

SQL days konferenz

What are index statistics

... and why should the developer care?

Thomas Hütter
Senior Developer

What are index statistics ...and why should the developer care?



SQLDays Erding 2021

Thomas Hüttner

What are index statistics ...

Thomas Hütter, Diplom-Betriebswirt

- Application developer, consultant, accidental DBA, author
- Worked at consultancies, ISVs, end user companies
- SQL Server > 6.5, former „Navision“ > 3.0, R > 3.1.2
- Speaker at SQL events around Europe

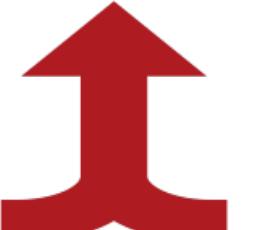


 @DerFredo <https://twitter.com/DerFredo>

 de.linkedin.com/in/derfredo

 www.xing.com/profile/Thomas_Huetter



sqlbits  **JOIN!**
THE DATABASE CONFERENCE

SQLdays
konferenz

Pure expertise at the
SQL Server Konferenz 2018
FEB 26th - 28th, 2018 | DARMSTADT

Sign up now at sqlkonferenz.de



Agenda

- So, what's the problem?
- Where your data „lives“: SQL Server database, tables and indexes
- Queries, the optimizer and the query plans
- What the optimizer looks at: statistics
- So, what could be the real problem - and how do we solve it?
- Some additional aspects of cardinality estimation
- Round-up, resources & credits

So, what's the problem?

Sales Order Processor - Microsoft Dynamics NAV client - connected to Dynamics 365 Business Central

CRONUS AG Home Search (Strg+F3)

ACTIONS

Sales Order Processor

Role Center

- Sales Orders
- Sales Orders - Microsoft Dyna...
- Sales Quotes
- Blanket Sales Orders
- Sales Invoices
- Sales Return Orders
- Sales Credit Memos
- Items
- Customers
- Item Journals
- Sales Journals
- Cash Receipt Journals
- Transfer Orders

Home

Sales

Purchasing

Inventory

Posted Documents

Self-Service

Departments

Activities

For Release

- Sales Quotes - Open 0
- Sales Orders - Open 20

Sales Orders Released Not Shipped

- Ready To Ship 6
- Partially Shipped 0
- Delayed 14
- Average Days Delayed 7,5

Returns

- Sales Return Orders - Open 0
- Sales Credit Memos - Open 1

My User Tasks

- Pending User Tasks 0

Self-Service

Time Sheets

- Open Time Sheets 0

Pending Time Sheets

- Submitted Time Sheets 0
- Rejected Time Sheets 0
- Approved Time Sheets 0

Approvals

- Requests to Approve 0

Power BI Reports

- Select Report
- Expand Report
- Previous Report
- Next Report
- Refresh Page

Get started with Power BI

My Items

Manage List Open New Find Filter Clear Filter

Item No. Description

CRONUS AG Donnerstag, 27. Januar 2022 WIN10SQL\VM

So, what's the problem?

Customers - Microsoft Dynamics NAV client - connected to Dynamics 365 Business Central

CRONUS AG Home Customers

Search (Strg+F3)

HOME ACTIONS NAVIGATE REPORT CRONUS AG ?

New Edit View Delete Cash Receipt Journal Customer - Order Summary Customer - Balance to Date Customer - Sales List Customer - Payment Receipt Reminder Sales Order Sales Quote Sales Credit Memo Aged Accounts Receivable Sales Invoice New Document Customer View Show Attached Notes Links Refresh Clear Filter Find

Role Center

- Sales Orders
- Sales Orders - Microsoft Dyna...
- Sales Quotes
- Blanket Sales Orders
- Sales Invoices
- Sales Return Orders
- Sales Credit Memos
- Items
- Customers**
- Item Journals
- Sales Journals
- Cash Receipt Journals
- Transfer Orders

