

电网生产管理数据仓库的设计与实现^①

张 平^{1,2}, 吕 立², 孙大嵬²

¹(中国科学院 研究生院, 北京 100049)

²(中国科学院 沈阳计算技术研究所, 沈阳 110168)

摘 要: 为了满足对生产管理进行决策分析的业务需求, 对某电网公司生产管理进行了需求分析, 对生产管理数据仓库进行了总体设计, 制定了数据从源系统到数据仓库中进行存储的 ETL 技术方案, 最后利用展现工具查询和分析数据, 实现决策支持, 为生产管理的决策提供了可靠的依据, 同时也验证了数据仓库的设计。

关键词: 电网; 生产管理; 数据仓库; 总体设计; ETL

Design of Grid Production Management Data Warehouse and Its Implementation

ZHANG Ping^{1,2}, LV Li², SUN Da-Wei²

¹(Graduate University, Chinese Academy of Sciences, Beijing 100049, China)

²(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

Abstract: In order to meet the business needs of decision analysis for production management. In this paper, a grid company conducted a needs analysis production management for the overall data warehouse design to develop ETL technology solutions for the data stored from source systems to a data warehouse, and finally make use of the display tools to query and analyze data so that achieve decision-making support, for production management decisions to provide a reliable basis, and also verify the data warehouse design.

Key words: grid; production management; data warehouse; general design; ETL(extraction-transformation-loading)

随着市场竞争的日益激烈, 某电网为实现国网建设“一强三优”(电网坚强、资产优良、服务优质、业绩优秀)的现代公司战略目标, 成功实施和应用了 ERP 系统, 在使得企业流程规范化的同时, 也使得企业中的管理系统处理的数据量快速增长, 电网管理者迫切需要从这些大量的数据中获取有用的信息, 用来支持决策, 而数据仓库能够满足电网管理者对数据分析的需要和提供决策支持。

本文对电网生产管理主题域进行了需求分析, 描述了电网生产管理数据仓库的主题设计、多维模型设计以及建模过程, 提出了 ETL 技术方案, 并对生产管理的数据进行了展现和分析。

1 数据仓库概述

数据仓库之父 William.H.Inmon 在 1991 年出版的

“Building the Data Warehouse”一书中提出和定义了数据仓库(Data Warehouse)的概念, 即数据仓库是一个面向主题的 (Subject Oriented)、集成的 (Integrated)、相对稳定的 (Non-Volatile)、反映历史变化 (Time Variant) 的数据集合, 用于支持管理决策(Decision Making Support)^[1]。它可以智能化地管理整个企业信息系统的海量数据信息并从中挖掘出最有价值的信息, 以向企业管理者提供强有力的决策支持, 从而使企业对市场反应更灵敏快捷, 整个企业也更具有生命力和竞争力。

2 数据仓库设计

2.1 需求分析

设计数据仓库成败的关键在于数据仓库的需求分析是否能够准确反映用户的决策支持需求。在进行需

① 收稿时间:2011-07-13;收到修改稿时间:2011-08-20

求分析时，需要定义访谈的参与者，并确定访谈的内容，如：用户最关心哪些业务，最需要掌握哪些业务信息，目前这些业务信息分别来自哪些部门，部门主管领导最关心哪些分析数据，部门的关键绩效指标又有哪些，这些指标通过哪些途径可以获得等。在此以电网生产管理为例，对其进行了需求分析。通过分析可以得出电网生产管理需要了解设备和项目大修的情况，需要对设备台帐、设备工单、设备维修费用、大修项目管理、设备生命周期等进行决策支持分析。根据电网生产管理的需求可以将其分为 3 个主题：维护工单分析、设备生命周期资产价值分析、设备信息分析。

2.2 设备信息分析主题设计

设备信息分析主题包含了所有类型设备的相关信息。基于该主题，用户能够对不同电压等级、处于不同状态以及不同种类的设备的相关信息进行分析，可以了解不同设备某一段时间在各个子公司的使用情况，也可以统计各类设备的台帐数以及购置费用等。因此，该主题可以满足用户对不同设备使用年限、不同设备的购置费用以及各种设备的台帐数进行分析等功能。

