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 S… DBA & DB Deve…
- …er …
- …oup …
- …ion – Director-at-Large

MVP Microsoft® Most Valuable Professional Reconnect

PURESTORAGE®

@SQLBek – sqlbek@gmail.com

https://sqlbek.wordpress.com/

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

PASS Data Community SUMMIT 2022

How does SQL Server come up with estimates?

Why are SQL Server's estimates often ridiculously inaccurate?

Why won't running UPDATE STATISTICS solve everything

Today's Agenda

- Statistics Overview
- Statisti
- Demos
- A Slight Tangent
- Demos - 2

Intermediate Level Session

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

Statistics Overview

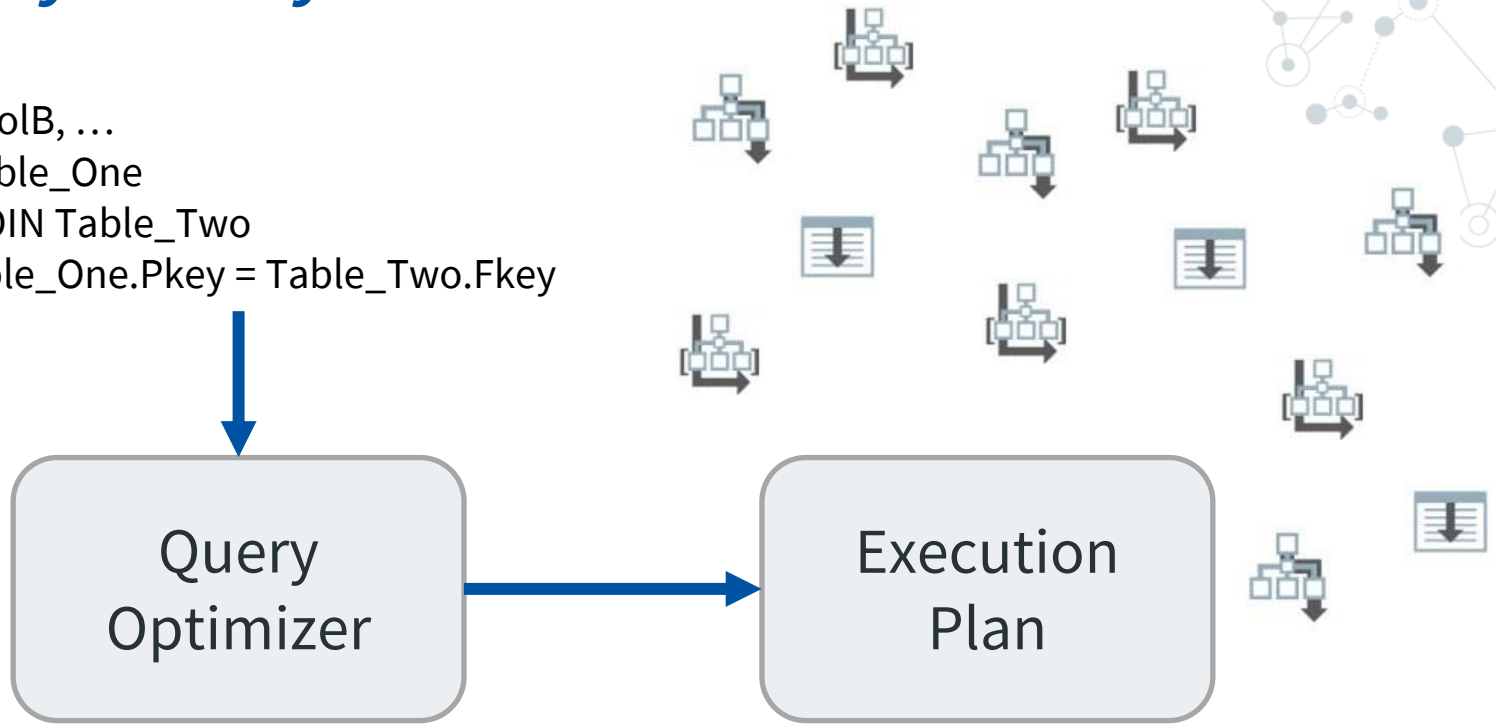
Even a General Understanding
Will Benefit You

