



This PDF is generated from authoritative online content, and is provided for convenience only. This PDF cannot be used for legal purposes. For authoritative understanding of what is and is not supported, always use the online content. To copy code samples, always use the online content.

Genesys Predictive Routing

dataset-`<name>` Section

5/8/2025

dataset-<name> Section

The **dataset-<name>** section is used to configure Datasets, in combination with the corresponding **schema-<name>** section. The <name> part of the section name should be replaced by a Dataset name of your choice.

Create multiple **dataset-*** sections, each containing the necessary options. Each one corresponds to a Dataset. Note that a Dataset configured in a **dataset-<name>** section can include fields joined from other Datasets, and/or from the Agent Profile and Customer Profile. This joining is also controlled by options configured in that section.

Special Dataset Names and Configuration

The following section names are reserved for the Agent Profile dataset, which is created by direct import from Genesys Info Mart, and the Customer Profile dataset. These reserved names cannot be used for any other purpose. Each has particular configuration options that must be included, while others can be omitted. If an option has a value specified, that option requires that value when you in that section.

- **dataset-agents-gim**
 - **sql-query**
 - **data-type**=agents
 - **enforce-schema-on-joined-data**
 - **join**
 - **join-type**
 - **upload-dataset**
- **dataset-customers**
 - **csv-separator**
 - **data-type**=customers
 - **join-keys**
 - **location**
 - **upload-dataset**

Sections configuring uploads of interaction data from the Genesys Info Mart database do not have a mandatory name, but they do have mandatory options, and some of those options have mandatory values. For example, you might name this section **dataset-interactions-aht**. Interaction datasets require the following options:

- - **chunk-size**
 - **sql-query**
 - **data-type**=interactions

- **enforce-schema-on-joined-data**
- **join**
- **join-type**
- **upload-dataset**

Note: Interaction datasets are only uploaded from the direct connection to the Genesys Info Mart database. To add other interaction-related data, such as feedback or outcome data not stored in the Genesys Info Mart database, upload a CSV file with the **data-type** option set to outcomes.

Sections configuring CSV uploads of agent data do not have a mandatory name, but they do have mandatory options, and some of those options have mandatory values.

- For example, you might name this section **dataset-agents-csv**. It requires the following options:
 - **csv-separator**
 - **data-type**=agents
 - **join-keys**
 - **location**
 - **upload-dataset**

See [Configure Data Loader to Upload Data](#) for a comprehensive discussion of how the **dataset-<name>** and **schema-<name>** sections, and their options, work together to configure data for upload.

- | | | |
|---|-----------------------------------|---|
| • chunk-size | • join-keys | • teradata-sql-query |
| • csv-separator | • join-type | • trigger-pipeline-execution |
| • data-source | • kpi-type | • update-period |
| • data-type | • location | • upload-dataset |
| • end-date | • num-days-upload | • upload-schedule |
| • enforce-schema-on-joined-data | • sql-query | • use-cloud-feature-engineering |
| • join | • start-date | • vq-filter |

chunk-size

Default Value: PT15M (15 minutes)

Valid Values: String in ISO 8601 duration format, PT1S (1 second) or higher

Changes Take Effect: Immediately

Defines the chunk size, defined as interactions that happened within the specified length of time, used when extracting data from the Genesys Info Mart Database for a dataset of the *interactions* data type. Interactions that started within the chunk, as defined by this option value, are uploaded to the GPR Core Services platform as a single file and are then appended to the previously uploaded data for the associated dataset.

The following table provides sample formats that are supported for the chunk-size option.

Value	Description
PT20.345S	Specifies 20.345 seconds
PT15M	Specifies 15 minutes
PT10H	Specifies 10 hours
P2D	Specifies 2 days
P2DT3H4M	Specifies 2 days, 3 hours and 4 minutes
P-6H3M	Specifies -6 hours and +3 minutes
-P6H3M	Specifies -6 hours and -3 minutes
-P-6H+3M	Specifies +6 hours and -3 minutes

csv-separator

Default Value: comma

Valid Values: comma, tab

Changes Take Effect: Immediately

Indicates the separator used in CSV data files that Data Loader is to upload, as indicated in the **location** option configured in the same section. This option is necessary only if you are configuring a Dataset that is to be created by uploading a CSV file.

data-source

Default Value:

Valid Values: gim, csv, teradata

Changes Take Effect:

Introduced: 9.0.021.00

The data source.

