Telmai User Manual

TΣLMAI

HISTORY OF CHANGES

Date	**Changes**
2023-11-01	Document created
2023-12-19	Added Out-of-box metrics
2024-03-15	Group of data sources
2024-05-24	New alert policy and custom metrics
2024-06-10	Canceling scans and create & use projects
2024-07-09	Jira integration
2024-07-17	Lightweight Scans
2024-08-13	RBAC Support
2024-09-23	PII Rules & Jira via API Key
2024-10-30	Notifications settings, admin page, and PII changes
2024-11-18	Updated policies
2024-12-18	Update Top Values policy & data binning

Telmai Definitions & Key Terms

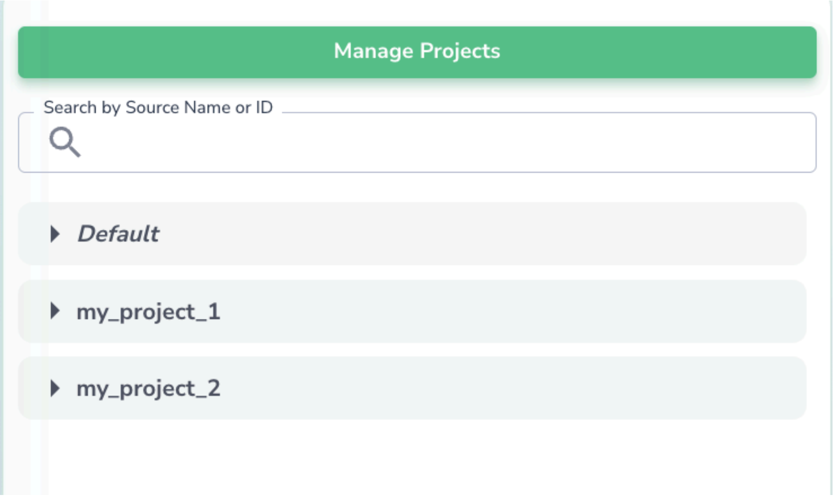
Term	Definition
Monitored Metric	A measurable observation on a data tracked by Telmai ● Metrics defined by Telmai are called “Health Metrics” ● Metrics defined by user are called “Custom Metrics” ● Metric has scope (column, record(s) or table)
Correctness Rules	Means to calculate correctness metric ● Correctness rules are calculated at a record level ● Rules check if record is correct (value based expectation); ex: value is in list
Alerting Policy	Alerting policy is When and How issues/alerts are identified. This is done by defining ● A threshold applied against a metric ● Associated action on threshold violation (notification, other)
Alert/Anomaly	Any metric that violates a policy threshold on a given scan. ● Alerts are created for each of the configured segments ● Alert priority is calculated based on significance
Threshold	A rule to apply against a numeric aggregation. The result of this rule is true or false

Data Integration

Adding a new dataset

Telmai allows you to connect a wide range of dataset types. To connect a new data source:

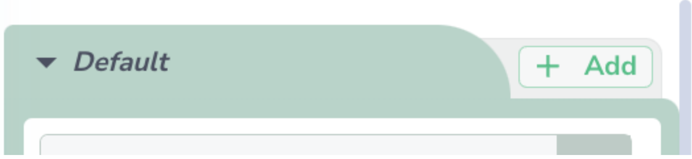
1. Navigate to “Configuration Page”
2. On the left side, you will see list of projects in your given tenant similar to below image, select the desired project (or create a new project first)



Manage Projects

Search by Source Name or ID

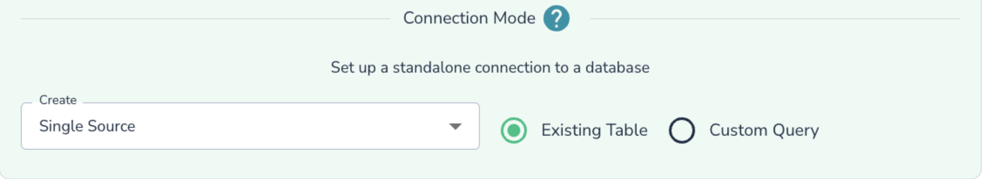
3. Click “+Add button next to the associated project name



4. You will be prompted to select the data source type
5. Once type is selected, you will be asked for more connection details
6. [Optional] If your source type supports grouping, you will be prompted to select the Connection Mode

Connection Mode

Telmai allows you to connect to a Single Source or share credentials to connect to a Group of Sources, ex. several tables in the same dataset.
Single Source connection can be created against any source type; however Grouped Sources can only be created for non-file based sources (ex: BigQuery, Databricks, Snowflake, etc.) You can pick the Connection Mode in the wizard as shown in below images:

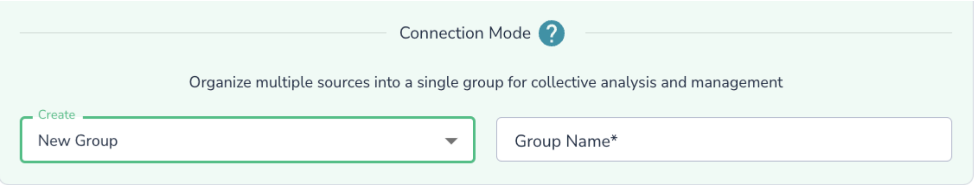


Connection Mode

Set up a standalone connection to a database

Create: Single Source

● Existing Table ○ Custom Query



Connection Mode

Organize multiple sources into a single group for collective analysis and management

Create: New Group

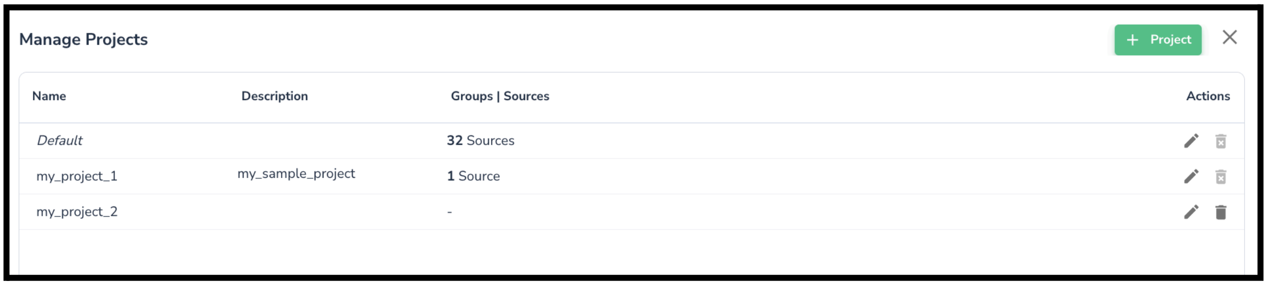
Group Name*

Create a new Project

A project is simply a set of sources and group-of-sources. It's a way to organize datasets in your tenant. Currently users have access to all projects within a tenant, but future Telmai versions will allow setting different permissions per project.

To create a new project, you need to:

1. Click on the "Manage Projects" button; this will launch a menu where you can create, modify or delete a project similar to the image below..



Manage Projects

Name	Description	Groups | Sources	Actions
Default		32 Sources	
my_project_1	my_sample_project	1 Source	
my_project_2		-	

2. Click "+Project" button
3. You will be prompted to provide a project name and an optional description
4. Click "Save"

You can modify project description at any time from the edit button, but can only delete it using the delete button if no sources exist within it.

Permissions can be set at any time for the project. Please refer to RBAC support section.

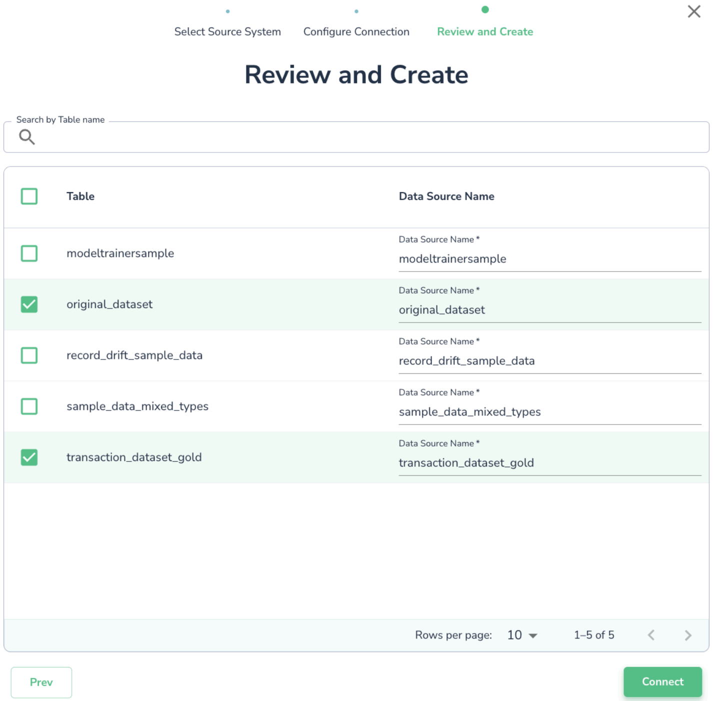
Creating a new Group

Grouping datasets allows you to connect to multiple tables at once, instead of one by one:

1. Reuse same credential to simplify configuration
2. Enable same scan schedule
3. Scan grouped sources in an optimized fashion to increase throughput

To create a new group:

1. Set the Connection Mode to “New Group” and provide group name This name is only used to cluster your sources. A group-id will also be generated that you can use in your API calls.
2. Click “Next” to select sources
3. List of available sources will be displayed



Review and Create

Search by Table name

□	Table	Data Source Name
□	modeltrainersample	Data Source Name * modeltrainersample
☑	original_dataset	Data Source Name * original_dataset
□	record_drift_sample_data	Data Source Name * record_drift_sample_data
□	sample_data_mixed_types	Data Source Name * sample_data_mixed_types
☑	transaction_dataset_gold	Data Source Name * transaction_dataset_gold

Rows per page: 10 1–5 of 5

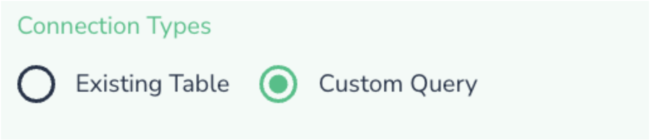
4. Select the desired sources, and click “Connect”

Note: You can add/remove datasets from a given group at any later time.

Using a custom query (only for single source)

For sources that support running SQL queries (example: Databricks, Snowflake, MySQL, etc), you can run a custom SQL query to transform data or combine tables before pulling them into T Telmai. To run a custom query, you will need to do extra steps while connecting the new data source:

1. In “Connection Types”, select “Custom Query”



Connection Types

○ Existing Table ● Custom Query

2. Next step will ask you to write the SQL query

Note: This is only supported when creating a new source. If the source already exists, you will need to delete the existing source and create a new one.

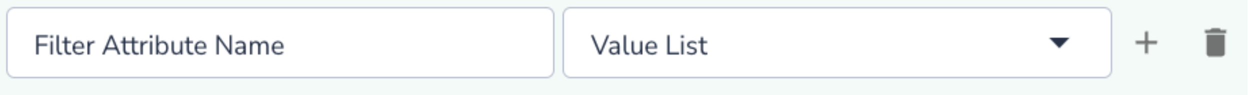
Advanced Configuration

Advanced configurations allow you to get more out of the box features when scanning and analyzing a dataset.

Filter Dataset:

When to use? This configuration should be used when you want to monitor a portion of your data or filter data that you are no longer using.

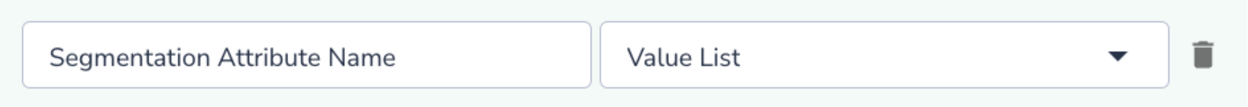
This configuration allows you to filter data based on a specific attribute and allowed values. You can utilize this by specifying an attribute name and allowed values list. You can do this for multiple attributes.



Segmentation

When to use? This configuration can be useful when you have skewed datasets. It helps you detect anomalies for small populations (segments) without getting masked by larger population segments.

Segmentation configuration tells Telmai that you want to analyze the behavior for each segment separately. This means if an anomaly is detected in multiple segments, you will get an anomaly detection per segment. Telmai will still analyze the dataset as a whole regardless.
Segmentation can only be done on a single attribute.



ID Attribute

When to use? This configuration must be defined when wanting to observe data uniqueness or run data binning (future releases of Telmai will not require ID attribute for binning).

The ID attribute is the primary key of your table.

Timestamp Attribute

When to use? This configuration must be defined when wanting to observe data freshness or scan data deltas only.

The timestamp attribute must be an ISO date format. It is used to calculate record freshness.

Delta Only (Checkbox)

When to use? This configuration must be defined when you want to scan data deltas only from your last scan.

Lightweight Scan (Checkbox)

When to use? This configuration tells Telmai to only check the table metadata rather than scanning the whole dataset. To learn more about Lightweight Scans, click here.

Define Monitored Attributes

Once you define your connection details, Telmai will scan the dataset to evaluate the schema and dataset's metadata. You will then be able to define your monitored attributes and custom expressions as needed.

Monitored attributes

Monitored attributes need to be selected to let Telmai know which attributes should be assessed and governed.

Custom Expressions

Custom expressions is a way that allows you to combine multiple attributes to form a new attribute. This can be used to create new columns that didn't exist in your original table. More info can be found here.

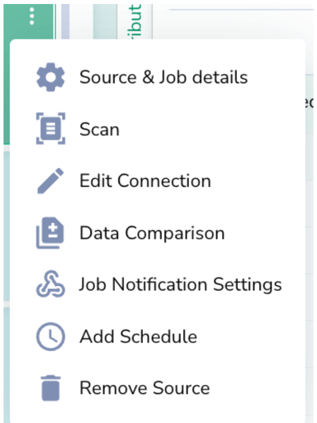
Data Diff

Telmai allows you to compare the differences between your datasets. This feature is useful when you want to ensure data consistency across different tables, and can be used in data reconciliation or migration cases.

