



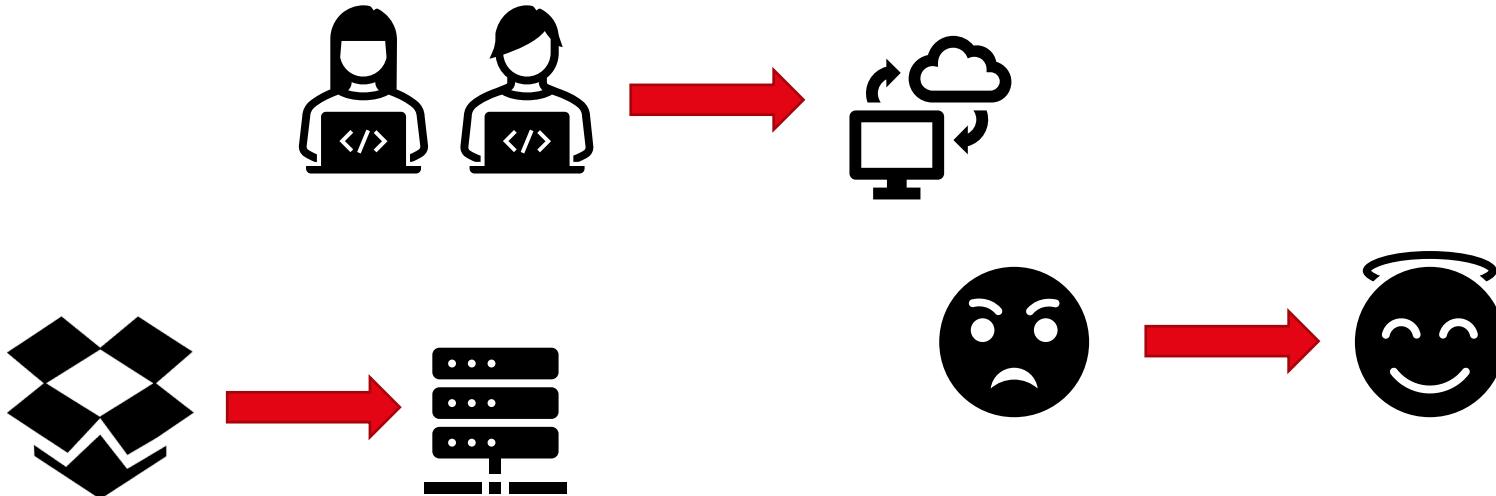
# BUILDING AN ETL PLATFORM FOR SOCIAL SCIENCE RESEARCH

