

# 数据库系统原理

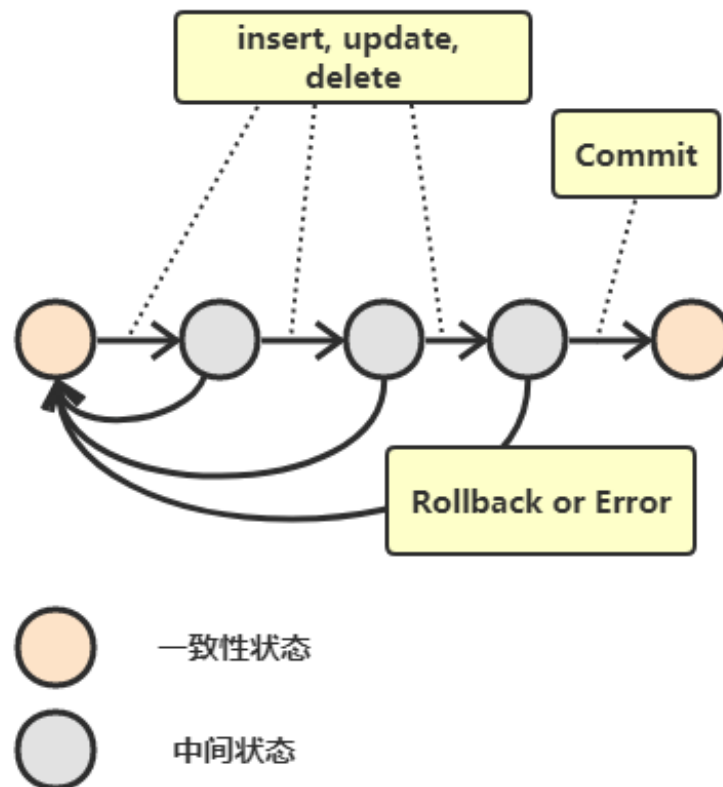
原作者github: <https://github.com/CyC2018/Interview-Notebook>

PDF离线制作github: <https://github.com/sjsdfg/Interview-Notebook-PDF>

希望各位不吝star

## 一、事务

### 概念



### ACID

## 1. 原子性 ( Atomicity )

事务被视为不可分割的最小单元，事务的所有操作要么全部提交成功，要么全部失败回滚。

回滚可以用日志来实现，日志记录着事务所执行的修改操作，在回滚时反向执行这些修改操作即可。

## 2. 一致性 ( Consistency )

数据库在事务执行前后都保持一致性状态。

在一致性状态下，所有事务对一个数据的读取结果都是相同的。

## 3. 隔离性 ( Isolation )

一个事务所做的修改在最终提交以前，对其它事务是不可见的。

## 4. 持久性 ( Durability )

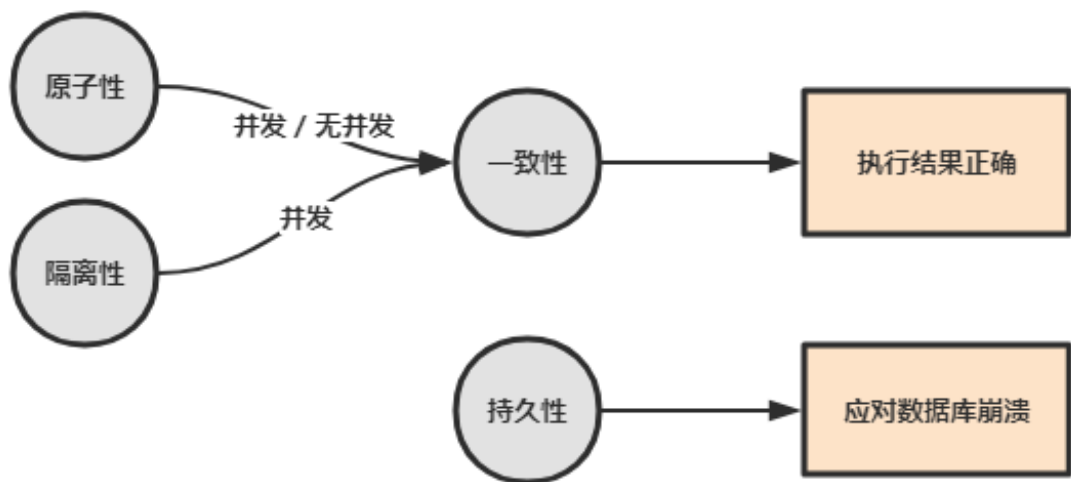
一旦事务提交，则其所做的修改将会永远保存到数据库中。即使系统发生崩溃，事务执行的结果也不能丢失。

可以通过数据库备份和恢复来实现，在系统发生奔溃时，使用备份的数据库进行数据恢复。

---

事务的 ACID 特性概念简单，但不是很好理解，主要是因为这几个特性不是一种平级关系：

- 只有满足一致性，事务的执行结果才是正确的。
- 在无并发的情况下，事务串行执行，隔离性一定能够满足。此时要只要能满足原子性，就一定能满足一致性。
- 在并发的情况下，多个事务并发执行，事务不仅要满足原子性，还需要满足隔离性，才能满足一致性。
- 事务满足持久化是为了能应对数据库奔溃的情况。