Home Sales Purchasing Inventory Posted Documents Self-Service Departments

Customers

No.	Name	Contact	Balance (LCY)	Balance Due (LCY)	Sales (LCY)	Payments (LCY)
01121212	Spotsmeyer's Furnishings	Mr. Mike Nash	0,00	0,00	0,00	0,00
01445544	Progressive Home Furnishin...	Mr. Scott Mitchell	2.324.952,63	0,00	2.324.952,63	0,00
01454545	New Concepts Furniture	Ms. Tammy L. McDonald	344.207,60	0,00	0,00	0,00
01905893	Candoxy Canada Inc.	Mr. Rob Young	0,00	0,00	0,00	0,00
01905899	Elkhorn Airport	Mr. Ryan Danner	0,00	0,00	0,00	0,00
01905902	London Candoxy Storage C...	Mr. John Kane	0,00	0,00	0,00	0,00
10000	Möbel-Meller KG	Herr Michael Emanuel	30.734.326,65	0,00	26.788.489,19	222.829,80
20000	Blütenhaus GmbH	Herr Paul West	13.740.452,39	1.237.956,95	12.414.279,43	190.060,72
20309920	Metatorad Malaysia Sdn Bhd	Mrs. Azleen Samat	0,00	0,00	0,00	0,00
20312912	Highlights Electronics Sdn B...	Mr. Mark Darrell Boland	0,00	0,00	0,00	0,00
20339921	Trax Tonic Sdn Bhd	Mrs. Rubina Usman	0,00	0,00	0,00	0,00
21233572	Somadis	M. Syed ABBAS	0,00	0,00	0,00	0,00
21245278	Maronegoce	Mme. Fadoua AIT MOUSSA	0,00	0,00	0,00	0,00
21252947	ElectroMAROC		0,00	0,00	0,00	0,00
27090917	Zanlan Corp.	Mr. Derik Stenerson	0,00	0,00	0,00	0,00
27321782	Karoo Supermarkets	Mr. Pieter Wycoff	0,00	0,00	0,00	0,00
27489991	Durbandit Fruit Exporters	Mr. Eric Lang	0,00	0,00	0,00	0,00
30000	Gilde Jupiter Versicherungs ...	Frau Claudia Mangel	10.780.278,42	170.399,26	9.459.149,70	0,00
31505050	Woonboulevard Kuitenbrou...	Maryann Barber	0,00	0,00	0,00	0,00
31669966	Meersen Meubelen	Michael Vanderhyde	0,00	0,00	0,00	0,00
31987987	Candoxy Nederland BV	Rob Verhoff	0,00	0,00	0,00	0,00
32124578	Nieuwe Zandpoort NV	Kevin Verboort	0,00	0,00	0,00	0,00
32656565	Antarcticopy	Michael Zeman	4.004.250,25	4.004.250,25	4.004.250,25	0,00
32789456	Lovaina Contractors	Hans Visser	0,00	0,00	0,00	0,00
33000019	Francematic	M. Herve BOURAIMA	0,00	0,00	0,00	0,00
33002984	Parmentier Boutique	M. Jean E. TRENARY	0,00	0,00	0,00	0,00

Customer Statistics

Balance (LCY):	13.740.452,39
Sales	
Outstanding Orders (LCY):	9.627,00
Shipped Not Invd. (LCY):	0,00
Outstanding Invoices (LCY):	0,00
Service	
Outstanding Serv. Orders (LCY):	0,00
Serv Shipped Not Invoiced(LCY):	0,00
Outstanding Serv.Invoices(LCY):	0,26
Payments	
Payments (LCY):	190.060,72
Last Payment Receipt Date:	13.01.2022
Total (LCY):	13.750.079,65
Credit Limit (LCY):	0,00
Overdue Amounts (LCY) as of 27.01.22:	1.237.956,95
Total Sales (LCY):	12.414.279,43
Invoiced Prepayment Amount (LCY):	0,00

Links

Link Address	Description	Cre...
There is nothing to show in this view.		

