

文章编号:1006-5911(2010)02-0423-08

基于模型驱动体系架构的 OLAP 开发方法

宋旭东^{1,2}, 胡墨谦¹, 刘晓冰²

(1. 大连交通大学 软件学院, 辽宁 大连 116028; 2. 大连理工大学 CIMS 中心, 辽宁 大连 116024)

摘 要:传统的联机分析处理开发方法难以很好满足企业的决策需求,复杂性和高效性是 OLAP 系统开发面临的两大主要问题。为此,在数据仓库系统统一集成建模框架下,提出了一种新的基于模型驱动体系架构的联机分析处理开发方法。使用 UML Profile 和公共仓库元模型规范,建立了基于模型驱动体系架构的联机分析处理平台无关模型元模型和联机分析处理平台相关模型元模型,并采用查询/视图/转换规范的 Relation 语言建立了一套平台无关模型元模型到平台相关模型元模型的转换规则,在模型转换工具 MediniQVT 和 IBM 的 EMF 框架下,实现了平台无关模型到平台相关模型的转换,通过自定义的从平台相关模型实例到结构化查询语言转换的 Java 类生成联机分析处理结构化查询语言代码,实现了模型驱动的联机分析处理软件开发过程。最后,通过某企业决策信息系统实际应用,验证了该方法的可行性和高效性。

关键词:数据仓库;联机分析处理;模型驱动体系架构;查询/视图/转换

中图分类号:TP311 **文献标识码:**A

On-line analysis processing development method based on model driven architecture

SONG Xu-dong^{1,2}, HU Mo-qian¹, LIU Xiao-bing²

(1. Software School, Dalian Jiaotong University, Dalian 116028, China;

2. CIMS Center, Dalian University of Technology, Dalian 116024, China)

Abstract: Traditional On-Line Analysis Processing (OLAP) development method was difficult to offer suitable solutions for enterprises' decision demands. Complexity and efficiency were the major problems concerning the development of OLAP system. To tackle these problems, a new OLAP development method based on Model Driven Architecture (MDA) in an integrated Data Warehouse (DW) modeling framework was proposed. By using UML Profile and Common Warehouse Metamodel (CWM), OLAP Platform Independent Model (PIM) Meta-model and OLAP Platform Specific Model (PSM) Meta-model were constructed, a set of transformation rules from OLAP PIM meta-model to OLAP PSM meta-model by using the Query/ View/ Transformation (QVT) relation language was proposed, transformation between PIM model and PSM model under the model transformation tool MediniQVT and IBM Eclipse Modeling Framework (EMF) was implemented, OLAP Structured Query Language (SQL) code by using user-defined transformation Java class between PSM model and SQL code was generated so that model-driven OLAP development was accomplished. Finally, this method was applied in practice and proved to be feasible and effective.

Key words: data warehouse; on-line analysis processing; model driven architecture; query/ view/ transformation

收稿日期:2009-03-11;修订日期:2009-08-31。Received 11 Mar. 2009;accepted 31 Aug. 2009.

基金项目:国家自然科学基金资助项目(70572098)。**Foundation item:** Project supported by the National Natural Science Foundation, China (No. 70572098).

作者简介:宋旭东(1969-),男,辽宁本溪人,大连交通大学软件学院副教授,大连理工大学 CIMS 中心博士研究生,主要从事软件开发方法和智能技术应用的研究。E-mail: xudongsong@126.com。

0 引言

目前,基于数据仓库(Data Warehouse,DW)进行决策分析主要有联机分析处理(On-Line Analysis Processing,OLAP)和数据挖掘两种方式。OLAP 是一种归纳型的决策过程,包括切片、切块、旋转、上卷、下钻等基本操作,其主要作用是使企业的决策者能灵活地操纵企业的数据,以多维的形式从多方面和多角度来观察企业的状态,了解企业的变化,通过快速、一致、交互地访问各种可能的信息视图,帮助管理人员掌握数据中的规律,实现对数据的归纳、分析和处理,帮助组织完成相关的决策。OLAP 具备了强大的决策支持能力,在决策支持系统中具有重要的地位。

虽然人们在 OLAP 方面开展了大量的研究工作,但主要还集中在多维数据模型^[1-3]和 OLAP 操作的有效算法^[4-7]等方面,对面向最终决策用户的 OLAP 应用开发未引起足够的重视,基于数据仓库的 OLAP 应用开发的研究还很不够。

OLAP 应用系统开发是一项复杂性工作,系统的可用性和可维护性是构建 OLAP 系统面临的主要问题。目前多数 OLAP 的应用开发过于依赖具体业务的表结构(行、列)和数据仓库的形式^[8],缺少一个直观的、集成的和统一标准规范的开发框架,因此很难借用以往的 OLAP 建模和设计成果,OLAP 与数据仓库系统其他层之间的互操作性和集成性难以保证。Jesus Pardilo 等人^[9]给出了一种模型驱动的 OLAP 元数据自动获取方法,该工作仅限于 OLAP 立方体,并没有给出 OLAP 操作模型表示。笔者在前期工作中提出了应用 UML Profile 设计 OLAP 元模型的方法^[10],实现了 OLAP 在概念层上的设计,提供了一个具备通用性和重用性的 OLAP 元模型,但没有给出这些概念模型到逻辑模型,乃至物理实现的具体转换方法,没有提供基于模型转换的 OLAP 实现。