## AUTOCOMMIT

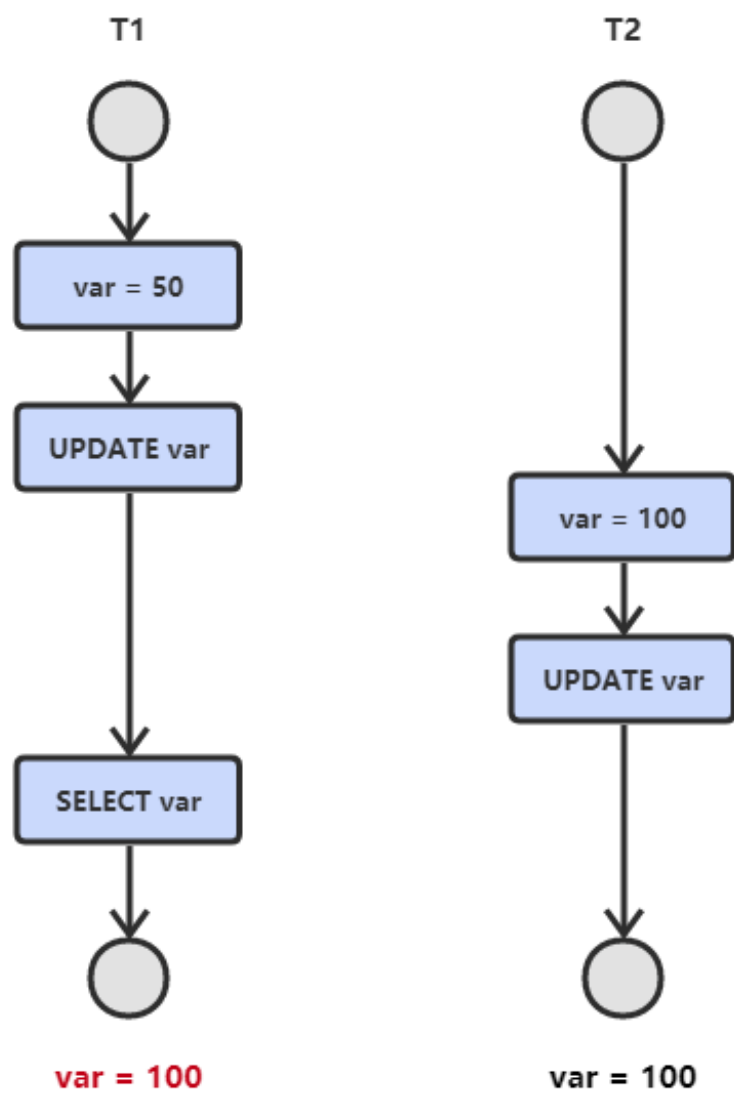
MySQL 默认采用自动提交模式。也就是说，如果不显式使用 `START TRANSACTION` 语句来开始一个事务，那么每个查询都会被当做一个事务自动提交。

