

数据库第六次作业

22371437 张智威

1.什么是索引?

答：索引是主表上的一种辅助数据结构，它是某个表中一列或若干列值的集合和相应的指向表中物理标识这些值的数据页的逻辑指针清单，对数据库表中一个或多个列的值进行排序，从而用于提高主表的查询速度。如果按照特定列的值来查找数据，则与在表中搜索所有行相比，使用索引可以减少磁盘的 I/O 操作此时，从而有助于更快地获取信息。

2.简述B树索引，Hash索引及聚簇索引的实现原理和优缺点。

- B+树索引
 - 实现原理：将索引键组织成一颗平衡树，即从树根到树叶的所有路径一样长，数据(指向基本表记录存储位置的指针存储在叶节点)，最底层的叶节点包含每个索引键和指向被索引行的指针。设计通道从而允许叶节点间的平行查询，每个叶节点和磁盘页面大小一致。
 - 优点：B+树的插入删除操作不会引起过多 I/O 操作，插入数据后B+树仍是平衡的，且保持了很好的性能，更新维护的代价较小。一般B+树保持在三层，使得很大的数据集也只需要三次 I/O 操作就能获取数据。
 - 缺点：当要查询的记录数占记录总数的百分比非常大的时候，不用索引比用索引更快。
- Hash索引
 - 实现原理：根据给定索引值，用一种算法将记录分散存储到多个“桶”中（一般一个桶就是一个数据块，块中内容用一次磁盘操作就可以读取到内存中），如果记录数很多，可以增加一个溢出块到桶的链上。当要查找记录时，用相同算法算出该记录所在的桶，读取整个桶的数据到内存中，然后在桶中顺序查找要找的记录。
 - 优点：Hash索引只需一次磁盘 I/O 操作就可以查到数据记录。在等值查找时速度很快。
 - 缺点：Hash索引无法用于范围查找。不适合在重复值很多的列上建立Hash索引。重构代价很大。
- 聚簇索引
 - 实现原理：大多数关系表以堆组织表的形式存放。建立聚簇索引后，数据根据索引中键值的逻辑顺序形成表中相应行的物理顺序，形成索引组织表。
 - 优点：在聚簇索引列上的查询速度比B+树索引快，在使用包含范围检查或group by或order by的查询时，查询速度很快。
 - 缺点：一张表只能有一个聚簇索引，DML频繁的表如果使用聚簇索引，会带来大量索引数据维护的开销。