

Are *you* a Runstats Master?

Dirk Johann – Senior Consultant

Author: Roy Boxwell – Senior Software Architect

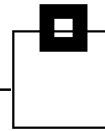
 SOFTWARE ENGINEERING GMBH

BMODUG / Canadian DB2 Users Group



SOFTWARE ENGINEERING

SOFTWARE ENGINEERING – What's new 2006:



Anniversary:


"We are in the Tools Business" – since 1986

Current seminars & Workshops:

- DB2 Migration to V8 – Be Prepared



Current development highlights:

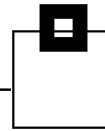
- DB2 V8 EarlyPrecheck ✓ (GA since January 2006)
- DB2 V9 Test & Development (Research since Feb. 2005)
- Feeding the Optimizer – Best in, Best Out
-  **Recovery HealthCheck**



SOFTWARE ENGINEERING



SOFTWARE ENGINEERING – What's new 2006:



New  **HealthCheck series™ for DB2 z/OS** product family:

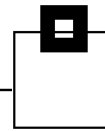
-  **Statistics HealthCheck**
-  **Recovery HealthCheck**
-  **Exception HealthCheck**



SOFTWARE ENGINEERING

Statistics HealthCheck is the freeware tool that I will now be discussing. Recovery HealthCheck is our new tool that checks that your RTOs are sustainable and reachable. Exception HealthCheck is our tool that checks for catalog inconsistencies like invalid packages, plans, superfluous incremental copies, non-available vsam clusters, objects in a bad status etc.

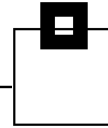
AGENDA



- RUNSTATS basics and DB2 Optimizer relevant statistics
- IBM recommendations through the ages
- RUNSTATS walk through
- SYSCOLDIST explained
- RUNSTATS Q & A
- Statistics HealthCheck
- Notes & Questions



Basic RUNSTATS knowledge



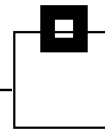
- The RUNSTATS utility
 - gathers summary information about the characteristics of data in table spaces, indexes and partitions
 - Invalidates the dynamic statement cache
 - Optionally
 - Reports the statistics
 - Updates the DB2 catalog
 - Updates the DB2 catalog history tables



SOFTWARE ENGINEERING

In V8 you can now use the UPDATE NONE REPORT NO feature to very quickly kill your DSC

Basic RUNSTATS knowledge



- Different types of RUNSTATS
 - RUNSTATS tablespace
 - RUNSTATS index
 - REORG / LOAD with Inline RUNSTATS
- Different types of statistics
 - Pure access path statistics
 - Those used by BIND in its process of optimization to determine access path
 - Parallelism access path statistics
 - Those used by BIND in its process of optimization to determine the degree of parallelism
 - Space statistics
 - Those used by the DBA to monitor space usage; to assist in capacity planning; to determine frequency of reorg; etc.



Catalog tables used for access path

- SYSIBM.SYSCOLDIST
- SYSIBM.SYSCOLSTATS *
- SYSIBM.SYSCOLUMNS
- SYSIBM.SYSINDEXES
- SYSIBM.SYSINDEXPART
- SYSIBM.SYSROUTINES
- SYSIBM.SYSTABLES
- SYSIBM.SYSTABLESPACE
- SYSIBM.SYSTABSTATS

* degree of parallelism only

SOFTWARE ENGINEERING

© 2006 SOFTWARE ENGINEERING GMBH

7

Note that SYSIBM.SYSINDEXSTATS that was in the documentation as being used up to and including V7 is not really used. In Vnext there are some new tables here (Possibly)

Columns used for access path

<u>SYSCOLDIST</u>	3.1	<u>SYSINDEXES</u>		<u>SYSROUTINES</u> 6.1
• CARDF	5.1	• CLUSTERING*	3.1	• CARDINALITY*
• COLGROUPOCOLNO	5.1	• CLUSTERRATIOF		• INITIAL_INSTS*
• COLVALUE		• FIRSTKEYCARDF		• INITIAL_IOS*
• FREQUENCYF		• FULLKEYCARDF		• INSTS_PER_INVOC*
• NUMCOLUMNS	5.1	• NLEAF		• IOS_PER_INVOC*
• STATSTIME		• NLEVELS		
• TYPE	5.1			
 		<u>SYSINDEXPART</u>		<u>SYSTABSTATS</u> 3.1
<u>SYSCOLUMNS</u>		• LIMITKEY*	4.1	• CARDF 4.1
• COLCARDF				• NPAGES 4.1
• HIGH2KEY		<u>SYSTABLES</u>		
• LOW2KEY		• CARDF		<u>SYSTABLESPACE</u>
		• EDPROC*		• NACTIVEF
		• NPAGESF		
		• PCTROWCOMP	3.1	
If not marked then table/column used in DB2 2.3				* This column is not updated By RUNSTATS

© 2006 SOFTWARE ENGINEERING GMBH

8

Administration Guide Chapter 5.9.1 / 5.10.1 Understanding statistics used for access path selection

As you can readily see, DB2 is pretty stable on the columns of its access path! There was one new table for Routines in V6 and a revamp of the SYSCOLDIST in V5 and then the last major change was the partitioning support in V4....(One note is that SYSCOLDIST used to be called SYSFIELDS in 2.3 but it was not updateable)

The SYSINDEXES CLUSTERING, SYSINDEXPART LIMITKEY, SYSTABLES EDPROC and none of the columns in SYSROUTINES are updated by RUNSTATS.

