

* 如何在11g上收集 Optimizer 统计信息 (文档 ID 1523938.1)

文档内容

目标

解决方案

[快速重建统计信息的推荐步骤](#)

[关于收集optimizer统计信息的重要文档](#)

[收集对象的统计信息](#)

[使用足够大的sample size](#)

[收集所有对象的统计信息](#)

[收集字段的统计信息/数据非均匀分布列的histogram](#)

[收集分区对象的Global Statistics\(全局统计信息\)](#)

[收集System Statistics\(系统统计信息\)](#)

[从早期版本升级到11g](#)

[缺省设置](#)

[统计信息收集命令的例子](#)

[在单一的表收集统计信息](#)

[收集某一schema下的所有对象的统计信息](#)

[收集database中所有对象的统计信息:](#)

参考

适用于:

Oracle Database - Enterprise Edition - 版本 11.1.0.6 到 11.2.0.3 [发行版 11.1 到 11.2]

Oracle Database - Standard Edition - 版本 11.1.0.6 到 11.2.0.3 [发行版 11.1 到 11.2]

Oracle Database - Personal Edition - 版本 11.1.0.6 到 11.2.0.3 [发行版 11.1 到 11.2]

本文档所含信息适用于所有平台

目标

这篇文档概述了在Oracle 11g中收集CBO (Cost Based Optimizer) 的统计信息的推荐方法。对于其它版本请参见:

[Document 1226841.1 How To: Gather Statistics for the Cost Based Optimizer](#)

注：下面这个文档中有个标题为"Best Practices for Managing Optimizer Statistics"的章节，该章节提供了关于收集统计信息的一些其它建议

[Document:1380043.1 Selected Performance Related Seminars from Oracle Openworld](#)

解决方案

快速重建统计信息的推荐步骤

为快速删除和重建某张表及它的索引的统计信息 (包含非均匀分布列的统计信息) ,推荐如下：

```
exec dbms_stats.delete_table_stats(ownname=>'user_name',-
  tabname=>'table_name',cascade_indexes=>true);

exec dbms_stats.gather_table_stats(ownname=>'user_name',-
  tabname=>'table_name',-
  estimate_percent => DBMS_STATS.AUTO_SAMPLE_SIZE,-
  cascade=>true,-
```

```
method_opt=>'for all columns size AUTO');
```

随后我们会解释这样推荐的原因。同样，后边会有更多收集统计信息的例子。
需要注意的是，从10gR2 开始,下边的方法可以用来恢复统计信息:

[Document 452011.1](#) * Restoring table statistics in 10G onwards

关于收集optimizer统计信息的重要文档

- 这些推荐适用于于大多数的数据库。
- 这些推荐的目的是为了尽可能生成准确的统计信息，为此我们推荐使用的sample size (采样率) 为100%。这是因为sample size 的减少可能会导致生成的统计信息不准确。一般来说使用100%的sample size会使收集信息的时间增长，所以需要考虑调整当前使用的maintenance window，使统计信息的收集能在一个maintenance window完成。
- 如果可能，我们推荐使用预设好的偏好设置使收集不同对象的过程标准化。不同于对不同的对象使用不同的设置，在 Database/Schema/Table 一级设置默认偏好能使设置集中化，且不需要对不同对象单独设置。自动统计信息收集操作也会使用设置好的默认偏好。关于如何设置默认偏好请详见：

[Document 1493227.1](#) How to Change Default Parameters for Gathering Statistics in Oracle 11g

尽管默认值对大多数系统都适用，如果您发现某些特定设置比默认值更适合您的系统，那么请自行添加至默认设置中。

- 收集新的optimizer统计信息一般会保持或改进已存在的执行计划，但是也可能使某些查询的性能下降。所以从10gR1开始系统默认会保留最近30天的统计信息，这期间如果遇到问题可以用如下方法恢复到之前的统计信息:

[Document 452011.1](#) * Restoring table statistics in 10G onwards

- 收集新的optimizer的统计信息可能使shared pool中的cursor无效，所以在数据库负载低的时候执行统计信息收集将是明智的选择，比如maintenance window。
- 除了object statistics (比如业务表和索引的统计信息)，我们推荐对dictionary objects (数据字典对象) 也收集统计信息:

[Document 457926.1](#) How to Gather Statistics on SYS Objects and 'Fixed' Objects?

- 对于非常大的系统，统计信息的收集会是非常消耗时间和资源的工作。在这种情况下需要小心选择sample size的大小，确保在可接受的时间范围、资源限制和维护窗口内完成收集。请参见以下文档：

[Document 44961.1](#) Statistics Gathering: Frequency and Strategy Guidelines

对于这种系统，我们也推荐使用基于改变的统计信息的收集，避免重复收集不必要的信息。和10g相比，11g的自动统计信息的过程已经发生了改变，详见以下文档：

[Document 237901.1](#) Gathering Schema or Database Statistics Automatically - Examples

[Document 756734.1](#) 11g: Scheduler Maintenance Tasks or Autotasks (Doc ID 756734.1)

[Document 743507.1](#) Why Has the GATHER_STATS_JOB been removed in 11g? (Doc ID 743507.1)