另一方面,模型驱动体系架构(Model Driven Architecture,MDA)是对象管理组织(Object Management Group,OMG)发布的一个软件开发框架^[11],它以模型作为软件开发的核心元素,并为其提供了多种技术规范作为基础框架。MDA 的核心思想是抽象出与实现技术无关、完整描述系统的平台无关模型(Platform Independent Model,PIM),针对不同实现技术制定变换定义;通过制定映射规则,将 PIM 转换成与具体实现技术相关的平台相关

模型(Platform Specific Model,PSM);最后,再通过转换工具将 PSM 自动转换成代码。MDA 最大的好处在于可以节省软件开发的时间和精力,提高开发效率,目前在数据仓库系统开发中有一定应用。Jose Norberto Mazon 等人^[12]提出了基于 MDA 的数据仓库开发方法,给出了数据仓库 PIM,PSM 等相关模型及具体转换方法,并主要论述了数据仓库资源层的开发。Leopoldo Zepeda 等人^[13]为数据仓库系统概念结构设计提出一种混合驱动的方法,应用 MDA 方法完成了实体关系(ER)模式到 OLAP 模式的映射,他们的工作仅限于数据仓库系统的概念结构设计,没有给出 OLAP 操作模型表示及实现方法。Octavio Glorio 等人^[14]应用 MDA 方法为支持地理 OLAP 查询实施了地理数据仓库系统开发,但没有给出 OLAP 应用开发的实现方法。Jesus Pardillo 等人^[15]提出了基于 MDA 的数据挖掘建模方法,给出了面向数据仓库的数据挖掘概念模型及相应模型转换框架。笔者提出了基于 MDA 的决策支持系统开发框架^[16]。然而,将 MDA 方法应用到 OLAP 开发中的研究报道并不多见,唯一的一项相关研究是 Jesus Pardillo 等人^[17]提出采用对象约束语言(Object Constraint Language,OCL)为 OLAP 操作实施平台无关概念建模的方法,同时给出 MDA 模型转换框架,但没有给出 OLAP PIM 元模型及 OLAP PSM 元模型的定义,没有给出模型映射规则等关键技术描述。

鉴于此,本文在数据仓库系统统一建模框架下,从模型驱动的角度实施 OLAP 应用开发,提出了基于 MDA 架构的 OLAP 应用开发方法,该方法构建了基于 UML Profile 和公共仓库元模型(Common Warehouse Metamodel,CWM)的 OLAP PIM 元模型和 OLAP PSM 元模型,定义了一套基于查询/视图/转换(Query/View/Transformation,QVT)规范的 PIM 元模型到 PSM 元模型转换的规则。在模型工具 MediniQVT 及 IBM 的 Eclipse 模型框架(Eclipse Modeling Framework,EMF)下实现了 PIM 模型到 PSM 模型,直至结构化查询语言(Structured Query Language,SQL)代码的转换,从而实现了基于模型驱动的 OLAP 开发。这种方法一方面在统一数据仓库框架下能很好解决集成和互操作问题,模型表示简洁,直接面向最终用户需求。采用 MDA 规范体系中的标准定义各层模型以及模型之间的转换规则,保证了模型和转换规则描

述的准确性和通用性;另一方面,可以让 OLAP 较早地伴随数据库系统进入设计阶段,开发的重点在 OLAP 概念层 PIM 模型的设计,而不必关心概念层 PIM 模型到最终 SQL 实施代码的转换,进而减少开发的时间和代价,提高 OLAP 应用开发效率,为决策支持系统实施提供有力的依据。

1 基于模型驱动体系架构的联机分析处理集成开发框架

引入模型驱动体系架构的思想,将 OLAP 的开发纳入整个数据库系统的统一集成开发框架下,提出了基于 MDA 的 OLAP 集成开发框架,如图 1 所示。

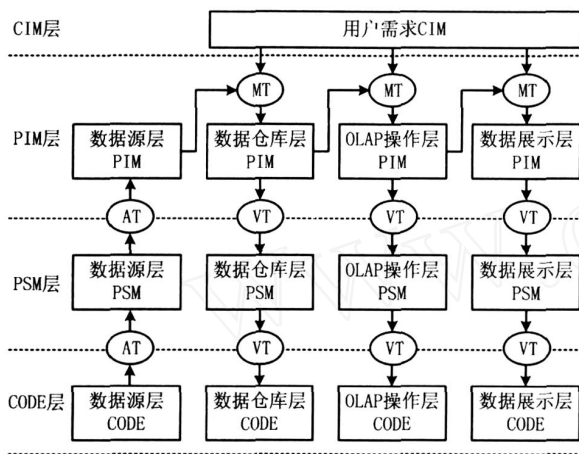


图1 基于MDA的OLAP集成开发框架

从图1可以看出,此框架的纵向是 OLAP 系统的每一个层次,横向是 MDA 开发的几个视图,每个 OLAP 系统层都是根据这三个不同的视图进行设计的。对于每一层次的每个视图,都有不同的模型与之相对应。为了获得目标模型,此框架提供了三类转换:

(1)垂直转换 指源模型和目标模型处在不同的抽象层上,它又可以分为正向垂直转换(VT)和逆向垂直转换(AT)。例如,开发出 PIM 后,就可以通过运用相应的转换得到其余的模型,直到得到 OLAP 系统的最终实现,这些是垂直的转换,PIM 到 PSM 是正向垂直转换,PSM 到 PIM 是逆向垂直转换。

(2)水平转换 指源模型和目标模型处在同一个抽象层次上的转换。如数据仓库层的 PIM 模型,可以用在 OLAP 操作的 PIM 模型制定上。

(3)合并转换(MT) 多个模型转换为一个目标模型。

下面对每一个层次的转换进行详细说明:

(1)数据源层 PIM 表示数据源的概念模型,此 PIM 只有在逻辑模式或数据结构有效的情况下才能获得,因此其转换是从 CODE 层或数据源层的 PSM 模型获得 PIM,转换类型设定为 AT。

(2)数据仓库层 PIM 表示数据仓库或者数据集市的概念模型,这一层的主要目标是在用户需求的主题下,针对数据源层 PIM,而不用考虑具体的平台技术,由 CIM 和数据源层 PIM 通过合并转换获得。它的 PSM 层和 CODE 层可借助数据仓库层 PIM 通过垂直转换得到。

(3)OLAP 操作 PIM 表示 OLAP 操作的概念模型,如切片、切块建模。它的 PIM 模型由数据仓库层 PIM 和 CIM 通过合并转换获得,其他层可借助 OLAP 操作层 PIM 通过垂直转换得到。

(4)数据展示 PIM 表示终端用户查看分析结果数据显示形式的概念模型,它的 PIM 模型可通过定制计算机集成制造(Computer Integrated Manufacturing, CIM)和 OLAP 操作层 PIM 的合并转换获得,其他层可借助数据展示层 PIM 通过垂直转换得到。

集成开发框架的优点:

(1)高效性 PSM 可以从 PIM 自动生成,提高了软件开发效率。同时 OLAP 系统的开发者可以把工作重点和注意力从逻辑和技术细节转移到 CIM 和 PIM,更多关注解决业务领域的问题,进而可以提升 OLAP 系统的质量。

(2)可移植性 同样的 PIM 可以转换为多个基于不同技术的 PSM。PIM 层对于不同目标平台是可移植的。

(3)可复用性 转换代表可重复的设计思路,所以一个转换开发后,可以把它运用到多个不同的 PIM 的实例中。

(4)适应性 对于新的需求,可以通过调整 CIM 和 PIM 来表述,这些变换会自动映射到 PSM 和 OLAP 系统中。对于新技术的出现,无需改变 PIM 模型,可以仅关注 PIM 和新的 PSM 之间的转换。

在该框架下,基于 MDA 的 OLAP 模型转换如图 2 所示。OLAP PIM 模型可以对应多个 OLAP PSM 模型。OLAP PSM 模型和特定技术相关,可以是关系型的或多维型的。如果采用 PSM 为关系型的,则 OLAP 的这些基本操作可以认为是一种扩展了的 SQL 操作(SQL 代码)。由于 SQL 语句与具体的工具相关,会有不同版的 SQL 代码,如 Oracle,SQL Server,DB2 等。

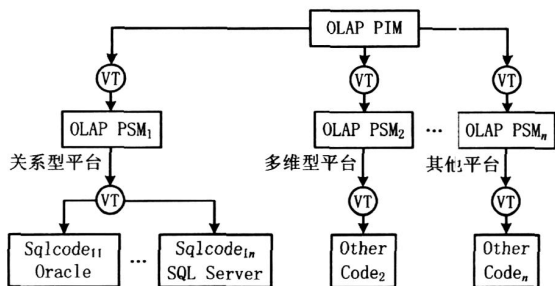


图2 基于MDA的OLAP模型转换

2 基于模型驱动体系架构的联机分析处理开发

模型转换是基于 MDA 的 OLAP 开发的核心内容,常用的模型转换方法有标记转换法、元模型转换法、模型转换法、模式转换法和模型合并法^[11]。

由于元模型转换法是基于元对象机制,表达模型更具有通用性,采用元模型转换法实施基于 MDA 的 OLAP 开发。OLAP 操作是 OLAP 的核心内容,以下重点研究基于 MDA 的 OLAP 操作模型开发。