These are with two exceptions ALL of the data that the optimizer uses. What are the exceptions? The sizes of the various Buffer/Rid pools and the CPU you are running on.

So what?

OK, we now know all the info that DB2 uses to choose its access path. What can we do with this info?

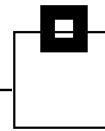
- We can change it to, hopefully, improve an SQL.
- We can delete some of it to, hopefully, improve an SQL.
- We can insert into it to, hopefully, improve an SQL.
- We can ignore it and simply trust that DB2 „Knows what it is doing...“
- We can mess it all up.
- We can use clever add-ons to *really* mess it all up!
- We can use freeware to see what state all these statistics are currently in. Good, Bad, or Awful.

SOFTWARE ENGINEERING

The clever-add on is Visual Explain with its Statistics Advisor and the freeware is, of course, Statistics HealthCheck. So lets take a look at what IBM tell us we should look out for and be VERY careful about actually doing...

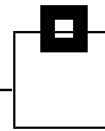
AGENDA

- RUNSTATS basics and DB2 Optimizer relevant statistics
- **IBM recommendations through the ages**
- RUNSTATS walk through
- SYSCOLDIST explained
- RUNSTATS Q & A
- Statistics HealthCheck
- Notes & Questions



SOFTWARE ENGINEERING

IBM Recommendations DB2 2.3



Correlations in the catalog (DB2 Administration Guide)

Relationships exist among certain columns of certain tables:

- Columns within SYSCOLUMNS
- Columns in the tables SYSCOLUMNS and SYSINDEXES
- Columns in the tables SYSCOLUMNS and SYSCOLDIST (well actually SYSFIELDS in those days!)



If you plan to update some values, keep in mind the following correlations:

- COLCARDF and FIRSTKEYCARDF/FULLKEYCARDF
- COLCARDF, LOW2KEY and HIGH2KEY. For non-default COLCARDF



SOFTWARE ENGINEERING



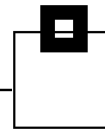
And that was it in 2.3. OK they did add bits after FIRSTKEYCARDF FULLKEYCARDF stating that they must be equal for the 1st column and full if a single column index. And they

Said that the LOW and HIGH key must be filled with data if the COLCARDF was not equal -1 (However the eight bytes, seven if you had NULL, were not that good for predicate

Filter factors on ranges....)

You had to work out nearly everything yourself

IBM Recommendations DB2 3.1



Correlations in the catalog (DB2 Administration Guide)

No change

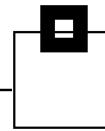


SOFTWARE ENGINEERING

IBM Recommendations DB2 4.1

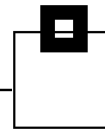
Correlations in the catalog (DB2 Administration Guide)

No change



SOFTWARE ENGINEERING

IBM Recommendations DB2 5.1



Correlations in the catalog (DB2 Administration Guide)

Relationships exist among certain columns of certain tables:

- Columns in the tables SYSCOLUMNS, SYSCOLDIST, and SYSINDEXES



If you plan to update some values, keep in mind the following correlation:

- **CARDF** in SYSCOLDIST. **CARDF** is related to **COLCARDF** and **FIRSTKEYCARDF** and **FULLKEYCARDF**. It must be the minimum:
 - A value between **FIRSTKEYCARDF** and **FULLKEYCARDF** if the index contains the same set of columns
 - A value between **MAX(colcardf of each col)** and the product of all the columns **COLCARDFs** in the group



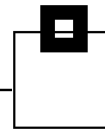
SOFTWARE ENGINEERING

Ah ha! A change! They added a relationship and a correlation. They also did something quite sneaky in 5.1 they changed the way that RUNSTATS works with SYSCOLDIST. In prior releases the runstats deleted any data in the syscoldist and then if requested put it back. With 5.1 they stopped the auto delete. This is now coming back to haunt them as they MUST find out somehow to get rid of the old frequency data in the syscoldist somehow (In Vienna there was a show of hands about this! In Vnext should it be automatic or a new option on the runstats control cards.....)

