Think Like The Cardinality Estimator

Taiob Ali
He/Him

Taiob Ali

Data Solutions Manager, GMO LLC

http://sqlworldwide.com/

/sqlworldwide

@sqlworldwide

taiob@sqlworldwide.com

Data Professional

Microsoft Data Platform MVP. 15 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 Founder of DBA virtual group. Organizer of Boston Data Saturday. Frequent speaker at local and virtual user groups, Data Saturdays and Azure conferences.

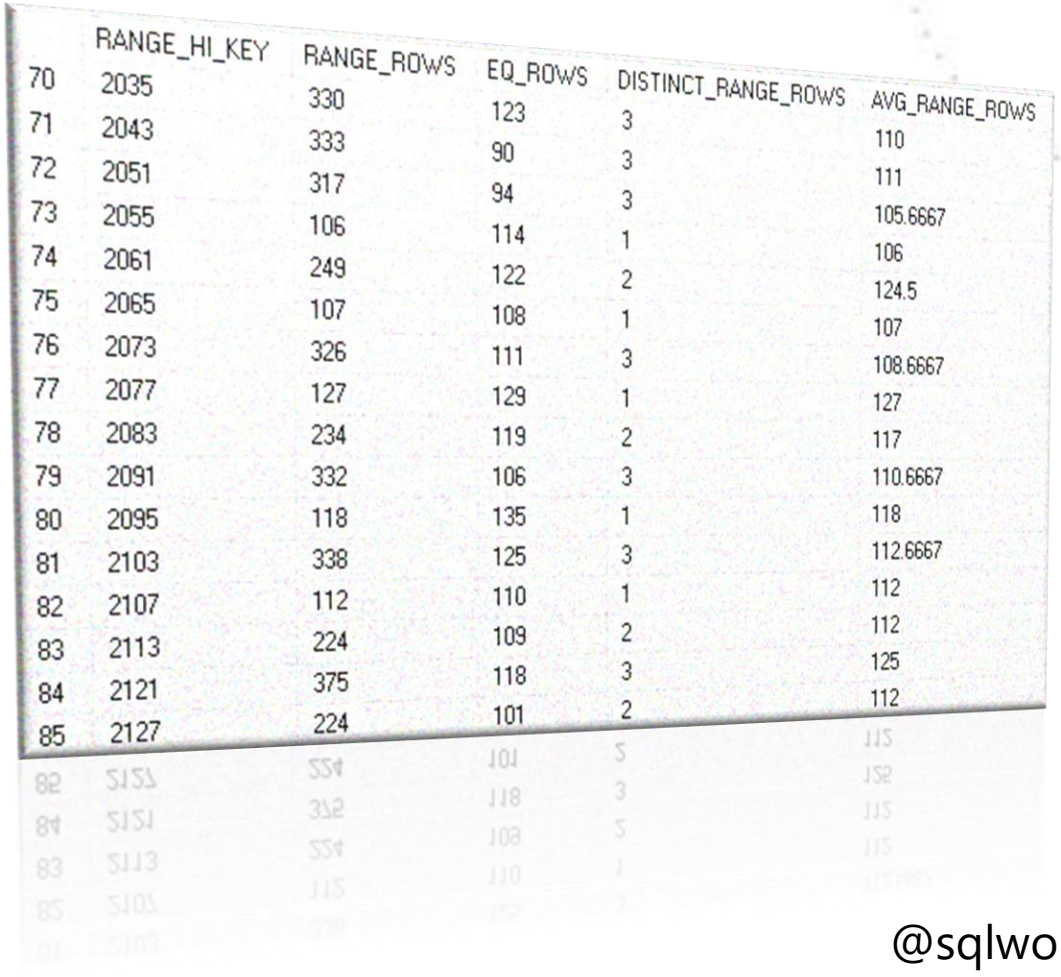
When Not Working

Running – 1x26.2 and 30+x 13.1, Learning US history. Shuttling 3 kids.