## 二、并发一致性问题

在并发环境下，事务的隔离性很难保证，因此会出现很多并发一致性问题。

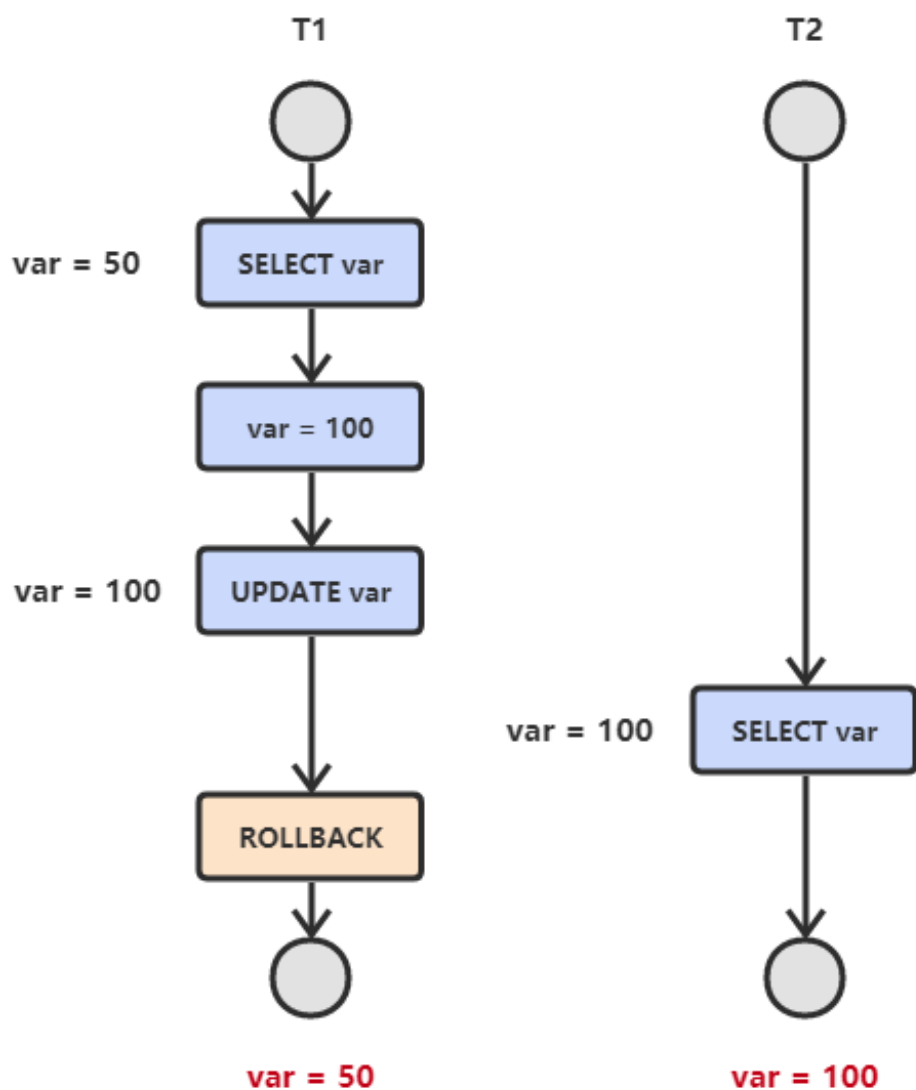
### 丢失修改

$T_1$  和  $T_2$  两个事务都对一个数据进行修改， $T_1$  先修改， $T_2$  随后修改， $T_2$  的修改覆盖了  $T_1$  的修改。



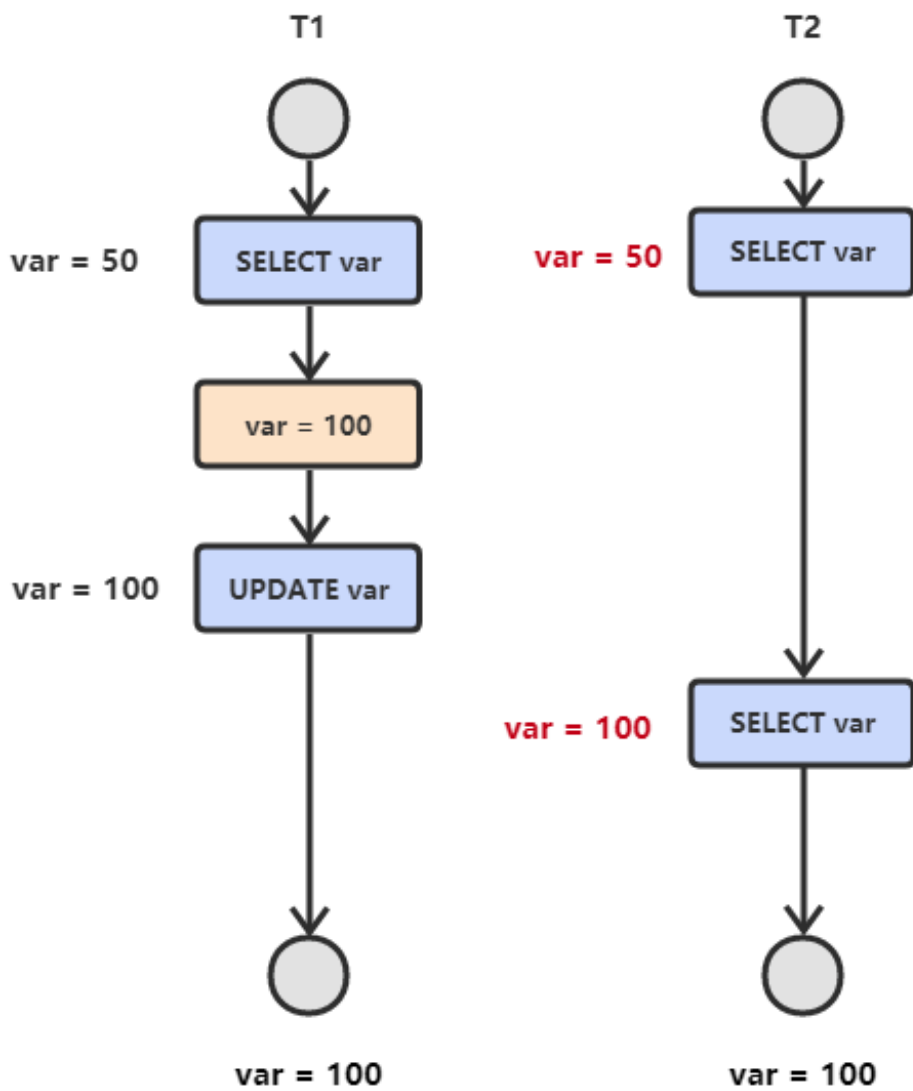
## 读脏数据

T<sub>1</sub> 修改一个数据，T<sub>2</sub> 随后读取这个数据。如果 T<sub>1</sub> 撤销了这次修改，那么 T<sub>2</sub> 读取的数据是脏数据。



