

第 18 讲:多版本并发控制

15-445/645 数据库系统 (2022 年秋季)

<https://15445.courses.cs.cmu.edu/fall2022/>

卡内基梅隆大学安迪·帕夫洛

1 多版本并发控制

多版本并发控制(MVCC)是一个比并发控制协议更大的概念。它涉及 DBMS 设计和实现的所有方面。MVCC 是数据库管理系统中应用最广泛的方案。在过去的 10 年里,它几乎被用于每一个新的 DBMS。甚至一些不支持多语句事务的系统(例如 NoSQL)也在使用它。

使用 MVCC, DBMS 在数据库中维护单个逻辑对象的多个物理版本。当事务写入对象时, DBMS 创建该对象的新版本。当事务读取对象时,它读取的是事务开始时存在的最新版本。

MVCC 的基本概念/好处是写者不阻塞写者,读者不阻塞读者。这意味着一个事务可以修改一个对象,而其他事务则读取旧版本。

使用 MVCC 的一个优点是只读事务可以在不使用任何类型的锁的情况下读取数据库的一致快照。此外,多版本的数据库管理系统可以很容易地支持时间旅行查询,这是基于数据库在其他时间点的状态的查询(例如,在数据库上执行 3 小时前的查询)。

有四个重要的 MVCC 设计决策:

- 1.并发控制协议
- 2.版本存储
- 3.垃圾收集
- 4.指标管理

并发协议的选择是在之前的讲座中讨论的方法之间(两阶段锁定,时间戳排序,乐观并发控制)。

快照隔离

快照隔离涉及在事务启动时为事务提供数据库的一致快照。快照中的数据值只包含来自自己提交事务的值,并且事务在完成之前与其他事务完全隔离地运行。这是只读事务的想法,因为它们不需要等待来自其他事务的写。写操作在事务的私有工作区中维护,只有在事务成功提交后才对数据库可见。如果两个事务更新同一个对象,则第一个写入者获胜。

当两个并发事务修改不同的对象导致竞争条件时,快照隔离中可能发生写倾斜异常。

2 版本存储

这就是 DBMS 如何存储逻辑对象的不同物理版本，以及事务如何找到对它们可见的最新版本。

DBMS 使用元组的指针字段为每个逻辑元组创建一个**版本链**，它本质上是一个按时间戳排序的版本链表。这允许 DBMS 在运行时找到对特定事务可见的版本。索引总是指向链的“头”，根据实现的不同，它可能是最新的版本，也可能是最旧的版本。一个线程遍历链，直到找到正确的版本。不同的存储方案决定了每个版本的存储位置/内容。

方法 1:仅追加存储

逻辑元组的所有物理版本都存储在同一个表空间中。版本在表中混合在一起，每次更新只是将元组的一个新版本追加到表中，并更新版本链。这个链可以是*旧到新*(O2N)，需要在查找时进行链式遍历，也可以是*新到旧*(N2O)，需要为每个新版本更新索引指针。

方法 2:时间旅行存储

DBMS 维护一个单独的表，称为时间旅行表，用于存储旧版本的元组。每次更新时，DBMS 都会将旧版本的元组复制到时间旅行表中，并用新数据覆盖主表中的元组。主表中的元组指针指向时间旅行表中的过去版本。

方法三:增量存储

类似于时间旅行存储，但不是整个过去的元组，DBMS 只存储增量，或元组之间的变化，即所谓的增量存储段。然后，事务可以通过迭代增量来重新创建旧版本。这导致写入速度比时间旅行存储更快，但读取速度更慢。

3 垃圾回收

随着时间的推移，DBMS 需要从数据库中删除**可回收**的物理版本。如果一个版本没有活动的事务可以“看到”这个版本，或者它是由一个被中止的事务创建的，那么这个版本就是可回收的。

方法 #1:元组级 GC

使用元组级垃圾收集，DBMS 通过直接检查元组来查找旧版本。有两种方法可以实现这一点：

- 后台吸尘**:单独的线程定期扫描表并寻找可回收的版本。这适用于任何版本存储方案。一个简单的优化是维护一个“脏页位图”，记录自上次扫描以来哪些页被修改过。这使得线程可以跳过没有修改过的页面。
- 协同清理**:工作线程在遍历版本链时识别可回收版本。这只适用于 O2N 链。

方法 2:交易级 GC

在事务级垃圾收集中，每个事务负责跟踪它们自己的旧版本，因此 DBMS 不必扫描元组。每个事务维护自己的读/写集合。

当一个事务完成时，垃圾收集器可以使用它来标识要回收哪些元组。DBMS 确定完成事务创建的所有版本何时不再可见。

4.索引管理

所有主键(pkey)索引总是指向版本链头。DBMS 更新 pkey 索引的频率取决于当元组更新时系统是否创建新版本。如果一个事务更新了一个(多个)pkey 属性，那么这被视为一个删除，然后是一个插入。

管理二级索引则更加复杂。有两种方法来处理它们。

方法一:逻辑指针

DBMS 对每个元组使用固定的标识符，该标识符不会更改。这需要一个额外的间接层，将逻辑 id 映射到元组的物理位置。然后，对元组的更新可以只更新间接层中的映射。

方法#2:物理指针

DBMS 使用版本链头的物理地址。这需要在版本链头更新时更新每个索引。