Assumption and Approach

Author: Yangyang Cai Date: 18/09/2021

Assumptions

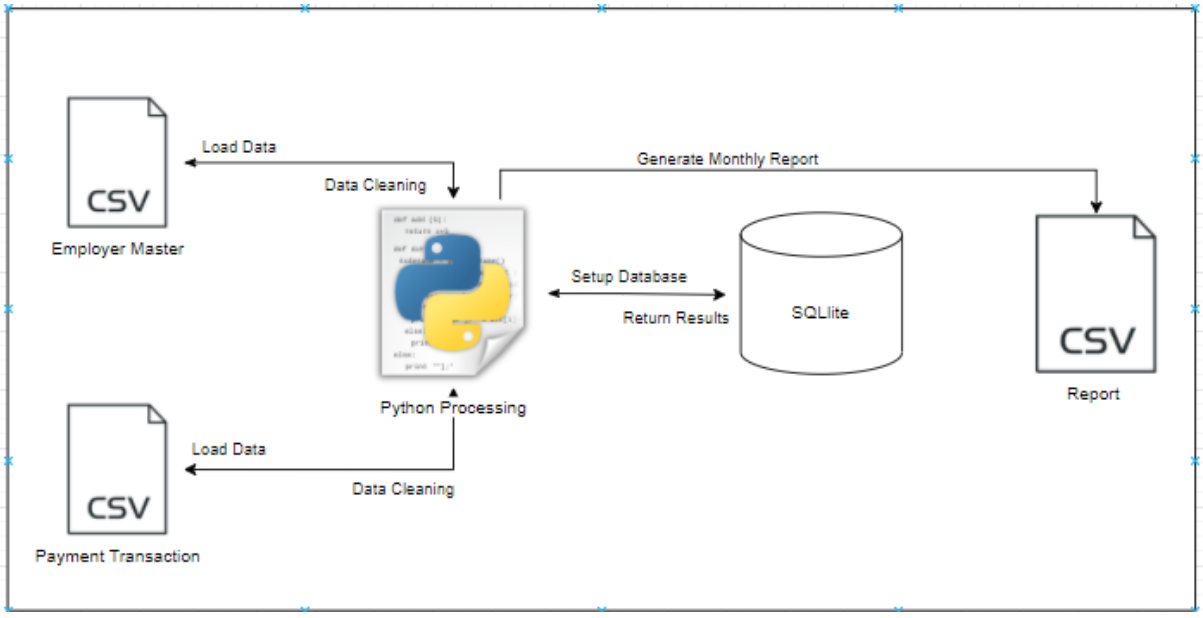
1. Employer Status means this employer's account is open between effective_from and effective_to period
2. Employer only has 4 type of tiers
3. Each employer may have multiple superannuation payments processed on the same day
4. Result report's month end date column needs to be %d/%m/%Y format
5. Result report's amount of payments column needs to keep 2 decimals

Approach

Folder Structure

- **Data Folder:** Employer master.csv, Payment transactions.csv, Payment transactions processed.csv, pythonsqlite.db
- **Result Folder**: monthly_employer_report.csv
- **Code Folder**: database_helper.py, Monthly Employer Analysis.ipynb
- Assumption and approach.pdf

Approach Pipeline



The approach designed is as database oriented, which has the following benefits:

1. **Repeatable:** You do not need setup SQLlite separately, just run Monthly Employer Analysis.ipynb, you could easily repeat the whole approach.
2. **Automation**: Data import, data processing and report generation have been automated.
3. **Flexible**: Provide dashboard helper module which enable you modify, extend, or customize your database.
4. **Extensible:**

a. Easy to add new report queries to Monthly Employer Analysis.ipynb to generate new reports if you want.
b. If you want to generate python dashboard based on the final csv result, you could just modify Monthly Employer Analysis.ipynb

The approach also has the following potential limits:

1. If we have high volume of data, SQLlite is lightweight database which may not be enough for data processing and data storage
2. If the query is very complex, running query using python is slower
3. The data validation needs some human check, it will be better to have some dashboards available could do cross checking