

# Think Like The Cardinality Estimator

Taioab Ali, Data Platform MVP

August 18, 2020

Tampa SQL User Groups



## Data Solutions Manager, GMO LLC



<http://sqlworldwide.com/>



/sqlworldwide



@sqlworldwide



Taiob@sqlworldwide.com

### Data Professional

Microsoft Data Platform MVP. 14 Years working with Microsoft Data Platform. Microsoft and MongoDB certified. Worked in ecommerce, healthcare and finance industry.

### Giving Back

Board member NESQL user group and PASS DBA virtual group. Organizer of Boston SQL Saturday. Frequent speaker at local and virtual user groups, SQL Saturdays, and azure events.

### When Not Working

Running – One 26.2 and many 13.1 miles. Learning US history. Shuttling 3 kids.

---

# Predicate, Density, Predicate Selectivity

---

## What is Cardinality

---

## Why Cardinality Matters

---

## DBCC SHOW\_STATISTICS

---

## Magic Numbers

---

# Predicate

Is an expression that evaluates to  
TRUE, FALSE, or UNKNOWN

- Join
- Filter
  - Where
  - Having

```
SELECT
    cus.CustomerID,
    COUNT(0) AS [NumOfOrders]
FROM
    sales.Orders AS ord
JOIN
    sales.Customers AS cus
ON
    ord.CustomerID=cus.CustomerID
WHERE
    ord.OrderDate='2013-01-01'
GROUP BY cus.CustomerID
HAVING COUNT(0) > 2 ;
GO
```

Join Predicate


Filter Predicate  
Where/Having

# Predicate Selectivity

$$\frac{[ \text{\# rows that pass the predicate} ]}{[ \text{total number of rows} ]}$$

```
SELECT
    COUNT(0) AS [NumOfOrders]
FROM
    sales.Orders
WHERE
    CustomerID=577;
GO
```

75 rows for  
customerID 577  
73595 total number  
of rows.  
 $75/73595=.0010$



# Density

How often duplicate values occur in a column.

$$\frac{1}{[\text{\# of distinct values in a column}]}$$

```
SELECT  
    COUNT(DISTINCT customerID) AS [DistinctCusId]  
FROM  
    sales.Orders;  
GO
```

Density  
663 distinct customerID  
 $1/663=0.00150830$



# Cardinality

## For Us

- Number of distinct values in a table column, relative to the number of rows in the table.

## Structured Query Language

- Uniqueness of data values contained in a particular column (attribute) of a database table.

## Math

- Cardinality of a set is a measure of the “number of elements of the set”.



## Clustered Index Seek (Clustered)

Scanning a particular range of rows from a clustered index.