Notes

CRONUS AG Donnerstag, 27. Januar 2022 WIN10SQL\VM

Where your data lives: db, tables and indexes

- The developer's view of the data structure

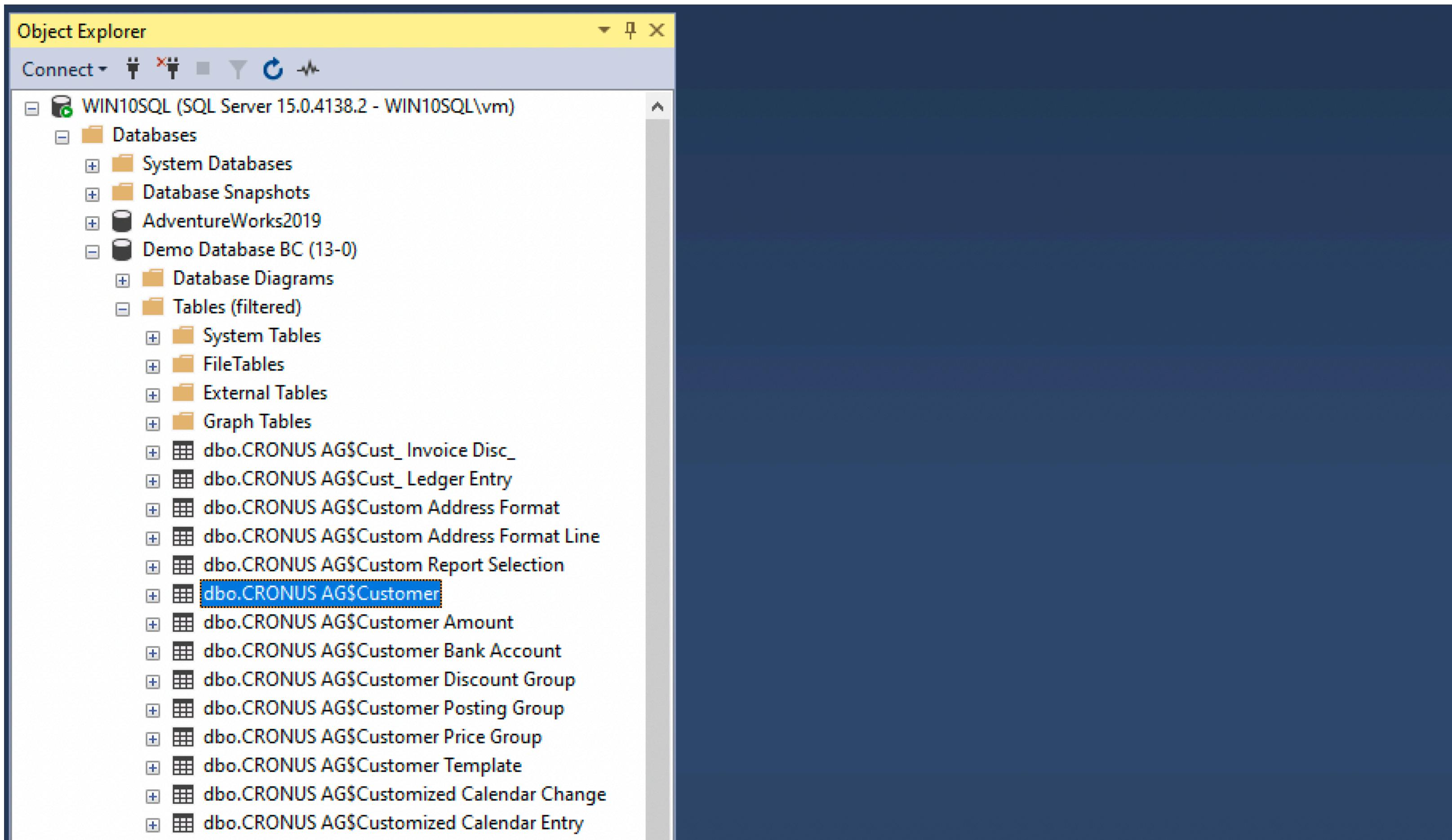
The screenshot shows the Microsoft Dynamics NAV Development Environment's Object Designer window. The title bar reads "Demo Database BC (13-0) - Microsoft Dynamics NAV Development Environment - [Object Designer]". The menu bar includes File, Edit, View, Tools, Window, and Help. The toolbar below has icons for剪切 (Cut), 复制 (Copy), 粘贴 (Paste), 搜索 (Search), and other navigation functions.

