

数据仓库的开发实践

Development Practice of Date Warehouse



1 引言

当前,企业所面临的市场环境复杂多变,市场竞争日益激烈,顾客需求日趋个性化。对市场机会作出快速响应,是企业赢得竞争的关键之一。因此,企业经营管理者对市场变化应迅速作出科学的决策,而大部分企业已经建立的各种业务管理系统面向的是低层业务处理,对中高层决策的支持作用很小。因此,很有必要构建数据仓库(Data Warehouse),并以此为基础进行联机分析处理(OLAP: Online Analytical Process)和数据挖掘(DM: Data Mining),从而为企业经营管理者决策提供科学依据。

数据仓库是一种信息管理技术,它能够将分布在企业网络中的各种数据集成起来,为决策者提供各种有效的数据分析,从而起到决策支持的作用。美国的W.H.Inmon在《Building the Data Warehouse》一书中对数据仓库做了如下定义:“数据仓库是一个面向主题的、集成的、稳定的、随时间不断变化的数据集,它用于支持管理中的决策制定过程。”

2 某企业数据仓库 SKSDW 的构建

在某企业计算机集成制造系统(SKSCIMS)的第一期工程中,开发了各种业务管理系统。它

摘要: 本文以某企业数据仓库SKSDW的开发为实例,介绍了数据仓库的体系结构、开发模式,并以主题的确定、数据的抽取和加载为重点,详细阐述了数据仓库的构建。

关键词: 数据仓库 数据集市 主题 抽取 加载

们分布于供应、生产和销售等主要业务部门,收集了日常业务中所产生的基础数据。数据仓库 SKSDW 正是在这些全面、系统数据库系统的基础上构建的。

2.1 SKSDW 的体系结构

SKSDW 采用了如图 1 所示的体系结构:

2.2 SKSDW 的开发模式

在开发模式上, SKSDW 采用有反馈的自底向上模式(见图 2)。

如图 2 所示,考虑到一开始就构建一个完整的、大型的数据仓库,所需投入较多的人力、物力和资金,而且周期较长,见效较慢。而且,用户的需求在数据仓库的开发过程中,一般会发生变化,有必要将其反馈于开发过程中。因此, SKSDW 采用有反馈的自底向上的开发模式。首先在供

