Data Loss and Component Failover

Data Flow from Logger to Historical Data Server

Assuming a full-duplex, fault-tolerant implementation, data is sent from CallRouter A to Logger A and from CallRouter B to Logger B.

The Logger Central Database forwards (replicates) historical data to corresponding historical tables on the Historical Database Server in summary intervals. These data populate the historical interval and daily reports.

Two Administration & Data Servers are typically set up as HDS machines. A similar fault-tolerant strategy applies to the HDS—when the primary HDS fails, the Administration Client automatically switches over to use the backup HDS.

Each Historical Data Server (HDS) is connected to a single Logger.

Recovery and Replication

Recovery Keys

The recovery key is the base key for all historical data tables. This key is always incremented by 1 before a new record is inserted into any historical table.

In a duplex configuration, the Logger that finishes initializing first is designated the primary Logger (although both the Loggers are always active and working in parallel). The recovery key is always initialized by the primary Logger. The recovery key is based on the current GMT date and time and always has a value greater than any previous value generated. This helps the recovery process to keep the Loggers in sync.

The replication process may have a latency of about one to five minutes because the Logger replicates data table-by-table on the HDS.

Temporary Tables

Each historical table on the Logger Central Database has two corresponding temporary tables that act as buffers to incoming historical data. As they have minimal indexes, the temporary tables speed up the process of inserting data into the corresponding actual table in the Logger Central Database.

Recovery Process

As the incoming historical data is written to the corresponding temporary tables by the Logger, the Recovery process reads the data from the temporary tables and performs a bulk insert operation of up to 2000 records into the actual historical tables.

In a duplex configuration, the recovery process keeps the historical data on the two Loggers in sync, using the recovery keys. The historical data between the Loggers is synced directly using actual tables; temporary tables are not used by the recovery process.

Replication

The Replication process is responsible for replicating data that has been committed to the Logger Central database to the HDS database.

The Replication mechanism consists of two processes: the Replication Server Process that runs on the Logger and the Replication Client Process that runs on the Distributor on which HDS has also been installed.

The Replication Client sends a request to the Replication Server requesting historical data that have associated Recovery Keys higher than those currently on corresponding historical tables. The Replication server sends the requested data back as a set of 2000 records each time.

The Replication server reads the historical data from the actual tables on the Logger and sends it to the Replication Client, which writes the historical data to the actual corresponding tables in the HDS database. Temporary tables are not used to replicate the data from the Logger to the HDS.

Possible Points of Delay or Inconsistency

If the Logger connected to the HDS goes offline, the HDS does not connect to a different Logger. For example, if the HDS is connected to Logger B and Logger B fails, the HDS does not connect to Logger A. When Logger B comes back up, it recovers data from Logger A and begins to receive current historical information. Once the Logger has recovered all of the data from Logger A, it begins to replicate this data to the HDS.

If reports are run from this HDS for recent intervals while the Logger is offline or while the Logger is in the process of recovering or replicating data, you might not see data for those intervals in reports. This situation is temporary, and you will see the data once the replication process for the tables used by the reports is complete. If you are using a fault-tolerant system with two HDS Administration & Data Servers, you can run reports using the backup HDS while the primary HDS is not receiving data.

If the HDS goes offline and you are using a fault-tolerant system with two HDS Administration & Data Servers, you can run reports using the backup HDS. When the HDS comes back up, it recovers data from the last HDS data backup and also replicates data from the Logger for the most recent data not available in the backup.

The recovery data replication is faster than regular Logger-HDS data replication. Once the HDS has recovered to its typical Logger-HDS latency of one to five minutes, data replication proceeds as usual.

If you are not using a fault-tolerant system, you will not see data in historical reports until the HDS is restored. You might also notice missing data as the replication process is in progress. This situation is temporary and you will see the data once the replication process for the tables utilized by the reports is complete.

Methods to Prevent Data Loss from Logger and HDS Failure

Data loss manifests as *data holes*, which are one or more missing records in an historical database table.

There are two types of data loss: temporary and permanent:

- A temporary data hole can happen during the Logger recovery process. For example, Logger A goes down, then comes back up and contacts Logger B to synchronize and recover historical data that was written while it was down.

  While this recovery process is going on, the reporting database on Logger A may have temporary data holes, which will be filled when the recovery process completes.

- A permanent data hole can happen during an Emergency Purge. For example, there can be permanent data loss if an emergency purge deletes records on one Logger that have not been sent to the other Logger or to the HDS.

It is possible to monitor and tune Unified CCE to minimize the occurrence of data loss.

Fault Tolerance

To protect your system, see the information on duplexed Unified CCE fault tolerance in the *Administration Guide for Cisco Unified Contact Center Enterprise*.

Data Retention and Backups

Another way to safeguard against loss is to configure the amount of time that data is stored on the Logger Central Database and in the HDS in relation to the schedule for HDS backups.

The Central database stores data for less time than the HDS. For example, you might store two weeks of data on the Logger and a year of data on the HDS.