I Love Analogies… and Pizza

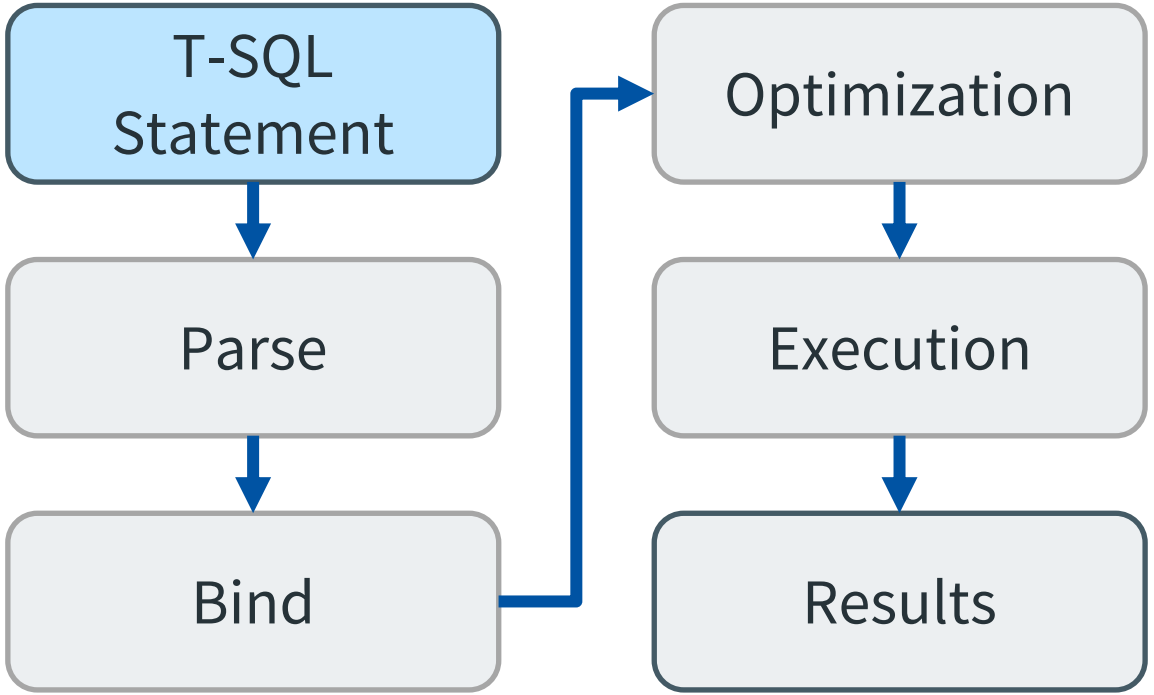
Makin Da Pizzas!

Query Lifecycle

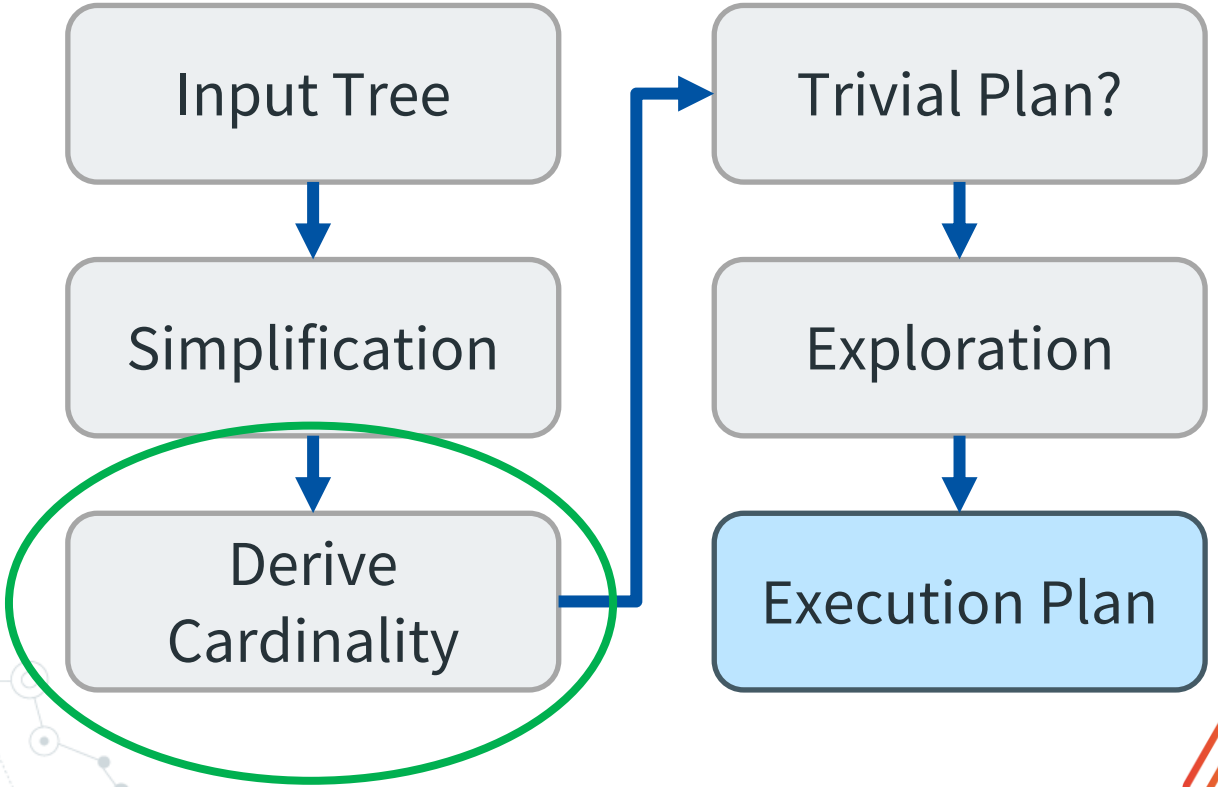
```
SELECT
   ColA, ColB, …
FROM Table_One
INNER JOIN Table_Two
   ON Table_One.Pkey = Table_Two.Fkey
```



