SQL SERVER Performance Tuning Notes

What are Statistics and what do they do?

Notes only-

1.   Statistics essentially "describe" the data what is there in the index.

2.   Statistics assigns buckets for an Index. We can have up to 200 buckets in an index. Buckets will define range of values.

   E.g. Bucket 1 will have values from 1 to 50.
   Bucket 2 will have values from 51 to 100 and so on.

   How many values roughly in each bucket, the SQL Server query optimizer can make some assumption depending on the values the query is expecting. Now we use these values for decision making like whether we can use Index scan or seek or anything else.

3.   SQL Server Query Optimizer uses statistics to estimate the distribution of values.

4.   Statistics helps query optimizer in making better decision making.

5.   In SQL Server we have Statistics on column and Index. We can create column statistics on a single column or multiple columns.

E.g. Use DBCC command to check statistics

```
USE AdventureWorks2012
GO


DBCC
SHOW_STATISTICS('[Sales].[SalesOrderHeader]','IX_SalesOrderHeader_CustomerID')
```

Output of the above query is given below-
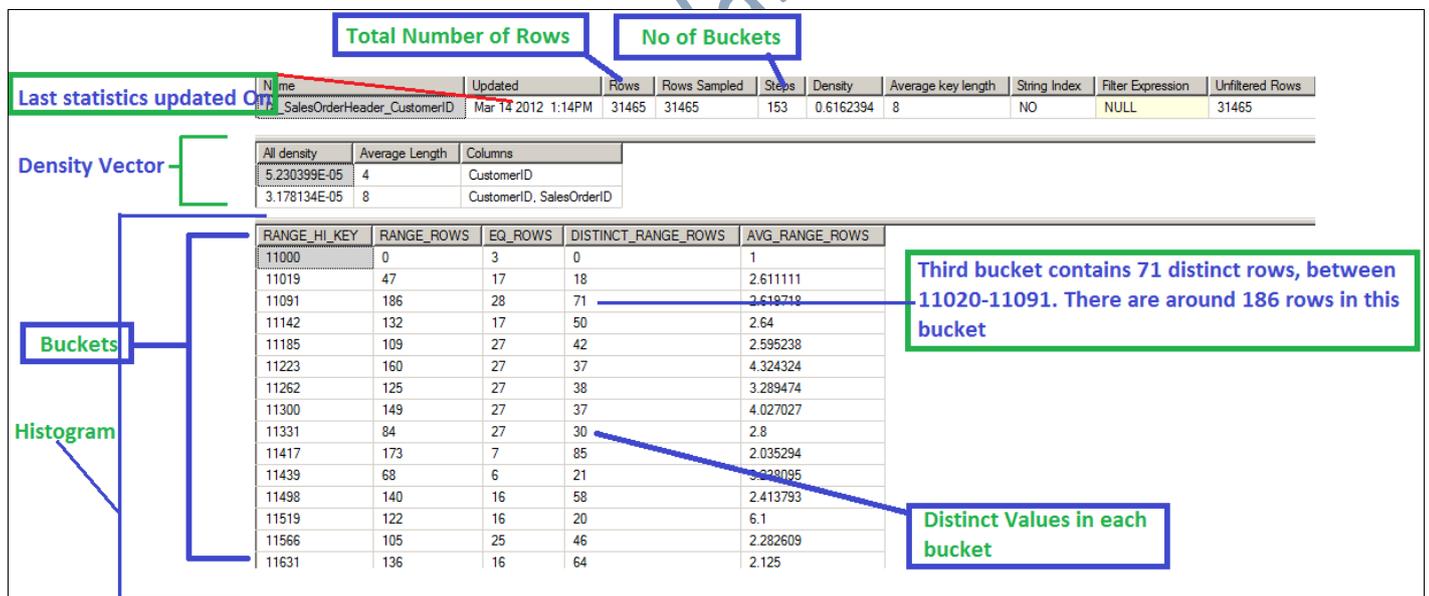
6.   In the above example we can see Density Vector and Histogram. Histogram contains distribution of values that's where all the buckets are. Here we have range rows, equality rows, distinct rows, and average rows.

7. Index stats are automatically created when an index is created.
   a. Clustered Index
   b. Non Clustered Index, with included columns & Filtered also.
   c. Spatial Index
   d. Memory optimized table index (In Memory OLTP)

8. Index stats are **NOT** automatically created on below indexes.
   a. Column store index
   b. XML index

9. Columns stats (This column is not part of an index key) are created when we use that column in our query statement. It can be part of the select, group by or order by. Means when we use that column in out column then only column stats will be created.

