



A Query Tuner's Practical Guide to Statistics

Andy Yun (he/him)

Field Solutions Architect – Pure Storage

sqlbek@gmail.com | @sqlbek

<https://sqlbek.wordpress.com>





Andy Yun

Field Solutions Architect

- SQL Server, T-SQL & DB Development



Group

Director-at-Large



@SQLBek – sqlbek@gmail.com

<https://sqlbek.wordpress.com/>

<https://www.github.com/sqlbek/>



A decorative network diagram at the top of the slide, featuring a series of interconnected nodes and lines. The nodes are represented by small circles, some of which are highlighted with a dashed border. The lines are thin and gray, creating a web-like structure that spans the width of the slide.

“

*How does SQL Server
come up with estimates?*

A decorative network diagram at the top of the slide, featuring a series of interconnected nodes and lines. The nodes are represented by small circles, some of which are highlighted with a dashed border. The lines are thin and gray, creating a web-like structure that spans the width of the slide.

“

*Why are SQL Server's
estimates often
ridiculously inaccurate?*

A decorative network diagram at the top of the slide, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are highlighted with larger, dashed outlines. The lines connecting them form a dense, branching structure. A central node is highlighted with a larger, dashed circle, and a blue double quote symbol is placed inside it.

“

*Why won't running
UPDATE STATISTICS
solve everything*

Today's Agenda

- Statistics Overview

- Statistics Intermediate Level Session

- Demos Will NOT cover the best ways to **MAINTAIN** statistics

- A Slight Tangent

Demos - 2

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are solid grey and others are hollow with a grey outline. The lines are thin and grey, creating a mesh-like structure that extends from the top-left towards the center of the slide.

Statistics Overview

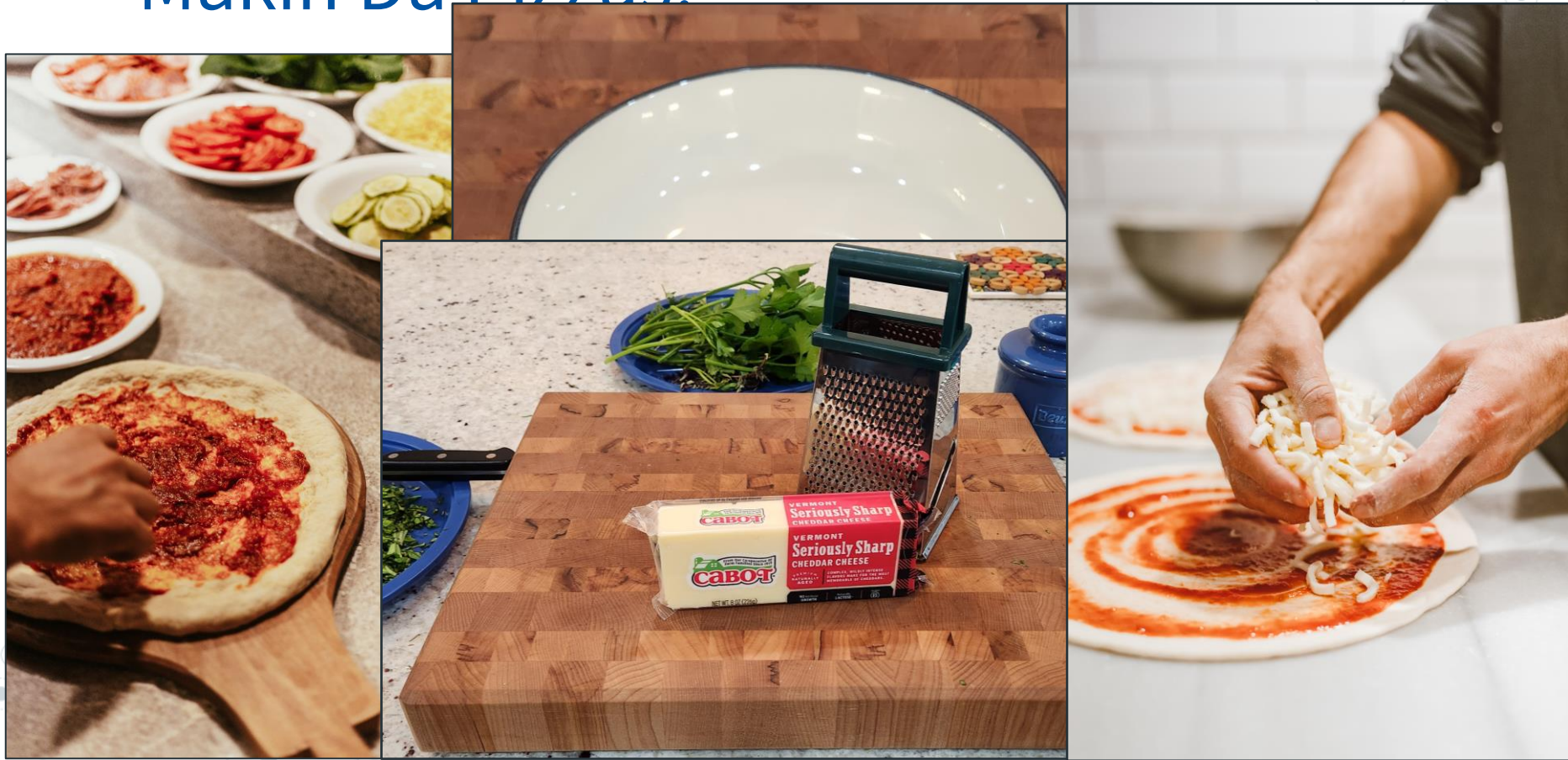
Even a General Understanding
Will Benefit You