2.1 联机分析处理模型开发过程

笔者提出的基于 MDA 的 OLAP

模型开发过程如图 3 所示,首先利用 UML Profile 建立 OLAP PIM 元模型;然后利用简化后的 CWM 设计 OLAP PSM 元模型;再利用基于 QVT 的 Relation 语言制定 OLAP PIM 元模型到 OLAP PSM 元模型的转换规则,最终实现 OLAP PIM 模型实例到 OLAP PSM 模型实例的自动转换;最后利用 IBM 的 EMF 框架,通过自行定义的转换规则实现从 OLAP PSM 模型实例到 SQL 的自动转换。

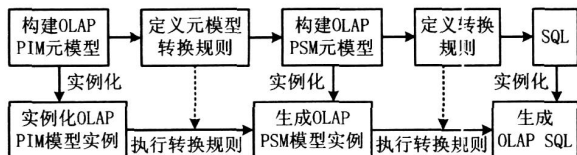


图3 基于MDA的OLAP模型开发过程

2.2 联机分析处理平台无关模型元模型定义

统一建模语言(Unified Modeling Language, UML)不能完整准确地对 PIM 模型进行描述,所以本文采用 UML Profile 中的版类(Stereotype)设计 OLAP PIM 元模型,利用其扩展机制保证对 PIM 描述的准确性,其元模型结构如图 4 所示。其中主

要包括 OLAP 操作模型(OLAP Operations),它是各种操作的描述模型,根据不同的 OLAP 操作,操作模型包括切片模型、切块模型、上卷模型、下钻模型和旋转模型五种。其他相关模型是事实模型(fact model)和目标模型(target model)。事实模型是 OLAP 各个基本操作的实体对象,包含分析的各个维(DIM)和分析对象(FA);目标模型则是在事实模型的基础之上满足用户要求的一个新的数据聚合或划分的实体。目标模型包含了各个 OLAP 基本操作所形成的条件标签({CONSTRAINTS}),以表示目标模型(«TargetModel»)较事实模型(«FactModel»)所产生的变化。通过 OLAP 操作模型、事实模型和目标模型,可以很明确地表示一个 OLAP 操作过程。OLAP PIM 元模型各版类继承于 UML 原有的版类,其层次结构如图 5 所示。

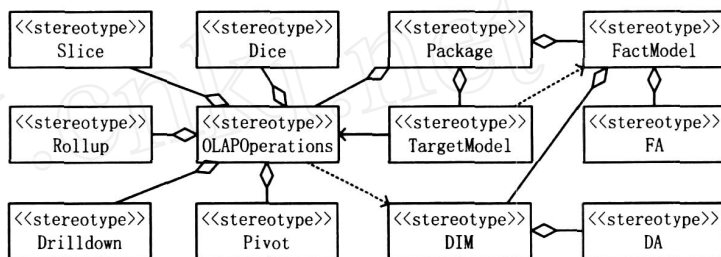


图4 OLAP PIM元模型结构

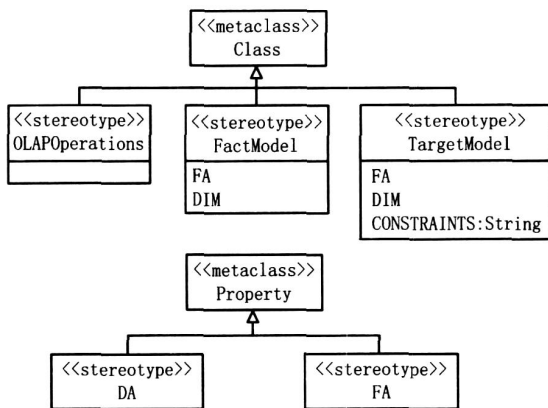


图5 OLAP PIM元模型版类层次结构

OLAP 操作模型详细描述如图 6 所示。

(1) 切片模型(«Slice»)

切片是对一个维的某个属性的值的确定,以获得满足这一属性值的事实属性的值的分布。切片模型需要一个进行切片操作的维属性的值({A})。

(2) 切块模型(«Dice»)

切块是对一个维的某个属性满足一个区间值的确定,该区间根据维属性的数据类型,可能是一个连

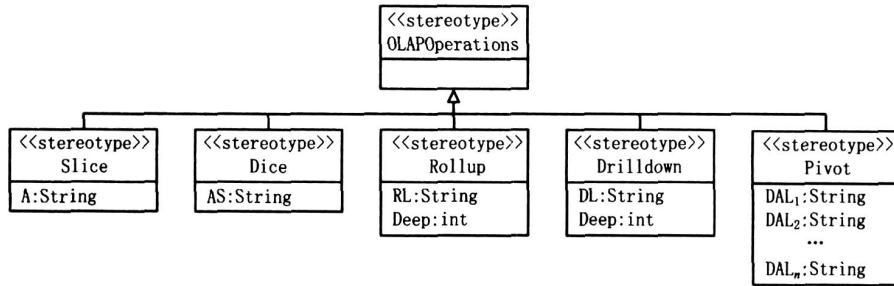


图6 OLAP操作模型

续的集合,或者是一个离散的集合。这个集合称为切块集({AS})。

(3) 上卷模型(<<Rollup>>)

