Genesys Info Mart User's Guide

Data-Quality Considerations

4/2/2025

Data-Quality Considerations

This page discusses the implications for data quality of certain challenging aspects of extract, transform, and load (ETL) processing:

- Partially merged calls
- Error handling in case of missing data
- High availability

Partially Merged Calls

In multi-site scenarios, Genesys Info Mart must merge data from the multiple sites. If an interaction moves from site to site during the handling process, Genesys Info Mart uses linkage data to integrate the data from various T-Servers into a single interaction. This merging can be disrupted for a number of reasons, causing data-quality issues.

There are three major cases when data from a site might be unavailable:

- The site is not monitored.
- The site is monitored, but information is missing.
- Information is delayed.

Data issues in a partially monitored environment

If you configure Genesys Info Mart to extract voice interaction data from topologies in which not all T-Servers or IVR Servers involved in the call flow are monitored by ICON, data inconsistencies can occur, such as incomplete and missing data.

> ### Important
>
> If you have an environment that includes intentionally unmonitored sites, note these sites in the GSYS_DNPREMOTELOCATION table, as described in Configuring Info Mart database for merge in the *Genesys Info Mart Deployment Guide*.

A partially monitored environment can result in missing data at the start, middle, or end of an interaction. The following interaction scenarios, or any combination of them, can affect the population of interaction data within Genesys Info Mart, resulting in data inconsistencies:

- The interaction originates in an unmonitored T-Server.

- The interaction terminates in an unmonitored T-Server.

- The interaction originates in a monitored T-Server, moves to an unmonitored T-Server, and then passes on to a monitored T-Server.

> ### Important
>
> Each time that the interaction moves from an unmonitored to a monitored T-Server, it appears to be a new interaction. For example, a single interaction might start on a monitored T-Server, be sent to an unmonitored T-Server, and then be sent to a monitored T-Server. This single interaction is represented in Genesys Info Mart as two interactions. When this type of interaction scenario occurs, the linkage information that ties an interaction together as it moves from T-Server to T-Server is incomplete, and ICON cannot associate what it sees as multiple calls into a single interaction.

A partially monitored deployment can result in data that is incorrect or missing from the following fact tables:

- `INTERACTION_FACT (IF)` — Some interaction facts will be missing where entire calls or parties are missing in the source data.

- `INTERACTION_RESOURCE_FACT (IRF)` — Some IRFs might not reflect accurate information about mediation resources; consultation, conference, and transfer metrics; technical descriptors; routing targets; and service level flags, such as `MET_SERVICE_OBJECTIVE_FLAG` and `SHORT_ABANDONED_FLAG`. This occurs because these fields are highly dependent on other resources that are involved in the interaction which might or might not be monitored. Because these fields are highly dependent on other resources that are involved in the interaction, incorrect data results when the other resources are not monitored.

- `IXN_RESOURCE_STATE_FACT(IRSF)` — In some IRSFs, the `STATE_DESCRIPTOR` and `STATE_ROLE` values of the referenced Interaction Resource State might not be accurate for the states that are generated for IRFs. Because these values are populated based on interaction-type information from T-Server, incorrect data results when this information changes as a result of an unmonitored T-Server in the environment.

- `MEDIATION_SEGMENT_FACT (MSF)` — Some MSFs might be missing or have incorrect technical descriptor values because the ETL cannot determine why the interaction was placed in the queue or virtual queue, or whether it was answered or abandoned after it was distributed from the queue or virtual queue.

Late Data

Late data from sources that Genesys Info Mart considers to be currently active can be a result of various issues, such as intermittent connectivity issues for an IDB. For example, you might currently have data for only the beginning of an interaction, but data from a second T-Server is anticipated to arrive "soon". In this case, soon means that data should arrive before the timeout set in the Genesys Info Mart extract-data-stuck-threshold configuration option, which specifies the allowable delay to wait for the missing data to become available, or before the merge timeout that is controlled by the max-call-duration option. If the delayed data arrives before the stuck threshold timeout expires, the interaction is processed normally. If the threshold expires before the data arrives, the data is treated as missing.