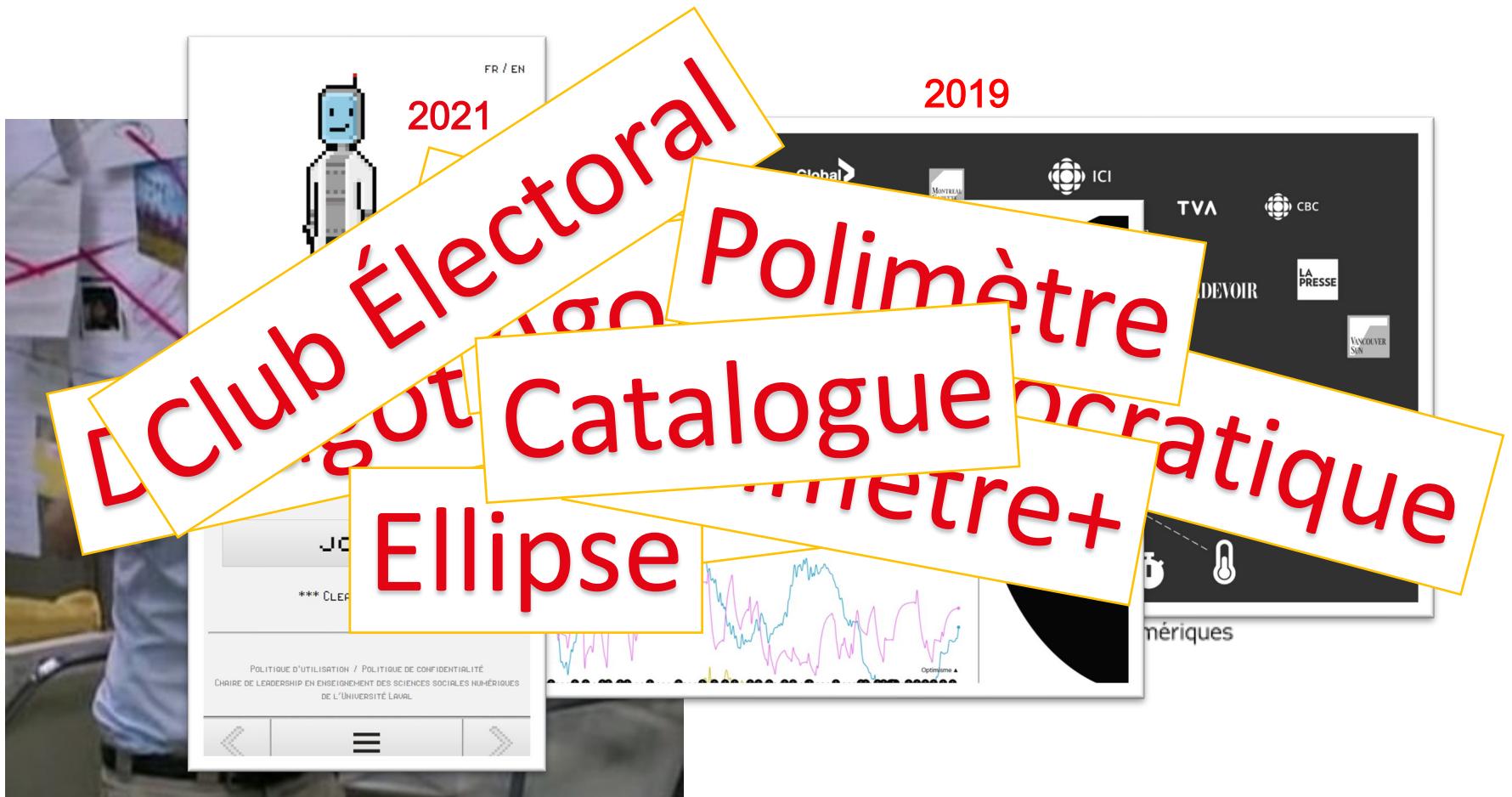
Olivier Banville

[olivier.chouinard-banville@dti.ulaval.ca](mailto:olivier.chouinard-banville@dti.ulaval.ca)



# The story





Pls stahp...

FR / EN



## DATAGOTCHI V.2021.09

UNE APPLICATION WEB ÉDUCATIVE ET LUDIQUE  
ÉMETTANT DES PRÉDICTIONS D'APRÈS VOS HABITUDES  
DE VIE.

JOUER

\*\*\* CLEAR STORAGE \*\*\*

POLITIQUE D'UTILISATION / POLITIQUE DE CONFIDENTIALITÉ  
CHIÈRE DE LEADERSHIP EN ENSEIGNEMENT DES SCIENCES SOCIALES NUMÉRIQUES  
DE L'UNIVERSITÉ Laval



PROJET  
QUORUM

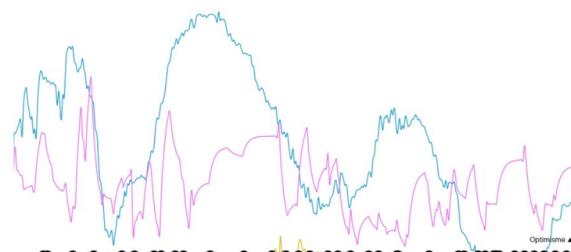
En

Menu

## DOSSIER ACTUEL La gestion de la crise de la COVID- 19 au Québec

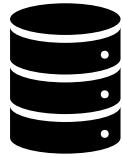
Plutôt optimiste ? Carrément pessimiste ? Situez-vous par rapport à la société québécoise et suivez en continu l'évolution de l'humeur de l'opinion publique, des médias et des décideurs publics

À PROPOS DE CE GRAPHIQUE ●  
HUMEUR DE L'OPINION PUBLIQUE ▲ 2.80% HUMEUR DES MÉDIAS ▲ 0.23% HUMEUR DES DÉCIDEURS PUBLICS ▲ 0.06%

