![Genesys logo]

Genesys Predictive Routing Deployment and Operations Guide

Set up data for import

8/26/2025

Contents

Certain requirements and limitations apply to the data that you upload to GPR. This topic explains these requirements, and also presents data security and anonymization.

Related documentation:

  •

Supported types of data

In general, you need the following types of data:

Interaction data

  • Data Loader automatically extracts interaction data from the Genesys Info Mart database to create datasets.

Agent Profile data

  • Data Loader automatically extracts agent data from the Genesys Info Mart Database. You can optionally add agent data from other sources by providing a .csv file for Data Loader to upload.

Customer Profile data

  • To create the customer profile, create a .csv file and upload it using Data Loader.

Outcome and other data

  • To use outcome data, or data any other sort you find to be relevant for predictive routing, such as results of an after-call survey, create a .csv file and upload it using Data Loader.

See Configure Data Loader to upload data for how to configure Data Loader to upload both Genesys Info Mart data and .csv data.

See the relevant portion of the Help for how to use the GPR application to view your uploaded data and append data to existing datasets.

.csv file size requirements

Use the following guidelines to construct .csv files for data uploads:

  • Data Loader uploads data in 512-MB chunks. If your dataset is larger than 512 MB, Data Loader

automatically breaks it into chunks for upload.

- The maximum number of columns in a dataset is 100; the maximum number of rows is 2.5 million. If you upload a file with more than 2.5 million rows, Data Loader uploads only the first 2.5 million and discards the remainder.

- The maximum length of a single column name in a .csv file for upload is 127 characters.

- The maximum length of a single column name that Data Loader will anonymize is 120 characters.

- The maximum number of rows in the agent profile is 20 thousand.

- The number of rows in the customer profile is 20 million.

- The maximum number of columns (features) in the agent and customer profile datasets is configured for each account. The default limit is 50 features for each profile dataset. Only a STAFF user can change this value.

If you try to upload more data than the data size limits allow, GPR generates an error and discards the remaining rows.

When you have reached the size limit, GPR does not add records. However, you can update data associated with previously uploaded records (as identified by the Agent or Customer ID). For example, if you have uploaded 20,000 agents, you cannot add any more. But you can upload the same agents with new values, such as skills or location, and GPR makes those updates.

To add records, you must remove some uploaded records using the GPR API */purge endpoints.

.csv data formatting requirements

- When you create the .csv data file for a dataset, agent profile, or customer profile, do not include the following in the column name for the ID_FIELD, the Agent ID, or the Customer ID:

- ID

- _id

- Any variant of the string ID that changes only the capitalization.

Using these strings in the column name results in an error when you try to upload your data.

- When you create the .csv data file for a dataset, agent profile, or customer profile, do not include the following reserved names in column names:

  - created_at

  - tenant_id

  - updated_at

  - acl

- In the agent profile, if you are using skill names that include a dot (period) or a space in them, use double quotation mark characters to enclose the skill name. For example, enter a skill named ''fluent spanish.8'' as "fluent spanish.8".

- GPR supports UTF-8 encoding. All responses and returned data arrives in UTF-8 encoding.

- If you use a Microsoft editor to create your .csv file, remove the carriage return (^M) character before uploading. Microsoft editors such as Excel, WordPad, and NotePad automatically insert this character. For tips on removing the character from Excel files, refer to How to remove carriage returns (line breaks) from cells in Excel 2016, 2013, 2010.

- GPR supports only one-dimensional dictionaries, with up to 200 key-value pairs where the key is a string and the value is int, float, or Boolean. GPR does not support nested dictionaries and lists.

- If you have dictionary-type fields that use comma separators, use tab separators for your .csv file.

- Fields of the dictionary (DICT) type are discovered correctly only if the quotes appear as in the following example, with double quotation marks outside a dictionary entry and single quotation marks for the values within it. This requirement applies to DICT fields in all datasets, including the agent and customer profile datasets.

    - "{'vq_1':0.54,'vq_2':6.43}"

- GPR does not support use of angle brackets () to signify "not" in SQL queries. Genesys recommends that you use the following symbols instead: **!=**.

Data size for models and scoring

The following size limits apply to model creation:

- Maximum number of active models per Tenant - 50

- Total cardinality limit for model training: no specific column count; has been tested up to 250 columns.

    - Total cardinality must be less than 2 to the power of 29.

    - *Total Cardinality* = the number of numeric columns plus the sum of the number of unique values across all string columns within a specified dataset.

- Record count limit for model training - not applicable; from a model-training perspective there is virtually no limit on the number of columns. The constraining issue is the possibility of compromising the model quality by ending up with a reduced number of samples for training.

    - The total number of records must be less than 2 to power of 29 (that is, 536870912) divided by total cardinality as defined above.

    - **Example 1:** You must to use ALL of the data for training the model . If the dataset contains 1 million records, the maximum total cardinality is 536 (536870912 divided by 1 million).

    - **Example 2:** You can *undersample* the data for training the model—that is, use fewer than the ideal number of records for training. You might take 10,000 as the total cardinality, but only 53,687 of your total of 1 million records will be used for training. The calculation to determine this is 10,000 * 53,687 = 536870912 (the maximum cardinality).

