

Statistics and the Query Optimization

Grant Fritchey
Product Evangelist
Red Gate Software

Goals

- ① Learn how SQL Server creates, stores and maintains statistics
- ① Understand how the optimizer consumes statistics to arrive at an execution plan
- ① Learn various methods for controlling statistics to take more direct control of your queries



grant@scarydba.com



www.scarydba.com



[@gfritchey](https://twitter.com/gfritchey)



www.linkedin.com/in/scarydba

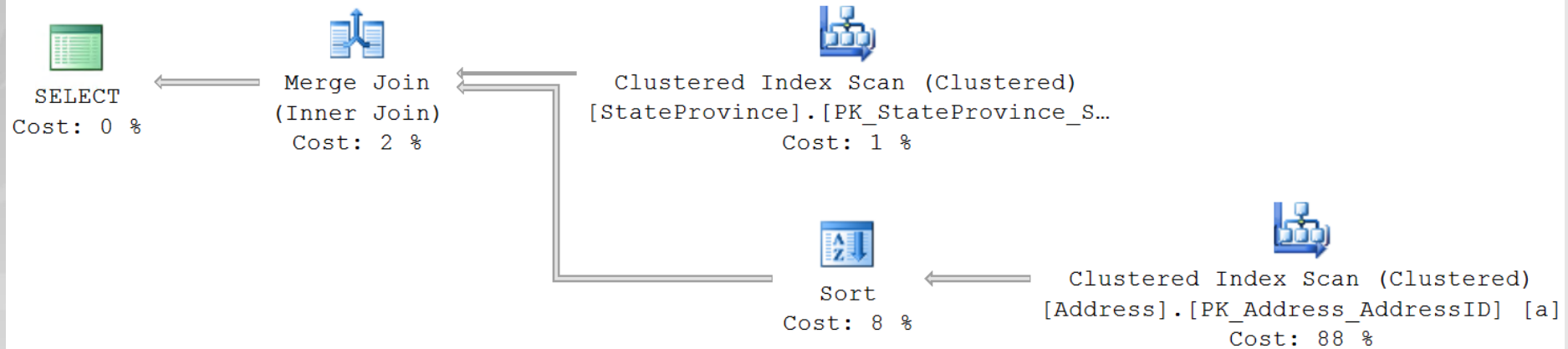
Grant Fritchey

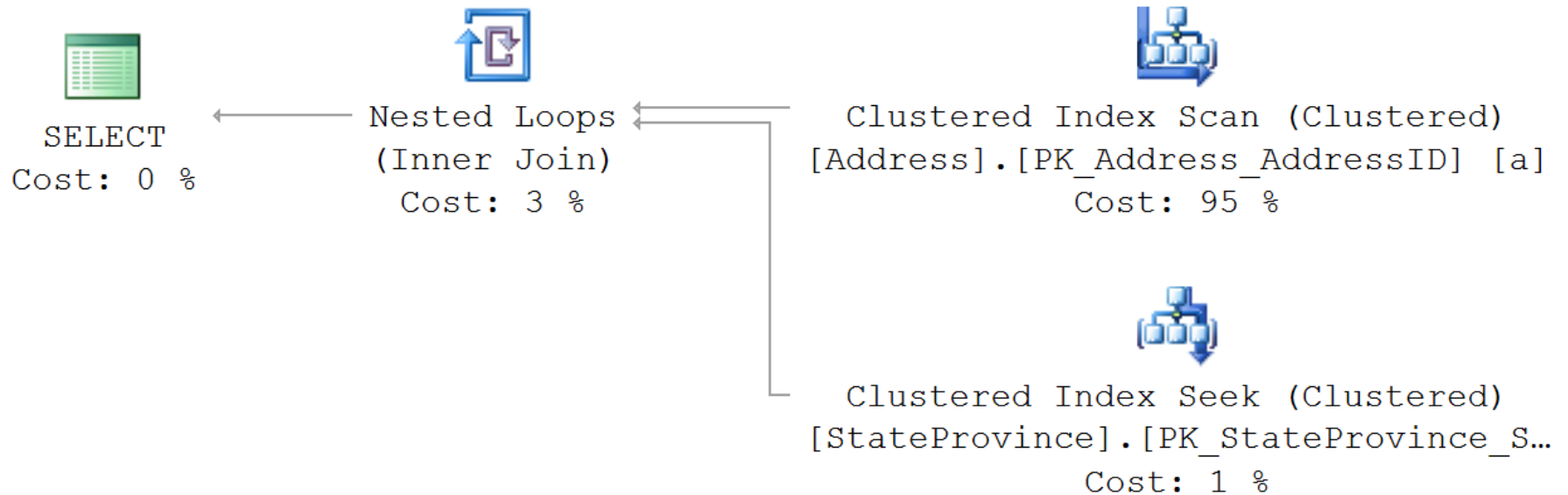
Product Evangelist, Red Gate Software



STATISTICS IN ACTION

```
CREATE PROC dbo.spAddressByCity @City NVARCHAR(30)
AS
    SELECT  a.AddressID,
            a.AddressLine1,
            a.AddressLine2,
            a.City,
            sp.[Name] AS StateProvinceName,
            a.PostalCode
    FROM    Person.Address AS a
            JOIN Person.StateProvince AS sp
    ON      a.StateProvinceID = sp.StateProvinceID
    WHERE   a.City = @City;
```





Actual Number of Rows	434
Actual Rebinds	0
Actual Rewinds	0
Defined Value	Estimated Value

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
91	Lincoln Acres	0	102	0	1
92	London	32	434	2	16
93	Long Beach	0	97	0	1

Estimated Number of Executions	1
Estimated Number of Rows	434

Actual Number of Rows	1
Actual Rebinds	0
Actual Rewinds	0
Defined Value	Estimated Value

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
100	Melton	0	105	0	1
101	Metchosin	10	96	6	1.666667
102	Metz	0	56	0	1

Estimated Number of Executions	1
Estimated Number of Rows	1.66667



WHAT ARE STATISTICS

DBCC

SHOW_STATISTICS('Person.Address',_WA_Sys_00000004_164452B1);

	Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expression	Unfiltered Rows
1	_WA_Sys_00000004_164452B1	Jan 4 2013 7:01AM	19614	19614	200	0.2336449	17.55359	YES	NULL	19614

	All density	Average Length	Columns
1	0.00173913	17.55359	City

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
1	Abingdon	0	1	0	1
2	Ballard	29	69	19	1.526316
3	Beaverton	9	213	5	1.8
					
198	Yakima	33	96	3	11
199	York	0	117	0	1
200	Zeeland	0	1	0	1

Key Statistic Terms

- ⦿ Rows Sampled
- ⦿ Steps
- ⦿ Density
- ⦿ Range_hi_key
- ⦿ Range_rows
- ⦿ Eq_rows
- ⦿ Avg_range_rows
- ⦿ Cardinality
- ⦿ Cardinality
- ⦿ Cardinality

Cardinality

- ⊙ Selectivity