2.3 数据仓库多维模型设计

为了创建企业数据仓库，同时出于分析应用的目的，需要使用多维数据模型，OLTP（联机事务处理）这种规格化的数据使其无法直接用来支持复杂的特殊数据分析，需要将其转换为适合 OLAP（联机分析处理）使用的数据。根据设备信息分析这个主题的设计得出多维数据星型模式的数据表如表 1 所示：

表 1 设备信息分析主题数据表

域名	数据类型	类型长度
设备号	CHAR	18
成本中心编号	CHAR	10
设备投入使用日期	DATS	8
投运年限分类	NUMC	3
设备类型	CHAR	10
设备状态	CHAR	5
电压等级	CHAR	3
设备数量	INT4	4
设备购置值	CURR	8
设备投运年限	INT4	4

在该数据表中，设备号、成本中心编号、设备投

入使用日期、投运年限分类、设备类型、设备状态、电压等级分别为各自维度表的维度键，同时一起构成了数据表的主键，可以唯一识别数据表中的所有数据记录，即所有事实数据，而设备数量、设备购置值、设备投运年限是数据表的事实数据。数据表通过主键与各个维度表进行连接，维度表则通过维度键来查看该维度下的事实数据。与上述数据表相关的维度表如表 2 所示：

表 2 对应数据表的所有维度表

域名	数据类型	类型长度
设备维度表		
设备号	CHAR	18
成本中心维度表		
成本中心编号	CHAR	10
成本控制范围	CHAR	4
公司代码	CHAR	4
负责人员	CHAR	12
时间维度表		
设备投入使用日期	DATS	8
会计年度	NUMC	4
会计年度/期间	NUMC	7
投运年限分类维度表		
投运年限分类	NUMC	3
设备类型维度表		
设备类型	CHAR	10
设备状态维度表		
设备状态	CHAR	5
电压等级维度表		
电压等级	CHAR	3

2.4 数据仓库建模

在 SAP BW 中，所有数据存储结构的定义、ETL 流程的设计以及运行和监控都在同一个工作平台上完成，即数据仓库工作台。在数据仓库工作平台上建模首先需要定义信息对象，其次是定义和设计信息提供者（InfoProvider）。数据仓库中的信息对象是数据的最小存储单位，用来创建存储数据的结构和表，使信息能够以结构化的形式建模，可分为特性和关键值等，关键值提供要评估的值，如数量、金额等，特性则是用来分析关键值的业务参考对象。所谓的信息提供者是由多个信息对象组合起来的数据存储模型，包括标

准信息立方体 (InfoCube)、虚拟信息提供者 (VirtualProvider)、多个信息提供者 (MultiProvider) 以及数据存储对象 (DSO)。DSO 旨在存储事务级别的记录, 它只需引用所需的信息对象, 无需做额外的定义, 因为在 DSO 中数据以平面结构存储。标准信息立方体以星型模型设计的数据表和维度表为基础来组织相应的关键值型信息对象和特性型信息对象。

首先, 根据多维模型设计的数据表和维度表来定义各种信息对象。定义信息对象的过程如图 1 所示:

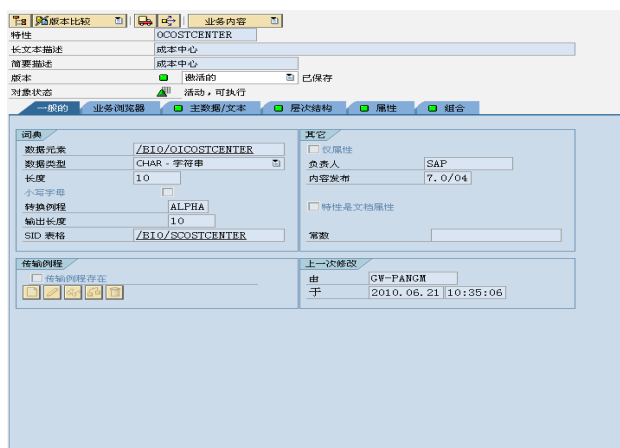


图 1 信息对象的定义界面

对于数据表中的事实数据设备数量、设备购置值、设备投运年限这三个字段, 分别将其定义为关键值型信息对象, 信息对象的类型和长度按照数据表中的对应字段的类型和长度进行设置。对于与数据表相关联的各个维度表, 将维度表中的各个字段定义为特性型信息对象, 它们的类型和长度也与维度表中对应的字段的类型和长度相同。除此之外, 由于公司之间具有从属关系, 所以公司代码信息对象应该带有层次结构, 因此在该信息对象界面选定“层次结构”选项卡, 在这里选定“带有层次结构”选项, 并将层次结构的属性设置为“与时间无关的层次结构”; 在定义成本中心编号信息对象时, 还需要在信息对象定义的界面的“属性”选项卡上将成本控制范围、公司代码这两个信息对象设置为该信息对象的导航属性, 这样这两个信息对象在执行查询时能够具备导航的功能, 对数据进行筛选, 而将负责人员信息对象设为描述属性, 以在查询需要时提供负责人员的信息。

其次, 按照数据分析的需要对定义好的信息对象进行组织, 即定义和设计信息提供者。数据仓库中的

数据存储分为三层, 底层为 PSA 层, 该层的数据来源于源系统, 不需要进行设计, 其它两层属于信息提供者数据存储模型, 依次为 DSO 层和 InfoCube 层。首先定义和设计 DSO 层, 在该层创建 DSO, 创建 DSO 的界面如图 2 所示:

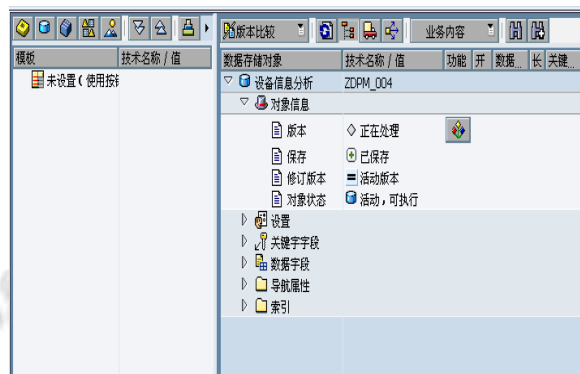


图 2 DSO 创建界面

在 DSO 界面上右击“关键字段”, 选择“信息对象直接输入”将设备号信息对象添加到“关键字段”中, 而将成本中心编号、设备投入使用日期、设备类型、设备状态、电压等级、设备数量、设备购置值、设备投运年限这些信息对象添加到“数据字段”中, 并将“导航属性”中的所有信息对象的导航开关选中, 以使这些信息对象在查询过程中具有导航的功能。设计完成后激活该 DSO。下一步则是定义和设计 InfoCube 层的信息立方体, 创建标准信息立方体的界面如图 3 所示:

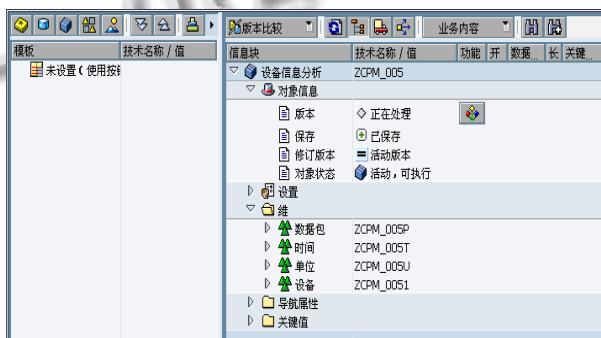


图 3 标准信息立方体创建界面

在定义了标准信息立方体的名称之后, 只需按照星型模型的设计结果在标准信息立方体定义界面上创建所需要的维度, 并将信息对象添加到“维度”和“关键值”中即可。在该标准信息立方体中定义了设备、成本中心、时间、设备投运年限分类等维度, 在每个维度

下添加相应的信息对象,如成本中心维度下,有成本中心编号、成本控制范围、公司代码三个信息对象,而在“关键值”中添加设备数量、设备购置值、设备投运年限三个信息对象。在该标准信息立方体中的信息对象与 DSO 中的信息对象的区别是其多拥有一个设备投运年限分类信息对象。设计完成后激活。至此,完成了数据仓库的建模,建立了数据仓库^[2-4]。