The left sidebar contains a tree view with categories: Table (selected), Page, Report, Codeunit, Query, XMLport, and MenuSuite. Below the tree is an "All" button. The main area is a grid table with columns: Type, ID, Name, and Modified (which is empty). The grid lists various system objects:

Type	ID	Name	Modified
Table	6	Customer Price Group	
Table	7	Standard Text	
Table	8	Language	
Table	9	Country/Region	
Table	10	Shipment Method	
Table	11	Country/Region Translation	
Table	13	Salesperson/Purchaser	
Table	14	Location	
Table	15	G/L Account	
Table	17	G/L Entry	
Table	18	Customer	
Table	19	Cust. Invoice Disc.	
Table	21	Cust. Ledger Entry	
Table	23	Vendor	
Table	24	Vendor Invoice Disc.	
Table	25	Vendor Ledger Entry	
Table	27	Item	
Table	30	Item Translation	
Table	31	Item Picture Buffer	
Table	32	Item Ledger Entry	

Where your data lives: db, tables and indexes

- The DBA's view of the data structure



Where your data lives: db, tables and indexes

- No more native database since Nav 2013, only SQL Server

- No object IDs - Naming convention for tables in SQL db:

<Schema name>.<Company name>\$<Table name>\$<App-GUID>

If DataPerCompany = Yes: dbo.CRONUS AG\$Customer

If DataPerCompany = No: dbo.Company

- Special characters in names get converted to “_”

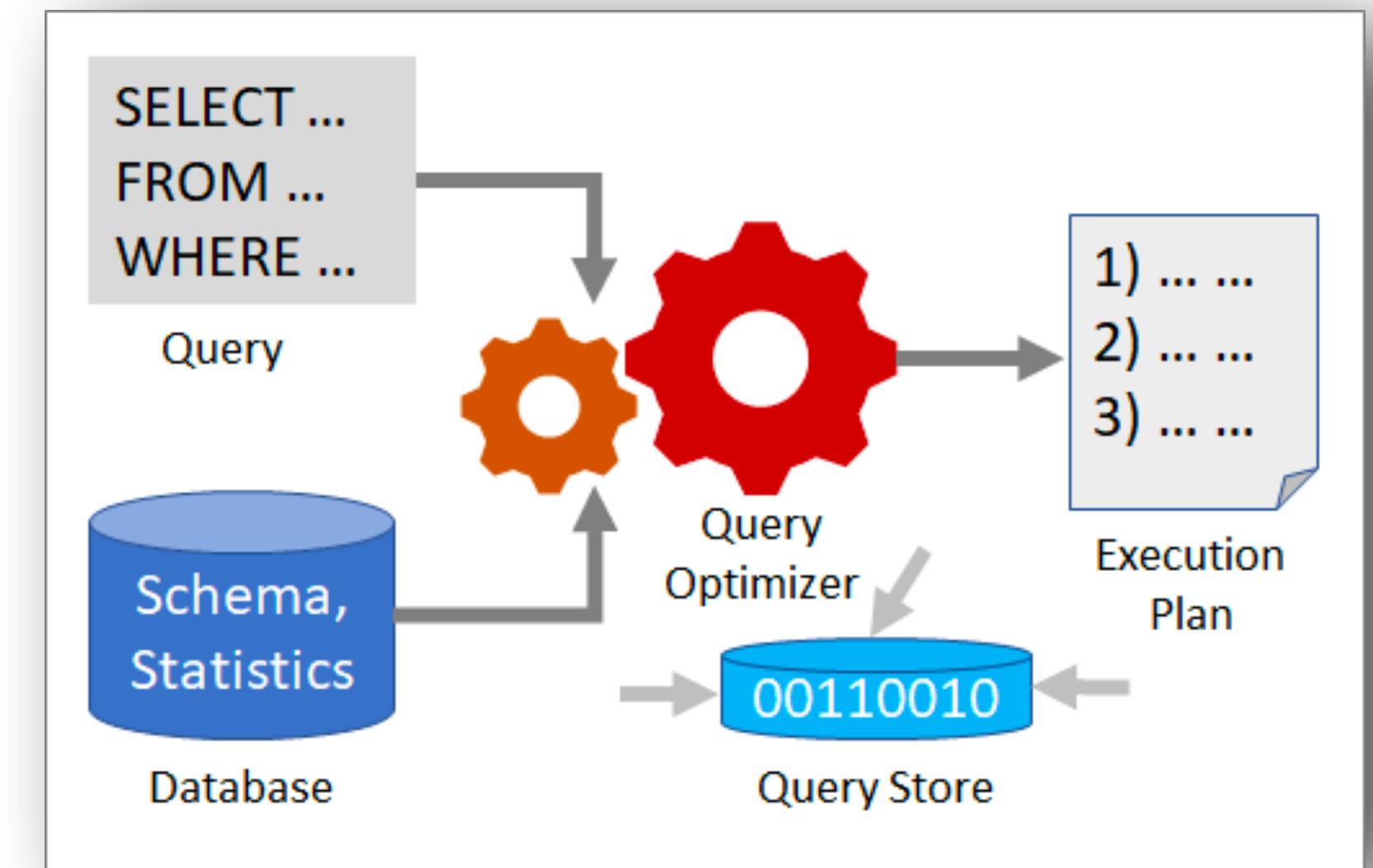
(according to: Alter database - Integration - Convert identifiers):

