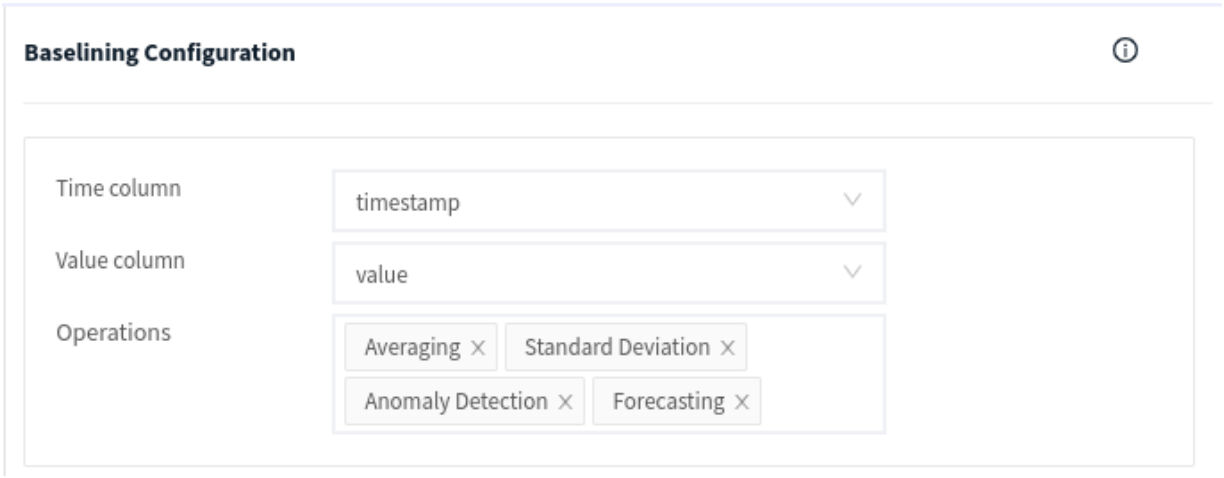
Baselining or ML on Query Data

Baselining feature studies the query data in time series format using machine learning algorithms to implement Anomaly detection and forecasting of the data based on historical data. This feature also includes the mathematical implementation of moving averages and moving standard deviation on the query data. The above feature makes it easy for the user to visualize the behavior of the data from a time-series perspective.

This feature can be mainly divided into four categories namely Anomaly detection, Forecasting, moving average, and moving standard deviation.

Basic Baselining Configuration Guide:

The baselining configuration dialog box appears on the right side of the query page. This configuration mainly needs a time column to indicate the time series data column in the dataset, and a value column to indicate the target value column dependent on the time series. The feature studies the dependent value column from the perspective of the time column and produces the required results which can be visualized using the corresponding meaningful charts. The required baselining features to be implemented are selected from the operations selection box. multiple operation selections can be done at the same time.



Baselining Configuration

Time column: timestamp

Value column: value

Operations: Averaging × | Standard Deviation × | Anomaly Detection × | Forecasting ×

Averaging Configuration

MA window: 96

Anomaly Detection Configuration

10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

STD Configuration

Window size: 48

STD Fraction: 1

Forecasting Configuration

Select Forcast Range in Days

(days)

Select Forcast Range in Hours

(hours)

Reset

Anomaly Detection

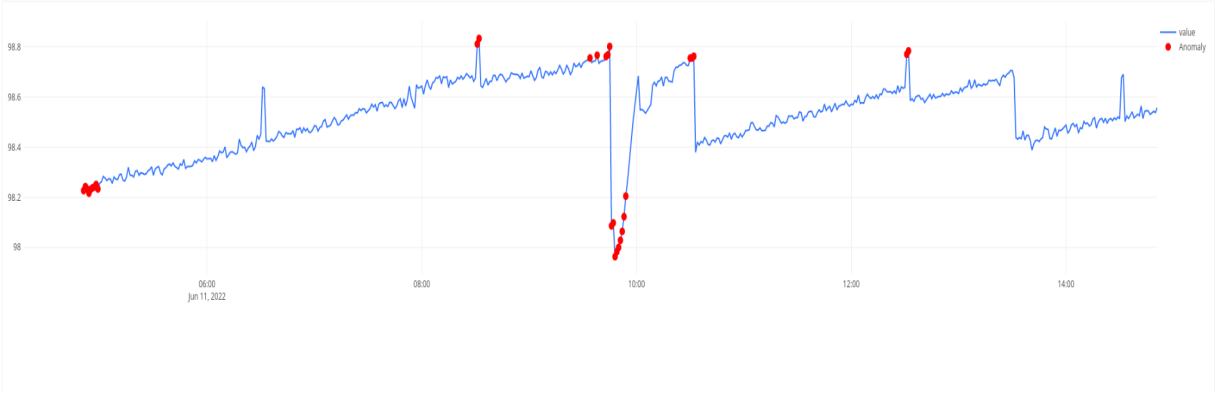
Anomaly Detection determines the anomalous data points from the selected value columns. Anomalous data points are the values in the dataset which deviate substantially from the regular pattern of the target data. Use the anomaly detection slider and select the appropriate percentage of anomalies. 10% gives fewer anomalous points, and 100% detects more anomalous points. The anomaly column is added to the results. Use the anomaly column as the error column in the visualization configuration.