Agenda

- Definition
- What is Cardinality?
- Why Cardinality Matters?
- DBCC SHOW_STATISTICS
- Magic Numbers

Definition

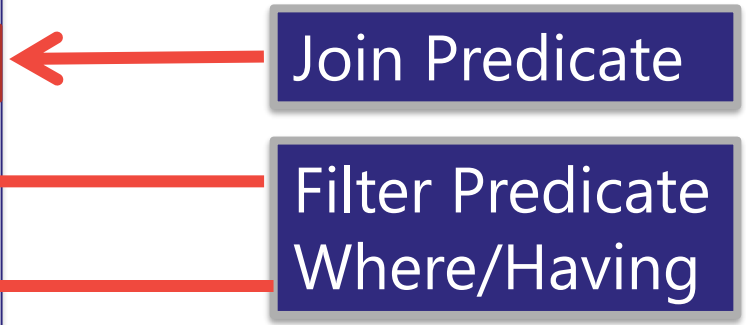


	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
70	2035	330	123	3	110
71	2043	333	90	3	111
72	2051	317	94	3	105.6667
73	2055	106	114	1	106
74	2061	249	122	2	124.5
75	2065	107	108	1	107
76	2073	326	111	3	108.6667
77	2077	127	129	1	127
78	2083	234	119	2	117
79	2091	332	106	3	110.6667
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

