Statistics

Like Chennai SQL Server User Group
Blog: http://sql-articles.com/category/cssug/
MailDL : cssug@googlegroups.com
FB : www.facebook.com/groups/cssug/

Ramkumar Gopal

Living For SQL Server

Blog:

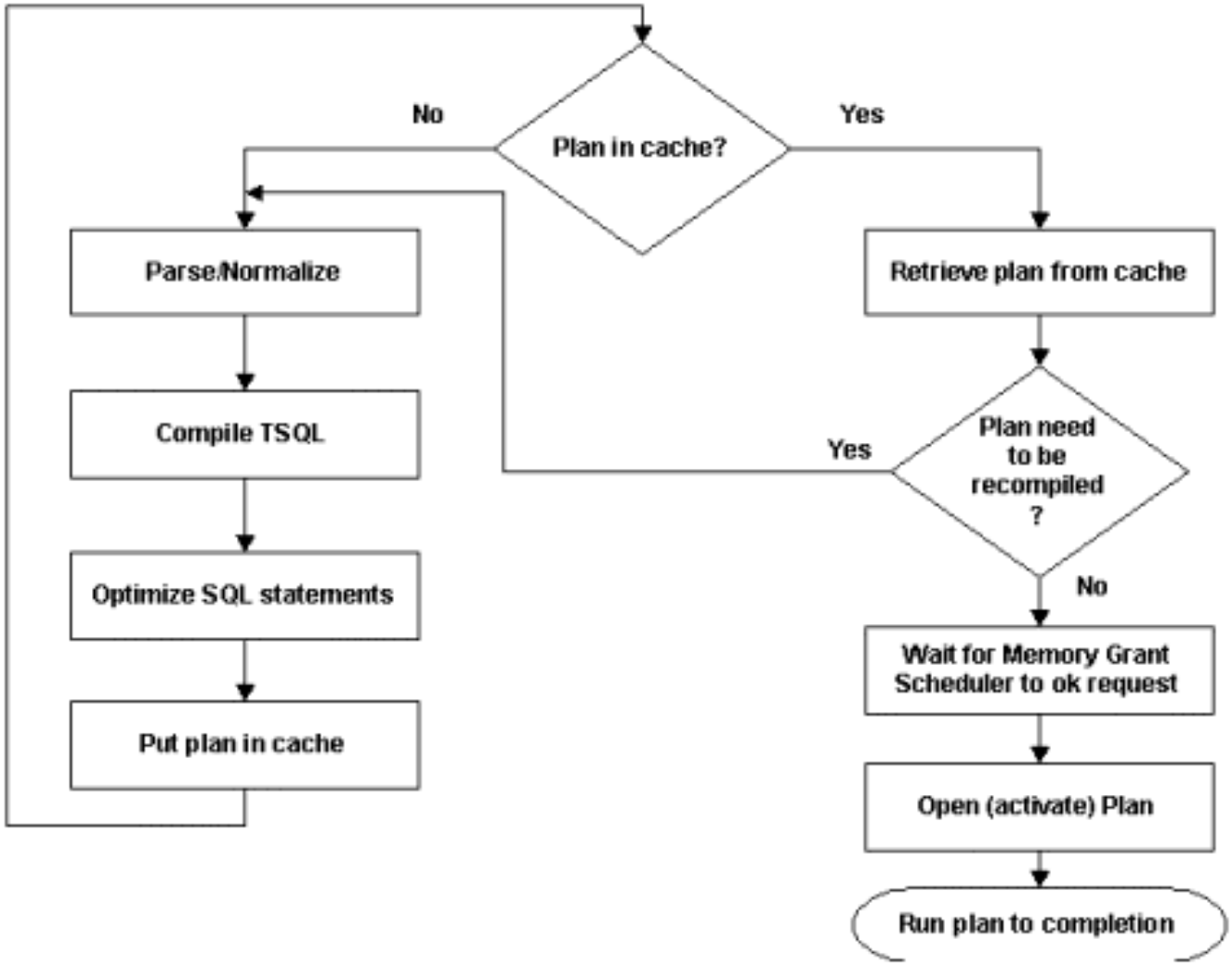
http://www.sqlservercentral.com/blogs/livingforsqlserver/

