



PORTO MAPPING COMMUNITY SPRING SESSION

Photo by [Dulcineia Dias](#) on [Unsplash](#)

COOL VISUALIZATION TOOLS AND ETL PROCESSES AT DOUROECI



João Coelho
DouroECI

LinkedIn: [jmcoelho](#) | Twitter: [@jcoelho_pt](#)

The goal of this presentation

To introduce you to the open source tools and ideas I will show you today

To motivate you to try and apply them on your projects

About me

- Geographic Engineer
- > 10 years helping to develop, implement and manage information systems for water management
- Since 2017 at DouroECI - Project Manager / Product Owner



joao.coelho@douroeci.com



[jmcoelho](https://www.linkedin.com/in/jmcoelho)



[jcoelho_pt](https://twitter.com/jcoelho_pt)

Who are we



Engineering | Consulting | Innovation

- A company with a clear focus on **water utilities management and optimization**
- Covering all water cycle: from **Engineering**, to **Consulting & Innovation**

Europe:
HQ Portugal

Middle-East:
Working across the region

Cape Verde

Africa:
Key presence in Portuguese Speaking Countries and working across the continent

Angola

Mozambique

We work across the globe, with an experienced and multidisciplinary team

Working in project development, implementation and optimization at **any scale**

METITO



- Global presence across water stressed markets in all segments of water & wastewater treatment
- Developed over **3000 successful projects** to date, including **34 concessions and long-term O&M contracts**
- Headquarters at Dubai, and regional offices in Sharjah, Cairo, Jakarta and Shenzhen.
- Strategic partnership with **Mitsubishi Group**, one of the key shareholders of Metito

strategic partner

DOUROECI



ENGINEERING



CONSULTING



INNOVATION

Hydraulic modelling

Water supply

ASSET MANAGEMENT

Water Resources Management

Wastewater treatment

NON-REVENUE WATER

GIS

Storm water drainage planning

INFILTRATION AND INFLOWS

Hydrology

Technology

Water quality

DESIGN

Water treatment

Institutional support

Information systems

www.douroeci.com

ETL & Data Viz at DouroECI

A new product

The development of an innovative **Maintenance and Asset Management System (M&AM)** for water utilities

M&AM for physical infrastructure continuously generate and requires a huge amount of spatial and temporal data, that needs to be combined with other data types and data sources