3 数据仓库实现

3.1 ETL 技术方案

ETL 的工作就是要将源系统中符合要求的数据抽取到 BW 系统中,进行转换和清洗,并最终存储到 BW 的信息提供者中。在 SAP BW 中利用信息包来将源数据抽取到 PSA 层进行预存储;在层与层之间创建一套转换规则,利用数据传输进程(DTP)执行转换,完成数据的转换和加载工作,从而完成整个 ETL 过程。因此,在数据仓库建模后的存储结构的基础上,通过设计信息包进程、转换以及数据传输进程来完成数据的 ETL 流程^[5]。下面以设备信息分析主题为例来设计 ETL 流程。

3.1.1 选择和抽取源数据

1) 选择源数据。设备信息分析所需要的数据为 SAP ERP 系统数据库中的数据,因此,只需选择 SAP ERP 系统作为与数据仓库连接的源系统。在选定源系统后还必须创建和选择数据源,数据源是逻辑上组合在一起的数据的集合,对于数据表中的数据需求,在设备信息汇总表中存储着关键值设备购置值的数据,记录着设备的投入使用日期,由该日期可以计算设备的投运年限,维度表所需要的数据一部分存储在设备信息汇总表中,如设备号、设备类型等,一部分则存储在设备工单表中,如成本中心。因此,分别基于设备信息汇总表和基于设备工单表创建数据源,并选择这两个数据源作为设备信息分析的数据的来源。

2) 抽取源数据。将选定的数据源中的数据抽取到 BW 的 PSA 层进行预存储,如图 4 所示:

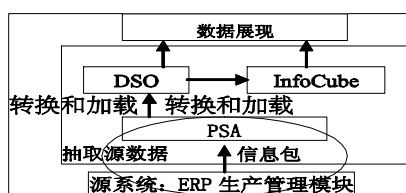


图 4 SAP BW 中的 ETL 流程图

从图中可以看出,这个过程是利用信息包来完成的,信息包的定义界面如图 5 所示:



图 5 信息包定义界面

信息包用来定义和管理数据从源系统到 PSA 的复制工作,能够对数据抽取进行具体的筛选和设置,定义更新的方式等,因此,在数据仓库工作平台上对选定的两个数据源分别创建信息包,由于是第一次对数据进行更新,所以选择更新方式为完全更新,设置完成后将这两个信息包分别激活,成功激活后数据就从源系统复制到 PSA 层中了。下面具体对两个数据源中的核心数据源设计数据的转换和加载。

3.1.2 建立各层之间的转换

转换是确定层与层之间信息对象的对应关系,这种对应关系遵循一定的规则,这些规则主要有直接分配、公式、例程、常数。直接分配是指将转换的源信息对象中的数据直接赋值给转换的目标信息对象,这说明源信息对象符合目标信息对象的要求;公式是指将源信息对象经过一些公式运算后赋值给目标信息对象;例程则是通过运用 ABAP 语言对源信息对象进行特定的更改后赋值给目标信息对象;常数是指信息对象自身的值是固定不变的。

1) 建立 PSA 层到 DSO 层之间的数据转换。定义转换规则的界面如图 6 所示。

对于 PSA 层的信息对象,设备号、成本中心编号、设备类型、设备状态、电压等级和设备购置值的数据类型和性质在 DSO 层中都能够找到与其一一对应的完全符合的信息对象,因此,在规则类型中选择“直接分配”即可。DSO 中的信息对象“设备数量”是对设备的计数,而 DSO 中的关键字段是设备号,惟一确定数据字段中设备数量的值,因此将其设为常数 1。对于 PSA 中的信息对象设备投入使用日期,需要进行两种

转换,一种是直接分配给 DSO 中的信息对象“设备投入使用日期”,一种则是通过使用公式转换后赋值给 DSO 的“投入使用年限”信息对象,使用的公式为: DATE_YEAR(当前日期)-DATE_YEAR(INBDT))+1,其中, INBDT 是设备投入使用日期的技术名称,该公式通过计算当前年与设备投入使用年份的差值得出设备的使用的年数。至此,完成了 PSA 层到 DSO 层的转换的创建。

