

oracle认证:Oracle中的成本评估机制解析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_ti2020/584/2021_2022_oracle_E8_AE_A4_E8_c102_584569.htm 数据库管理员编写的SQL语句并不会马上得到解析，而是需要先经过Oracle数据库的优化。优化程序会将最初的SQL语句转换成可以返回相同结果、但执行起来更有效率的SQL语句。如可以通过视图合并、谓词下推等等来减少扫描、连接或者聚合作业，来提高SQL语句的执行性能。但是Oracle数据库毕竟不是人，不会像人的大脑那样去判断什么样的SQL语句才具有更高的执行效率。为此在数据库设计的时候，必须要设置一个评判的标准，让数据库优化器能够以这个规则去判断，去选择更有效率的SQL语句。在Oracle数据库中，提供了非常丰富的数据库结构、分区和索引技术，很多的连接方法。为此通常一个SQL语句经过优化器优化之后会生成几个执行计划。虽然这些执行计划采用不同的访问路径、连接方法、连接顺序的组合方式，可以用不同的方法访问和处理数据，但是最后产生的结果都是相同的。其实数据库最终只执行一条执行计划。那么数据库会执行哪一条呢?这就要看Oracle数据库中定义的成本评估规则了。也就是说，要评估这些执行计划的成本，并选择其中成本最低的执行计划，优化程序依赖于组成SQL语句执行的单独操作的成本评估。这些评估要尽量的准确。在Oracle数据库中的成本评估模型，包含了数据库数据结构、对象级和系统统计数据访问方法、性能信息等等。那么这个成本模型到底是如何运作的呢?笔者就以Oracle数据库推荐的成本模型CPU IO成本模型来谈谈其运作机制以及维护的要点。 一、根据执行

时间来评估运行成本。在判断执行计划的优劣时，根据其执行所需要花费的时间来判断无疑是最快速的。而影响这个执行计划的时间主要有两个方面的因素。一是IO操作的数据和类型。因为数据库的大部分操作都要涉及到硬盘的读写。而从硬盘中读取速度与比内存中读取速度要慢的多。为此IO操作的数据和类型对于SQL执行计划的时间具有很大的影响。二是CPU的周期数，这也是影响SQL语句执行计划时间的重要因素。当然还有其它方面的因素，不过其他因素对于执行计划时间关系不大。为此为了在最短时间内估计出给定查询的执行时间，CPU IO成本模型下的查询优化程序会估算IO操作的数据和类型，并且在查询执行过程中数据库可能执行的CPU周期数。然后利用系统统计数据将这些CPU周期数和IO操作转换为SQL执行计划时间。然后根据选择一个执行计划时间最短的执行计划来对数据库进行操作。可见在CPU IO模型中，其只考虑CPU周期数与IO操作这两个因素。不过对于大部分数据库应用来说，这已经足够了。从这个成本模型中，也给我们提高数据库性能做了一些暗示。CPU周期数的话我们很难改善，除非采用性能更好的CPU，但是这需要花费比较大的代价。但是改善IO操作的话，还是有比较多的方法。同可以通过磁盘阵列技术，让系统同时向多块硬盘中存取数据以缩短IO操作的时间，这倒是可行的。虽然Oracle数据库的成本模型是其内部操作的，但是深入研究其成本统计的方法，对我们通过改善硬件配置来提高数据库性能也是很有帮助的。