Critical requirements

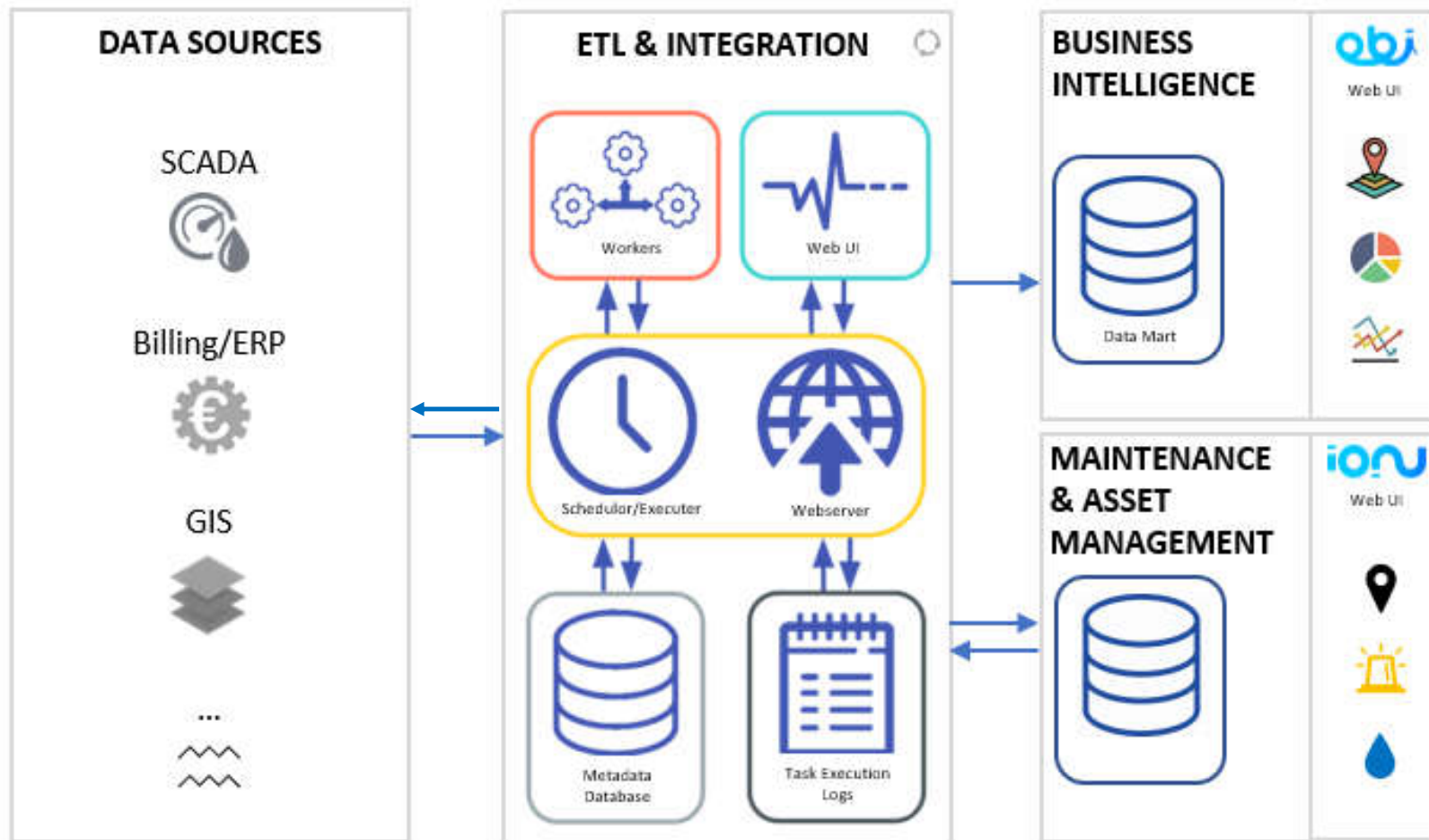
E
T
L

- To integrate with external data sources
- To collect, process and publish combined and enriched data
- To automate and monitor required processes

D
a
t
a
V
i
z

- To provide interactive and powerful data visualization and exploration tools

Architecture

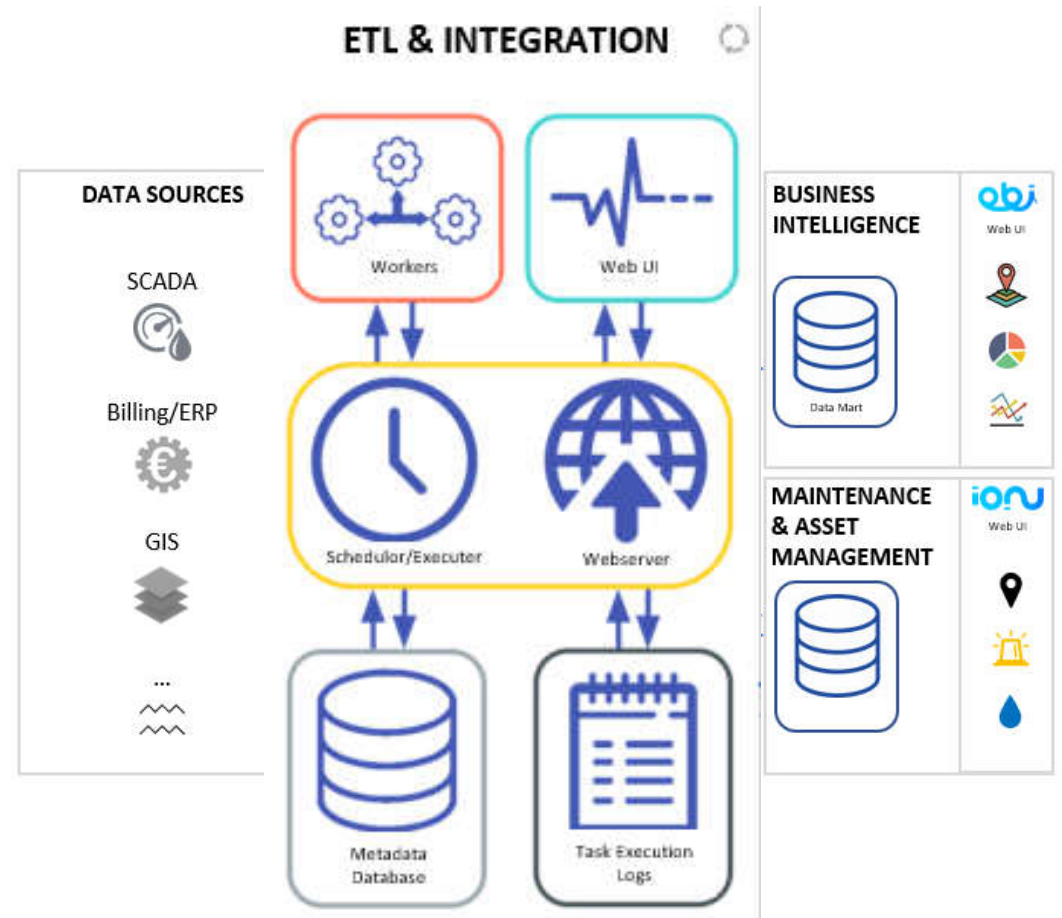


ETL TECH STACK

The technology stack

ETL for spatial and temporal data

- Apache Airflow (DAGs)
- Pandas (data analysis and manipulation)
- Geopandas (geoprocessing)
- Python (code)