10. Where do you find the statistics in DB? You can find this information in your DB. Check out the screen shot below from object explorer.
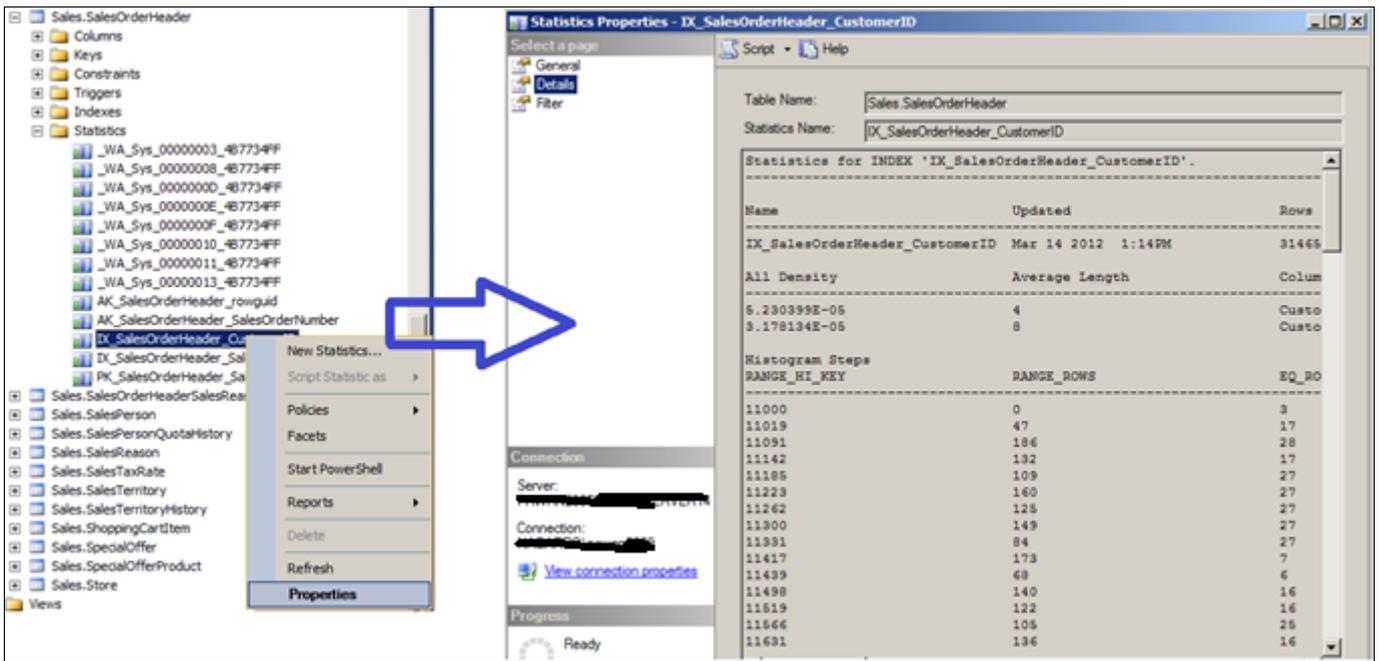
11. Use below command to see more information about statistics

```sql
SELECT * FROM sys.stats

SELECT * FROM sys.dm_db_stats_properties(34,17)
```

12. Internally all these stats information is stored in sys.sysidxstats. We cannot query it.

13. When to Update Statistics
    a. By default in SQL Server we have Auto Updates statistics on.
    b. Update statistics will happen based on the number of rows changed in an index.

Total Number of rows in a table or in an index	Minimum # of changes required for updating statistics
Insertion of first row	1
Rows < 500	500
Rows > 500	500 rows or more + 20% no of rows

14. We can update statistics synchronously and asynchronously. Default is synchronously. Here when SQL Server tries to execute a query before executing it, SQL Server will check the threshold value against the # of changes happened in the table and if the threshold has crossed than SQL Server first update the statistics and then the query will be executed. So depending on the table size and the other processes running on the server update statistics may take some time and slow down you query.

15. In asynchronous case the query will be executed first and then the update statistics will happen. In this case the benefit is that our query will not have to wait for update statistics to happen. The disadvantage of this case is that our query may get executed with less qualified execution plan as the statistics are not up to date. Most of the cases people don't ON this feature.

16. Rebuild Indexes will update statistics on indexes.

17. On larger tables we should create a schedule to update stats.

18. Now if you want to update statistics on all the tables and on all the indexes then use sp_updatestats. Well it is not recommended though.