GENESYS™

Genesys Predictive Routing Deployment and Operations Guide

Configure Data Loader to upload data

2/8/2026

Contents

Learn how to configure Data Loader to upload data to your instance of the GPR Core Platform for analysis and agent scoring.

Related documentation:

-

Overview

Data Loader uploads data to the GPR Core Platform for analysis and agent scoring. GPR uses the following types of data:

- Agent Profile data and interaction data, which Data Loader uploads using a direct connection to Genesys Info Mart. Once you have configured the connection parameters, Data Loader automatically uploads this data and refreshes it periodically.

- Customer Profile data, which can come from several different sources, depending on your environment. To upload customer data, you must collect it into a .csv file with a consistent schema for Data Loader to upload it.

- Optionally, you can upload more agent data, interaction outcome data, and any other data you find useful. Compile this data into .csv files for Data Loader to upload it.

To upload data, you must create, and specify values for, the configuration options on the **Options** tab of the Data Loader `Application` object. You configure the configuration sections and options in your configuration manager application (GAX or Genesys Administrator).

- The Predictive Routing Options Reference contains a complete list of all Predictive Routing options, with descriptions, default and valid values, and a brief explanation of how to use them.

This topic focuses specifically on data-upload configuration and provides a task-based perspective, enabling you to find which options you must configure for specific upload types and giving more context to help you decide on the appropriate values for your environment.

NOTE: Once you have configured a schema in the Data Loader `Application` object and uploaded your data to the GPR Core Platform, you cannot modify the schema.

Data upload basics

This section explains how to configure Data Loader to upload the essentials for Predictive Routing. These essentials are the Agent Profile dataset, the Customer Profile dataset, and an interaction dataset.

- Create an Agent Profile dataset and a Customer Profile dataset before trying to join interaction, outcome, or other data. The sections that follow cover more in-depth scenarios, including adding other

types of data and joining data.

To upload to Data Loader, open the Data Loader `Application` object in your configuration manager (Genesys Administrator or GAX) and set up the following configuration sections and options for each essential category of data.

1. Agent Profile

To create your Agent Profile dataset, configure the following two sections, and include the options specified within each:

- **dataset-agents-gim**
    - *sql-query* - Use the default query provided with Data Loader (the **agents_data_gim.sql** file in the **/dl** folder) or create your own. If you plan to use your own SQL file, be sure to include the mandatory fields listed in Create your own SQL query (below).
    - *data-type* - The value for this option must be `agents`.
    - *upload-dataset* - This option is set to `true` by default, which tells Data Loader to start uploading data at the interval specified in the *update-period* option. To pause uploading, set this option to `false`.
    - *upload-schedule* (in release 9.0.018.00 and higher) or *update-period* - Specifies how often Data Loader runs an upload process to get new and updated data.
- **schema-agents-gim** - In this section, create one option for each column in your Agent Profile dataset. You must include the columns/options listed below. If you modified your SQL query to read values from additional fields, you must also add those fields as options in this section. The option values are the datatype, and, if the field contains sensitive or PII data, the value anon.
    - **dbID**
    - **EMPLOYEE_ID** (note that this field name must use exactly this format, all caps with an underscore)
    - **groupNames**
    - **skills**
    - **tenantDbID**

2. Customer Profile

To create your Customer Profile dataset, configure the following two sections, and include the options specified within each:

- **dataset-customers**
    - *csv-separator* - Tells Data Loader what separator type you are using so it can correctly parse the .csv file.
    - *data-type* - The value for this option must be `customers`.
    - *location* - Enter the Data Loader Docker container path to the .csv file containing your Customer Profile data.
    - *upload-dataset* - This option is set to `true` by default, which tells Data Loader to start uploading data at the interval specified in the *upload-schedule* (in release 9.0.018.00 and higher) or *update-period* option. To pause uploading, set this option to `false`.

- **schema-customers** - In this section, create one option for each column in your Customer Profile dataset. The option values are the datatype, and, if the field contains sensitive or PII data, the value anon.

  - **CustomerID** - The only required option/column.

3. Interaction Dataset

To create a basic interaction dataset, complete the following steps. In a production environment, you would typically join the interaction data to the Agent and/or Customer Profile datasets. The sections that follow explain how to join data. This section gives you the minimum necessary information to create an initial interaction dataset.

- **dataset-name>** - When you create this configuration section, replace name> with a dataset name of your choice. For example, you can create a section called **dataset-interactions**.

  - **num-days-upload** - (In release 9.0.019.01 and higher) Specifies the number of days of data to upload in your initial data upload from Genesys Info Mart.

  - **chunk-size** - Defines how many interactions you want Data Loader to upload at once by specifying a period of time. Data Loader uploads all interactions that occur during this amount of time in a single batch.