Apache Airflow

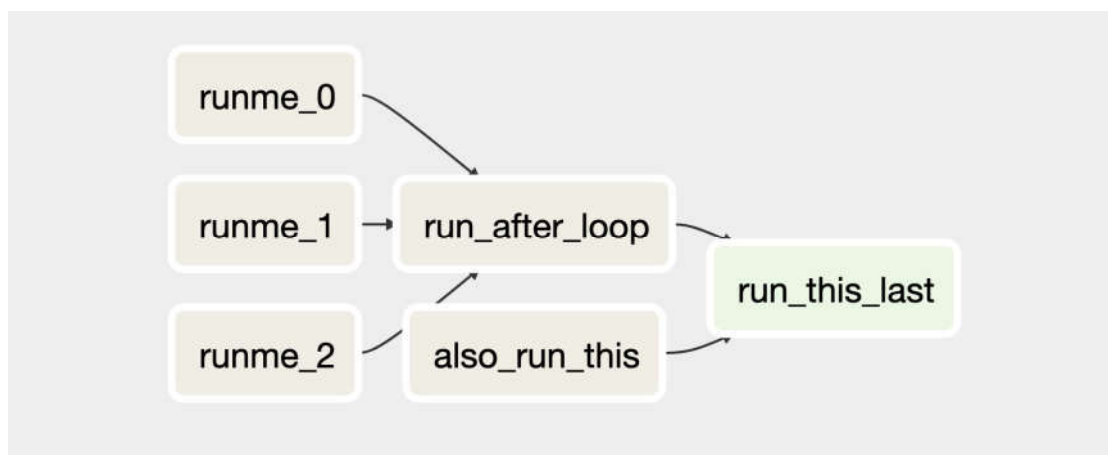
<https://airflow.apache.org/>

- Open source platform to programmatically author, schedule and monitor workflows of tasks
- Scalable to infinity (orchestrate an arbitrary number of workers)
- Pipelines configuration as code (Python), allows dynamic pipeline generation
- Useful UI to monitor, schedule and manage your workflows

Just google images for the expression [*airflow data example*](#)

Apache Airflow

| Airflow | | | | | | | | | | | |
|---------|----|-------------------------------------|-----------|------------------|------|-------|-------------------------|--|----|---|--|
| DAGs | | Data Profiling | Browse | Admin | Docs | About | 2020-06-04 00:00:16 UTC | | | | |
| | On | ION_BR_OrdensTrabalho_IntAbast | 25 20 *** | Python Operators | 5 | | 2020-06-02 20:25 | | 10 | 7 | |
| | On | ION_BR_OrdensTrabalho_Requisicoes | 15 20 *** | Python Operators | 5 | | 2020-06-02 20:15 | | 10 | 7 | |
| | On | ION_BR_OrdensTrabalho_Tarefas | 10 20 *** | Python Operators | 5 | | 2020-06-02 20:10 | | 10 | 7 | |
| | On | ION BR OrdensTrabalho causas conseq | 40 21 *** | Python Operators | 5 | | 2020-06-02 21:40 | | 10 | 7 | |



```
# -- Airflow packages
import airflow
from airflow.models import DAG
from airflow.operators.python_operator import PythonOperator
from airflow.operators.bash_operator import BashOperator
from airflow.contrib.operators.ssh_operator import SSHOperator

# -- Data manipulation packages
from sqlalchemy import create_engine
import datetime
import pandas as pd

# -- DAG arguments
args = {
    'owner': 'joao.coelho',
    'start_date': airflow.utils.dates.days_ago(1),
}

dag=DAG(
    dag_id='ION_MC_Sync_SAP',
    description='Synchronization process SAP --> ION Used Materials',
    default_args=args,
    schedule_interval='10 20 * * *',
)
```

Pandas

<https://pandas.pydata.org/>

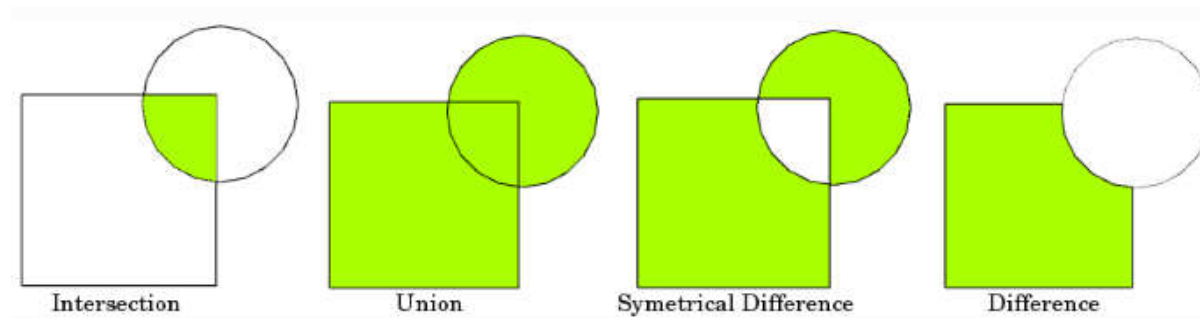
- A fast, powerful, flexible and easy to use open source data analysis and manipulation tool
- It offers data structures and operations to manipulate numerical tables and time series
- Built on top of the Python programming language

GeoPandas

<https://geopandas.org/>

GeoPandas extends the datatypes used by pandas to allow:

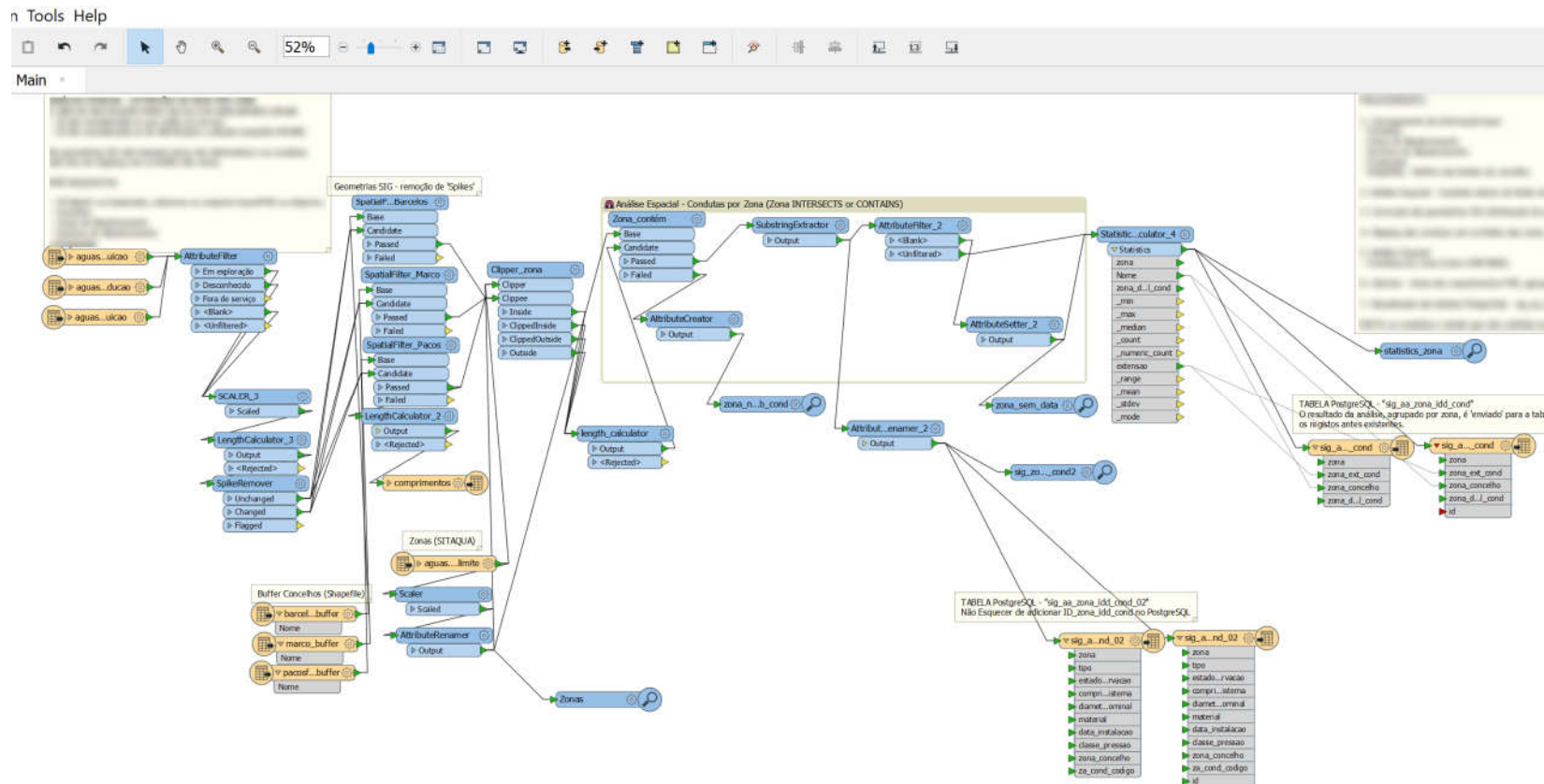
- Spatial operations on geometric types
- Create and plot maps
- Manage projections
- Set-Operations with overlay
- Aggregation with dissolve
- Merging data
- Geocoding



ETL EXAMPLE

ETL process with spatial data (an example)

The challenge: replace an ETL process developed with a proprietary software



ETL process with spatial data (an example)












Goal: Aggregate linear assets length by age, material and geographic area

Tasks:

- read linear features and geographic areas (from Databases, Webservices, Files, ...)
- clip by boundary
- calculate the length of linear features
- classify by geographic area (Measurement and Control Zones)
- aggregate length by age, type of material, and geographic area

Execution: daily [*schedule_interval='@daily'*]

ETL process with spatial data (an example)

| |  DAG | Schedule | Owner | Recent Tasks  | Last Run  | DAG Runs  | Links |
|---|--|----------|-----------------|--|--|---|---|
|  |  ION_AssetLength_vs_ZMC | @daily | Ignacio Ramirez |  | 2020-06-03 00:00  |   |  |

```
# -- airflow packages:
import datetime as dt
import airflow
from airflow.models import DAG
from airflow.operators.python_operator import PythonOperator
from airflow.operators.bash_operator import BashOperator

# --- main packages:
from sqlalchemy import create_engine
import pandas as pd
import numpy as np

# -- geo analysis packages:
import geopandas as gpd
from earthpy import clip as cl

# -- airflow arguments:
args = {
    'owner': 'Ignacio Ramirez',
    'start_date': airflow.utils.dates.days_ago(1),
}
```

```
# -- get asset:
asset = get_shapefile(system, asset_name)

# -- apenas ativos "Em Exploração":
asset = asset.loc[asset.ESTADO == 'Em exploração']
```

```
asset_zd['AssLength'] = asset_zd.to_crs({'init': 'epsg:3763'}).geometry.length
```

```
# -- classification
asset_zd = get_linestring_vs_polygon(
    asset, zd, 'zona_abastecimento')[['CODIGO', 'geometry', 'zona_abastecimento']]

asset_sa = get_linestring_vs_polygon(
    asset, sa, 'setor_abastecimento')[['CODIGO', 'geometry', 'setor_abastecimento']]
```