Cust. Ledger Entry becomes dbo.CRONUS AG\$Cust_Ledger Entry

- Tools for managing your data platform: SQL Server Management Studio (Windows), Azure Data Studio (all platforms)

Queries, the optimizer, query plans and query store

- Query: an inquiry to the server in T-SQL
 - DML: store, modify, retrieve or delete data
 - DDL: create, modify or delete DB objects
- Query optimizer: analyzes your query and determines the most efficient plan in terms of CPU, memory and I/O.
- Query / execution plan: a sequence of steps that the SQL Server engine has to follow in order to return the data the user asked for.
- Query store (introduced with SQL Server 2016, SSMS 16): captures the history of queries, plans, runtime statistics per database.
- Statistics: the magic that helps the optimizer with cardinality estimation



What the optimizer looks at: statistics

„Statistics for query optimization are binary *large* objects (BLOBs) that contain statistical information about the distribution of values in one or more columns of a table or indexed view.“ [Microsoft]

Properties window or **DBCC Show_Statistics**:

- General
 - Table name, Statistics name, Statistics columns (editable), Time last updated
- Details
 - Name, time updated, Rows in table, Rows sampled, Steps (in histogram), Avg key length, String Index, Filter props, Persisted Sample Percent
- Density vector
 - All density = $1 / \text{number of distinct values in this level}$,
 - Average length of key values, Columns in key level

What the optimizer looks at: statistics

- Histogram is computed only on the first column, max 200 steps:
 - Range_Hi_Key = column value at the top of this step
 - Range_Rows = number of rows within this interval (excl. high and low key)
 - Eq_Rows = number of rows equalling high key
 - Distinct_Range_Rows = number of distinct values within interval (excl. high/low)
 - Avg_Range_Rows = average number of rows per distinct value
- Filter tab
 - Any filter expressions used
- **EXEC sp_helpstats <table name>, 'ALL'** -> deprecated
- **sys.stats and sys.stats_columns**
- **sys.dm_db_stats_properties and sys.dm_db_stats_histogram**

So, what could be the real problem and how do we solve it?