Predicate

- **Expression** = TRUE, FALSE, 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
```



Predicate Selectivity

Fraction of rows from the input set of the predicate that satisfy the predicate

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

Predicate Selectivity

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

75 rows for customerID (577)

73595 rows for all customer

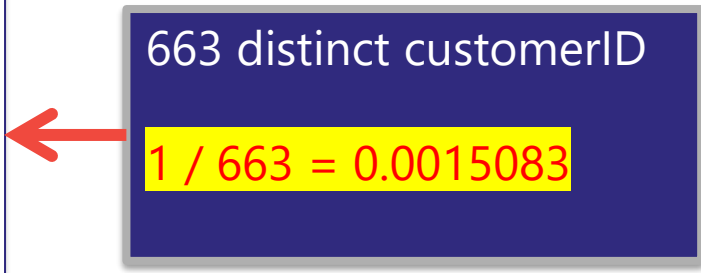
75 / 73595 = .0010

Density

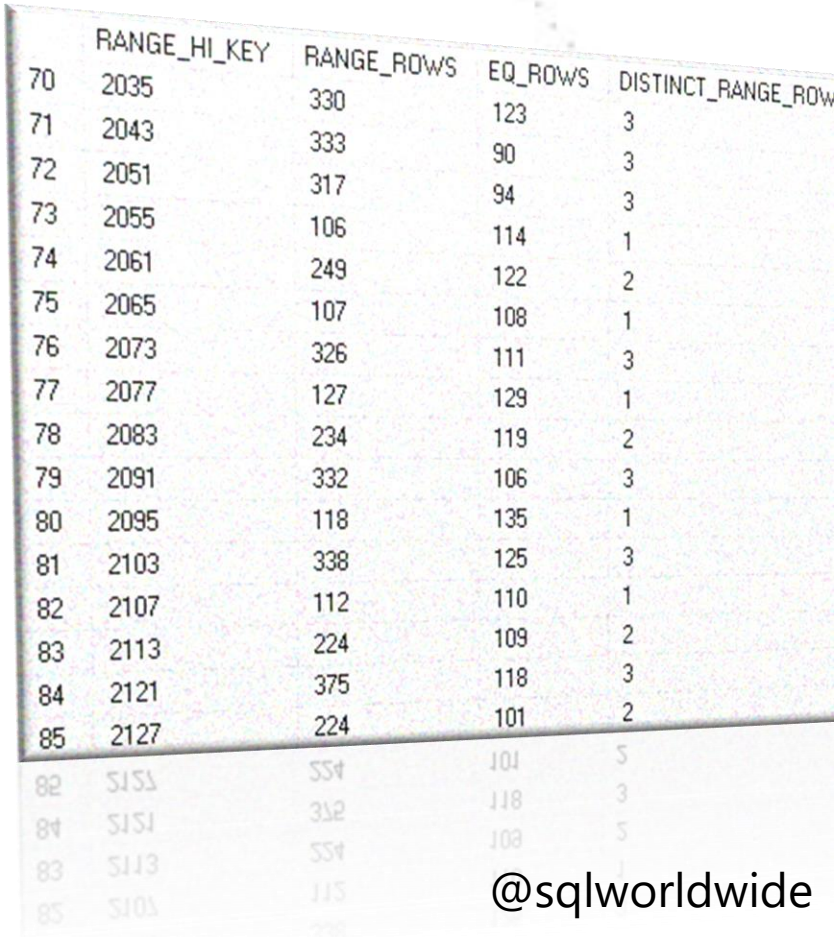
How often duplicate values occur in a column

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