In the end, processed data is uploaded to a Data Mart where it can be accessed by data visualization and exploration tools

Our experience so far

- In one of our clients ~80 DAGs executed everyday
- Processing and combining data from GIS, Billing, ERP and CMMS
- Airflow has proven to be extremely useful
- However, you can come across certain pitfalls, which can cause occasional errors or DAGs not being executed as scheduled

We are still learning best practices and new ways to take advantage of the full potential of Airflow and showcased python libraries

DATA VIZ

Solutions

Web Mapping

- OpenLayers to M&AM Desktop and Mobile apps

Interactive Data Visualization & Exploration

- Apache Superset for Business Intelligence

OpenLayers

<https://openlayers.org/>

A high-performance, feature-packed library for all your mapping needs

FEATURES

Tiled Layers

Pull tiles from OSM, Bing, MapBox, Stamen, and any other XYZ source you can find. OGC mapping services and untiled layers also supported.



Cutting Edge, Fast & Mobile Ready

Leverages Canvas 2D, WebGL, and all the latest greatness from HTML5. Mobile support out of the box. Build lightweight custom profiles with just the components you need.



Vector Layers

Render vector data from GeoJSON, TopoJSON, KML, GML, Mapbox vector tiles, and other formats.



Easy to Customize and Extend

Style your map controls with straight-forward CSS. Hook into different levels of the API or use [3rd party libraries](#) to customize and extend functionality.



OpenLayers – Tile & Vector Maps

MENU

AC200340

REPARAÇÃO DE FUGA DE ÁGUA

ACTIVO: AdB - Rede AA

INÍCIO: 2020 Jun 03 13:30

VER

TAREFAS

ESCONDER MAPA

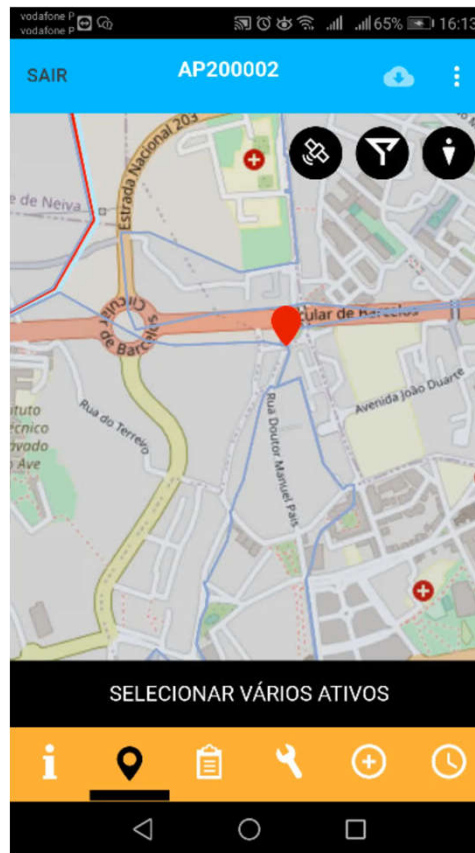
SELECIONE VÁRIOS ACTIVOS

● activos interveniçoados

● activos seleccionados

| ACTIVO | TIPO DE TAREFA | TAREFAS | ESTADO |
|-----------------------------|----------------|-----------------------------|----------|
| CD00087867 Fibrocimento ... | INTERVENÇÃO | REPARAÇÃO FUGA CONDUTA ÁGUA | Completo |
| VA00048791 | INTERVENÇÃO | MANOBRA DE VÁLVULA | Completo |

OpenLayers – Tile & Vector Maps



Apache Superset

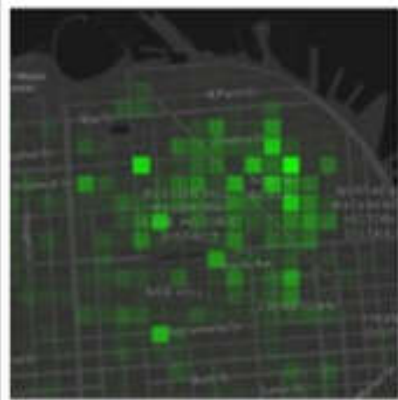
<https://superset.incubator.apache.org/>

- Data Visualization & Exploration tool initially designed by Airbnb and later open sourced for the community
- Designed to be visual, intuitive and interactive
- It allows to slice, dice and visualize data
- It's written in python and uses Flask as web framework library
- Includes **Mapbox & Deck.gl maps**
- It comes with advanced security features