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It consists of a cluster of interconnected nodes and lines. The nodes are small circles, some solid grey and some hollow with a grey outline, connected by thin grey lines. This diagram is positioned in the bottom-right corner, mirroring the style of the top-left decoration.

I Love Analogies... and Pizza

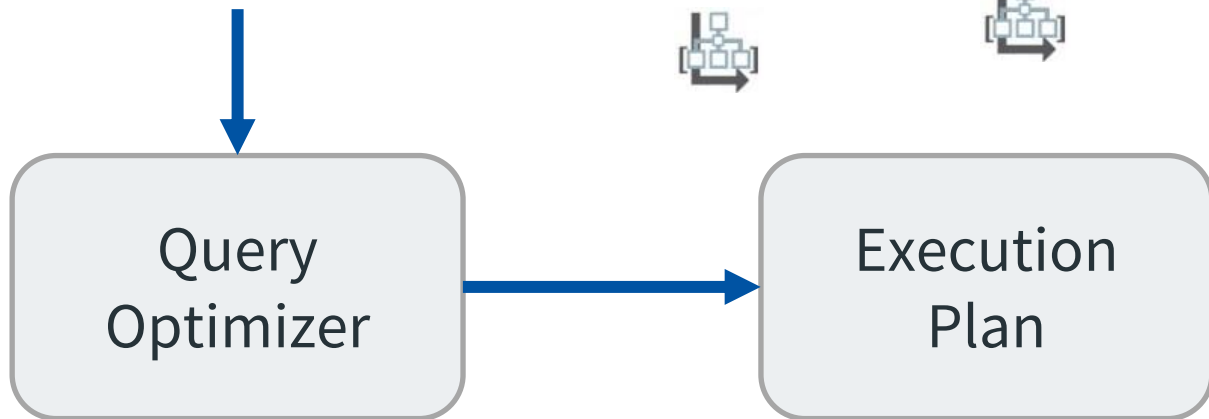


Makin Da Pizzas!

