

## 一种基于 IIMS 的关系数据库融合方案

张渊源,夏雪挺,陶宏才

(西南交通大学 计算机与通信工程学院,四川 成都 610031)

(zyy67@163.com)

**摘 要:**提出了一个基于 IIMS 平台融合关系数据库的实现方案,并针对数据库数据建立元数据标准的方法进行了初步探讨,最后给出了一个基于该方案的应用实例。

**关键词:**IIMS;关系数据库;数据库融合;数据共享;元数据

**中图分类号:**TP311.13 **文献标识码:**A

## Scheme of interfusing relational database based on IIMS

ZHANG Yuan-yuan, XIA Xue-ting, TAO Hong-cai

(School of Computer Science and Communication Engineering, Southwest Jiaotong University, Chengdu Sichuan, 610031, China)

**Abstract:** A new scheme of interfusing relational database based on IIMS was proposed. With this scheme, a preliminary study of the way to build meta data for relational database was study, and a typical application case by using this scheme was given.

**Key words:** IIMS(Information Integrated Management System); relational database; database interfusing; data sharing; meta data

### 0 引言

现实环境下,政府和企业中的大量数据呈现这种状况:1)大量的数据都是非结构化的,除了一些特殊的或专门的数据保存在结构化的数据库(如:关系数据库)中,大量的数据以各种形式保存,如 Word、图片、视频文件等,这些数据没有统一的标准;2)这些二进制的图片或视频文件数据量巨大,往往大到几十兆,这与结构化的数据库中的文本字段有着显著区别;3)数据呈分散性,大量数据并不是集中存放,数据分散在不同部门和个人计算机中;4)数据还涉及安全性、共享、更新等方面的问题。

由日本地质环境网络公司(Geo Environmental Net Co., Ltd.)开发的信息综合管理系统 IIMS(Information Integrated Management System),是目前国际上所能见到的功能完善且实用性很强的信息共享应用系统开发平台,是实现数据综合管理与数据共享的分布式管理系统,有效地解决了上述问题。该系统以元数据和数据集的概念为基础,以空间数据的共享为目的,包括 GIS 数字地图、遥感影像、多媒体文件、办公文档等,主要用来实现分散于网络上的空间数据的综合管理及共享。

但对于结构化数据库,如关系数据库, IIMS 没有提供直接的互连融合平台。目前关系数据库仍是应用的主流平台,政府和企业已经基于关系数据库建立了完整的应用系统和业务流程。这种情况下,完全用 IIMS 来替代关系数据库的作用,重建新的应用系统,在人力、成本各方面都是不合实际的。本文提出了一种基于 IIMS 平台来融合关系数据库的实现方案,并给出了一个利用该实现方案的应用实例。

### 1 数据库融合方案的设计

#### 1.1 设计思想及任务

IIMS 的数据管理是建立在元数据和数据集概念基础之

上的,为实现 IIMS 上的关系数据库融合,可以利用 IIMS 提供的二次开发接口,将关系数据库中的数据转化为元数据及数据集数据,上传至 IIMS 中,从而实现与关系数据库的融合。

按照此思想,具体实现时应完成如下任务:1)应该为关系数据库定义合理的元数据结构;2)针对元数据和数据集,应为其实现完整的增加、删除、更新及查询功能;3)新增元数据或数据集还应设定用户及用户组的权限分配,这样既可以实现多元信息的充分共享,又保证了这些原始数据的安全与完整;4)应实现定义视图等功能。

#### 1.2 实现框架

关系数据库的表结构定义一般为两种情况:1)纯文本型,即记录中不包含相关的图片及文字;2)记录可能与一个或多个图片、文档、视频信息等相关联。数据库设计者可能会用一个字段来保存这些图片、文档或视频数据的存放路径,也可能专门用一表来保存这些路径信息。考虑到以上两种情况,针对 IIMS 平台,作者建议的元数据定义的方法为:元数据的结构与关系数据库的表结构对应,各个元数据项目数据类型与表结构中的定义保持一致,不定义路径信息项目,通过将相关的图片、文档或视频数据上传至数据集服务器,由 IIMS 来管理元数据和数据集的对应关系。

