The following is the architecture diagram for the project. All the input files are expected to be in CSV format, so I tried to use pandas for faster file read.

Currently shared memory is used for storing data across process. It will not work once data becomes huge.

I tried using TinyDB as a lightweight data base. But it does not work with Parallel systems.

Code standard: Currently code is performed using PEP8 standard.

Checking is done using following

Isort \*.py

flake8 --ignore=E501 \*.py

**Test:**

A test module is there to test stuff.

**Run:**

python worker.py -i /Users/sandipnandi/MYASSN/Assignments/percolate/back\_end\_rolodex/data.in -o output.json

**Design.**

