

12.3 数据库的备份与恢复

主要考点

- 1、数据库系统故障种类
- 2、数据库备份
- 3、数据库恢复

数据库系统故障的种类

1、**事务故障**：是由于程序执行错误而引起事务非预期的、异常终止的故障。通常有如下两类错误引起事务执行失败：

（1）逻辑错误。如非法输入、找不到数据、溢出、超出资源限制等原因引起的事务执行失败。

（2）系统错误。系统进入一种不良状态（如死锁），导致事务无法继续执行。

数据库系统故障的种类

2、系统故障：是指硬件故障、软件（如DBMS、OS或应用程序）漏洞的影响，导致丢失了内存中的信息，影响正在执行的事务，但未破坏存储在外存上的信息。

3、介质故障：是指数据库的存储介质发生故障，如磁盘损坏、瞬间强磁场干扰等。这种故障直接破坏了数据库，会影响到所有正在读取这部分数据的事务。

数据库备份

- 数据转储是将数据库自制到另一个磁盘或磁带上保存起来的过程，又称为数据备份。

(1) **静态转储和动态转储**。静态转储是指在转储期间不允许对数据库进行任何存取、修改操作；动态转储是在转储期间允许对数据库进行存取、修改操作，因此，转储和用户事务可并发执行。

(2) **海量转储和增量转储**。海量转储是指每次转储全部数据；增量转储是指每次只转储上次转储后更新过的数据。

(3) **日志文件**。在事务处理的过程中，DBMS把事务开始、事务结束以及对数据库的插入、删除和修改的每一次操作写入日志文件。

(4) **数据库镜像**。为了避免磁盘介质出现故障影响数据库的可用性，许多DBMS提供数据库镜像功能用于数据库恢复。

数据库恢复

要使数据库在发生故障后能够恢复，必须建立冗余数据，在故障发生后利用这些冗余数据实施数据库恢复，常用的是数据转储和日志文件。

1、故障恢复的两个操作：

(1) **撤销事务(UNDO)**：将未完成的事务撤销，使数据库恢复到事务执行前的正确状态。

撤销事务的过程：反向扫描日志文件（由后向前扫描），查找事务的更新操作；对该事务的更新操作执行逆操作，用日志文件记录中更新前的值写入数据库，插入的记录从数据库中删除，删除的记录重新插入数据库中；继续反向扫描日志文件，查找该事务的其它更新操作并执行逆操作直至事务开始标志。

(2) **重做事务(RED0)**：将已提交的事务重新执行。

重做事务的过程：从事务的开始标志起，正向扫描日志文件，重新执行日志文件登记的该事务对数据库的所有操作，直至事务结束标识。

数据库恢复

- 2、故障恢复策略：

(1) **事务故障的恢复**：事务故障是事务在运行至正常终止点（SUMMIT或ROLLBACK）前终止，日志文件只有该事务的开始标识而没有结束标识。对这类故障的恢复通常是通过撤销（UNDO）产生故障的事务，使数据库恢复到该事务执行前的正确状态来完成的。

具体做法：

- 1、反向扫描日志文件，查找该事务的更新操作。
- 2、对事务的更新操作执行逆操作。
- 3、继续反向扫描日志文件，查找该事务的其他更新操作，并做同样的处理，直到事务的开始标志。

注：事务故障的恢复是由系统自动完成的，对用户是透明的。

数据库恢复

(2) 系统故障的恢复：系统故障会使数据库的数据不一致：

一是未完成的事务对数据库的更新可能已经写入数据库；

二是已提交的事务对数据库的更新可能还在缓冲区来不及写入数据库。

因此对于系统故障，恢复操作是UNDO+REDO：

1、撤销故障发生时未完成的事务（UNDO）。

2、重做已经提交的事务（REDO）。

(3) 介质故障的恢复：介质故障时数据库遭到破坏，需要重装数据库，一般需要DBA的参与，装载故障前最近一次的备份和故障前的日志文件副本，再按照系统故障的恢复过程执行撤销（UNDO）和重做（REDO）来恢复。

数据库恢复

例：系统重启后，由DBMS根据（ ）对数据库进行恢复，将已提交的事务对数据库的修改写入硬盘。
输入数据违反完整性约束导致的数据库故障属于（ ）

- | | | | |
|---------|----------|---------|----------|
| A. 日志 | B. 数据库文件 | C. 索引记录 | D. 数据库副本 |
| A. 系统故障 | B. 事务故障 | C. 介质故障 | D. 网络故障 |

数据库恢复

检查点机制 (CHECKPOINT)：在日志中设置检查点，当发生故障需要利用日志文件恢复时，反向扫描日志文件，找到检查点，确认检查点时刻正在执行的事务（活动事务），即检查点前有事务开始标志但没有事务结束标志。

对于检查点后提交的事务，执行REDO（重做）

对于检查点后未提交的事务，执行UNDO（撤销）

【说明】

如果一个数据库恢复系统采用检查点机制，且其日志文件如表5-1所示，第一列表示日志记录编号，第二列表示日志记录内容， $\langle T_i, \text{START} \rangle$ 表示事务 T_i 开始执行， $\langle T_i, \text{COMMIT} \rangle$ 表示事务 T_i 提交， $\langle T_i, D, V_1, V_2 \rangle$ 表示事务 T_i 将数据项 D 的值由 V_1 修改为 V_2 ，请回答以下问题。

【问题1】（6分）

假设系统开始执行前 $X=100, Y=50, Z=10$ ，系统出错恢复后， X, Y, Z 各自的数值是多少？

【问题2】（4分）

系统发生事务故障时，故障恢复有撤销事务（UNDO）和重做事务（REDO）两个操作，请给出系统恢复时需要REDO的事务列表和需要UNDO的事务列表。

【问题3】（5分）

请简要描述系统出错后，基于检查点的恢复过程。

LSN1	$\langle T_1, \text{START} \rangle$
LSN2	$\langle T_1, X, 100, 1 \rangle$
LSN3	$\langle T_2, \text{START} \rangle$
LSN4	$\langle T_2, X, 1, 3 \rangle$
LSN5	$\langle T_3, \text{START} \rangle$
LSN6	$\langle T_2, Y, 50, 6 \rangle$
LSN7	$\langle T_3, Y, 6, 8 \rangle$
LSN8	$\langle T_3, Z, 10, 9 \rangle$
LSN9	CHECKPOINT
LSN10	$\langle T_1, \text{COMMIT} \rangle$
LSN11	$\langle T_3, Z, 9, 10 \rangle$
LSN12	CRASH