

IN查询分析:

`select * from A where id in (select id from B)`

等价于: 1. `select id from B` ----->先执行in中的查询。

2. `select * from A where A.id = B.id`

以上in()中的查询只执行一次, 它查询出B中的所有id并缓存起来, 然后检查A表中查询出的id在缓存中是否存在, 如果存在则将A的查询数据加入到结果集中, 直到遍历完A表中所有的结果集为止。

1. 当A表中有100条记录, B表中有1000条记录, 那么最多可能遍历 $100 * 1000$ 次, 效率很差。

2. 当A表中有1000条记录, B表中有100条记录, 那么最多可遍历 $1000 * 100$ 次, 内循环次数减少, 效率大大提升。

虽然 $100 * 1000 = 1000 * 100$, 但当主查询表A有索引时, $1000 * 100$ 的效率要高于 $100 * 1000$

结论: IN() 查询适合B表数据比A表数据小的情况, IN() 查询是从缓存中取数据。

Exists查询分析:

`select * from A where exists (select * from B where B.id = A.id)`

以上查询等价于:

1. `select * from A;`

2. `select * from B where B.id = A.id`

exists是对外表做loop循环, 每次loop循环再对内表进行查询, 内表的查询不会将结果进行缓存, 只会返回一个布尔值true或者false (可以用hash), 因为外表的循环不可避免, 所以外表尽量的小。

结论: exists适合B表比A表数据大的情况

当A表数据和B表数据一样大的时候, in与exists效率差不多, 可任选一个。