

Виды оптимизации:

- Оптимизация запросов (rule-based optimization)
- Оптимизация плана исполнения (cost-based optimization)

- Статистика по таблице
 - Число строк
 - Число блоков
 - Средняя длина строки
- Статистика по столбцу
 - Число различных значений
 - Число null-значений
 - Гистограмма распределения данных

- Статистика по индексу
 - Число листьев
 - Число уровней
 - Фактор кластеризации
- Системная статистика
 - Производительность и использование ввода-вывода
 - Производительность и использование CPU

DBCC SHOW_STATISTICS("SalesOrderDetail", NCI_SalesOrderDetail_ProductID)
WITH HISTOGRAM;

	RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RANGE_ROWS	AVG_RANGE_ROWS
1	707	0	30830	0	1
2	708	0	30070	0	1
3	710	1880	440	1	1880
4	711	0	30900	0	1
5	712	0	33820	0	1
6	713	0	4290	0	1
7	714	0	12180	0	1
8	715	0	16350	0	1
9	716	0	10760	0	1
10	717	0	2180	0	1
11	718	0	2190	0	1
12	722	440	3920	1	440
13	725	520	3740	1	520
14	726	0	2880	0	1
15	729	480	3750	1	480
16	730	0	2880	0	1

- RANGE_HI_KEY – верхняя граница интервала (шага гистограммы), в котором считается количество значений
- RANGE_ROWS – ожидаемое количество строк, имеющих значение столбца внутри интервала
- DISTINCT_RANGE_ROWS – ожидаемое количество различных значений внутри интервала
- EQ_ROWS – This represents the estimated number of rows whose column value equals the upper bound of the histogram step.
- AVG_RANGE_ROWS ($\text{RANGE_ROWS} / \text{DISTINCT_RANGE_ROWS}$, если $\text{DISTINCT_RANGE_ROWS} > 0$) – ожидаемое среднее число дубликатов

Варианты обновления статистики:

- Автоматическая
- Вручную
- Сбор на ходу (во время выполнения запроса)
 - Сбор полной статистики
 - Сбор статистики, основанный на сэмплировании

Когда автоматический сбор статистики может не справляться:

- Часто изменяемые таблицы, значения в которых постоянно удаляются или переписываются в течение дня
- Таблицы, в которые массово загружаются данные объёмом более 10% от уже содержащихся

Практическое использование статистики:

1. В статье [Statistics in MS SQL Server](#) показаны примеры создания и использования статистики с кодом запросов. Для работы необходимо скачать и восстановить из .bak файла базу данных [Adventure Works 2014](#).
2. В статье Importance of Statistics and How It Works in SQL Server ([часть 1](#), [часть 2](#)) используются данные из базы данных Adventure Works 2008R2, которая уже не доступна для скачивания с официального сайта тестовых баз данных для MS SQL Server. Тем не менее, в статье приводятся результаты работы запросов и данные из планов их исполнения, а так же исходный код запросов, который можно изменить для использования более новой версии этой базы, ссылка на которую есть в пункте 1.

Использованные ресурсы:

1. <https://www.simple-talk.com/sql/performance/sql-server-statistics-basics/>
2. <https://dba.stackexchange.com/questions/808/what-are-database-statistics-and-how-can-i-benefit-from-them>
3. https://docs.oracle.com/cd/A87860_01/doc/server.817/a76992/stats.htm
4. https://docs.oracle.com/cd/B19306_01/server.102/b14211/stats.htm#i41506
5. <http://www.databasejournal.com/features/mssql/importance-of-statistics-and-how-it-works-in-sql-server-part-1.html>
6. <http://www.databasejournal.com/features/mssql/importance-of-statistics-and-how-it-works-in-sql-server-part-2.html>
7. <http://www.oracle.com/technetwork/database/bi-datawarehousing/twp-optimizer-stats-concepts-110711-1354477.pdf>