二、CPU IO成本模型可以重新排序谓词。

在Oracle数据库中，成本模型有很多，为什么其大力推荐CPU IO成本统计模型呢？这其中很重要的原因就是采用CPU

IO的话，数据库在查询时可以采用重新排序谓词。笔者在文章一开头就谈到过，可以通过视图合并、谓词下推等等来减少扫描、连接或者聚合作业，来提高SQL语句的执行性能。而重新排序为此是其提高SQL语句执行性能的一个重要举措。这个措施只有在CPU IO成本模型下才能够实现。因为重排序谓词只有在每个谓词成本都可以估算的情况下才可以进行。而只有在CPU IO成本模型下才会对每个谓词进行成本估算，其他的成本统计模型都不会。所以采用这个成本模型的话，那么最终优化的执行计划效果会比较明显。根据一些官方的资料，采用CPU IO成本模型，比采用其他成本模型，如IO成本模型，会把执行计划的时间缩到70%。这是一个很可观的性能改善数字。为此对于一些性能要求比较高的数据库，可以考虑把成本模型设置为CPU IO。但是要注意天下没有免费的午餐。由于在CPU IO成本模型中，需要估算每个谓词的成本，需要考虑CPU的周期数与IO操作的数量与类型，这会增加数据库服务器一些额外的开销。可见也并不是在任何情况下都适合采用CPU IO成本统计模型。如在一些比较小的数据库中采用这种成本模型的话，反而起到的效果不会很明显。

三、通过取样来统计数据。在统计学中，往往会利用取样来节省统计的时间。这在Oracle数据库中也得到了充分的应用，毕竟如果要对全部数据进行统计需要耗费比较多的时间。在Oracle数据库中通过检查相关的数据样本并对其取样来收集需要的统计数据。这个样本可能是动态的，也可能是静态的。样本数据与查询在同一个事物处理中发生，可以与并行运行结合使用。因为Oracle数据库中的并行执行体系结构允许任何SQL语句以任何等级的并行度去执行。为了得到具有代表

性的样本与比较准确的统计数据，Oracle数据库会根据需要自动去获取相关的样本。但是为了避免一些不良样本对其造成的负面影响，Oracle数据库管理员仍然有不少的工作要做。如在把服务器投入到生产用之前，要对数据库中的数据进行清空作业。把数据库中的数据，包括日志信息等等全部删除。也就是说，在投入运行之前，数据库已经是干净的。在设计过程中因为测试所产生的一切信息都应该删除。

四、选择合适的优化模式。

在数据库初始化的过程中，数据库管理员不仅可以选择合适的成本模式，而且还可以选择合适的优化模式来影响优化程序的选择。在Oracle数据库中，提供了两种不同的优化模式。这两种优化模式分别为返回查询结果前N行时间的最小化与返回全部数据时间的最小化。返回查询结果前N行时间的最小化优化模式是指获得返回结合最先几行的最佳反应时间，这个不仅仅跟数据库有关，而且还跟应用程序的设计有关。返回全部数据时间的最小化的优化模式是指获得最佳的吞吐量。那么数据库管理员该如何选择这两种优化模式呢？笔者的下面这些建议虽然不能说“放之五湖四海皆准”，但是至少大家可以拿来参考。

一是在一些事务性的数据库中，最好采用返回查询结果前N行时间最小化优化模式。因为在事务处理系统中，虽然时间的推移，其数据会越来越多。但是用户平时要用到的数据可能是最新的一些记录，如最近半年或者一年的记录。而不需要全部的记录。为此采用这种优化模式的话，可能会取得比较好的效果。

二是数据库如果用来做决策支持系统或者数据仓库中，则最好采用返回全部数据时间最小化优化模式。因为在这些应用下，用户出于统计或者其他的目的，往往需要用到表中的全部记

录。此时在选择优化模式的时候，更应该关注数据库系统的吞吐量。所以在这些应用中，采用返回全部数据时间最小化优化模式比较适合。因为这种优化模式，可以提供最佳的吞吐量，能够满足数据库系统对整表数据统计的需要。三是需要注意，采用返回查询结果前N行时间最小化优化模式的时候，不仅跟其返回数据的多少有关，而且还跟应用系统的设计与操作系统有关。如同一个数据库系统(含有同样的数据与架构)，部署在不同的操作系统上(硬件配置相同)或者不同语言开发的应用系统上，其效果不经相同。为此在采用这种优化模式的时候，往往还需要获得系统工程师的帮助，通过调整系统配置来改善这种优化模式的效果。当数据库管理员苦苦想不出该如何改善数据库性能的时候，这个就非常有用。更多优质资料尽在百考试题论坛 百考试题在线题库 oracle认证更多详细资料 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