

数据仓库在医院的应用与实践

广州市红十字会医院 陈宁

摘要：本文介绍医院数据仓库的设计和开发过程，介绍了数据仓库在医院数据查询、统计分析、医疗质量分析、工作效率评估等方面的应用。

关键词：数据仓库、多维分析

医院信息系统经过多年的建设和发展，从挂号收费系统到临床信息系统，几乎覆盖了医院工作的各个方面。在这多年的应用中，积累了大量的原始数据，如何对这些信息进行再开发和利用，从中提取出有用的信息，为临床和管理服务，是我们信息化的目标和任务。

数据仓库技术为历史数据的分析和处理提供和很好的工具和方法。我院自 2005 年开始着手进行医院数据仓库的探索，到 2006 年上半年初步建立并投入使用，取得良好的效果。现把我院建立数据仓库的过程、基本思路及经验作个总结，和大家交流。

一、 系统建立目标

1、 整合各个应用系统的报表，做成统一的查询分析系统。目前医院已经有了比较完整全面的信息系统，有 HIS 系统，财务系统，病案统计系统、及仓库、设备等等，各系统各自为政，要查询报表只能到各自的系统，很不方便，关键的问题是各系统的统计口径不一样，出来的结果也很不相同，使数据的可靠性受到质疑。

2、 整合历史数据。我们自 1996 年就开始实施信息系统，有多年的数据积累和应用，从最开始的 DOS 操作系统，FOXBASE 数据库，到现在的 windows 系统，MSSQL 数据库，系统几经升级更换。系统更换后，以前的数据连查询都困难了，更不用说做深入的分析和应用了。

二、 系统总体设计

采用 B/S 结构，后台采用 MSSQL 数据库，建立专用于查询分析的数据库，中间件采用第三方的多维数据分析及展示工具，用户端用 IE 浏览器。

三、 数据源分析

1、 HIS 数据源：由于 HIS 系统经过几次升级和更新，以前的系统和现在的系统数据结构有很大的差异，就是现在的系统，为了提升速度，也对数据进行了拆分，分成了当前库和历史库，在当前库中又分了当前表和年表，分析不易。因此，我们把 HIS 数据源分成二部分处理，一

是当月数据，直接以当前的 HIS 数据库为数据源，只作较为简单的和必要的查询，二是隔月数据，定时导入到数据仓库中，数据仓库数据库在物理上与在线处理系统分离，这样做大数据量分析的时候就不会影响到在线处理系统的运行。

统计病案数据源：病案统计数据经过病案和统计人员层层把关和核对，数据比较完整，特别是病案首页信息，数据质量比较高，可以直接作多维分析的数据源。但是因为系统经过几次升级，病案统计中有一套独立的科室、病区、人员编码，与当前系统不同，无法与当前的 HIS 数据进行对比分析，我们只好把病案统计系统中的数据也定期转入到数据仓库中。

四、 数据库设计

虽然不同的公司提供的应用软件，数据结构有很大的差异，数据库表的设计有极大的不同，但都可以分为以下二类信息：

一是业务数据：该数据每时每刻都在变化，这类数据中在医院里最主要，数据量最大的有二个，一是收费明细记录，二是病人的就诊信息。

二是字典数据：这类数据，数据量不大，变动也不大。如科室字典、人员字典、药品字典，项目字典等。

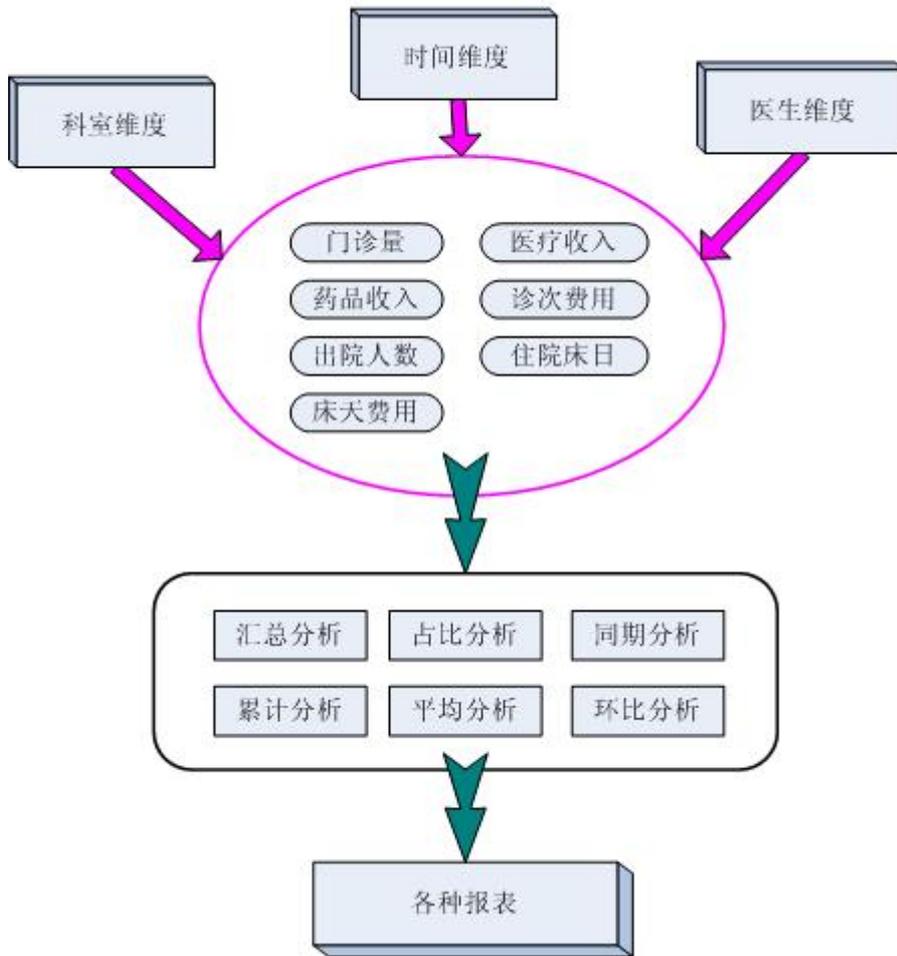
基于以上分析，我们在数据仓库中建立以下表，住院病人首页表，门诊病人就诊信息表，门诊收费明细表，住院收费明细表，以这些数据作为基本事实表。药品项目表，药品项目分类表，科室信息表，职工信息表，以这些数据作为维度。再根据各主题需要增加一些必要的数据库表。