上卷是在一个维中,对具有上下层级关系的属性,从下至上的事实属性的聚合,而聚合的程度依赖于维中具有层级关系的维属性。所以,上卷模型(<<Rollup>>)包括上卷表({RL})和上卷深度(Deep:int)。其中,上卷表({RL})应当包括所能聚合的各个层次的维属性,并按照聚合程度能力从小到大排列;而上卷深度(Deep)表示信息需要聚合到的级数。例如,对一个企业的销售量在时间维上进行上卷操作,上卷表中应当有日、月、年,并且需要获得年份的销售量统计情况,上卷深度为3。

(4) 下钻模型(<<Drilldown>>)

下钻可以看作是上卷的一个逆操作,因此下钻模型应该具备和上卷模型相似的结构,即包含一个下钻表({DL})和下钻深度(Deep:int)。其中,下钻表包含所能划分的各个层次的维属性,并按照划分详细程度从小到大排列;下钻深度(Deep)表示信息需要划分到的级数。使用上卷模型所用的例子,下钻表中应当有年、月、日,并且需要获得月份的销售量统计情况,下钻深度为2。

(5) 旋转模型(<<Pivot>>)

旋转即是改变维度的位置关系,从而得到不同视角的数据。这个维度位置关系的变化可以是不同维度之间的,也可以是同维度下的,但这些变化都是维属性的位置变化。旋转模型则具有相互改变的维属性({DAL₁},{DAL₂},{DAL₃},…,{DAL_n}),其顺序表示改变后的次序。基于数据可视化目的,使用表格展示数据时,可用多个表格来表示超过两个维度的情况;使用数据立方体时,通过多个立方体来表示超过三个维度的情况。

2.3 联机分析处理平台相关模型元模型定义

在OLAP开发模型中,一个PIM模型可以转

换为多个PSM模型。PSM模型往往是针对一种具体的数据库来设计的。目前普遍采用的是关系型数据库,所以本文的PSM模型设定为基于关系型数据库。由于CWM规范中所规定的OLAP元模型结构较为复杂,而本文进行转换后的PSM模型旨在最终转换成扩展的SQL操作,并对其进行合理的简化,以降低模型的理解难度。在保证对OLAP各个操作描述完整性和正确性的同时,对CWM中OLAP元模型进行简化,得到简化后的OLAP PSM元模型结构,如图7所示。

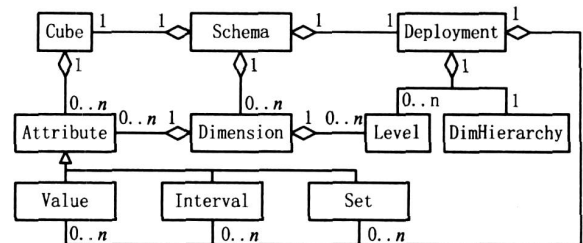


图7 OLAP PSM元模型结构

图中:Schema表示一个包含了所有OLAP模型元素的逻辑容器,它是一个指向OLAP模型的根元素;Cube表示具有多个分析角度的多个分析对象的聚合体;Dimension表示观察事实数据的角度;维包含了一般属性(Attribute)和具备层次的属性(Level),通过设置层次的值来实现维的一个结构;Deployment表示对Cube和Dimension进行的部署,以描述OLAP的各种基本操作;DimHierarchy表示对Cube进行观察的维的顺序结构,不同的顺序结构得到的分析数据自然也不一样;Level表示具备层次性的属性,以实现对分析对象的上卷和下钻;Attribute表示一般的属性,如维的属性或事实数据属性;Value表示一个属性取一个特殊值时的状态,以实现对分析对象的切片;Interval表示一个属性取线性值时的状态,以实现对分析对象的切块,表示线性取值;Set表示一个属性取离散值时的状态。

2.4 平台无关模型到平台相关模型的转换

MDA 中最主要的一个部分就是 PIM 到 PSM 的转换。由 OMG 所提出的 QVT 是目前较为完善的模型转换语言,在 QVT 规范中,定义了数据模型和数据仓库表结构的转换规则,说明该规范在数据仓库中应用的可靠性。采用 QVT 实施 PIM 到 PSM 的转换符合 MDA 规范体系的要求,使得模型具备集成性和重用性。QVT 可以将模型转换规则描述为一组源模型元素与目标模型元素之间的关系 (Relation)。每个 Relation 包含几个 domain,这些 domain 指向参与转换的模型元素,每个 domain 包含一组 pattern,通过匹配 pattern 可以选择出符合转换规则的模型元素。如果 pattern 不能匹配,则通过修改、删除、创建目标模型元素来达到匹配,同时完成模型转换。Relation 是一种规则型的、声明性的模型转换语言,它的语义涵盖了几乎所有规则型或者声明性模型转换语言的语义^[18]。本文根据 Relation 语言制定由 OLAP 的 PIM 元模型到 PSM 元模型的转换规则,它们分别是 Package To Schema, Model To Cube, FA To Attr, DIM To Dimension, DA To DimAttr, DA To LevelAttr, Dim To DimId, Operation To Deployment, Slice To Value, Dice To LOV, Dice To Interval, Drilldown To Level, Rollup To Level, Pivot To DimHierarchy。