 - » Returned values / Total Values
- ⊙ Histogram
- ⊙ Density
 - » $1/\text{distinct values}$
- ⊙ Compound Columns
 - » Selectivity * Selectivity
 - » Density * Density
- ⊙ + Trace Flags which we'll talk about

Estimates vs. Cardinality

- ⊙ Comparing variables
 - ⊙ Using != and NOT
 - ⊙ Predicates comparing columns within a table
 - ⊙ Functions w/o constant value
 - ⊙ Joins using arithmetic or string operations
 - ⊙ No statistics!
-
- ⊙ Skewed distribution (eh)

Statistics are used to...

- ⊙ Determine cardinality which is used to...
 - » Determine number of rows processed which is used to...
 - Determine cost of the operation in the plan which is used to...
 - Pick the operations used in the plan which is used to
 - » Return your data in an efficient manner which is used for whatever the heck the business wants



CAPTURING BEHAVIOR

Capture Mechanisms

- ⦿ Static
 - » DBCC SHOW_STATISTICS
 - » Execution Plans (sort of)
- ⦿ Dynamic
 - » Trace Events
 - » Extended Events

Auto_stats

Selected events:

Name		
auto_stats	1	✓
missing_column_statistics	1	✓
query_optimizer_estimate_card...	1	✓

auto_stats

Occurs when index and column statistics are automatically updated. This event can be generated multiple times per statistics collection when the update is asynchronous.

Event configuration options:

Global Fields (Actions)			Filter (Predicate)	Event Fields
Name	Description	Data Type		
count	Indicates the number ...	uint32		
database_id	The ID of the databas...	uint32		
☒ database_name	Name of the databas...	unicode_string		
duration	The time (in microsec...	uint64		
index_id	The ID for the index o...	uint32		
job_id	The number that repr...	int32		
job_type	Indicates the type of t...	qp_job_type		
last_error	Indicates the last erro...	int32		
object_id	The object for which ...	int32		
retries	Indicates the number ...	uint16		

Missing_column_statistics

Selected events:

Name ^		
auto_stats	1	✓
missing_column_statistics	1	✓
query_optimizer_estimate_card...	1	✓

missing_column_statistics

Occurs when a query accesses a column that does not have statistics available that could have been useful for query optimization. Use this event to identify queries that may be using sub-optimal plans. This event is only generated when the Auto Create Statistics option is turned off.