Steps to configure anomaly detection for a query:

1. Open a query page in edit query mode.
2. Select a meaningful time-series column and a value column.
3. Select the Anomaly Detection option from the operations selection box.
4. An anomaly detection box appears. This box contains a slider using which the percentage of the outlier points to be detected can be set. Usually, a 10% to 30% outlier detection is ideal.
5. Save the configuration and execute the query.
6. The resulting dataset contains a column called Anomaly which indicates if the value is an outlier or not.

Steps to visualize anomaly detection for a query:

1. Create a new visualization or open an already existing visualization in edit mode.
2. Select the chart type as a line for better visualization.
3. Select the time column as the X column and the value column as the Y column. Select the Anomaly column as the Error column. A graph similar to the following picture must be rendered.



Averaging

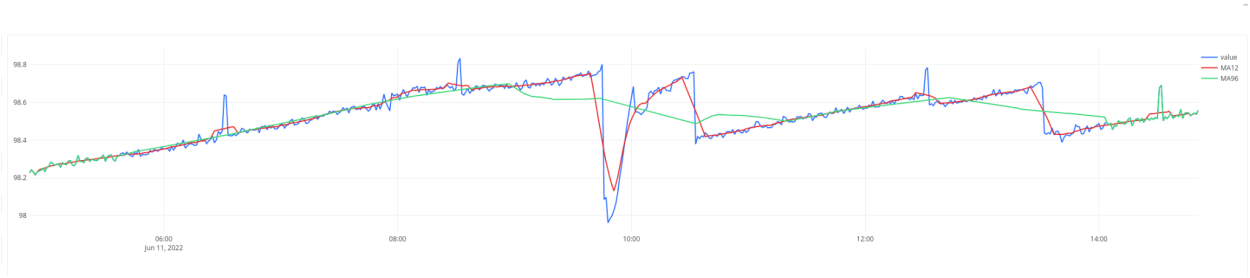
The averaging feature adds a moving average of the target value column to the dataset**.** Two new columns are added to the dataset whose names start with MA (Moving Average). MA12 is added by default, and the second column is configurable. The configuration sets up the moving average window size.

Steps to configure averaging for a query:

1. Open a query page in edit query mode.
2. Select a meaningful time-series column and a value column.
3. Select the Averaging option from the operations selection box.
4. An averaging configuration box appears. This box contains a dropdown called ma window using which the window size of the moving average can be set. The bigger the window size the more curvilinear the graph becomes
5. Save the configuration and execute the query.
6. The resulting dataset contains two columns MA12 whose window size is 12 and another column with MA whose window is previously configured.

Steps to visualize averaging for a query:

1. Create a new visualization or open an already existing visualization in edit mode.
2. Select the chart type as a line for better visualization.
3. Select the time column as the X column and the value column as the Y column. Add the moving average columns i.e; columns with the prefix MA, into the Y column multi-selection box. A graph similar to the following picture must be rendered.



Standard Deviation(STD)

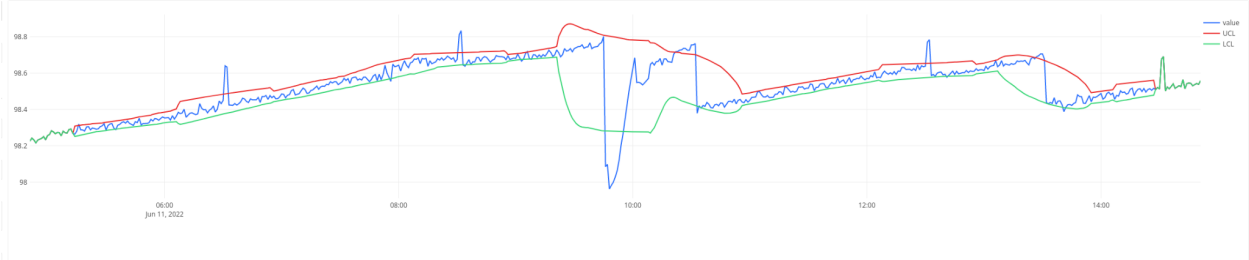
Standard Deviation adds two columns UCL(Upper Control Limit), and LCL(Lower Control Limit), to the dataset. These columns determine the moving standard deviation of the value column selected. The standard deviation is the mathematical expression that defines the data point variation from the dataset average. This feature is used to detect the sudden peaks or crusts in the data.