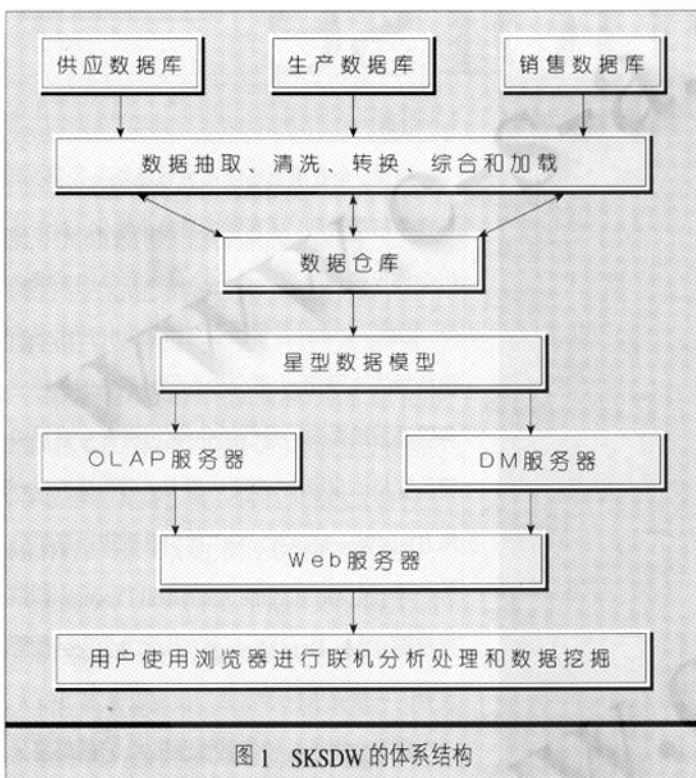


图 1 SKSDW 的体系结构

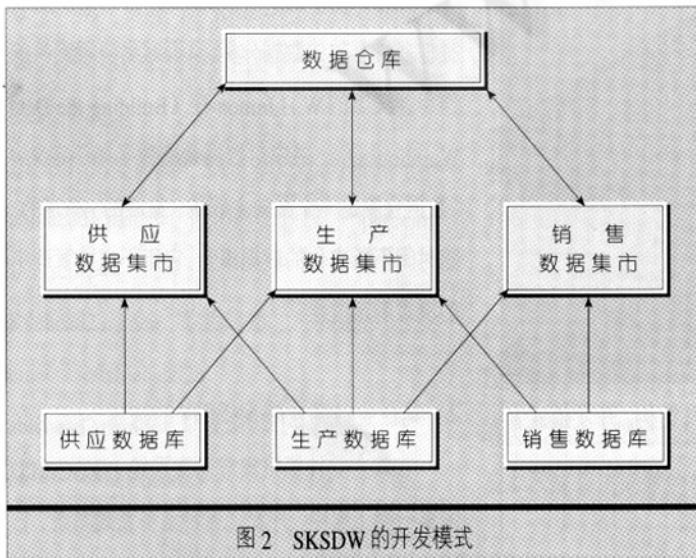


图 2 SKSDW 的开发模式

应、生产和销售部门数据库的基础上,构建部门数据集市,然后在数据集市的基础上,构建整个数据仓库。在构建数据集市的过程中,如果用户的需求发生变化,则对数据集市作出相应调整。因为数据集市规模较小,调整起来相对容易。用这种模式开发出来的数据仓库,能较好地满足用户的需求,投入使用后用户需求变化相对较少。

2.3 SKSDW 的开发过程

数据仓库的建设不是一蹴而就的,而是一项复杂而艰巨的工作。构建 SKSDW 时,根据 SKS 企业的具体情况,按以下步骤进行:

(1) 需求的收集和分析:构建数据集市和数据仓库前,先对供应、生产和销售这三个部门的主管和有关决策、分析人员进行访谈,收集并分析其需求,从而确定每一主题的主题域。

(2) 启动 SKSDW 工程:确定开发数据仓库的目标并制定工程计划。

(3) 建立技术环境:选择实现数据仓库的软硬件资源。SKSDW 以 Windows NT 为网络操作系统,以 Oracle 8 为数据库管理系统,以 IIS 为 Web 服务器,以 ASP 为数据仓库与 Web 服务器集成的实现技术。

(4) 确定主题进行数据建模:SKSDW 面向供应链管理,因此其主题确定为供应、生产和销售。供应主题描述了 SKS 企业的原材料采购、入库、出库、库存情况以及供应商信息,生产主题描述了原材料消耗量、产品产量等生产情况,销售主题描述了产品库存、销售情况以及客户信息。经分析,这三个主题包含的内容如下:

① 供应主题:供应商基本信息包括:供应商编码、供应商名称、国家名、省份名、城市名、电话号码等;原材料基本信息包括:原材料编码、原材料名称、原材料类别、计划价格等;原材料采购信息包括:原材料编码、供应商编码、实际采购价格、采购日期、采购数量等;原材料入库信息:原材料编码、入库单号、仓库编码、入库日期、入库数量等;原材料出库信息包括:原材料编码、出库单号、仓库编码、出库日期、出库数量等;原材料库存信息包括:原材料编码、仓库编码、日期、库存量等;

② 生产主题:产品基本信息包括:产品编码、产品名称、产品类别等;产品生产信息包括:产品编码、生产日期、生产车间、产量、原材料编码、消耗量等;

③ 销售主题:客户基本信息包括:客户编码、客户名称、国家名、省份名、城市名、电话号码等;销售合同信息包括:客户编码、合同编码、起始日期、终止日期、产品编码、销售数量、销售价格等;产品销售信息包括:产品编码、客户编码、合同编号、销售日期、付款方式、销售单价、销售数量等。

确定上述主题和主题域后,应设计数据模型,作为数据的组织方式。SKSDW 采用星型模式。如上述销售主题可设计成如下的星型模式:

(5) 选择数据源: SKSDW 的数据主要来自于原材料供应管理系统、生产管理系统和销售管理系统。由于这三个系统与其它业务管理系统之间或多或少发生数据交换或调用, SKSDW 中必然有一部分数据来自其它业务部门甚至企业外部, 如供应链上其他合作企业所提供的共享信息等。

(6) 设计数据仓库中的数据结构 数据仓库中的数据结构是在现有业务系统数据结构基础上, 针对管理信息的特征: 时间特性和汇总特性, 对数据的名称、类型、描述及关联等进行重新定义, 主要包括: 统一数据类型、调整数据长度和增加时间属性。

① 统一数据类型。同一数据有不同数据类型时, 必须统一为同一数据类型。例如日期字段, 有的系统定义为日期型数据, 有的定义为字符型数据。可以统一将其定为字符型数据。

② 调整数据长度。当同一数据的长度不一致时, 调整为统一数据长度。

③ 增加时间属性。时间性是数据仓库的重要特性之一。为此, 我们在数据仓库的表中设置日期字段。这样使得变量都带上了时间属性, 实现了数据的长期存放和信息的历史分析, 并可按时间维进行不同程度的汇总。

(7) 粒度的划分: 粒度反映了数据的详细程度和级别, 直接影响数据仓库的数据量以及所适合的查询类型。不同粒度级别的数据用于不同类型的分析处理。如: SKSDW 中有关销售主题的数据在产品维上按产品小类和 大类汇总, 在客户维上按客户代码汇总, 在时间维上按周、月、季、年进行汇总。为了提高查询速度, 可按产品维、客户维和时间维综合汇总, 生成导出数据, 以满足某些经常性的查询。

