AUTOMATING THE COLLECTION OF DATA
# Table of Contents

**Understanding Data Collection Triggers** ........................................................................................................3
  - What are Data Collection Triggers? .............................................................................................................. 4
  - What options are available for triggering Data Collection? ....................................................................... 5

**Setting Up Data Collection Triggers** ........................................................................................................... 6
  - Create a Data Collection Trigger (Scheduled / Externally Triggered)......................................................... 7
  - Configure Data Dependencies for a Data Collection Trigger ...................................................................... 11
  - Establish dependencies between Data Collection Triggers ...................................................................... 16
  - Recollect Data for a Data Collection Trigger ............................................................................................ 18

**Setting Up Data Collection Notifications** .................................................................................................. 23
  - How to receive a Notification if Data Collection does not start ............................................................... 24
  - Request an email with errors encountered during Data Collection ......................................................... 26

**Optimizing the Performance of Data Collection** ............................................................................................ 29
  - Collect Metric Data Incrementally ........................................................................................................ 30
Understanding Data Collection Triggers
What are Data Collection Triggers?

Data collection Triggers are used to control when content should be updated within Metric Insights. The following types of objects within Metric Insights can be populated through a data collection trigger:

1. Metrics and Reports sourced from a Data Source
2. External report images (for example, Tableau Dashboard Images)
3. Dynamically populated Calendar Events
4. Dynamically populated Dimension Values

Where are Data Collection Triggers Referenced

For dynamically populated content, the associated data collection trigger can be found immediately before the fetch command in the relevant editor.
What options are available for triggering Data Collection?

Data Collection can be initiated within Metric Insights via an external process or based upon a schedule.

Additional dependencies can be established based on:

- Other data collection events (e.g., run the weekly metric collection once the daily collection has completed for the day)
- Data dependencies (e.g., run the daily metric data collection only once there is a full day of data in the warehouse for the prior day)

Common Scenarios

The following common requirements would be satisfied with data collection trigger setup as follows:

1. **Trigger collection once ETL processing is complete**: Setup externally triggered dependency and include trigger URL as last step in ETL processing.
2. **Trigger collection when a full day of data exists in my sales table**: Setup schedule-based data collection trigger with a data dependency that checks that sales transactions for the prior day have been successfully collected at any time since the previous Trigger run.
3. **Trigger collection when a full day of data exists in my sales table processed during the current collection cycle**: Setup schedule-based data collection trigger with a data dependency that checks that sales transactions for the prior day have been successfully loaded within the current collection cycle of the Trigger's run.
4. **Trigger collection when ETL processing flags completion in a control table**: Setup scheduled data collection trigger with a data dependency that verifies that data in an ETL control table has been updated.
Setting Up Data Collection
Triggers
Create a Data Collection Trigger (Scheduled / Externally Triggered)

This article outlines how to create a new data collection trigger to run based on a pre-specified schedule. The same process can be used when defining existing collection triggers.

1. Access Admin menu > Data Collection Triggers

At the bottom of the page click [New Data Collection Trigger]
2. Define the basics of a new trigger

Add Data Collection Trigger

1. Data Collection Scope: This trigger can be applied either to elements (Reports, Metrics, etc.) or Dimension Values. This choice can be made at this step only and once saved cannot be altered.

2. Name: should be unique and descriptive. NOTE: In versions prior to 4.1 trigger names could not contain spaces, caps and other characters.

3. Data collection based on:
   - scheduled
   - external Trigger

The Data Collection Trigger pop-up opens.

1. **Data Collection Scope**: This trigger can be applied either to elements (Reports, Metrics, etc.) or Dimension Values. This choice can be made at this step only and once saved cannot be altered.

2. **Name**: should be unique and descriptive. **NOTE**: In versions prior to 4.1 trigger names could not contain spaces, caps and other characters.

3. **Data collection based on**:
   - **scheduled**
   - **external Trigger**
1. Provide the Interval and unit that will determine how often the trigger is executed.
2. Specify whether the trigger should run as early as possible in the period (day, week, etc..) or if it should only be run after a specific time.