Event configuration options:

Global Fields (Actions)		Filter (Predicate)	Event Fields
	Name ^	Description	Data Type
☒	column_list	Provides the list of co...	unicode_string

Query_optimizer_estimate_cardinality

Selected events:

Name ▲	⚡	🔍
auto_stats	1	✓
missing_column_statistics	1	✓
query_optimizer_estimate_card...	1	✓

query_optimizer_estimate_cardinality ▲

Occurs when the query optimizer estimates cardinality on a relational expression.

▼

Event configuration options:

⚡ Global Fields (Actions)

🔍 Filter (Predicate)

Event Fields

Name ▲	Description	Data Type
calculator	The strategy used for ...	xml
creation_time	Time at which the qu...	filetime
input_relation	The input relation on ...	xml
query_hash	Hash of the query tha...	uint64
stats_collection	The stats collection g...	xml
stats_collection_id	ID of the stats collecti...	uint32



STATS ARE CREATED

```
SELECT  s.name,  
        s.auto_created,  
        s.user_created,  
        s.filter_definition,  
        sc.column_id,  
        c.name AS ColumnName  
FROM    sys.stats AS s  
        JOIN sys.stats_columns AS sc ON sc.stats_id = s.stats_id  
AND sc.object_id = s.object_id  
        JOIN sys.columns AS c ON c.column_id = sc.column_id  
AND c.object_id = s.object_id  
WHERE   s.object_id = OBJECT_ID('dbo.Address2')
```

	name	auto_created	user_created	filter_definition	column_id	ColumnName
1	_WA_Sys_00000001_4242D080	1	0	NULL	1	AddressID
2	_WA_Sys_00000004_4242D080	1	0	NULL	4	City
3	_WA_Sys_00000006_4242D080	1	0	NULL	6	PostalCode
4	_WA_Sys_00000009_4242D080	1	0	NULL	9	ModifiedDate

Event: auto_stats (2013-12-19 07:47:44.0279754)

Details

Field	Value
count	1
database_id	5
database_name	
duration	51453
index_id	0
job_id	0
job_type	StatsUpdate
last_error	0
object_id	1159675179
retries	0
sql_text	SELECT addressid FROM address2 WHERE addressid = 42;
statistics_list	Created: AddressID
status	Other
success	True

Event: auto_stats (2013-12-19 07:47:44.0280771)

Details

Field	Value
count	1
database_id	5
database_name	
duration	0
index_id	2
job_id	0
job_type	StatsUpdate
last_error	0
object_id	1159675179
retries	0
sql_text	SELECT addressid FROM address2 WHERE addressid = 42;
statistics_list	Loading without updating: address2._WA_Sys_00000001_451F3D2B
status	Loading stats without updating
success	True

	Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expression	Unfiltered Rows
1	_WA_Sys_00000004_4707859D	Dec 20 2013 5:14AM	19614	19614	200	0.2336449	17.55359	YES	NULL	19614

	All density	Average Length	Columns
1	0.00173913	17.55359	City

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
95	Lynnwood	2	101	1	2
96	Malabar	32	81	2	16
97	Marysville	3	93	3	1
98	Matrville	1	83	1	1
99	Melbourne	5	110	2	2.5
100	Melton	0	105	0	1
101	Metchosin	10	96	6	1.666667
102	Metz	0	56	0	1
103	Mill Valley	14	87	4	3.5
104	Milsons Point	4	80	2	2
105	Milton Keynes	0	63	0	1
106	Milwaukie	1	105	1	1
107	Morangis	46	31	10	4.6

What?

`_WA_Sys_00000001_0A1E72EE`

What?

_WA_Sys_00000001_0A1E72EE



Hexidecimal Object ID

What?

_WA_Sys_00000001_0A1E72EE

Hexidecimal Object ID

Table Column Number

What?

_WA_Sys_00000001_0A1E72EE

Hexidecimal Object ID

Table Column Number

System

What?