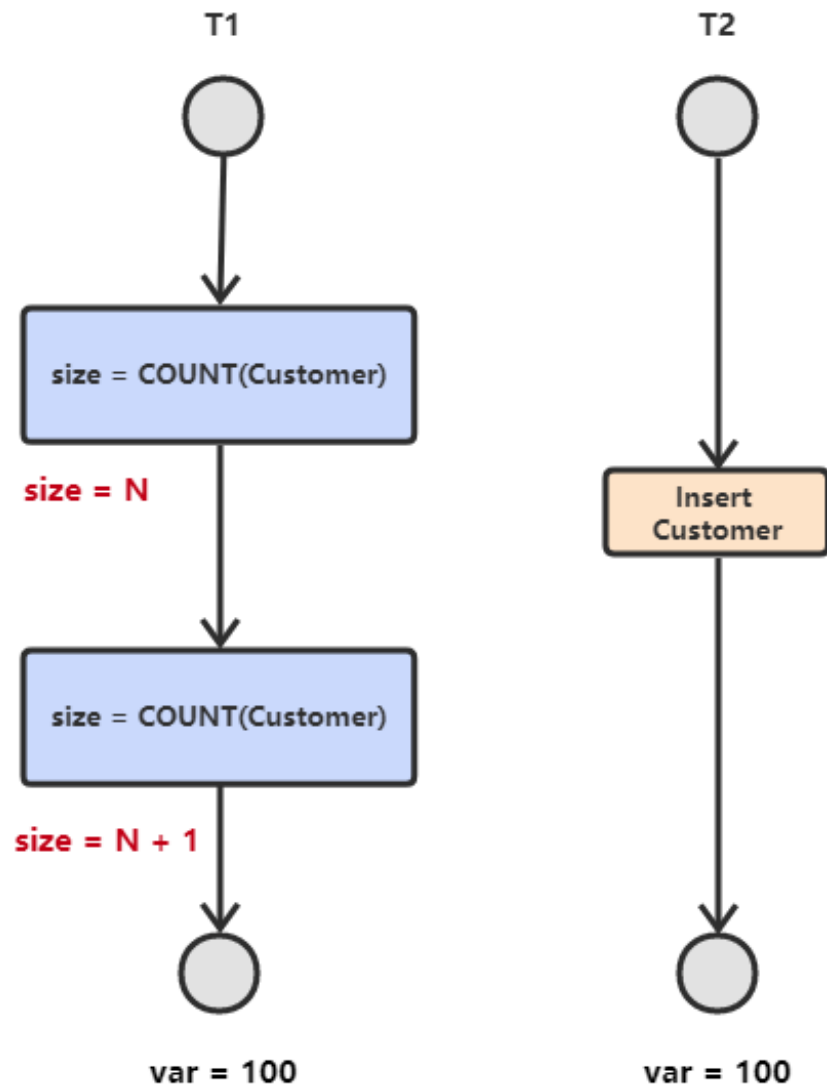
## 不可重复读

$T_2$  读取一个数据， $T_1$  对该数据做了修改。如果  $T_2$  再次读取这个数据，此时读取的结果和第一次读取的结果不同。



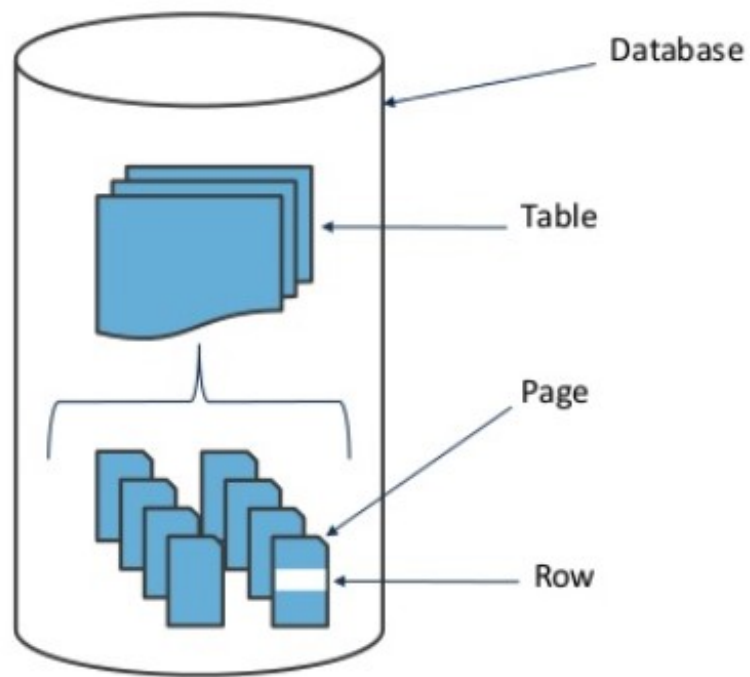
## 幻影读

$T_1$  读取某个范围的数据， $T_2$  在这个范围内插入新的数据， $T_1$  再次读取这个范围的数据，此时读取的结果和第一次读取的结果不同。



### 三、封锁

封锁粒度



## 封锁类型

### 1. 读写锁

- 排它锁 ( Exclusive ) ，简称为 X 锁，又称写锁。
- 共享锁 ( Shared ) ，简称为 S 锁，又称读锁。

有以下两个规定：

- 一个事务对数据对象 A 加了 X 锁，就可以对 A 进行读取和更新。加锁期间其它事务不能对 A 加任何锁。
- 一个事务对数据对象 A 加了 S 锁，可以对 A 进行读取操作，但是不能进行更新操作。加锁期间其它事务能对 A 加 S 锁，但是不能加 X 锁。

锁的兼容关系如下：

-	X	S
X	NO	NO
S	YES	YES



-	X	S
S	NO	YES

## 2. 意向锁

使用意向锁 ( Intention Locks ) 可以更容易地支持多粒度封锁。

在存在行级锁和表级锁的情况下，事务 T 想要对表 A 加 X 锁，就需要先检测是否有其它事务对表 A 或者表 A 中的任意一行加了锁，那么就需要对表 A 的每一行都检测一次，这是非常耗时的。