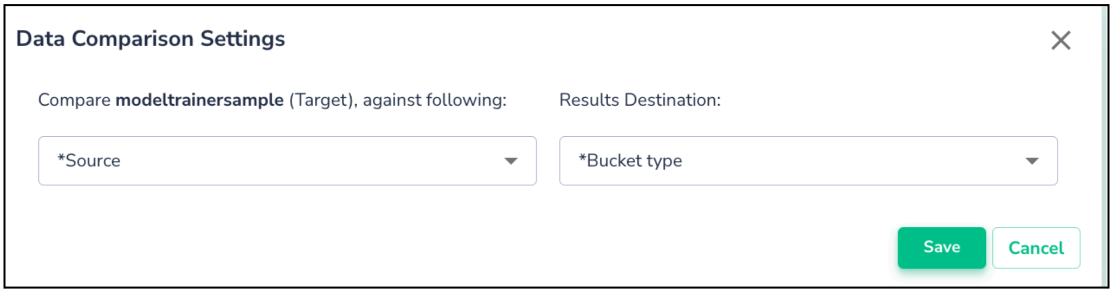
As part of scanning a dataset, Telmai checks for the differences between source and target tables (if configured). In the app, Telmai will provide summary stats on the number of new, missing and changed records. Outside the app, Telmai will store the changed or different records for further usage or analysis.

Telmai's Data Diff feature runs as part of tables' regular scan. You will first need to connect the two datasets, you want to compare, to Telmai using the steps defined here. Then, you will need to update the configs for the table you want to compare (target table) using these steps:

1. Define the ID attribute(both source and target)
2. From the 3 dot menu, click "Data Comparison" option



3. You will be prompted to fill details:


Data Comparison Settings

Compare **modeltrainersample** (Target), against following:

Results Destination:

*Source

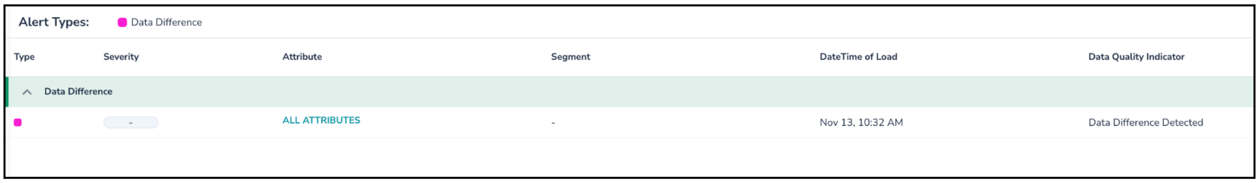
*Bucket type

   a. Source table: Dataset you want to compare to
   b. Result Destination: Output for parquet files (S3, Azure Blob, or GCP storage)
4. Once the details is selected, you will be prompted to fill more details on associated bucket

5. Next scan will analyze the deltas between both datasets, and alerts will be created if deltas exist

Data Diff Alert Example

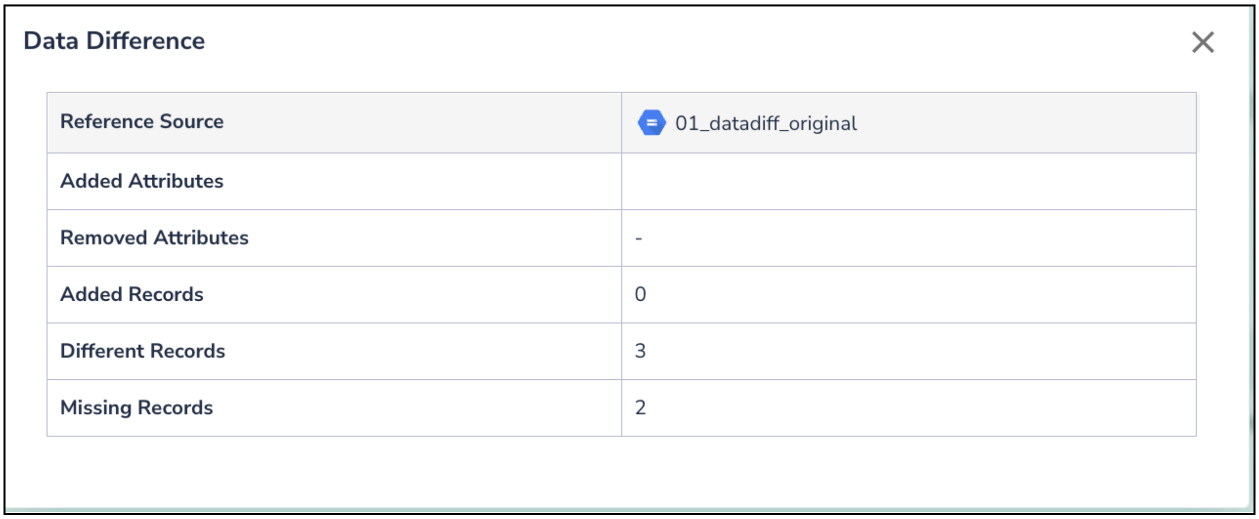
If any differences is detected across the source and target datasets, a “Data Difference” alert is created similar to below picture:



Alert Types: ■ Data Difference

Type	Severity	Attribute	Segment	DateTime of Load	Data Quality Indicator
Data Difference					
■	-	ALL ATTRIBUTES	-	Nov 13, 10:32 AM	Data Difference Detected

Clicking on the alert, will show more details on changed schema and records similar to picture below:



Data Difference

Reference Source	01_datadiff_original
Added Attributes	
Removed Attributes	-
Added Records	0
Different Records	3
Missing Records	2

Lastly, navigating to the output parquet files, you can see more details on changed records.

Job Notification Settings

Telmai allows you to configure webhooks that get triggered when processing job status changes. If defined, the callback gets triggered when:

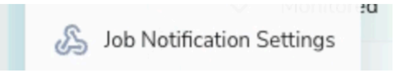
- Job is running
- Intermediate job state changes
- Job is completed

The following params are sent as part of the body:

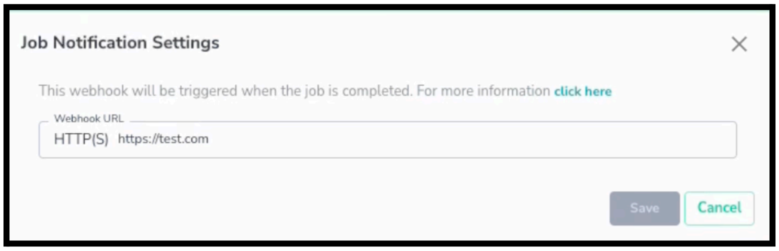
Param	**Description**
tenant	Associated tenant ID where the callback happened
source	Associated dataset source ID
jobId	Associated job ID
status	Job status (IN_PROGRESS, FINISHED, FAILED, PREPARING, BATCH_WAIT)
details	Description of job status

To use this feature, you will need to:

1. Navigate to the configuration page
2. Open the settings for associated dataset
3. Click on Job Notifications Settings



4. In the field for Webhook URL, specify your API URL


Job Notification Settings

This webhook will be triggered when the job is completed. For more information click here

Webhook URL: HTTP(S) https://test.com

5. Once configured, the URL will be called back with the associated values automatically

Important: parameters in URL are not supported

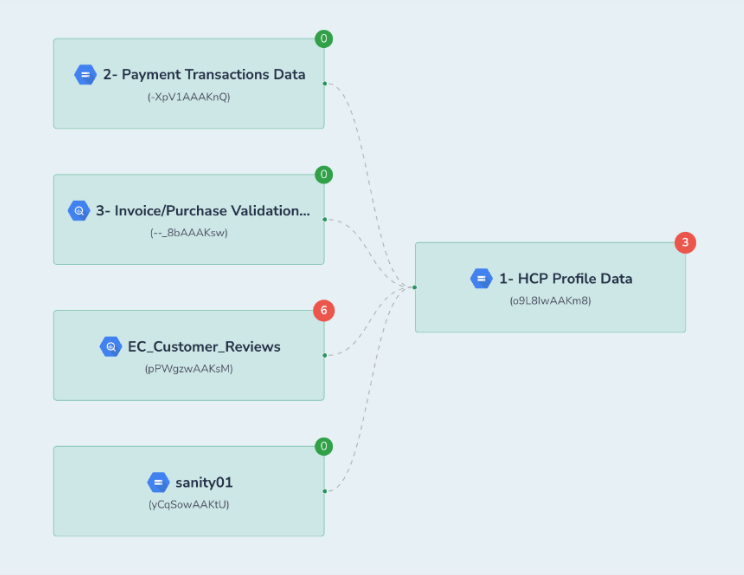
Define Data Lineage

Telmai allows you to define the lineage across your dataset regardless of their type. Telmai uses this information to learn more about your dataset, to calculate correlation metrics, and to alert you about any observed issues.

To define the data lineage,

1. Navigate to the configuration page
2. Click “Lineage” tab
3. Select desired data source
4. Define parent(s) and children as desired
5. Repeat steps 3 & 4 for the rest of the dataset

Telmai will use that information to construct the lineage graph that gets used when analyzing the data:



Note: The data lineage is required to calculate for “Record count correlation” & “Uniqueness correlation” drifts.

Scan and Process Data

Telmai loads your dataset to look for anomalies. Telmai learns about your data properties with each scan. This allows Telmai to run its ML models (and other checks) to detect anomalies within the data.

Regular Data Scan

Telmai supports different ways of scanning data for governance. Here're the different ways to scan your data.

- **Full dataset scan**
  If the "Delta Only" checkbox is not checked, Telmai will scan the defined dataset in full for analysis and monitoring.
- **Delta data scan**
  If the "Delta Only" checkbox is checked, Telmai will scan the defined dataset based on the timestamp of the last scan. In the first scan, the full dataset is scanned

Lightweight Scans

Telmai allows you to scan only a table's metadata instead of scanning the whole table. This helps you monitor high level information about your table's health, and still allow you to alert for quality issues.
To enable this feature, just check the "Light Scan" option under advanced options as defined here.

□ Light Scan

In case of group, the option will be under "Group Settings".

Note: This feature is only supported for BigQuery and Databricks. Future Telmai versions will support more sources.

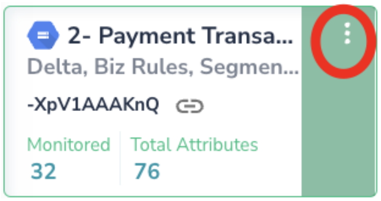
Triggers

Data scan and processing jobs can be triggered through:

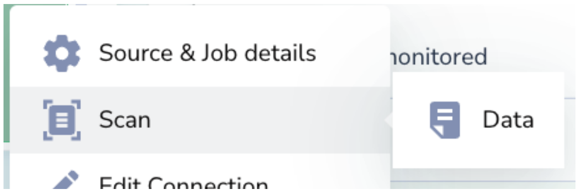
- Telmai's UI
- Telmai's APIs
- Scheduler

Trigger job using Telmai UI:

1. Navigate to configuration page
2. Click the 3 dot options tab for the corresponding data source



3. A window will be prompted, navigate to Scan->Data

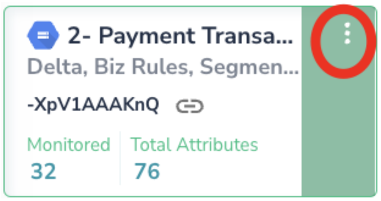


Trigger job using Telmai API

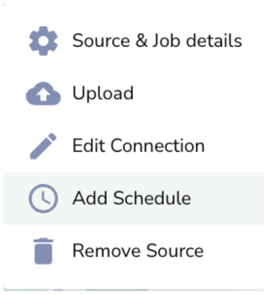
Please refer to documentation here:
https://docs.telm.ai/telmai/upload-data/upload-data-api

Trigger job using Scheduler

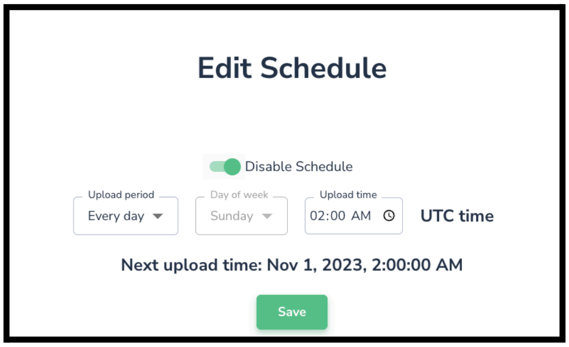
1. Navigate to configuration page
2. Click the 3 dot options tab for the corresponding data source



3. Click “Add Schedule”



4. You will be prompted to enable the schedule and with schedule detail to fill



Edit Schedule

Disable Schedule

Upload period: Every day

Day of week: Sunday

Upload time: 02:00 AM

UTC time

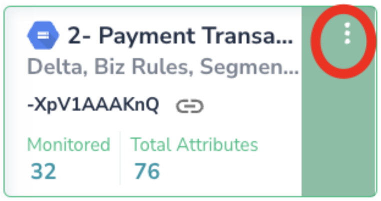
Next upload time: Nov 1, 2023, 2:00:00 AM

Historic data scan

This feature can be utilized when you have an existing dataset that is partitioned by date and want to look for anomalies in the last couple of days or train Telmai's ML model faster.

To utilize this feature,

1. Have a datasource that uses a SQL engine (Databricks, Snowflake, etc)
2. Define timestamp attribute
3. Navigate to configuration page
4. Click the 3 dot options tab for the corresponding data source



2- Payment Transa...

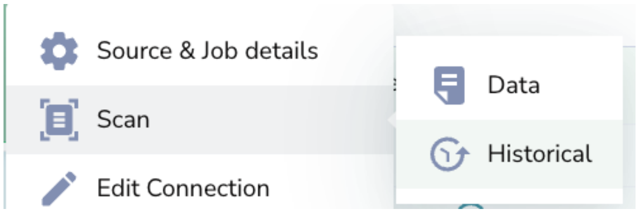
Delta, Biz Rules, Segmen...