IBM Recommendations DB2 6.1

Correlations in the catalog (DB2 Administration Guide)

No change

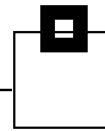


SOFTWARE ENGINEERING

IBM Recommendations DB2 7.1

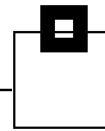
Correlations in the catalog (DB2 Administration Guide)

No change



SOFTWARE ENGINEERING

IBM Recommendations DB2 8.1



Correlations in the catalog (DB2 Administration Guide)

If you plan to update some values, keep in mind the following correlations:

- The COLCARDF, LOW2KEY, and HIGH2KEY gained a note:
„If the COLCARDF is 1 or 2 DB2 uses LOW2KEY and HIGH2KEY as domain statistics to generate frequencies“
- CARDF in SYSTABLES. CARDF must be equal or larger than any other cardinalities, such as COLCARDF, FIRSTKEYCARDF, FULLKEYCARDF, and CARDF in SYSCOLDIST
- FREQUENCYF and COLCARDF or CARDF. The number of frequencies collected must be less than or equal to COLCARDF for the column or CARDF for the column group
- FREQUENCYF. The sum of frequencies collected for a column or column group must be less than or equal to 1



Ah ha! A change! They added three correlations all to do with CARDF and FREQUENCYF this also proves the theory that in V8 that the optimizer likes using these statistics much more than it used to in V7 and before!

IBM Recommendations DB2 8.1

In the „Filter factors and catalog statistics“ chapter (DB2 Administration Guide)

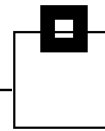
Recommendation: If query performance is not satisfactory, consider the following actions:

- Collect cardinality statistics on all columns that are used as predicates in a WHERE clause.
- Collect frequencies for all columns with a low cardinality that are used as COL op literal predicates.
- Collect frequencies for a column when the column can contain default data, the default data is skewed, and the column is used as a COL op literal predicate.
- Collect KEYCARD on all candidate indexes.
- Collect column group statistics on all join columns.

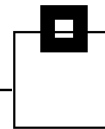
In this chapter they have added a whole new section, again proof that V8 is much more sensitive to SYSCOLDIST data. My favorite quote at the moment is from Susan Lawson from “The DB2 Cocktail hour #11” where when asked about statistics and runstats in V8 answers “Do ‘em, Need ‘em, Gotta have ‘em” and then Dan Luksetich says “Statistics are everything in DB2 V8” and “The Optimizer is more than ever dependant on statistics!”

AGENDA

- RUNSTATS basics and DB2 Optimizer relevant statistics
- IBM recommendations through the ages
- **RUNSTATS walk through**
- SYSCOLDIST explained
- RUNSTATS Q & A
- Statistics HealthCheck
- Notes & Questions



RUNSTATS walk through DB2 7.1 & 8.1



Object Details:

Database ROYTEST2
Tablespace ROYTEST2
Table ROYTEST2
Columns COL1, COL2, COL3, COL4, COL5
all CHAR(16) NOT NULL
Indexes COL1, COL2 & COL1, COL3 & COL2 & COL1
(All Unique)

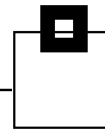


All RUNSTATS run with UPDATE ACCESSPATH.



SOFTWARE ENGINEERING

RUNSTATS walk through DB2 7.1 & 8.1



RUNSTATS ROYTEST2.ROYTEST2

Updates in the DB2 catalog (as listed in the RUNSTATS output):

- SYSTABLES
- SYSTABLESPACE



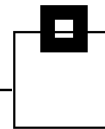
Probably the worst RUNSTATS you could actually do!



SOFTWARE ENGINEERING



RUNSTATS walk through DB2 7.1 & 8.1



RUNSTATS ROYTEST2.ROYTEST2 TABLE(ALL)

Updates in the DB2 catalog (as listed in the RUNSTATS output):

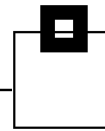
- SYSTABLES
- SYSCOLUMNS - all columns
- SYSTABLESPACE

Still pretty bad!



SOFTWARE ENGINEERING

RUNSTATS walk through DB2 7.1



RUNSTATS ROYTEST2.ROYTEST2 TABLE(ALL) INDEX(ALL)

Updates in the DB2 catalog (as listed in the RUNSTATS output):

- SYSTABLES
- SYSCOLUMNS not for COL1
- SYSTABLESPACE
- SYSINDEXES for Index 01
- SYSCOLUMNS for COL1

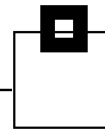
- SYSINDEXES Repeated block for indexes 02 & 03
- SYSCOLUMNS for COL2 & COL3
- SYSCOLDIST for COL2 & COL3 – Frequencies only



SOFTWARE ENGINEERING

Much better. Note that single column unique indexes do not get any frequencies using the INDEX(ALL) as they are, by definition, "unique" values. This is also true if this Column is then used as the leading column in any other index. So if you have indexes COL1 (Unique), COL1 & COL2, COL1 & COL2 & COL3 you will not get any leading column frequencies!