根据上述方法,同时利用 IIMS 提供的二次开发接口,即可制定出如图 1 所示的关系数据库融合方案的实现框架。

图 1 中,XML 配置文件描述了两方面的信息:1) IIMS 中的元数据和关系数据库中表的对应关系;2)每个元数据中的元数据项目和表字段的对应关系及其数据类型定义。

为了实现关系数据库中记录的可维护性,需要修改表结构定义,增设标志位字段,包括上传标志位,更新标志位。同时,为了实现对已导入 IIMS 中的元数据及其数据集的可维护

性,增设两个辅助字段,用来保存每条记录对应的元数据 ID 及数据集文件 ID。

对于图 1 中的数据导入系统,其数据导入方法如下:首先在关系数据库中查询,把查询结果组织为类对象,这些对象携

带了所对应记录的所有信息,然后利用 IIMS 提供的二次开发接口,写入对象值,即每新增一条元数据,为其定义相应的数据集,实现导入 IIMS 的功能,最后回填本地数据库,更新标志位及其他辅助性字段。

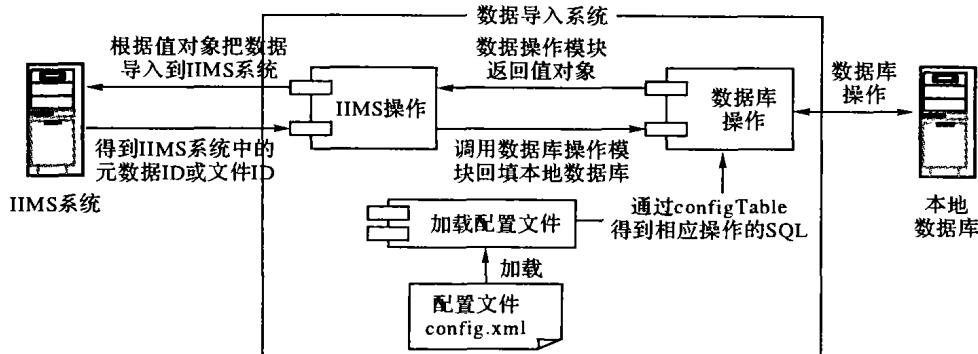


图 1 方案的实现框架

2 应用实例

本节将根据上述方案实现框架,用一个实例来说明基于 IIMS 的关系数据库融合。假设存在一个 SQL 关系数据库,包含 People 和 Photo 两张表,将 People 及其对应的照片信息分开存放。现要求将 People 表中所有人的信息及其照片数据融合到 IIMS 中,从而实现由 IIMS 来统一管理。

2.1 修改 SQL 数据库表结构

按实现框架要求,应修改表结构。对于 People 表,增加 ims\_flagup、ims\_batch\_no、ims\_meta\_id 三个字段,如表 1 所示。对 Photo 表,增加 ims\_flagup、ims\_flag\_new、ims\_batch\_no、ims\_meta\_id、ims\_file\_id 五个字段,如表 2 所示。

表 1 People 表

字段名	类型	备注
P_id	int(自增加)	些字段为原 People 表中的字段
P_name	varchar	
P_sex	varchar	
...	...	...
Ims_flagup	varchar	上传是否成功,成功为 1
Ims_batch_no	varchar	ims 的批次号
Ims_meta_id	varchar	ims 元数据 id

表 2 Photo 表

字段名	类型	备注
Pho_id	int(自增加)	这些字段为原 Photo 表中的字段
P_id	Int	
Pho_name	varchar	
...	...	...
Ims_flagnew	varchar	是否需要更新,1 为需要更新,缺省为 0
Ims_batch_no	varchar	ims 的批次号
Ims_meta_id	varchar	ims 元数据 id
Ims_file_id	varchar	ims 数据集 id

People 表字段描述人的基本特征,Photo 表将字段 P\_id 作为外关键字与之关联。People 表中的一条记录成功导入到 IIMS 后,回填 People 表,即将 ims\_flagup 置为 1,并将 ims\_batch\_no 字段写入上传批次,ims\_meta\_id 置为上传至 IIMS 中后获取的元数据 id;如果 Photo 表中有此人对应的照片信息,则还要将上传批次和元数据 id 写入 Photo 表中对应记录的

ims\_batch\_no 和 ims\_meta\_id 字段;根据全文表中的 ims\_meta\_id 字段,就可以完成上传数据集的工作,将此人的照片全部上传至数据集服务器,对应的元数据 id 为 ims\_meta\_id 字段所记录的值。上传完毕后,再回填全文表,将 ims\_file\_id 置为 0 从 IIMS 获取的文件 id 号,该文件 id 号唯一标识了所上传的图片文件。