五、 主题应用。

（一） 一般应用分析

此类应用侧重点在医院组织架构的层次维度（医院各业务科室、人员）、时间层次维度（年/季度/月/时段），通过对工作量、医疗收入、门诊诊次费用、住院病床使用情况、用药情况等量度进行汇总分析、占比分析、同期分析、累计分析、平均分析、排行榜分析、环比分析。提供多个视野角度由汇总到明细的层层分析，便于医院高层和各科室主任及时掌握医院和科室的工作状况。

多维分析模型涉及的维度、量度、分析方法图



1、门诊医疗分析

维度：时间、科室、医生。

度量值：门诊量、医疗收入、药品收入、诊次费用。

2、住院医疗分析

维度：时间、科室、医生。

度量值：出院人数、住院床日、医疗收入、药品收入、床天费用。

3、住院收入分析：

维度：时间、科室、医生。

度量值：按财务大项分类。

4、用药分析：

维度：时间、科室、药品（一级分类、二级分类）。

度量值：数量、金额。

5、项目收入分析：

维度：时间、科室、收费项目（项目类别）。

度量值：数量、金额。

（二）统计报表

此功能是与省卫生厅统计信息系统相联，把省厅统计信息系统中的数据作为数据源，把其中的一些报表移植过来，做进一步的查询和分析。主要有：医院工作报表（门诊部分），医院工作报表（住院部分），医院出院病人疾病分类报表，医院出院病人疾病年龄构成报表，医院出院病人手术分类报表，医院出院病人病人来源报表，医院出院病人部分病种报表，医院经济效益报表。这些报表转入后也可以按时间、科室作为维度，以相应的指标作为度量值进行多维分析。

（三）深层应用分析

1、多元函数变量因素分析

多元函数变量因素分析是描述各因素对事物影响程度的统计方法，强调了因素间影响程度的主次，可找出影响变量变化的主要原因。应用多变量因素分析法，对医院医疗收入状况进行分析，以了解相关因素对医疗收入的影响。

2、科室医疗质量综合评价分析（TOPSIS 法）

在医院的医疗质量评价中，被评价的各指标之间往往出现此好彼差的现象，因此，就必须采用客观、可行、定量、准确的多指标综合评价方法。利用 TOPSIS 分析法，把表示医院医疗质量的指标：治愈率、病死率、病床周转次数、病床使用率、平均住院日、入出院诊断符合率、危重病人抢救成功率、无菌手术感染率等指标。组成多指标综合评价体系。通过归一化处理后找出各项指标的最优和最劣向量，然后通过计算各指标与最优向量和最劣向量之间的距离。求出各指标与最优指标的相对接近程度，计算出综合评价体系的总评分，以此来对科室的医疗质量指标进行综合评价。

3、科室病床利用情况的综合评价（RSR 法）

选用科室病床使用率、病床周转次数、平均病床工作日、出院病人平均住院日 4 个常见统计指标组成综合评价指标体系，采用秩和比（RSR）法进行综合评价分析，并得出 RSR 质控图。RSR 值反映被评价科室之间得相对优劣程度，RSR 值比较接近，表明被评价科室的水平相当，RSR 值差异较大，表明被评价得科室病床使用情况存在较大差异。

4、引入病床工作效率分析科室病床配置是否合理

根据任一时间段的 RSR 质控值、实际占用总床日数、出院人数、平均开放病床数，给定病床工作效率的控制值，得出开放病床数的合理区间，作为科室配置病床数的科学依据。

六、 结语

数据仓库技术引入医院信息系统，从根本上改变了原来的报表方式，由原来被动的、固定的报表模式，到现在各级领导可根据自己的想法，从现有的模型中主动挖掘出自己想要数据，能够充分挖掘医院信息系统的信息资源，提高医院信息系统的的应用分析水平。

数据仓库应用是一个不断深入、不断完善的过程，完成了一个主题，一个应用，又会提出更多的应用需求。我们目前仅仅是初步的应用，已经取得了令人兴奋的效果，接下来我们将对病人信息分析，医疗信息分析、绩效评估等方面进行更深入的研究和实践。

参考文献

- [1] 陈京民. 数据仓库与数据挖掘技术. 北京: 电子工业出版社, 2002, 95-96.
- [2] 张燕 高非 衡反修 医院数据仓库的设计和应用

地址: 广州市同福中路 396 号 邮编: 510220

E_mail: chenning3@hotmail.com 电话: 020-34403667