RUNSTATS walk through DB2 8.1



RUNSTATS ROYTEST2.ROYTEST2 TABLE(ALL) INDEX(ALL)

Updates in the DB2 catalog (as listed in the RUNSTATS output):

- SYSTABLES
- SYSCOLUMNS not for COL1
- SYSTABLESPACE

- SYSCOLUMNS for COL1, COL2 & COL3
- SYSINDEXES for Index 01 – 03

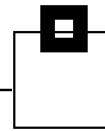
- SYSCOLDIST for COL2 & COL3 – Frequencies only



SOFTWARE ENGINEERING

Same as V7 apart from the report looks better and now if there is a single column unique index it will have frequencies! So if you have indexes COL1 (Unique), COL1 & COL2, COL1 & COL2 & COL3 you **will** get frequencies for COL1 (in this case twice once for the col1 col2 and once again for the col1 col2 col3!)

RUNSTATS walk through DB2 7.1



RUNSTATS ROYTEST2.ROYTEST2 TABLE(ALL) INDEX(ALL KEYCARD)

Updates in the DB2 catalog (as listed in the RUNSTATS output):

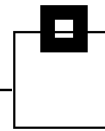
- SYSTABLES
- SYSCOLUMNS not for COL1
- SYSTABLESPACE
- SYSINDEXES for Index 01
- SYSCOLUMNS for COL1

- SYSINDEXES Repeated block for indexes 02 & 03
- SYSCOLUMNS for COL2 & COL3
- SYSCOLDIST Frequencies for COL2 & COL3
Cardinality for COL3 on COL3, COL2



Ah super! And the SYSCOLDIST data? Yep it is the Frequencies and the Cardinalities that the optimizer needs. We now have the top ten leading column index frequencies for the two columns COL2 and COL3 and a type C cardinality record. Only one? The reason is simple the FIRSTKEYCARDF contains the first column and the FULLKEYCARDF contains the complete index so it only needs the combinations between in this case col3 and col2

RUNSTATS walk through DB2 8.1



RUNSTATS ROYTEST2.ROYTEST2 TABLE(ALL) INDEX(ALL KEYCARD)

Updates in the DB2 catalog (as listed in the RUNSTATS output):

- SYSTABLES
- SYSCOLUMNS not for COL1
- SYSTABLESPACE
- SYSCOLUMNS for COL1, COL2 & COL3
- SYSINDEXES for Indexes 01, 02 & 03
- SYSCOLDIST Frequencies for COL2 & COL3
 Cardinality for COL3 on COL3, COL2

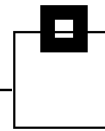


SOFTWARE ENGINEERING

Same as V7 apart from the report looks better. Now that the SYSCOLDIST actually contains data what does it actually all mean?

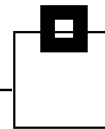
AGENDA

- RUNSTATS basics and DB2 Optimizer relevant statistics
- IBM recommendations through the ages
- RUNSTATS walk through
- **SYSCOLDIST explained**
- RUNSTATS Q & A
- Statistics HealthCheck
- Notes & Questions



SOFTWARE ENGINEERING

SYSCOLDIST contents explained



The SYSCOLDIST is used for two separate functions:

- Frequencies
- Cardinalities



The column TYPE contains either a „C“ or a „F“ (new in V8 is „N“ for non-padded frequency values)

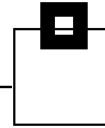


Regardless of the TYPE value the columns TBOWNER, TBNAME, NAME, COLGROUPCOLNO, NUMCOLUMNS and STATSTIME are used for the same purpose.



SOFTWARE ENGINEERING

SYSCOLDIST contents explained



TBOWNER and TBNAME are self explanatory.
NAME is the always just the first column name.
These three columns are also the non-unique index.



COLGROUPOCOLNO for a single column object is an empty string, for a multi-column object it contains a string of two byte smallint fields which contain the column numbers from the original table.



NUMCOLUMNS is the number of columns in this group.

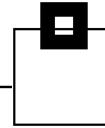
STATTIME is the time when RUNSTATS inserted this entry. It is also used when there is a complete duplicate so that DB2 uses the last inserted value.



SOFTWARE ENGINEERING

The single column object COLGROUPOCOLNO is not 100% correct I will show you why later

SYSCOLDIST contents explained



For frequency data the following columns are used:

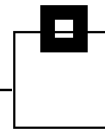
COLVALUE contains the actual data from the 1 – n columns in the group simply concatenated together. This data might be readable or it might not. Caution must be used when inserting or changing this data as numeric and date, time forms must be the internal DB2 format and not the external format (E.g. The high bit must be flipped for numbers) and watch out for NULLable columns!

FREQUENCYF contains a floating point value between 0.0 and 1.0 which is the frequency that this value occurs for this group in the table.

If you have V8 you can also use **MOST**, **BOTH**, or **LEAST** on your **RUNSTATS** to collect the most, both, or the least often.

SOFTWARE ENGINEERING

SYSCOLDIST contents explained



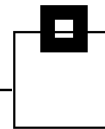
For cardinality data the following column is used:

CARDF contains a floating point value which is how many different values for this group occur in the data.



SOFTWARE ENGINEERING

SYSCOLDIST contents explained



SYSCOLDIST in a nut shell:-

- Frequencies are good for - COL op literal
- Cardinalities are good for - everything!

And remember that a frequency without a cardinality is like a warm beer. Nice to look at, perhaps, but the optimizer will not touch it!