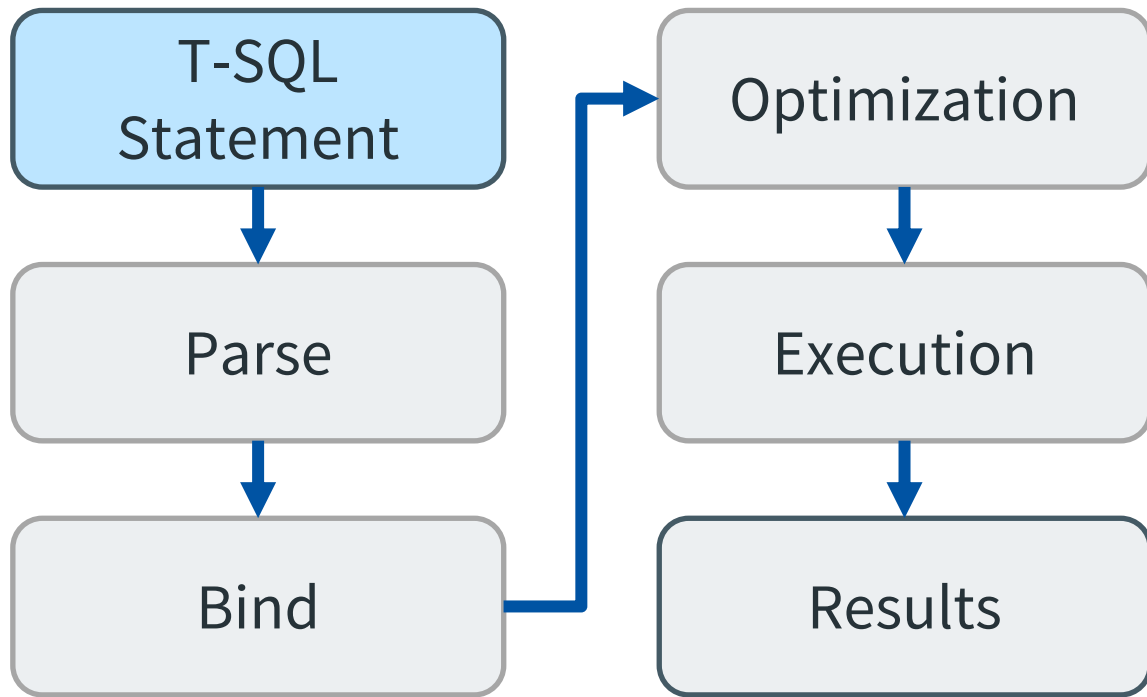


Query Lifecycle

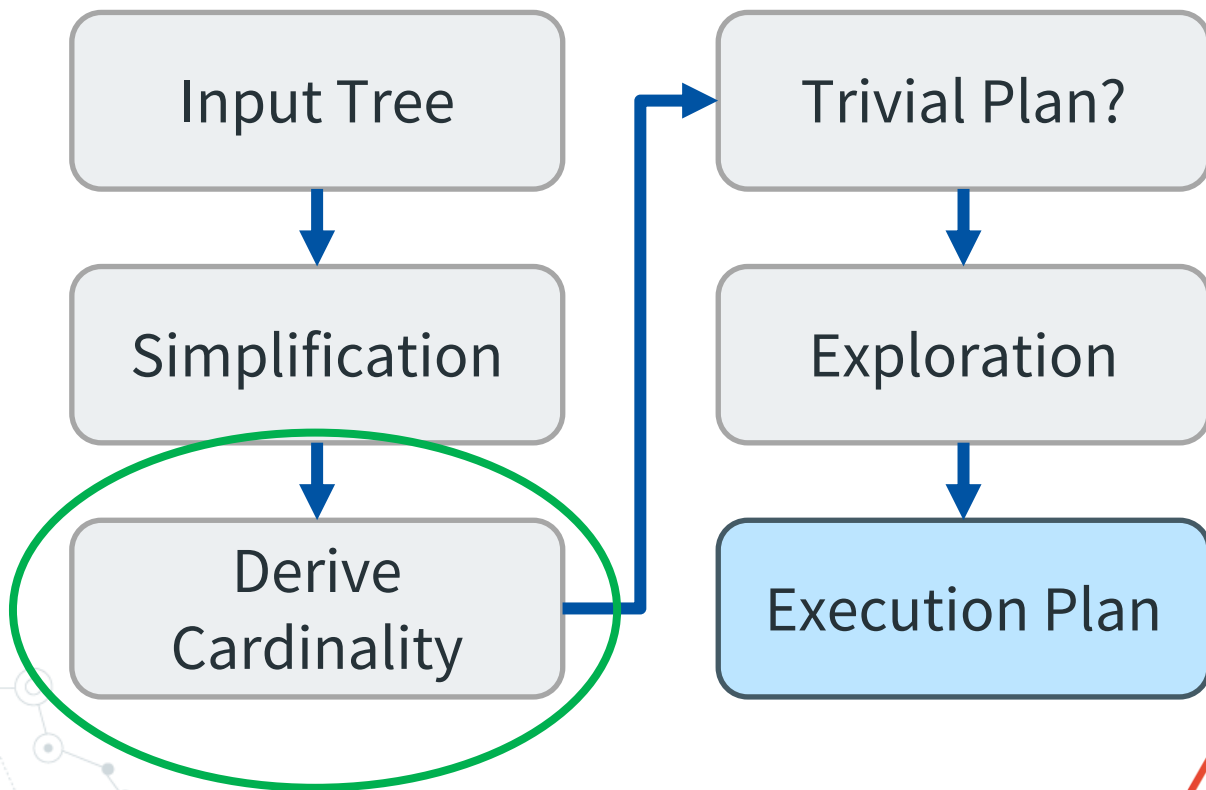
```
SELECT  
  ColA, ColB, ...  
FROM Table_One  
INNER JOIN Table_Two  
  ON Table_One.Pkey = Table_Two.Fkey
```