The US State of Washington... yes I'm serious

↓
_WA_Sys_00000001_0A1E72EE

↑
Hexidecimal Object ID

↑
Table Column Number

↑
System



YOU CAN CREATE STATS

```
CREATE STATISTICS MyStats
ON Person.Person (Suffix)
WITH FULLSCAN;
```

	name	auto_created	user_created	filter_definition	column_id	ColumnName
1	PK_Person_BusinessEntityID	0	0	NULL	1	BusinessEntityID
2	IX_Person_LastName_FirstName_MiddleName	0	0	NULL	7	LastName
3	IX_Person_LastName_FirstName_MiddleName	0	0	NULL	5	FirstName
4	IX_Person_LastName_FirstName_MiddleName	0	0	NULL	6	MiddleName
5	AK_Person_rowguid	0	0	NULL	12	rowguid
6	_WA_Sys_00000008_693CA210	1	0	NULL	8	Suffix
7	MyStats	0	1	NULL	8	Suffix

```

CREATE STATISTICS MyJrStats
ON Person.Person (Suffix)
WHERE Suffix = 'Jr.'
WITH FULLSCAN;

```

	name	auto_created	user_created	filter_definition	column_id	ColumnName
1	PK_Person_BusinessEntityID	0	0	NULL	1	BusinessEntityID
2	IX_Person_LastName_FirstName_MiddleName	0	0	NULL	7	LastName
3	IX_Person_LastName_FirstName_MiddleName	0	0	NULL	5	FirstName
4	IX_Person_LastName_FirstName_MiddleName	0	0	NULL	6	MiddleName
5	AK_Person_rowguid	0	0	NULL	12	rowguid
6	_WA_Sys_00000008_693CA210	1	0	NULL	8	Suffix
7	MyStats	0	1	NULL	8	Suffix
8	MyJrStats	0	1	([Suffix]='Jr.')	8	Suffix


```
CREATE STATISTICS MyComboStats
ON Person.Person (Title,Suffix)
WHERE Suffix = 'PhD'
WITH FULLSCAN;
```

	Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expression	Unfiltered Rows
1	MyComboStats	Jan 14 2014 10:41AM	2	2	1	0	12	YES	[(Suffix)='PhD']	19972

	All density	Average Length	Columns
1	1	6	Title
2	1	12	Title, Suffix

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
1	Ms.	0	2	0	1

...Or Drop Them

```
DROP STATISTICS Person.Person.MyStats;  
DROP STATISTICS Person.Person.MyJrStats;  
DROP STATISTICS Person.Person.MyComboStats;
```



STATS ARE MAINTAINED

Dungeons and Dragons

- ⦿ Add 1 Row when 0
- ⦿ Add > 500 Rows when < 500
- ⦿ Add 20% + 500 Rows when > 500

```
CREATE INDEX AddressCity  
ON dbo.Address2 (City);
```

```
DBCC SHOW_STATISTICS(Address2,AddressCity)
```

	Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expression	Unfiltered Rows
1	AddressCity	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
	All density	Average Length		Columns						
	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS					

```
SELECT * FROM dbo.Address2 AS a
WHERE City = 'Springfield';
```

	Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expression	Unfiltered Rows
1	AddressCity	Dec 20 2013 5:18AM	1	1	1	0	22	YES	NULL	1

	All density	Average Length	Columns
1	1	22	City

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
1	Springfield	0	1	0	1

	AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	SpatialLocation	rowguid
1	501	1313 Mockingbird Lane		Springfield	30	02134	NULL	D7842560-EB19-4DC9-959E-2467A6E
2	1	1313 Mockingbird Lane		Springfield	30	02134	NULL	D3D2DF5E-D6BD-423C-AAE5-8DAFD

<	III							
---	-----	--	--	--	--	--	--	--

	Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expression	Unfiltered Rows
1	AddressCity	Dec 20 2013 5:41AM	501	501	199	1	27.54491	YES	NULL	501

	All density	Average Length	Columns
1	0.002	27.54491	City

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
1	100SpringField	0	1	0	1
2	103SpringField	2	1	2	1
3	105SpringField	1	1	1	1
4	107SpringField	1	1	1	1
5	109SpringField	1	1	1	1
6	110SpringField	1	1	1	1
7	112SpringField	1	1	1	1
8	114SpringField	1	1	1	1

Advanced D&D

- ⦿ Trace Flag 2371
 - » SQL Sever 2008 R2 Sp1 and above
 - » After 25,000 rows
 - Percentage changed based on number of rows
 - Reducing as the number grows
- ⦿ Trace Flag 4137
 - » Minimum selectivity instead of multiplication on AND predicates
- ⦿ Trace Flag 9471 (SQL Server 2014)
 - » Minimum selectivity on AND and OR predicates
- ⦿ Trace Flag 9472 (SQL Server 2014)
 - » Independence (pre-2014 behavior)