SOFTWARE ENGINEERING

SYSCOLDIST contents explained

In this example you can see the importance of cardinality and frequency. If no frequency data existed in the SYSCOLDIST DB2 would assume that all values are equally distributed.

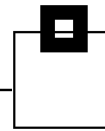
Here the default filter factor would be $1/5$ ($1/\text{COLCARDF}$) or 20%

Category	Frequency
Infant	5%
Child	15%
Adolescent	25%
Adult	40%
Senior	15%

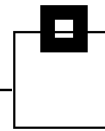
This would lead the optimizer to underestimate by 50% for ADULT and to over-estimate by 400% for INFANT.

AGENDA

- RUNSTATS basics and DB2 Optimizer relevant statistics
- IBM recommendations through the ages
- RUNSTATS walk through
- SYSCOLDIST explained
- **RUNSTATS Q & A**
- Statistics HealthCheck
- Notes & Questions



Runstats Q & A



We now know which data is used and where, so now comes a list of RUNSTATS questions:

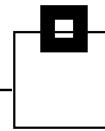
- `FREQVAL NUMCOLS 3 COUNT 10` – What does this do?
- `FREQVAL NUMCOLS 3 COUNT 0` - What does this do?
- Use of `COLGROUP` in DB2 V8
- Use of `SAMPLE`
- Use of `REOPT(ONCE)`
- Does use of `REORG INDEX` with inline statistics cause problems?
- What happens to frequencies & cardinalities when not specified in the RUNSTATS utility run?



SOFTWARE ENGINEERING

We have seen performance increases of 10% to 20% in V7 and V8 just by adding `KEYCARD` to all RUNSTATS!

Runstats Q & A



FREQVAL NUMCOLS 3 COUNT 10

According to the DB2 documentation the NUMCOLS is the number of leading index columns to sample so you actually need to give:

RUNSTATS ROYTEST2.ROYTEST2 TABLE(ALL) INDEX(ALL KEYCARD

FREQVAL NUMCOLS 1 COUNT 10

FREQVAL NUMCOLS 2 COUNT 10

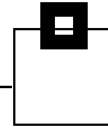
FREQVAL NUMCOLS 3 COUNT 10)

To get the results you expect!

SOFTWARE ENGINEERING

And here is a problem – This will generate a single column object COLGROUPOCOLNO with a COLGROUPOCOLNO of one column. This is not allowed according to IBM...whether or not the optimizer has a problem here I cannot state, my hope is that a one in numcols with no COLGROUPOCOLNO is treated the same as a filled COLGROUPOCOLNO. There is also a difference in behavior between V7 and V8 in CM or NF mode and that is in V7 for a single column index that was also the leading column in any other index using FREQVAL NUMCOLS 1 did not create a SYSCOLDIST record but in V8 it does! Note that with any use of FREQVAL you MUST add the NUMCOLS 1 otherwise no leading column FREQS are collected!

Runstats Q & A



FREQVAL NUMCOLS 3 COUNT 0

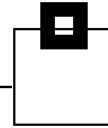
This is more of a „hidden feature“ and it deletes all of the multi-column frequencies.

However it does **NOT** delete any entries created by COLGROUP processing in V8.



SOFTWARE ENGINEERING

Runstats Q & A



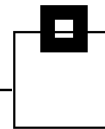
Use of COLGROUP in DB2 V8

This is a very powerful addition to DB2 V8 as it completely removes the need for the old DSTATS program. DSTATS had serious performance problems and column type limitations which made its use complex, time consuming, and error prone.



SOFTWARE ENGINEERING

Runstats Q & A



Use of SAMPLE

According to the DB2 Utilities documentation this is only for non-indexed columns, however from my tests in V7 and V8 it has a major impact on indexed columns as well.

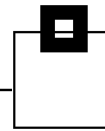
The problem with SAMPLE is that even using SAMPLE 100 DB2 still gets it wrong...

Col	1	5	10	20	21	22	23	24	25	99	100
TBNAME	62	121	162	178	178	180	180	178	180	180	180
CLNAME	11647	11647	5442	4854	11647	5057	5340	3293	6707	7210	7168

The actual values were 181 and 6871. So be very careful when using this!

Sample default is 25 and the range is 1 – 100. 25 is a very good as it gives nearly perfect results. In this example there was a five column index the first two columns had a COLCARDF of one then the third had a COLCARDF of 180. So for these three columns the multi-column cardf is calculated as being $1 \times 1 \times 180 = 180$. However the actual multi-cardf was 181 (Which is correct as $1 \times 1 \times 181 = 181$). This is a trivial example but could well lead to problems! My tests were in a V7 and V8 CM system and both TBNAME and CLNAME were varchar which I think is the problem here! This means that you might need to write a whole bunch of `SELECT COUNT(DISTINCT coln) FROM xxxxx` code to compare the actual values with the sampled values. The DB2 admin guide also states that "if you know a more accurate number then you can update the COLCARDF"

Runstats Q & A



Use of REOPT(ONCE) in DB2 V8

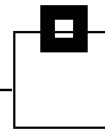
This is a very interesting addition to DB2 V8 as it enables DB2 to do its dynamic SQL mini-bind only once. This can be very good for performance... or not...

