

# 商务智能在保险数据分析和决策支持中的设计与实现<sup>①</sup>

陈鸿雁 (中国人寿保险股份有限公司重庆市分公司 信息技术部 重庆 渝中 400011)

**摘要:** 介绍了商务智能在保险数据分析和决策支持中的设计与实现。系统实现了数据仓库的建设、数据挖掘模型的开发、多维数据集的开发,以及用 Sharepoint Server 2007 开发的信息门户实现对 BI 展示层应用的集成,最终实现自动化智能决策。详细阐述了该系统的解决方案、关键技术及实现、展示工具及展示效果,目前该系统已在中国人寿保险部分省分公司得到了广泛深入应用。最后提出成功应用 BI 的关键是要从管理层进行推动并培养应用队伍。

**关键词:** 商务智能; 数据仓库; 数据挖掘; BI 展示

## Design and Implementation in Insurance Data Analysing and Decision-Making Based on Business Intelligence

CHEN Hong-Yan (Information Technology Department, Chinalife Insurance Co. Ltd Chongqing  
Branch, Yuzhong 400011, China)

**Abstract:** This paper introduces the design and implementation in insurance data analysing and decision-making based on business intelligence. This system implements the construction in data-warehouse, the development of data-mining and multidimensional data collection, and the implementation of integration of BI lay-out-layer using Sharepint Server 2007. Then it introduces the resolution of this system, key technology and implementation, lay-out tools and effect with all the details and particulars. It has been widely and deeply used in some filiale of China Life insurance Co.Ltd. At last this paper provides that the key to implementing of BI is bringing up the application team with the promotion of management layer.

**Keywords:** BI; data-warehouse; data-mining; BI tools

### 1 引言

经过多年来的业务发展和信息系统建设,保险公司积累了大量的历史数据,如何充分利用这些数据,把它转化成商机,一直是各家保险公司思考的问题。商业智能技术提供了一个良好的方法,把分散的数据集成到一个数据仓库系统中,利用在线分析技术和数据挖掘技术能帮助公司对业务进行分析,寻求业务规律,为领导层提供决策支持。因此,商业智能走进保险行业,已成为一种趋势和必然。

商务智能(Business Intelligence, 简称 BI)应用

可分为数据报表、数据分析和数据挖掘三个层次,目前我国保险公司的 BI 应用多数还处于数据报表阶段,属于 BI 的低端应用,而国外大部分企业已进入了中端 BI(即数据分析层次),少数企业也已经进入高端 BI(即数据挖掘层次),这给我国保险业带来了新的挑战。本文将结合自己在对中国人寿某分公司保险数据分析、支持公司决策方面就数据分析和数据挖掘应用与大家进行分享。

### 2 保险商务智能解决方案

随着保险实业的逐步增多,保险竞争日益激烈,

<sup>①</sup> 收稿时间:2010-03-09;收到修改稿时间:2010-05-01

保险公司管理层希望通过保险数据分析、挖掘,寻找隐藏数据背后的深层次信息,辅助公司管理决策,提高公司的商务决策能力。根据保险业的行业特点及管理层的决策信息需求,提出了全面解决保险决策的商务智能解决方案,该方案以先进的数据仓库(DW)、联机分析处理(OLAP)、数据挖掘(DM)技术为基础,以数据抽取、查询、分析、知识发现为表现形式,为管理者决策层提供代理机构/代理人分析、销售主力跟踪分析、客户忠诚度分析、客户预测、利润/成本分析、产品分析、经营风险分析、市场情报分析等,提升企业风险控制和经营决策能力,提高了管理决策的科学性、及时性、准确性,增强公司综合竞争力。

### 2.1 保险商务智能解决方案

保险商务智能系统由 OLTP 数据源、ODS、数据仓库、OLAP 服务器及挖掘模型、前端展示五部分组成,如下图所示:

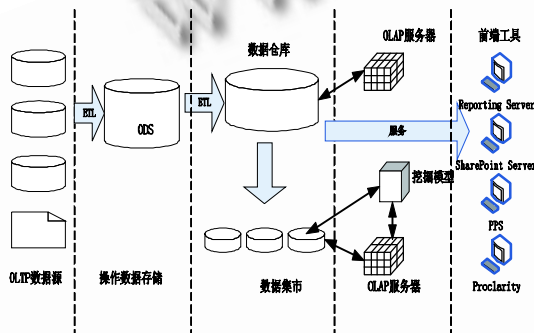


图 1 BI 系统系统结构图

① OLTP 数据源:即所有保险联机事务处理数据库,以及其它非结构化数据。

② ODS 层:即操作数据存储,是对多个异构 OLTP 数据源经过 ETL(即数据抽取、转换、装载)过程按照主题进行有效集成,定期刷新,包含当前有效数据,具备四个特点:面向主题、集成性、近实时数据发布、当前数据。

③ 数据仓库:包含大量从 ODS 层传送来的历史数据,传入数据一般不再修改,它是面向分析型数据处理,支持分析决策,不同于操作型数据库,具备四个特点:面向主题、集成的、相对稳定的、反映历史变化。

④ OLAP 服务器[4]和数据挖掘模型层: OLAP 服务器是支持和管理多维数据结构的数据处理引擎,实现多角度、多层次的分析;数据挖掘模型用于发现

先前求知的、有效的、可实现的信息,用于预测未来趋势及行为。

⑤ 前端展示层:主要包括各种数据分析、数据挖掘工具集,再通过 Sharepoint Server 整合信息,形成统一门户,成为用户的唯一信息获取点,方便用户获取各种信息。

### 2.2 保险商务智能解决方案的特点

① 完整、有效、统一的数据体系:集成管理所有影响企业管理决策的内外部数据,经过 ETL,按照保险行业标准指标定义体系,形成统一规范化数据。

② 及时、灵活获取信息:用户可在无须信息人员支持下,按照自己需求灵活制作精美报表,按照自己对问题的理解进行多角度、多层次分析,及时发现问题本质,也可输入进行数据挖掘、趋势预测、发现知识,促进管理决策。

③ 信息展示直观化、多样化:通过图形直观展现,使管理能快速获取 KPI 和风险点异常变动,及时调整战略。

④ 深挖信息,预测未来:从大量历史数据中找到潜在的、有价值的信息和规则,使保险公司可以预测未来,掌握先机。

## 3 关键技术及实现

保险商务智能体系在实施过程中,涉及 ETL 技术、维度设计、索引策略、MDX 开发、DMX 开发、数据存储分区等诸多技术及各种 BI 展示工具的组合应用,在此仅对一些关键技术与大家进行分享:

### 3.1 ETL 增量数据抽取控制

常见的 ETL 增量获取方式有触发器、时间戳、全表对比和日志对比,在该项目中,为保证数据抽取的准确性,我们对 OLTP 数据源采取触发器方式抽取,抽取到中间临时表,再对中间临时表中的“轨迹”数据按业务键抽取最近一条数据至 ODS 层中,ODS 到数据仓库、数据仓库到数据集市的数据抽取全部采取时间戳方式进行。

### 3.2 数据仓库模型的反规范设计

数据仓库是采取星型模式或雪花模式和面向主题的设计关系数据库,减少 I/O 开销,提高 I/O 效率,为在设计中存在许多反规范化的设计:

① 合并表[2]设计,即将数据量少且经常被相关访问的字段组合成一个新维度表;

② 数据数组<sup>[5]</sup>设计, 即当数列中值的数据稳定、数据是按顺序访问的、数据的创建与修改在统计上是以非常有规律方式存在, 创建一个数组;

③ 导出(即已计算出的)数据设计, 创建的创造性索引或创造性概要文件, 或在月末处理时, 计算出月平均、月最大、月最小等值, 生成导出数据表;

④ 分表<sup>[2]</sup>设计, 通过对数据稳定性进行分析, 对表中各字段访问率, 将访问频繁相同的字段组合成一个新表。

### 3.3 数据仓库的 ETL 设计技巧

数据仓库的 ETL 设计的目的是让数据仓库的自动化实现, 即维度表和事实表的生成:

① 顺序上, 先远端维度表, 再近端维度表, 最后生成事实表, 为方便管理, 均统一添加整型代理键和时间戳;

② 常规维度表的抽取流程, 如下图所示:

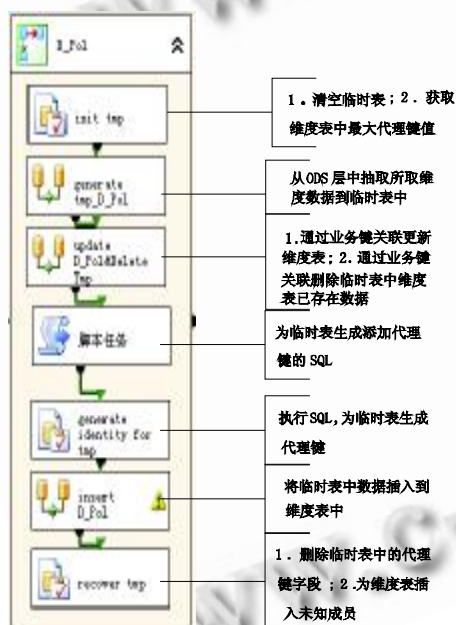


图 2 常规维度表抽取流程

说明:

“脚本任务”中为临时表添加代理键的 SQL 如下:

```
alter table tmp_d_pol add key_pol_id int
identity(9000,1)
```

其中“9000”为流程中的第一步计算而来的最大代理键值加 1 计算而来, 作为生成键的种子, 避免与维度表中的代理键值重复;

“recover tmp”任务的 SQL 如下:

```
if not exists (select 1 from d_pol where
key_pol_id=-1)
insert into d_pol (key_pol_id,pol_gfjt_
flag,pol_branch_no,pol_name,pol_cbps_code) values
(-1,!,'660000','未知,!);--未知成员
alter table tmp_d_pol drop column key_pol_id;--删除
代理键
```

③ 渐变维度表<sup>[1]</sup>的抽取是将图 2 流程的第二、三步合并成一步, 通过数据流转换中的“渐变维度”插件实现, 根据需求, 需要将个人代理人和客户用渐变维度实现, 另外, 渐变维度的时间戳应该有两个, 分别表示开始有效时间和结束有效时间;

④ 父子结构维度表<sup>[1]</sup>的抽取则需要将图 2 流程中最后一部添加对父代理键的关键更新, 语句类似如下:

```
update a set a.pri_dept_id=b.key_dept_id
from d_saleddept a,d_saleddept b
where a.pri_dept_id is null --父代理键
为空白
and a.key_dept_id!= -1
and a.pri_dept_code=b.dept_code; --业务
键关联
```

### 3.4 两种数据挖掘模型的介绍

根据海量保险客户资料, 分析客户: 性别、年龄、婚否、职业、收入、教育程度等特性, 建立数据挖掘模型, 预测已购买保险的客户可能还会购买哪些保险, 还需要哪些保险责任, 未购买保险的新客户, 根据其客户特性, 预测其最可能购买哪些保险及哪些保险责任, 这些预测信息可用于老客户保险深度挖掘, 也可为新入司销售人员提供业务指导。在该项目中, 我们用了决策树和关联规则进行预测, 还分别介绍如下:

① 决策树<sup>[3]</sup>是把根据条件(不论是离散还是连续)自动分解为多个离散的类别, 典型应用为一个或多个变量预测目标, 变量对目标的重要程度, 如: 预测客户是否会购买某种产品, 适用于分类、预测, 预测潜在客户;

