

海量数据查询优化Oracle认证考试 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/645/2021_2022__E6_B5_B7_E9_87_8F_E6_95_B0_E6_c102_645438.htm 数据库系统是管理信息系统的核心，基于数据库的联机事务处理（OLTP）以及联机分析处理(OLAP)是银行、企业、政府等部门最为重要的计算机应用之一。从大多数系统的应用实例来看，查询操作在各种数据库操作中所占据的比重最大，而查询操作所基于的SELECT语句在SQL语句中又是代价最大的语句。举例来说，如果数据的量积累到一定的程度，比如一个银行的账户数据库表信息积累到上百万甚至上千万条记录，全表扫描一次往往需要数十分钟，甚至数小时。如果采用比全表扫描更好的查询策略，往往可以使查询时间降为几分钟，由此可见查询优化技术的重要性。在应用项目的实施中发现，许多程序员在利用一些前端数据库开发工具（如PowerBuilder、Delphi等）开发数据库应用程序时，只注重用户界面的华丽，并不重视查询语句的效率问题，导致所开发出来的应用系统效率低下，资源浪费严重。因此，如何设计高效合理的查询语句就显得非常重要。本文以应用实例为基础，结合数据库理论，介绍查询优化技术在现实系统中的运用。分析问题许多程序员认为查询优化是DBMS（数据库管理系统）的任务，与程序员所编写的SQL语句关系不大，这是错误的。一个好的查询计划往往可以使程序性能提高数十倍。查询计划是用户所提交的SQL语句的集合，查询规划是经过优化处理之后所产生的语句集合。DBMS处理查询计划的过程是这样的：在做完查询语句的词法、语法检查之后，将语句提交给DBMS

的查询优化器，优化器做完代数优化和存取路径的优化之后，由预编译模块对语句进行处理并生成查询规划，然后在合适的时间提交给系统处理执行，最后将执行结果返回给用户。在实际的数据库产品(如Oracle、 Sybase等)的高版本中都是采用基于代价的优化方法，这种优化能根据从系统字典表所得到的信息来估计不同的查询规划的代价，然后选择一个较优的规划。虽然现在的数据库产品在查询优化方面已经做得越来越好，但由用户提交的SQL语句是系统优化的基础，很难设想一个原本糟糕的查询计划经过系统的优化之后会变得高效，因此所写语句的优劣至关重要。下面重点说明改善查询计划的解决方案。

解决问题 下面以关系数据库系统Informix为例，介绍改善用户查询计划的方法。

1. 合理使用索引 索引是数据库中重要的数据结构，它的根本目的就是为了提高查询效率。现在大多数的数据库产品都采用IBM最先提出的ISAM索引结构。索引的使用要恰到好处，其使用原则如下：
在经常进行连接，但是没有指定为外键的列上建立索引，而不经常连接的字段则由优化器自动生成索引。
在频繁进行排序或分组（即进行group by或order by操作）的列上建立索引。
在条件表达式中经常用到的不同值较多的列上建立索引，在不同值少的列上不要建立索引。比如在雇员表的“性别”列上只有“男”与“女”两个不同值，因此就无必要建立索引。如果建立索引不但不会提高查询效率，反而会严重降低更新速度。
如果待排序的列有多个，可以在这些列上建立复合索引（compound index）。
使用系统工具。如Informix数据库有一个tbcheck工具，可以在可疑的索引上进行检查。在一些数据库服务器上，索引可能失效或者

因为频繁操作而使得读取效率降低，如果一个使用索引的查询不明不白地慢下来，可以试着用**tbcheck**工具检查索引的完整性，必要时进行修复。另外，当数据库表更新大量数据后，删除并重建索引可以提高查询速度。

2. 避免或简化排序应当简化或避免对大型表进行重复的排序。当能够利用索引自动以适当的次序产生输出时，优化器就避免了排序的步骤。以下是一些影响因素：

- 索引中不包括一个或几个待排序的列；
- `group by`或`order by`子句中列的次序与索引的次序不一样；
- 排序的列来自不同的表。

为了避免不必要的排序，就要正确地增建索引，合理地合并数据库表（尽管有时可能影响表的规范化，但相对于效率的提高是值得的）。如果排序不可避免，那么应当试图简化它，如缩小排序的列的范围等。

3. 消除对大型表行数据的顺序存取 在嵌套查询中，对表的顺序存取对查询效率可能产生致命的影响。比如采用顺序存取策略，一个嵌套3层的查询，如果每层都查询1000行，那么这个查询就要查询10亿行数据。避免这种情况的主要方法就是对连接的列进行索引。例如，两个表：学生表（学号、姓名、年龄……）和选课表（学号、课程号、成绩）。如果两个表要做连接，就要在“学号”这个连接字段上建立索引。还可以使用并集来避免顺序存取。尽管在所有的检查列上都有索引，但某些形式的where子句强迫优化器使用顺序存取。下面的查询将强迫对orders表执行顺序操作：

```
SELECT *
FROM orders WHERE (customer_num=104 AND order_num
```

100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com