Senior data engineer take home

The goal of this task is to showcase a full ETL pipeline, from a datasource to a Data warehouse, preferably using python and SQL.

We would like to extract from the World Bank data related to the volume of remittances and migration between countries. In the end we would like this data to be inside a postgresql database that could answer this questions:

- Top 10 country_to_country by number of migrants
- Top 10 country_to_country by volume of remittances
- Top 10 sending countries
- Top 10 receiving countries
- Top 10 Net senders
- Top 10 Net receivers

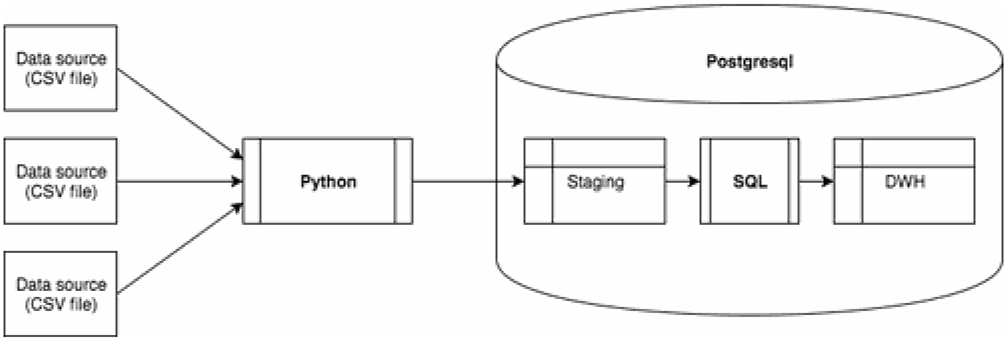
The datasets can be found here:

- http://www.knomad.org/sites/default/files/2017-11/bilateralremittancematrix2016_Nov2017.xlsx
- http://pubdocs.worldbank.org/pubdocs/publicdoc/2015/10/38881445543162029/bilateral-migration-matrix-2013-0.xlsx

This data can be downloaded manually and extraction to a better format (like CSV) can be manual, but to prepare data for the database, any data manipulation and transformation should be done using python. The structure on the staging area does not need to be an exact replica of the structure inside the excel file.

Once the data is in a database, any needed transformation should be done in SQL.

This is the proposed architecture for a solution:



Proposed tables for the staging schema:

- country
- remittance
- migration

Proposed tables for the DWH schema:

- country
- corridor (from_country, to_country, remittance_value, migration_value)

This solution and database schemas do not have to be used, you can present a completely different way to solve this problem. Just remember that we are looking for data engineers that should be able to move data around and that any solutions presented should be able to be automated, scalable and maintainable.

And although we do like and use these technologies, we also need to be able to work without them (and often do in production code), so please:
- No pandas.
- No python notebooks.
- Keep it simple & understandable

BONUS (really not a requirement): Import this data, for all available years (available here: http://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data) and adapt the schema accordingly.

Deliverables: Python code, SQL code, description of the implemented solution, answers to the questions asked.