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



What is Cardinality Estimation?

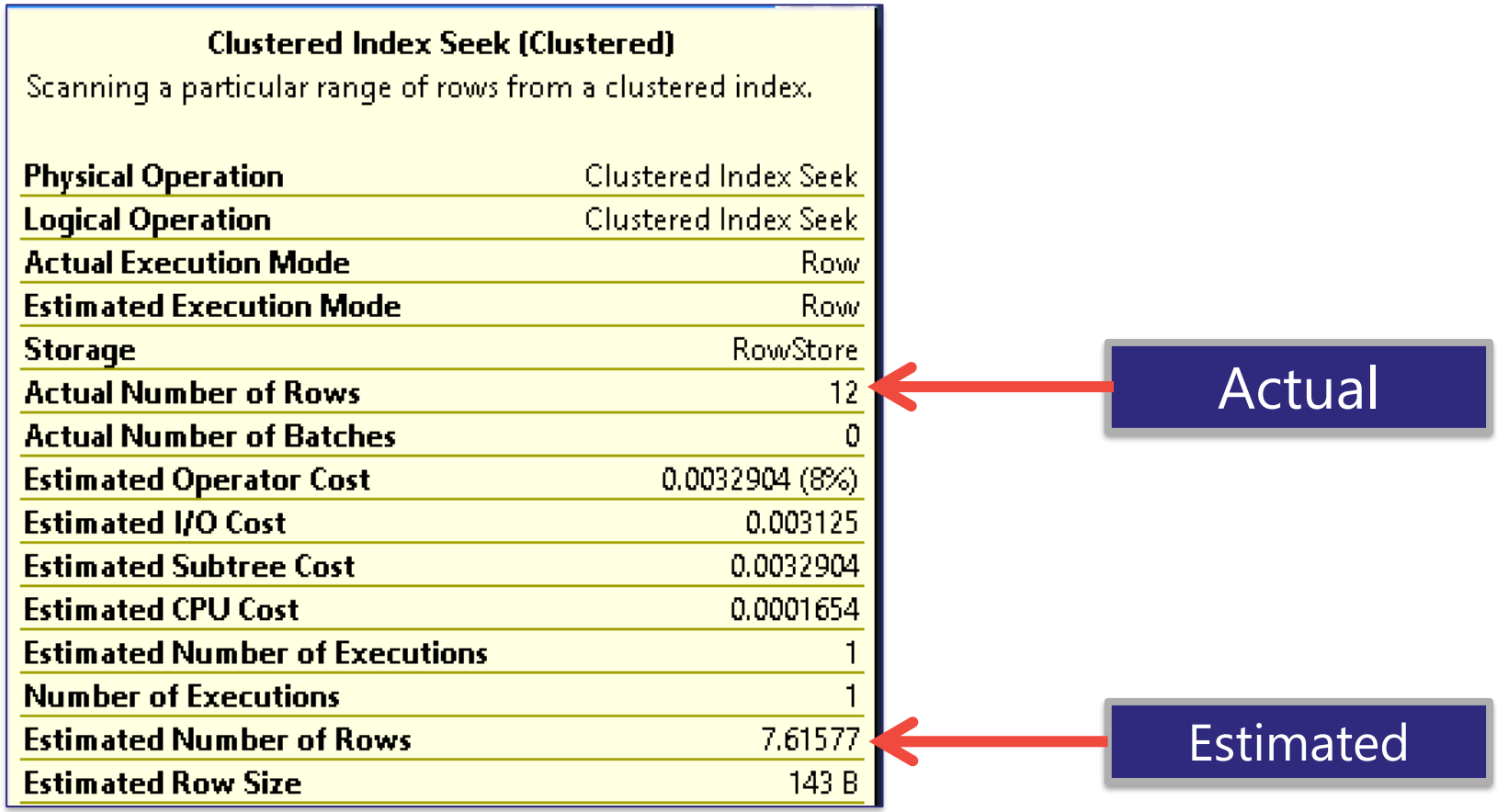


	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROW
70	2035	330	123	3
71	2043	333	90	3
72	2051	317	94	3
73	2055	106	114	1
74	2061	249	122	2
75	2065	107	108	1
76	2073	326	111	3
77	2077	127	129	1
78	2083	234	119	2
79	2091	332	106	3
80	2095	118	135	1
81	2103	338	125	3
82	2107	112	110	1
83	2113	224	109	2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

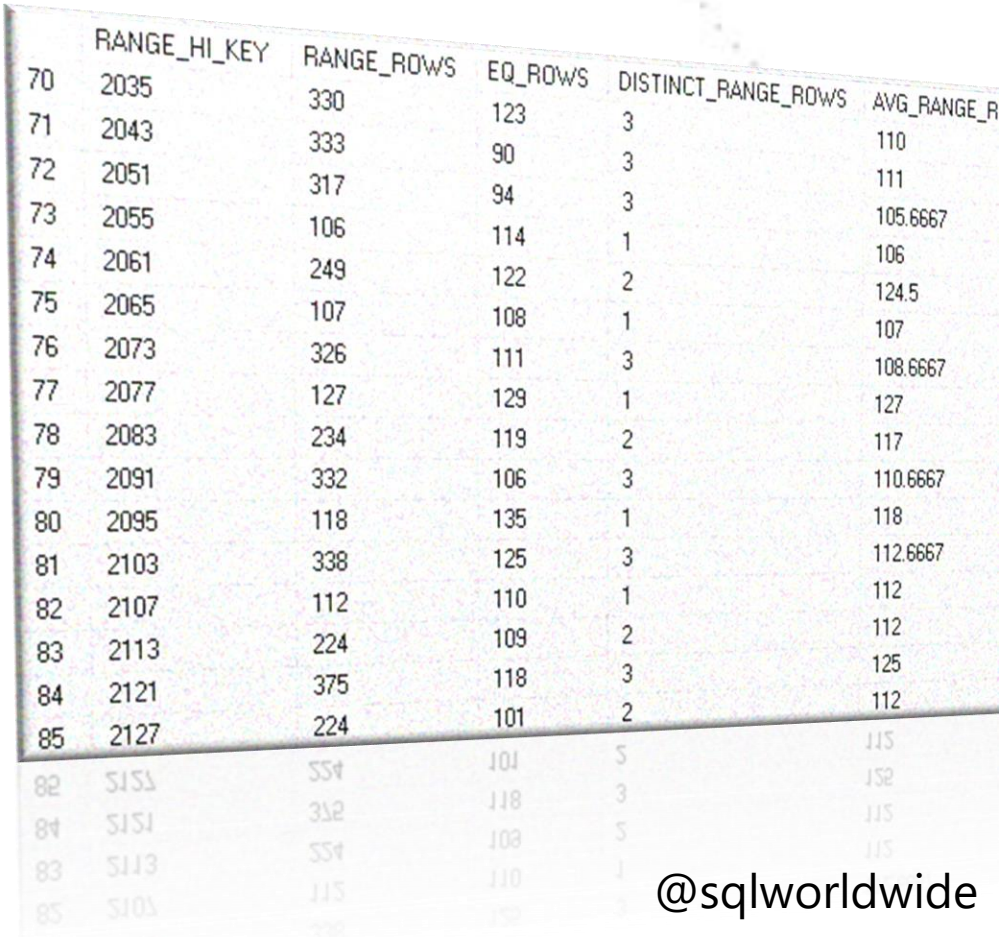
@sqlworldwide

Structured Query Language

