Statistics Gathering and Histograms

DOs ✔ and DON’Ts ✗

Wolfgang Breitling
www.centrexcc.com

Copyright 2009 Centrex Consulting Corporation. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising and promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from Centrex Consulting.
Who Am I

Independent Consultant since 1996

Specializing in Oracle and Peoplesoft Setup, Administration, and Performance Tuning

Member of the Oaktable Network

30+ Years in Database Management

DL/1, IMS, ADABAS, SQL/DS, DB2, Oracle

Oracle since 1993 (7.0.12)

OCP Certified DBA - 7, 8, 8i, 9i
DON’ts

✘ Do not use ANALYZE ANYMORE

✘ Do not use DBMS.Utility.ANALYZE_
  {DATABASE | SCHEMA | PART_OBJECT}

✘ Do not use DBMS_DDL.ANALYZE_OBJECT (pre 10g)
DO

✔ use DBMS_STATS.GATHER_xxx_STATS
✔ from 11g on use auto_sample_size
✔ in 10g or less

✔ if you can’t use estimate_percent 100,
  gather table stats with cascade=>false

✔ then gather index stats separately with
  estimate_percent=>100.
### Effect Of Sample Size

<table>
<thead>
<tr>
<th>OBJ#</th>
<th>SAMPLESIZE</th>
<th>ROWCNT</th>
<th>BLKCNT</th>
<th>ANALYZETIME</th>
<th>SAVE_TIME</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>OBJ#</th>
<th>SAMPLESIZE</th>
<th>ROWCNT</th>
<th>LVL</th>
<th>LEAFCNT</th>
<th>DISTKEY</th>
<th>LBLKKEY</th>
<th>DBLKKEY</th>
<th>CLUFAC</th>
<th>ANALYZETIME</th>
<th>SAVE_TIME</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>OBJ#</th>
<th>SAMPLESIZE</th>
<th>ROWCNT</th>
<th>LVL</th>
<th>LEAFCNT</th>
<th>DISTKEY</th>
<th>LBLKKEY</th>
<th>DBLKKEY</th>
<th>CLUFAC</th>
<th>ANALYZETIME</th>
<th>SAVE_TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>177225</td>
<td>210652</td>
<td>39647136</td>
<td>3</td>
<td>210609</td>
<td>1041176</td>
<td>1</td>
<td>16</td>
<td>16733511</td>
<td>2011-07-16 10:54:18</td>
<td>2011-07-16 16:30:17</td>
</tr>
</tbody>
</table>
Column Statistics Gathering

Do not use
“for all indexed columns ...”

Do use
“for all columns size 1”
then
“for columns size 254 col1[,col2,...]”

not “for columns col1[,col2,...] size 254” *

for columns requiring histograms

* see www.centrexc.cc/Active%20Statistics.ppt.pdf
**When To Use Histograms**

- Histograms are stored in the dictionary.
- There is a maintenance and space cost for using histograms.
- **Only** create histograms on columns that are used in WHERE clauses of queries and have a highly-skewed data distribution.
- `method_opt "size auto"` is supposed to do just that but does not always get the second part right.
When Not To Use Histograms

Histograms are **not** useful if

* all predicates on the column use bind variables
* the column data is uniformly distributed
* the column is not used in WHERE clauses of queries
* the column is unique and is used only with equality predicates
method_opt “size auto”

Histograms are **not** gathered on columns for which there is no predicate information in the col_usage$ table.

col_usage$ is only populated

* when queries are hard parsed
* with columns referenced in predicates
Bind Peeking

A cursor will be moved out of the Shared Pool and therefore subsequently requires a hard parse in a number of circumstances:

* Database shutdown/restart
* Cursor not in use by any session, and aged out by LRU algorithm
  It will **not** get moved out by flushing the Shared Pool if it is pinned (i.e. in use)
* Change to the structure of any referenced object (e.g. alter table)
* Granting/revoking privileges on a referenced object.
* Change to the stats associated with any referenced object (e.g. following a gather stats job).
  However, that may not be immediate if no Invalidate is AUTO
Since 10g whenever optimizer statistics are modified using the DBMS_STATS package, old versions of the statistics are saved automatically for future restoration, so the above is not necessary.

However, after 31 days they DO get purged.

See Note 452011.1 "Restoring table statistics in 10G onwards".

✔ Do export your (good) statistics regularly.
DO delete object statistics before importing them.

Always.
‘gather auto’ vs. ‘gather stale’

gather auto

* ignores all other options and uses
  * ‘for all columns size skewonly’ until 11.1.0.7
  * ‘for all columns size auto’ for 11.1.0.7 and up

gather stale

* other parameters specified will also be taken
* does not gather statistics on objects which do not have statistics. (modifications can only be detected relative to existing num_rows)
ANALYZE command gathers incorrect num_distinct in every release if the length of character column is more than 32 bytes.

DBMS_STATS with METHOD_OPT => 'FOR ALL COLUMNS SIZE 1' gathers correct num_distinct.

From 11g on, DBMS_STATS with METHOD_OPT => 'FOR ALL COLUMNS SIZE N' (N >1) generates correct num_distinct. But, it still generates incorrect histograms.
Index Rebuild Problem

ALTER INDEX ... REBUILD COMPUTE STATISTICS deletes histograms of the leading column.

This is due to the Bug 6040988
To save histogram you should set event 8130 at system or session level.

```
ALTER {session|system} SET EVENTS
'8130 TRACE NAME CONTEXT FOREVER';
```

Reported for 9.2.0.8, appears fixed in 10g
Index Rebuild Problem

• Gather statistics of the table
• [ wait a few minutes ]
• Rebuild the index
• Restore the statistics giving a time between gathering the statistics and the rebuild.

You end up with no statistics at all on the index. Not the ones gathered with the table nor the ones gathered as a result of the rebuild.
The statistics are just wiped out.

Fixed in Oracle 11.2.0.1 and up
Deleting Histograms

Prior to 11g there is no simple way to delete just the column histogram statistics, leaving the basic column statistics, without re-analyzing.

Workaround:

Use `DBMS_STATS.GET_COLUMN_STATS` to get all column stats, change the bucket count in SREC to reflect no histogram and use `DBMS_STATS.SET_COLUMN_STATS` to re-set them.

Oracle 11g adds a new argument (COL_STAT_TYPE) to `DBMS_STATS.DELETE_COLUMN_STATS` to provide the ability to delete histogram statistics only.

**COL_STAT_TYPE** can be

* **ALL** is the default behaviour which deletes all column statistics
* **HISTOGRAM** will only delete the histograms for the column
Extended Stats and Histograms

Poor (overestimated) table cardinality estimates are possible when using multi-column statistics if there were predicates on columns with histograms that were not part of the column group used to calculate selectivity (bug 9069046).

Affected: 10.2.0.4 <= Oracle version < 11.2.0.2

Fixed in 11.2.0.2
Wolfgang Breitling
breitliw@centrexcc.com
Centrex Consulting Corp.
www.centrexxcc.com