Like http://www.facebook.com/LivingForSqlServer

Agenda

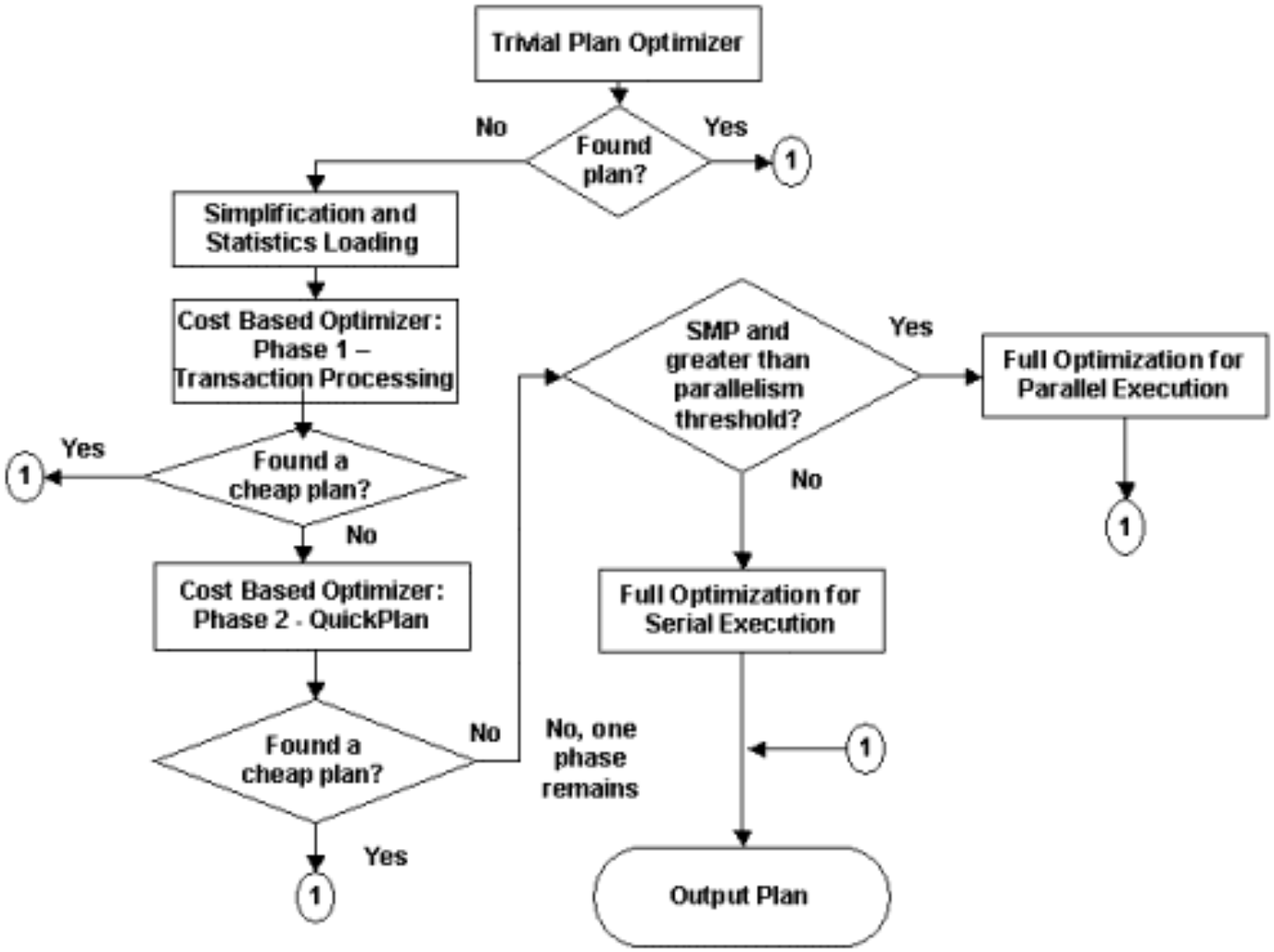
- Query Processor Stages
- Query Optimization Stages
- Statistics – Sample
- Statistics related commands