- Outdated statistics
 - **Auto_Update_Statistics** is set to off
applies to Index stats, single-column stats, manual stats
 - Recompilation threshold („Tipping point“) not hit

Table type	Table cardinality	Recomp threshold <= SQL 2014 (120)	Recomp threshold >= SQL 2016 (130) *
Temporary	$n < 6$	6	6
Temporary	$6 \leq n \leq 500$	500	500
Permanent	< 500	500	500
Temporary or Permanent	> 500	$500 + 0.2 * n$	$\text{Min}(500 + 0.2 * n, \sqrt{1000 * n})$

* use Trace Flag 2371 in SQL Server 2008R2 through 2014/compat level <= 120

So, what could be the real problem and how do we solve it?

- Incomplete (sampled) statistics
 - Tables < 8 MB get full scan, > 8 MB use a non-linear algorithm with decreasing sample rate for increasing table size
 - **Update Statistics <Table name> <Stats name> With**
 - **Fullscan** scans the whole table vs **Sample n Percent | Rows**
 - **Resample** uses the most recent sample rate for each statistic
 - **Persist_Sample_Percent = On** retains sample rate for subsequent statistic updates
 - **Incremental = On** One statistic per partition
 - **All | Columns | Index**
 - **EXEC sp_updatestats;** updates all stats in the database

Some additional aspects of cardinality estimation

- Parameterized queries:
estimated number of rows = density * total number of rows
- Inequality predicate:
estimated number of rows = 30% of total number of rows
- For missing statistics:
estimated number of rows = square root of total number of rows
- For correlated columns „exponential backoff“ is used:
 $C_{table} * S_0 * \sqrt[2]{S_1} * \sqrt[4]{S_2} * \sqrt[8]{S_3} \dots$

Some additional aspects of cardinality estimation

- Missing statistics might be due to
 - AUTO CREATE STATISTICS is OFF
 - Using table variables
 - XML or spatial data types
 - Remote queries / linked servers
 - Read-only databases, readable secondaries, db snapshots

What the optimizer looks at: statistics

Options for statistics, at the database level:

- **Auto_Create_Statistics**

Query optimizer creates statistics on individual columns (that do not already have a histogram), as necessary. Stat name starts with `_WA`

- **Auto_Update_Statistics**

Optimizer updates out-of-date stats (determined by the recompilation threshold), only when they are used by a query

- **Auto_Update_Statistics_Async**

Off = Sync: optimizer waits for stats update before compiling/executing

On = Async: compiles/executes first, updates stats as background task thereafter; only the following queries benefit from updated stats