Cardinality estimation (CE) is the process by which the Query Optimizer derives the estimated number of rows for a query plan

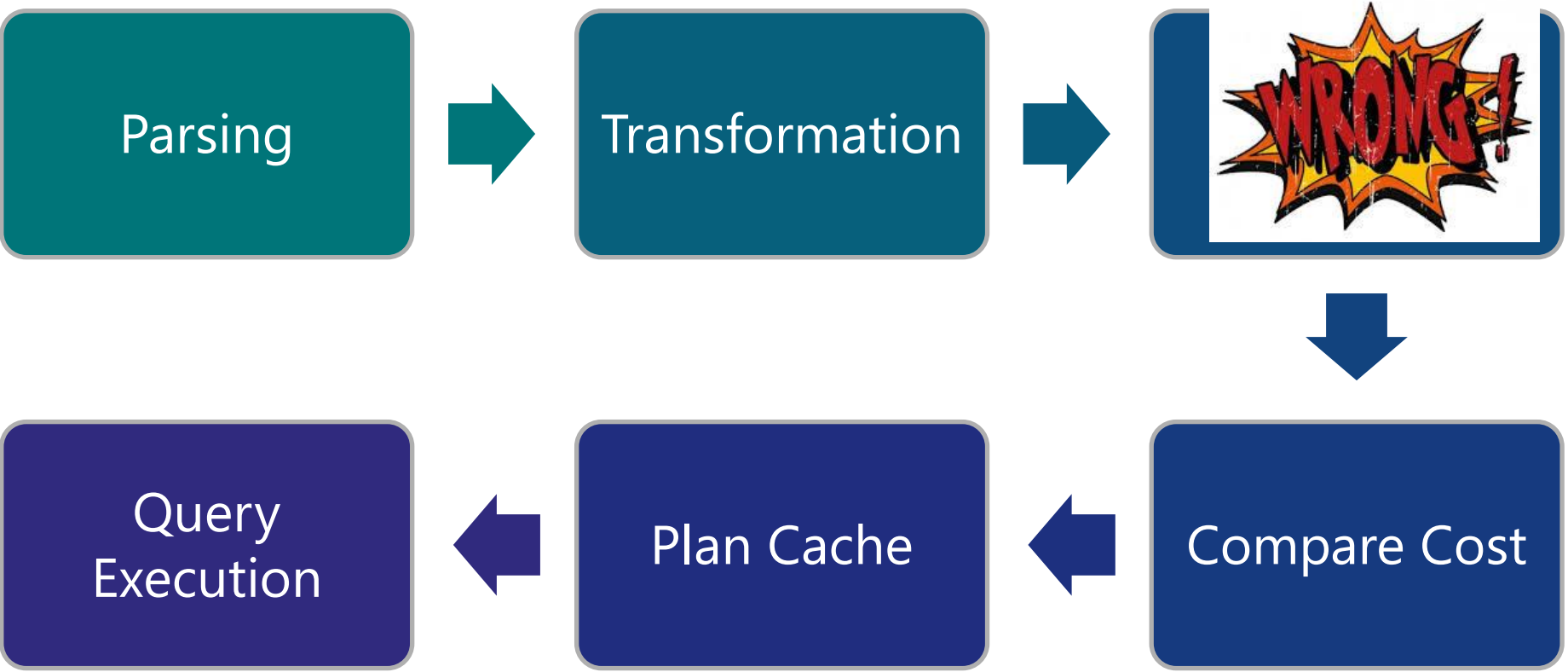

Clustered Index Seek (Clustered)
Scanning a particular range of rows from a clustered index.
Physical Operation Clustered Index Seek
Logical Operation Clustered Index Seek
Actual Execution Mode Row
Estimated Execution Mode Row
Storage RowStore
Actual Number of Rows 12
Actual Number of Batches 0
Estimated 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

Why Cardinality Matters?

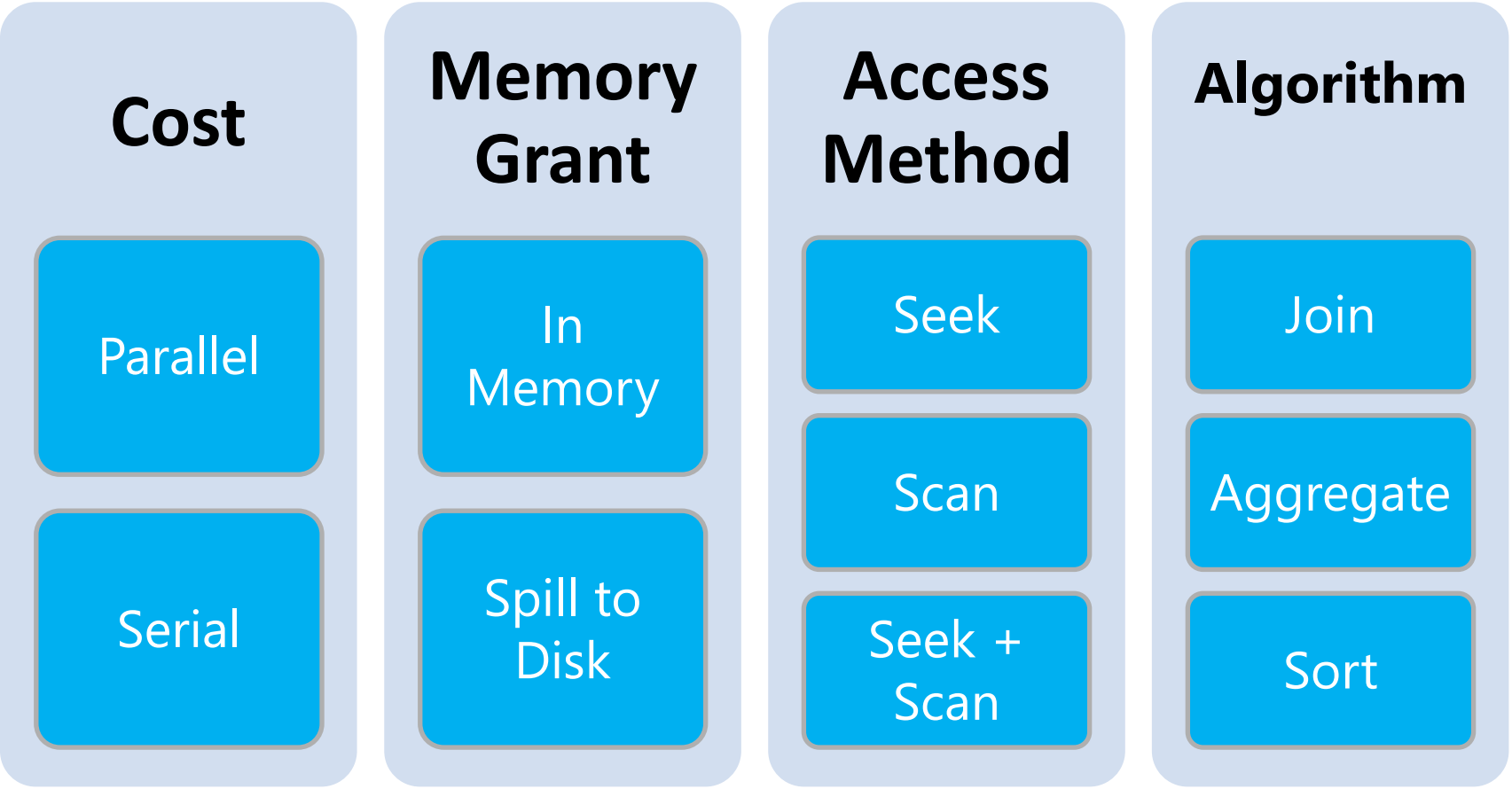


	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_R
