ASSUMPTIONS

- IDs in the reference list of pickup locations are congruent with IDs in the complaints files, once they both came from the 4Flow system.
- IDs in the delivery events files are not congruent with both reference and complaints data.
- Delivery events and complaint files are in two separated folder in the central file system.
- Pickup stations that are removed from the reference table are deactivated in the dimensional SCD2 table.
- ID or fingerprint (name + address) identify a pickup station.

Data sources

List Pickup Stations

Central file system

/delivery_event/
XML | CSV
JSON

/complaints/
XML | CSV
JSON

Copies files from central file system into ADLS updating the *landing.landing_file* table

Trigger (schedule) >>> Get Metadata >>> ForEach file >>> Copy Activity + Script Sctivity (INSERT)

Landing zone

Daily runs a SQL query to bring data from the reference table into a staging table.

/landing/delivery_events/ <timestamp>/<filename> or */landing/complaints/ <timestamp>/<filename>*

Python parser in a container or function

Parses CSV/XML/JSON file content, standardizing columsn names, normalizing text data and calculating registry hashs.

Trigger (schedule) >>> LookUp RECEIVED >>> Condition >>> Custom activity (Python)

Inserts data into *landidng.landing_file*

Updates *landidng.landing_file* changing RECEIVED to PARSED/FAILED according to its status.

Staging zone

Inserts data into *staging.stg_pickup_station*

Daily runs a sequence of SQL queries to add new stations, update existing stations and deactivate removed stations.

Inserts data into *staging.stg_delivery_events*

Inserts data into *staging.stg_complaints*

DW zone

Updates data into *dw.dim_pickup_station*

After updating *dw.fact_complaints*, updates a pickup station data integration monitoring table

Updates data into *dw.monitoring_complaints*

Daily, after the *dim_pickup_station* was updated, updates the *dw.fact_complaints* and *dw.fact_delivery_events* inserting new entries.

Updates data into *dw.fact_complaints*

Updates data into *dw.fact_delivery_events*

REPORT 1 (VIEW) Identifier, Name and Zip Code of Pickup locations affected by claims together with the according number of claims within the last 14 calendar days

REPORT 2 (VIEW) Name and address of final recipients with more than 5 complaints in the last two months and the given pickup locations of the according complaints per recipient