The following limitation applies to scoring requests:

- Maximum number of agents that can be scored in one scoring request - 1,000.

Data anonymization

PII, or personally identifiable information, and sensitive data, such as passwords, must be hidden when you upload it to the GPR Core platform. To ensure that sensitive data is secured, instruct Data Loader to anonymize the fields containing such data.

After Data Loader anonymizes the fields you identified as PII, it uploads it securely using TLS.

Note the following points about anonymized data in GPR:

- You can anonymize up to 20 fields in each dataset.

- You cannot anonymize fields *after* you have uploaded data.

- Once you have uploaded data with anonymized fields, you cannot de-anonymize them.

- Anonymizing Numeric or Boolean fields changes them to String fields. This change has some effect on how the fields are weighted in the Feature Analysis report and during scoring.

- Each Tenant has its own unique salt for anonymization.

NOTE: If you anonymize a field, you must anonymize it in every dataset in which it appears. For example, if you anonymize a customer phone number in the customer profile, you must also anonymize it in any dataset in which it appears. If there is an inconsistency, GPR cannot correctly map agents and, as a result, cannot build models for them.

GPR uses the following steps to ensure secure data handling:

1. When Data Loader starts up, it generates a unique 64-character salt string that will be used for anonymization. It stores this string in the **anon-salt** option in the **[default]** section on the **Annex** tab of the primary and backup Data Loader `Application` objects and the Predictive_Route_DataCfg `Transaction List` object.

   - When you open these options in GAX, or any other configuration manager application you use, you cannot see the salt value itself. What you see is an obfuscated version of the salt string.

   - **WARNING!** Do not edit or delete the value Data Loader sets for the **anon-salt** options. If you try to modify a salt value, GPR generates an alarm message and Data Loader restores the original salt value. *If for some reason, Data Loader cannot restore the original salt value, your predictors become unusable for scoring and routing.* To rectify this situation you must recreate the agent and customer profiles, reload all interaction datasets, and retrain your models. If you do not recreate the agent and customer profiles and datasets exactly, you must also create and train new predictors and models. Therefore, Genesys strongly recommends that you do not modify or delete the salt values.

2. Before uploading the dataset to the GPR Core Platform, Data Loader uses this salt to anonymize the fields you specified as sensitive or PII data when you configured the schema.

3. The anonymized data is uploaded to the GPR Core Platform using TLS for secure data transport. The uploaded data is used for creating predictors and models.

4. After you create a predictor and one or more models, and begin using them to route interactions, the GPR subroutines retrieve the list of sensitive or PII features that are included in the active predictor. This list of features is stored in the URS Global Map.

5. The GPR Subroutines access the on-premises instance of your data to use in scoring requests. As a result, the Subroutines anonymize all sensitive fields included in the predictor you are using for scoring, based on the salt value stored in the Predictive_Route_DataCfg `Transaction List` object.

6. If one of the anonymized fields is the EMPLOYEE_ID, after the ActivatePredictiveRouting subroutine receives the response to the score request, it maps the agent scores back to the non-anonymized versions of the employee IDs so that routing can proceed.

7. Before the GPRIxnCleanup subroutine reports the routing outcome to the GPR Core Platform, it anonymizes all fields marked as PII that are included in the score outcome report. It then sends the results to the score log, which is stored in the cloud.

Unsupported characters in column names

The following characters are not supported for column names. If GPR encounters these characters in a .csv file, it reads them as column delimiters and parses the data accordingly.

- The pipe character
- `\t` (the TAB character)
- `,` (the comma)

Workaround: To use these characters in column names, add double quotation marks (" ") around the entire affected column name, except in the following situations:

- If you have a comma-delimited .csv file, add double quotations marks around commas within column names; you do *not* need quotations for the \t (TAB) character.
- If you have a TAB-delimited .csv file, add double quotations marks around TAB characters within column names; you do *not* need quotations for the , (comma) character.
- You must *always* use double quotations for the pipe character.

Data retention policies

GPR follows standard Genesys data retention guidelines for Genesys Multicloud CX as outlined in Section 14 of the Genesys Multicloud CX User Guide.

Most objects and data are deleted automatically after 90 days during which they are inactive. These include the following:

- Dataset data and the dataset object - Deleted after 90 days of idle time, which means no new files were appended and the dataset was not used to generate any data for predictors in that period.
- File upload object - Deleted after 90 days of idle time. Here idle time means this file was not used to generate any data for predictors in that period.
- Agent / Customer Profiles - Deleted after 90 days of idle time, which means the profile was not updated in the last 90 days.
- Model - Deleted after 90 days of idle time, which means the model was not used for any score requests in last 90 days.
- Predictor generated data and the predictor object - Deleted after 90 days of idle time, which means that no associated model was used for a score request in last 90 days.

The following data uses different retention policies:

- Uploaded anonymized files - Deleted 7 days after upload.

- Files stored for billing purposes - Deleted 60 days after creation.