2.2 元数据结构定义

仅以 People 元数据为例,People 元数据中增加的项目 ims\_batch\_no,其值与 People 表中对应记录相同,该字段用来方便 IIMS 对元数据的维护工作,如表 3 所示。

表 3 People 元数据

项目	类型
P_name	String
P_sex	String
...	...
ims_batch_no	Integer

2.3 配置文件定义

本方案采用 XML 配置文件来描述关系数据库表与 IIMS 元数据,以及表中字段与元数据项目之间的对应关系。

配置文件中数据库表与 IIMS 元数据的示例描述如下:

```
<? xml version = "1.0" encoding = "GB2312"? >
<! DOCTYPE config [
<! ELEMENT table (iimsTableName, orgTableName, field + ) >
<! ELEMENT iimsTableName ( # PCDATA ) >
<! ELEMENT orgTableName ( # PCDATA ) >
<! ELEMENT field (orgFieldName, iimsFieldName) >
<! ELEMENT orgFieldName ( # PCDATA ) >
<! ELEMENT iimsFieldName ( # PCDATA ) >
<! ATTLIST table isContainFile #REQUIRED > ] >
```

配置文件中表中字段与元数据项目间的对应关系示例描述如下:

```
< config >
< table isContainFile = true >
< iimsTableName > people </iimsTableName >
< orgTableName > people </orgTableName >
< field >
< orgFieldName > P_name </orgFieldName >
< iimsFieldName > P_name </iimsFieldName >
</field >
< field >
```

```

<orgFieldName> P_sex </orgFieldName>
<iimsFieldName> P_sex </iimsFieldName>
</field>
<field>
<orgFieldName> ims_batch_no </orgFieldName>
<iimsFieldName> ims_batch_no </iimsFieldName>
</field>