If you are a SAP customer then the change that SAP did in ecc5 to use the REOPT(ONCE) could be a CPU killer! The problem is that the first run SQL might not actually reflect the normal SQL that is executed over the day. The way out of this problem is to actually delete all frequency records from the SYSCOLDIST for the relevant tables.

SOFTWARE ENGINEERING

The problem is actually very hard to find because when the DBA explains the query it can return "normal" results. The frequencies in "col op literal" enable DB2 to see that no rows will be returned and so it can use a bad index to get no rows. However using the bad index to get 100000 rows is not good!

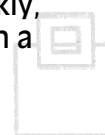
Runstats Q & A



Does use of REORG INDEX with inline statistics cause problems?

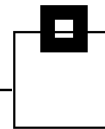
YES!

The problem is that the inline stats will *only* update the index statistics and *not* any of the table ones. This leads, very quickly, to the statistics „drifting apart“. The solution here is either run a tablespace runstats after an index reorg, manually update the table statistics, or do not use inline statistics.



SOFTWARE ENGINEERING

Runstats Q & A



What happens to your frequencies and cardinalities when you do a RUNSTATS run without using FREQVAL and/or KEYCARD?

For frequencies the leading column information is replaced (and then the COLGROUPOCOLNO is correctly set to an empty string) and all the other data stays in the catalog. This can be a good and a bad thing!

The cardinalities stay there! Again this can be a good and a bad thing!

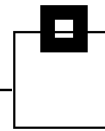
This means that if you have ever run a RUNSTATS with either FREQVAL or KEYCARD and since then *without* these keywords you probably have „interesting“ data in the SYSCOLDIST.

SOFTWARE ENGINEERING

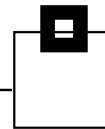
As I mentioned before from V5 RUNSTATS does not do an auto-delete anymore so be careful!

AGENDA

- RUNSTATS basics and DB2 Optimizer relevant statistics
- IBM recommendations through the ages
- RUNSTATS walk through
- SYSCOLDIST explained
- RUNSTATS Q & A
- **Statistics HealthCheck**
- Notes & Questions



Statistics HealthCheck



Now that you have learned everything that RUNSTATS does, the question is:

How do I know that the statistics I have are correct?



The answer is:

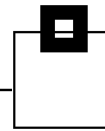
Run  **Statistics HealthCheck**



SOFTWARE ENGINEERING



Statistics HealthCheck



Here is the list of critical violations:

```
Critical violations
Frequency < 0 . . . . . : 0
Frequency > 100 . . . . . : 0
Frequency count > cardf . . . . . : 0
Frequency count > colcardf . . . . . : 0
Frequency sum < 0 . . . . . : 0
Frequency sum > 100 . . . . . : 0
SYSCOLDIST cardf outside allowable range . . . . . : 1
SYSCOLUMNS - low2key high2key empty . . . . . : 0
SYSCOLUMNS aggregate colcardf < SYSCOLDIST cardf : 1
SYSCOLUMNS 1st cardf <> SYSINDEXES firstkeycardf : 0
SYSCOLUMNS 1st cardf <> SYSINDEXES fullkeycardf : 0
SYSINDEXES 1 col ix firstkeycardf <> fullkeycardf : 0
SYSTABLES cardf < SYSCOLDIST cardf . . . . . : 0
SYSTABLES cardf < SYSCOLUMNS colcardf . . . . . : 0
SYSTABLES cardf < SYSINDEXES firstkeycardf . . . . : 0
SYSTABLES cardf < SYSINDEXES fullkeycardf . . . . : 0
SYSTABLES no RUNSTATS . . . . . : 72
```

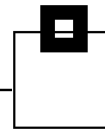


The output of the Statistics HealthCheck is a detail report listing the exact violation that has been detected and a summary report at the end. This is the list of critical checks that currently exist. These critical checks are all based on the DB2

Administration guide "Correlations in the catalog". We then also have Serious and these are from Statistics Advisor. Then come Warnings and these are IBM recommendations and finally come Informational messages. We also output up to four extra counts and these are the number of not-used in index CARD and FREQ records, the number of CLUSTER Y but CLUSTERING N indexes and any V8 index that contains a varying length column that is defined as PADDED. In V8 it could make sense to ALTER these to NON PADDED and rebuild the index. Of course the way to "fix" these problems is to simply run a RUNSTATS with the correct keywords or delete data from the SYSCOLDIST and run a RUNSTATS. Note that there are some DB2 APARs that correct runstats processing that you might need to apply first!

As an example of how critical these problems are the frequency count > cardf means that the SYSCOLDIST entries will be ignored and defaults used!

