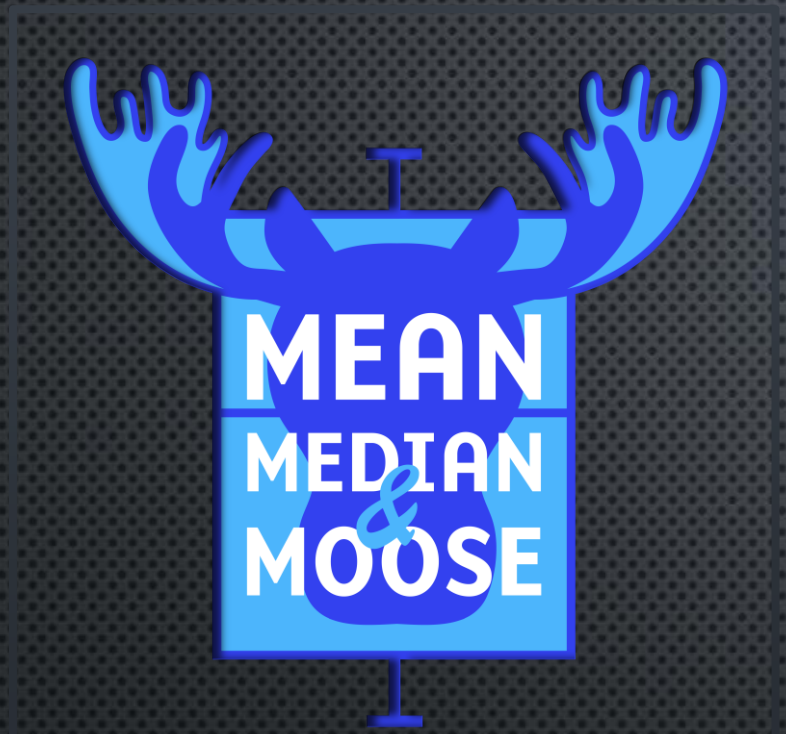
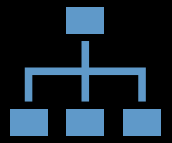


ETL WITH PYTHON

Border City Data

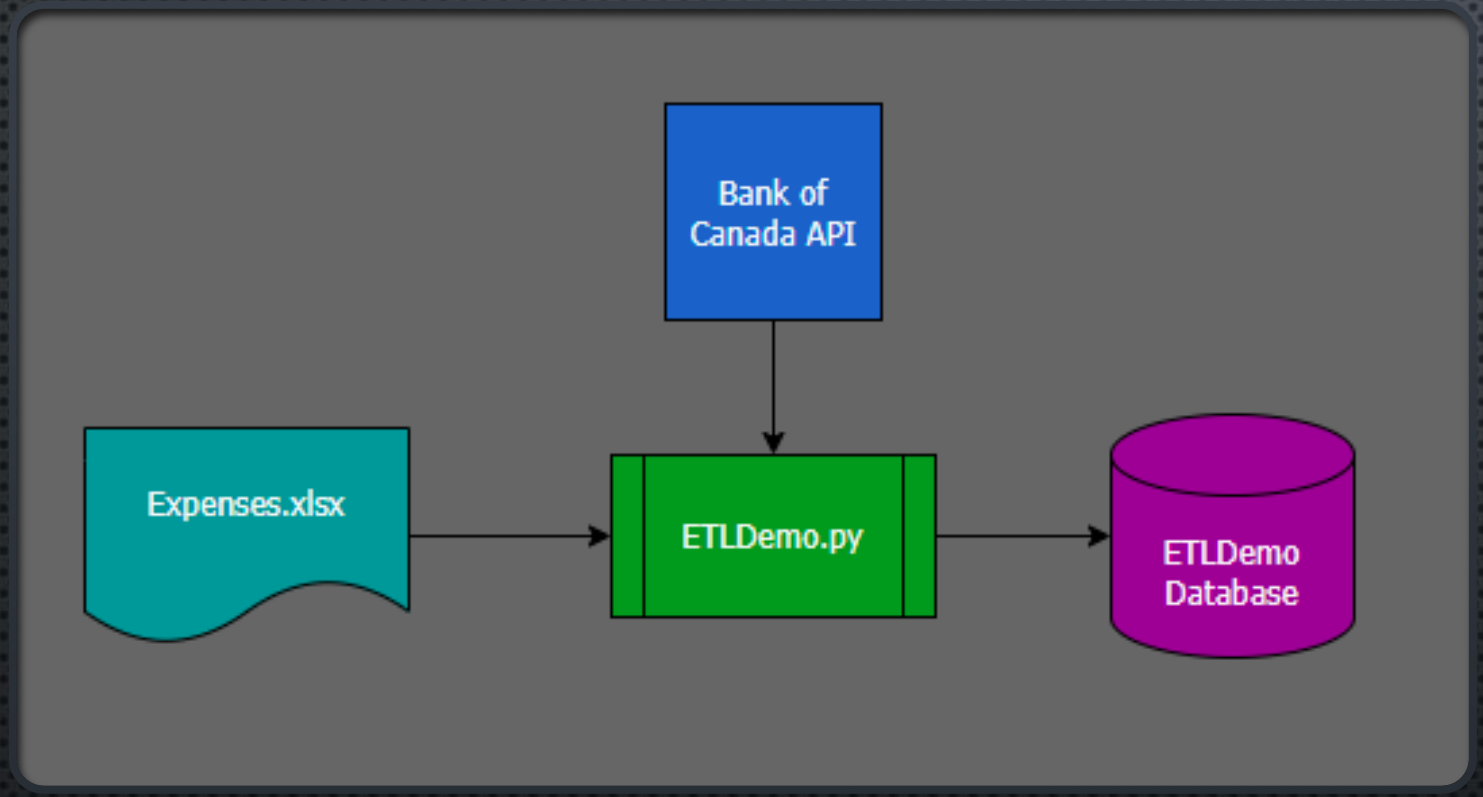
Extraction of data from one or more sources