意向锁在原来的 X/S 锁之上引入了 IX/IS，IX/IS 都是表锁，用来表示一个事务想要在表中的某个数据行上加 X 锁或 S 锁。有以下两个规定：

- 一个事务在获得某个数据行对象的 S 锁之前，必须先获得表的 IS 锁或者更强的锁；
- 一个事务在获得某个数据行对象的 X 锁之前，必须先获得表的 IX 锁。

通过引入意向锁，事务 T 想要对表 A 加 X 锁，只需要先检测是否有其它事务对表 A 加了 X/IX/S/IS 锁，如果加了就表示有其它事务正在使用这个表或者表中某一行的锁，因此事务 T 加 X 锁失败。

各种锁的兼容关系如下：

-	X	IX	S	IS
X	NO	NO	NO	NO
IX	NO	YES	NO	YES
S	NO	NO	YES	YES
IS	NO	YES	YES	YES

解释如下：

- 任意 IS/IX 锁之间都是兼容的，因为它们只是表示想要对表加锁，而不是真正加锁；
- S 锁只与 S 锁和 IS 锁兼容，也就是说事务 T 想要对数据行加 S 锁，其它事务可以已经获

得对表或者表中的行的 S 锁。

# 封锁协议

## 1. 三级封锁协议

### 一级封锁协议

事务 T 要修改数据 A 时必须加 X 锁，直到 T 结束才释放锁。

可以解决丢失修改问题，因为不能同时有两个事务对同一个数据进行修改，那么事务的修改就不会被覆盖。

T <sub>1</sub>	T <sub>2</sub>
lock-x(A)	
read A=20	
	lock-x(A)
	wait
write A=19	.
commit	.
unlock-x(A)	.
	obtain
	read A=19
	write A=21
	commit
	unlock-x(A)

### 二级封锁协议

在一级的基础上，要求读取数据 A 时必须加 S 锁，读取完马上释放 S 锁。

可以解决读脏数据问题，因为如果一个事务在对数据 A 进行修改，根据 1 级封锁协议，会加 X 锁，那么就不能再加 S 锁了，也就是不会读入数据。

T <sub>1</sub>	T <sub>2</sub>
lock-x(A)	
read A=20	
write A=19	
	lock-s(A)
	wait
rollback	.
A=20	.
unlock-x(A)	.
	obtain
	read A=20
	commit
	unlock-s(A)

### 三级封锁协议

在二级的基础上，要求读取数据 A 时必须加 S 锁，直到事务结束了才能释放 S 锁。

可以解决不可重复读的问题，因为读 A 时，其它事务不能对 A 加 X 锁，从而避免了在读的期间数据发生改变。

T <sub>1</sub>	T <sub>2</sub>
lock-s(A)	
read A=20	

T	T
	lock-x(A)
	wait
read A=20	.
commit	.
unlock-s(A)	.
	obtain
	read A=20
	write A=19
	commit
	unlock-X(A)

## 2. 两段锁协议

加锁和解锁分为两个阶段进行。

可串行化调度是指，通过并发控制，使得并发执行的事务结果与某个串行执行的事务结果相同。

事务遵循两段锁协议是保证可串行化调度的充分条件。例如以下操作满足两段锁协议，它是可串行化调度。

```
1. lock-x(A) ... lock-s(B) ... lock-s(C) ... unlock(A) ... unlock(C) ... unlock(B)
```

但不是必要条件，例如以下操作不满足两段锁协议，但是它还是可串行化调度。

```
1. lock-x(A) ... unlock(A) ... lock-s(B) ... unlock(B) ... lock-s(C) ... unlock(C)
```

## MySQL 隐式与显示锁定

MySQL 的 InnoDB 存储引擎采用两段锁协议，会根据隔离级别在需要的时候自动加锁，并且所有的锁都是在同一时刻被释放，这被称为隐式锁定。

InnoDB 也可以使用特定的语句进行显示锁定：

```
1. SELECT ... LOCK In SHARE MODE;  
2. SELECT ... FOR UPDATE;
```

## 四、隔离级别

### 未提交读 ( READ UNCOMMITTED )

事务中的修改，即使没有提交，对其它事务也是可见的。

### 提交读 ( READ COMMITTED )

一个事务只能读取已经提交的事务所做的修改。换句话说，一个事务所做的修改在提交之前对其它事务是不可见的。

### 可重复读 ( REPEATABLE READ )

保证在同一个事务中多次读取同样数据的结果是一样的。

### 可串行化 ( SERIALIZABLE )

强制事务串行执行。

---

隔离级别	脏读	不可重复读	幻影读
未提交读	YES	YES	YES
提交读	NO	YES	YES
可重复读	NO	NO	YES
可串行化	NO	NO	NO

## 五、多版本并发控制

多版本并发控制 ( Multi-Version Concurrency Control, MVCC ) 是 MySQL 的 InnoDB 存储引擎实现隔离级别的一种具体方式，用于实现提交读和可重复读这两种隔离级别。而未提交读隔离级别总是读取最新的数据行，无需使用 MVCC；可串行化隔离级别需要对所有读取的行都加锁，单纯使用 MVCC 无法实现。