Statistics HealthCheck



Here is an example from the detail report:

```
VIOLATION DA.S1N PA.T1N
I - CLUSTERING IX NOT CLUSTERED . : PA.X1NPK
S - MISSING INDEX KEYCARD . . . . : PA.X1N022
S - MISSING INDEX KEYCARD . . . . : PA.X1N042
S - MISSING LEADING COLUMN FREQ . : PA.X4010E ADRESS_STREET
S - MISSING LEADING COLUMN FREQ . : PA.X1N017 LBS
W - MISSING INDEX FREQVAL NUMCOLS : PA.X1N017 NUMCOLS 00002
S - MISSING LEADING COLUMN FREQ . : PA.X1N022 POSTCODE
W - MISSING INDEX FREQVAL NUMCOLS : PA.X1N022 NUMCOLS 00002
W - MISSING INDEX FREQVAL NUMCOLS : PA.X1N022 NUMCOLS 00003
S - MISSING LEADING COLUMN FREQ . : PA.X1N042 ADRESS_CITY
W - MISSING INDEX FREQVAL NUMCOLS : PA.X1N042 NUMCOLS 00002
W - MISSING INDEX FREQVAL NUMCOLS : PA.X1N042 NUMCOLS 00003
W - MISSING INDEX FREQVAL NUMCOLS : PA.X1N042 NUMCOLS 00004
W - MISSING INDEX FREQVAL NUMCOLS : PA.X1N042 NUMCOLS 00005
```



SOFTWARE ENGINEERING

© 2006 SOFTWARE ENGINEERING GMBH

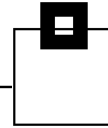
46

In this example you can see that one index defined as clustering is not clustered – here you would need a reorg to get the data and index back in line.

Two index keycards are missing for 22 and 42. (17 is not here as you can see later that 17 has only two columns and so no cardinality record can be there)

The three indexes 17, 22 and 42 have no frequency data (Or the syscolldist data was deleted).

Garbage in the SYSCOLDIST



A colleague is on holiday and the telephone rings with end-users or management saying that the system performance is bad. You grab your RUNSTATS JCL, submit the job and all is well.



Or is it?



SOFTWARE ENGINEERING



Garbage in the SYSCOLDIST

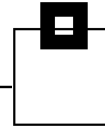
Here is the query that you have found:

```
SELECT A.ICTYPE , A.TIMESTAMP
FROM SYSIBM.SYSCOPY A
WHERE A.ICTYPE IN ('D','I','F','S','W','Y','R','X','Z')
      AND A.ICBACKUP NOT IN ('RP','RB')
      AND A.DBNAME = ?
      AND A.TSNAME = ?
      AND (A.DSNUM = ? OR (? > 0 AND A.DSNUM = 0))
      AND A.TIMESTAMP = (SELECT MAX(B.TIMESTAMP)
                          FROM SYSIBM.SYSCOPY B
                          WHERE B.ICTYPE IN
                             ('D','I','F','S','W','Y','R','X','Z')
                             AND B.ICBACKUP NOT IN ('LB','RP','RB')
                             AND B.DBNAME = ?
                             AND B.TSNAME = ?
                             AND (B.DSNUM = ? OR (? > 0 AND B.DSNUM = 0))
                          )
WITH UR
;
```

SOFTWARE ENGINEERING

Now try and work out the RUNSTATS....

Garbage in the SYSCOLDIST



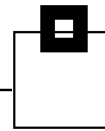
This is the actual RUNSTATS you should run...

```
RUNSTATS TABLESPACE DSNDB06.SYSCOPY
TABLE (SYSTEM.SYSCOPY)
COLGROUP (ICTYPE, TIMESTAMP, TSNAME)
COLGROUP (DBNAME, TIMESTAMP)
COLGROUP (ICTYPE, TSNAME)
COLGROUP (DBNAME, TIMESTAMP, TSNAME)
COLGROUP (TIMESTAMP, TSNAME)
COLGROUP (ICBACKUP)  FREQVAL COUNT 10
COLGROUP (ICTYPE)  FREQVAL COUNT 1
COLGROUP (DBNAME, ICTYPE, TIMESTAMP, TSNAME)
COLGROUP (ICBACKUP, TSNAME)
COLGROUP (DBNAME, ICBACKUP, TSNAME)
COLGROUP (ICBACKUP, TIMESTAMP, TSNAME)
COLGROUP (DBNAME, ICTYPE, TSNAME)
COLGROUP (ICTYPE, TIMESTAMP)
COLGROUP (DSNUM)  FREQVAL COUNT 1
COLGROUP (DBNAME, ICTYPE, TIMESTAMP)
COLGROUP (DBNAME, ICBACKUP, TIMESTAMP)
COLGROUP (ICBACKUP, TIMESTAMP)
COLGROUP (DBNAME, ICBACKUP, TIMESTAMP, TSNAME)
SORTDEVT SYSDA
SHRLEVEL CHANGE REPORT YES
```



Would you have guessed this? Actually I am being a bit unfair on Statistics Advisor here as this was the output in VE 1.0.8 the output in 1.0.9 looks like this -

Garbage in the SYSCOLDIST



This is the actual RUNSTATS you should run...

```
RUNSTATS TABLESPACE DSNDB06.SYSCOPY
          TABLE(SYSIBM.SYSCOPY)
          COLGROUP(ICBACKUP)  FREQVAL COUNT 10
          COLGROUP(ICTYPE)    FREQVAL COUNT 1
          COLGROUP(DSNUM)     FREQVAL COUNT 1
          SORTDEVT SYSDA
SHRLEVEL CHANGE REPORT YES
```



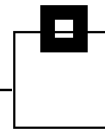
SOFTWARE ENGINEERING

In VE 1.0.9 it now only has the three COLGROUPs with FREQVAL keywords in them. But this does show the usage of SA and how the COLGROUP defs will play a major role in the future!