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

Statistics Internals

Internals You Ought To Know

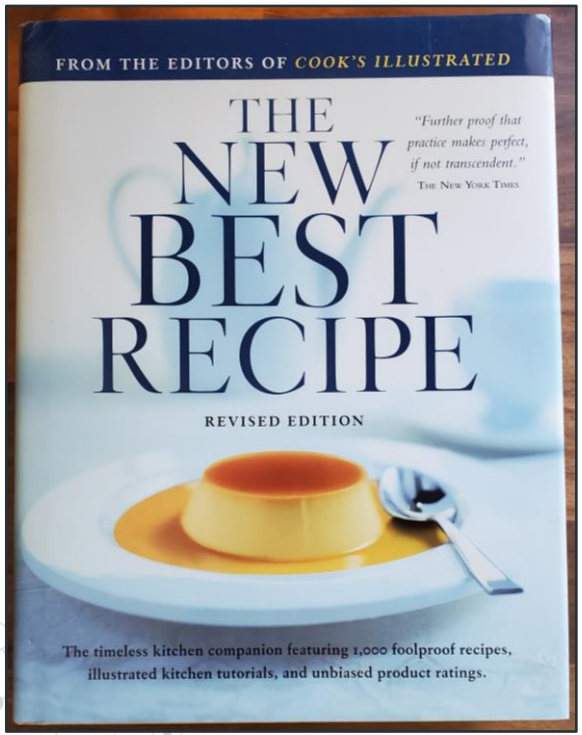
Statistics Are Stored About What?

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

How Can We Query Statistics?

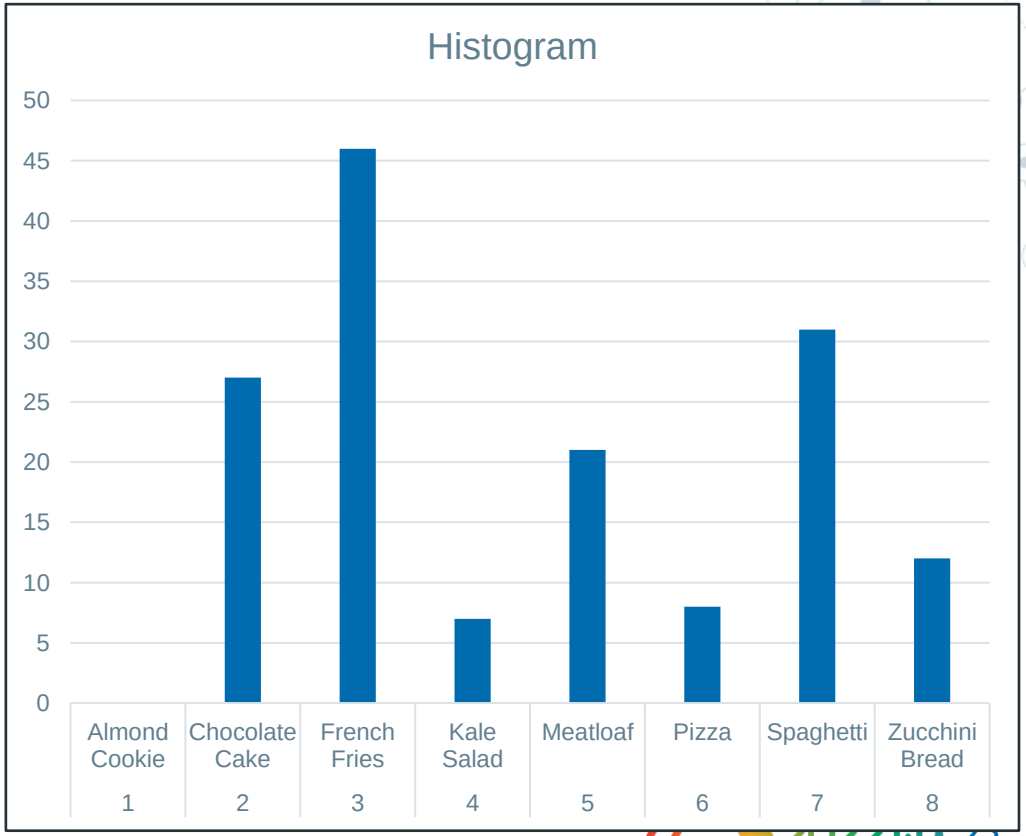
- sys.stats
- sys.stats_columns
- sys.dm_db_stats_properties()
- sys.dm_db_stats_histogram()
  - DBCC SHOW_STATISTICS()