-XpV1AAAKnQ

Monitored: 32

Total Attributes: 76

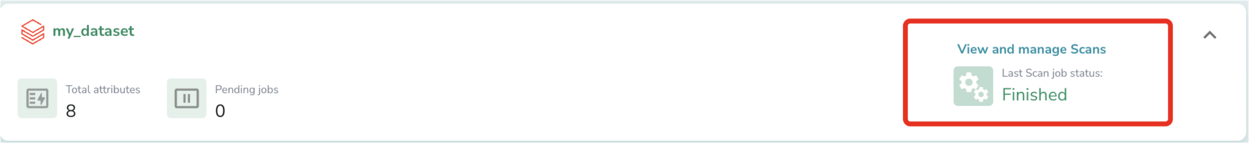
5. A window will be prompted, navigate to Scan->Historical



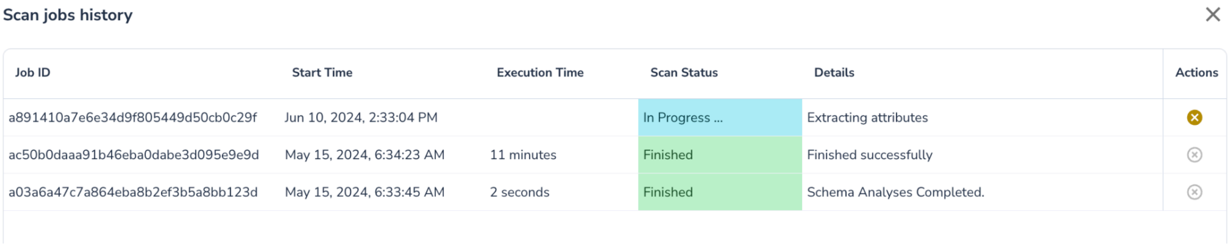
Note: This feature can only be used with SQL based engines datasources.

View and Manage Scans

In case you like to see the status of past or active scans, you can click "View and manage Scans" on the associated source similar to the image below:



Clicking this button will launch a new window were you will be able to view the status of all associated scans.


Scan jobs history

Job ID	Start Time	Execution Time	Scan Status	Details	Actions
a891410a7e6e34d9f805449d50cb0c29f	Jun 10, 2024, 2:33:04 PM		In Progress ...	Extracting attributes	⊗
ac50b0daaa91b46eba0dabe3d095e9e9d	May 15, 2024, 6:34:23 AM	11 minutes	Finished	Finished successfully	⊗
a03a6a47c7a864eba8b2ef3b5a8bb123d	May 15, 2024, 6:33:45 AM	2 seconds	Finished	Schema Analyses Completed.	⊗

Canceling scans

In cases where the user wants to stop any unnecessary scans or if a scan was triggered incorrectly, the user can click on the cancel button “x” for the associated scan to request cancellation.
Telmai will register the cancellation request and attempt cancellation. This can take a few minutes to complete based on the stage the job was in.

Note: Latest state for the job will always be displayed.

Data Profiling

Once data is processed by Telmai, you can immediately profile your data at column and row level. That data is available in the Data Health and Investigator pages.

Calculated Health Metrics

Telmai automatically calculates a set of health metrics out of the box. These metrics are:

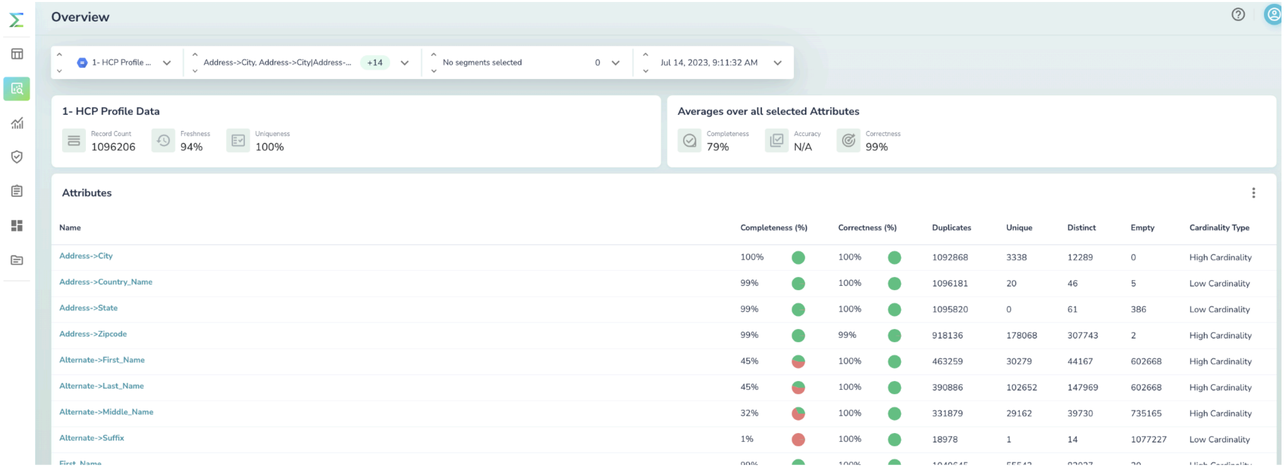
Column	**Description**
Total Record Count	The total size of the monitored table. This is only available when CDC is enabled (i.e., the “delta only” flag is on). If the flag is off or the monitored source is not an SQL database, this field will be set to N/A.
Record Count	If CDC is enabled it reflects the size of the delta, otherwise, it's the size of the entire set of data.
Completeness	Percentage of not null/missing/placeholder values. Completeness is tracked both at the data source level as well as the attribute level: ● Attribute level: percent of records where the attribute value is not null, not empty, or not one of the user-defined placeholders like N/A. ● Data Source level: the compounded average of all attribute level completeness within a data source.
Correctness	Correctness is tracked both at the data source and attribute levels. Correctness is calculated only for attributes where Expectations are set, otherwise, the default is 100%. ● Attribute level: percent of records where the attribute value meets all expectations, set for the attribute. ● Source level: the compounded average of all attribute level correctness within a data source
Freshness	Freshness is tracked both at data source and record level. ● Record level freshness is defined by setting an expectation on a timestamp attribute within the data

	source (e.g.,. “Record Update Date” attribute to be no more than 1 month from now). To mark an attribute as a timestamp use the Advanced section of the Edit Connection menu. If the timestamp attribute is not configured this KPI will be N/A. ● Table level freshness is based on the time since the table was last updated
Uniqueness	Uniqueness of the records based on an ID attribute. The KPI measures the ratio of records with unique id versus the total number of records. For example, if out of 10 records, there are 2 records sharing the same value the uniqueness is 80%. An attribute can be marked as an ID attribute in the Advanced section of the Edit Connection menu. If ID attribute is not configured this KPI will be N/A
Accuracy	The accuracy of the values, based on historical data analysis, and detects when current values don’t match predictions. For example, if the revenue for the company Acme Inc was slowly growing from $4M to $5M over the past year, but in today’s observation it’s 20M, then such a value is considered inaccurate. Accuracy is only calculated for attributes that have “Custom Metrics” Configured, otherwise, the default is 100%. ● Attribute level: the ratio between custom metrics dimensions where no drift was detected to the total number of dimensions. If multiple custom metrics are defined for an attribute, then the minimum accuracy is picked for this attribute. ● Data Source level: the compounded average of all attribute level accuracy within a data source.

These metrics can be found in different places across the application depending on the scope you’re looking at. In case lightweight scan is enabled, only a subset of these metrics are measured based on the source type.

Data Health Overview Page

Dataset Health Overview page can be utilized to understand the status of a given dataset and its attributes at a given scan.



Overview

1- HCP Profile ... | Address->City, Address->City|Address-... +14 | No segments selected 0 | Jul 14, 2023, 9:11:32 AM

1- HCP Profile Data

Record Count: 1096206
Freshness: 94%
Uniqueness: 100%

Averages over all selected Attributes

Completeness: 79%
Accuracy: N/A
Correctness: 99%

Attributes

Name	Completeness (%)		Correctness (%)		Duplicates	Unique	Distinct	Empty	Cardinality Type
Address->City	100%	●	100%	●	1092868	3338	12289	0	High Cardinality
Address->Country_Name	99%	●	100%	●	1096181	20	46	5	Low Cardinality
Address->State	99%	●	100%	●	1095820	0	61	386	Low Cardinality
Address->Zipcode	99%	●	99%	●	918136	178068	307743	2	High Cardinality
Alternate->First_Name	45%	●	100%	●	463259	30279	44167	602668	High Cardinality
Alternate->Last_Name	45%	●	100%	●	390886	102652	147969	602668	High Cardinality
Alternate->Middle_Name	32%	●	100%	●	331879	29162	39730	735165	High Cardinality
Alternate->Suffix	1%	●	100%	●	18978	1	14	1077227	Low Cardinality

In this page, you can see the Health KPIs defined in Calculated Health Metrics. You can also see attribute level metrics:

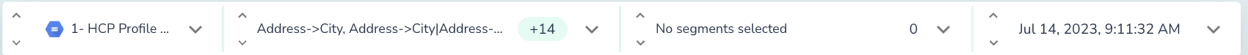
KPI	Description
Duplicates	How many records have a duplicate value (appear more than once) in the attribute
Unique	How many records have a unique value in the attribute
Distinct	Count of distinct values in the attribute
Empty	Count of empty values, i.e. nulls or empty strings in the attribute
Cardinality	Cardinality of the values: Low or High

Note: for all attribute level metrics, table level metric is calculated as average across different columns.

Selector Component

This component allows you to select:

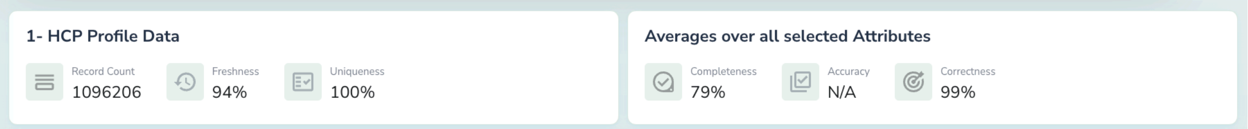
- Dataset
- Attributes to calculate metrics on
- Segments to filter on
- Time of the data scan



1- HCP Profile ... | Address->City, Address->City|Address-... +14 | No segments selected 0 | Jul 14, 2023, 9:11:32 AM

Health KPIs Summary

This component provides a summary of key Data Quality KPIs on table and attribute level



1- HCP Profile Data

Record Count 1096206

Freshness 94%

Uniqueness 100%

Averages over all selected Attributes

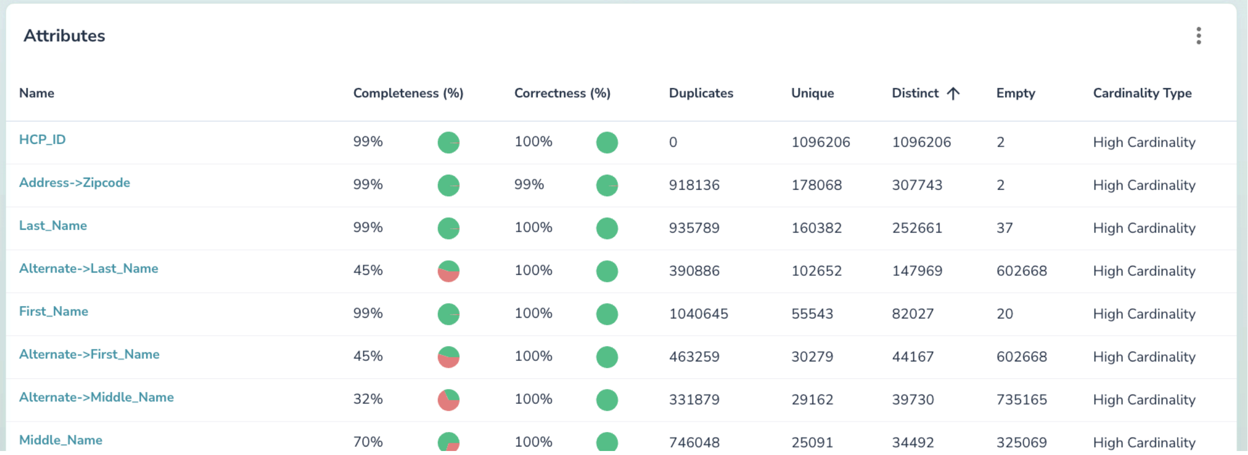
Completeness 79%

Accuracy N/A

Correctness 99%

Attribute(s) KPIs

For every attribute (or column) we are calculating the corresponding profiling metrics

