**SQL Server statistics and the role it plays**

* What is the SQL query optimizer?
* What is the function of the query optimizer?
* What are SQL Server Statistics?
* How does it help up in performance?
* How do you create statistics?
* How do you view them?
* Best practice for statistics

A query is a request for information from a database; this can be either a simple request or a complex request such as using joins

Since database structures are complex and the need to access the data can be achieved in different ways, (such as a table scan or use of index) the processing time to get that data may vary and SQL may need ‘help’ to retrieve that data. **Statistics** in SQL Server refers specifically to information that the server collects about the distribution of data in columns and indexes

This is where the **query optimizer** comes in. Note, the query optimizer cannot be accessed directly by users, but the query optimizer attempts to determine the most efficient way to execute a given query by considering the possible [query plans](https://en.wikipedia.org/wiki/Query_plan), such as using a table scan or the index, uses statistics to “find the best path” in its query optimization and factors such as number of records, density of pages, histogram, or available indexes all help the SQL Server optimizer in “guessing” the most efficient way to retrieve data

Thus query optimization typically tries to approximate the optimum processing time it will take to retrieve the data request

**So how are statistics created?**

Statistics can be automatically created when you create an index. If the database setting auto create stats is on, then SQL Server will automatically create statistics for non-indexed columns that are used in queries

See images

* Or they can be manually created by TSQL or GUI

See image

USE [AdventureWorks2014]

GO

CREATE NONCLUSTERED INDEX [NonClusteredIndex-20160203-192425]

ON [dbo].[People2]

([Fname] ASC,[Lname] ASC)

DBCC SHOW\_STATISTICS ('people2','NonClusteredIndex-20160203-192425')

WITH HISTOGRAM

**How are statistics updated?**

The default settings in SQL Server are to auto create and auto update statistics.  
  
***Auto Update Statistics*** basically means, if there is an incoming query but statistics are stale, SQL Server will update statistics first before it generates an execution plan.

***Auto Update Statistics Asynchronously*** on the other hand means, if there is an incoming query but statistics are stale, SQL Server uses the stale statistics to generate the execution plan, then updates the statistics afterwards.

What configuration settings should we set?

Automatic Statistics

By default, SQL Server databases automatically create and update statistics. The information that gets stored includes:

* The number of rows and pages occupied by a table's data
* The time that statistics were last updated
* The average length of keys in a column
* Histograms showing the distribution of data in a column