Datawarehousing para datos genéticos, socioeconómicos y fenotípicos, con visualización 3D

SciPy 2018

Luciano Serruya Aloisi
Pablo Toledo Margalef

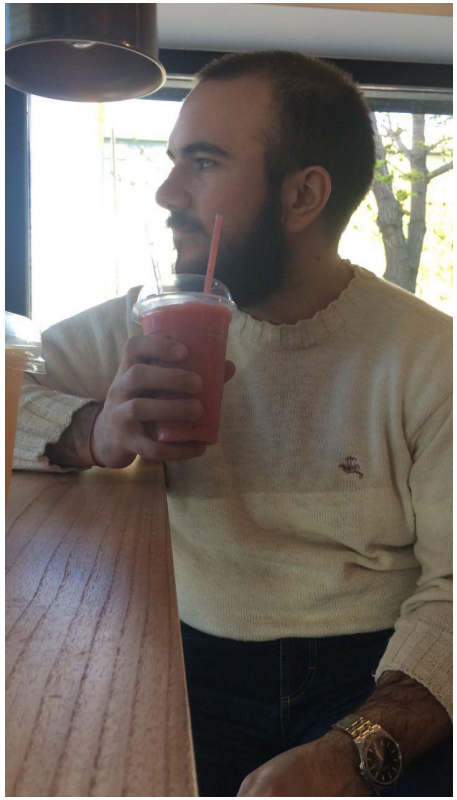
Universidad Nacional de la Patagonia San Juan Bosco

November 24, 2018

Roadmap

- ✓ Introduction 👋👋
- ✓ A little bit of software engineering 👨‍💻👨‍🔬
  - What we did
  - Why we did it that way
  - How we did it
- ✓ Demo time 😎🚀🔥
- ✓ Conclusions 🤔
- ✓ The end 🙏

whoami



Linux, Python, and Javascript

@LucianoSerruya

@LucianoFromTrelew

Linux and Python. FP enthusiast

@T_Papablo

@PaPablo

Both students at **UNPSJB, Trelew**

Getting in context

- **IPCSH-CONICET** studies how traditions, manners, and ancestry heritage relate themselves with medical interest physical variables.
- **RAICES Project** (**IPCSH-CONICET**) aims to build a Patagonian *biobank* (a genetic data bank)
- It is intended to help futures applications and design of public health politcs

Getting in context (cont.)

- RAICES Project sampling consists of a poll made to the volunteers (people who have been born in Argentina) and others several different exams
- These polls were (and still are) completed and then exported to a `xls` (Excel file) 👀
- The exams also output their own files

Getting in context (cont.)

- Sampled data:
  - Phenotypic data (whole-body videos, 3D scans)
  - Socioeconomic data (monthly income, lifestyle, how many home appliances do they have)
  - Ancestry (where do their parents and grandparent come from, what languages do they speak)
  - Drinking, smoking, eating habits
  - ...

Internship @ CENPAT

- This work comes out as a internship at CENPAT-CONICET
- Researchers of the project needed a software to handle all that data and files 😱
- That is where we come on in! To develop a datawarehouse and a web application to fulfill their needs

Design decisions 📐

- Web application
- REST Architecture 🏗
- noSQL Database (Mongo)

Development

- Datawarehouse (Python 🐍 + Mongo = Mongoengine FTW)
- Pandas 🐼 for data processing
- DRF + Vue.js
- THREE.js 👁
- Bokeh 📊

Problems we came across 🙈

- Missing data 😒
- Incorrectly formatted data 🤦
- Lack of documentation about Django + Mongo 🤨
- THREE.js API is not ES6 friendly 🤸

IT'S DEMO TIME

Conclusions & Resolutions

- You will have to integrate your frontend applicaciont with your API sooner or later 🤷
  - Decoupled architectures generates **coupling** if there is **a lack of communication** between both development teams 🙊
- Working with files and REST it is not the happiest thing to do (it was not for us at least)

Conclusions & Resolutions (cont.)

- If you get to work with 3D visualization, keep an eye on the following
  - Size and scale of your mesh
  - Camera and mesh position and angle
  - **Lighting** (if you are working with textures)

DOCKER IS NOT THE SOLUTION FOR EVERYTHING

I would like to know some more about it, please

- This slides - `https://github.com/LucianoFromTrelew/scipy2018-raices-dw.git`
- RAICES Project - `https://twitter.com/raices_proyecto`
- IPCSH - `https://ipcsh.conicet.gov.ar/`
- CENPAT - `http://www.cenpat-conicet.gob.ar/`

We listen to your questions 🤠

The end 👼

- ¡Muchas gracias!
- Muito obrigado!
- Thank you very much!

UNPSJB logo