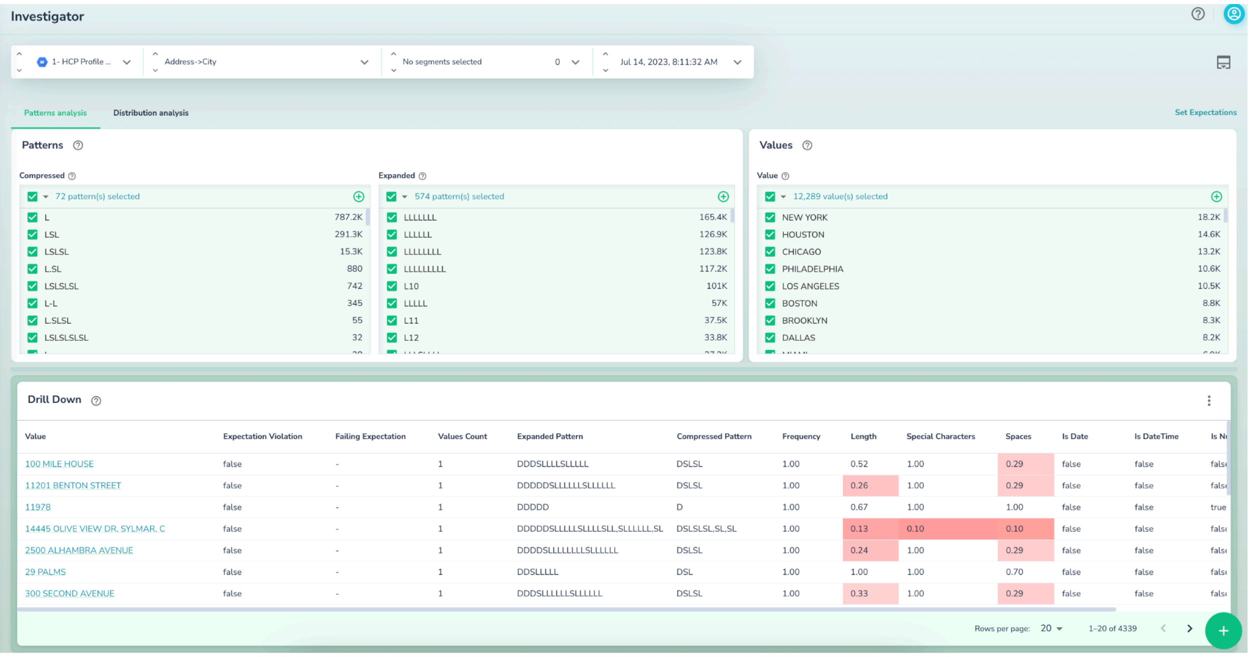


Attributes

Name	Completeness (%)	Correctness (%)	Duplicates	Unique	Distinct ↑	Empty	Cardinality Type
HCP_ID	99% ●	100% ●	0	1096206	1096206	2	High Cardinality
Address->Zipcode	99% ●	99% ●	918136	178068	307743	2	High Cardinality
Last_Name	99% ●	100% ●	935789	160382	252661	37	High Cardinality
Alternate->Last_Name	45% ●	100% ●	390886	102652	147969	602668	High Cardinality
First_Name	99% ●	100% ●	1040645	55543	82027	20	High Cardinality
Alternate->First_Name	45% ●	100% ●	463259	30279	44167	602668	High Cardinality
Alternate->Middle_Name	32% ●	100% ●	331879	29162	39730	735165	High Cardinality
Middle_Name	70% ●	100% ●	746048	25091	34492	325069	High Cardinality

Data Investigator

Another page where you can drill down your dataset is the Investigator page:



1- HCP Profile ... | Address->City | No segments selected 0 | Jul 14, 2023, 8:11:32 AM

Patterns

Compressed — ☑ 72 pattern(s) selected

☑	L	787.2K
☑	LSL	291.3K
☑	LSLSL	15.3K
☑	L.SL	880
☑	LSLSLSL	742
☑	L-L	345
☑	L.SLSL	55
☑	LSLSLSLSL	32

Expanded — ☑ 574 pattern(s) selected

☑	LLLLLLL	165.4K
☑	LLLLLL	126.9K
☑	LLLLLLLL	123.8K
☑	LLLLLLLLL	117.2K
☑	L10	101K
☑	LLLLL	57K
☑	L11	37.5K
☑	L12	33.8K

Values

Value — ☑ 12,289 value(s) selected

☑	NEW YORK	18.2K
☑	HOUSTON	14.6K
☑	CHICAGO	13.2K
☑	PHILADELPHIA	10.6K
☑	LOS ANGELES	10.5K
☑	BOSTON	8.8K
☑	BROOKLYN	8.3K
☑	DALLAS	8.2K

Drill Down

Value	Expectation Violation	Failing Expectation	Values Count	Expanded Pattern	Compressed Pattern	Frequency	Length	Special Characters	Spaces	Is Date	Is DateTime	Is N
100 MILE HOUSE	false	-	1	DDDSLLLLSLLLLL	DSLSL	1.00	0.52	1.00	0.29	false	false	false
11201 BENTON STREET	false	-	1	DDDDDSLLLLLLSLLLLLL	DSLSL	1.00	0.26	1.00	0.29	false	false	false
11978	false	-	1	DDDDD	D	1.00	0.67	1.00	1.00	false	false	true
14445 OLIVE VIEW DR, SYLMAR, C	false	-	1	DDDDDSLLLLLSLLLLSLL,SLLLLLL,SL	DSLSLSL,SL,SL	1.00	0.13	0.10	0.10	false	false	false
2500 ALHAMBRA AVENUE	false	-	1	DDDDSLLLLLLLLSLLLLLL	DSLSL	1.00	0.24	1.00	0.29	false	false	false
29 PALMS	false	-	1	DDSLLLLL	DSL	1.00	1.00	1.00	0.70	false	false	false
300 SECOND AVENUE	false	-	1	DDDSLLLLLLSLLLLLL	DSLSL	1.00	0.33	1.00	0.29	false	false	false

Rows per page: 20 | 1–20 of 4339

Similar to the overview page, this page has a similar Selector Component.
Other components can be seen in this page:

Patterns

This section shows the currently selected data scan's data distribution in terms of compressed and expanded patterns. This can be used to understand the data, as well as, build more data restrictions.

Values

Sample of the datasets' top values.

Drill Down

Sample of the datasets' top values that have similar properties. Clicking on any of the values in this table shows your more details on the value properties.
This table can also be used to understand data policy violations that are described in a later section.

In this table, you would find the following columns per group/cohort of values

Column	**Description**
Value	Sample data that has same characteristics
Expectation Violation	True if value violates a correctness expectation
Failing Expectation	Associated correctness rule (only valid for violations)
Values Count	Count of appearances of the value within its attribute
Expanded Pattern	Anomaly score, based on value's pattern, i.e. representation of the value string but with each character that belongs to one of these types: alphabet/letter (L), digit (D), or space (S) being replaced by the character representing the type. A long sequence of the same type character is represented by the character and a number indicating the length of the sequence
Compressed Pattern	Anomaly score, based on value's short pattern, i.e. representation of the value string but with each sequence of characters that belong to one of these types: alphabet/letter (L), digit (D), or space (S) being reduced to the single character representing the type
Frequency	Anomaly score, based on how many records share the same value. 1 being normal, 0 being abnormal.
Length	Anomaly score, based on count of characters in the value; 1 being normal, 0 being abnormal.
Special Characters	Anomaly score, based on how many special characters detected in the value; 1 being normal, 0 being abnormal.
Spaces	Anomaly score based on how many whitespace characters detected in the value; 1 being normal, 0 being abnormal.
Is Date	True if value is ISO date, false otherwise
Is DateTime	True if value is date, false otherwise
Is Number	True if value is number, false otherwise
Is Alpha	True if value is alphabetical characters, false otherwise
Is Credit Card	True if value matches pattern for credit card
Is IP Address	True if value matches pattern for IP Address

Is Phone Number	True if value matches pattern for phone number
Is Social Security	True if value matches pattern for SSN (Social Security Number)
Is Zip Code	True if value matches pattern for Zip Code

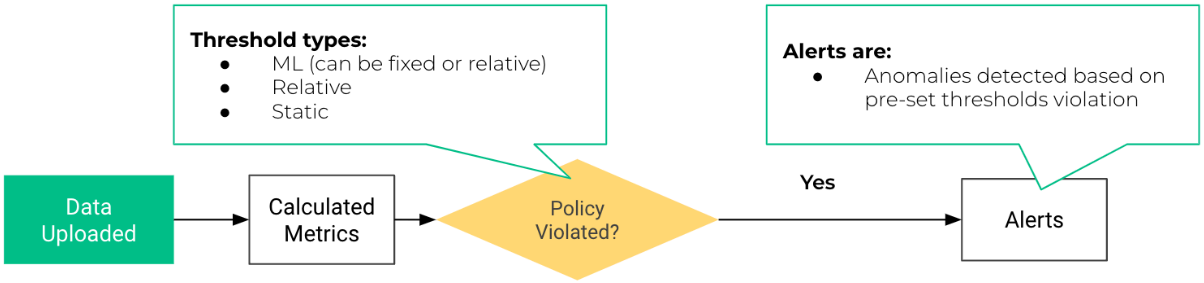
Creating and Viewing Policies

Telmai creates alerts when an anomaly or issue is detected. Alert, by definition, is a warning that something is not as expected.

What is monitored?

1. Out of the box, Telmai monitors a bunch of metrics around the table meta data and Health KPI (mentioned in Data Health KPIs)
2. Custom metrics that can be defined by the users

Each of the monitored metrics is validated against a set of policies (thresholds). If any of the metrics is violated, it becomes an alert. The below flow diagram shows how alerts are generated.

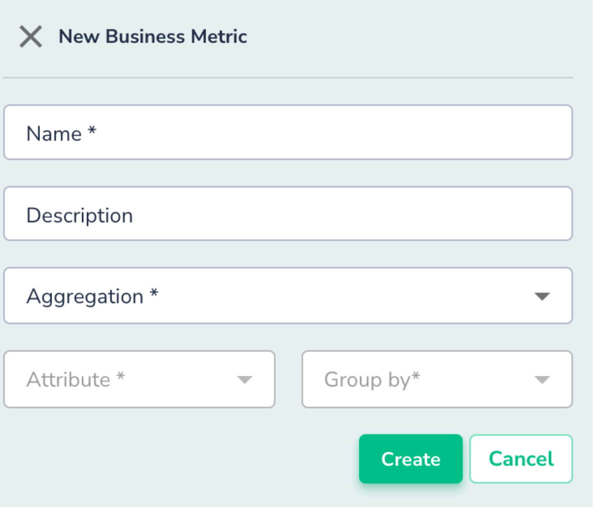


On top of the Calculated Health Metrics, Telmai allows you to define your own set of custom metrics that you need to track.

Creating Custom Metrics

Users have the ability to add custom metrics that they need to track for anomalies. To add a new Custom Metric,

1. Select the dataset you want to add metric for
2. Navigate to “Alerting Policies” page and choose the “Metrics” tab
3. In “Custom Metrics” secretion, Click the “+ Custom Metric” button
4. A new window will appear to allow you to define the metric



5. Fill the parameters:
    a. **Name**: Metric name (this metric will get automatically generated as you define the other params)
    b. **Description**: Useful explanation of what that metric is
    c. **Aggregation**: Aggregation function (SUM, AVG, ect.)
    d. **Attribute**: Column that will be used for aggregation
    e. **Group by:** Grouping attribute. Choose “None”, if grouping is not required
6. Click Create
7. Telmai will start monitoring this metric in future scans
   IMPORTANT: Metric being defined here means it’s only available for tracking & alerting. But no alerts will be generated until a policy is created using this metric.

Defining Correctness Rules

Users can define a set of expectations of what an attribute values are expected to be (example: values in column A has to be between 1 and 100). An attribute row-value is marked incorrect if it doesn't match the expectation. This incorrectness starts getting tracked as part of the correctness KPI.

To define expectations on a given attribute, you will need to:

1. Navigate to "Correctness Rules" page
2. Select dataset you want to add expectation for
3. Click the + button to pick which attribute you want to add expectation for
4. Under the new attribute, click the + button to start adding expectations. You will be prompted to enter:
    a. Expectation type
    b. Expectation metadata
5. Repeat steps 4 & 5 as needed
6. Telmai will start monitoring this expectation against correctness metric in future scans

Here is the list of possible expectation rules you can define:

Rule	Description	Example
Accepted Compressed Patterns	List of allowed compressed patterns	"LD"
Accepted Expanded Patterns	List of allowed expanded patterns	"LLLDDD"
Accepted Values	List of allowed values	"Aa" "Bb" "Cc"
Date Time Range	Allowed time range	From: 1d - To: 1M From: 1d - To: @now
Empty Values (note this specific case is handled for completeness)	Values to be treated as empty	"_EMPTY_" "null"
Frequency	Frequency is how many times a distinct value is expected to occur	From: 0 - To: 1
Is Date	Value is ISO Date (yyyy-mm-dd)	True/False
Is DateTime	Value is ISO Datetime (yyyy-mm-ddTHH:MM:SSZ)	True/False

Length	String length	From: 5 - To: 5
Numeric Value Range	Value range	From: 0 - To: 100
Regex	Regular expression	“^helloWorld\d”
Rejected Compressed Patterns	List of not allowable compressed patterns	“LD”
Rejected Expanded Patterns	List of not allowable expanded patterns	“LLLDDD”
Rejected Values	Not allowable values	“Aa” “Bb” “Cc”
Spaces Count	Count number of spaces	From 0 - To: 10
Special Characters Count	Count of special characters in the value	From: 0 - To: 10
Tokens Count	Count of tokens in the value, splitted by any number of spaces or punctuations	From: 0 - To: 10
Trimmed String	If the string has no leading and trailing spaces	True/False
Value Is Alpha	Boolean check for alphabetical letter only	True/False
Value Is Digit	Boolean check for single digit only	True/False
Value Is Email	Boolean check for valid email string format	True/False
Value Is Number	Boolean check for numerical value	True/False
Value Is URL	Boolean check for valid URL string format	True/False
Words Count	Number of words, i.e split by whitespace characters	From 5 - To 10
Contains PII	Value contains PII pattern for either of Credit cards - IP Addresses - Social security numbers - Phone numbers - Zip codes - Email	True/False

	- Credit card	
Contains IP Address	Value contains PII pattern for IP Address	True/False
Contains Social Security Number	Value contains PII pattern for SSN	True/False
Contains Phone Number	Value contains PII pattern for Phone Number	True/False
Contains Zip Code	Value contains PII pattern for Zip Code	True/False
Contains Credit Card Number	Value contains PII pattern for Credit card number	True/False

Creating Alert Policies

Telmai has a set of out of the box policies that are automatically created when adding any new dataset plus the user has the ability to add their own policies.

Out-of-box Policies

These policies focuses on:

- **Metrics drifts:** Telmai uses its preparatory ML models to detect when metrics have deviated from their normal trends.