Query Processor Stages



http://msdn.microsoft.com/en-us/library/aa226174(v=sql.70).aspx

Query Optimizer Stages

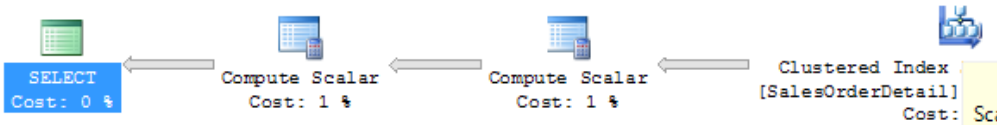


Execution Plan - Cost

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT * FROM [AdventureWorks2008].[Sales].[SalesOrderDetail]
```

Results | Messages | Execution plan

```
Query 1: Query cost (relative to the batch): 100%
SELECT * FROM [AdventureWorks2008].[Sales].[SalesOrderDetail]
```



Clustered Index Scan (Clustered)

Scanning a clustered index, entirely or only a range.

Physical Operation	Clustered Index Scan
Logical Operation	Clustered Index Scan
Actual Number of Rows	121317
Estimated I/O Cost	0.916458
Estimated CPU Cost	0.133606
Number of Executions	1
Estimated Number of Executions	1
Estimated Operator Cost	1.05006 (98%)
Estimated Subtree Cost	1.05006
Estimated Number of Rows	121317
Estimated Row Size	95 B
Actual Rebinds	0
Actual Rewinds	0
Ordered	False
Node ID	2

Object

[AdventureWorks2008].[Sales].[SalesOrderDetail].
[PK_SalesOrderDetail_SalesOrderID_SalesOrderDetailID]

Execution Plan – Cost (Zoom)

```
red Index Scan (Clustered)
derDetail].[PK_SalesOrderDe…
     Cost:
```

Clustered Index Scan (Clustered)

Scanning a clustered index, entirely or only a range.

Physical Operation	Clustered Index Scan
Logical Operation	Clustered Index Scan
Actual Number of Rows	121317
Estimated I/O Cost	0.916458
Estimated CPU Cost	0.133606
Number of Executions	1
Estimated Number of Executions	1
Estimated Operator Cost	1.05006 (98%)
Estimated Subtree Cost	1.05006
Estimated Number of Rows	121317
Estimated Row Size	95 B
Actual Rebinds	0
Actual Rewinds	0
Ordered	False
Node ID	2

Object

[AdventureWorks2008].[Sales].[SalesOrderDetail].
[PK_SalesOrderDetail_SalesOrderID_SalesOrderDetailID]

Statistics – Sample Output

SQLQuery1.sql - (...ramkumar-g (52))*

```
DBCC SHOW_STATISTICS ('sales.SalesOrderDetail','PK_SalesOrderDetail_SalesOrderID_SalesOrderDetailID')
```

Results | Messages

	Name	Updated	Rows	Rows Sampled	Steps	Density	Average key length	String Index	Filter Expre
1	PK_SalesOrde...	Mar 29 2012 1:52PM	121317	121317	150	0.2694389	8	NO	NULL

	All density	Average Length	Columns
1	3.178134E-05	4	SalesOrderID
2	8.242868E-06	8	SalesOrderID, SalesOrderDetailID

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
1	43659	0	12	0	1
2	43692	282	28	32	8.8125
3	43898	716	28	205	3.492683
4	44079	403	27	180	2.238889
5	44288	766	34	208	3.682692
6	44488	570	27	199	2.864322
7	44538	614	35	49	12.53061
8	44570	370	32	31	11.93548
9	44758	409	27	187	2.187166
10	44801	504	26	42	12
11	45047	306	26	245	1.24898

Query executed successfully. | (local) (10.50 RTM) | HCLISD\ramkumar-g (52) | AdventureWorks2008 | 00:00:00 | 1 ro

Statistics - Sample

VALUE RETURNED	DESCRIPTION
Updated	The date and time the index statistics were last updated.
Rows	The total number of rows in the table.
Rows Sampled	The number of rows sampled for index statistics information.
Steps	The number of distribution steps.
Density	The selectivity of the first index column prefix.
Average key length	The average length of the first index column prefix.
All density	The selectivity of a set of index column prefixes.
Average length	The average length of a set of index column prefixes.
Columns	The names of index column prefixes for which All density and Average length are displayed.
RANGE_HI_KEY	The upper bound value of a histogram step.
RANGE_ROWS	The number of rows from the sample that fall within a histogram step, not counting the upper bound.
EQ_ROWS	The number of rows from the sample that are equal in value to the upper bound of the histogram step.
DISTINCT_RANGE_ROWS	The number of distinct values within a histogram step, not counting the upper bound.
AVG_RANGE_ROWS	The average number of duplicate values within a histogram step, not counting the upper bound (where RANGE_ROWS / DISTINCT_RANGE_ROWS for DISTINCT_RANGE_ROWS > 0).

http://www.quest.com/whitepapers/tuning_article_1_final.pd

Statistics

Useful Indexes (look at WHERE, JOIN, GROUP BY, ORDER BY, covering)

Statistics = density information about a table kept by SQL

Help SQL "estimate" how many rows a certain query will return

Stored as a BLOB, but you can query the data using "DBCC SHOW_STATISTICS"

Courtesy: http://blogs.technet.com/b/josebda/archive/2009/03/18/sql-server-statistics.aspx

Statistics

Statistics cannot be right for every case, or else
it would be actual data

By definition, it is not all the data, it's a sample.

How frequently do you update?
Depends on how the data is changing.

Generating statistics is usually not so expensive, it's done with sampling

The query optimizer will actually create statistics on the fly, as queries come in.

Courtesy:
http://blogs.technet.com/b/josebda/archive/2009/03/18/sql-server-statistics.aspx

Statistics – Update types