[Option 2] Data collection initiated by External Trigger
If you choose to perform data collection upon the External Trigger, a URL is to be displayed at the field below. You can use this URL with other integrations.

3. Save the Changes

Note that the trigger will not automatically start at the pre-specified schedule if there are Trigger Dependencies or Data Dependencies which have not been satisfied.

A schedule-based trigger will be started after the specified time ONLY once all associated dependencies have been satisfied.

For more information see: Establish dependencies between Data Collection Triggers

What would you like to do next?

- Create a Report Undimensioned / Dimensioned with this Data Collection Trigger

NOTE: The same logic applies to automatic update of Dimensioned / Undimensioned Metrics
Configure Data Dependencies for a Data Collection Trigger

Data Dependencies are typically associated with a Data Collection Trigger in order to ensure that data is available and current in the source system through the collection cycle specified in that Dependency.

PREREQUISITES:

• An enabled Data Collection Trigger

Examples of Data Dependency Checks:

1. Check that data in a Warehouse Customer Order table contains records post midnight of the prior day (which indicates that all of the prior day's orders have been loaded).
2. Verify that an ETL control table specifies that ETL processing has been completed since the last data collection cycle.
3. Verify that volume in a data table is at least X% of the volume of data in a prior day as a data qualify check before loading data.

Video Tutorial
1. Add a New Data Dependency

1. Click [+ New Data Dependency] in the Data Collection Trigger Editor Dependencies tab
2. The Add Data Dependency pop-up opens
3. From the drop-down list select the Add New Data Dependency option.

2. Define the basics for the New Data Dependency

Add Data Dependency

Name: Full Day of Customer Order Data has been loaded
Description: Verify that we have customer orders through the current day

Save or cancel

After the new Data Dependency is created, the Data Dependency Editor opens.
3. Enter Data Dependency Specifications

1. Make sure your **Data Dependency** is 'enabled'
2. Specify the **Data Source** for the Data Dependency
3. Provide a fetch command that will return a positive number when the dependency is satisfied (you can use `:last_dependency_satisfied` as a bind variable)
4. **Validate** your fetch command

### 3.1. Specify Last Dependency substitution

**Parameters Substitution**

- **Last Dependency satisfied**: 2016-02-22 06:26:45

If `:last_dependency_satisfied` was included in the fetch command, provide a value for testing the data dependency logic
4. Below the statement specify when and how often to perform the Data Dependency check

1. Specify how often the dependency check fetch command should be run.
2. Provide a time window during which checks should be performed.
3. Specify a time period after which the dependency check should be expired.

Enable Dependency at the top right corner of the screen.

This Data Dependency can now be added to the Data Collection Trigger

NOTE:

- A given dependency can be shared among multiple Data Collection Triggers
- Multiple dependencies can be specified for a Data Collection Trigger and all associated dependencies must be satisfied before the data collection trigger starts
Control when Data Dependency of Data Collection Trigger is Satisfied

If your Data Collection Trigger's interval is an hour or longer, you can control the circumstances under which an associated Data Dependency must be satisfied by accessing the Data Collection Trigger Editor and editing an existing dependency.
Establish dependencies between Data Collection Triggers

It is sometimes necessary to set the completion of one data collection trigger as a dependency for starting a second trigger.

Typical situations where this functionality is required are:

1. You want to ensure that weekly and monthly data collection triggers start after the daily collection trigger finishes so that all the daily data is available for summaries that are based on collected data in Metric Insights.
2. You want to define a daily event which contains composite reports and you want to ensure that the data for the source reports has already been collected so you make completion of the core reporting event a dependency to the composite reporting data collection event.