Below is the list of predefined policies

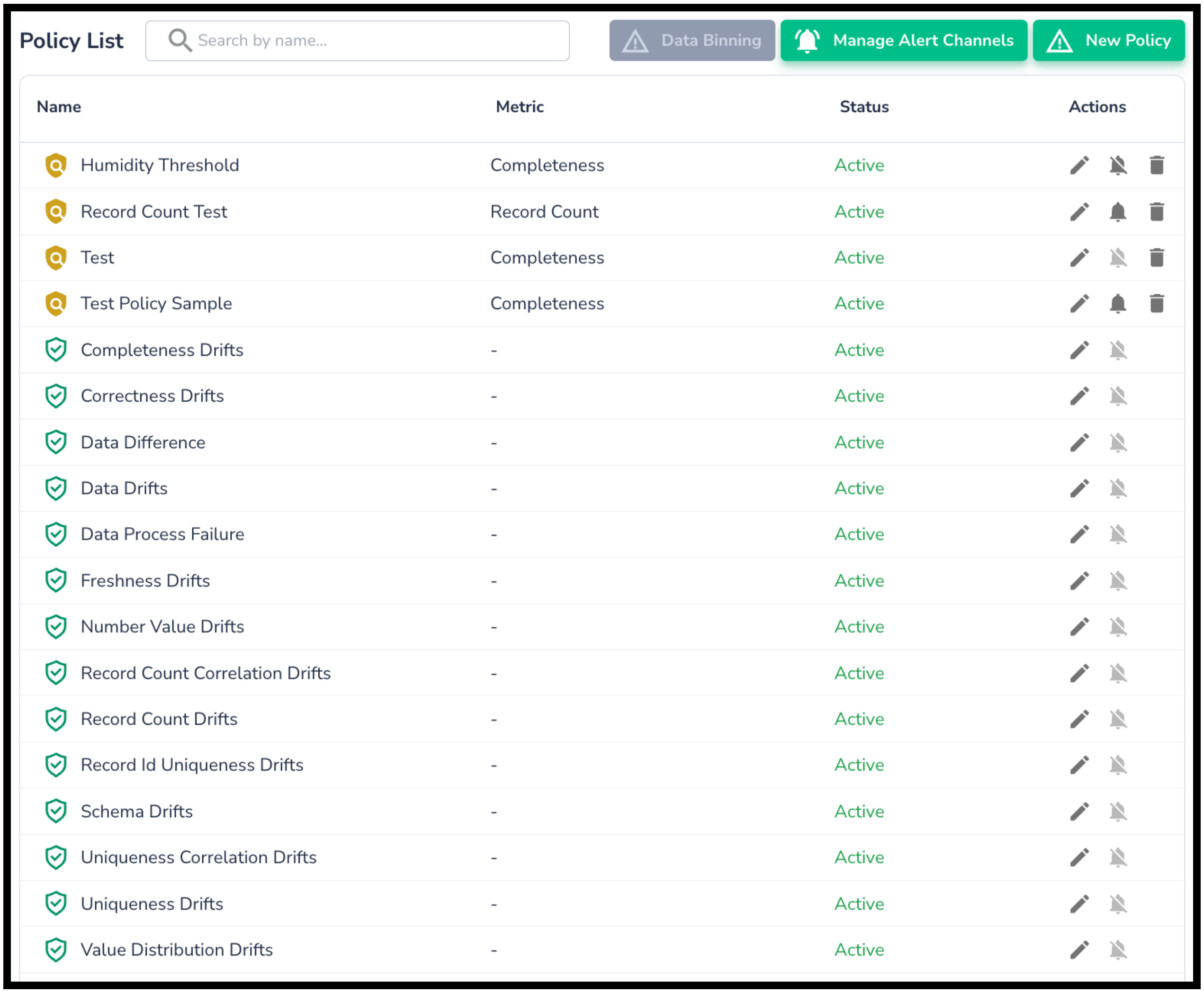
Policy	**What is Monitored?**
Completeness Drifts	Changes in percentage of empty records
Correctness Drifts	Changes in percentage of correct records based on defined Expectations
Correctness Rules Violation	Detection for any rule violation
Data Difference	Added, removed or different records across tables. Note: Data diff must be enabled first
Data Drifts	Various changes of values format (length, spec characters, number of tokens etc) Note: This policy is disabled by default
Freshness Drifts	Table and record level freshness drifts Note: For this policy to take effect, timestamp attribute must be identified
Numeric Value Drifts	Changes in averages over numeric values for the attribute Note: This policy is disabled by default
PII Exposure	Presence of PII data in your dataset (email, IP, etc). Note: This policy is disabled by default
Record Count Correlation Drifts	Changes in data volume across correlated tables with defined lineage.

	Note: For this policy to take effect, you need to define the tables lineage
Record Count Drifts	Changes in number of scanned rows
Record Id Uniqueness Drifts	Changes in data uniqueness of record Id Note: For this policy to take effect, you need to enable table Id
Schema Drifts	Changes in schema; example: column added or removed
Top Values Drifts	This feature monitors the changes for the top 10 values. An alert is generated if new values are added or removed
Uniqueness Correlation Drifts	Changes in data uniqueness across tables with defined lineage. Note: For this policy to take effect, you need to define the tables lineage
Uniqueness Drifts	Changes in data uniqueness across attributes
Value Distribution Drifts	Changes in distributions of the top 20 most frequent categorical values

Modifying/Reviewing Existing Policies

At any time, you can see existing policies, modify them or disable them. This can be done by navigating to the “Alerting Policies” page and selecting the corresponding dataset. The corresponding policies will be seen in as table with the following properties:

- Name: Policy name (User defined policies are marked in yellow & Telmai out-of-box policies are marked in green)
- Metric: Monitored metric
- Status: Enable status
- Actions:
    - Edit: Modifying policy properties or scope
    - Notifications On\Off
    - Delete



Name	Metric	Status	Actions
Humidity Threshold	Completeness	Active	✎ 🔕 🗑
Record Count Test	Record Count	Active	✎ 🔔 🗑
Test	Completeness	Active	✎ 🔕 🗑
Test Policy Sample	Completeness	Active	✎ 🔔 🗑
Completeness Drifts	-	Active	✎ 🔕
Correctness Drifts	-	Active	✎ 🔕
Data Difference	-	Active	✎ 🔕
Data Drifts	-	Active	✎ 🔕
Data Process Failure	-	Active	✎ 🔕
Freshness Drifts	-	Active	✎ 🔕
Number Value Drifts	-	Active	✎ 🔕
Record Count Correlation Drifts	-	Active	✎ 🔕
Record Count Drifts	-	Active	✎ 🔕
Record Id Uniqueness Drifts	-	Active	✎ 🔕
Schema Drifts	-	Active	✎ 🔕
Uniqueness Correlation Drifts	-	Active	✎ 🔕
Uniqueness Drifts	-	Active	✎ 🔕
Value Distribution Drifts	-	Active	✎ 🔕

User Defined Policy

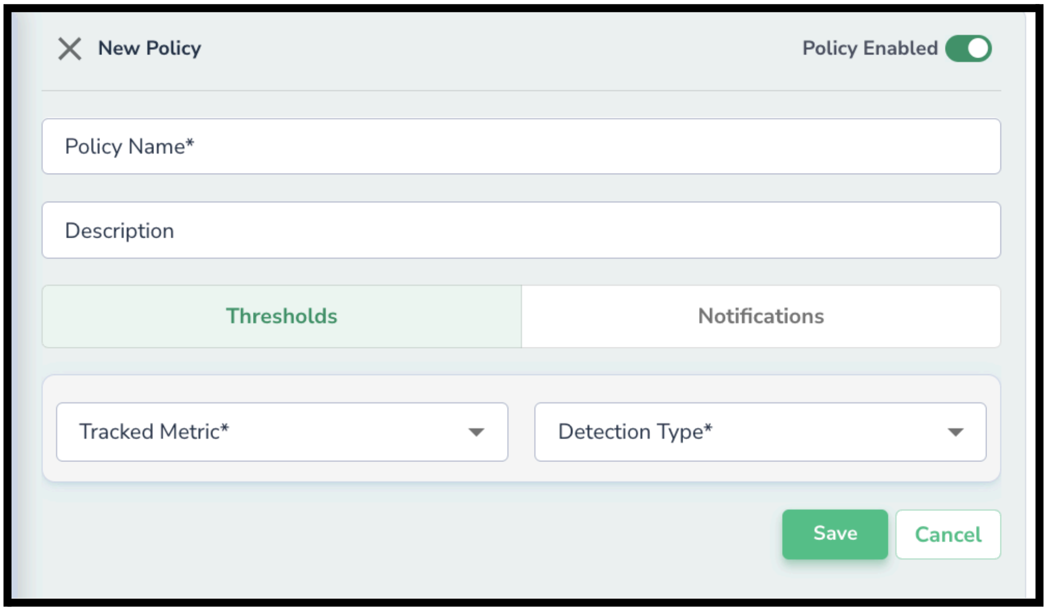
Users can define their policy to track any of the calculated metrics. The metrics are tracked by specifying the threshold that creates an alert when violated. Below is the list of supported thresholds.

Supported Threshold

- **Telmai-ML**
  Telmai's out-of-box anomaly detection using Machine Learning
- **Relative Drift**
  Percentage change in data compared to past scan (or average of scans)
- **Acceptable Range**
  Check a metric within a given range

To create the policy, you will need to:

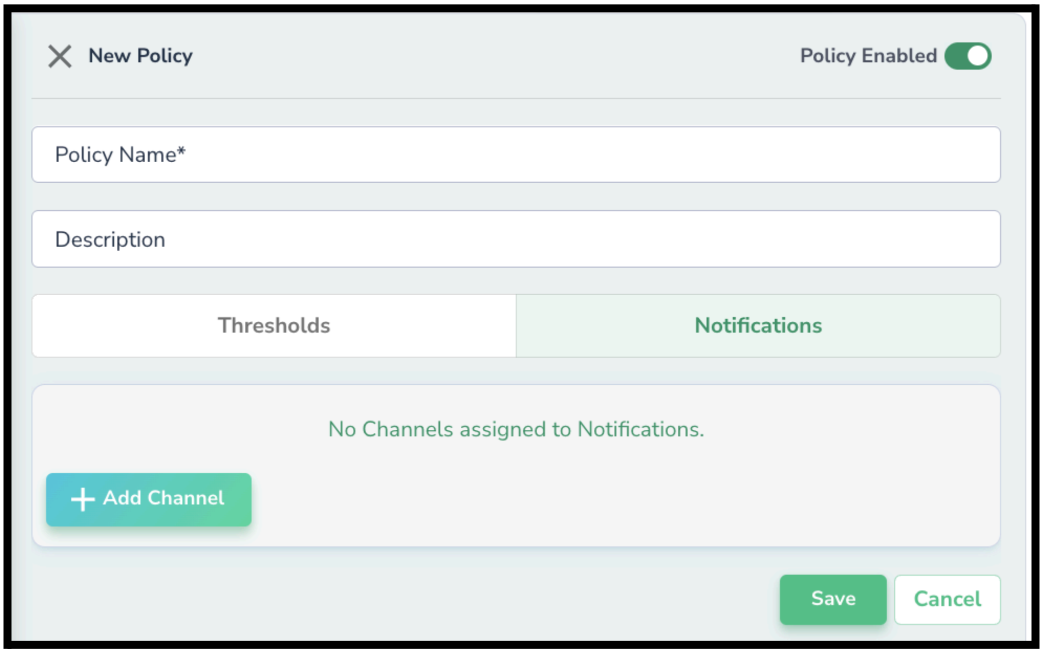
1. Navigate to "Alerting Policies" page & chose the "Policies" Tab
2. Click ⚠ New Policy to start defining your policy
3. A side-menu will appear similar to this image



4. Select desired "Tracked Metric" from list of
   - User defined metrics, or
   - Health metric calculated by Telmai
5. Based on Metric selection, user may need to select tracked attributes (ex: correctness, user will need to specify associated attributes)
6. Select detection type
   Note: In case of Relative and Range detections, user can specify how to handle the metric when not available (replace by zero or average)

<Only follow below steps if you want to be notified when an alert is generated>
Click Next to set notification destinations

7. Navigate to Notifications tab within the New Policy page



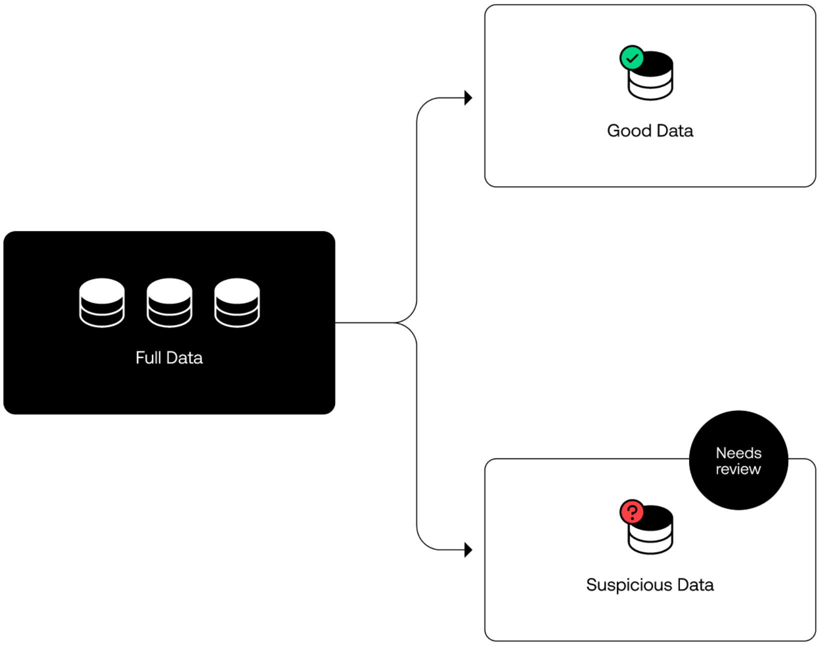
No Channels assigned to Notifications.

8. Click +Add Channel
9. List of defined channels will appear
10. Repeat step 2&3 for all desired channels
11. After providing the required info, click Save

After setting up notifications channels for your policy, you will start getting alert notifications via the desired channel.

