# Project\_1

**Type 1 data load to a database**

* File will be placed in an inbound folder path
* File will be read and data will be written into a Dataframe
* Target data will be read (if exists) into a Dataframe
* A hash column will be added to both source and target dataframe.
* Both dataframes will be joined with the hash values by full outer join
* These data will be loaded to final table:

1. Not joined data of target dataframe (filter the data where source hash is null)
2. Not joined data of source dataframe (filter the data where target hash is null)
3. Joined data of source dataframe (filter the data where source and target hashes are not null)

* These data will not be loaded to final table:

1. Joined data of source dataframe

* New data will completely overwrite the older data (Truncate and load)