目录

问题 1：估算对关系 x 做排序的代价

使用两阶段多路合并排序（Two-Pass Multiway Merge Sort）算法来估算：

1. **第一阶段（分割阶段）：**
    1. 将关系 x 分割成多个小于或等于内存大小的部分，每部分包含 `M` 块，因此共需要分成 `ceil(bx/M)` 个部分。
    2. 每个部分被读入内存，进行内部排序，然后写回磁盘。这一过程的读写代价是 `2 * bx`（每块读和写各一次）。
2. **第二阶段（合并阶段）：**
    1. 假设内存能同时容纳所有分割后的部分的第一个元素即块，即 `ceil(bx/M) <= M`，则可以在一轮合并中完成排序。
    2. 每轮合并涉及到所有子表的一个块的读和写操作。

读和写每个块两次，即 `2 * bx`，读写代价总和则是 `4* bx`。

问题 2：ACID 特性的含义及其在 DBMS 中的实现

ACID 是数据库管理系统事务的四个基本特性：

1. 原子性（Atomicity）： 原子性意味着事务是一个不可分割的最小工作单元，整个事务中的所有操作要么全部提交成功，要么全部失败回滚。这意味着事务中的操作如果失败，将回滚到事务开始之前的状态，以确保数据的一致性。在 MySQL 中，通过支持事务的提交和回滚来实现原子性。
2. 一致性（Consistency）： 一致性是指事务必须使数据库从一个一致性状态变换到另一个一致性状态。一致性状态的含义在于，如果事务开始之前和结束之后数据库处于一致性状态，那么这个事务就是正确的。例如，在银行转账的例子中，一致性确保了即使在执行某几条语句之间时系统崩溃，信用卡

账号也不会有任何损失，因为事务最终没有提交，所以事务中所做的修改也不会保存到数据库中，保证数据一致性。

3. 隔离性（Isolation）： 隔离性意味着多个事务并发执行时，一个事务的执行不影响其他事务的执行。在 MySQL 中，通过使用事务的隔离级别来实现隔离性。例如，使用读未提交（Read Uncommitted）、读提交（Read Committed）、可重复读（Repeatable Read）或串行化（Serializable）等隔离级别，可以避免多个事务并发执行时的相互干扰。

4. 持久性（Durability）： 持久性意味着一旦事务提交，则其所做的修改就会永久保存到数据库中，之后的其他操作或故障都不会对事务的结果产生影响。在 MySQL 中，通过将数据写入磁盘以及使用日志文件来实现持久性。

问题 3：关系 r 和 s 的自然连接场景分析

1. 嵌套循环连接（Nested Loop Join）：

   通过双层循环比较数据来获得结果，相当于两表直接做笛卡尔积，逐条记录比较，非常消耗性能；在嵌套循环中，内表被外表驱动，外表返回的每一行都要在内表中检索找到与它匹配的行，因此使用简单嵌套循环，整个查询返回的结果集不能太大。

2. 块嵌套循环连接（Block Nested Loop Join）：

   当两个表都相对较小，可以部分地加载到内存中时适用，即一次性可以加载一个块、多条记录通过一次性缓存外层表的多条数据，以此来减少内层表的扫表次数，从而达到提升性能的目的。

3. 索引嵌套循环连接（Index Nested Loop Join）：

   说 Index Nested-Loop Join 通过外层表匹配条件 直接与内层表索引进行匹配，避免和内层表的每条记录去进行比较， 这样极大的减少了对内层表的匹配次数，从原来的匹配次数=外层表行数 * 内层表行数,变成了 外层表的行数 * 内层表索引的高度，极大的提升了 join 的性能。但是该算法的前提是匹配的字段必须建立了索引。

4. 哈希连接（Hash Join）：

   散列连接是大数据集连接时的常用方式，优化器使用两个表中较小的表（或数据源）利用连接键在内存中建立散列表，然后扫描较大的表并探测散列表，找出与散列表匹配的行。

   这种方式适用于较小的表完全可以放于内存中的情况，这样总成本就是访问两个表的成本之和。但是在表很大的情况下并不能完全放入内存，这时优化

器会将它分割成若干不同的分区，不能放入内存的部分就把该分区写入磁盘的临时段，此时要有较大的临时段从而尽量提高 I/O 的性能。

散列连接不需要索引，这是它最大的优势，并且与循环嵌套连接相比，散列连接更容易处理大结果集，但是它只能用作等值连接，同时 Hash Join 由于需要计算和存储 hash 值，会消耗更多的 cpu 资源和内存。Hash Join 是一种典型的用空间和 cpu 资源来换取时间的查询方法。

5. 归并连接（Merge Join）：

适用于已经排序或可以高效排序的表。此方法将两个表按排序键归并，行与行之间比较以找到匹配项；对于两个表都很大且无索引的情况，Merge join 要比 Nested Loop Join 要好。