Query Optimizer



Optimization Steps



Cardinality? You mean the bird?

- ◎ Uniqueness of the elements in a set of data
- ◎ High Cardinality
 - Many unique values
- ◎ Low Cardinality
 - Fewer unique values



Cardinality Synonyms

- ◎ Uniqueness – *Not “Unique”*
- ◎ Selectivity – High or Low
- ◎ Distinct Different Values – Many or Few

Cardinality Examples – High or Low?

Of everyone here,
what is your... ?

- ◎ Job Title
- ◎ First Name
 - + Last Name?

Other Examples

- ◎ Social Security Number
- ◎ Birthdate
 - With or without year?

Cardinality Impacts...

- ◎ What index to use
- ◎ Whether to seek or scan
- ◎ How much memory is needed
 - i.e. Memory Grants

Remember

Most of the choices made
by the optimizer are driven
by cardinality estimation.

-Paul White



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels of connectivity or importance. The lines are thin and gray, creating a mesh-like structure.

Statistics Internals

Internals You Ought To Know

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with some nodes being larger and more prominent than others, indicating a hierarchical or central structure within the network.

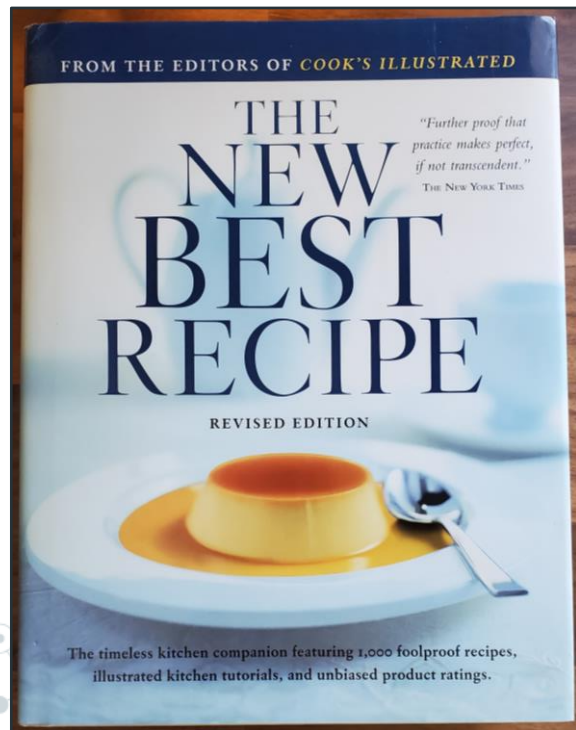
Statistics Are Stored About What?

- ◎ Indexes – clustered & non-clustered
- ◎ Columns
- ◎ What about heaps?

How Can We Query Statistics?

- ◎ sys.stats
- ◎ sys.stats_columns
- ◎ sys.dm_db_stats_properties()
- ◎ sys.dm_db_stats_histogram()
- ◎ DBCC SHOW_STATISTICS()

Recipes Galore!



