

# SQL 执行计划与直方图关系研究<sup>①</sup>

张开基

(中石化石油工程设计有限公司, 东营 257000)

**摘要:** sql 语句调优是数据库性能调优的重要方面. 要达到同样的执行结果, sql 语句有多种写法, 不同的写法其性能差别很大. 即使同一个 sql 语句, oracle 也有多种途径去执行, 即有多个执行计划. oracle 比较这多个执行计划的性能优劣, 耗费资源多少, 来选择最优的执行计划. oracle 在评估各个执行计划的性能时, 需要借助 sql 语句执行的环境, 即统计信息, 来计算出每个执行计划耗费资源的多少. 因此, 尽可能收集准确的统计信息, 对于 oracle 能否选择最优的执行计划, 至关重要. 其中, 直方图的收集与否起着很重要的作用. 本文通过实验来验证直方图对 sql 执行计划的影响, 从而明确何种情况下需要收集直方图.

**关键词:** sql 调优; 执行计划; 统计信息; 直方图

引用格式: 张开基. SQL 执行计划与直方图关系研究. 计算机系统应用, 2017, 26(10): 246-250. <http://www.c-s-a.org.cn/1003-3254/6008.html>

## Research on Relationship between SQL Execution Plan and Histogram

ZHANG Kai-Ji

(Sinopec Petroleum Engineering Corporation, Dongying 257000, China)

**Abstract:** Sql statement tuning is an important aspect of database performance tuning. To achieve the same effects, sql statement has a variety of wording, with different performance for the different wording. Even with only one sql statement, the oracle also has a variety of ways to implement. That is, there are multiple execution plans. The oracle compares the performance of these multiple execution plans, the cost of resources, to select the optimal execution plan. In assessing the performance of each implementation plan, the oracle needs the implementation of sql statement with the environment, that is statistical information to calculate the cost of the number of resources for each implementation plan. Therefore, it is critically important for the oracle to choose the best implementation plan to collect as much as possible accurate statistical information. Among them, the collection of the histogram plays a very important role. The experiment verifies the impact of the histogram for the sql implementation plan, which clears the circumstances under which it needs to collect histograms.

**Key words:** sql tuning; implementation plan; statistics; histogram

### 1 引言

对 oracle 数据库的性能调优是数据库管理员日常工作的重要内容. 调优方法有多种, 包括对数据库内存的调整, 对数据库数据存储的优化等等. 其中, sql 调优是数据库管理员做的最多的一项工作, 也是效果比较明显的一种调优方法<sup>[1]</sup>.

应用开发人员在开发过程中, 往往只关注执行结

果是否正确, 而忽略了不同的实现方法之间可能存在的性能差异. 因此, 基于 oracle 应用系统的很多性能问题, 是由应用系统的 sql 语句性能较差引起的, 所以, 对 sql 语句的调优, 往往是数据库管理员性能调优的重要手段.

在 sql 语句的调优过程中, 通过 sql 执行计划来了解 sql 语句的性能如何, 是必须掌握的内容. 而 oracle

<sup>①</sup> 收稿时间: 2017-01-22; 采用时间: 2017-02-23

对 sql 执行计划的选择,受多种因素的影响,比如有无索引,有无统计信息,当前的优化器模式是哪种等等<sup>[2]</sup>.其中, `cursor_sharing` 参数值的选择,直方图的有无,直接影响了 oracle 对 sql 执行计划的确定. 搞清直方图对 sql 执行计划的影响,对于 sql 语句的调优,至关重要.

## 2 SQL 执行计划与直方图简介

### 2.1 执行计划描述

为了执行一条 sql 语句, oracle 需要执行某些步骤的操作,每一步骤可能是从数据库中物理检索数据行,或者用某种方法准备数据行,供发出语句的用户使用<sup>[3]</sup>. Oracle 用来执行语句的这些步骤的组合即为执行计划. 执行计划是 sql 优化中最为复杂也是最为关键的部分,只有知道了 oracle 在内部到底是如何执行该 sql 语句的,才能知道优化器选择的执行计划是否是最优的.

### 2.2 执行计划的获取

获取执行计划的方法有很多,常用的有两种,一种是利用 `autotrace` 命令,前提是用户拥有 `plustrace` 角色,这样用户就可以利用 `set autotrace` 命令来执行 sql 语句查看执行计划. 这种方式在 oracle9i 及以下版本中是理论上的执行计划,不一定是 oracle 实际选择的,而在 oracle10g 之后,这种方式获取的执行计划已比较准确,本文采取的是这种方式;一种是利用 `sql_trace` 跟踪用户会话获取用户的跟踪文件,跟踪文件中详细列出了用户执行的 sql 语句和 sql 语句的执行计划,这样获取的执行计划是 oracle 实际选择的执行计划. 另外,还可以查询动态性能视图,从内存中直接获取语句的执行计划.

