

Configuring Workbook Settings

You can configure workbook settings when you create a new workbook and you can edit them as needed, later. The system administrator can also edit the workbook settings.

- [Configuring Workbook Settings](#)
 - [Trigger options](#)
 - [Data driven options](#)
 - [Scheduled options](#)
 - [Data retention](#)
 - [Purge historical data](#)
 - [Never delete historical data](#)
 - [Export only](#)
 - [Description](#)
 - [Saving results and time based partitions](#)
 - [Error handling](#)
 - [Notifications](#)
 - [Advanced workbook settings](#)
 - [Run workbook after saving or updating the workbook](#)
- [Changing Workbook Settings](#)

Configuring Workbook Settings

On the **Workbook Settings** page, you can configure the following options.

Trigger options

The trigger settings specify how often the workbook data is calculated. It has the following options:

- **Manually.** You determine to run the workbook as needed. If a workbook is triggered manually and the workbook is currently running, the new run is queued to run after the current run has completed.
- **Data driven.** Calculates workbook when any specified datasource is updated. This suitable for a datasource with periodic updates. It isn't recommended if data changes constantly. If a workbook is triggered when data comes in, the new run begins running as soon as the new data is available.
- **Scheduled.** Workbooks run at a specified time. If a workbook is triggered by a schedule and the workbook is currently running, the new run is skipped.

Data driven options

- **Run when any of the selected data sources are updated.**
 - The entire workbook is recalculated any time selected data sources are updated.
- **Trigger the workbook to run when any of the selected data sources are updated only if the following conditions are true.**
 - Conditions can be created for each data source in the workbook that must be fulfilled before the workbook calculates.
 - Triggering condition measurements are performed after the data source set to run has completed. The triggering conditions are only measured when the the job (workbook data source) has a Complete or Complete with Warnings status.

For example, this workbook recalculates when the data source 1 or 2 update in Datameer unless Data_Source_1 has updated less than 2 days from today or Data_Source_2 has updated within the past 3 hours. The workbook doesn't run if Data_Source_3 updates.

▼ Trigger

Define Workbook Triggers:

Manual

Data Driven

Run when any of the following are updated:

Data_Source_1

Data_Source_2

Data_Source_3

If all of the following conditions are met:

Data_Sourc... is not older than TODAY()-2d —
+

Data_Sourc... is not older than NOW()-3h —
+

Scheduled

Using this mode triggers the workbook when any of the selected data sources are updated with new data, and all of the [optional] conditions about the age of the latest available data are true. Add conditions using formulas with fixed or relative timings, such as **TODAY()-5d**, **NOW()-3h**, or **ASDATE("18-07-2014 19:00:23")-5d**

Scheduled options

- Hourly through Monthly. Choose the interval suitable for your needs.
- Custom Cron Pattern. The format is a series of numbers separated by spaces to indicate the following: minute hour DayOfMonth Month DayOfWeek. Learn more about [Scheduling with Custom Cron Patterns](#).

For example, 0 9-15 * * * indicates that the task should launch once every hour between 9am to 3pm everyday.

▼ Trigger

Define Workbook Triggers:

Manual

Data Driven

Scheduled

Time Settings:

Hourly

Daily

Weekly

Monthly

Custom Cron Pattern

Cron Pattern:

0 9-15 * * *

The job runs every 0th minute past the 9, 10, 11, 12, 13, 14 and 15th hour

Set the schedule defining how often this workbook should be triggered.

Data retention

Options on how to store workbook data.

Purge historical data

Setting to delete previous results so as to free available resources.

▼ Data Retention

Store Historical Results:

Purge historical data

Keep last results

Purge results after they're generated

Never delete historical data

Export only (Never keep data)

If "store last n results" and "expire after" both contain values, at least n results are kept, even if one or more of them are expired.

The "Export only" option should only be selected if the user wants to save disk space in HDFS (This option does not maintain workbook data as a duplicate)

Never delete historical data

Any time a workbook runs, the previous data is stored until manually discarded.

Export only

This option is available for those that don't want to keep workbook calculations on HDFS. All data is deleted and is only able to be accessed from where it was exported.

▼ Data Retention

Store Historical Results:

Purge historical data

Never delete historical data

Export only (Never keep data)

• All associated data will be deleted from HDFS.

• "Show Full Data" will not be available for the exported workbook after the export has been completed.

If "store last n results" and "expire after" both contain values, at least n results are kept, even if one or more of them are expired.

The "Export only" option should only be selected if the user wants to save disk space in HDFS (This option does not maintain workbook data as a duplicate)

Description

Option to give a description for your workbook.

▼ Description

Path:
/Analytics/Workbooks/Site_Users_WB.wbk

Description:

Workbook contains site information and analysis on click-through data.

Saving results and time based partitions

Saving results specifies which workbook sheets are saved and visible. By saving only the sheets that you need you can improve performance and computation time. If you want to retain smart sampling on filtered sheets, make sure to save them. Smart sampling is disabled if a sheet isn't saved because it can slow down other processes.

