

用混合数据库系统实现科学与工程数据库

任 涛 丛小杰 翁志成 陈志勇 姜会林*
杨大任* 李共德* 周秀良** 于德英** 杨妹清**

(应用光学国家重点实验室)

摘要: 有很多数据库应用项目, 如工程、科学数据库, 其数据对象都包含有图形, 图像数据或复杂关联的非结构数据。对于这样的数据对象, 只靠数据库管理系统 (DBMS) 很难 (或者说不可能) 实现组织管理。然而, 文件系统却很容易实现对这些数据的存贮组织。事实上, 大多数CAD系统, 都是基于文件之上的。本文讨论了如何结合数据库管理系统的强有力的数据操纵和事物处理能力与文件系统的存贮组织能力构造一个能胜任工程、科学数据要求的数据库系统。

一、前 言

数据库是近年来发展最为迅速的电子计算机软件之一, 是数据信息管理的新技术, 广泛地用于国民经济、文化教育、军事情报、科学研究、人工智能等领域。近年来, 关系型数据库取得了飞速发展。它以方便、直观、高效, 越来越受人们的青睐。

然而, 目前很多应用项目要求的服务, 超出了常规数据库管理系统所能提供的范围。例如, 工程和科学数据的组织管理。这些数据中有大量未结构和松散结构在一起的信息, 或异常长信息, 如图形、图像数据。DBMS不能满足这些应用的要求, 主要表现在以下几个方面:

1. 数据异常长。如图像、图形数据, 一般都在几十k至几百k, 甚至几M。而一般DBMS不支持这一类型数据, 如DBASE-III最长数据为16k, RDB和ORACLE最长数据为64k。
2. 对异常复杂和动态结构关系的有效支持, 如大量动态的图解。
3. 版本历史的支持, 记录数据更改的历史, 恢复对象到以前的状态。

对于上述DBMS的缺陷, 正是文件系统的长处。首先, 文件可以存贮任意长的数据 (只要存贮介质允许); 其次, 对于存贮于文件的数据, 不要求必需是结构化的, 可以是动态联系的; 最后, 文件系统具有版本历史的记载能力。根据版本的不同, 可以将数据恢复到任意一个以前的状态。事实上大多数CAD系统都是基于文件之上的。

但是, 对于一个应用来说, 仅靠文件系统的存贮能力是不行的, 我们还必须充分地利用数据库系统的灵活高效的数据操纵能力、多用户并发访问控制能力、事务处理控制能力、数据安全保密能力和数据库恢复能力。因此, 我们希望把文件系统和数据库系统的能力结合起来, 构造一个更适合工程、科学数据的系统。我们称这样一个系统为混合数据库系统。本文以VMS文件系统和ORACLE关系数据库管理系统为例, 论述混合数据库的数据存贮、事

注: * 长春光学精密机械学院; **长春光机所

务处理、安全保密等问题。本文所阐述的设计思想和方法已在“中国光学镜头数据库”中实现。

二、关系的扩展

我们都知道，所谓关系就是一个二维表。在现有的关系数据库管理系统中，任何一个都不能直接将文件做为其关系中的一个域。换句话说，任何一个 DBMS 都不支持文件这种数据类型。如 ORACLE 所支持的数据类型有 CHAR（字符）、NUMBER（数字）、DATE（日期）和 LONG（长型）。为了能使文件系统与 RDBMS 有机地结合在一起，我们必须对 RDBMS 的关系进行扩展，使其支持文件。为了扩展关系，我们先来考查一下关系的定义。

设有属性 A_1, A_2, \dots, A_k ，它们分别在值域 D_1, D_2, \dots, D_k 中取值，按照集合论的观点，这些值域构成一个笛卡尔乘积空间 $D = D_1 \times D_2 \times \dots \times D_k$ ， D 中的任意一个子集 D' 称为一个关系，记为 R 。其关系框架是属性 A_1 的一个有序集合，记为 $R(A_1, A_2, \dots, A_k)$ 。 D' 中的任一个点 (t_1, t_2, \dots, t_k) 称为 R 的一个元组。