Add a New Data Collection Trigger Dependency

1. Click [+ New Trigger Dependency]
2. The Add Trigger Dependency pop-up opens
3. Choose a dependency from the drop-down list

Save your entry.
Specify the Trigger Dependency that must be satisfied

1. Select what Data Collection Trigger must complete before the current event can start
2. Save your changes

Note that you can add multiple dependencies to a data collection event and that ALL dependencies must be satisfied before the event starts.
Recollect Data for a Data Collection Trigger

Sometimes historical data in a source system is re-stated or corrected. This means your Metrics and Reports sourced from this data will need to be re-run to reflect these changes. One simple way to do this in Metric Insights is to manually re-run a data collection trigger using the Refetch historical data feature.

**ALTERNATIVE METHODS:**

If you find your historical data is often changing and would rather have your elements automatically refresh their historical instance, use one of these methods:

- the ':last_measurement_time' bind variable for metrics
- the ':measurement_time' bind variable for report

**To recollect data for just a single Report.**

**To recollect data for just a single Metric.**

1. Access a Data Collection Trigger Editor

   Access the **Data Collection Trigger Editor** through your **Admin** menu and page down to the bottom of the Editor:

   1. Click **Refetch historical data** to recollected historical data for the elements associated with this Trigger
   2. The **Trigger now** button is used to run this trigger now and collect only current data
2. Specify data to be collected and for which Elements

Delete and Refetch Historical Data

Delete Data Collected Since: 2015-10-12 17:32:00

Regenerate Charts following Data Deletion

Select or cancel

1. **Delete Data Collected Since**: Enter the date of how far back the process should start. For Hourly and Minute data collection intervals, also enter time. All data from the specified date/time will be removed and replaced by newly collected data.

2. Click **Select** to initiate the process of data collection for the selected elements.

2.1. Review the data to be collected

Clear Collected Values

Delete Data Since 2015-03-09 05:00:00

**Metrics**

The following metrics will have the indicated number of historical metric measurement values deleted and refetched:

- Hourly Sales: 25
- bk Hourly Sales: 25

**Reports**

The following reports can NOT be backfilled and will not be affected:

- Hourly users (markdown) 25

**OK** or cancel

A **Clear Collected Values** pop-up is displayed showing:

- The list of elements affected
- How many Metric values and/or Report instances will be deleted and re-collected

Once you click **OK** to confirm, the system displays a message while deleting the data that will be re-stated.
2.2. Review the rules to be applied

Data collection for this Trigger has been queued up for execution. Affected report and metric charts will continue to display existing data until the data collection process is complete. Upon completion, the data from the specified starting date/time forward will have been deleted and new data collected.

The following rules are applied to the affected Elements:
- For snapshot reports all instances that have a measurement time after the re-fetch date/time will be re-stated.
- Non snapshot reports will be fully re-generated.
- Metrics that fetch data incrementally (last_measurement_time is included in the fetch command) will be re-stated for values collected after the re-fetch time.
- Metrics that do not fetch data incrementally will be fully re-stated.
- Multi-metric charts will be re-stated if one or more of the included Metrics have

Click OK

3. Review Data Collection Trigger Editors buttons and grids

Once you have clicked OK to proceed, the buttons change as follows:

1. **Abort Collection** button: While the Trigger is running, you can abort at any time; the collection in-process for an Element completes and no other Element fetches are started.
2. **Trigger Queued To Start** button: Changes to **Trigger Running Now** as soon as the first Element's fetch is in progress.
3. A banner showing "Data Collection Trigger Queued to Start" changes to a green with "Data Collection Trigger Running" once the trigger is executing.
The grid titled *All Elements in Data Collection Trigger* changes to *Current Running Data Collection* once it starts to run. The grid lists each Element’s fetch process as it begins and is updated as each collection starts and includes a *Completion Estimate* time.

