

Using the Column Statistics and Histograms tab (Oracle and SQL Server)

Histograms are special statistics that exist for a limited number of columns and are created by the database administrator. Column histograms should be created only when there are highly- skewed values in a column, such as is the case of an order details table with an Order Status column where the number of closed orders for a business operating for several years is far greater than the number of open orders. The Order Status column therefore meets the criteria of a useful target for a histogram because the data is highly skewed. Using histograms the optimizer determines that a full-scan is recommended when searching for closed orders, but an index scan is more useful when searching for open orders.

Oracle-specific Column Statistics and Histograms tab example

DB Optimizer looks at the columns that have histograms and using statistics tries to determine whether the column is a good or bad candidate for a histogram and presents this information on the Column Statistics and Histograms tab.

Object		Table Name	Column Name	Histogram	# Buckets	Filter Type	Indexed	Median Value Deviation	# Distinct Value
<input type="checkbox"/>	<input type="checkbox"/>	PS_PAY_CALENDAR	PAY_CONFIRM_RUN	NONE	1	Literal Equality	Yes		
<input type="checkbox"/>	<input type="checkbox"/>	PS_RETROPAY_RQST	RETROPAY_SEQ_NO	NONE	1	Join Equality	Yes		1
<input type="checkbox"/>	<input type="checkbox"/>	PS_RETROPAYFORM TEL	OFF_CYCLE	NONE	1	Join Equality	No		
<input type="checkbox"/>	<input type="checkbox"/>	PS_PAY_CALENDAR	PAYGROUP	NONE	1	Join Equality	Yes		
<input type="checkbox"/>	<input type="checkbox"/>	PS_RETROPAYFORM TEL	RETROPAY_SEQ_NO	NONE	1	Join Equality	Yes		

The row shading indicates the following:

- **Green:** Good histogram candidate
- **Red:** Bad histogram candidate
- **No shading:** Not determined to be a good or bad histogram candidate

Median value deviation

For columns that have histograms, the median value deviation is presented. Understanding the median value deviation can help you determine whether an index scan or a full-table scan would be more efficient.

The median value deviation represents the number of values that have duplicates away from the median. In the case or the Order Status column, there are only three possible values, open, processing, and closed. Consider the following:

- 10 open orders
- 100, 000 closed orders
- 1 order in processing


In this case the median is the middle value, 10. The number of closed orders is 10,000 times the median which indicates that the column data is highly skewed. In this case the value in the Median Value Deviation column would be presented as:

1, 0, 0, 0, 1, 0, 0, 0

There are 1's at the first and 5th spot in the median value deviation field indicating one column value (value of orders in the *processing* state which appears once) is 1 factor of 10 away from the median and there is a 1 at the 5th position indicating there is a column value (orders in the *closed* state) that appears 5 factors of 10 more often (10,000) than the median value of 10.

A column with a median value deviation of 0, 0, 0, 0, 0, 0, 0, 0 indicates that the column data is not skewed and it is a bad candidate for a histogram, and therefore a full scan of the table would more efficiently satisfy a query than an index scan.

To update the statistics of any object, you can select **Gather** for that column and then click **Display Statistics** or **Collect Statistics**.
To stop gathering statistics for an object, such as a bad candidate for a histogram, select **Drop** for that column and then click **Display Statistics** or **Collect Statistics**.



If you are gathering statistics for a column for which the statistics were missing or out- of-date, then once the statistics collection is complete, you should return to the Overview tab and rerun the cases, because the characteristics of the column may have changed, so the hints to improve performance would also change.