使用集合论的符合，上述定义可表示为：

$$R = \{ (t_1, t_2, \dots, t_k) \mid (t_1, t_2, \dots, t_k) \in D' \subset D \} \text{ 或 } R = [T_k \mid T_k \in D']$$

其中 K 为关系 R 的元数， $t_k = (t_1, \dots, t_k)$ 为 R 的 K 元元组变量。

从上述定义中可知，关系框架是属性的有序集合。对于 ORACLE 数据库管理系统，可知关系只与属性的顺序有关，与该属性对应的数据的存贮无关。因此我们可以把“文件”看做 ORACLE 表中的一个属性，对其原有的属性加以扩展，即把文件看成是 ORACLE 表中的一列。扩展后的关系如下表 1

表 1

ORACLE列 ₁	列 ₂	...	列 _n	文件列
----------------------	----------------	-----	----------------	-----

这样 ORACLE 表被扩展到具有文件属性，或者说具有文件列。扩展后的数据的操纵也必须由 ORACLE 管理系统和 VMS 共同完成。

三、扩展的数据操纵

关系被扩展后，数据被分为 VMS 管理下的文件数据和 RDBMS 管理的表数据两部分。而他们之间又是紧密地联系在一起的一个整体（一个个元组）。因此，每一个数据操纵动作都要由 RDBMS 动作和 VMS 文件管理动作两步组成。

下面以“中国光学镜头数据库”中的一个扩展关系 LENS 为例，说明扩展后数据的操纵。其中 LENS_T 为扩展关系 LENS 的 ORACLE 部分的表名，LENS\$DIR 为扩展关系。文件部分的文件目录名。

LENS _T			LENS\$DIR
列 ₁	列 ₂

关系 LENS

1. 增加一个新记录 (新元组)

增加新记录, 就是向扩展关系中加入新元组, 包括 ORACLE 表部分和 VMS 文件部分。

增加ORACLE部分:

```
INSERT INTO LENS$ VALUES (.....)
```

增加VMS文件部分:

```
COPY 文件 LENS$ DIR
```

2. 删除一个记录

删除ORACLE部分:

```
DELETE FROM LENS$ WHERE ...
```

删除VMS文件部分:

```
DELETE 文件
```

3. 修改一个记录

修改记录数据不象增加、删除记录那样必须由ORACLE和VMS两个动作完成。若要修改的数据, 只在ORACLE表部分, 那么只需ORACLE动作即可以。

```
UPDATD LENS$ SET ... WHERE ...
```

如果需要修改的只是VMS文件部分, 则只需对文件修改。一般情况下, 对文件部分的修改都是通过CAD系统完成的。(文件也是由CAD系统产生的) 可直接启用CAD系统对其进行修改。

在多数情况下, 如果文件被修改了, ORACLE类部分的数据往往也会发生变化, 因此两个动作常常一起执行。

4. 查询操作

我们利用RDBMS的优越的查询检索能力来完成查询操作, VMS 文件做为一个不做查询条件的域存在, 就象ORACLE中的长型 (LONG) 域、RDB中的子段型 (SEGMENTED STRING) 域一样。查询到的文件的引用有两种形式: 直接引用库中的文件和先做副本备份, 然后对副本操作, 可根据应用项目自己设计。

四、混合数据库的事务处理

DBMS具有事务处理控制能力, 以保证数据的一致性, 特别象 ORACLE 这样的大型DBMS具有非常强的事务处理控制能力。混合数据库的数据一致性是指关系表部分的数据, 和其文件部分一一对应。仅靠DBMS的事务处理, 还不能保证混合数据库的数据一致。

例如对关系LENS增加一个记录过程:

① COPY 文件 LENS\$ DIR

② INSERT LENS\$ VALUES (...)