Transformation by cleansing, aggregating, standardizing, applying business rules…

Load data into target system

"EXTRACT, TRANSFORM AND LOAD"



PROCESS FLOW

BANK OF CANADA VALET API

"Programmatic access to a range of global financial data"

- Hundreds of data sets
- Exchange rates
- Currency statistics
- Economic data
- Survey responses

Base URL https://www.bankofcanada.ca/valet

Formats XML, JSON, CSV

EXCHANGE RATE DATA

DAILY USD/CAD RATES

HTTPS://WWW.BANKOFCANADA.CA/VALET/OBSERVATIONS/FXUSDCAD/JSON?START_DATE=


```
{
"terms":{
    "url": "https://www.bankofcanada.ca/terms/"
},
"seriesDetail":{
"FXMUSDCAD":{"label":"USD/CAD","description":"US dollar to Canadian dollar monthly exchange rate"}
},
"observations":[
{"d":"2017-01-01","FXMUSDCAD":{"v":1.3193}},
{"d":"2017-02-01","FXMUSDCAD":{"v":1.3107}},
...
]
}
```

	A	B
	date	USD
	2020-01-11	$25.00
	2020-01-30	$7.00
	2020-02-11	$25.00
	2020-02-28	$7.00
	2020-03-11	$25.00
	2020-03-30	$7.00
	2020-04-11	$25.00
	2020-04-30	$4.00
	2020-05-11	$20.00
	2020-05-30	$4.00
	2020-06-11	$20.00
	2020-06-17	$3.20
	2020-06-30	$4.00

EXPENSES.XLSX

```
USE ETLDemo

DROP TABLE IF EXISTS Expenses

CREATE TABLE Expenses
(
    date datetime,
    USD money,
    rate DECIMAL(6,5),
    CAD money
)
```