### 版本号

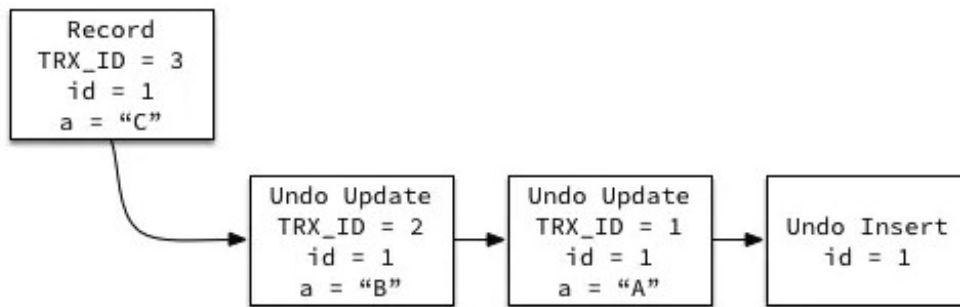
- 系统版本号：是一个递增的数字，每开始一个新的事务，系统版本号就会自动递增。
- 事务版本号：事务开始时的系统版本号。

InnoDB 的 MVCC 在每行记录后面都保存着两个隐藏的列，用来存储两个版本号：

- 创建版本号：指示创建一个数据行的快照时的系统版本号；
- 删除版本号：如果该快照的删除版本号大于当前事务版本号表示该快照有效，否则表示该快照已经被删除了。

### Undo 日志

InnoDB 的 MVCC 使用到的快照存储在 Undo 日志中，该日志通过回滚指针把一个数据行 ( Record ) 的所有快照连接起来。



## 实现过程

以下实现过程针对可重复读隔离级别。

### 1. SELECT

当开始新一个事务时，该事务的版本号肯定会大于当前所有数据行快照的创建版本号，理解这一点很关键。

多个事务必须读取到同一个数据行的快照，并且这个快照是距离现在最近的一个有效快照。但是也有例外，如果有一个事务正在修改该数据行，那么它可以读取事务本身所做的修改，而不用和其它事务的读取结果一致。

把没有对一个数据行做修改的事务称为 T，T 所要读取的数据行快照的创建版本号必须小于 T 的版本号，因为如果大于或者等于 T 的版本号，那么表示该数据行快照是其它事务的最新修改，因此不能去读取它。

除了上面的要求，T 所要读取的数据行快照的删除版本号必须大于 T 的版本号，因为如果小于等于 T 的版本号，那么表示该数据行快照是已经被删除的，不应该去读取它。

### 2. INSERT

将当前系统版本号作为数据行快照的创建版本号。

### 3. DELETE

将当前系统版本号作为数据行快照的删除版本号。

### 4. UPDATE

将当前系统版本号作为更新后的数据行快照的创建版本号，同时将当前系统版本号作为更新前的数据行快照的删除版本号。可以理解为先执行 DELETE 后执行 INSERT。

## 快照读与当前读

### 1. 快照读

使用 MVCC 读取的是快照中的数据，这样可以减少加锁所带来的开销。

```
1.  select * from table ...;
```

### 2. 当前读

读取的是最新的数据，需要加锁。以下第一个语句需要加 S 锁，其它都需要加 X 锁。

```
1.  select * from table where ? lock in share mode;  
2.  select * from table where ? for update;  
3.  insert;  
4.  update;  
5.  delete;
```

## 六、Next-Key Locks

Next-Key Locks 也是 MySQL 的 InnoDB 存储引擎的一种锁实现。MVCC 不能解决幻读的问题，Next-Key Locks 就是为了解决这个问题而存在的。在可重复读（REPEATABLE READ）隔离级别下，使用 MVCC + Next-Key Locks 可以解决幻读问题。



## Record Locks

锁定整个记录（行）。锁定的对象是记录的索引，而不是记录本身。如果表没有设置索引，InnoDB 会自动在主键上创建隐藏的聚集索引，因此 Record Locks 依然可以使用。

## Gap Locks

锁定一个范围内的索引，例如当一个事务执行以下语句，其它事务就不能在 t.c 中插入 15。

```
1. SELECT c FROM t WHERE c BETWEEN 10 and 20 FOR UPDATE;
```

## Next-Key Locks

它是 Record Locks 和 Gap Locks 的结合，不仅锁定一个记录，也锁定范围内的索引。在 user 中有以下记录：

1.		id		last_name		first_name		age	
2.		-----		-----		-----		-----	
3.		4		stark		tony		21	
4.		1		tom		hiddleston		30	
5.		3		morgan		freeman		40	
6.		5		jeff		dean		50	
7.		2		donald		trump		80	
8.		-----		-----		-----		-----	

那么就需要锁定以下范围：

1.	(-∞, 21]
2.	(21, 30]
3.	(30, 40]
4.	(40, 50]
5.	(50, 80]
6.	(80, ∞)

## 七、关系数据库设计理论

### 函数依赖

记  $A \rightarrow B$  表示 A 函数决定 B，也可以说 B 函数依赖于 A。

如果  $\{A_1, A_2, \dots, A_n\}$  是关系的一个或多个属性的集合，该集合函数决定了关系的其它所有属性并且是最小的，那么该集合就称为键码。

对于  $A \rightarrow B$ ，如果能找到 A 的真子集  $A'$ ，使得  $A' \rightarrow B$ ，那么  $A \rightarrow B$  就是部分函数依赖，否则就是完全函数依赖；

对于  $A \rightarrow B, B \rightarrow C$ ，则  $A \rightarrow C$  是一个传递函数依赖。

### 异常

以下的学生课程关系的函数依赖为  $Sno, Cname \rightarrow Sname, Sdept, Mname, Grade$ ，键码为  $\{Sno, Cname\}$ 。也就是说，确定学生和课程之后，就能确定其它信息。