When the HDS recovers after going offline, it retrieves all of the data on the Logger for the interval for which data is missing from the backup. You must manually restore the rest of the data from the last HDS backup.

The amount of data retained on the Logger should cover, at a minimum, the time period between HDS backups. For example, if the Logger stores data for two weeks, then you need to back up at least every other week to ensure that you can recover all historical data.

CPU Utilization

It is possible that the process on one of the Loggers is slow because of space issues or an overload of the SQL Server. In this situation, the data on the Logger with the slower SQL Server will lag in persistence of the historical data with respect to the other Logger. This causes the HDS on the corresponding side to lag as well.

As a consequence, if both sides have an HDS set up and the same reports are run from both HDSs, the reports might differ. This is usually a temporary inconsistency, since the condition that causes the SQL server process to slow might be remedied. Autogrowing of the database and load conditions often remediate. The Loggers and the HDSs eventually catch up and are in sync. Running the reports later will result in consistent reports.

However, it the database server runs of disk space, the situation is quite serious and might cause data to be out of sync for a longer duration until the problem is remedied. A permanent loss of data can occur when data is purged from the peer Logger and never replicated on the slower side.

Scheduled Purge and Retention Settings on Loggers

The goal of the scheduled purge is to free up database space by purging the oldest data.

There are several reasons for data loss during a scheduled purge:

- **Retention settings on Loggers**

  Data inconsistencies and permanent data loss can occur if the number of days to retain the data differs on the Loggers.

  Assume that Logger A is set to retain 7 days' worth of data, while Logger B is set to retain 15 days worth of data.

  If Logger B is down for 6 days, a temporary data discrepancy exists when it is brought back up, until the Recovery process synchronized the data from Logger A. However, if Logger B is down for 10 days, when it comes back up, it can synchronize only the last 7 days worth of data, based on Logger A's retention setting. Three days are lost permanently from Logger B.

  

  Note The data might be lost from the system permanently, if the historical data was copied to the HDS database associated with Logger A. Although this situation appears as a discrepancy in the reports that are run from HDS servers that connect to side B, the system is functioning in a predictable manner. It can be considered as an issue of perception.

  To avoid this situation, make sure that the retention settings are the same on both Loggers are the same.

- **Scheduled purge and Peripheral Gateway failure**

  If multiple Peripheral Gateways (PGs) are configured, and if one of the PGs goes down for a brief period, it is possible to lose historical data permanently.

  Assume that there are three PGs in the system and that one goes down for a day and then comes back online. When that PG comes back online, it sends historical data for activity that occurred prior to it going offline.

  If the scheduled purge mechanism activates and determines that the oldest one hour of data needs to be purged, it is possible that the purge will delete data that was sent by the PG after it came online but before it was replicated to the HDS.

  Permanent data loss can occur is the HDS is down and the scheduled purge on the Logger deletes data that has not yet been replicated to the HDS.

Emergency Purge

The emergency purge mechanism is triggered when the Logger Central Database becomes full or reaches a configured threshold size. Its objective is to free up space by purging data from the historical tables so that the database has more free space than the allowed minimum.

The emergency purge goes through each historical table in a predefined order one at a time and purges one hour's worth of data from the table. As data is purged from each historical table, a check is made to verify if the free space is more than the minimum threshold value. Once adequate space has been recovered, the emergency purge procedure stops. Otherwise, it continues through to the next historical table and keeps looping as necessary.

Permanent loss of historical data can occur if the emergency purge removes historical data that has not yet made it to an HDS and has also not been replicated to the peer Logger that is “down” or in the recovery process.

Database used percentage is displayed as a normal status message in the replication process every few minutes. You can occasionally monitor this value to make sure that is does not grow too often or too fast. Emergency purge occurs when the percentage used is greater than the configured value (usually 90%).

Data Loss from PIM Failure and Reporting

Here are some reporting considerations when you experience data loss from PIM failure.

The Peripheral Interface Manager (PIM) is the process on the Peripheral Gateway responsible for the actual connection to the peripheral and for normalizing the CTI interface on behalf of Unified CCE .

If a PIM fails, if the link between the PIM and the ACD goes down, or if the ACD goes down, then all of the reporting data that has been gathered for the peripheral associated with the PIM is deleted.

When the PIM failures occur, the peripheral is marked offline to the central controller.

The state of all agents on that peripheral is set to *logged out* and is reported as such to the CallRouter.

The CallRouter has no way of determining what was going on at the ACD while the PIM was out of contact with the ACD. When the PIM reconnects to the ACD, the ACDS does not send the PIM sufficient information to allow the recording of accurate historical reporting data for the interval(s) in which the disconnect took place.

Note When the PIM reconnects to the ACD, most ACDs do pass information to the PIM about each agent's state and duration in that state. While this is not enough to allow accurate historical reporting data to be recorded, it is enough to allow the CallRouter to make accurate call routing decisions.

When the PG is duplexed, either the Side A or Side B PIM is active for each peripheral. If one side loses connection, the other comes up and activates.