当第一步完成后, 如果第二步产生错误 (如库空间不足, 重复关键字符), 则向关系表LENS\$中插入的记录被回退, 这样就产生了文件与表记录不对应的错误。

那么怎样控制混合数据库的事务处理过程, 保证数据一致性呢? 我们注意到, DBMS具有较强的事务处理能力, 充分利用DBMS的能力, 正确安排文件操作与DBMS操作的顺序, 可以保证混合数据库事务处理的正确性。例如我们安排事务处理过程如下:

①开始。②DBMS事务处理；如果失败，转⑥。③文件处理；如果失败，转⑤。④呈交DBMS事务，转⑦。⑤清理文件现场（指清理临时工作文件）。⑥回退DBMS事务。⑦结束。

按上述事务处理过程，我们就保证了混合数据库事务处理的正确性。如删除一个记录时，首先删除DBMS表部分（注意并没有呈交），如果不成功，就回退DBMS的删除；然后，删除文件，如果不成功，清理文件现场，再回退DBMS的删除；如果前两步都成功，呈交DBMS删除，结束事务处理过程。这样每一次事务处理完了，无论其成功、还是失败，总能保证数据的一致性。

五、锁机制的运用

DBMS都提供有丰富的锁机制，以解决多用户并发访问冲突问题。例如 ORACLE 就提供有共享（SHARE）锁、共享更新（SHARE UPDATE）锁和独占（EXCLUSIVE）锁。对于混合数据库，由于其数据库具有两个部分，即使是查询也要加锁。下面的例子说明了不加锁的后果。

设有两个用户A、B，用户A查询一个记录，用户B做删除这个记录。其执行顺序如下：

①用户A检索出一个记录R（未加锁）。②用户B请求删除记录R（得到锁）。③用户B完成对记录R的删除。④用户A访问记录R的文件域。

当用户A试图访问它所检索的记录的记录的文件域时，这个文件已被删除。这是数据库的一个致命错误——数据不一致。

产生这种数据不一致的根本原因在于没有采用锁机制，DBMS 提供了读的一致性，而文件系统却没有，为对文件锁定，就必须对DBMS表部分加锁。因此，对于混合数据库，即使是只读检索也必须加锁，以保证数据的一致性。

六、混合数据库的安全保密

对于任何一个数据库，数据的安全保密都是一项不可忽视的工作。因此现代 DBMS 都提高了数据的安全保密手段，如ORACLE提供了对数据的查询(SELECT)、增加(INSERT)、修改(UPDATE)、删除(DELETE)等授权保护功能，未经授权的用户不能对数据进行操作。对于工程、科学数据库，其数据涉及到尖端科技、专利技术和军事国防应用等领域，安全保密尤其重要。

利用DBMS的数据安全保密机制可以很容易地实现对关系表部分数据的安全保密。然而混合数据库中还存在有大量文件，仅靠 DBMS 还不能对整个数据库进行保护。必须把DBMS和文件系统的保护机制结合起来实现混合数据库的分级保护。我们可以把用户分做几个级别，分别授予不同的ORACLE操作级别。再根据文件保护级别的不同（不同UIC组号，具有不同级别的保护），授与用户相应的UIC号。

七、结束语

针对于科学、工程数据库的研究，目前在计算机软件界是一个极活跃的领域。面向对象的数据库管理系统是实现工程、科学数据库的较为理想的工具。但面向对象的数据库技术正

在发展之中，还不完善，还没有一个商品化的产品。在这种情况下，采用混合数据方法，实现工程、科学数据库不乏是一种捷径。特别是保留了文件属性更易于与CAD系统连接，也可谓是一优点。另一方面对用户操作来说，它与纯DBMS建立的数据库并无两样。

参 考 文 献

[1] 郑若忠、刘鸿武，《数据库原理与方法》，湖南科学技术出版社，1984

A Method of Realization Science and Engineering Database a Mixed Database System

Ren Tao, Cong Xiaojie

Abstract

Some database application demands to deal with different types of data, including unformatted numbers, texts as well as graphs, which are difficult to be organized by a normal DBMS whereas easy to be managed by a documental system. Based on this application peculiarity, and taking COLDB as an example, the paper puts forward a feasible method to establish a science and engineering database with different types of data by using a hybrid management system a DBMS combined with a documental system.