Data Binning Policy

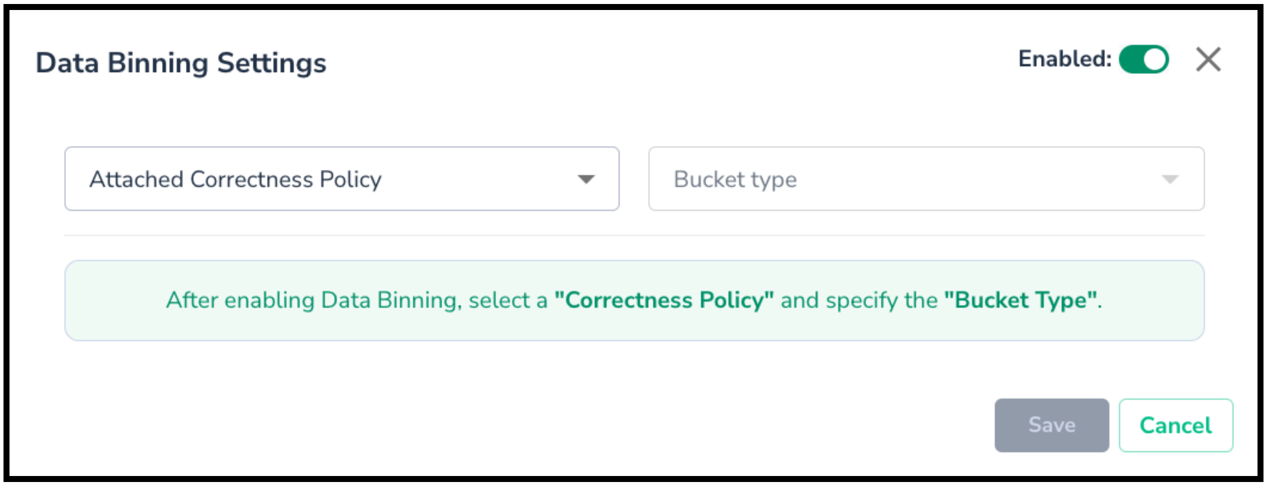
To ensure handling data quality issues at the source, Telmai allows you to utilize a feature called Data Binning. This feature allows defining a policy where Telmai monitors your data correctness and splits your data into good and bad. Good data can continue to be used within your pipeline, but bad or suspicious data can be reviewed and accessed.



This feature can help you make sure only good (or expected) data is flowing into your ecosystem. By doing so, you can ensure your costly pipelines are only running on healthy datasets.

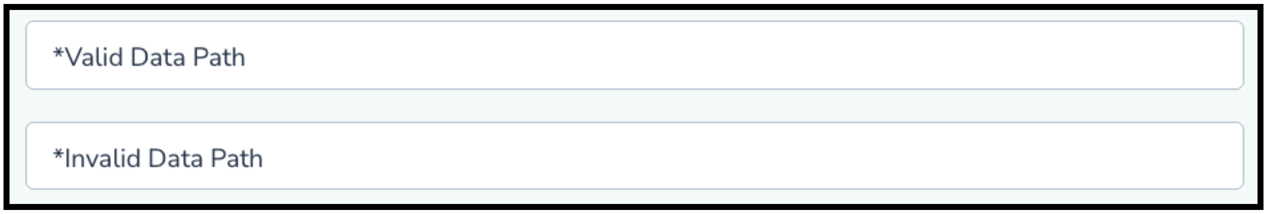
To enable this feature for a connected data source, you will need to:

1. Configure ID attribute for the data source
2. Set data expectation rules in Correctness Rules page
3. Navigate to the Alert page, you will need to create a correctness policy. This policy will only be used for scoping
    a. You are now able to set your Data Binning policy
4. Click Enable Data Binning, a prompt will ask you to define the policy details:



After enabling Data Binning, select a **"Correctness Policy"** and specify the **"Bucket Type"**.

5. Select previously created correctness policy
6. Pick desired bucket type (AWS-S3, GCP-Storage or Azure-Blob)
    - Once selected, you will need to enter the credentials
7. You will then need to define:
    - *Valid Data Path:* Path for good data (correct data)
    - *Invalid Data Path:* Path for bad data (incorrect data)



This binning will automatically take effect in your next data processing job.

Note: Data binning will write the data with the following partitioning:

- __telm_scan_date: Date the scan has run
- __telm_jobId: Job Id for associated scan

Reviewing Health KPIs and Alerts

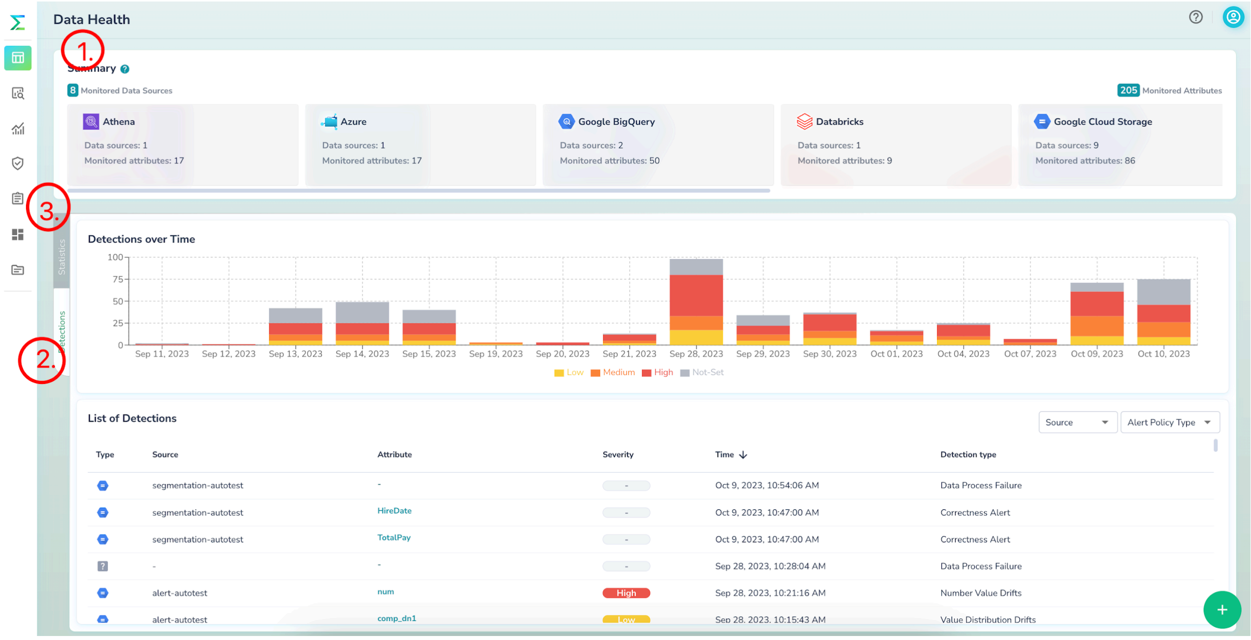
There are 3 levels where you can learn about your data health and alerts:

1. Data Health page
2. Trends page
3. Investigator page

Data Health Page

To monitor the health of your data on an ongoing basis, Telmai provides you with a business dashboard about all your data sources and their respective health metrics. This dashboard consists of three sections:

1. Summary Section
2. Detections Overtime
3. DQ KPIs Statistics



Data Health

Summary

8 Monitored Data Sources

205 Monitored Attributes

Athena	Azure	Google BigQuery	Databricks	Google Cloud Storage
Data sources: 1	Data sources: 1	Data sources: 2	Data sources: 1	Data sources: 9
Monitored attributes: 17	Monitored attributes: 17	Monitored attributes: 50	Monitored attributes: 9	Monitored attributes: 86

Detections over Time

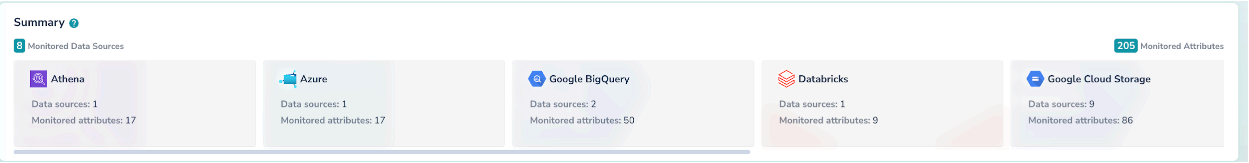
List of Detections

Type	Source	Attribute	Severity	Time ↓	Detection type
	segmentation-autotest	-	-	Oct 9, 2023, 10:54:06 AM	Data Process Failure
	segmentation-autotest	HireDate	-	Oct 9, 2023, 10:47:00 AM	Correctness Alert
	segmentation-autotest	TotalPay	-	Oct 9, 2023, 10:47:00 AM	Correctness Alert
?	-	-	-	Sep 28, 2023, 10:28:04 AM	Data Process Failure
	alert-autotest	num	High	Sep 28, 2023, 10:21:16 AM	Number Value Drifts
	alert-autotest	comp_dn1	Low	Sep 28, 2023, 10:15:43 AM	Value Distribution Drifts

Here're the details of what you can find on this page.

Summary Section

This section summarized the datasets that are connected to Telmai. You will see:



Summary

8 Monitored Data Sources

205 Monitored Attributes

Athena	Azure	Google BigQuery	Databricks	Google Cloud Storage
Data sources: 1	Data sources: 1	Data sources: 2	Data sources: 1	Data sources: 9
Monitored attributes: 17	Monitored attributes: 17	Monitored attributes: 50	Monitored attributes: 9	Monitored attributes: 86

1. Total count of data source(s).
2. Total count of monitored attribute(s).
3. A tile for each data source type currently connected. Each tile will have:
    1. Count of datasets monitored.
    2. Count of attributes monitored.

Detections Overtime

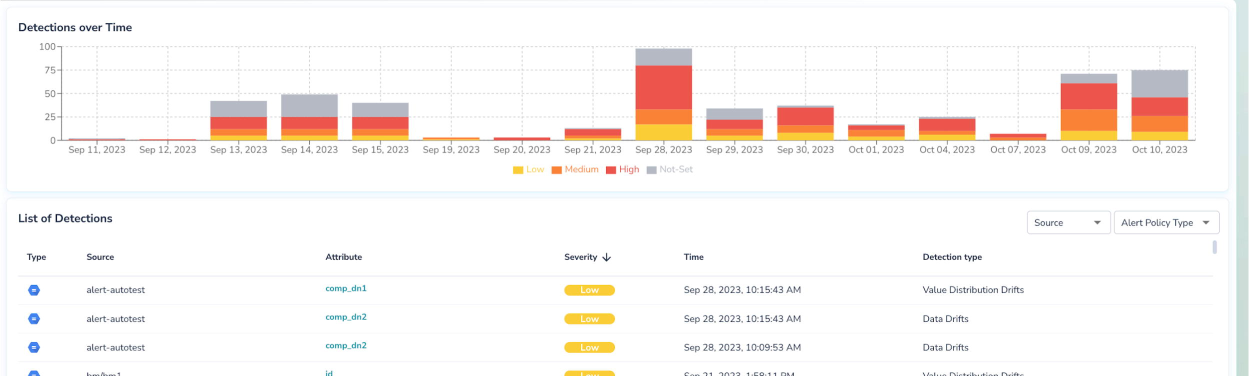
This component allows you to visualize detections (or alerts) that were identified over time across all your data sources. The detections are grouped by their severity (High, Med, Low).

This is done by showing:

1. A graph that summarizes the count of alerts every day by severity.
2. A table that lists the top 100 alerts based on current selection.

Clicking anywhere in the graph or the table selector will allow exploring more of these alert:

1. You can click on the graph to see alerts on a given time.
2. You can use the dropdown selector to look at a specific data source.
3. You can use the dropdown selector to look at a specific alert.
4. You can use a combination of these filters as you will.
5. Finally, clicking on any attribute alert, will navigate you to the trends section to get more details on the corresponding alert.



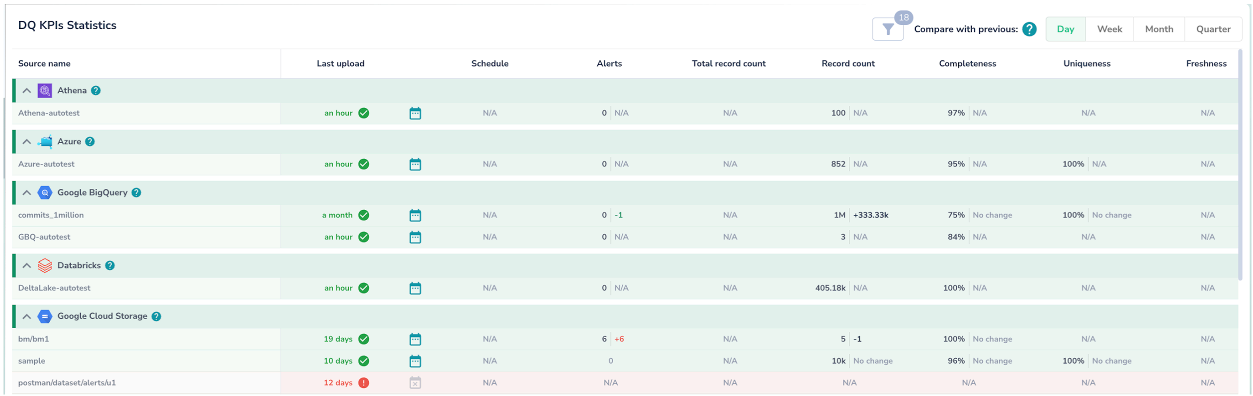
Detections over Time

List of Detections

Type	Source	Attribute	Severity ↓	Time	Detection type
⊜	alert-autotest	comp_dn1	Low	Sep 28, 2023, 10:15:43 AM	Value Distribution Drifts
⊜	alert-autotest	comp_dn2	Low	Sep 28, 2023, 10:15:43 AM	Data Drifts
⊜	alert-autotest	comp_dn2	Low	Sep 28, 2023, 10:09:53 AM	Data Drifts
⊜	bm/bm1	id	Low	Sep 21, 2023, 1:58:11 PM	Value Distribution Drifts

DQ KPIs Statistics

This component summarized the data scan status, as well as, health KPIs across all datasets.