### 2.3 直方图描述

sql 执行计划的确定,受数据库对象统计信息的影响. 统计信息主要是描述数据库中表、索引的大小、规模、数据分布状况等的一类信息. 比如,表的行数、块数、平均每行的大小、索引的 leaf blocks、索引字段的行数、不同值的大小等,都属于统计信息. CBO 正是根据这些统计信息数据,计算出不同访问路径下,不同 join 方式下,各种执行计划的成本,最后选择出成本最小的执行计划<sup>[4]</sup>.

在 oracle 中直方图是一种对数据分布质量情况进行描述的工具. 它会按照某一列不同值出现数量的多少,以及出现的频率高低来绘制数据的分布情况,以便能够指导优化器根据数据的分布做出正确的选择. 在某些情况下,表列中的数值分布将影响优化器使用索引还是执行全表扫描. 当 where 子句的值具有不成比

例数量的数值时,将出现这种情况,使得全表扫描比索引访问的成本更低. 这种情况下如果 where 子句的过滤谓词列上有一个合理正确的直方图,将会对优化器做出正确的选择发挥巨大的作用,使得 SQL 语句执行成本最低从而提升性能.

通俗地说,oracle 中的直方图就是描述表中列值的数据分布情况. 在表列数据的唯一值分布不均匀的情况下,收集直方图信息可以使得 oracle 根据数据分布情况选择更准确的执行计划. 若是数据分布均匀,直方图的收集没有意义.

## 3 SQL 执行计划与直方图关系实验

### 3.1 实验意义

对于同一个查询,可能有几个执行计划都符合要求,都能得到符合条件的数据. 例如,参与连接的表可以有多种不同的连接方法,这取决于连接条件和优化器采用的连接方法. 为了在多个执行计划中选择最优的执行计划,优化器必须使用一些实际的指标来衡量每个执行计划使用的资源(I/O 次数、CPU 等),这些资源也就是我们所说的代价(cost). 如果一个执行计划使用的资源多,我们就说使用执行计划的代价大. 以执行计划的代价大小作为衡量标准,优化器选择代价最小的执行计划作为真正执行该查询的执行计划,并抛弃其它的执行计划.

由于一系列因素都会影响语句的执行,优化器综合权衡各个因素,在众多执行计划中选择最佳的执行计划. 但是,很多情况下,优化器不能得到较真实的执行环境,就有可能选择次优的执行计划,这样,oracle 性能就会受到影响. 尤其在数据量较大的环境下,不能选择最优的执行计划,会使得 oracle 极其耗费系统资源,影响系统的响应时间,继而影响用户体验.

因此,尽可能准确的收集 oracle 对象的统计信息,搞清在不同的数据库环境下直方图的收集与否,对于提高 sql 语句执行的性能,至关重要.

### 3.2 实验过程

实验分以下几种情况:

实验以上六种组合下,sql 语句执行计划受直方图的影响,六种组合保证不会相互影响.

实验环境: oracle 版本: 10.2.0.4, 表空间本地管理, 段自动管理. 创建实验表 `tab_1`, 插入实验数据.

组合一: (`cursor_sharing` 为 `exact`, 实验字段上无直方图). 确认目前数据库 `cursor_sharing` 值为 `exact`, 如图 1.

表 1 参数值组合

	cursor_sharing	直方图
组合一	exact	无
组合二	exact	有
组合三	force	无
组合四	force	有
组合五	similar	无
组合六	similar	有

```
SOL> show parameter cursor_sharing
NAME                                TYPE      UALUE
cursor_sharing                       string    EXACT
```

图 1 cursor\_sharing 值

查询表 tab\_1 的数据分布情况:



图 2 表 tab\_1 数据分布情况

查询表 tab\_1, 发现表 tab\_1 数据分布不均匀, 列 B 唯一值个数是 10 个, 其中, 值为 5 的记录共有 9991 行, 其它只有一行. 我们已在列 B 上创建索引. 实验在这种情况下, 有无直方图对执行计划的影响.

利用 sql 语句分析表, 收集表的统计信息, 不收集直方图, 然后利用 autotrace 命令查看统计信息, 执行结果及执行计划如图 3.

```
SQL> select * from tab_1 where b='5';
```

从结果可以看到, 查询 b 值等于 5 的记录, 在无直方图的情况下, sql 执行走了索引, 因为 b 值等于 5 的记录共有 9991 行, 这种情况下走全表扫描比走索引效率更高, 在无直方图的情况下, sql 走了索引, 用了性能较差的执行计划.

在实验 b=3 的情况下, sql 如何选择执行计划:

这种情况下, sql 照例走了索引.