Sno	Sname	Sdept	Mname	Cname	Grade
1	学生-1	学院-1	院长-1	课程-1	90
2	学生-2	学院-2	院长-2	课程-2	80
2	学生-2	学院-2	院长-2	课程-1	100
3	学生-3	学院-2	院长-2	课程-2	95

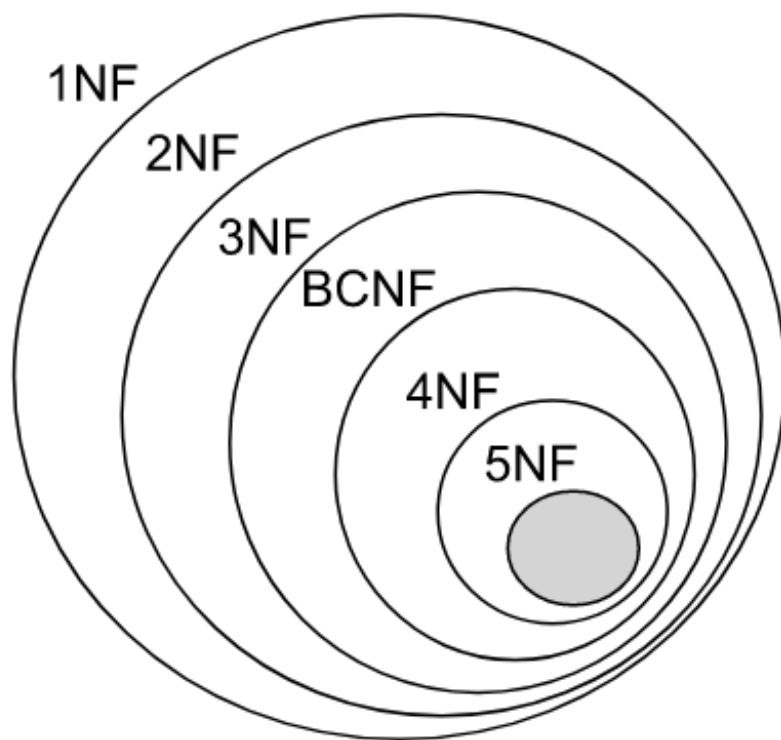
不符合范式的关系，会产生很多异常，主要有以下四种异常：

- 冗余数据：例如 学生-2 出现了两次。
- 修改异常：修改了一个记录中的信息，但是另一个记录中相同的信息却没有被修改。
- 删除异常：删除一个信息，那么也会丢失其它信息。例如如果删除了 课程-1，需要删除第一行和第三行，那么 学生-1 的信息就会丢失。
- 插入异常，例如想要插入一个学生的信息，如果这个学生还没选课，那么就无法插入。

# 范式

范式理论是为了解决以上提到四种异常。

高级别范式的依赖于低级别的范式，1NF 是最低级别的范式。



## 1. 第一范式 (1NF)

属性不可分；

## 2. 第二范式 (2NF)

每个非主属性完全函数依赖于键码。

可以通过分解来满足。

## 分解前

Sno	Sname	Sdept	Mname	Cname	Grade
1	学生-1	学院-1	院长-1	课程-1	90
2	学生-2	学院-2	院长-2	课程-2	80
2	学生-2	学院-2	院长-2	课程-1	100
3	学生-3	学院-2	院长-2	课程-2	95

以上学生课程关系中，{Sno, Cname} 为键码，有如下函数依赖：

- Sno -> Sname, Sdept
- Sdept -> Mname
- Sno, Cname-> Grade

Grade 完全函数依赖于键码，它没有任何冗余数据，每个学生的每门课都有特定的成绩。

Sname, Sdept 和 Mname 都部分依赖于键码，当一个学生选修了多门课时，这些数据就会出现多次，造成大量冗余数据。

## 分解后

### 关系-1

Sno	Sname	Sdept	Mname
1	学生-1	学院-1	院长-1
2	学生-2	学院-2	院长-2
3	学生-3	学院-2	院长-2

有以下函数依赖：

- Sno -> Sname, Sdept, Mname
- Sdept -> Mname

### 关系-2

Sno	Cname	Grade
1	课程-1	90
2	课程-2	80
2	课程-1	100
3	课程-2	95

有以下函数依赖：

- Sno, Cname -> Grade

### 3. 第三范式 (3NF)

非主属性不传递函数依赖于键码。

上面的 关系-1 中存在以下传递函数依赖：Sno -> Sdept -> Mname，可以进行以下分解：

关系-11

Sno	Sname	Sdept
1	学生-1	学院-1
2	学生-2	学院-2
3	学生-3	学院-2

关系-12

Sdept	Mname
学院-1	院长-1
学院-2	院长-2

## 八、ER 图

Entity-Relationship，有三个组成部分：实体、属性、联系。

用来进行关系型数据库系统的概念设计。

### 实体的三种联系

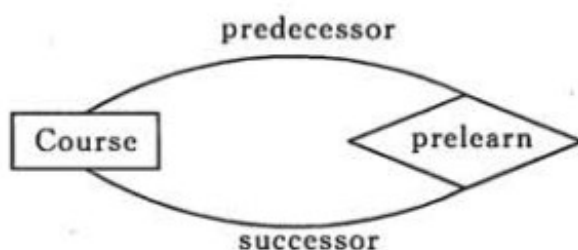
包含一对一，一对多，多对多三种。

如果 A 到 B 是一对多关系，那么画个带箭头的线段指向 B；如果是一对一，画两个带箭头的线段；如果是多对多，画两个不带箭头的线段。下图的 Course 和 Student 是一对多的关系。