Package To Schema 表示一个整体的 PIM 到 PSM 的转换,Package 和 Schema 都是 PIM 和 PSM 对各自模型的整体性描述,包含各模型整体结构的说明,它通过后置条件 (where 子句) 来触发关系 Model To Cube, Operation To Deployment。Model To Cube 是将 PIM 中事实模型和目标模型转换成 PSM 中 Cube 的规则,其后置条件触发关系 FA To Attr, DIM To Dimension, Dim To DimId 分别实现事实属性到属性转换、维属性的转换,以及维的转换。DA To DimAttr 和 DA To LevelAttr 分别表示 PIM 中维属性到 PSM 中一般维属性和层次维属性的转换,这两个关系都是关系 DIM To Dimension 的后置条件,依赖它来触发实现。另外,这两个关系都使用 level 的取值作为前置条件,以判断 PIM 中的 DA 进行何种转换。Operation To Deployment 表示将 PIM 中的操作模型转换成 PSM 中的 Deployment。其六个后置条件实现了 OLAP 基本操作从 PIM 到 PSM 的转换。Slice To Value 通过设置一个属性的值,并指定该属性所从属的维,实现切片模型的转换。Dice To LOV 和 Dice To Interval 在判断 is-Continuous 作为前置条件的条件下,分别实现了离

散取值和线性取值情况的切块模型转换。Drill-down To Level 和 Rollup To Level 分别实现了下钻模型和上卷模型的转换。Pivot To DimHierarchy 将旋转模型转换到 DimHierarchy,通过对重定序转换的描述实现旋转操作的转换。

下面以 PIM 中的操作模型转换为 PSM 中的 Deployment 模型为例,说明基于 QVT 的多维数据 PIM 到 PSM 模型转换的转换规则。

```
Relation OperationToDeployment {
    pn:String;
    checkonly domain pim p:OLAP_PIM::Package{
        olapOperations = oper:OLAP_PIM::OLAPOperations{
            name = pn
        }
    };
    enforce domain psms:OLAP_PSM::Schema{
        deployment = dp:OLAP_PSM::Deployment{
            name = pn
        }
    };
    where{
        SliceToValue(oper,dp);
        DiceToLOV(oper,dp);
        DiceToInterval(oper,dp);
        DrilldownToLevel(oper,dp);
        RollupToLevel(oper,dp);
        PivotToDimHierarchy(oper,dp);
    }
}
```

总映射分为 Package 和 Schema 两个域,表示发生映射的两端。PIM 中的类 olapOperations 属于 Package 中的一个引用,要转换到 PSM 中的 deployment,deployment 同时属于 Schema 的一个引用。这个 name = pn 的 olapOperations 必须与要转换成的 deployment 的 name 相匹配。

Slice To Value (oper, dp), Dice To LOV (oper, dp), Dice To Interval (oper, dp), Drilldown To Level (oper, dp), Rollup To Level (oper, dp), Pivot To DimHierarchy (oper, dp) 这几条语句定义在 Where 子句中,用来触发类 olapOperations 中的各类基本操作到 deployment 中所匹配的属性的映射。

2.5 平台相关模型到结构化查询语言代码的转换

IBM 的 EMF 是基于 Eclipse 的模型框架,它可以将模型转换成高效的、正确的和易于定制的 Java 代码。采用 EMF 可以建立 PSM 元模型中各个模型元素相对应的 Java 类,再通过自定义的从 PSM 模型实例到 SQL 转换的 Java 类,实现 SQL 代码的自动生成。

在 EMF 框架下,定义了用于实现 PSM 到 SQL

转换的两个Java类:pretreatment和convert。其中预处理类pretreatment负责读取由PIM到PSM转换所生成的PSM模型实例,并将其转换为相对应的Java对象实例;转换类convert负责把生成的对象实例转换成相对应的SQL语句,实现SQL代码的自动生成。