If data-source is *teradata*, the **chunk-size** option will be ignored. Support of chunk-size has been removed for Teradata dataset uploads. By default it will be *P1D*.

data-type

Default Value: No default value

Valid Values: agents, customers, interactions, outcomes

Changes Take Effect: Immediately

Specifies the type of Dataset you are uploading.

- **agents** - Data Loader uploads the data to the Agent Profile on the GPR Core Services platform. This data can come from Genesys Info Mart or from a CSV file. You can also join it with a Dataset of the "interactions" type.
- **customers** - Data Loader uploads the data to the Customer Profile on the GPR Core Services platform. Its source is a CSV file. You can also join it with a Dataset of the "interactions" type.
- **interactions** - The Dataset contains interactions extracted from Genesys Info Mart database, which Data Loader uploads to the GPR Core Services platform. This data can optionally be joined with the Datasets of the "agents", "customers", or "outcomes" types before it is uploaded.
- **outcomes** - The Dataset contains information extracted from sources other than the Genesys Info Mart database and that is provided as a CSV file, which Data Loader uploads to the GPR Core Services platform. This data can optionally be joined with the Datasets of the "interactions" type.
Note: This data type is used for *any* data that is not of the "interactions" type and that is being uploaded to a Dataset with a user-specified name (that is, a Dataset other than **dataset-agents-gim** or **dataset-customers**). The data you are uploading does not have to be literal outcome data.

end-date

Default Value: 1970-01-01

Valid Values: date in YYYY-MM-DD format

Changes Take Effect: After 15 min timeout

The last date in the period for which Data Loader should retrieve data for a dataset. This date can be in the future.

- Change the default value to a date suitable for your environment. For example, you might enter 2020-11-04.

This option is required for datasets of the **interactions** and **outcomes** types. It is not used for datasets of the **customers** and **agents** types.

enforce-schema-on-joined-data

Default Value: true

Valid Values: true, false

Changes Take Effect: After 15 min timeout

- If set to **true**, all fields are joined to the "interactions" Dataset from the Datasets listed in the **join** option configured in the same section.
- If set to **false**, all fields from the Datasets listed in the **join** option configured in the same section are added to the "interactions" Dataset.

join

Default Value: No default value

Valid Values: a comma-separated list of section names containing dataset configurations

Changes Take Effect: After 15 min timeout

Specifies the list of the Datasets of the "agents", "customers", or "outcomes" types to join with the current "interactions" Dataset prior to upload to the GPR Core Services platform.

This join can be inner or outer depending on the value of the **join-type** option configured in the same section. The following examples show what you join depending on the value or values specified for this option:

- **agents** - Joins interaction data obtained from Genesys Info Mart with agent information uploaded from Genesys Info Mart or from a CSV file.
- **customers** - Joins interaction data obtained from Genesys Info Mart with customer information uploaded from a CSV file.
- **fcr** - Joins interaction data obtained from Genesys Info Mart with first call resolution (FCR) interaction outcome data provided in a CSV file. For this value to be valid, you must have configured **dataset-fcr** and **schema-fcr** sections.
- **agents, customers** - Joins interaction data obtained from Genesys Info Mart with the data from the Agents and Customer profiles.
- **agents, customers, fcr** - Joins interaction data obtained from Genesys Info Mart with the data from the the Agents and Customer profiles and the FCR interaction outcome data.

join-keys

Default Value: No default value

Valid Values: Comma-separated list of column names

Changes Take Effect: After 15 min timeout

A comma-separated list of the column names defined in the **schema-agents-gim** section that contain key values by which to join the data from this agent Dataset to an interaction type Dataset.

join-type

Default Value: inner

Valid Values: inner, outer

Changes Take Effect: After 15 min timeout

- **inner** - Only the records successfully joined with the specified Datasets (outcomes, agents, customers, FCR, and so on) are uploaded to the GPR Core Services platform. This is the the typical value used in production environments.
- **outer** - All interaction records are uploaded to the GPR Core Services platform. Any missing data is replaced with **null** values. Typically this value is used only for troubleshooting.

kpi-type

Default Value:

Valid Values: aht, sales, xfer, nca, fcr, csat, churn

Changes Take Effect: After 15 minutes

Introduced: 9.0.021.00

This option is mandatory only when **use-cloud-feature-engineering** is set to *true*.