TARGET DATABASE

In repository:
DemoDBDDL.sql

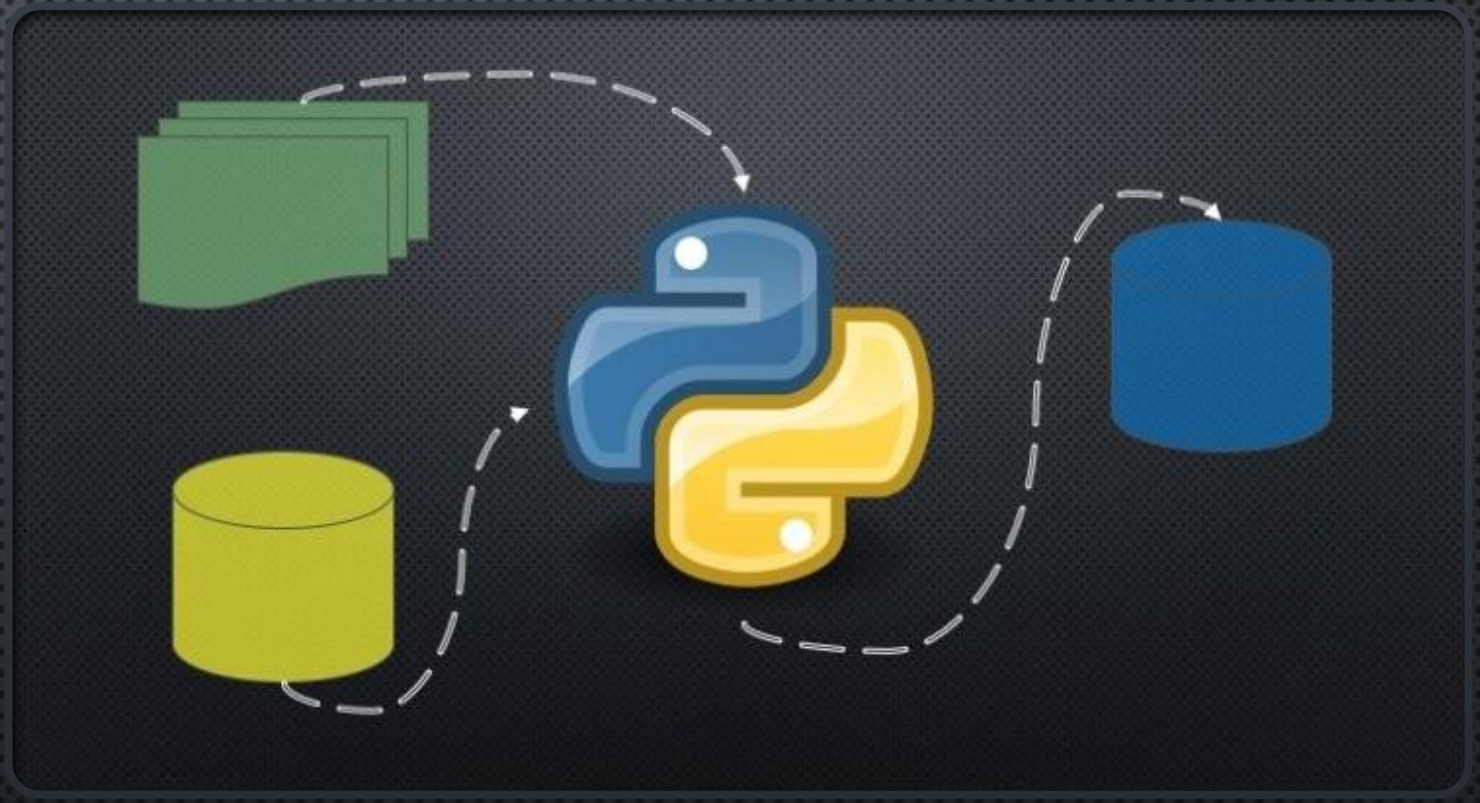
PETL

A general-purpose Python package for extracting, transforming and loading tables of data.

Docs:
https://petl.readthedocs.io/en/stable/

Source:

https://github.com/petl-developers/petl



Other options

- Pandas (pandas.pydata.org)
  - Data analysis with Python
- Bubbles (bubbles.databrewery.org)
  - Framework for data processing
- Bonobo (bonobo-project.org)
  - Minimalistic ETL toolkit
- Pygrametl (chrthomsen.github.io/pygrametl)
  - Data warehouse focused

ETLDEMO.PY

SOURCE:

HTTPS://GITHUB.COM/DSARTORI/ETLDEMO

```
import os
import petl
import pymssql
import configparser
import requests
import datetime
import json
import decimal
```

DEMO

CODING THE SOLUTION