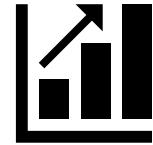


# A centralized alternative

Store



Transform



Access



# ClessnHub v1



First attempt

The good

- Django
- Openshift

The bad

- Slow API
- CSV output



# ClessnHub v2

Second attempt

The good

- Faster filtering
- More format support

The bad

- Slow pagination

# Database pagination



# Database pagination

Limit offset pagination



# Database pagination

Server-side cursor



# Database pagination

Ordered index cursor pagination

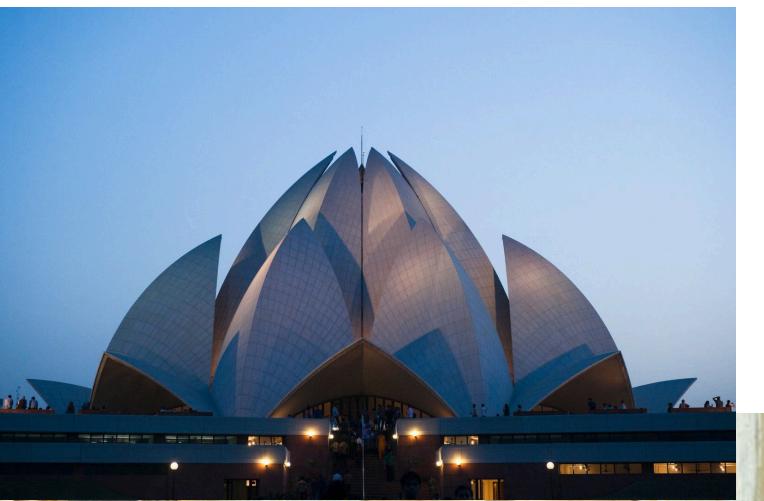


ordered  
indexed  