2.6 基于模型驱动体系架构的联机分析处理模型实现

选择支持QVT规范的MediniQVT开源模型

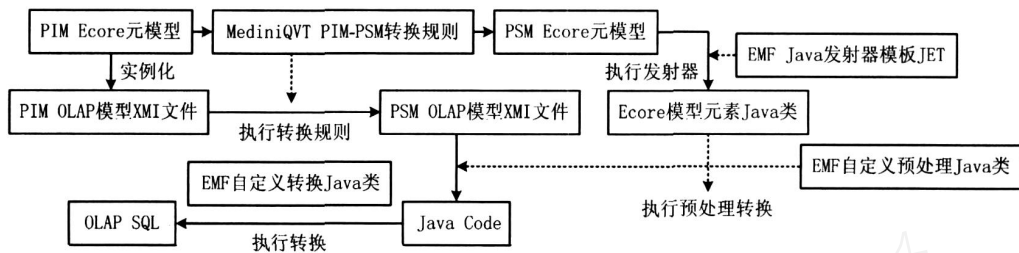


图8 基于MDA的OLAP模型实现

3 实例应用

目前,在东北特钢集团基于数据仓库的决策信息系统的项目中,已经成功地应用了本文提出的基于MDA方法来构建高层OLAP操作,收到了很好的效果,下面通过一个具体实例进行描述。企业集团战略管理处定期对企业各生产分厂进行考核,要实时统计分析销售合同执行情况,要求决策信息系统提供实时查询提前入库合同的能力。为此,针对合同跟踪事实表,首先进行提前入库合同OLAP PIM概念建模,并通过基于MDA的模型转换生成相应的PSM模型,直至生成SQL代码。

在本实例中,提前入库OLAP操作中仅涉及切块操作,所以Operation To Deployment中只有Dice-ToLOV关系会在转换过程发生作用,而且Dice-ToLOV关系会将Dice模型中的对入库日期区间(2008-12-28~2008-12-31)及提前入库(实际入库年月比合同中计划排产年月早的合同)的描述转换为在Deployment中的描述。利用支持QVT规范的MediniQVT开源模型转换工具,实现了OLAP操作元模型的定义和提前入库OLAP操作模型转换。具体实例中定义了提前入库OLAP PIM模型实例(source.xmi),利用模型转换生成OLAP PSM模型实例(target.xmi)。同时在IBM EMF框架下,使用自定义的类pretreatment.Java读取所生成的PSM模型实例(target.xmi),再使用类convert.Java将

转换工具,实现了OLAP元模型及其转换规则的定义,元模型的定义通过创建EMF的Ecore文件完成,元模型转换规则通过定义QVT文件完成,模型实例通过定义XML元数据交换(XML Metadata Interchange, XMI)文件完成。图8给出了在MediniQVT及EMF框架下基于MDA的OLAP模型实现过程。首先进行OLAP PIM概念建模,并通过基于MDA的模型转换生成相应的OLAP PSM模型,直至OLAP SQL代码生成。

其转换成相对应的提前入库合同查询SQL语句。最后,将生成的SQL代码应用到笔者开发的决策信息系统用报表系统中,可以自动生成提前入库合同查询报表,如图9所示。

日期	合同号	生产号	单位	数量	金额	备注
2008-12-28	200801-001	200801001	东北特钢集团	1000	100000	
2008-12-29	200801-002	200801002	东北特钢集团	2000	200000	
2008-12-30	200801-003	200801003	东北特钢集团	3000	300000	
2008-12-31	200801-004	200801004	东北特钢集团	4000	400000	

图9 提前入库合同查询结果

利用提前入库合同查询结果,战略管理处管理人员可以准确定位所有提前入库合同详细信息,正确评价合同实际完成情况,辅助制定科学、高效的企业生产计划。

应用结果表明,使用基于MDA的OLAP开发方法,一方面可以大大提高OLAP查询开发效率,开发效率比传统软件开发方法提高了75%,支持模型重用;另一方面,通过模型驱动的方法可以让设计人员不必关心OLAP查询实现细节,使得他们更多关注模型的概念设计,进而提高模型的适用性和准确性,更加高效满足最终客户查询需求。

4 结束语

基于 MDA 的软件开发方法研究是一个具有广泛用途的领域, 本文将其应用到 OLAP 系统开发中是一个较新的尝试, 该方法在数据库系统统一建模框架下, 将 OLAP 设计从逻辑层提升到概念层, 在概念层实现 OLAP 的 PIM 建模, 通过 PIM 模型到 PSM 模型及 PSM 模型到 SQL 代码的转换实现系统开发。该方法可以提高 OLAP 开发效率, 减少开发代价, 提高了 OLAP 模型复用性, 为决策支持系统的数据分析提供有力的支持。目前, 该方法已应用到某企业决策信息系统开发中, 取得了良好的效果, 为 MDA 在 OLAP 及决策支持系统的应用开发提供了宝贵的经验。下一步将考虑针对生成的 SQL 语句如何进行优化的问题。同时, 考虑将基于 MDA 的方法应用到数据挖掘系统建模中, 形成一个面向决策主题的数据挖掘系统 MDA 开发集成框架。

参考文献:

- [1] LI Qi, BAI Yingcai. Modeling and query of dimension for data warehouses[J]. Journal of Computer Research and Development, 2002, 39(5): 612-618 (in Chinese). [李琪, 白英彩. 数据库中的建模和查询[J]. 计算机研究与发展, 2002, 39(5): 612-618.]
- [2] LI Jianzhong, GAO Hong. Multidimensional data modeling for data warehouse[J]. Journal of Software, 2000, 11(7): 908-917 (in Chinese). [李建中, 高宏. 一种数据仓库的多维数据模型[J]. 软件学报, 2000, 11(7): 908-917.]
- [3] LI Sheng'en, WANG Shan. Multidimensional data model ER (H) [J]. Chinese Journal of Computer, 2005, 28(12): 2059-2067 (in Chinese). [李盛恩, 王珊. 多维数据模型 ER (H) [J]. 计算机学报, 2005, 28(12): 2059-2067.]
- [4] HU Kongfa, CHEN Ling, GU Qi, et al. Rapid multi-dimension hierarchical aggregation algorithm in data warehouse system[J]. Computer Integrated Manufacturing Systems, 2007, 13(1): 196-201 (in Chinese). [胡孔法, 陈峻, 顾颀, 等. 数据仓库系统中一种高效的多维层次聚集算法[J]. 计算机集成制造系统, 2007, 13(1): 196-201.]
- [5] HU Kongfa, DONG Yisheng, XU Lizhen, et al. A novel aggregation algorithm for online analytical processing queries evaluation based on the dimension hierarchical encoding [J]. Journal of Computer Research and Development, 2004, 41(4): 608-614 (in Chinese). [胡孔法, 董逸生, 徐立臻, 等. 一种基于维层次编码的 OLAP 聚集查询算法[J]. 计算机研究与发展, 2004, 41(4): 608-614.]
- [6] SARAWAGI S, AGRAWAL R, MEGIDDO N. Discovery-driven exploration of OLAP data cubes[C]// Proceedings of International Conference on Extending Database Technology. London, UK:Spring-Verlag, 1998:168-182.
- [7] THEODORA TOS D, TSOIS A. Heuristic optimization of OLAP queries in multidimensionally hierarchically clustered databases[C]// Proceedings of the 4th ACM International Workshop on DOLAP. New York, N. Y., USA: ACM Press, 2001:48-55.
- [8] LUJIAN-MORA S, TRUJILLO J, SONG I Y. A UML profile for multidimensional modeling in data warehouses[J]. Data & Knowledge Engineering, 2006, 59(3): 725-769.
- [9] LUJIAN-MORA S, TRUJILLO J, SONG I Y. Model-driven metadata for OLAP cubes from the conceptual modeling of data warehouses[J]. Data & Knowledge Engineering, 2006, 59(3): 725-769.
- [10] SONG Xudong, HU Moqian, LIU Xiaobing. Design OLAP metamodel based on UML profile[J]. Computer Engineering and Design, 2009, 30(9): 2187-2190 (in Chinese). [宋旭东, 胡墨谦, 刘晓冰. 基于 UML Profile 设计 OLAP 元模型[J]. 计算机工程与设计, 2009, 30(9): 2187-2190.]
- [11] Object Management Group (OMG). MDA Guide 1.0.1[EB/OL]. [2009-02-15]. <http://www.omg.org/cgi-bin/apps/doc?omg/03-06-01.pdf>.
- [12] MAZÓN J N, TRUJILLO J. An MDA approach for the development of data warehouses[J]. Decision Support Systems, 2008, 45(1): 41-58.
- [13] ZEPEDA L, CELMA M, ZATARAIN R. A mixed approach for data warehouse conceptual design with MDA[C]// Proceedings of International Conference on Computational Science and its Applications, Part II. Berlin, Heidelberg, Germany:Spring-Verlag, 2008:1204-1217.
- [14] GLORIA O, TRUJILLO J. Designing data warehouses for geographic OLAP querying by using MDA[C]// Proceedings of International Conference on Computational Science and its Application, Part I. Berlin, Germany:Spring-Verlag, 2009: 505-519.
- [15] PARDILO J, MAZÓN J N, ZUBCOFF J. Applying MDA to integrate mining techniques into data warehouses: a time series case study[C]// Proceedings of the 2nd KDD workshop on Mining Multiple Information Sources 2008. Las Vegas, Nev., USA:ACM Press, 2008:47-53.
- [16] SONG Xudong, XU Lianpeng, TIAN Hong. An MDA approach for the development of decision support system[J]. Advances in Information and Systems Sciences, 2009, 3(1/2): 399-406.
- [17] PARDILO J, MAZÓN J N, TRUJILLO J. Bridging the semantic gap in OLAP models: platform-independent queries [C]// Proceedings of ACM the 11th International Workshop on Data Warehousing and OLAP. New York, N. Y., USA: ACM, 2008:89-96.
- [18] WANG Xuebin, WANG Huaimin, WU Quanyuan, et al. A weaving framework for model transformation[J]. Journal of Software, 2006, 17(6): 1423-1435 (in Chinese). [王学斌, 王怀民, 吴泉源, 等. 一种模型转换的编织框架[J]. 软件学报, 2006, 17(6): 1423-1435.]