  - **upload-schedule** (in release 9.0.018.00 and higher) or **update-period** - Specifies how often Data Loader runs an upload process to get new and updated data.

  - **sql-query** - Use the default query provided with Data Loader (the **interaction_data_aht.sql** file in the **/dl** folder) or create your own. If you plan to use your own SQL file, be sure to include the mandatory fields listed in the option description. See Create your own SQL query for instructions.

  - **data-type** - The value for this option must be `interactions`.

  - **upload-dataset** - This option is set to `true` by default, which tells Data Loader to start uploading data at the interval specified in the **upload-schedule** (in release 9.0.018.00 and higher) or **update-period** option. To pause uploading, set this option to `false.`

- **schema-name>**- When you create this configuration section, replace name> with the same dataset name you entered for the **dataset-name>** section. If your dataset section is called **dataset-interactions**, this section is **schema-interactions**.

How to join data

Joining data enables you to pull together information about many aspects of your environment, enabling more powerful and nuanced predictions. When you join the data, the interaction dataset is the one containing joined fields. That is, the interaction dataset uploaded to the GPR application displays all its own fields, plus the fields from the joined datasets.

This section explains the basic procedure for joining the Agent Profile and/or Customer Profile dataset with an interaction dataset. The "Advanced upload scenarios" section on this page presents further examples.

Data upload sequence

When joining features from the Agent Profile or Customer Profile datasets to the interactions dataset,

the order in which you upload the datasets initially is important. Upload data in the following sequence:

1. Upload Agent Profile data by setting the **[dataset-agents-gim].upload-dataset** option to `true`. This setting starts upload of Agent Profile data from Genesys Info Mart.

2. After Data Loader starts the agent data upload, configure the Customer Profile data configuration sections and start the Customer Profile data upload by setting the **[dataset-customers].upload-dataset** option to `true`.

3. Then configure the interactions data upload. The procedure below explains how to specify the Agent Profile and Customer Profile dataset fields to be joined to the interaction data. Once this configuration is complete, start the upload of the interaction dataset by setting the **[dataset-name>].upload-dataset** option to `true`.

Data joining procedure

To enable joining, perform the following steps:

1. In the Data Loader `Application` object, create or edit the **dataset-name>** section for your interaction dataset. These instructions refer to the **dataset-interactions** section used as an example in the "Data upload basics" procedures, above. If your dataset has a different name, use your dataset name throughout.

2. Add the ***join*** option. As the option value, specify the names of the datasets to be joined.

   • For example, if you would like to join data from the Agent Profile and Customer Profile datasets with **dataset-interactions**, set the value of the ***join*** option to: **agents-gim**, **customers**, **interactions**.

3. Specify the fields to be joined to the interaction dataset by adding the names of the columns from the joined datasets into the Data Loader **schema-interactions** configuration section.

   • Each joined column must be a separate option in the **schema-interactions** configuration section. The value for each of these options must be the name of the **schema-** section from which the data comes.

   • For example, if you join fields from the Customer Profile (which you created using the **dataset-customers** and **schema-customers** configuration sections) with your interaction dataset, the option value for all joined fields is **schema-customers**.

4. Specify the desired values for the following additional configuration options:

   • ***join-type*** - Configured in the **dataset-interactions** section. Specifies the join type, whether inner (only successfully joined records are uploaded) or outer (all interaction records are uploaded to the platform and the missing data is replaced with `null` values).

   • ***join-keys*** - Configured in the **dataset-** sections of the datasets you are joining to the interaction dataset. Specifies a comma-separated list of the column names by which to join the data from this dataset to the interaction dataset.

   • ***enforce-schema-on-joined-data*** - Specifies whether fields that are not already in the interaction dataset are joined. If set to true, only data from fields already in the interaction dataset are joined. If set to **false**, ALL columns from the joined datasets are added to the interaction dataset. If you set this option to **false**, be careful not to exceed the maximum allowed number of 100 columns in the joined dataset, or Data Loader generates an error and exits the upload process.

The following example shows a **schema-** configuration section for an interaction dataset with joins to the Agent and Customer Profile datasets and an outcome dataset called **feedback**:

[schema-interactions-joined]

- **AHT**=numeric

- **Customer_location**=schema-customers

- **CUSTOMER HOLD DURATION**=numeric

- **CUSTOMER TALK DURATION**=numeric

- **CUSTOMER_ACW_DURATION**=numeric

- **Customer_language**=schema-customers

- **Customer_ANI**=schema-customers

- **EMPLOYEE_ID**=schema-agents-gim

- **Feedback**=schema-feedback

- **groupNames**=schema-agents-gim

- **InteractionID**=string

- **Is_Resolved**=schema-feedback

- **skills**=schema-agents-gim

- **START_TS**=timestamp,created_at

Advanced upload scenarios

In addition to the Agent Profile, Customer Profile, and interactions datasets described in the preceding sections, you can upload other data from your environment. This section presents examples showing how to upload data from additional sources.