unique  
immutable  
pagination

# Hublot

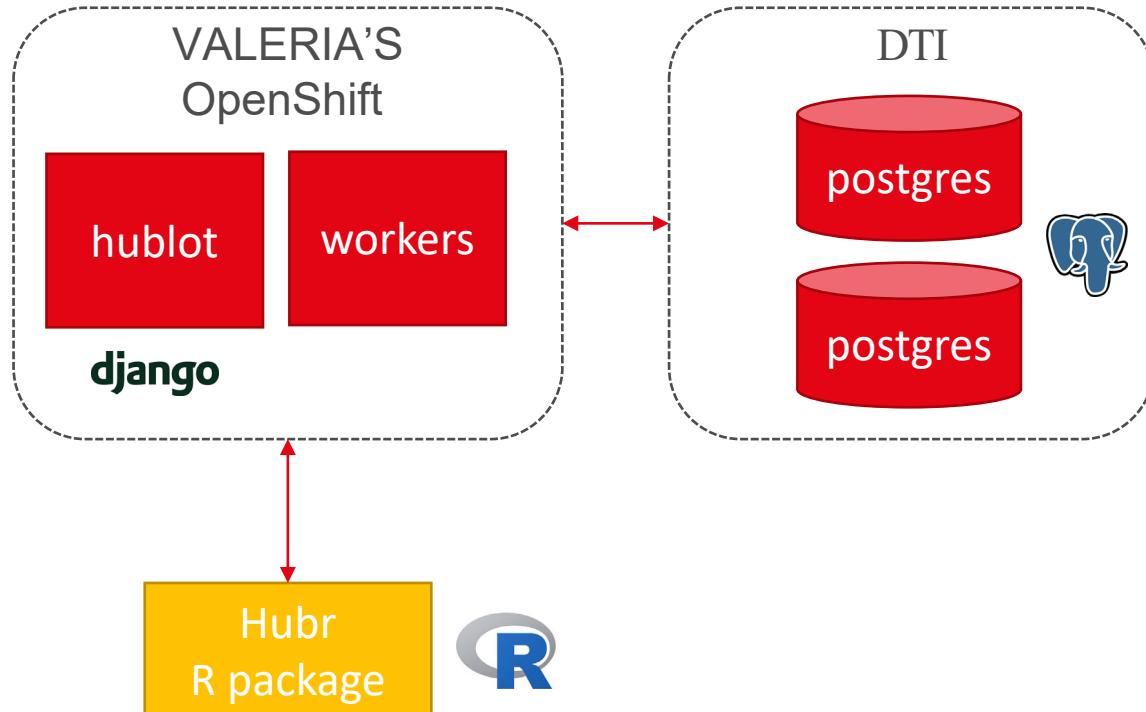
Third and last attempt

- Django 4
- Ordered index cursor pagination
- More RAM
- Independent API



cat tax

# Design



**WARNING**  
Technical stuff

# Dynamic Tables

Creating new database tables on the spot in a model-based framework



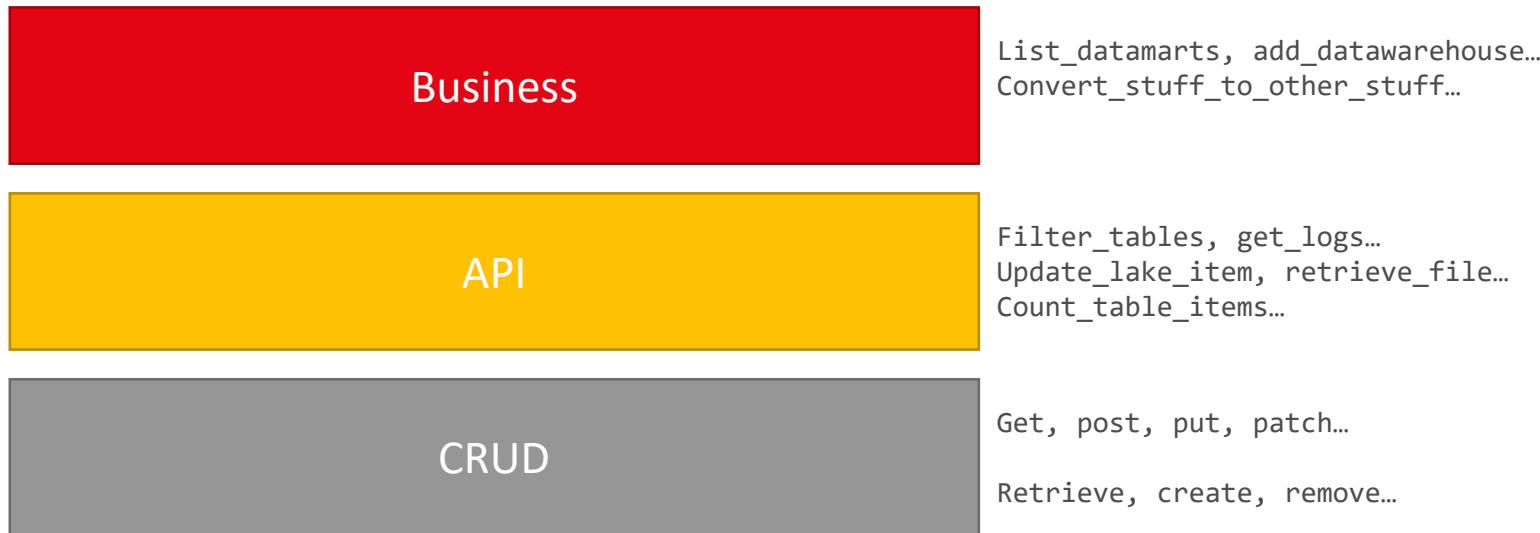
Django dynamic tables

- Dynamically create a new model class
- Dynamically create related serialized, admin page and api viewset
- Force refresh admin
- Force refresh routing table
- Profit