DQ KPIs Statistics

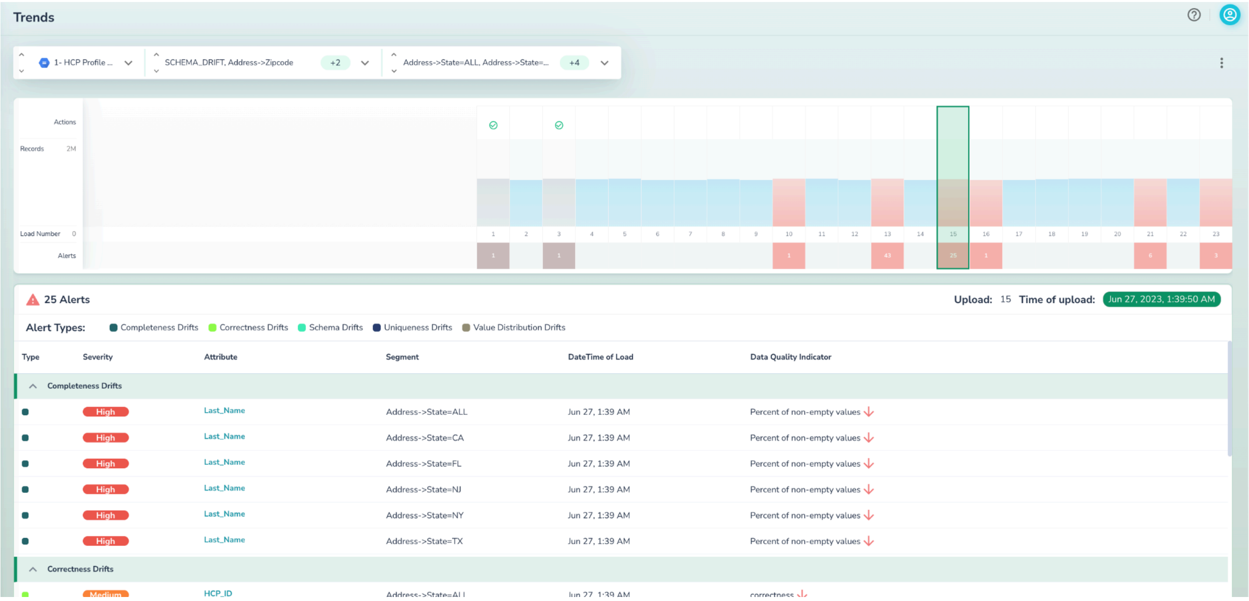
Source name	Last upload	Schedule	Alerts	Total record count	Record count	Completeness	Uniqueness	Freshness
Athena ?								
Athena-autotest	an hour ✓	N/A	0 N/A	N/A	100 N/A	97% N/A	N/A	N/A
Azure ?								
Azure-autotest	an hour ✓	N/A	0 N/A	N/A	852 N/A	95% N/A	100% N/A	N/A
Google BigQuery ?								
commits_1million	a month ✓	N/A	0 -1	N/A	1M +333.33k	75% No change	100% No change	N/A
GBQ-autotest	an hour ✓	N/A	0 N/A	N/A	3 N/A	84% N/A	N/A	N/A
Databricks ?								
DeltaLake-autotest	an hour ✓	N/A	0 N/A	N/A	405.18k N/A	100% N/A	N/A	N/A
Google Cloud Storage ?								
bm/bm1	19 days ✓	N/A	6 +6	N/A	5 -1	100% No change	N/A	N/A
sample	10 days ✓	N/A	0	N/A	10k No change	96% No change	100% No change	N/A
postman/dataset/alerts/u1	12 days !	N/A	N/A	N/A	N/A	N/A	N/A	N/A

In this table you will be able to see:

- Last scan: Time since last scan with status green check mark that means success, and red alert icon for processing failure.
- Schedule: Data scan schedule if setup
- Alerts: Number of alerts in latest data scan
- Health KPIs: please refer to Calculated Health Metrics for more details

Data Trends Page

This page allows you to explore detections (alerts) within a given dataset:



25 Alerts

Upload: 15 Time of upload: Jun 27, 2023, 1:39:50 AM

Alert Types: ● Completeness Drifts ● Correctness Drifts ● Schema Drifts ● Uniqueness Drifts ● Value Distribution Drifts

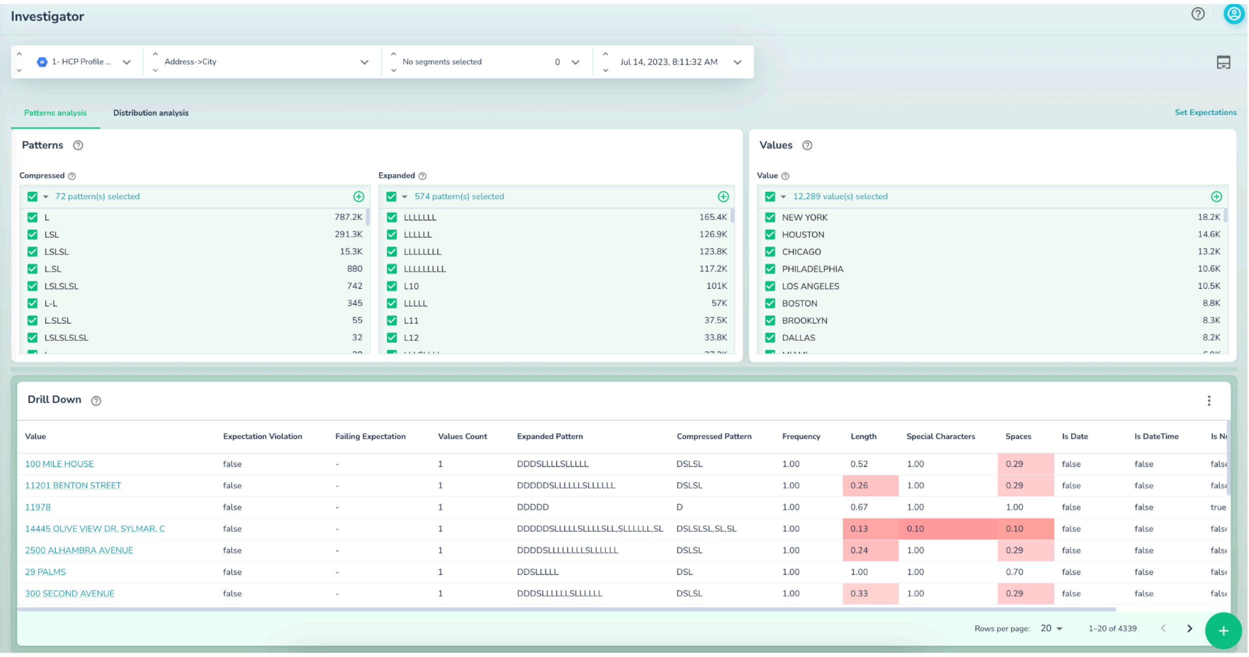
Type	Severity	Attribute	Segment	DateTime of Load	Data Quality Indicator
Completeness Drifts					
●	High	Last_Name	Address->State=ALL	Jun 27, 1:39 AM	Percent of non-empty values ↓
●	High	Last_Name	Address->State=CA	Jun 27, 1:39 AM	Percent of non-empty values ↓
●	High	Last_Name	Address->State=FL	Jun 27, 1:39 AM	Percent of non-empty values ↓
●	High	Last_Name	Address->State=NJ	Jun 27, 1:39 AM	Percent of non-empty values ↓
●	High	Last_Name	Address->State=NY	Jun 27, 1:39 AM	Percent of non-empty values ↓
●	High	Last_Name	Address->State=TX	Jun 27, 1:39 AM	Percent of non-empty values ↓
Correctness Drifts					
●	Medium	HCP_ID	Address->State=ALL	Jun 27, 1:39 AM	correctness ↓

This page has the following sections:

- **Data Scan & Alert Type Selector**
  - This selector allows you to pick
    - Dataset to be checked
    - Filter on alert types (if exists)
    - Filter on segments (if exists)
- **Data Scans Stats Graph**
  - This graph allows you to:
    - See record count per scan
    - Number of alerts per scan
    - Exclude scan from analysis to not be used in Telmai's learning about your dataset
  - Each bar is clickable to look at graphs within the corresponding scan
- **Alerts**
  - List of alerts within the corresponding scan
  - If applicable, each alert will have a clickable component that visualizes the detection
- **Tickets creation**
  - Telmai now allows you to create Jira tickets for existing alerts. This is defined in the Jira integration section

Data Investigator Page

Another page where you can drill down your dataset for profiling and investigating data violation is the Investigator page:



Investigator

1- HCP Profile ... | Address->City | No segments selected 0 | Jul 14, 2023, 8:11:32 AM

Patterns

Compressed

☑ 72 pattern(s) selected

	Pattern	Count
☑	L	787.2K
☑	LSL	291.3K
☑	LSLSL	15.3K
☑	L.SL	880
☑	LSLSLSL	742
☑	L-L	345
☑	L.SLSL	55
☑	LSLSLSLSL	32

Expanded

☑ 574 pattern(s) selected

	Pattern	Count
☑	LLLLLLL	165.4K
☑	LLLLLL	126.9K
☑	LLLLLLLL	123.8K
☑	LLLLLLLLL	117.2K
☑	L10	101K
☑	LLLLL	57K
☑	L11	37.5K
☑	L12	33.8K

Values

Value

☑ 12,289 value(s) selected

	Value	Count
☑	NEW YORK	18.2K
☑	HOUSTON	14.6K
☑	CHICAGO	13.2K
☑	PHILADELPHIA	10.6K
☑	LOS ANGELES	10.5K
☑	BOSTON	8.8K
☑	BROOKLYN	8.3K
☑	DALLAS	8.2K

Drill Down

Value	Expectation Violation	Failing Expectation	Values Count	Expanded Pattern	Compressed Pattern	Frequency	Length	Special Characters	Spaces	Is Date	Is DateTime	Is N
100 MILE HOUSE	false	-	1	DDDSLLLLSLLLLL	DSLSL	1.00	0.52	1.00	0.29	false	false	false
11201 BENTON STREET	false	-	1	DDDDDSLLLLLLSLLLLLL	DSLSL	1.00	0.26	1.00	0.29	false	false	false
11978	false	-	1	DDDDD	D	1.00	0.67	1.00	1.00	false	false	true
14445 OLIVE VIEW DR, SYLMAR, C	false	-	1	DDDDDSLLLLLSLLLLSLL,SLLLLLL,SL	DSLSLSL,SL,SL	1.00	0.13	0.10	0.10	false	false	false
2500 ALHAMBRA AVENUE	false	-	1	DDDDSLLLLLLLLSLLLLLL	DSLSL	1.00	0.24	1.00	0.29	false	false	false
29 PALMS	false	-	1	DDSLLLLL	DSL	1.00	1.00	1.00	0.70	false	false	false
300 SECOND AVENUE	false	-	1	DDDSLLLLLLSLLLLLL	DSLSL	1.00	0.33	1.00	0.29	false	false	false

Rows per page: 20 1–20 of 4339

Similar to the overview page, this page has a similar Selector Component.
Other components can be seen in this page:

Patterns

This section shows the currently selected data scan's data distribution in terms of compressed and expanded patterns. This can be used to understand the data, as well as, build more data restrictions.

Values

Sample of the datasets' top values.

Drill Down

Sample of the datasets' top values that have similar properties. Clicking on any of the values in this table shows your more details on the value properties.
This table can also be used to understand data correctness policy violations by looking at the "Expectation Violation" column.

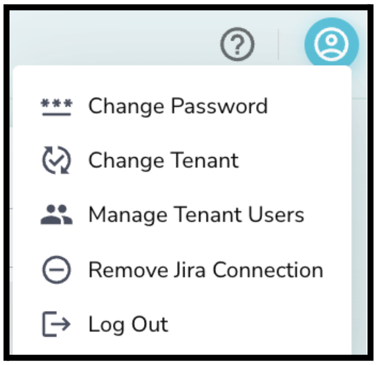
RBAC Support

Telmai supports project-scoped permissions. Tenant admins are able to modify these permissions accordingly:

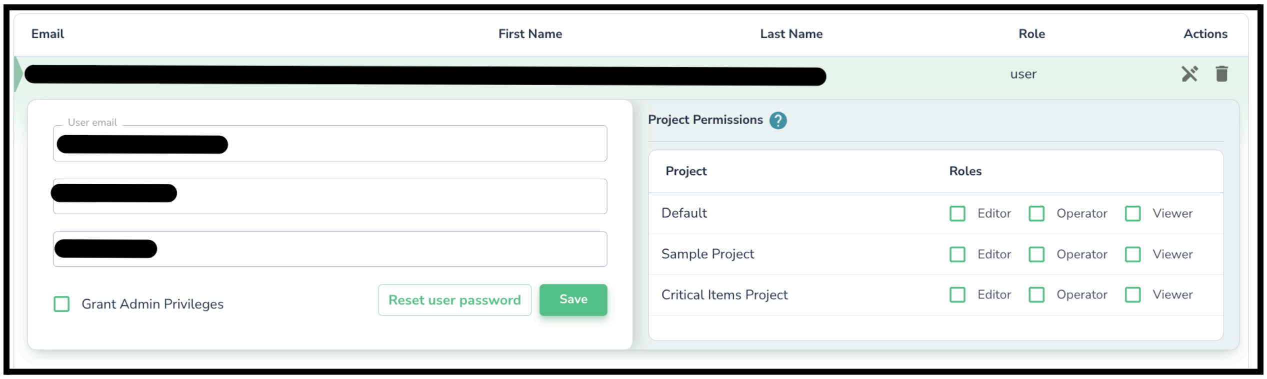
Role	Add/Modify Users	Add, Edit or Delete Source	Scan source/ Schedule Scans	View scan results
Tenant Admin	x	x	x	x
Editor		x	x	x
Operator			x	x
Viewer				x

To modify user roles,

1. Click “Manage Tenant Users” under user menu



2. Click on the user you like to modify permissions for
3. “Project Permissions” table with different roles



Email	First Name	Last Name	Role	Actions
[redacted]			user	

User email
[redacted]
[redacted]
[redacted]

□ Grant Admin Privileges

Project Permissions

Project	Roles
Default	□ Editor □ Operator □ Viewer
Sample Project	□ Editor □ Operator □ Viewer
Critical Items Project	□ Editor □ Operator □ Viewer

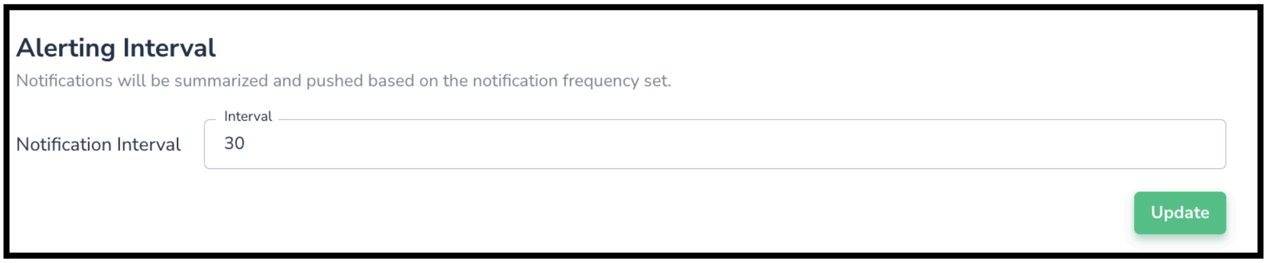
4. Select appropriate roles
5. Click Save

Notification Settings

Telmai allows you to set different channels that get notified when an alert is generated. Notifications are summarized per “notification interval”, and are sent to the configured destination.

