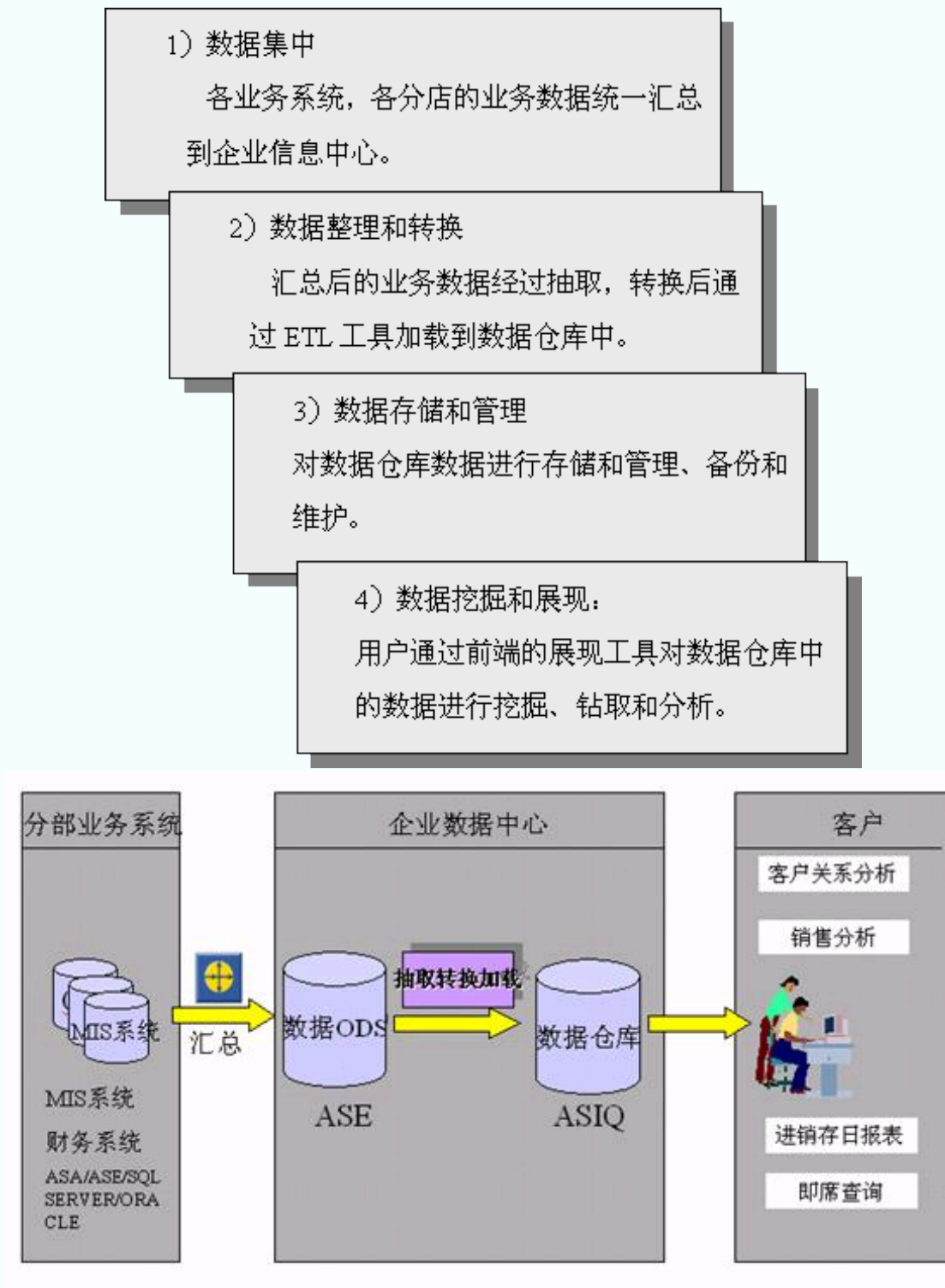


图1 数据仓库决策分析的业务处理流程

### 5.1.1 业务数据集中

数据仓库必须建立在完整、统一的业务数据基础上，而在大部分企业内部，业务信息分散在不同部门的不同业务系统中，这也是为什么原有的业务信息不能共享或信息不一致的主要原因。因此在建立数据仓库之前，有必要将所有需要的信息整合到一起。然后再在此基础上进行整理和分析。

在以前的业务系统数据传送中，总部对分支机构信息的获得大部分仅限于汇总信息，对于销售流水等明细信息并

不进行整理和传送。但对于数据仓库系统的即席查询等需求而言，很可能钻取到最小粒度，因此数据仓库的数据传送必须细化到最小粒度。这是数据仓库对信息传送的重要要求。同时随着数据仓库的不断深入，企业需要分支机构提供数据会有所变化，因此企业必须提供灵活的数据传输和同步方案，满足数据仓库当今和未来发展的需求。同时由于不同业务系统平台和软件可能完全不同，在数据集中过程中还必须考虑到异构平台和异构数据源的同步问题。