Steps to configure standard deviation for a query:

1. Open a query page in edit query mode.
2. Select a meaningful time-series column and a value column.
3. Select the standard deviation option from the operations selection box.
4. The standard deviation configuration box appears. This has two dropdowns namely window size and std fraction. Window size is used to select the moving standard deviation window size and std fraction is used to amplify the limits by the fraction selected.
5. The resulting dataset contains two columns LCL and UCL. LCL is used to detect the crusts and UCL is used to detect peaks.

Steps to visualize the LCL and UCL:

1. Create a new visualization or open an already existing visualization in edit mode.
2. Select the chart type as a line for better visualization.
3. Select the time column as the X column and the value column as the Y column. Add the standard deviation columns i.e; LCL and UCL, into the Y column multi-selection box. A graph similar to the following picture must be rendered.



Forecasting

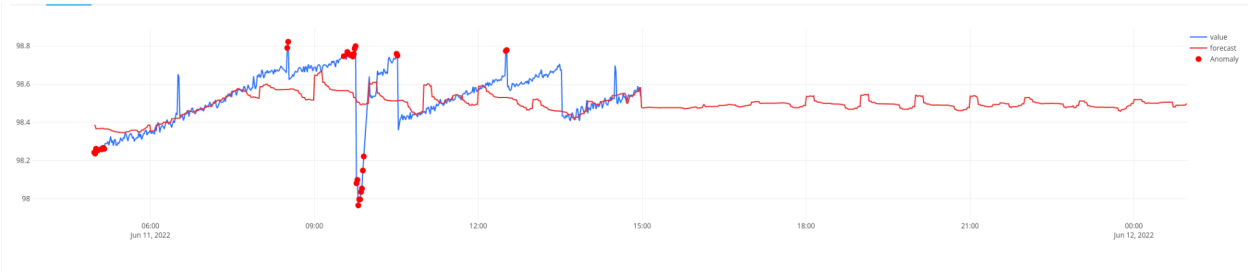
Forecasting determines the forecast of the graph given based on mathematical regression techniques. This feature appends two columns to the dataset, AnomalyForecast, and forecast. AnomalyForecast is a local forecast of the data where the data seems anomalous. The forecasting is done for the window selected in the forecasting configuration.

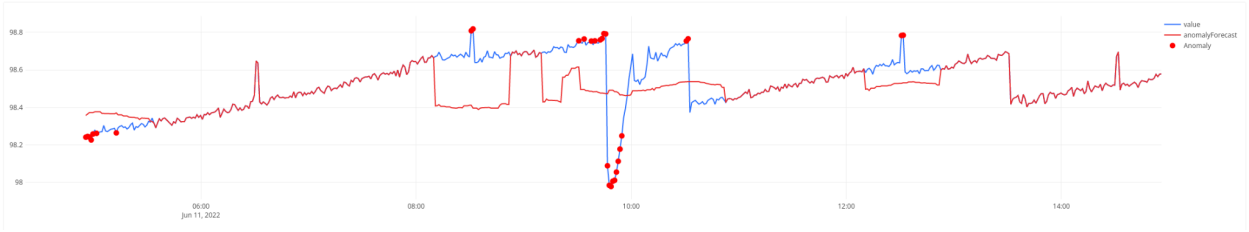
Steps to configure standard deviation for a query:

1. Open a query page in edit query mode.
2. Select a meaningful time-series column and a value column.
3. Select the forecasting option from the operations selection box.
4. The forecasting configuration box appears. This has two sliders namely hours and minutes. Use the sliders to create a time segment in the future for future forecasting.
5. The resulting dataset contains two columns forecast and anomaly-forecast.

Steps to visualize the Forecasting:

1. Create a new visualization or open an already existing visualization in edit mode.
2. Select the chart type as a line for better visualization.
3. Select the time column as the X column and the value column as the Y column. Add the standard deviation columns i.e; forecast and anomaly-forecast, into the Y column multi-selection box. A graph similar to the following picture must be rendered.


value
forecast
Anomaly
98.8
98.6
98.4
98.2
98
06:00
Jun 11, 2022
09:00
12:00
15:00
18:00
21:00
00:00
Jun 12, 2022


value
anomalyForecast
Anomaly
98.8
98.6
98.4
98.2
98
06:00
Jun 11, 2022
08:00
10:00
12:00
14:00