

1.0 Data Recovery

Data Recovery is a process to restore the data in case of accidental deletes, intentional deletes from the tables/schemas/databases, any issue due to loading of data in the tables, . There are different ways to recover the data for tables for different databases. But for snowflake there is a unique feature “TIME TRAVEL” to recover the data for the tables.

2.0 Snowflake Time Travel

Snowflake Time Travel enables accessing historical data (i.e. data that has been changed or deleted) at any point within a defined period. It serves as a powerful tool for performing the following tasks:

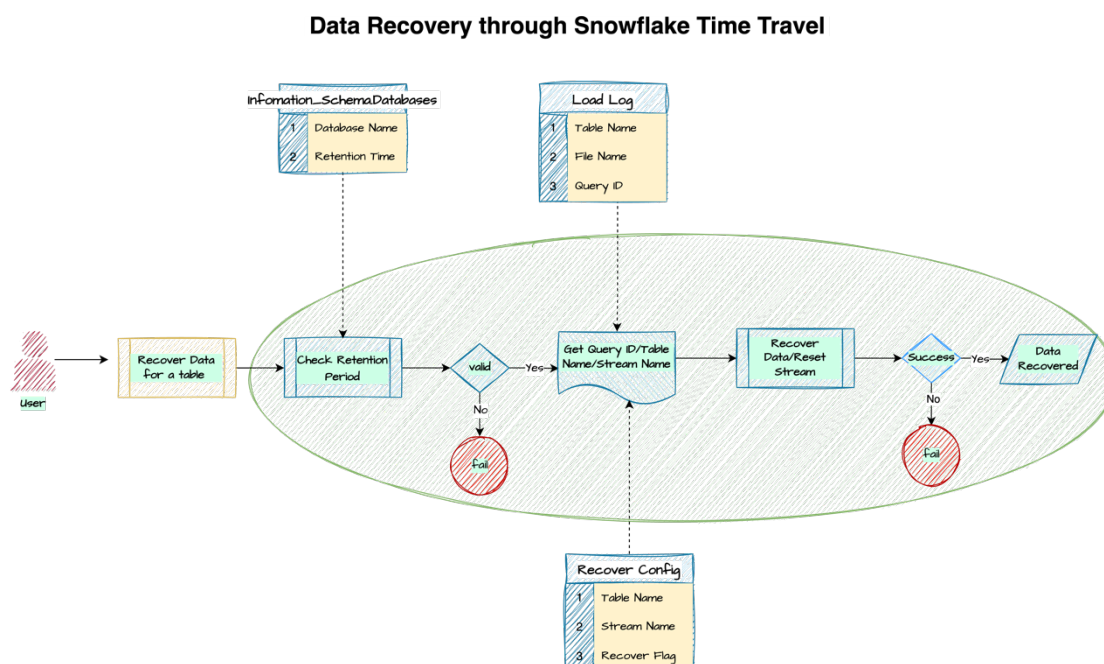
- Restoring data-related objects (tables, schemas, and databases) that might have been accidentally or intentionally deleted.
- Duplicating and backing up data from key points in the past.
- Analyzing data usage/manipulation over specified periods of time.

3.0 Data Recovery Process in PX Analytics

Recovery Process for PX Analytics Datalake has been designed with Snowflake Time Travel feature. PX Analytics Data Lake has mainly 3 set of tables RAW/STG/TRFN. The load sequence of the tables are as follows:

Source Files → RAW Table → STG Table → TRFN Table
→ Reset Streams

The recovery process has been explained with the following figure.



The recovery process of the tables are controlled through the following native components of the Snowflake.

- Snowflake Log tables of the PX Analytics Datalake (ING_PX_LOAD_LOG).
- Snowflake Config Table (C_ING_PX_RECOVER_CONFIG).
- Snowflake Stored Procedure (SP_RECOVER_DATA).

3.1 Snowflake Log Table

This table holds all the logs for RAW table loads with the query id for loading. This query id is used in recovering the data for all the table in the set (RAW/STG/TRFN). The following figure shows the table and values of the fields in the table.

Field Name	Data Type	Nullability
TABLE_NM	Varchar	Yes
FILE_NAME	Varchar	Yes
CODE	Varchar	Yes
MESSAGE	Varchar	Yes
QUERYID	Varchar	Yes
QUERYTEXT	Varchar	Yes
SNAPSHOT_DT	Date	Yes
FILE_TS	Varchar	Yes
FILE_SEQ	Varchar	Yes
LOAD_TS	Timestamp	Yes
LOAD_STATUS	Varchar	Yes

3.2 Snowflake Recover Config Table

This table holds the values and dates of the tables to be configured for recovery. In this table the configuration values to maintained are the set of table names, stream name (if exists) and a flag (T/F) to identify if the table needs to be recovered or not. The following figure shows the table and values of the fields in the table.

Field Name	Data Type	Nullability
TBL_ID	Varchar	Yes
RAW_TBL_NM	Varchar	Yes
STG_TBL_NM	Varchar	Yes
TRFN_TBL_NM	Varchar	Yes
STRM_NM	Varchar	Yes
RECVR_FLG	Varchar	Yes

3.3 Snowflake Stored Procedure

- The stored procedure controls the entire recovery process. The stored procedure expects 4 parameters “Database Name”, “Source Schema Name”, “Target Schema

Name”, “Load Date”. First it reads the Information_Schema.Databases table to get the retention time.

- If retention time is \leq the date provided for recovery, then it will go for the next step for recovery or else it will stop.
- If recovery is feasible w/ to the retention time, then it reads the config table and identify the tables/streams to be recovered.
- It reads the load log table w/ to the values fetched from the config table and identify the load date from the parameter passed. It fetches the query id on the basis of the date passed and the data will be recovered for all the tables as specified in the config table before the specified query id fetched on the basis of corresponding load date passed in the parameter.
- It will reset the streams corresponding to the tables that has been recovered.