```

```

...
</table>
</config>

```

## 2.4 数据导入的 UML 描述

类图是一个静态模型,用来描述系统中类及组件之间的固定关系和依赖性,图2为数据导入的类图。

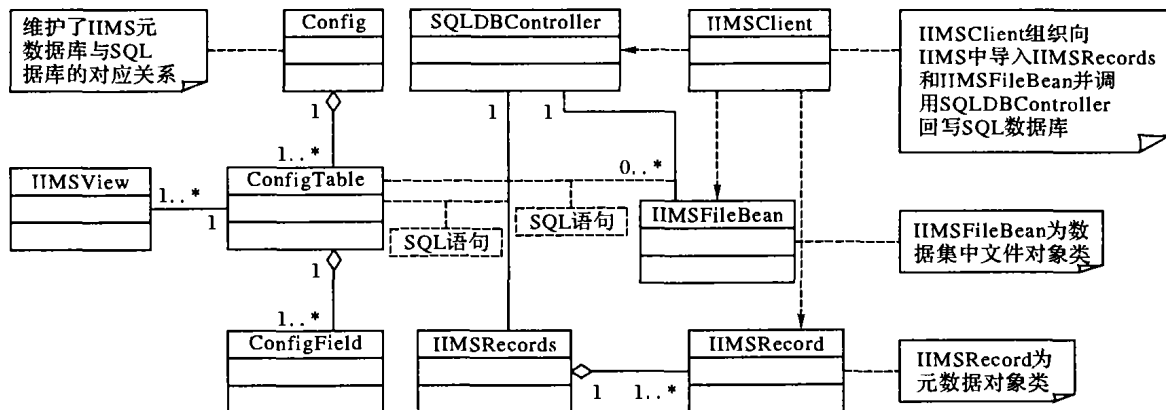


图2 数据导入类图

除了静态模型外,还可动态模型来描述对象间的协作关系。动态模型一般可用四种图来描述,即:顺序图、协作图、

状态图和活动图。限于篇幅,此处仅给出如图3所示的数据导入顺序图。

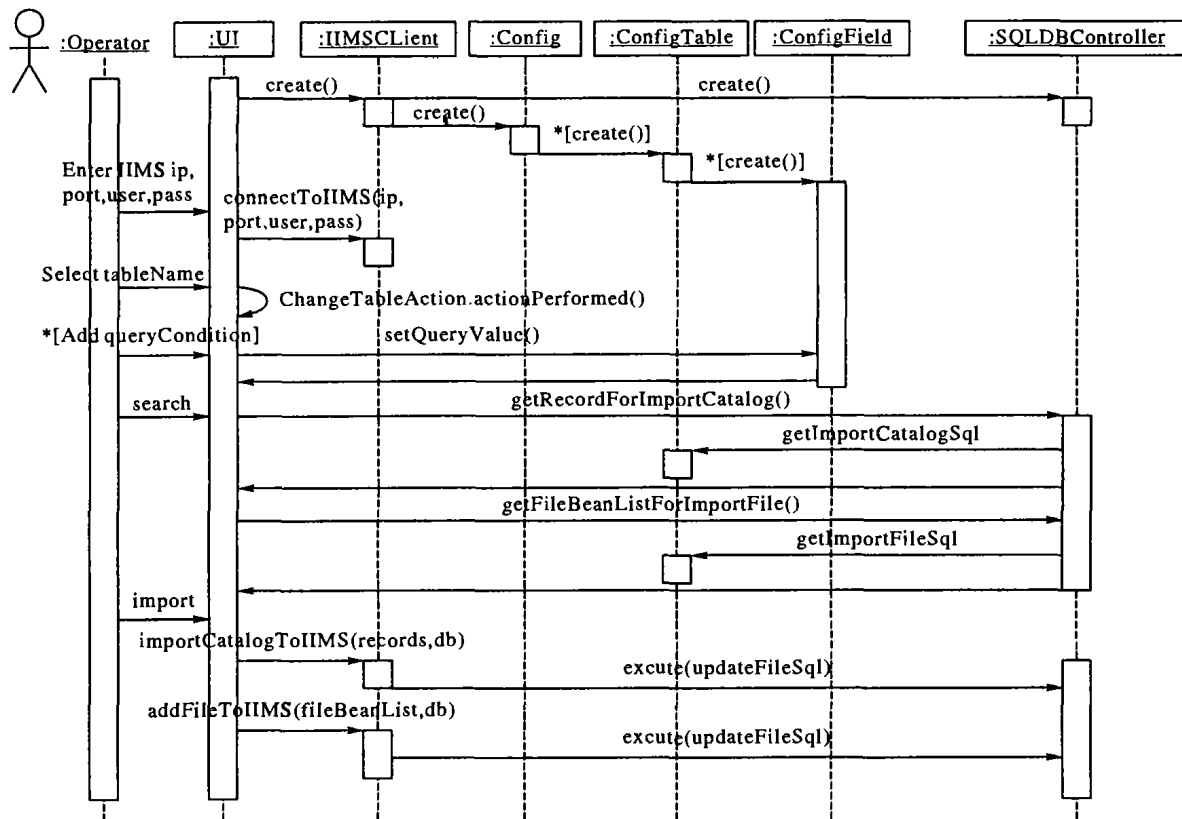


图3 数据导入顺序图

## 3 结语

关系数据库在当前仍是数据库产品的主流, IIMS 的数据库后台虽然仍然是结构化的数据库,但在管理模式上引入了元数据的概念,近年来得到越来越多的应用,因此,有必要对这两者之间的互联融合进行深入的探讨。本文提出的方案可用于实现 IIMS 平台融合关系数据库的目的,在保留原有关系数据库系统的基础上,建立以 IIMS 为业务核心的应用系统,

并保证了原有数据的安全性和完整性。

### 参考文献:

- [1] (美) FILEV A, LOTON T. VS. NET UML 建模高级编程——应用 Visio for Enterprise Architects [M]. 北京: 清华大学出版社, 2003.
- [2] 刘炜, 张亮. 数字图书馆的体系结构与元数据方案 [J]. 情报学报, 2003, (2).
- [3] (美) LARMAN C. UML 和模式应用: 面向对象分析与设计导论 [M]. 北京: 机械工业出版社, 2002.