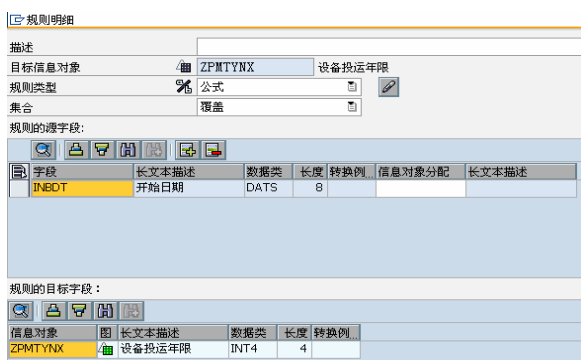


图 6 转换规则定义界面

2) 建立 DSO 层到 InfoCube 层的转换。InfoCube 层与 DSO 层中的信息对象的区别在于 InfoCube 层中多了一个“投运年限分类”信息对象,该信息对象用于标识设备使用时间的长短,将使用时间分为三个等级,分别为 005、010、015,代表使用年限在五年内、十年内和十五年内,该信息对象的值能够通过 DSO 层的投入使用年限信息对象进行转换得到,由于这种转换涉及到逻辑问题,只能通过例程这种方式来完成转换,该例程的 ABAP 代码如下所示:

```
METHOD compute_ZPMNXFL.  
DATA:  
MONITOR_REC???TYPE?rsmonitor.  
if (SOURCE_FIELDS-/BIC/ZPMTYNX>0and  
SOURCE_FIELDS-/BIC/ZPMTYNX<6)  
RESULT='005'.  
elseif(SOURCE_FIELDS-/BIC/ZPMTYNX> 5  
and SOURCE_FIELDS-/BIC/ZPMTYNX<11).  
RESULT='010'.  
elseif(SOURCE_FIELDS-/BIC/ZPMTYNX>11).  
RESULT='015'.  
endif.  
ENDMETHOD.
```

这段代码中将所有使用年限大于 10 年的设备统一归类为 015,说明设备的使用年限在十五年内,其中的 ZPMTYNX 是 DSO 层中的投入使用年限信息对象的技术名称。DSO 层中的其余信息对象在 InfoCube 层中都有与其可以一一对应的信息对象,因此,只需将 DSO 层中的其它信息对象通过直接分配的转换方式赋值给它们所对应的 InfoCube 层中的对象即可。至此,完成了从 DSO 层到 InfoCube 层的转换的创建。

3.1.3 建立各层之间的数据传输进程

数据传输进程(DTP)可控制特定转换的实际数据流,用于执行转换,它包含提取模式(增量或完全)、过滤条件,以及错误处理相关的参数的各项设置。从 PSA 层到 DSO 层,从 DSO 层到 InfoCube 层,以及直接从 PSA 层到 InfoCube 层,都需要相应地建立独立的数据传输进程。这样设计的好处在于用户可以全面地掌握和控制数据流在 BW 各层之间的传输,当发现数据传输错误时,用户可以逐层检查 DTP,定位故障点;每个 DTP 都能够独立执行一套转换规则,能够帮助用户更灵活地对数据进行处理;每个 DTP 都可以进行过滤条件、错误处理机制的设置等,如果某一层的数据加载出现了错误,用户可以直接就这一层进行数据的重读,而不会影响到整个数据流,从而减小了错误影响的范围同时也加快了消除错误的过程。

1) 创建从 PSA 层到 DSO 层的数据传输进程(DTP)。将这两层之间的 DTP 的类型设置为实时数据获得的数据传输协议,这样数据可以实时地从 PSA 层更新到 DSO 层,以满足用户对事务级别数据的实时查询和分析功能;创建 DTP 的界面如图 7 所示:

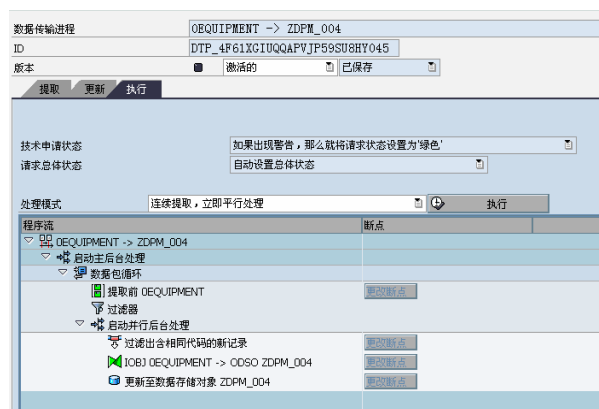


图 7 DTP 定义界面

在“提取”选项卡上,选择提取模式为 Delta,即增