**WARNING**  
Technical stuff

## Layered approach to API wrapping



**WARNING**  
Technical stuff

ETL

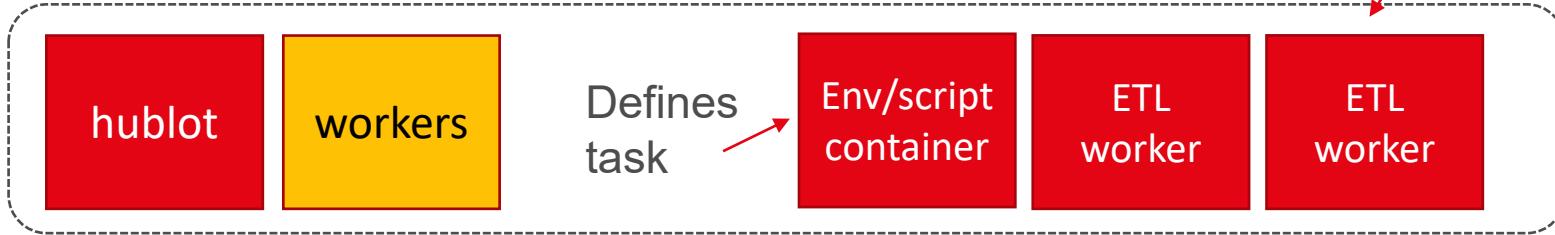
A



B

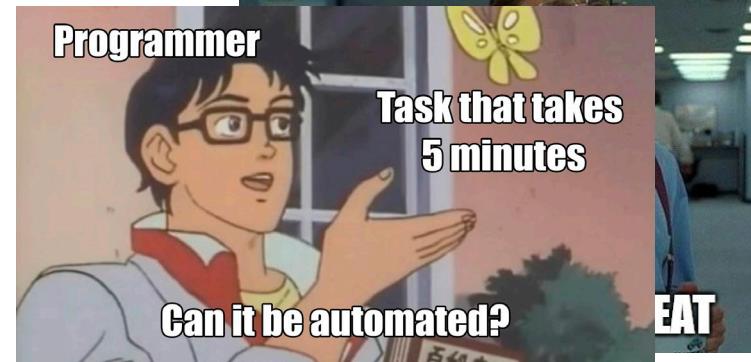
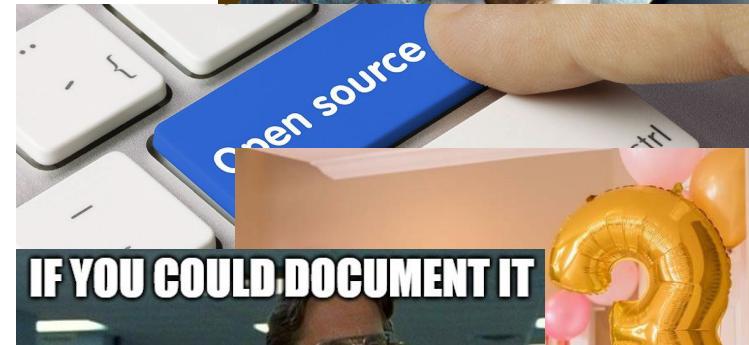


C



# The future

- Business Frontend
- Ready project for opensourcing
- Third party
- Documentation
- ETL

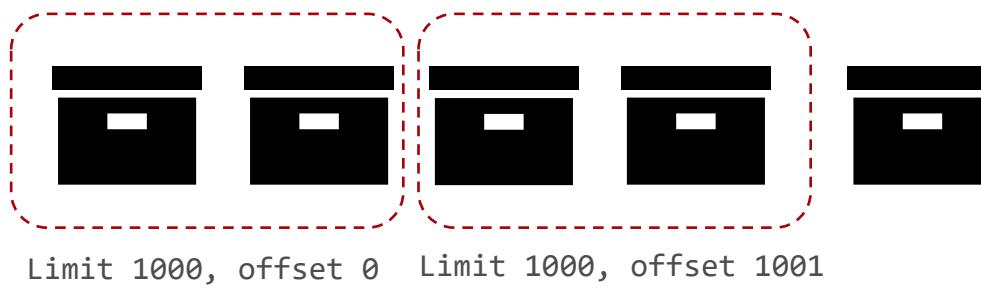


# Database pagination

Get millions of rows, bit by bit



- Lazy method: limit - offset
  - Ask for all rows, to a limit of 1000, offset 0
  - Ask for all rows, to a limit of 1000, offset 1001
  - ...

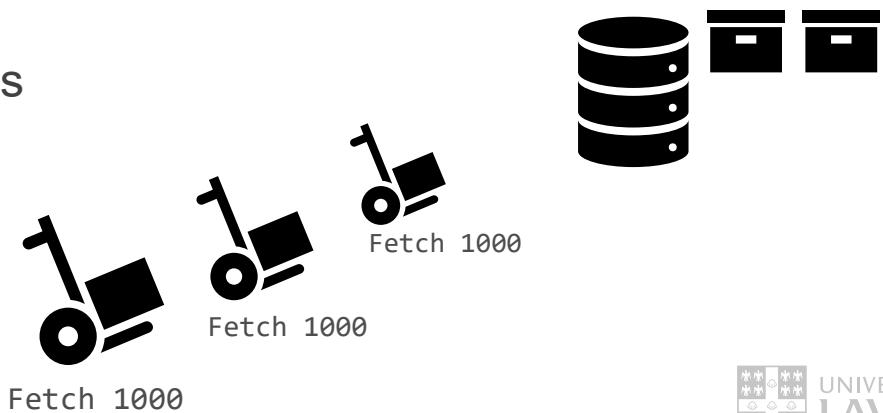


# Database pagination

Get millions of rows, bit by bit



- Server method
  - Create server-side cursor for a specific query
  - Ask for 1000 items
  - Ask for another 1000 items
  - ...



# Database pagination

Get millions of rows, bit by bit

**WARNING**  
Boring technical stuff

- Inbetween: Immutable-ordered pagination
  - Order by an immutable field
  - Ask for all rows, limit 1000, check highest value of immutable field
  - Ask for all rows where field > last highest limit 1000
  - ...