总结: 在数据分布不均匀的情况下, 若是不收集直方图信息, oracle 无法获得数据的分布情况, 因此不能得到最佳的执行计划.



图 3 组合一执行计划 1

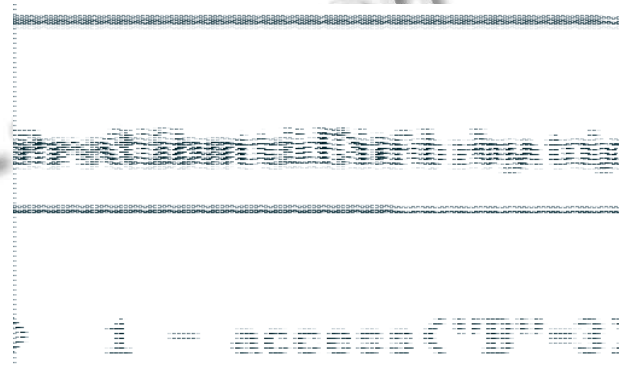


图 4 组合一执行计划 2

组合二: (cursor\_sharing 为 exact, 实验字段上有直方图).

利用 sql 语句分析表, 收集表的统计信息, 并收集直方图, 然后利用 autotrace 命令查看统计信息, 执行结果及执行计划如图 5.

```
SQL> select * from tab_1 where b='5';
```

在实验 b=3 的情况下, sql 如何选择执行计划:

总结: 这种情况下, 查询 b 等于 5 和 b 等于 3, oracle 选择了不同的执行计划. B 等于 5 走全表扫描, b 等于 3 走了索引. 这是符合理论的. 因为 oracle 收集了直方图信息, 获取了表数据的具体分布情况, 因此能够根据执行计划的代价大小, 比较准确的获取最佳的执行计划.

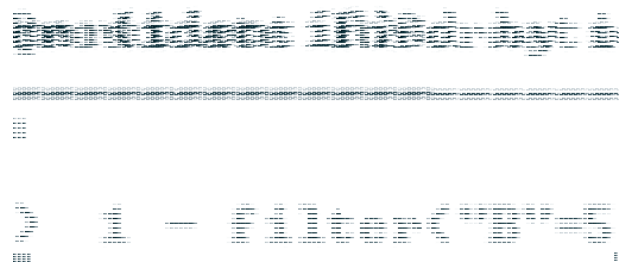


图 5 组合二执行计划 1

以上两种情况下, cursor\_sharing 为 exact, 在字段



数据分布不均匀的情况下,有直方图可得到预期的执行计划;无直方图,oracle无法判断数据分布情况,无论检索何值都走了索引.

组合三: (cursor\_sharing 为 force, 实验字段上无直方图). 首先修改 cursor\_sharing 值为 force:

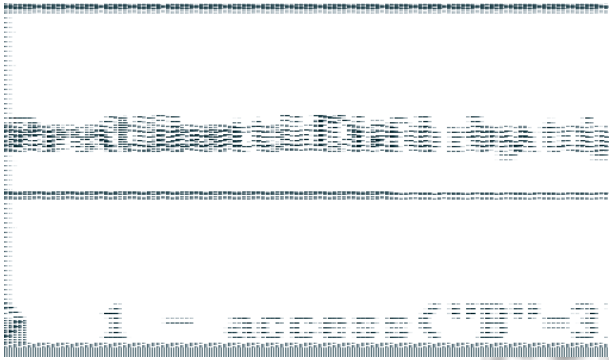


图6 组合二执行计划2



图7 修改 cursor\_sharing 值

cursor\_sharing 取值 force, oracle 会强制 sql 绑定变量, 无论实验字段上有无直方图.

收集统计信息, 不收集直方图, 然后利用 autotrace 命令获得列 b 等于 5 和等于 3 两种值下的执行计划如图 8、9.

```
SQL> select * from tab where b='5';
```

这种情况, 不论列数据值如何分布, 执行计划无法获得数据值的分布情况, sql 也都走了索引.

组合四: (cursor\_sharing 为 force, 实验字段上有直方图).

收集统计信息, 并收集直方图, 然后利用 autotrace 命令获得列 b 等于 5 和等于 3 两种值下的执行计划如图 10、11.