Notification Interval

Notification Interval is a setting to limit how often notifications are sent. Example: if the notification interval is set to 30 minutes, you will not get more than one notification sent every 30 minutes.
To modify this interval, you can navigate to the Administration page, and select the “Notification Interval” setting. The setting can be set with a minimum value of 1 minute and maximum for 300 minutes.



Alerting Interval

Notifications will be summarized and pushed based on the notification frequency set.

Notification Interval — Interval: 30

Notification Info

Once a notification is sent, it will include the following details:

- Tenant where alerts are generated
- Number of sources with alerts
- Alert count by priority
- A link to the full report in Telmai

Adding Notification Destination

To add a new notification destination, you will need to:

1. Navigate to Alerting Policies page
2. Click Manage Alert Channels button
3. Click +Alert channel
   Note: A channel can have multiple receivers
4. Click “Save”

Note: At any time you can test the channel by clicking the “Test” or play buttons

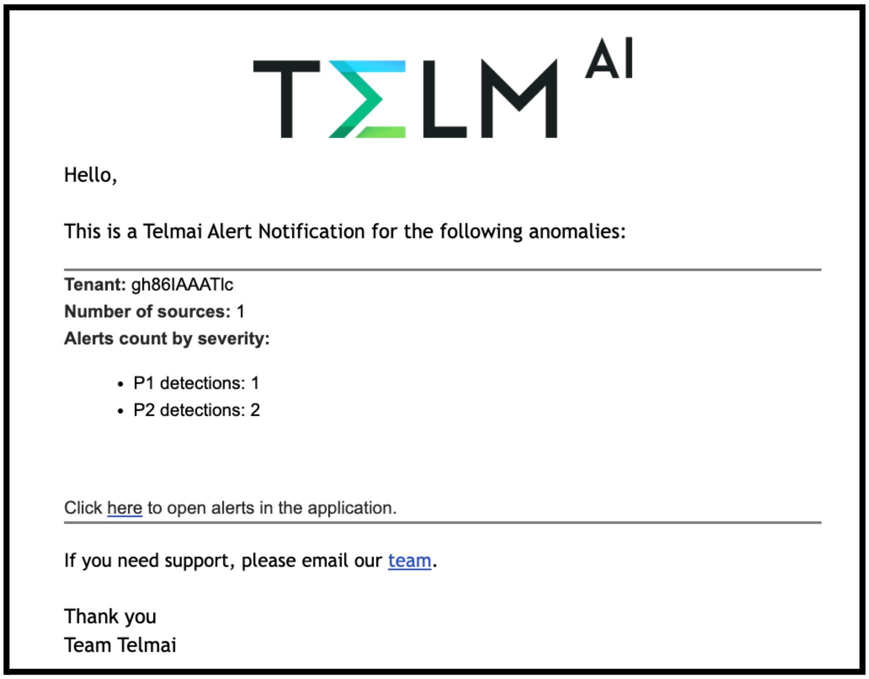
Supported Notification Destinations

Currently the following destinations are supported:

- Slack
- Email
- Microsoft Teams

Sample Notifications

Email Sample



TELM AI

Hello,

This is a Telmai Alert Notification for the following anomalies:

Tenant: gh86IAAATlc
Number of sources: 1
Alerts count by severity:

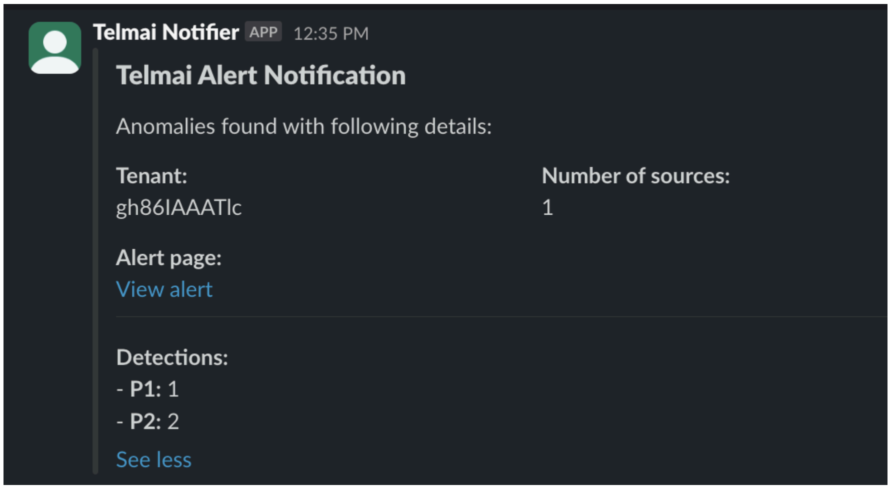
- P1 detections: 1
- P2 detections: 2

Click here to open alerts in the application.

If you need support, please email our team.

Thank you
Team Telmai

Slack Sample



Telmai Notifier APP 12:35 PM

Telmai Alert Notification

Anomalies found with following details:

Tenant:
gh86IAAATlc

Number of sources:
1

Alert page:
View alert

Detections:
- **P1:** 1
- **P2:** 2

See less

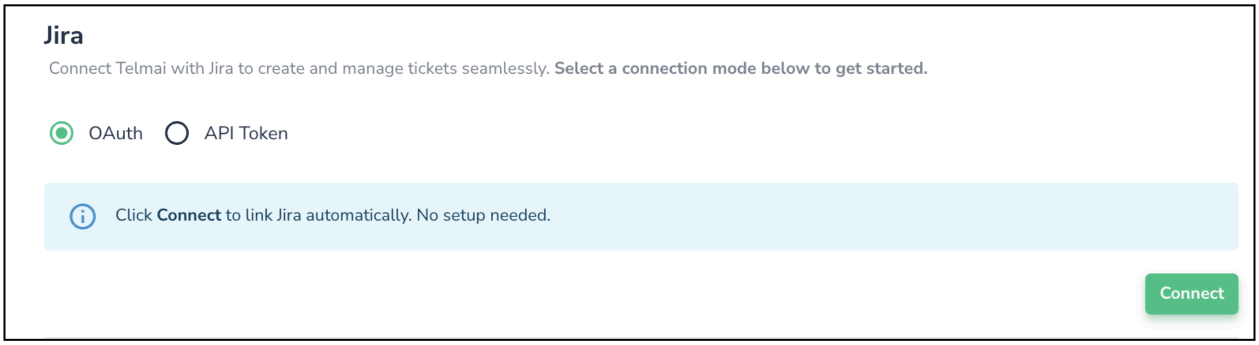
Jira Integration

You can now integrate your Jira instance with Telmai. This will simplify and ease issues creation and tracking. Simply connect to your Jira instance and create tickets from generated alerts.

Connect to Jira

To integrate your Jira instance:

1. Navigate to administration menu from the left tab menu (you will need to be admin to perform this action)
2. You will see setting for Jira as follows



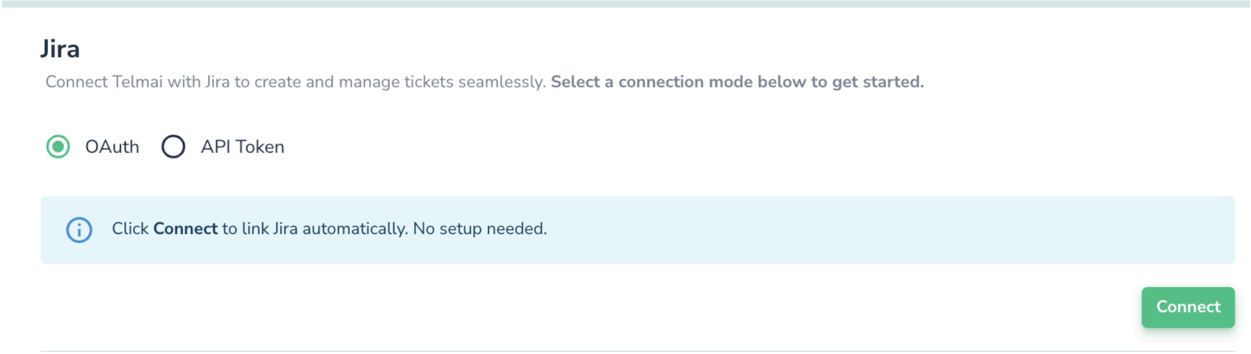
Jira

Connect Telmai with Jira to create and manage tickets seamlessly. **Select a connection mode below to get started.**

● OAuth ○ API Token

Click **Connect** to link Jira automatically. No setup needed.

3. Select how you like to connect to Jira
    a. OAuth (only supported in Telmai PaaS instance)
    b. API token
4. Provide the necessary information as prompted
5. You have now connected to your Jira instance
   Note: you can always come back to admin page to modify the Jira connection settings.



Jira

Connect Telmai with Jira to create and manage tickets seamlessly. **Select a connection mode below to get started.**

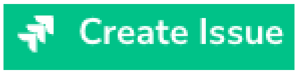
● OAuth ○ API Token

Click **Connect** to link Jira automatically. No setup needed.

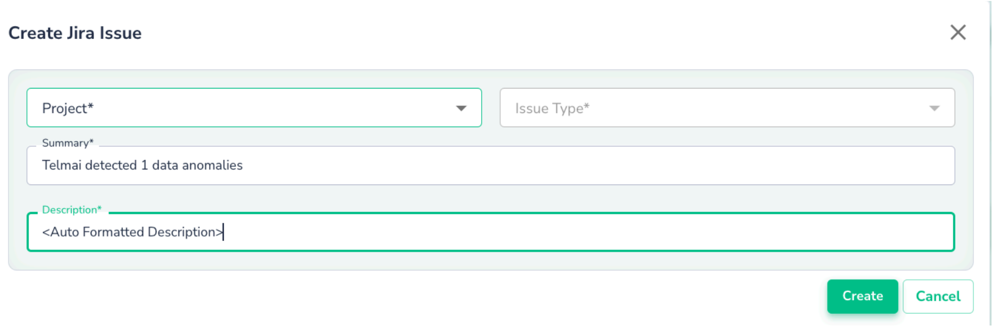
Create Jira Tickets

To create Jira tickets:

1. Navigate to “Trends & Alerts” page
2. Chose a scan that has alerts

3. Click  button

4. A modal will appear to request more details from you



Create Jira Issue

Project*

Issue Type*

Summary*
Telmai detected 1 data anomalies

Description*
<Auto Formatted Description>

5. Specify the “project” and “issue type”
6. Telmai will auto fill the “summary” and “description” but you can manually change them
7. Click “Create”

View & Manage Created Ticket

After creating Jira tickets, you are able to see the created tickets & unlink them from Telmai. To do so,

1. Navigate to “Trends & Alerts” page
2. Click “See Issues” button
3. Telmai will display the list of open tickets through Telmai
4. Options:
    a. Click the ticket to navigate to your Jira instance
    b. Unlink the ticket from Telmai
    c. View ticket status

Pipeline Integration Examples

Using Telmai's services and APIs, you can start building smarter ETL pipelines that mitigate known issues.

Data Binning for Suspicion Data

Scenario 1: Bad version data

Data: Let's consider a scenario where we are tracking the user count of an app that's available.
Situation: A bug in the latest release version of the app caused the user count to double. App version has been rolled back, but some users are still reporting from the rogue version.

How to use data binning?

1. User defines a correctness rule. This correctness rules sets an expectation that excludes the rogue app version.
2. The user sets up the data binning policy to use the created correctness rule

Result:

- Good data will keep flowing into the good data bucket
- "Suspicious" data is not isolated in a separate bucket, and can be cleaned up before using it in business reports

The capability to automate this process of segregating 'good' and 'suspicious' data in real-time significantly reduces the manual effort and time required to ensure data quality, thereby accelerating the time-to-value in your operations.

API Integrations

Using Telmai's APIs, users can query information about their data health. That is querying any of the health KPIs or checking for ongoing policy violations.

One use case of this is circuit breaker.

Circuit Breaker

Circuit breaker is a case where you want to switch off the ETL operation because bad data has been introduced into your system. This could be because processing costs are very high or because of other factors.
Here're the steps you can follow to create the circuit breaker:

1. Trigger Telmai data scan (api reference)
2. Check for Telmai job completion
    a. Polling mechanism (api reference)
    b. Using webhooks for job notification
3. Check for alerts (api reference)
4. Decide best action

These APIs can be integrated in ETL orchestration tools like Airflow or other.