Apache Superset – Maps



Deck.gl - Grid



Deck.gl - Screen grid



Deck.gl - Scatter plot



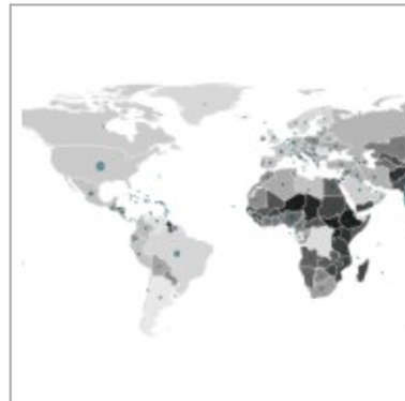
Deck.gl - Hexagons



Mapbox



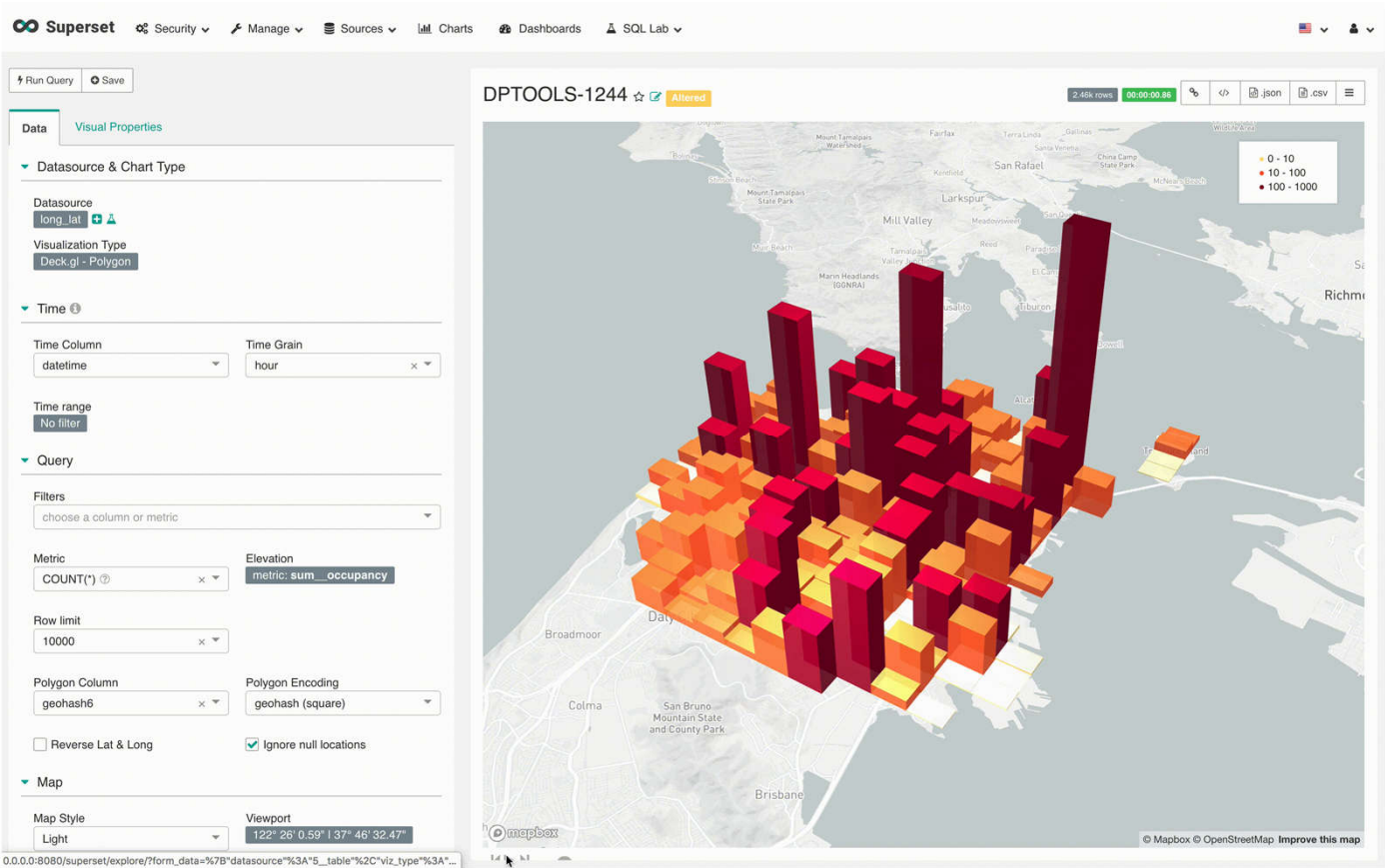
Country Map



World Map

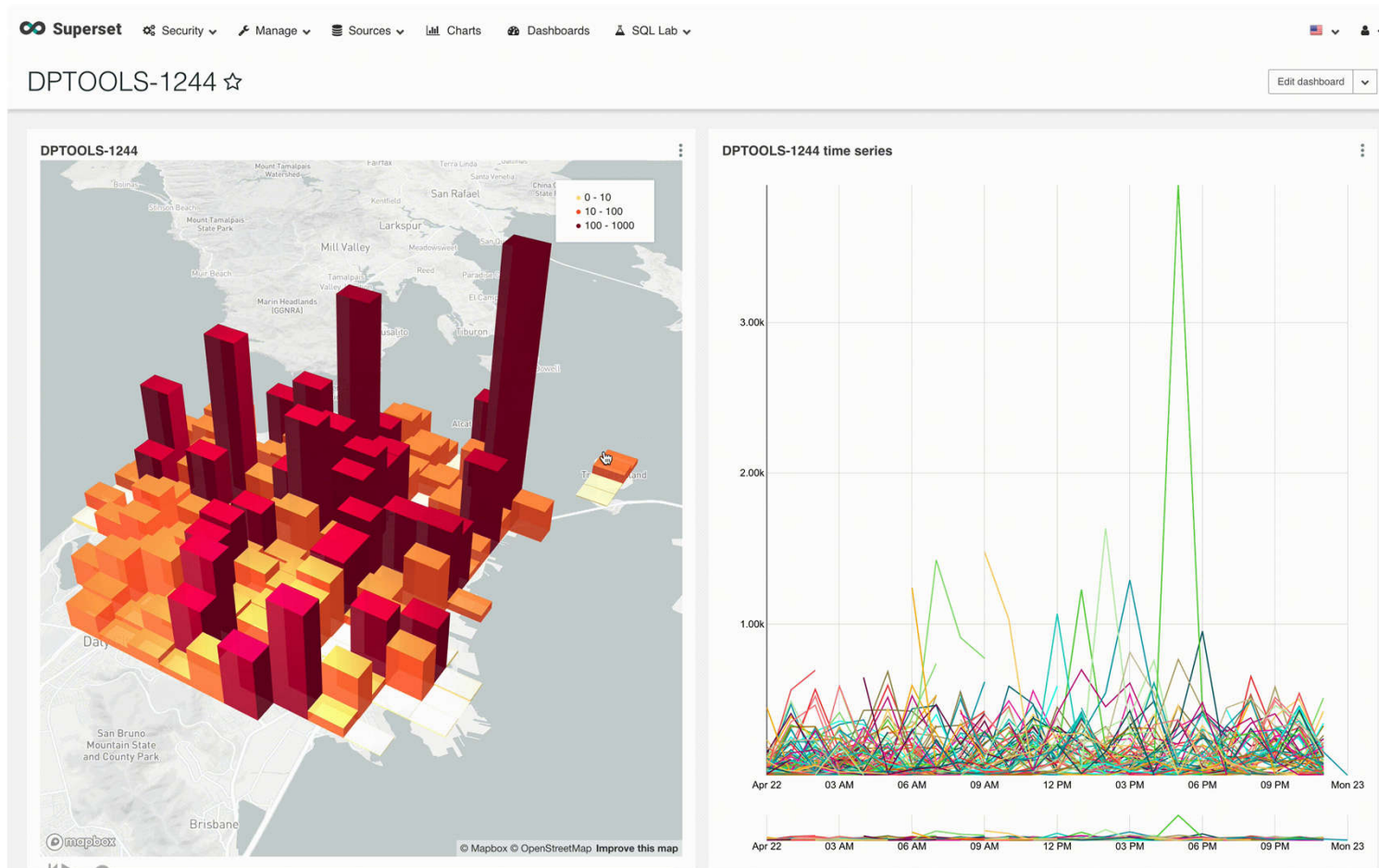