图8 组合三执行计划 1



图9 组合三执行计划 2

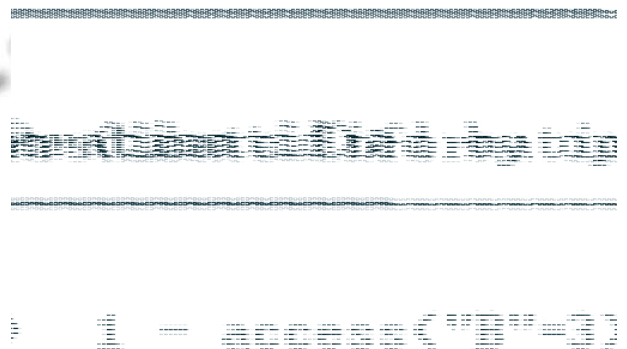


图10 组合四执行计划 1

```
SQL> select * from tab where b='5';
```

这种方式下, 无论先执行哪条语句, 都能得到预期的执行计划.

在 cursor\_sharing 值为 force 的情况下, 和 cursor\_sharing 值为 exact 的情况下一样. sql 根据统计信息的直方图收集情况, 来选择不同的执行计划.

组合五: (cursor\_sharing 为 similar, 实验字段上无直方图). 首先修改 cursor\_sharing 值为 similar:



图11 组合四执行计划 2

理论上, cursor\_sharing 的取值, 当表的字段被分析过存在直方图的时候, similar 的表现和 exact 一样; 当

表的字段没被分析,不存在直方图的时候, similar 的表现和 force 一样. 这样避免了一味地如 force 一样转换成变量形式. 因为有直方图的情况下转换成变量之后容易产生错误的执行计划, 没有利用上统计信息, 因此 similar 综合了两者的优点. 实验如图 12 所示.



图 12 修改 cursor\_sharing 值

收集统计信息, 不收集直方图, 然后利用 autotrace 命令获得列 b 等于 5 和等于 3 两种值下的执行计划如图 13、14.



图 13 组合五执行计划 1

SQL> select \* from tab where b='5';  
两种情况都走了索引, 与理论相符.

组合六: (cursor\_sharing 为 similar, 实验字段上有直方图). 收集统计信息, 并收集直方图, 然后利用 autotrace 命令获得列 b 等于 5 和等于 3 两种值下的执行计划如图 15、16.

SQL> select \* from tab where b='5';

cursor\_sharing 为 similar, 实验字段上有直方图, 优化器计算出了最优的执行计划, 查询 b 等于 5 和等于 3 的数据, 走了不同的执行计划.

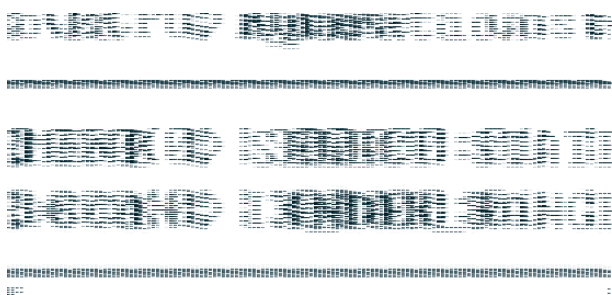


图 14 组合五执行计划 2

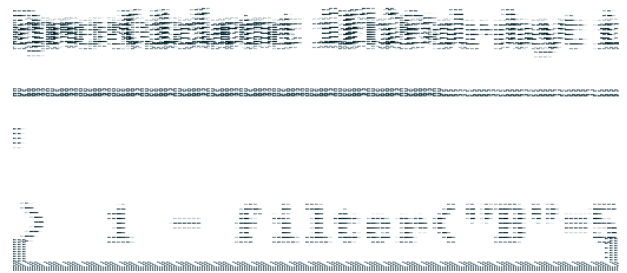


图 15 组合六执行计划 1



图 16 组合六执行计划 2

#### 4 结语

通过以上实验可以得出结论: cursor\_sharing 无论取何值, oracle 根据直方图的有无获得的执行计划与理论一致: 有直方图, 可以根据表列值的数据分布情况决定走全表扫描还是走索引; 无直方图, sql 选择索引. 所以, 在日常维护中, 数据库管理员应根据表数据的分布情况, 来决定是否收集直方图. 某一列数据分布不均匀, 应针对这一列收集直方图, 使得数据库的执行计划优化器可以根据列的数据分布情况来选择效率最高的执行计划. 另外, 收集直方图有系统开销, 对于数据分布比较均匀的表, 为节省系统开销, 可以不收集直方图.

#### 参考文献

- 1 曾实. ORACLE 数据库优化技术研究. 科技信息, 2011, (27): 80, 52.
- 2 韩云波, 宋莉. Oracle 性能调整技术研究. 电脑知识与技术, 2010, 6(7): 1554-1556.
- 3 戴小平. Oracle9i 数据库性能调整与优化. 安徽工业大学学报, 2006, 23(3): 315-319.
- 4 高攀, 施蔚然. 基于 Oracle 数据库的 SQL 语句优化. 电脑编程技巧与维护, 2010, (22): 38-39. [doi: 10.3969/j.issn.1006-4052.2010.22.015]