What does "Create Statistics" do in SQL Server 2005?

Asked 16 years, 3 months ago   Modified 10 years, 10 months ago

Viewed 16k times

43

The Database Tuning Advisor is recommending that I create a bunch of statistics in my Database. I'm something of a SQL n00b, so this was the first time I'd ever come across such a creature. The entry in MSDN was a little obtuse - could someone explain what exactly this does, and why it's a good idea?

sql-server   t-sql   statistics   database-tuning-advisor

Share

Improve this question

Follow

edited Jan 9, 2013 at 14:48

Alex KeySmith
17.1k ● 11 ● 82 ● 157

asked Sep 24, 2008 at 16:14

Electrons_Ahoy
38.5k ● 37 ● 106 ● 127

5 Answers

Sorted by: Highest score (default) ⇕

Cost Based Query Optimisation is a technique that uses histograms and row counts to heuristically estimate the

38

cost of executing a query plan. When you submit a query to SQL Server, it evaluates it and generates a series of Query Plans for which it uses heuristics to estimate the costs. It then selects the cheapest query plan.
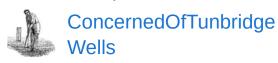
Statistics are used by the query optimiser to calculate the cost of the query plans. If the statistics are missing or out of date it does not have correct data to estimate the plan. In this case it can generate query plans that are moderately or highly sub-optimal.

SQL Server will (under most circumstances) generate statistics on most tables and indexes automatically but you can supplement these or force refreshes. The query tuning wizard has presumably found some missing statistics or identified joins within the query that statistics should be added for.

Share   Improve this answer

Follow

answered Sep 24, 2008 at 16:58

ConcernedOfTunbridge Wells

66.5k ● 15 ● 148 ● 198

The last paragraph is only slightly technically true: SQL Server will not create a Persistent Index autonomously (i.e. `CREATE INDEX`) but it can cache some data in an "Index Spool" but this data is not shared between different query executions. see this QA: dba.stackexchange.com/questions/43772/… – Dai Apr 27, 2016 at 22:33

▲

8

▼

Statistics are used by the optimizer to determine whether to use a specific index for your query. Without statistics, the optimizer doesn't have a way to know about how many of your rows will match a given condition, causing it to have to optimize for the "many rows" case, which could be less-than-optimal.

Share Improve this answer

Follow

answered Sep 24, 2008 at 16:21

Jonathan Rupp

15.8k ● 5 ● 48 ● 61

▲

3

▼

In a nutshell, it prepares your database to work effectively. By having prepared statistics, your database knows (before it needs to figure out an execution plan) what is likely to be its most efficient route.

Share Improve this answer

Follow

answered Sep 24, 2008 at 16:16

nathaniel

1,145 ● 2 ● 8 ● 20

2

Basically just keeps SQL updated with what type of indexing you have, row count, etc. This helps SQL better estimate how to execute your queries. Keeping the statistics updated is a good thing.

Share   Improve this answer

Follow

Kevin Fairchild
10.9k ● 6 ● 34 ● 52

2

From the BOL...

> Creates a histogram and associated density groups (collections) over the supplied column or set of columns of a table or indexed view. String summary statistics are also created on statistics built on char, varchar, varchar(max), nchar, nvarchar, nvarchar(max), text, and ntext columns. The query optimizer uses this statistical information to choose the most efficient plan for retrieving or updating data. Up-to-date statistics allow the optimizer to accurately assess the cost of different query plans, and choose a high-quality plan.

Share   Improve this answer

Follow

Galwegian
42.2k ● 16 ● 113 ● 158