The Time Based Partition column displays which sheets have a date type column. If not all partitions need to be saved, set specific date ranges to save.

Save Results and Time Based Partition

Save Results from:

Sheets:	Time Based Partition:	
<input type="checkbox"/> Data_Source_1		
<input type="checkbox"/> Data_Source_2		
<input type="checkbox"/> Data_Source_3		
<input type="checkbox"/> group_ranking	+	
<input type="checkbox"/> group_bookRanking	+	
<input checked="" type="checkbox"/> Joined	+	
<input checked="" type="checkbox"/> Sorted	+	
<input type="checkbox"/> sortRanked	+	
<input type="checkbox"/> concat_author_title	+	
<input checked="" type="checkbox"/> user_age	+	
<input checked="" type="checkbox"/> user_age_grouped	+	
<input checked="" type="checkbox"/> user_countries	+	
<input checked="" type="checkbox"/> user_countries_dist	+	

Check only the sheets which contain usable results. This saves on storage and can dramatically improve workbook computation time.

Error handling

Specifies what to do if Datameer encounters a row that possibly contains invalid or missing data.

You can choose to:

- Ignore. Ignores error calculations leaving the cell blank and continues processing the workbook.
- Skip. Skips the error and calculate rows with valid data. The invalid data is dropped. (You can view statistics on dropped data. See [Importing Data](#) to learn more.)
- Abort. Stops the calculation. This option stops processing the workbook.

Error Handling

Action:

Ignore - leaves the cell(s) blank, logs an error

Skip - drops the row and continues processing

Abort - cancels the entire workbook

Bad data or errors in worksheet formulas sometimes result in rows that cannot be calculated.

Max # of errors to log:

1000

The maximum numbers of errors to be logged.

Notifications

You can also specify who should be informed if there is a problem or when a workbook is successfully run. Separate multiple email addresses with a comma.

▼ Notifications

Please note
Datameer can only send email notifications if an email server has been configured.
Please contact your system administrator for assistance.

Error Email Notification:

support@datameer.com

Email addresses to be notified. Use commas to separate multiple addresses.

Successful Job Email Notification:

support@datameer.com

Email addresses to be notified. Use commas to separate multiple addresses.

Advanced workbook settings

The advanced settings allow custom key/value properties and logging severity/customization options.

- Custom properties. add [custom properties](#).

Custom Properties:

das.sampling.lookahead.maxdepth=0
das.job.concurrent-mr-job=6
das.join.map-side.memory.max-file-size=5000000

One key/value pair per line. Multiple values are comma separated.
E.g. key=value1,value2

- Default log severity. Specify the severity of logging messages to be recorded.

Default log severity:

ERROR

Default log severity for the ROOT logger. Messages above the specified minimum severity will be logged.
(For non-embedded Hadoop mode only)

- Logging customization. Add custom [LOG4J settings](#).

Logging Customization:

log4j.category.mypackage.MyClass=DEBUG

Custom LOG4J logging settings. In log4j.properties format.
e.g.
log4j.category.mypackage.MyClass=DEBUG
(For non-embedded Hadoop mode only)

Run workbook after saving or updating the workbook

Select the checkbox to run the workbook directly after saving or updating the workbook configuration page.

▼ Run Workbook After Save

Process now:

Start calculation process immediately after save

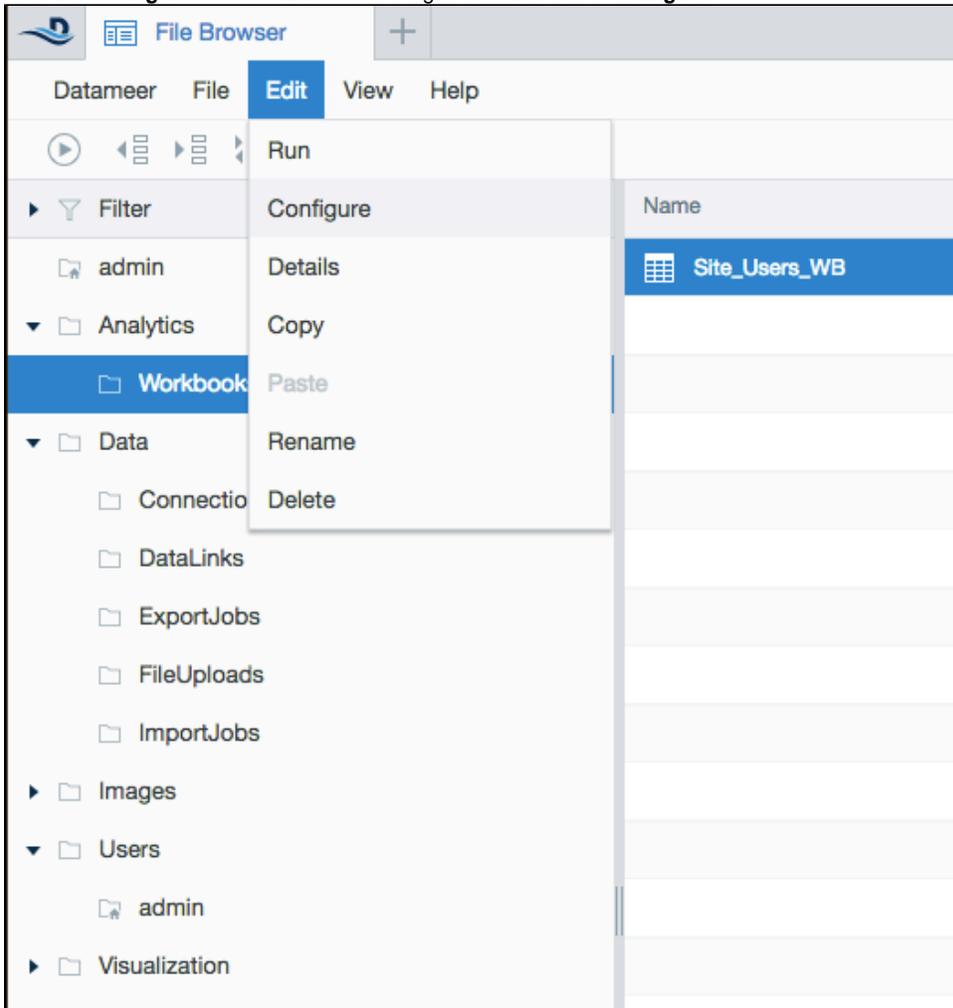
Close

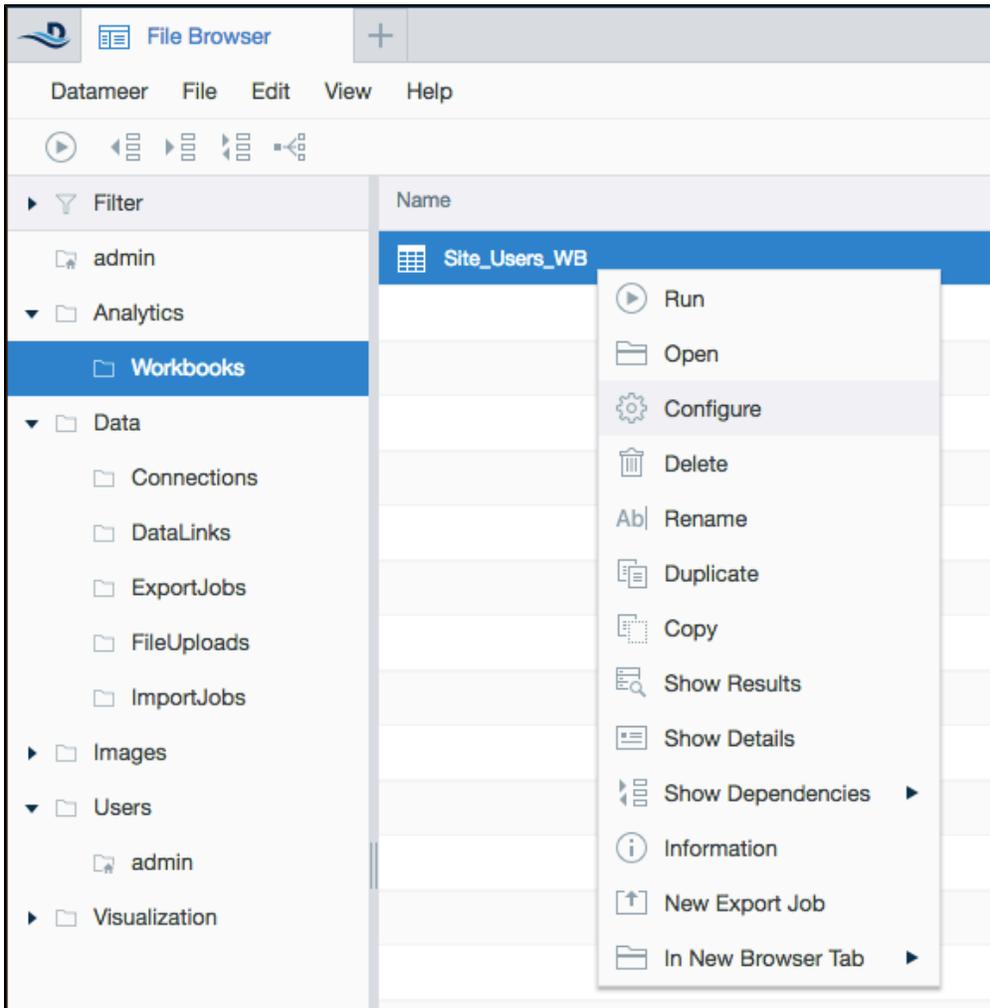
Save

Changing Workbook Settings

To change workbook settings:

1. Select the workbook from the **File Browser**.
2. Click the **Configure** from the **Edit** menu or right-click and select **Configure**.





3. Make changes and click **Save**.