Rebuilding WiNDC

Adam Christensen
May 28, 2019

Big Picture

- Fully open source
- Modular/Robust
- Documented

How are we achieving these goals?...
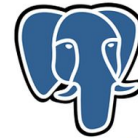
Fully Open Source...

Previous Version

- GAMS w/ NEOS integration
- Many GAMS utilities
  - GDXXRW
  - CSV2GDX
  - XLSDUMP

Rebuild

- Python 3
- PostgreSQL
- GAMS w/ NEOS integration
- Julia/JuMP
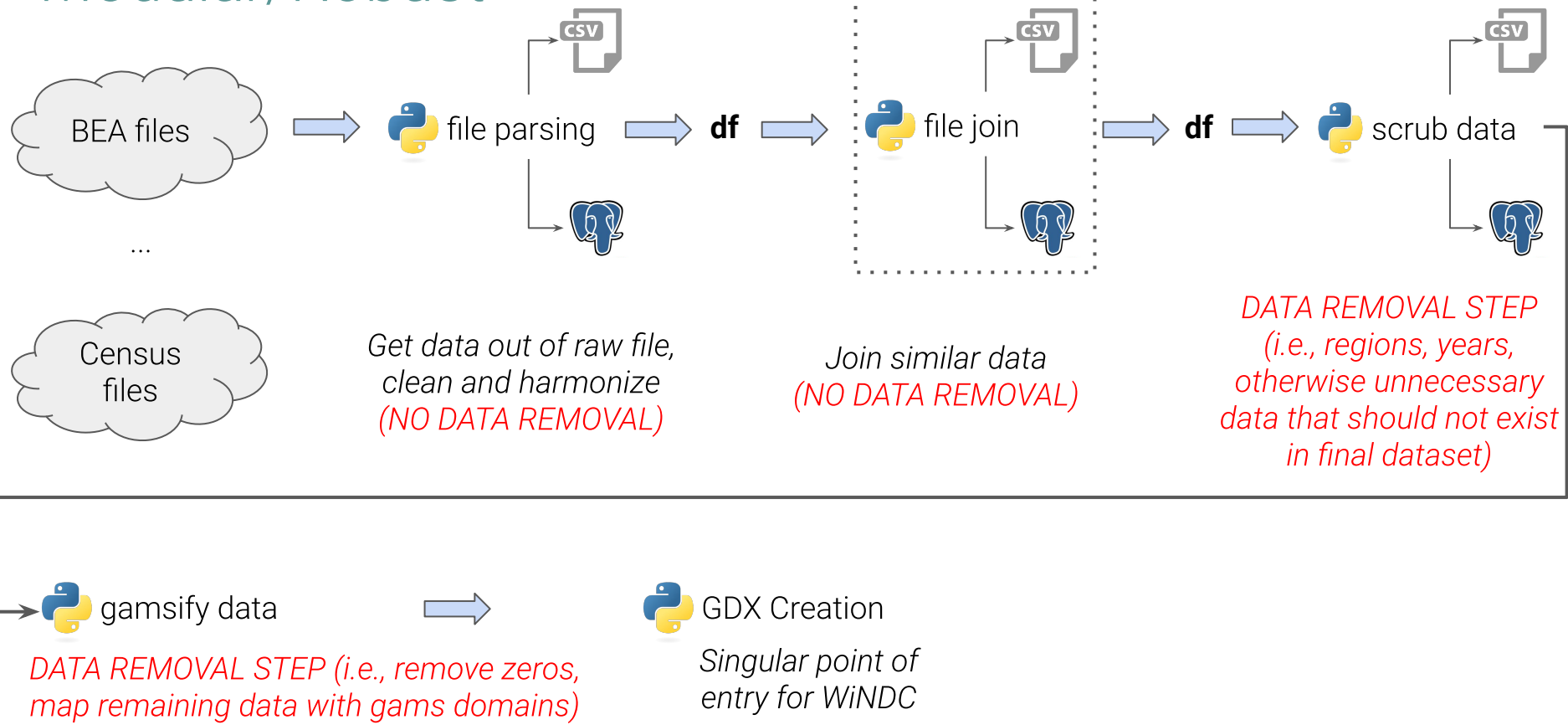- Auto PEP8 style

Modular/Robust

Previous Version

- Modular design w/ caveats
    - Data processing and model solution were interconnected
    - No singular point of entry for data (i.e., years, states, etc.)
- GAMS did all the heavy lifting
    - GAMS is not a data processing language
    - No data typing in GAMS (not as robust)
    - Relabeling/harmonization is possible with maps, but syntax is hard to parse
    - No metadata

Rebuild

- Python is an ideal for data processing
    - Openpyxl for reading in Excel files
        - Does not read .xls files only .xlsx
    - Python/Pandas for reading in CSVs
    - Lots of data cleaning tools available
    - More human readable
    - Data typing is straightforward
    - Relabeling is easy with .map methods
- Bright lines between major data process steps
- Easy Python/PostgreSQL integration

Modular/Robust

BEA files

...

Census files

file parsing

Get data out of raw file, clean and harmonize (NO DATA REMOVAL)

df

file join

Join similar data (NO DATA REMOVAL)

df

scrub data

DATA REMOVAL STEP (i.e., regions, years, otherwise unnecessary data that should not exist in final dataset)

gamsify data

DATA REMOVAL STEP (i.e., remove zeros, map remaining data with gams domains)

GDX Creation

Singular point of entry for WiNDC

Documentation

- We are committed to documentation
  - Script log generation
  - User manuals
  - Maintenance of these documents