For more information about the Genesys Info Mart timeouts, see the *Genesys Info Mart Configuration Options Reference*. For the definition of what Genesys Info Mart considers to be active data sources, see Genesys Info Mart Terminology Conventions in the *Deployment Guide*.

Error Handling in Case of Missing Data

For various reasons, information from a data source for a specific time range might be missing from a monitored site. In an HA environment, a single failure does not result in loss of data. If, in exceptional circumstances, multiple failures occur, the HA environment becomes, in effect, a non-HA environment.

For a full discussion of the potential points of failure in a non-HA environment and the implications for data quality, see the section on error handling in the chapter about ETL processing in the *Genesys Info Mart 8.1 Deployment Guide*.

Configuration options for missing-data behavior

For transformation, two configuration options are important for controlling the handling of missing data:

- error-policy-islink-dangling
- error-policy-irf-exception

The first option enables you to determine whether missing data is handled as an exception. If it is, the second option enables you to specify how such an exception is handled.

For more information about Genesys Info Mart error-handling, see the section on error handling in the chapter about ETL processing in the *Genesys Info Mart 8.1 Deployment Guide* and the **error-policy-*** option descriptions in the *Genesys Info Mart Configuration Options Reference*.

Missing Configuration Objects

Genesys Info Mart checks the list of known configuration objects during data transformation. If **Job_TransformGIM** notes a missing configuration object during transformation of configuration facts data, Genesys Info Mart records the information in the STG_IDB_FK_VIOLATION table.

During transformation of data types other than configuration data, Genesys Info Mart treats such missing configuration objects as late-arriving and creates placeholders for the missing objects based on the configuration object type and its unique ID. When the missing configuration objects arrive from Configuration Server, these placeholders are populated with missing data. The unique configuration ID prevents accidental duplication of configuration objects.

High availability

Deploying a high availability (HA) configuration can greatly reduce the possibility of data loss and

other issues with data quality. Genesys recommends using HA throughout the data chain, from data sources through Interaction Concentrator. Genesys Info Mart is designed to take advantage of HA configurations in order to determine and draw on the most complete and reliable data available.

In an HA configuration, each HA set of Interaction Concentrator instances consists of two or more redundant ICONs populating redundant IDBs. Genesys Info Mart selects available data in such a way that it takes the most complete and correct set of data from one of the redundant IDBs with no duplications. To accomplish this, Genesys Info Mart uses data-session information that ICON stores in each IDB to identify whether the data for the time period that is to be extracted is complete and correct. Genesys Info Mart extracts data from whichever one of the redundant IDBs has the most complete and reliable data for a particular time range.

This approach to identifying the best data for any period eliminates the need for resource-consuming double-extraction, analysis, and deduplication processes.

Criteria for choosing a better IDB

The rules that govern how the ETL process determines the time slices within an extraction cycle, as well as how it selects the best IDB source for each time slice, are designed to minimize both data loss and the number of switchovers from one IDB to another. For more information about the criteria for choosing the best IDB source for a particular time period, see the High Availability chapter in the *Genesys Info Mart 8.1 Deployment Guide*.

Important

Genesys Info Mart requires ICON to write session information to the IDBs, whether or not your environment is HA. For information about the ICON configuration settings that Genesys Info Mart requires, see Preparing the ICON Application in the *Genesys Info Mart Deployment Guide*.

Preventing data quality issues when restarting ICON

An HA configuration can eliminate multimedia data-quality issues that might arise as a result of setting the **calls-in-the-past** ICON configuration option to the required value of 1 (or `true`).

If you need to restart a multimedia ICON — for example, to install an upgrade — and you do not have an HA configuration, information about previous parties and first values of user data keys might be missing or inaccurate. Genesys recommends that you use an HA configuration, which eliminates these issues.

For additional information about potential data-quality issues for multimedia ICONs, see the discussion about special considerations when restarting a Multimedia ICON on the Managing ICON and data sources page in the *Genesys Info Mart Operations Guide*.