This option defines how DataLoader will categorize the datasets and predict the right input S3 locations for the glue job to pick.

If any of the configured dataset section does not include this option, the data upload for that particular dataset configuration will fail with an error, *Missing kpi-type*.

location

Default Value: No default value

Valid Values: A valid path name string for a file containing a dataset in CSV format

Changes Take Effect: After 15 min timeout

Specifies the path to a CSV file containing a dataset. Required for the datasets provided as CSV files.

Configure the file location as described in the following steps:

1. Place the file itself in the Data Loader IP folder structure using the following path:
`<ip_folder>/ai-data-loader-scripts/scripts/datasets_<dataset_type>`
2. The value for the location option is the path inside the Data Loader Docker container. Specify only the final part of the full path as given below:
`/datasets/<dataset_type>/<dataset file name>.csv`

The possible dataset types are **agents**, **customers**, and **outcomes**.

Example:

- The folder path for the Customer Profile dataset is: `<ip_folder>/ai-data-loader-scripts/scripts/datasets_customers`
- The **location** option value for this file is `datasets/customers /<dataset_file_name>.csv`

Note: Interactions are only uploaded using the direct Genesys Info Mart-Data Loader connection. If

you are uploading additional interaction data from a CSV file, use the **outcome** dataset type.

If you want to update the dataset using a new CSV file, it must have the same file name or the option value must be changed to reflect the new file name. In either case, the folder where the file is located must remain the same.

num-days-upload

Default Value: 120

Valid Values: integers 1 to 180

Changes Take Effect: on initial startup

Introduced: 9.0.019.01

Specifies the number of days of data Data Loader should upload from Genesys Info Mart when it starts for the first time. The data for each day is uploaded to a separate file.

Warning! Do not change the default value without consulting with your Genesys representative to avoid unintended results.

After this initial upload, Data Loader uploads daily on the schedule you configure using the upload-schedule and chunk-size options.

Note: Do not change the **chunk-size** option value from the default, which is one day.

To avoid uploading partial data for calls in progress, Data Loader does not upload data from the current day. If Data Loader does miss any in-progress interactions, it uploads them the following day.

Genesys strongly recommends that you run uploads between the hours of 00:05am and 5:00am. This enables GPR to retrain your models on the new data during a period that is typically less active.

sql-query

Default Value: No default value

Valid Values: A string starting with "file:" and followed by a valid path to a file in the Data Loader Docker container containing an SQL query

Changes Take Effect: After 15 min timeout

Modified: 9.0.017.01

You need to configure this option only when you are using a customized query to extract data from the Genesys Info Mart database for the Agent Profile and interactions datasets. You do not need to configure the **sql-query** option to create datasets from .csv files, such as for Customer Profile data, outcomes data, and agent data from sources other than Genesys Info Mart.

Two example SQL queries are provided in the Data Loader Docker container for your reference:

- **/dl/interaction_data_aht.sql** - the query used to collect average handling time (AHT) data for Data Loader to upload to the interactions dataset.
- **/dl/agents_data_gim.sql** - the query used to collect data to populate the default Agent Profile dataset.

For instructions to create your own SQL query, see [Create your own SQL query](#) in the *Deployment*

and Operations Guide.

The following is an example of a valid value for this option: **file:/datasets/outcomes/my_interactions_data_gim.sql**

If you do not configure this option in the **[dataset-agents-gim]** or **[dataset-interactions-gim]** sections, Data Loader uses the appropriate default query.

start-date

Default Value: 1970-01-01

Valid Values: date in YYYY-MM-DD format

Changes Take Effect: After 15 min timeout

The earliest date in the period for which Data Loader should retrieve data for a dataset.

- Change the default value to a date suitable for your environment. For example, you might enter 2018-11-29.

This option is required for datasets of the **interactions** and **outcomes** types. It is not used for datasets of the **customers** and **agents** types.

teradata-sql-query

Default Value:

Valid Values: A string starting with "file:" and followed by a valid path to a file in the Data Loader Docker container containing an SQL query

Changes Take Effect: After 15 minutes

Introduced: 9.0.021.00

The SQL query to access Teradata for Outcomes handling. This option is required when the **datasource** option is *teradata*.