Add agent data from sources other than Genesys Info Mart

If your business has employee data relevant for Predictive Routing that is not available from Genesys Info Mart, create a .csv file with that data and configure the Data Loader `Application` object with the following sections and options:

- **dataset-agents-name>**

  - *csv-separator* - Tells Data Loader what separator type you are using so it can correctly parse the .csv file.

  - *data-type* - The value for this option must be `outcomes`.

  - *location* - Enter the Data Loader Docker container path to the .csv file containing your Agent Profile data.

  - *upload-dataset* - This option is set to `true` by default, which tells Data Loader to start uploading data at the interval specified in the **update-period** option. To pause uploading, set this option to `false`.

- **schema-agents-name>** - Shows how to configure the schema for an agent Dataset to be uploaded from a .csv file. It has only one mandatory field, but you must also configure options specifying the data types for all other fields. Add the anon parameter if the field must be anonymized.

- EMPLOYEE_ID=string,id,anon

Upload outcome data

To upload outcome data from a post-interaction survey, use the following configuration. You can adapt this example to create a dataset based on nearly any data source. In this example, we are uploading a dataset called **feedback**.

- **dataset-feedback**
  - ***csv-separator*** - Tells Data Loader what separator type you are using so it can correctly parse the .csv file.
  - ***data-type*** - The value for this option must be `outcomes`.
  - ***location*** - Enter the Data Loader Docker container path to the .csv file containing your Feedback data.
  - ***upload-dataset*** - This option is set to `true` by default, which tells Data Loader to start uploading data at the interval specified in the ***update-period*** option. To pause uploading, set this option to `false`.

- **schema-feedback**

-
  - Customer_ANI=string,anon
  - EMPLOYEE_ID=string,anon
  - Feedback=string
  - InteractionID=string
  - Is_Resolved=boolean
  - Timestamp_Feedback=timestamp,created_at

Upload aggregated interaction data to your Agent Profile dataset

Note: In Data Loader release 9.0.017.01 and higher, the cloud-based feature engineering pipeline (CFEP) provides more robust aggregation capabilities. See Configure for Feature Engineering for details.

Aggregated features enable you to pull together data from various sources to create features that address specific scenarios. For example, you can define a feature that gives you the average handle time (AHT) for the interactions each agent in the Agent Profile dataset received from a certain virtual queue over a specified period.

To configure aggregated features in the Agent Profile, replace the **dataset-agents-gim** and **schema-agents-gim** configuration sections with the following:

- **dataset-agents-gim-ext** - *In addition to* the standard options configured for the **dataset-agents-gim** section, add the following options:
  - start-date
  - end-date
  - chunk-size

- **schema-agents-gim-ext** - Be sure to add all fields that Data Loader must aggregate as options in this configuration section.

Create your own SQL query

For your reference, the Data Loader container includes the following two default SQL queries:

- **/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.

To review these SQL queries, follow these steps:

1. Run the following command to locate the container ID:
   ```
   docker ps
   ```

2. Use the response from Step 1 to run the following command, which accesses the container:
   ```
   docker exec -it /bin/bash
   ```

3. Navigate to the **/dl/** folder in the container and use the following command to view the default SQL query files:
   ```
   cat
   ```

To create your own custom SQL query, use the following procedure:

1. Review the requirements for mandatory fields contained in this section (below) and follow the guidelines to create your queries.

2. Place your custom SQL query files into one of the subfolders in your installation directory. The installation directories are mapped to the container. The folders available on the host computer are the following:

   - **/ai-data-loader-scripts/scripts/datasets_customers**

   - **/ai-data-loader-scripts/scripts/datasets_agents**

   - **/ai-data-loader-scripts/scripts/datasets_outcomes**

3. For each dataset using a custom SQL query, use GAX to open the associated dataset configuration section in the Data Loader **Application** object. Specify the path to the query file in the *sql-query* option. The path must be given relative to the container folder structure.
   For a table containing the mapping between folders on the host and in the container, see Mapping between folders on the host and in the container.

Mandatory fields for custom SQL query files

The following fields are mandatory in your agent data SQL file:

- **EMPLOYEE_ID** - The id of the agent. Must match the name of the ID Field in the Agent Profile dataset.

- **dbID** - The agent's configuration database DBID. Obtained from the Genesys Info Mart RESOURCE_.RESOURCE_CFG_DBID field.

- **tenantDbID** - Obtained from the Genesys Info Mart RESOURCE_.tenant_key field.

- **groupNames** - List of group names assigned to the agent.
- **skills** - Dictionary containing the agent's current skills.