Recipes Galore!



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	range_rows / 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

Join Types

A Slight Tangent

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

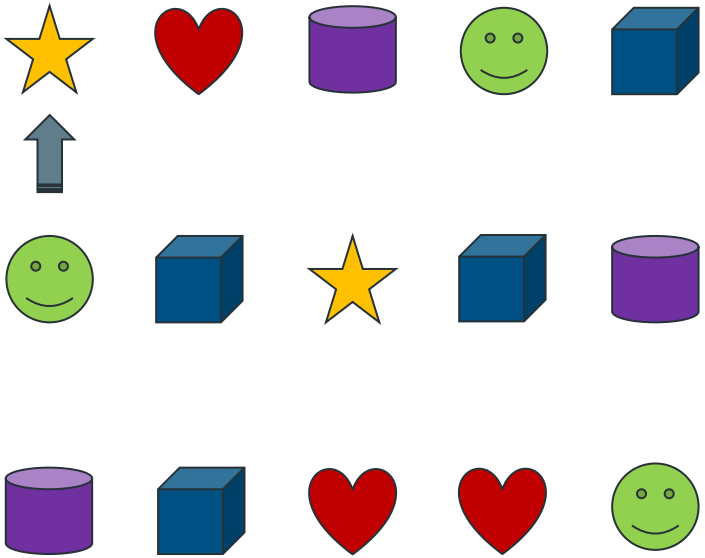
Nested Loop

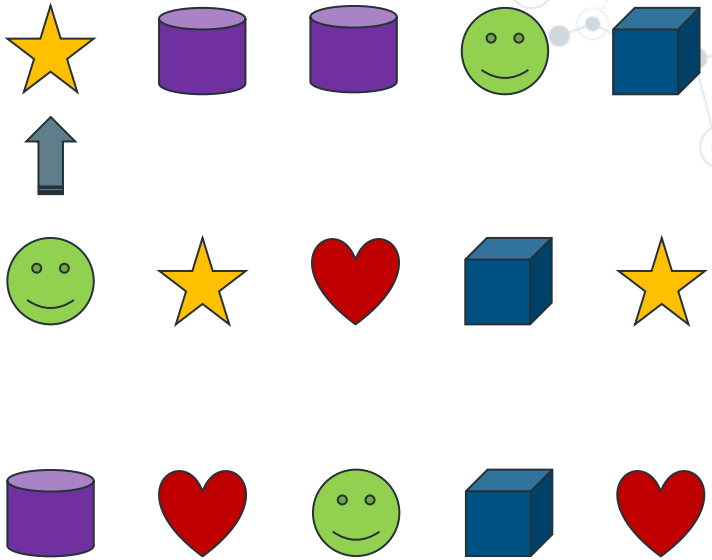
- Takes a value, compares to all others in dataset
- Preferred if one dataset is much smaller

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

Parting Thoughts

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-Y4IwY

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=XQynIdV3d8U

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=3I9mHjait2I

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)

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.PASSDataComminitySummit.com/evaluation