**~~YOU CAN~~ SHOULD
MAINTAIN STATS
MANUALLY**

Automatic Maintenance

- ⦿ AUTO_CREATE_STATISTICS
- ⦿ AUTO_UPDATE_STATISTICS
 - » Uses rules in previous section
- ⦿ AUTO_UPDATE_STATISTICS_ASYNC
 - » Test, test, test

sp_updatestats

```
ALTER procedure [sys].[sp_updatestats]
@resample char(8)='NO'
As
...
if ((@ind_rowmodctr <> 0) or ((@is_ver_current is not
null) and (@is_ver_current = 0)))
...
```

sp_updatestats

```
ALTER procedure [sys].[sp_updatestats]  
@resample char(8)='NO'
```

```
As
```

```
...
```

```
if ((@ind_rowmodctr <> 0)  
or ((@is_ver_current is not null) and (@is_ver_current =  
0)))
```

```
...
```

UPDATE STATISTICS

UPDATE STATISTICS Person.Address;

UPDATE STATISTICS

```
UPDATE STATISTICS Person.Address  
WITH FULLSCAN;
```

UPDATE STATISTICS

```
UPDATE STATISTICS dbo.Address2  
AddressCity;
```

UPDATE STATISTICS

UPDATE STATISTICS

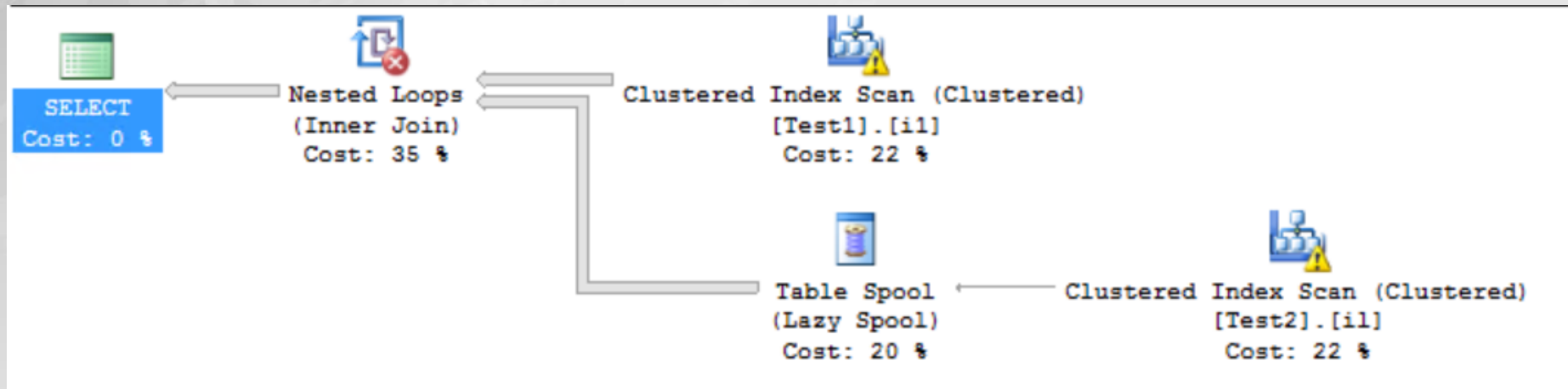
dbo.Address2 AddressCity

WITH FULLSCAN, NORECOMPUTE;



STATISTICS AND OPTIMIZER AT WORK

Statistics Matter



Compatibility Levels

```
ALTER DATABASE  
AdventureWorks2012 SET  
COMPATIBILITY_LEVEL = 120;
```

Optimizer Switches

```
OPTION (QUERYTRACEON 2312);
```

```
DBCC TRACEON(4199);
```

```
DBCC FREEPROCCACHE();
```

Recommendations

- ⦿ Compatibility setting
- ⦿ Automatic creation
- ⦿ Automatic update
- ⦿ Update asynchronous where necessary
- ⦿ Use appropriate sample rate

Goals

- ① Learn how SQL Server creates, stores and maintains statistics
- ① Understand how the optimizer consumes statistics to arrive at an execution plan
- ① Learn various methods for controlling statistics to take more direct control of your queries

Resources

- ⦿ Scarydba.com/resources
- ⦿ [Understanding SQL Server Cardinality Estimations](#)
- ⦿ [Fixing Cardinality Estimation Errors](#)
- ⦿ [Statistics Used by the Optimizer](#)
- ⦿ [First look at the
query_optimizer_estimate_cardinality XE Event](#)
- ⦿ [Changes to automatic update statistics inSQL
Server – traceflag 2371](#)
- ⦿ [Cardinality Estimation for Multiple Predicates](#)



grant@scarydba.com



www.scarydba.com



[@gfritchey](https://twitter.com/gfritchey)



www.linkedin.com/in/scarydba

Grant Fritchey

Product Evangelist, Red Gate Software

Questions?

How would you...?

What happens when...?

Why does...?

When do I...?