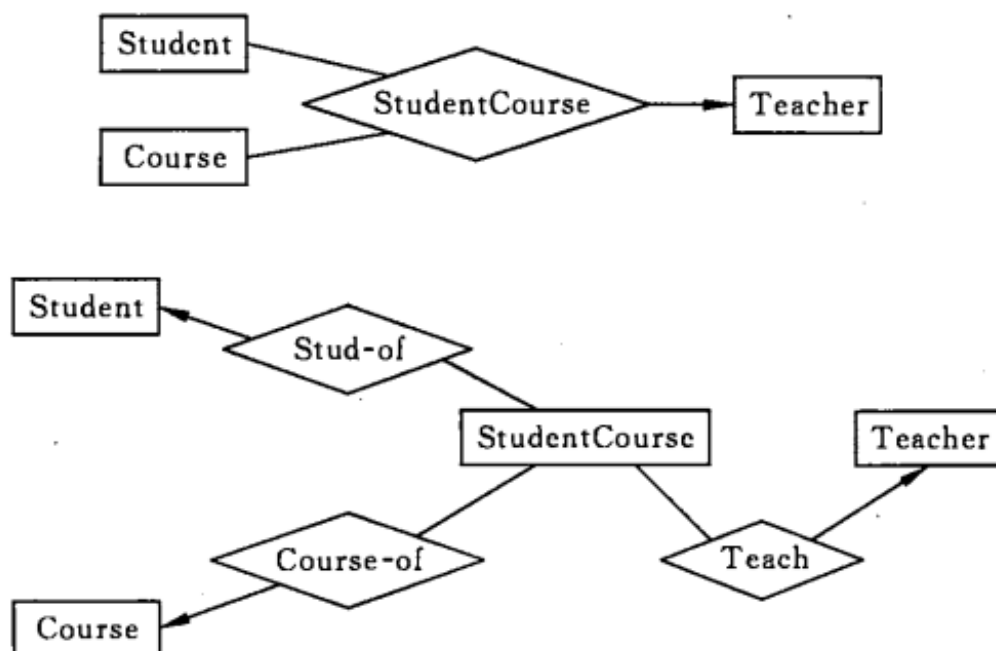
### 表示出现多次的关系

一个实体在联系出现几次，就要用几条线连接。下图表示一个课程的先修关系，先修关系出现两个 Course 实体，第一个是先修课程，后一个是后修课程，因此需要用两条线来表示这种关系。



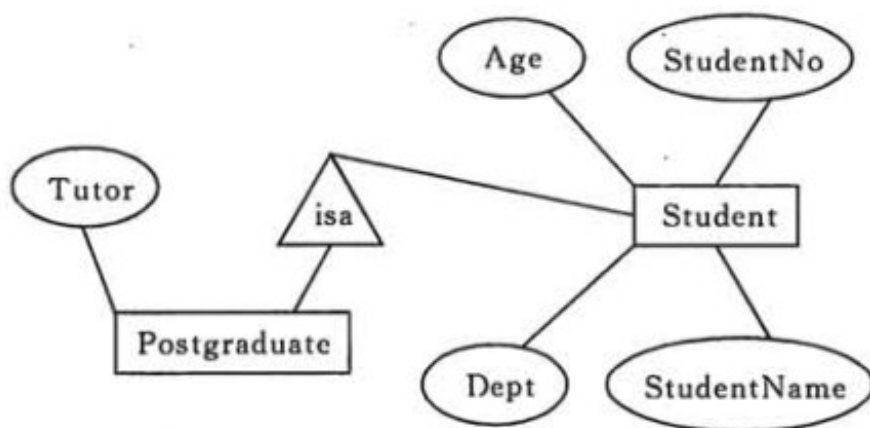
### 联系的多向性

虽然老师可以开设多门课，并且可以教授多名学生，但是对于特定的学生和课程，只有一个老师教授，这就构成了一个三元联系。



## 表示子类

用一个三角形和两条线来连接类和子类，与子类有关的属性和联系都连到子类上，而与父类和子类都有关的连到父类上。



## 参考资料

- AbrahamSilberschatz, HenryF.Korth, S.Sudarshan, 等. 数据库系统概念 [M]. 机械工业出版社, 2006.
- 施瓦茨. 高性能 MySQL(第3版)[M]. 电子工业出版社, 2013.
- 史嘉权. 数据库系统概论[M]. 清华大学出版社有限公司, 2006.
- [The InnoDB Storage Engine](#)
- [Transaction isolation levels](#)
- [Concurrency Control](#)
- [The Nightmare of Locking, Blocking and Isolation Levels!](#)
- [Database Normalization and Normal Forms with an Example](#)
- [The basics of the InnoDB undo logging and history system](#)
- [MySQL locking for the busy web developer](#)
- [深入浅出 MySQL 和 InnoDB](#)
- [InnoDB 中的事务隔离级别和锁的关系](#)

github: <https://github.com/sjsdfg/Interview-Notebook-PDF>