### 4. Information on the current Run

To follow the progress of data collection, open the *Current Run* tab:

1. The banner at the top of the tab provides the status of data collection procedure and the time it started
2. All elements for which data is currently being collected are shown in the *Currently Running Data Collection* grid
3. Elements for which the procedure is complete are shown in the *Processing complete* grid
5. Review completion

1. Once the trigger completes, a corresponding message is shown at the banner at the top of the tab, displaying the status of data collection procedure, time when completed, number of elements processed and if there have been any errors
2. A grid of elements Completed during last run is displayed
3. The buttons at the top of the Editor return to their normal labels and state

6. Data Collection Triggers statuses

Statuses of all Triggers can be quickly reviewed at the Data Collection Triggers table, with each Trigger being highlighted with a corresponding color.
Setting Up Data Collection
Notifications
How to receive a Notification if Data Collection does not start

You can configure a Data Collection Event to send an email notification if the event does not start by a specified point in time. This is useful to notify users or system administrators that visualizations are not getting updated due to a dependency that is not satisfied.

This feature is only available for data collection events that fire daily or less frequently.

1. Access Data Collection Triggers via Admin menu > Data Collection Triggers

Select the required Data Collection Trigger from the list.

The Data Collection Trigger Editor opens.
2. Configure Notification sending options

1. Open the **Trigger Configuration** tab
2. Set **Send Email if data collection does not start?** field to ‘yes’
3. Specify the time that should trigger the notification. In the above example, a notification email will be sent if the event does not start by 02:15 AM.
4. Enter the email address to which the notification is to be sent.

**Save** your changes.

3. Example of an email you will receive if the data collection event does not start by the expected time

An email like the one shown above will be sent to the specified email address if the data collection event does not start by the expected time

1. Clicking on link with allow you to investigate the problem
Request an email with errors encountered during Data Collection

The Data Collection Trigger Editor has an option of sending notifications on any errors occurring during Data Collection. This setting usually defaults to 'no', so you must update the Editor to initiate the notification sending process. This feature allows you to take follow-up action if there are problems in fetching data.

To create a Data Collection Trigger, see Setting Up Data Collection Triggers.

NOTE: The format of Data Collection Trigger Editor is slightly different for Version 3.

1. Access Admin > Data Collection Triggers

The Data Collection Trigger Editor opens.

Select a trigger from the list.
2. Update the setting to request error email

1. Open the **Trigger Configuration** tab
2. Set **Email error report** field to 'yes'
3. Enter the **Email address** to which the notification is to be sent

Save your changes.
3. Example of an email you will receive if Data Collection Trigger Fails

An email like the one shown above will be sent to the specified email address if the Data Collection process finds any errors.

The following errors occurred while processing the `daily-metric-refresh` event which was completed on 2012-12-31 00:24:08:

- Single Day Sales by some Country
- Single Day Sales by some Country
- Single Day Sales by some Country
- Single Day Sales by some Country
- Total Daily Sales (1010)

Fetch command:
```plaintext
<tabu label="Sales by Day" breaks="date"><break col="date" sort="up"> <col source="sales" fun="sum" label="Sum of Sales"> <tabu col="cell value"/>
```
Optimizing the Performance of Data Collection
Collect Metric Data Incrementally

Data collection speed can be increased and load on source systems can be decreased by collecting metric data incrementally. Incremental data collection provides the ability to fetch either new or recent records from the source system and process only those records in Metric Insights (rather than the full historical dataset). This approach works well when data is only changing incrementally.

This approach can also be used when historical data is changing in the source system but you only need to worry about changes to recent history (for example, only changes in the last 7 days need to be processed and changes to older data can be ignored).

### 1. Using :last_measurement_time in fetch command to specify incremental data collection

![Metric Insights Interface](image)

- **Data Source**: Demo DB (SQL)
- **Data Collection Trigger**: 10-minute-refresh
- **SQL statement**: `select calendar_date, sum(total_amount) from daily_order_summary where calendar_date < :last_measurement_time group by 1`

In your Metric Data Fetch Command (**SQLStatement** text box above), include a `:last_measurement_time` bind variable which will be substituted at run-time with the last data point collected for the metric.
2. Re-fetching recent History for a Metric

If you want to re-fetch data going back a certain number of days before the last measurement time collected, you can do so by setting an offset in the Data Selection tab of the Metric Editor. This offset will be applied to the last measurement time associated with the metric to determine the :last_measurement_time substitution variable.