A

All-Purpose Gravy, 188
 All-Season Blueberry Cobbler
 (with Frozen Blueberries),
 942
Almond(s)
 -Apricot Oatmeal Scones, 716
 -Apricot Sour Cream Coffeecake,
 840
 Granola Bars
 Coconut and Sesame, 822
 Crunchy, 821-22
 Dried Cranberry and Ginger, 822
 Linzertorte, 925-28
 Nut Oatmeal Cookies, 783
 -Orange Biscotti, 805-6
 and Orange Oatmeal Cookies, 783
 Peanuts, and Pumpkin Seeds,
 Mexican-Spiced, 3
 -Raspberry Coffeecake, 838
 Sandies, 785
 Toasted, and Coconut, Chocolate
 Chip Cookies with, 777
 Toasted, and Orange, Rice Pudding
 with, 956
 Toasted Slivered, Fish Meunière
 with, 501
**American Potato Salad with
 Hard-Boiled Eggs and
 Sweet Pickles, 98-100**
American Sandwich Bread, 725-29
 Buttermilk, 729
 Kneaded in a Food Processor, 729
 Oatmeal, 729
 Slow-Rise, 729
**American-Style Soda Bread with
 Raisins and Caraway Seeds,
 698-99**
Anchovy(ies)
 and Basil, Deviled Eggs with, 23
 Broccoli, and Garlic, Spaghetti with,
 265-66
 Garlic, and White Wine, Braised
 Cauliflower with, 154-55
 -Garlic Butter with Lemon
 and Parsley, 506
 mincing, 266

Anchovy(ies) (cont.)

and Mustard, Madeira Pan Sauce
 with, 391
 Spaghetti Puttanesca, 248-50
Angel Food Cake, 825-27
Appetizers
 Almonds, Peanuts, and Pumpkin
 Seeds, Mexican-Spiced, 3
 Asparagus Wrapped with
 Prosciutto, 19
 Bruschetta
 with Arugula, Red Onion,
 and Rosemary-White Bean
 Spread, 18
 with Sautéed Sweet Peppers,
 17-18
 Toasted Bread for, 17
 with Tomatoes and Basil, 17
 Buffalo Wings, 26-28
 Cashews and Pistachios,
 Indian-Spiced, with Currants, 4
 Cheese Straws, 8-9
 cheese trays, preparing, 5-6
 Crudités, 9-11
 Dates Stuffed with Parmesan, 7
 Deviled Eggs
 with Anchovy and Basil, 23
 Classic, 22-23
 with Tuna, Capers, and Chives,
 23
 Goat Cheese, Marinated, 6
 Melon and Prosciutto, 18-19
 Olives, Black and Green, Marinated,
 4-5
 Pecans, Spiced, with Rum Glaze,
 2-3
 Pita Chips, 16
 Quesadillas, Spicy Roasted
 Red Pepper, 23-24
 Shrimp, Herb-Poached, 19-21
 Tarts
 Tomato and Mozzarella, 28-30
 Tomato and Mozzarella,
 with Prosciutto, 30
 Tomato and Smoked Mozzarella,
 30
 see also Dips and spreads