70	2035	330	123	3	110
71	2043	333	90	3	111
72	2051	317	94	3	105.6667
73	2055	106	114	1	106
74	2061	249	122	2	124.5
75	2065	107	108	1	107
76	2073	326	111	3	108.6667
77	2077	127	129	1	127
78	2083	234	119	2	117
79	2091	332	106	3	110.6667
80	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

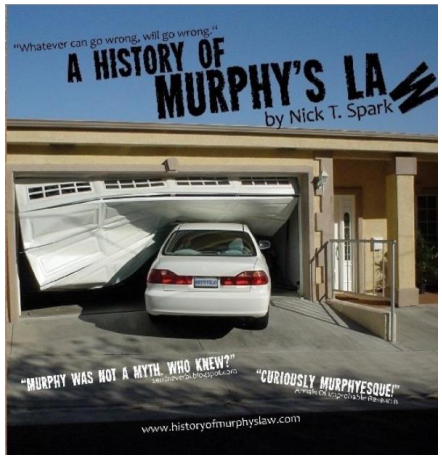
@sqlworldwide



@sqlworldwide



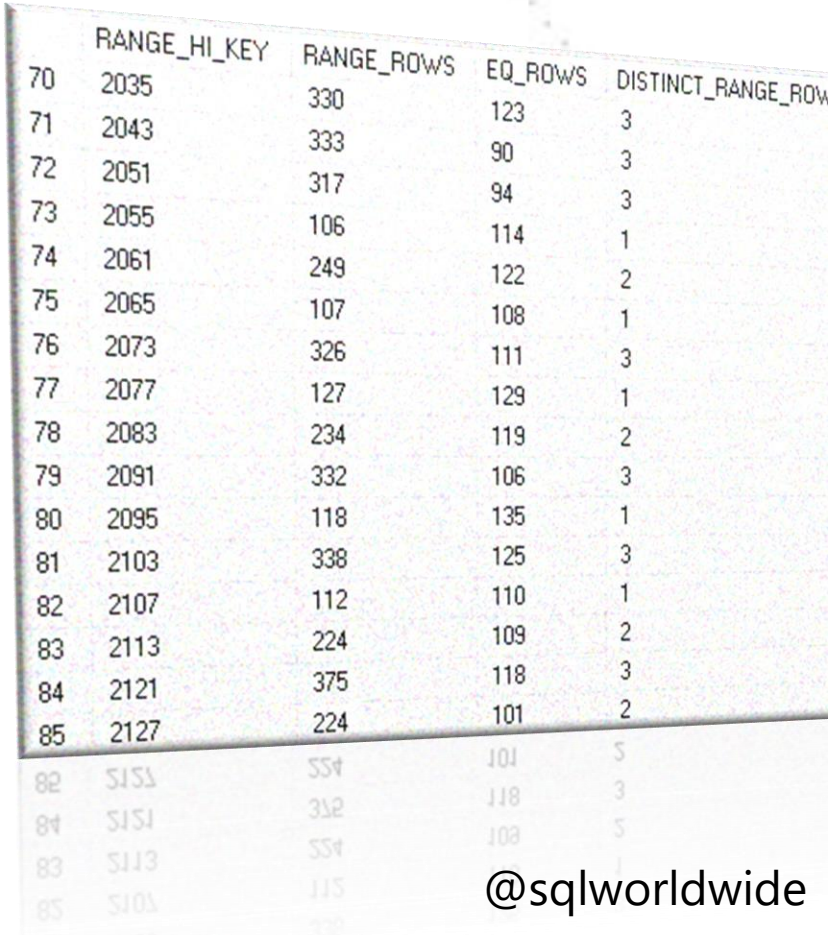
@sqlworldwide



http://www.historyofmurphyslaw.com

- Missing Statistics
- Stale Statistics
- Inadequate Statistics Sample Rate
- Maximum 200 Steps in Histogram
- Parameter Sniffing
- Out-of-Model Query Constructs

DBCC SHOW_STATISTICS



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

@sqlworldwide

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
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
2083	334	119	2	117
2091	332	106	3	110.6667

RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
2083		**119**		
Between 2084 and 2090	**332**		**3**	**110.66**
2091		**106**		

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

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

DEMO

SQL 2019 CU9
SSMS 18.8

Reference

- 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

taiob@sqlworldwide.com


THANK
YOU