Apache Superset - Maps



The polygon spatial viz can be animated using the play slider.

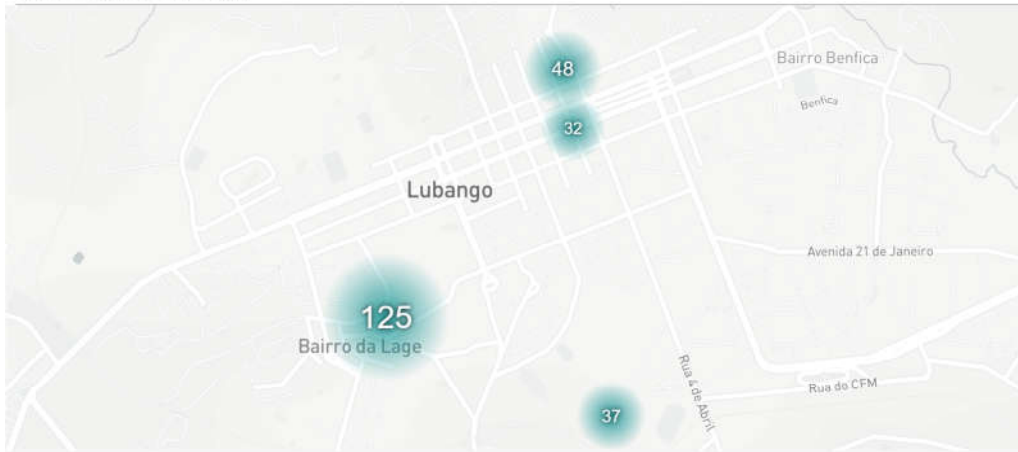
Apache Superset – Maps



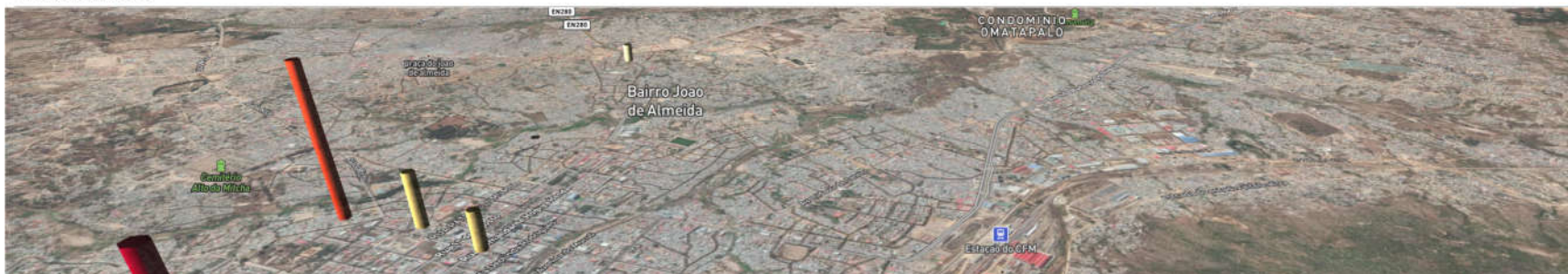
And interacts with another visualization