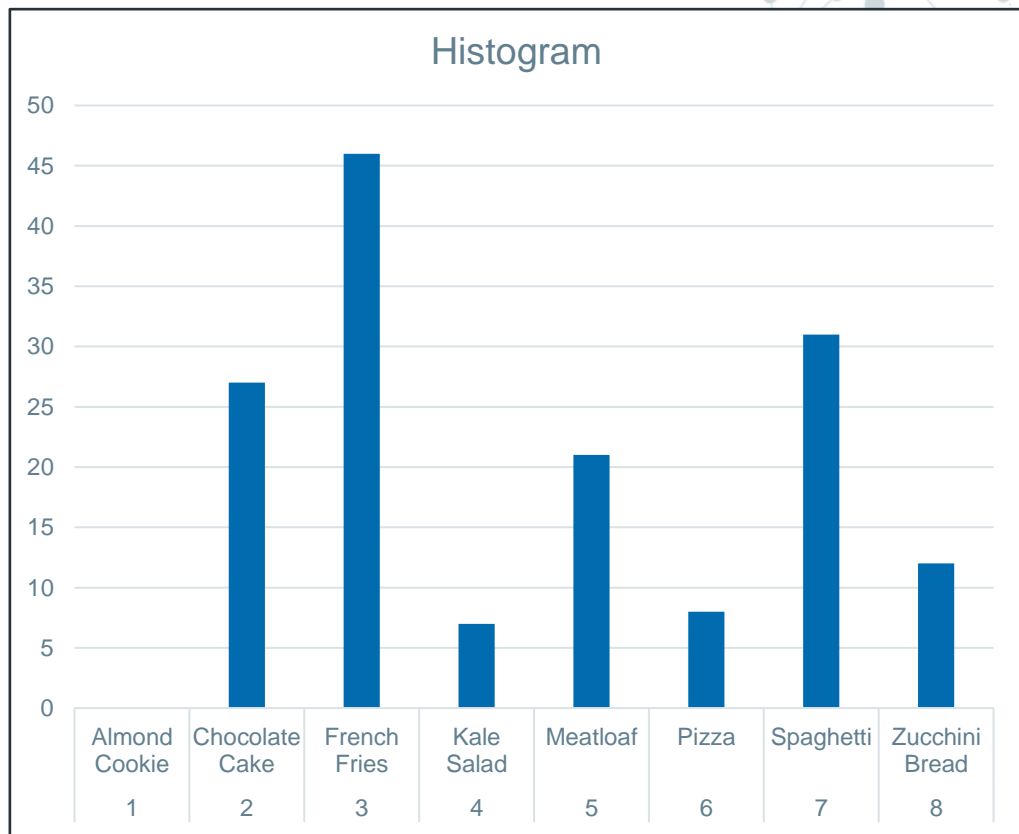
Apple Cider and Brown Sugar Glaze, 482-83

Apple(s)

Bacon, Sage, and Caramelized
 Onions, Bread Stuffing with,
 369
 and Cabbage Braised in Cider,
 148
 and Cider, Smothered Pork Chops
 with, 463
 -Cinnamon Coffeecake, 837
 coring, 935
 Cream, and Sage, Sautéed Pork
 Tenderloin Medallions with,
 469
 Crisp, 930-31
 and Currants, Curried Tuna Salad
 with, 119
 Curried, Macaroni Salad with, 104
 Modern Minicmeat Pie, 898-99
 Pancake, German, 936-37
 Pancake, German, with Caramel
 Sauce, 937
 Pandowdy, 931-33
 Pie
 Classic, 886-88
 with Crystallized Ginger, 888
 with Dried Fruit, 888
 Dutch, 888-90
 Dutch, Quick, 890
 with Fresh Cranberries, 888
 Strudel, Quick, 938-40
 Tarte Tatin, 914-17
 Waldorf Chicken Salad, 121
Apricot(s)
 -Almond Oatmeal Scones, 716
 -Almond Sour Cream Coffeecake,
 840
 and Corn Muffins with Orange
 Essence, 708
 dried, cutting, tip for, 560
 Dried, Sausage, and Pecans,
 Bread Stuffing with, 368
 Lamb Tagine, 558-60
Artichokes
 about, 125-26
 Baby, Roasted, 128

Let's Sample & Summarize

	Recipe Name	How Many In Between?
1	Almond Cookies	0
2	Chocolate Cake	27
3	French Fries	46
4	Kale Salad	7
5	Meatloaf	21
6	Pizza	8
7	Spaghetti	31
8	Zucchini Bread	12



sys.dm_db_stats_histogram()

range_high_key	Actual Data/Key Value (R.H.K)
equal_rows	Est. # of rows equal to R.H.K
range_rows	Est. # of rows between R.H.K and prior R.H.K.
distinct_range_rows	Est. # of distinct VALUES between R.H.K and prior R.H.K.
average_range_rows	$\text{range_rows} / \text{distinct_range_rows}$

Let's Apply That...

	Recipe Name	How Many In Between?
1	Almond Cookies	0
2	Chocolate Cake	27
3	French Fries	46
4	Kale Salad	7
5	Meatloaf	21
6	Pizza	8
7	Spaghetti	31
8	Zucchini Bread	12

range_high_key – ?

equal_rows – ?

range_rows – ?

distinct_range_rows – ?

average_range_rows – ?