```
ALTER DATABASE dbName SET AUTO_UPDATE_STATISTICS ON

ALTER DATABASE dbName SET AUTO_UPDATE_STATISTICS_ASYNC
ON
```

Sync:
- Query must wait until the statistics update is complete before compiling

Async:
- Handled by a background process
- The query will not get the benefit of the statistics update, however future queries will.

Statistics – When Statistics are updated?

>> colmodctr

determine stale statistics:

If the table contained more than 500 rows,
statistics is considered as to be stale if at least 20% of the column's data and additional 500 changes have been observed.

Tables with less or equal than 500 rows
a statistic is stale if at least 500 modifications have been made.

Every time the table changes its row count from 0 to a number greater than zero, all of a table's statistics are outdated.

For temporary tables, an additional check is been made. After every 6 modifications on a temporary table, statistics on temporary tables are also invalidated.

Table variable

There is no statistics for Table Variables

Courtesy:

SQL Server Statistics By Holger Schmeling

Statistics – Maximum Statistics per Table

SQL 7, 2000

Max Clustered Index: 1
Max Non Clustered Index: 249
Statistics on Non Indexed Columns: ?

SQL 2005

Max Clustered Index: 1
Max Non Clustered Index: 249
Statistics on Non Indexed Columns: 2000

SQL 2008/2012

Max Clustered Index: 1
Max Non Clustered Index: 999
Statistics on Non Indexed Columns: 30000

Filtered Statistics

```
CREATE STATISTICS ContactPromotion1
ON Person.Person (BusinessEntityID, LastName,
EmailPromotion)
WHERE EmailPromotion = 2 WITH SAMPLE 50 PERCENT
```

- In Large Tables having millions of rows

http://blogs.msdn.com/b/psssql/archive/2010/09/28/case-of-using-filtered-statistics.aspx

Issues with Statistics

There is no statistics object at all

- if Auto Create/Update Stats if OFF

- Bad Execution Plan

The database is read only

- The optimizer can't add missing statistics

- Create statistics before changing db to readonly

Providing expressions in predicates

Wrong: select c1,c2 from T0 where sqrt(c1) = 100

Right: rewrite queries. where c1 = 10000

Courtesy:

SQL Server Statistics By Holger Schmeling

Issues with Statistics

Insufficient sample size

Create/Update Statistics Manually (With FULLSCAN)
Note:
Be careful with Rebuild Index.
This will overwrite FULLSCAN WITH SAMPLE

No automatic generation of multi-column statistics

statistics are always single-column statistics
You have to add multi-column statistics manually

Courtesy:

SQL Server Statistics By Holger Schmeling

Courtesy and Reference

http://msdn.microsoft.com/en-us/library/aa226174(v=sql.70).aspx

http://www.quest.com/whitepapers/tuning_article_1_final.pdf

http://blogs.technet.com/b/josebda/archive/2009/03/18/sql-server-statistics.aspx

http://blogs.technet.com/b/josebda/archive/2009/03/18/sql-server-statistics.aspx

Redgate ebook SQL Server Statistics By Holger Schmeling