The following fields are mandatory in your interaction data SQL file:

- **InteractionID** - Obtained from the Genesys Info Mart INTERACTION_FACT.media_server_ixn_guid field.
- **EMPLOYEE_ID** - field that defines the Agent ID, which must match the name of the id_field in the agents dataset
- **WHERE clause** - Must include the following:
  - irf.start_date_time_key - This value must fall between the dates set in the :start_ts and :end_ts parameters. Data Loader replaces the :start_ts and :end_ts fields with the start and end time of the interactions dataset chunk, in epoch format.
  - **Note:** The abbreviation **irf** in the WHERE clause indicates the INTERACTION_RESOURCE_FACT table in the Genesys Info Mart database.

To create a custom SQL query when using the cloud feature engineering pipeline (CFEP), set the **use-cloud-feature-engineering** option to **true** and create your query using the following SQL template placeholders:

- :user_data_fields in the SELECT clause and :user_data_joins in the WHERE clause to extract the user data stored in the Info Mart database.
- :vq_filter in the WHERE clause to enable filtering of the interaction records by the Virtual Queue names defined in the **vq-filter** option.
- If you use multiple mapping rules for one KVP, you need to specify how Data Loader should include them in the dataset.

NOTES:

- The field names are case-sensitive.
- Use of angle brackets to signify "not" is not supported. Genesys recommends that you use **!=** instead.

Configure the data upload schedule

How often Data Loader uploads data from Genesys Info Mart and how much data it uploads at once is controlled by configuration options. The options available depend on your Data Loader version.

- For releases 9.0.018.00 and higher, you can use the **upload-schedule** option to configure precise schedules in CRON format.
- For releases before 9.0.017.01, or if you choose not to use the CRON scheduling, two options, **update-period** and **chunk-size**, coordinate to control upload timing and size.

9.0.018.00 and higher

A CRON expression is a string consisting of six or seven subexpressions (fields) that describe individual details of the schedule. These fields, separated by white space, can contain any of the

allowed values with various combinations of the allowed characters for that field. The following table shows the fields in the order expected and the allowed values for each.

Name	Required	Allowed Values	Allowed Special Characters
Seconds	Y	0-59	, - * /
Minutes	Y	0-59	, - * /
Hours	Y	0-23	, - * /
Day of month	Y	1-31	, - * ? / L W C
Month	Y	0-11 or JAN-DEC	, - * /
Day of week	Y	1-7 or SUN-SAT	, - * ? / L C #
Year	N	empty or 1970-2099	, - * /

Example Cron Expressions

Cron expressions can be as simple as * * * * ? * or as complex as 0 0/5 14,18,3-39,52 ? JAN,MAR,SEP MON-FRI 2002-2010.

The table contains various Cron expressions and their meanings.

Expression	Means
0 0 12 * * ?	Start dataset upload at 12:00 PM (noon) every day
0 15 10 ? * *	Start dataset upload at 10:15 AM every day
0 15 10 * * ?	Start dataset upload at 10:15 AM every day
0 15 10 * * ? *	Start dataset upload at 10:15 AM every day
0 15 10 * * ? 2005	Start dataset upload at 10:15 AM every day during the year 2005
0 * 14 * * ?	Start dataset upload every minute starting at 2:00 PM and ending at 2:59 PM, every day
0 0/5 14 * * ?	Start dataset upload every 5 minutes starting at 2:00 PM and ending at 2:55 PM, every day
0 0/5 14,18 * * ?	Start dataset upload every 5 minutes starting at 2:00 PM and ending at 2:55 PM, AND fire every 5 minutes starting at 6:00 PM and ending at 6:55 PM, every day
0 0-5 14 * * ?	Start dataset upload every minute starting at 2:00 PM and ending at 2:05 PM, every day
0 10,44 14 ? 3 WED	Start dataset upload at 2:10 PM and at 2:44 PM every Wednesday in the month of March
0 15 10 ? * MON-FRI	Start dataset upload at 10:15 AM every Monday, Tuesday, Wednesday, Thursday and Friday
0 15 10 15 * ?	Start dataset upload at 10:15 AM on the 15th day of every month
0 15 10 L * ?	Start dataset upload at 10:15 AM on the last day of every month
0 15 10 ? * 6L	Start dataset upload at 10:15 AM on the last Friday

Expression	Means
	of every month
0 15 10 ? * 6L	Start dataset upload at 10:15 AM on the last Friday of every month
0 15 10 ? * 6L 2002-2005	Start dataset upload at 10:15 AM on every last Friday of every month during the years 2002, 2003, 2004, and 2005
0 15 10 ? * 6#3	Start dataset upload at 10:15 AM on the third Friday of every month
0 0 12 1/5 * ?	Start dataset upload at 12 PM (noon) every 5 days every month, starting on the first day of the month
0 11 11 11 11 ?	Start dataset upload every November 11 at 11:11 AM

9.0.017.01 and lower

The *update-period* option 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.