Demos - 1

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, while others are smaller and solid. The lines are thin and gray, connecting the nodes in a non-linear fashion.

Join Types

A Slight Tangent

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with some nodes being larger and having concentric circles, and others being smaller and solid. The lines are thin and gray.

Physical Join Operators



Nested Loop



Merge



Hash Match

Merge



- ◎ Most effective
- ◎ Requires both sets of data to be pre-sorted

Hash Match vs Nested Loop

Nested Loop



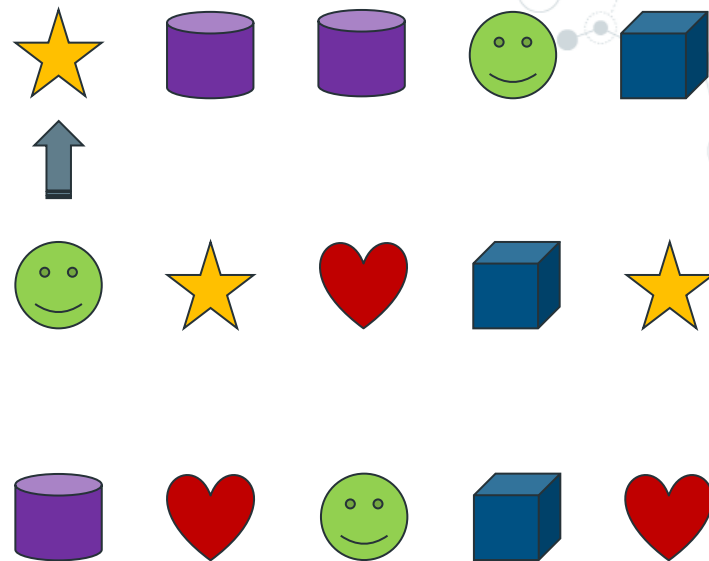
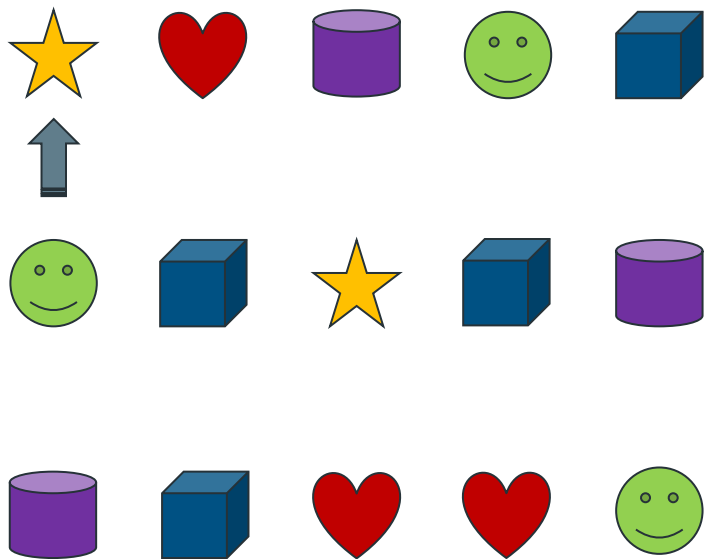
- ⦿ Takes a value, compares to all others in dataset
- ⦿ Preferred if one dataset is much smaller

Hash Match



- ⦿ Hashes all values
- ⦿ Preferred for two large, unsorted datasets
- ⦿ Subject to Spilling

SQL Server Is Set-Based... Right???



Remember

Nested Loops are great...

Until they're not...





Demos - 2

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting a hierarchical or central structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.

Parting Thoughts

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with some nodes being larger and more prominent than others, creating a sense of depth and complexity.

A decorative network diagram at the top of the slide, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are highlighted with larger, concentric circles. The lines connecting the nodes are thin and grey. The overall structure is a dense, interconnected mesh that spans the width of the slide.