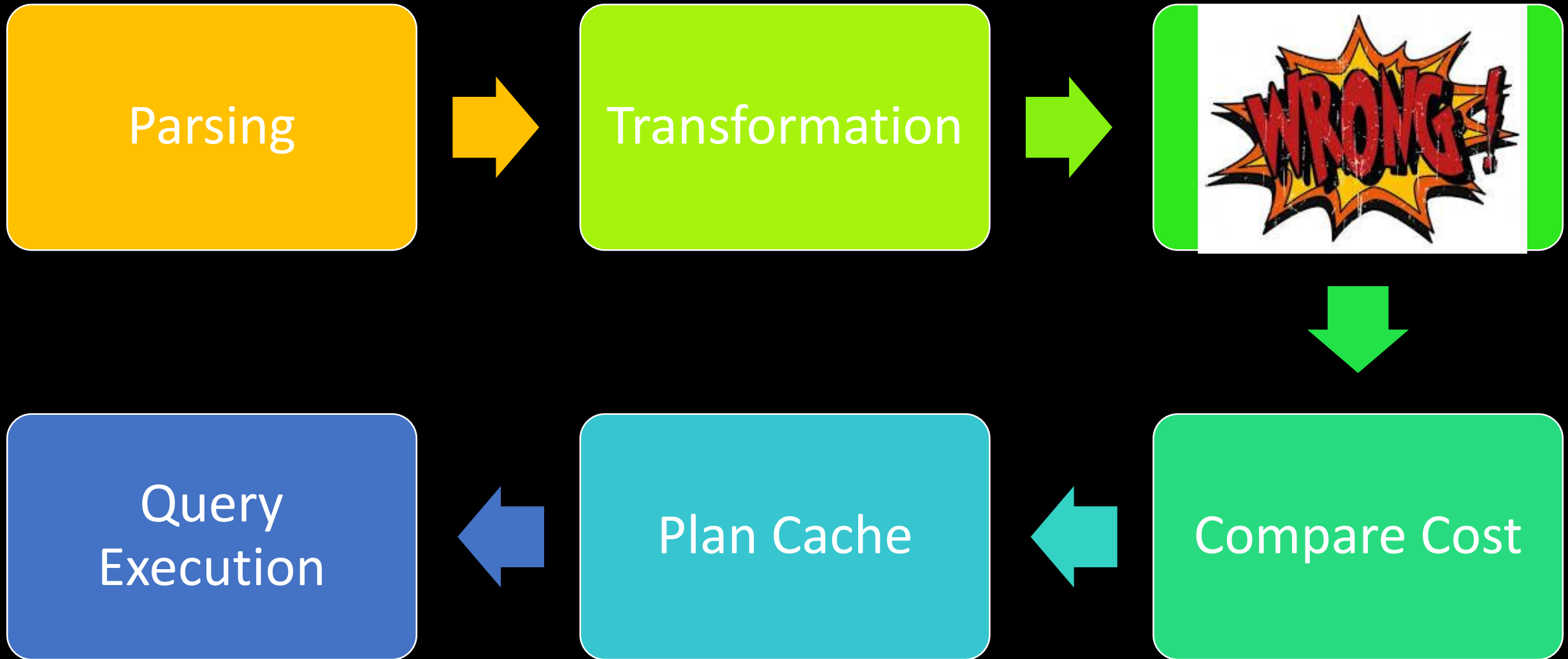
Physical Operation	Clustered Index Seek
Logical Operation	Clustered Index Seek
Actual Execution Mode	Row
Estimated Execution Mode	Row
Storage	RowStore
Actual Number of Rows	12
Actual Number of Batches	0
Estimated Operator Cost	0.0032904 (8%)
Estimated I/O Cost	0.003125
Estimated Subtree Cost	0.0032904
Estimated CPU Cost	0.0001654
Estimated Number of Executions	1
Number of Executions	1
Estimated Number of Rows	7.61577
Estimated Row Size	143 B

Actual



Estimated





## Cost

Parallel

Serial

## Memory Grant

In  
Memory

Spill to  
Disk

## Access Method

Seek

Scan

Seek +  
Scan

## Algorithm

Join

Aggregate

Sort

Missing Statistics

Stale Statistics

Inadequate Statistics Sample Rate

Maximum 200 Steps in Histogram

Parameter Sniffing

Out-of-Model Query Constructs

# DBCC SHOW\_STATISTICS

## Header

- Meta data about the statistics.

## Density Vector

- How many unique values are present within a column or columns?

## Histogram

- Frequency of data within the first key column of the statistics.

- STAT\_HEADER

Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expression	Unfiltered Rows
FK_Sales_Orders_ContactPersonID	Jun 2 2016 10:45AM	71583	71583	200	0.009388237	8	NO	NULL	71583

- DENSITY\_VECTOR

	All density	Average Length	Columns
1	0.001512859	4	ContactPersonID
2	1.39698E-05	8	ContactPersonID, OrderID

- HISTOGRAM

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
1	1001	0	128	0	1
2	1007	248	107	2	124
3	1013	231	130	2	115.5
4	1021	327	111	3	109
5	1025	101	85	1	101
6	1031	237	133	2	118.5

# STAT\_HEADER

Name	Updated	Rows	Rows Sampled	Steps
NCI_FilteredContactPersonID	Mar 31 2017 4:55PM	50299	50299	200

Deprecated

Density	Average key length	String Index	Filter Expression	Unfiltered Rows
0.00945746	8	NO	[(contactpersonid]>(2000))	73595