② 关联规则<sup>[3]</sup>是处理海量历史数据, 快速找出产品之间的关联, 找出“先购买什么?然后购买什么?接下来购买什么?”的规则。

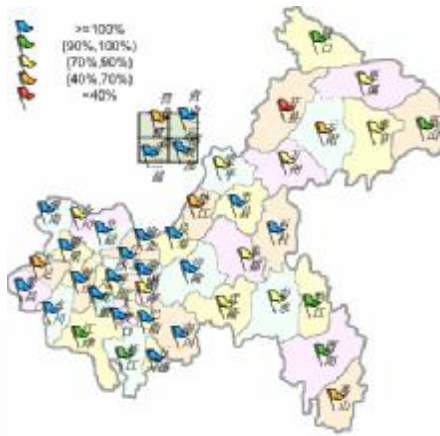
## 4 BI展示工具简介及展示效果

常见的有 Reporting sevice、Performance

Point、Proclarity、Dundas 等,实现柱状、饼状图、分解树、鸟瞰图、投影图、复合图、地图等多种形式的 BI 展示,下面对部分展示图进行介绍(下图中所用数据全为测试数据):

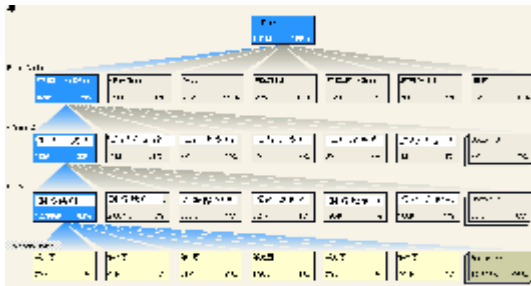
### 4.1 战略图

能从全局、直观掌握全市 KPI 的完成情况,及时发现风险点,以便及时进行计划调整,掌握全局主动权,适用于公司高层管理者、决策者使用。



### 4.2 分解树

可实现宏观到具体,一层层展开分析,可对一个指标进行任意多角度分析,是进行原因分析的主要手段,适用于公司各级管理层。



### 4.3 复合图

通过仪表盘对关键指标当前完成情况进行直观展示,通过柱状图对即将影响关键指标完成的相关指标进行预期展示,使管理者不但对当前任务完成情况有及时了解,而且对任务完成预期也有了及时掌控,适用于分公司、基层公司或团队对当前经营状况进行分析。

### 4.4 客户预测

通过数据挖掘模型实现新老客户的预测,可用于老客户保险的深度挖掘,也可用于新入司销售员的展业指导。



新客户及意向数据

新增客户: 1,000 | 意向客户: 2,500 | 潜在客户: 5,000

新增客户: 1,000 | 意向客户: 2,500 | 潜在客户: 5,000

新增客户: 1,000 | 意向客户: 2,500 | 潜在客户: 5,000

客户	客户名称	性别	职业类别	职业类别
客户1	张三	男	教师	教师
客户2	李四	女	医生	医生
客户3	王五	男	工程师	工程师
客户4	赵六	女	公务员	公务员
客户5	孙七	男	自由职业	自由职业

## 5 结束语

本文结合自身工作实践,介绍了商务智能在保险数据分析与决策支持中的设计与实现,但由于篇幅有限,许多技术细节,如:一元运算符、MDX、DMX及挖掘数据的回写等未能做一一介绍,旨在起到抛砖引玉的作用。商务智能系统是对传统管理思想和方式的转变,是让管理者彻底从“拍脑袋”决策向“数字决策”、“精准决策”的转变,其成功实施关键在于应用,而不在于技术本身,因此,我们应从管理层进行推动,并培养 BI 应用队伍,商务智能系统才能真正发挥其商业价值,也只有这样,才能真正实现企业智能化决策,商务智能系统才能真正成为企业迎接竞争、迎接挑战的有力利器。

### 参考文献

- 1 Ralph Kimball, Margy Ross 著.谭明明金译.数据仓库工具箱;维度建模的完全指南(第二版).北京:电子工业出版社. 2003.10.
- 2 William H.Inmon 著.王志海译.数据仓库.北京:机械工业出版社. 2007.12.
- 3 ZhaoHui Tang, Jamie MacLennan 著.邝祝芳,焦贤龙,高升译.数据挖掘原理与应用—SQL Server 2005 数据库.北京:清华大学出版社. 2007.1.
- 4 Brian Larson 著.赵志恒,武海峰译. Microsoft SQL Server 2005 商业智能实现.北京:清华大学出版社. 2008.3.
- 5 朱德利. SQL Server 2005 数据挖掘与商业智能完全解决方案.北京:电子工业出版社. 2007.10.