“

*How in-depth do I need
to know statistics to be
an effective query tuner?*

A Matter of Perspective

- ◎ Statistics are important, but...
- ◎ Focus and tune ONE query?
- ◎ Focus and tune a WORKLOAD?
- ◎ Statistics only one piece of the execution plan

Learn More: Query Related

Troubleshooting Plans: P1-Root Cause Analysis/Caching Probs – Kimberly Tripp Randal

<https://www.youtube.com/watch?v=Sbqg0U5ZBZc>

Troubleshooting Plans: P2-Statistics Problems – Kimberly Tripp Randal

<https://www.youtube.com/watch?v=76scR-Y4lwY>

Paying Attention to Estimates– Aaron Bertrand

<https://sqlperformance.com/2016/07/t-sql-queries/paying-attention-estimates>

Understanding Statistics in SQL Server – Erin Stellato

<https://www.youtube.com/watch?v=XQynldV3d8U>

Can I Use Statistics to Design Indexes? – Kendra Little

<https://littlekendra.com/2016/10/06/can-i-use-statistics-to-design-indexes-dear-sql-dba-episode-18/>

Cardinality Blogs Category – Erik Darling

<https://www.erikdarlingdata.com/category/cardinality/>

Learn More: Managing Statistics

Updating Statistics in SQL Server – Kendra Little

<https://www.youtube.com/watch?v=fzy8AnF6UQ8>

Updating Statistics in SQL Server: Maintenance Questions & Answers – Kendra Little

<https://littlekendra.com/2016/04/18/updating-statistics-in-sql-server-maintenance-answers/>

Updating Database Statistics Pt. 1– Marsha Pierce

<https://marshapiercedba.com/updating-database-statistics-pt-1/>

Statistics Blogs Category – Erik Darling

<https://www.erikdarlingdata.com/category/statistics/>

Learn More: Advanced Resources

Cardinality Estimation: Combining Density Statistics – Paul White

<https://sqlperformance.com/2017/08/sql-optimizer/combining-density>

Cardinality Estimation for Multiple Predicates – Paul White

<https://sqlperformance.com/2014/01/sql-plan/cardinality-estimation-for-multiple-predicates>

Cardinality Estimation for Disjunctive (OR) Predicates in SQL Server 2014 Onward – Paul White

<https://www.sql.kiwi/2014/04/cardinality-estimation-for-disjunctive-predicates-in-2014.html>

Query Optimizer Deep Dive – Paul White

<https://www.youtube.com/watch?v=3l9mHjait2I>

Statistics & Cardinality Estimator Model Variations – “SQL Scotsman”

<https://sqlserverscotsman.wordpress.com/2016/11/10/statistics-cardinality-estimator-model-variations/>

Optimizing Your Query Plans with the SQL Server 2014 Cardinality Estimator – Joe Sack

[https://learn.microsoft.com/en-us/previous-versions/dn673537\(v=msdn.10\)](https://learn.microsoft.com/en-us/previous-versions/dn673537(v=msdn.10))

Learn More: Demo References

<https://www.sqlskills.com/blogs/paul/how-are-auto-created-column-statistics-names-generated/>

<https://sqlserverscotsman.wordpress.com/2016/11/10/statistics-cardinality-estimator-model-variations/>

<https://sqlperformance.com/2014/01/sql-plan/cardinality-estimation-for-multiple-predicates>

<https://www.sql.kiwi/2014/04/cardinality-estimation-for-disjunctive-predicates-in-2014.html>

<https://dba.stackexchange.com/questions/312816/how-does-sql-estimate-the-number-of-rows-in-a-less-than-predicate>

<https://dba.stackexchange.com/questions/148523/cardinality-estimation-for-and-for-intra-step-statistics-value/169384#169384>

<https://www.brentozar.com/archive/2020/11/paul-white-explains-temp-table-caching-3-ways/>

<https://www.sql.kiwi/2012/08/temporary-tables-in-stored-procedures.html>

<https://www.sql.kiwi/2012/08/temporary-object-caching-explained.html>

Thank you

Any Questions?

Andy Yun

sqlbek@gmail.com | @sqlbek

<https://sqlbek.wordpress.com>

<https://github.com/sqlbek>



Session evaluation

Your feedback is important to us



Evaluate this session at:

www.PASSDataCommunitySummit.com/evaluation