# DENSITY\_VECTOR

1/Number of  
distinct values in  
column

Names of columns  
in the prefix

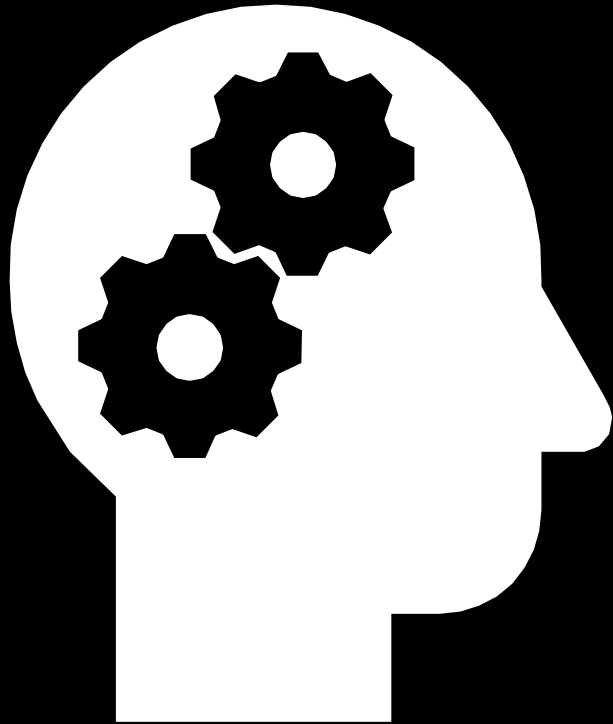
All density	Average Length	Columns
0.002164502	4	ContactPersonID
1.988111E-05	8	ContactPersonID, OrderID



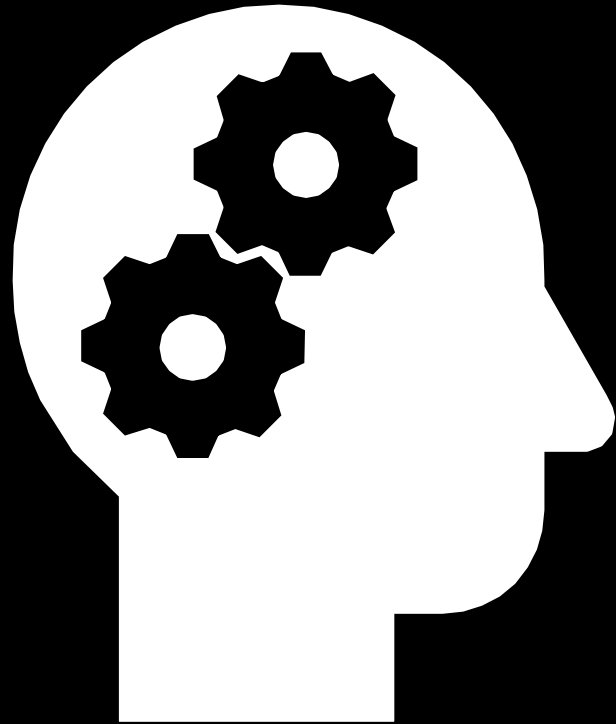
# Histogram

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
70	2035	330	123	3	110
71	2043	333	90	3	111
72	2051	317	94	3	105.6667
73	2055	106	114	1	106
74	2061	249	122	2	124.5
75	2065	107	108	1	107
76	2073	326	111	3	108.6667
77	2077	127	129	1	127
78	2083	234	119	2	117
79	2091	332	106	3	110.6667
80	2095	118	135	1	118
81	2103	338	125	3	112.6667
82	2107	112	110	1	112
83	2113	224	109	2	112
84	2121	375	118	3	125
85	2127	224	101	2	112

RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
<b>2083</b>		<b>119</b>		
<b>Between 2084 and 2090</b>	<b>332</b>		<b>3</b>	<b>110.66</b>
<b>2091</b>		<b>106</b>		



- Single Predicate
  - Histogram direct hit
  - Histogram intra step
  - Scaling
  - Distinct



- Multiple Predicates
  - Conjunction
  - Disjunction
- Parameter Sniffing
- Unknown
- Ascending Key



SQL2019 CU6  
SSMS 18.6

# Resource

- [Statistics](#)
- [Query Tuning Fundamentals](#)
- [DBCC SHOW\\_STATISTICS \(Transact-SQL\)](#)
- [13 Things You Should Know About Statistics and the Query Optimizer](#)
- [Cardinality Estimation for Multiple Predicates](#)
- [New Trace Flag to Fix Table Variable Performance](#)
- [Ascending key Issue TF 2389 and 2390](#)
- [Optimizing Query Plans with the SQL Server 2014 Cardinality Estimator](#)
- [Cardinality Estimation \(SQL Server\)](#)



@SqlWorldWide



linkedin.com/in/sqlworldwide



sqlworldwide.com



taio@sqlworldwide.com