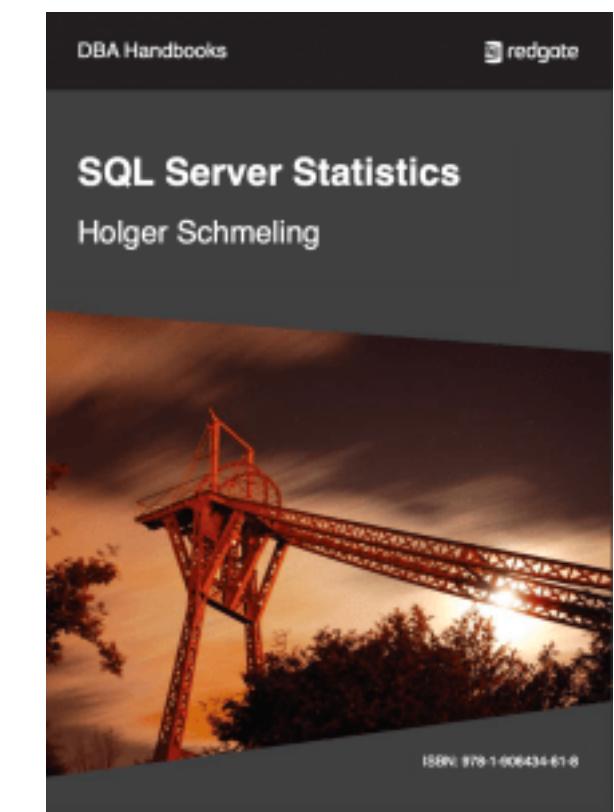
Resources

- Query processing architecture guide
<https://docs.microsoft.com/en-us/sql/relational-databases/query-processing-architecture-guide>
- Execution plans
<https://docs.microsoft.com/en-us/sql/relational-databases/performance/execution-plans>
- Microsoft docs on statistics
<https://docs.microsoft.com/en-us/sql/relational-databases/statistics/statistics>
- Search the web for „SQL Server“ and
 - statistics
 - cardinality estimation

Resources

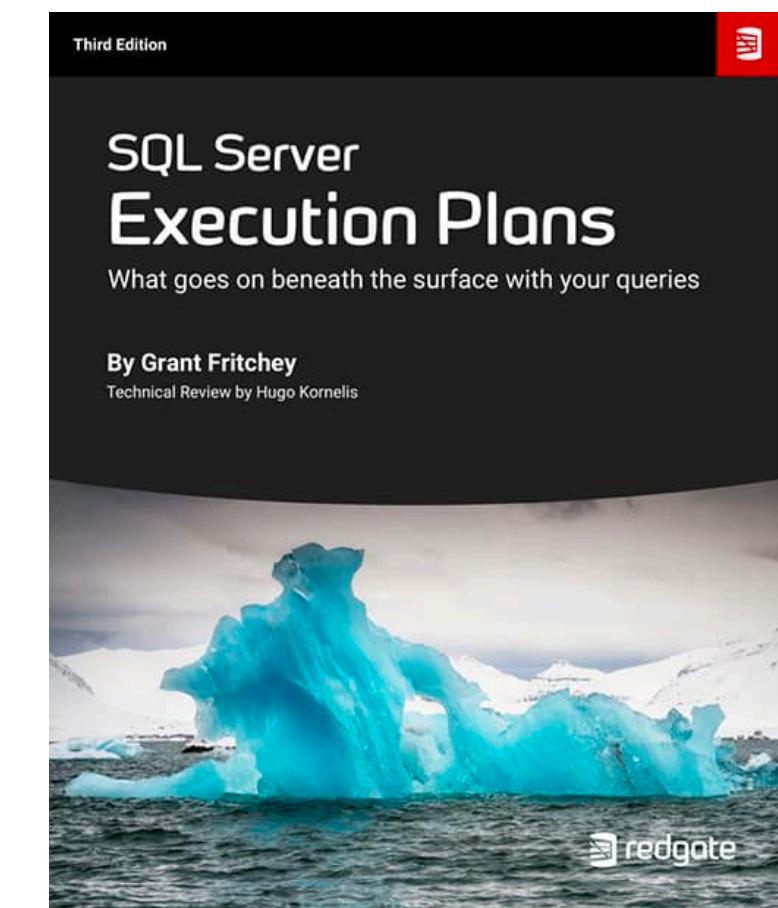
- eBook „SQL Server Statistics“, 44 pp,
by Holger Schmeling, 2010

<https://www.red-gate.com/library/sql-server-statistics>



- eBook „SQL Server Execution Plans“, 515 pp,
by Grant Fritchey, 3. ed 2018

<https://www.red-gate.com/products/dba/sql-monitor/entrypage/execution-plans>



What are index statistics ... and why should the developer care?

Thank you for your time and interest & keep in touch:

-  @DerFredo <https://twitter.com/DerFredo>
-  de.linkedin.com/in/derfredo
-  www.xing.com/profile/Thomas_Huetter



This file and the demo scripts can be found at:

<https://bit.ly/DerFredoErding21>

Vielen Dank!

Ich freue mich auf Feedback!
Euer Thomas

Bitte hier entlang:

<https://forms.office.com/r/4QuTgXUaQA>