收集对象的统计信息

The Cost Based Optimizer (CBO) 是基于统计信息生成特定查询的执行计划的。当sample size减小的时候，基于抽样出来的样本可能会生成不同的统计信息，这是因为这些样本有可能正好是用某些特定的方法录入的。

在11g中，收集统计信息推荐使用定期统计信息收集脚本。在绝大多数情况下系统默认脚本提供的采样率都是合适的，并兼顾了以下推荐：

- **使用足够大的sample size**

在11g中，对于ESTIMATE_PERCENT，我们建议使用缺省的 DBMS_STATS.AUTO_SAMPLE_SIZE。Oracle首先会判断如果对一个表使用100%的sample size是否能在一个maintenance window中收集完。如果能收集完那么就对这个表使用100%的sample size，尽管这意味着对这张表收集统计信息的频率会降低。如果100%的sample size并不可行，那么会至少使用30%作为sample size。11g使用了一种hash算法来计算统计信息，所以在大多数时候收集信息的性能都是可以接受的。通常，对大部分应用程序而言，相对于数据每天发生的一点点变化，统计信息整体上的准确度更重要一些。这个参数在早期版本的设置请参见之后的文档。

- **收集所有对象的统计信息**

确保所有对象（表和索引）都收集了统计信息。这个通过使用CASCADE参数可以很容易的实现。

- **收集字段的统计信息/数据非均匀分布列的histogram**

确保所有数据非均匀分布的列都收集了histogram,并且使用METHOD_OPT参数来确保histogram的精度。通常我们推荐使用缺省的设置“**AUTO**”。这意味着如果histogram的存在会有助于生成更准确的执行计划，DBMS_STATS会自动在需要字段上生成它。相对于在所有的列上收集histogram，只在确实需要的列上收集是更稳妥的做法。

[Document 390249.1](#) How To Quickly Add/Remove Column Statistics (Histograms) For A Column

注意：假如某个列上的数据是非均匀分布的，并且sample size不是100%（例如使用**AUTO**），那么这个列上的某些数据可能会被统计信息给漏掉。假如发生了这种情况，并且查询恰好使用了那些没有histogram bucket对应的值，那么这条查询就不可能使用真正准确的统计信息，进而可能产生不准确甚至是错误的执行计划。在这种情况下，使用100%的sample size是唯一能保证统计信息准确性的办法。假如这是不可行的，移除列上的histogram（只使用列的最大最小值）将使生成的执行计划更加一致，因为selectivity是不变的。

同样的道理，假如统计信息不是最新的，histogram的存在甚至会引入额外的问题，比如当查询引用的值是在histogram的范围之外，或者是“frequency” histogram的中间值（比如，列上添加了一个之前不存在的值或者某个特定值的行数发生了很大的变化）。在这样的情况下，optimizer将不得不做一些猜测（有可能是不正确的），从而产生不准确的执行计划。所以，需要通过应用程序测试不同的值以确定哪种统计信息能帮助产生最好的执行计划。

需要注意的是由于功能限制目前只有254个buckets可以用来存储列统计信息。如果您的distinct value的个数非常大而且上面数据的分布是不均匀的，列的统计信息仍然可能是不准确的。详见：

[Document 212809.1](#) Limitations of the Oracle Cost Based Optimizer

还有就是只有当你的应用确实能用到histogram的时候收集它才是有意义的。特别是当你使用绑定变量但不会做窥测(bind peek)的时候（例如**OPTIM_PEEK_USER_BINDS = FALSE**），optimizer不会有任何用来检索histogram的数据，从而也不会因为histogram的存在而更好地估算cardinality。

在早期的版本中，参数METHOD_OPT的缺省的设定是“**FOR ALL COLUMNS SIZE 1**”。这意味着我们只会收集列的最大最小值，而没有值的具体分布的信息。我们知道，有些时候histogram的存在反而不利于好的执行计划的产生，所以当从一个版本升级到另外一个版本后，用户可能希望先将这个参数设置为之前版本的值，运行一段时间后再调整为当前版本的值，使升级后的初始阶段更加平稳，详见以下文档：

[Document 465787.1](#) How to: Manage CBO Statistics During an Upgrade from 10g or 9i into 11g

- **收集分区对象的Global Statistics(全局统计信息)**

如果使用了partition（分区）技术，时间允许的话尽可能收集Global Statistics。Global Statistics是非常重要的，但因为涉及数据量及收集时间长度的影响，很多时候它都被略过了。假如不能采用100%的sample size，那么我们推荐至少使用1%来做收集。小的采样（例如0.001, 0.0001, 0.00001等）可以是非常有效的，但同时要知道大量能帮助optimizer生成准确执行计划的数据会被忽略掉的。参数ESTIMATE_PERCENT有一个非常弹性的有效的范围 [0.000001 -> 100]，对于非常巨大的表您可以使用非常小的值。不同的系统需要经过测试才能得出最适合这个系统的值。详见以下文档：

[Document 236935.1](#) Global statistics - An Explanation

11g同时提供了以增量方式收集Global Statistics的方法.详见:

Oracle Database Performance Tuning Guide
11g Release 1 (11.1)
Part Number B28274-02
Chapter 13 Managing Optimizer Statistics

• 收集 System Statistics (系统统计信息)

收集 System Statistics 来真实反映系统 CPU 负载情况。除了估算正常 I/O 部分的 cost 之外，这些信息会帮助 CBO 更好的估算 CPU 部分的 cost。详见：

[Document 470316.1](#) Using Actual System Statistics (Collected CPU and IO information)
[Document 149560.1](#) Collect and Display System Statistics (CPU and IO) for CBO usage
[Document 153761.1](#) Scaling the System to Improve CBO optimizer

• 从早期版本升级到 11g

如果您是从 Oracle 的早期版本升级到 11g，您可以参见以下关于统计信息管理的建议：

[Document 465787.1](#) How to: Manage CBO Statistics During an Upgrade from 10g or 9i into 11g/div>

缺省设置

不同版本收集统计信息时使用的缺省设置可能是不一样的，例如：

- **ESTIMATE_PERCENT** 的缺省值：
 - 9i : 100%
 - 10g : DBMS_STATS.AUTO_SAMPLE_SIZE (使用非常小的估算百分比)
 - 11g : DBMS_STATS.AUTO_SAMPLE_SIZE (使用更大的估算百分比 - 100%)
- **METHOD_OPT**: defaults:
 - 9i : "FOR ALL COLUMNS SIZE 1" 等同于没有列值的具体分布信息。
 - 10g and 11g : "FOR ALL COLUMNS SIZE AUTO" - 这意味着如果 histogram 的存在会有助于生成更准确的执行计划，DBMS_STATS 会自动在需要字段上生成它。

在 11g 中，AUTO_SAMPLE_SIZE 的缺省值是 100%，从而尽可能使统计信息是准确的。在之前的版本上，由于时间的制约，100% 的 sample size 有时候是不可能的，由于 11g 实施了一个新的 hash 算法避免了传统的排序（在 9i 和 10g 典型慢的部分是排序），这显著的减少了收集的时间并提升了资源使用效率。同样，是否收集列上的统计信息也是自动决定的，总之 11g 上还有很多其它类似的改进。

您可以使用以下文档中提到的方法来修改默认值：

[Document 1493227.1](#) How to Change Default Parameters for Gathering Statistics in Oracle 11g

统计信息收集命令的例子

在单一的表收集统计信息

```
>exec dbms_stats.gather_table_stats( -
  ownname => ' Schema_name ', -
  tabname => ' Table_name ', -
  estimate_percent => DBMS_STATS.AUTO_SAMPLE_SIZE, -
  cascade => TRUE, -
  method_opt => 'FOR ALL COLUMNS SIZE AUTO' );
```

注：如果确定 histogram 不会对产生更好的执行计划有帮助，那么更好的做法是将下边这一行：

```
method_opt => 'FOR ALL COLUMNS SIZE AUTO'
```

替换成

```
method_opt => 'FOR ALL COLUMNS SIZE 1'
```

注：您需要将 ' Schema_name ' 和 ' Table_name ' 替换成真正需要收集的 schema 和 table 的名字。

收集某一 schema 下的所有对象的统计信息

```
exec dbms_stats.gather_schema_stats( -
```

```
ownname => ' Schema_name ', -
cascade => TRUE, -
method_opt => 'FOR ALL COLUMNS SIZE AUTO' );
```

注：将 ' Schema_name ' 替换成需要收集的schema的名字。

收集database中所有对象的统计信息:

```
exec dbms_stats.gather_database_stats( -
cascade => TRUE, -
method_opt => 'FOR ALL COLUMNS SIZE AUTO' );
```

参考

- [NOTE:457926.1](#) - How to Gather Statistics on SYS Objects and 'Fixed' Objects?
 - [NOTE:465787.1](#) - How to: Manage CBO Statistics During an Upgrade from 10g or 9i into 11g
 - [NOTE:236935.1](#) - Global statistics - An Explanation
 - [NOTE:237901.1](#) - Gathering Schema or Database Statistics Automatically in 8i and 9i - Examples
 - [NOTE:44961.1](#) - Statistics Gathering: Frequency and Strategy Guidelines
 - [NOTE:470316.1](#) - Using Actual System Statistics (Collected CPU and IO information)
 - [NOTE:743507.1](#) - How to Benefit from Automatic Maintenance Tasks Following the Removal of the GATHER_STATS_JOB in 11g?
 - [NOTE:756734.1](#) - 11g: Scheduler Maintenance Tasks or Autotasks
 - [NOTE:1226841.1](#) - How To: Gather Statistics for the Cost Based Optimizer
 - [NOTE:1380043.1](#) - Selected Performance Related Seminars from Oracle Openworld (OOW)
 - [NOTE:149560.1](#) - How to Collect and Display System Statistics (CPU and IO) for CBO use
 - [NOTE:452011.1](#) - Restoring Table Statistics
 - [NOTE:390249.1](#) - How To Quickly Add/Remove Column Statistics (Histograms) For A Column
- 未找到您要查找的产品?