(8) 数据的抽取和加载: 在 SKSDW 中, 数据的抽取是通过视图来完成, 而抽取出来的数据的加载则是通过存储过程来实现。视图和存储过程都是在 Oracle 8 所提供的灵活而方便的 SQL*PLUS 工具中创建, 并在装有 Oracle 数据库管理系统的后台服务器上运行。下面将以仓库物资收发结存报表为例加以说明。

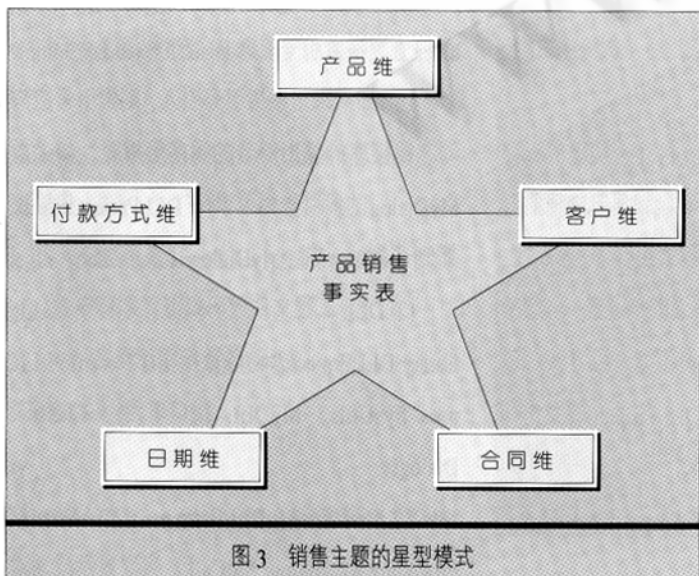


图3 销售主题的星型模式

首先, 创建名为 w_v_sfc 的视图, 从原材料供应管理系统中的 OLTP 数据库中提取数据。其 SQL 文件如下:

```
/* 产生本月的收发结存数量 */
```

```
create or replace view w_v_sfc (ny,dwbm,wzlb,wzbm,jhdj,yckcl,rkysl,
zkbrs,jnbrs,dbrksl,tlsj,zjsc,scwx,dx,jc,qt,llsl,dcksl,rssl,bfsl,bfje,
tzsl,tzje,yksl,ymkcl)
```

/* 创建的视图中包括年月、单位编码、物资类别、物资编码、计划单价、月初库存量、市场采购数量、调拨入库数量、直接生产消耗数量、生产维修消耗数量、大修消耗数量、技措和技改消耗数量、调拨出库数量、让售数量、报废数量、调整数量、盘点盈亏数量、月末库存量等信息, 这些信息对原材料采购供应部门作出采购决策和供应决策提供支持。在通过视图提取数据时, 一定要设置好时间条件限制。这主要是基于两方面的考虑: 一方面, 数据仓库中的数据本来就是历史数据, 其时间性很强; 另一方面, OLTP (联机事务处理: Online Transaction Process) 数据库中的数据量非常大, 若不加时间条件限制, 每次提取时都将所有与收发结存有关的数据统统提取出来并加以分组、排序、计算、统计等, 从而过多地消耗宝贵的系统资源, 且响应时间很长。

(8) 管理元数据: 元数据是关于数据的数据, 它描述了数据仓库的数据和环境, 在整个数据仓库中起着核心的作用。在 SKSDW 的初步实践中, 加强了对 Oracle 数据字典、数据仓库主题描述、数据逻辑模型定义以及视图、存储过程等数据处理规则的管理。

3 结语

SKSDW 的建立, 为供应、生产和销售等部门作出科学决策提供了现实可能性。但, 要真正实现这种决策支持功能, 还需在 SKSDW 的基础上, 运用 OLAP 和 DM 工具。OLAP 应支持用户进行多维分析, 如跨维、跨层次的计算, 切片、切块, 旋转分析等, 应该提供多视角查询、分析、预测和制作动态图的功能。而 DM 工具则利用数据仓库、方法库、模型库和知识库共同完成数据挖掘的过程, 从 SKSDW 中大量的分析型数据中挖掘出潜在的模式, 从而支持趋势分析和预测分析。目前, 在 SKSDW 的初步实践中, OLAP 工具和 DM 工具的运用, 正处于积极的探索当中。

参考文献

- 1 W.H.Inmon 著, Building the Data Warehouse, 北京机械工业出版社, 2000。
- 2 王珊等, 数据仓库技术与联机分析处理, 北京科学出版社, 1998。
- 3 Marc Demarest, Building the Data Mart, DBMS Magazine, 1999 7 (8)。