If this option is not configured when using the datasource as teradata, DataLoader will display an error when uploading outcomes and the data set upload will fail.

The timestamp fields, **start_ts** and **end_ts** are no longer supported for Teradata SQL. The **start_date** and **end_date** date fields must be used instead.

trigger-pipeline-execution

Default Value: False

Valid Values: True, False

Changes Take Effect: On next data upload

This option enables you to trigger the execution of the Cloud Feature Engineering pipeline. The pipeline execution happens after the next scheduled interactions data upload is complete (unless the

number of newly-uploaded records is 0).

update-period

Default Value: PT24H

Valid Values: String in ISO 8601 duration format, from PT15M to P30D

Changes Take Effect: After 60 sec timeout

Related Options: chunk-size

Specifies the interval at which Data Loader attempts to upload data, enabling fresh data stored in the Genesys Info Mart database to be automatically uploaded to the associated dataset. Used with **dataset-agents-gim** and the main interactions dataset, which are the datasets created directly from Genesys Info Mart data.

- If the **update-period** value is less than the value for the **chunk-size** option, Data Loader uploads all data after the watermark marking the end of the previous upload.
- If the **update-period** value is larger than the value of the **chunk-size** option, Data Loader uploads all data after the watermark, split into chunks of the size specified by the value of the **chunk-size** option.

Examples

NOTE: In the the examples below the value of the **end-date** option is set in the future.

- If **update-period** is set to 1 day (P1D) and **chunk-size** is set to one hour (PT1H), all the data after the previous watermark is uploaded in 1-hour chunks. This chunking is designed to prevent overloading your infrastructure.
- If you are uploading a dataset for the first time and set **start-date** to 90 days in the past, **update-period** to 1 day (P1D), and **chunk-size** to 30 days, Data Loader uploads the 90 days of data in three 30-day chunks.

upload-dataset

Default Value: see option description

Valid Values: true, false

Changes Take Effect: After 60 sec timeout

Modified: 9.0.017.01

Notifies Data Loader that the dataset is fully configured and the data processing for this dataset can be started. Data Loader checks every 60 seconds to see whether the value of this option has changed.

If set to `true`, Data Loader starts the dataset upload. If set to `false`, Data Loader does not upload data.

The default value for this option is pre-set to `true` for the **dataset-agent-gim** dataset and to `false` for the **dataset-interactions-gim** dataset. This configuration ensures that the agent profile, which needs to be in place first, is uploaded immediately.

NOTE: The **dataset-interactions-gim** configuration section was included in the default Data Loader configuration template starting in release 9.0.017.01. It is used with the cloud feature engineering

pipeline (CFEP).

upload-schedule

Default Value: No default value

Valid Values: A valid schedule in Cron format

Changes Take Effect: On the next upload

Introduced: 9.0.018.00

This option enables you to execute the upload of a dataset on a preconfigured schedule. In release 9.0.017.01 and lower, or if you do not set a value for the upload-schedule option, data upload scheduling is controlled using the update-period and chunk-size options.

The value for this option must be a Cron expression. For a complete explanation of how to create a Cron schedule, see the "Configure the data upload schedule" section of the "Configure Data Loader to upload data" topic in the Genesys Predictive Routing Deployment and Operations Guide.

use-cloud-feature-engineering

Default Value: True

Valid Values: True, False

Changes Take Effect: Immediately

Controls whether Data Loader should upload data to the cloud feature engineering pipeline (CFEP).

- **true** (the default) - Data Loader uploads your data to the GPR Core Platform via the CFEP, where it can be augmented with additional features and joined with other datasets before it is used for predictor and model creation, model training, and agent scoring.
- **false** - Data Loader uploads data as it did in previous releases, uploading it to the Agent Profile schema, Customer Profile schema, or a configured interactions or outcome dataset, depending on the value of the **data-type** option.

vq-filter

Default Value: No default value

Valid Values: a comma-separated list of valid virtual queue names

Changes Take Effect: On the next data upload

To have Data Loader upload data only from a subset of virtual queues (VQs) for inclusion in an interaction-type dataset, enter a comma-separated list of the VQs to include. Data Loader uploads records from the Genesys Info Mart database associated with the specified VQs.