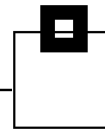
This also shows you the long term problems that exist in the handling of the SYSCOLDIST. If the data it contains deviates from the real state of the data your performance will not get any better after a RUNSTATS....again Statistics HealthCheck can come to the rescue by showing where differences exist within SYSCOLDIST for related data.

AGENDA

- RUNSTATS basics and DB2 Optimizer relevant statistics
- IBM recommendations through the ages
- RUNSTATS walk through
- SYSCOLDIST explained
- RUNSTATS Q & A
- Statistics HealthCheck
- **Notes & Questions**



Notes and Future development



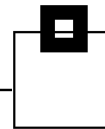
The Statistics HealthCheck is freeware and so has certain limitations:

- A maximum of 100 indexes on any one table
- No unicode table, index, or column names (long is OK!)



SOFTWARE ENGINEERING

Notes and Future development



Some possible future enhancements:

1. Add optional link to the Real-time Statistics tables to check that TOTALROWS and TOTALENTRIES is correctly mirrored in the CARD counts
2. SAP support
3. Direct link to our maintenance thresholds for RSA etc
4. Better input parameters for DB's to check or exclude
5. Generate corrective RUNSTATS statements
6. Generate corrective EXEC SQL statements for COLCARDF and SAMPLE problem



Go to WWW.NEONESOFT.COM and then the HealthCheck link for more details.



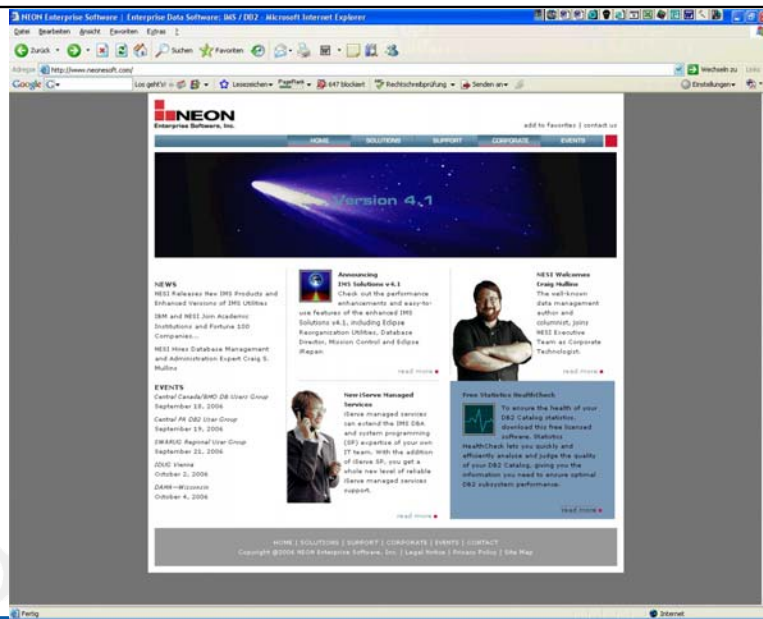
SOFTWARE ENGINEERING

RTS is not supported yet as it has still not become industry standard. If we get requests for this from five or more customers we will build this "link" in.

We will at some stage create a "payware" version of this software that will be easier to use and fully integrated into our Maintain and IQA product lines so some of these enhancements will not be for free!

The web site www.seg.de contains the Statistics HealthCheck forum where for the next two – three months you can cut-and-paste your summary reports and compare your output to those of your fellows. There will also be an Enhancements list a FAQ and a bug reporting system.

Notes and Future development

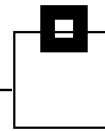


© 2006 SOFTWARE ENGINEERING GMBH

54

Website with Healthcheck link in the blue box.

Questions for you



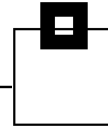
Now I have some questions for you all!

- How many of you use SAP on z/OS?
- How many of you use the DSNACC.EXCEPT_TBL to limit RUNSTATS?
- How many of you use a third-party or home-made tool to limit the execution of RUNSTATS?
- How many of you manipulate statistics in the Catalog?



SOFTWARE ENGINEERING

The End



Many thanks for your attention and if you have any questions I will be pleased to answer them!



By the way, you are all

official



RUNSTATS Masters!

SOFTWARE ENGINEERING