### 5.1.2 数据抽取转换和加载

由于汇总的数据来源于不同的业务系统，因此出现信息不一致的现象不可避免，因此有必要对汇总后的数据进行一定的内容整理和处理。并且原有的以业务为主的MIS系统的存储模式也不一定适合于以分析为主的数据仓库维模型，也需要重新进行整理和转换。因此必须对数据进行清洗转换后，才可以加载到数据仓库中，供业务人员和领导使用。只有正确的数据才能分析出正确的结果，本身就不一致的数据不能提供任何价值信息。

数据抽取转换的过程可以根据转换复杂度，资金投入等不同因素考虑不同的实现方式，如手工处理方式或专业的ETL工具等，企业可根据自己的需求合理选择。但不论是采用手工处理方式还是工具自动处理方式，在实施该阶段必须注意的一个重要问题是：该阶段在整个数据仓库实施过程中工作量所占百分比。很多企业认为该过程只是一个任务整理过程，所以对该阶段任务估计不够，时间分配不足，这种心理上的轻视很大程度上影响了数据仓库项目的成功。而实际上，该阶段占据了整个数据仓库实施的大部分时间，在有些项目中甚至高达80%。

### 5.1.3 数据存储和管理

抽取转换后的数据必须保存在一个高效的数据存储中心中，为用户的统计、分析和挖掘提供信息基础。与业务系统仅保存最近的业务数据不同，数据仓库保存的是大量的历史数据，少则3年，多则10年、20年，其数据量是相当大的。数据量大一方面为系统分析提供了更多的数据参考，使趋势分析更加准确和理性；另一方面也给系统带来了极大的负面影响：其一是影响系统查询性能，其二是加大系统的存储投资成本。因此在系统实施过程和数据仓库选择时，必须考虑到：如何保证系统在大数据量下的查询性能？如何降低大数据量的存储成本。

由于数据仓库主要负责历史进销存业务的统计、分析、报表和挖掘，其对数据的修改相对来说很小，而且大部分操作是对批量数据的统计，而不是单条业务的处理。如一般是某类商品的销售分析，而不是具体的某条销售流水，因此对其性能要求主要集中在查询性能，而不是删除，加载性能。这也是选择数据仓库存储时与选择传统的业务系统存储时的一个重要区别。

### 5.1.4 数据展现和分析

在完成了数据集中，转换和存储后，就已经为数据仓库分析准备了足够的信息，但如何从众多的信息中发现规律和价值，还必须依靠一个优秀的展现工具。

在传统的MIS系统中，所有查询的内容和报表都是固定的，如进货单、商品明细帐、总帐等，对于这些系统，传统的MIS开发工具，如PB、VB、Delphi都是很好的选择。但数据仓库的查询则是随机，灵活变化的，在具体查询前没有固定的格式，查询的结果也是不可预知的，在此情况下，如果仍采用传统的报表工具就会极大的限制查询的内容和灵活性，因此必须采用专业的数据仓库分析工具，用户可以根据自己的需要，随时进行各种方式的查询和分析。

正确的展现工具的选择也是影响数据仓库实施成功的重要因素。

## 5.2 影响数据仓库成功的关键因素？

在认真分析了数据仓库实施前提和实施过程之后，我们可以总结出如下影响数据仓库成功的关键因素：

- 1) 以业务需求为导向，而不是技术需求为导向，不是为建数据仓库而建立数据仓库；
- 2) 领导的大力支持以及业务部门和信息中心的通力配合；
- 3) 灵活的，支持各类平台和数据源的数据集中工具；

4) 用户和集成商对数据转换加载阶段的足够重视;

5) 选择支持大数据量查询的数据仓库软件, 以提高查询性能, 降低存储成本;

6) 选择专业的数据仓库工具, 而不是传统的报表工具。

7) 合理的数据仓库切入时间等等。

以此为指导, 我们可以大大提高企业数据仓库的成功概率, 降低实施风险。

### 5.3 数据仓库实施过程建议

设计和实施数据仓库是一个复杂的过程, 许多企业没有相应的经验和资源。为此, 必须提供了一个低风险、专业的解决方案, 从而为企业开发数据仓库提供了一个坚实的基础。数据仓库的应用实施建议分为两个阶段。

**第一阶段: 解决决策支持查询、提供多维分析。**

这一阶段转移OLTP系统中的部分决策支持查询和报表, 提供即席查询、灵活报表和多维分析, 并可以通过Client/Server和Internet/ Intranet等多种途径展现。应用模式包括: 静态报表、动态报表、即席查询、OLAP分析、EIS、DSS。

**第二阶段: 数据挖掘。**

对数据仓库进一步完善, 提供相关性分析、趋势分析、波动分析等。

根据以往实施数据仓库的经验, 分两个阶段实施可以使数据仓库边开发边见效, 降低项目决策者的疑虑, 在实际使用中不断检验和完善数据仓库系统。如果跨过第一步而将目标定得过高, 会造成系统应用面变窄, 项目启动困难, 失去数据仓库对决策支持普遍支持的重要意义, 人为增加项目开发的风险。

---

作者简介: 苏宏亮, 现任Sybase软件(中国)有限公司高级系统顾问。成功主持过多个大型信息化项目的建设。曾参加过国家烟草专卖局25个MIS系统的开发建设, 有丰富的烟草行业信息化建设经验。