Other Possible Points of Failover

Peripheral Gateway / CTI Manager Service Failover

If the agent's PG shuts down or the CTI Manager service shuts down, the agent is momentarily logged out. The agent might be logged in again automatically once the backup PG or CTI Manager comes into service. The agent Media Logout Status reports for the agent, agent skill group, agent team, and agent peripheral show a logout reason code of 50002.

Table 1: Agent State Before and After Peripheral Gateway/CTI Manager Service Failover

Agent State at Fail-Over	Agent State after Fail-over
Available	Available

Agent State at Fail-Over	Agent State after Fail-over
Not Ready	Not Ready
Wrap-up	Available, if in Available state before the call. Otherwise, the agent reverts to Not Ready.

Agent Desktop/Finesse Server Failover

If the agent desktop (Finesse desktop) shuts down or loses communication with Finesse server, or if the Finesse server shuts down, the agent is logged out of all MRDs supported by the peripheral that has lost communication with the contact center software.

The agent is logged in again automatically when one of the following occurs:

- The agent desktop comes back up or resumes communication with the Finesse server
- The agent is connected to the backup Finesse server

The agent Media Logout Status reports for the agent, agent skill group, agent team, and agent peripheral show a logout reason code of 50002.

The state to which the agent reverts after failover depends on the agent's state when the failover occurred, as described in the following table.

Table 2: Agent State Before and After Agent Desktop/Finesse Server Failover

Agent state at failover	Agent state after failover
Available	Available
Not Ready	Not Ready
Reserved	Available
Wrap-up	Available, if in Available state before the call. Otherwise, the agent reverts to Not Ready.

Application Instance / MR PG Failover

If the connection between the Application Instance and MR PG shuts down or either component shuts down, the Central Controller discards all pending NEW_TASK requests received from the application.

The Application Instance waits for the connection to be restored and continues to send messages regarding existing tasks and new tasks assigned by the Application Instance to the Agent PG CTI server. When the connection, MR PIM, or Application Instance is restored, the Application Instance resends any pending NEW_TASK requests for which it has not received a response from the Central Controller. The tasks that are assigned to the agent by the Application Instance while the connection is down and completed before the connection is restored do not appear in reports.

Note If the Application Instance shuts down, this situation also affects Agent PG CTI server connections.

If the connection between the MR PIM and the Central Controller shuts down or the Central Controller shuts down, the MR PIM sends a ROUTING_DISABLED message to the Application Instance that causes the Application Instance to stop sending routing requests to the Central Controller.

Any request sent while the connection is down is rejected with a NEW_TASK_FAILURE message. The Application Instance continues to send messages regarding existing tasks and new tasks assigned by the Application Instance to the Agent PG CTI server.

When the connection or Central Controller is restored, the MR PIM sends the Application Instance a ROUTING_ENABLED message that causes the Application Instance to start sending routing requests to the Central Controller again. The tasks that are assigned to the agent by the Application Instance while the connection is down and completed before the connection is restored do not appear in reports. If the connection between the Central Controller and the MR PG fails, the CallRouter deletes all pending new tasks. When the connection is restored, the application connected to MR PG will resubmit all the tasks.

Note If the Central Controller shuts down, this situation also affects the Application Instance/Agent PG CTI server interface.

Application Instance / Agent PG CTI Server / PIM Failover

If the connection between the Application Instance and Agent PG CTI server shuts down or either component shuts down, agents stay logged in. Tasks remain for a time, based on the task life attribute of the MRD. If the task life expires while the connection is down, tasks are terminated with the disposition code of 42 (DBCD_APPLICATION_PATH_WENT_DOWN).

Note For the email MRD, agents are not logged out automatically when the Agent PG CTI server or connection to CTI server shuts down. Instead the email Manager continues to record agent state and assign tasks to agents. When the connection is restored, the email Manager sends the updated agent state information on the peripherals serviced by the Agent PG CTI server to the CTI server, which sends the information to Unified CCE software. The software attempts to recreate historical data and corrects current agent state. If the connection or Agent PG CTI server is down for more than the time limit configured for the MRD, reporting on tasks might be ended prematurely and restarted with the connection is reestablished

The application instance can assign tasks to agents while the connection or CTI server is down and, if the connection to the MR PG is up, can continue to send routing requests to the central controller and receive routing instructions. However, no reporting data is stored for the tasks while the connection is down. Also, any tasks that are assigned and completed while the connection or CTI server is down do not appear in reports. If the connection between the Agent PG CTI server and the CallRouter shuts down or if the CallRouter shuts down, the application instance continues to send messages to the CTI server and agent activity is tracked. However, this information is not sent to the CallRouter until the connection or the CallRouter is restored, at which time the cached reporting information is sent to the central controller.

Note If the Central Controller shuts down, this situation also affects the Application Instance/MR PG interface.

If the PIM shuts down, voice media routing is unavailable for agents associated with the PIM. However, the Central Controller can continue to assign non-voice tasks to agents associated with the PIM, and the CTI server can continue to process messages and requests about agents associated with the PIM for non-voice MRDs. When the connection is restored, voice media routing is available again.