Apache Superset - Examples

Mapa - Duração total OT (h)



Mapa 3D N° de OT



Análise OT

Search:

| Sistema | Ativo | Tipo Ativo | Ordem de trabalho | Tarefa | Duração Real (h.) |
|----------|----------------------------|--------------------------|-------------------|-------------------------------|-------------------|
| Lubango | ZONA 06 - BAIRRO LAUREANOS | Rede | CO170065 | Correcção de fugas e rupturas | 61.0 |
| Quipungo | Cx. Manobras | Câmara/Caixa de manobras | CO180011 | NULL | 15.0 |
| Quipungo | Tubo em FFD DN80 | Tubagem | CO180012 | NULL | 15.0 |
| Chibia | Gerador Socorro 13kVA | Geradores | CO180013 | NULL | 15.0 |
| Quipungo | Reservatório 200 m3 | Reservatório | CO180010 | NULL | 10.0 |
| Lubango | ZONA 01 - BAIRRO MAPUNDA | Rede | PR170025 | NULL | 10.0 |
| Lubango | ZONA 07 - BAIRRO MARINGA | Rede | CO170122 | Correcção de fugas e rupturas | 9.00 |
| Lubango | Bairro Da | Rede | CO170007 | Correcção de fugas e | 8.00 |

Filtro

Since Until

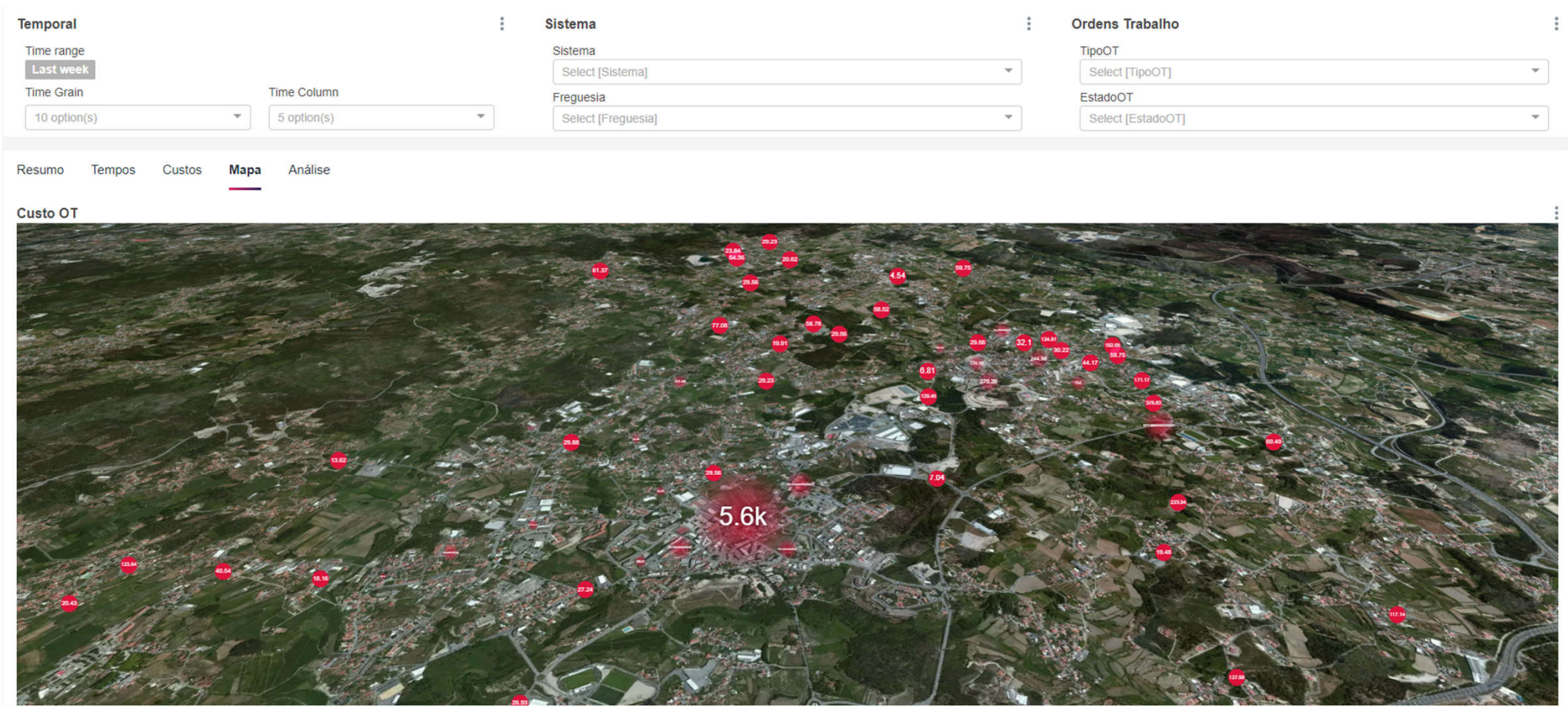
Sistema

Tipo Ativo

Tipo OT

Estado

Apache Superset - Examples



THANK YOU!