

Plataforma de adquisición de datos pluviométricos para la predicción y aviso de inundaciones

Arkaitz Oderiz Garin
Unai Perez Goya



Cambio Climático

Definición: Cambios a largo plazo de las temperaturas y los patrones climáticos

Principal causa: La actividad humana

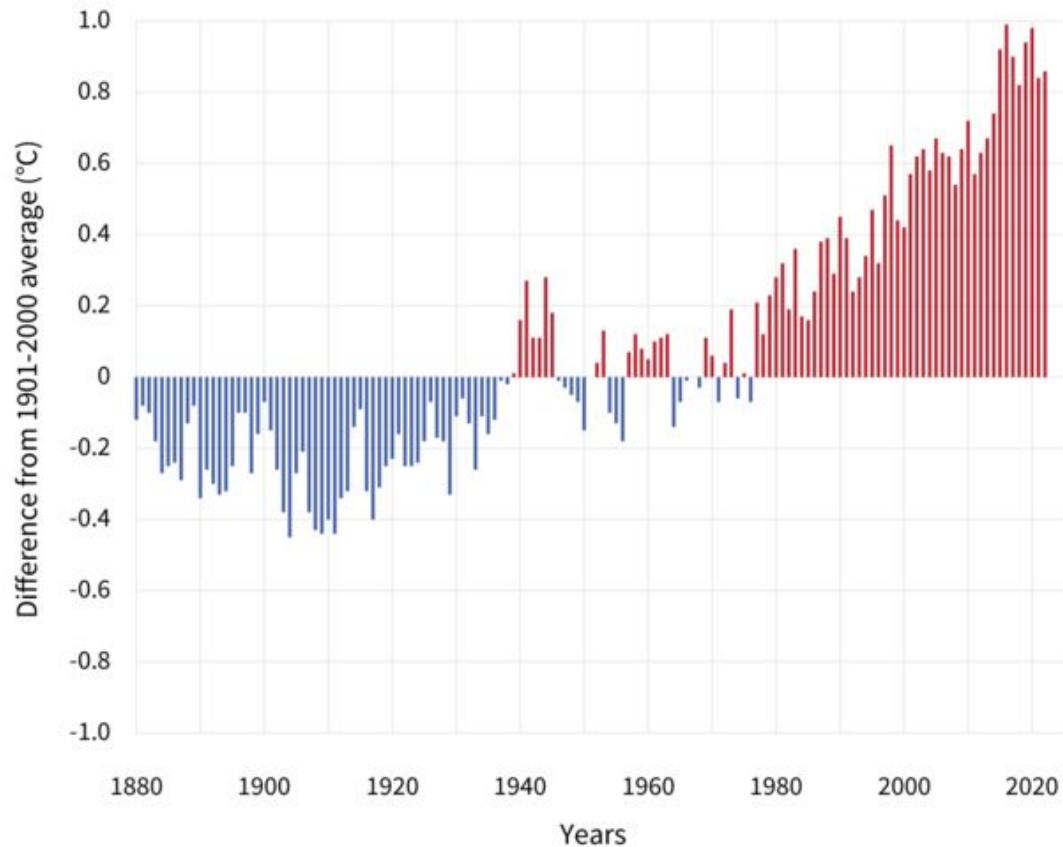
- Combustibles fósiles (Gases invernadero)
- Deforestación
- Sobrepoblación



Consecuencias

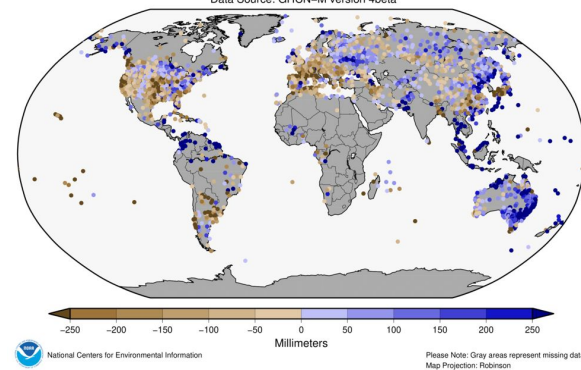
- Aumento de las temperaturas (Calentamiento Global)
 - Records mundiales
 - Subida nivel marino
 - Desertificación
- Fenómenos naturales extremos
 - Reducción del periodo de retorno
 - Precipitación media acumulada por debajo de la media
 - Intensificación de las tormentas y aguaceros
 - Riadas

GLOBAL AVERAGE SURFACE TEMPERATURE



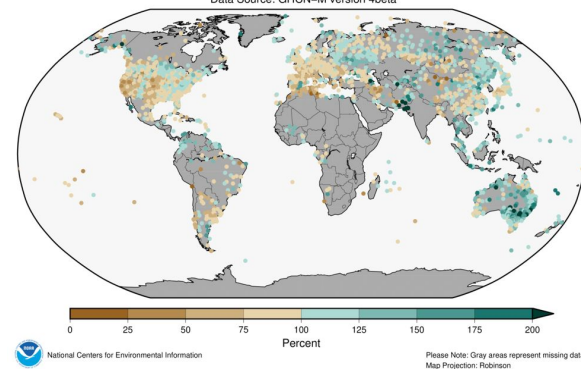
Land-Only Precipitation Anomalies Jan-Dec 2022
(with respect to a 1961-1990 base period)

Data Source: GHCN-M version 4beta



Land-Only Percent of Normal Precipitation Jan-Dec 2022
(with respect to a 1961-1990 base period)

Data Source: GHCN-M version 4beta





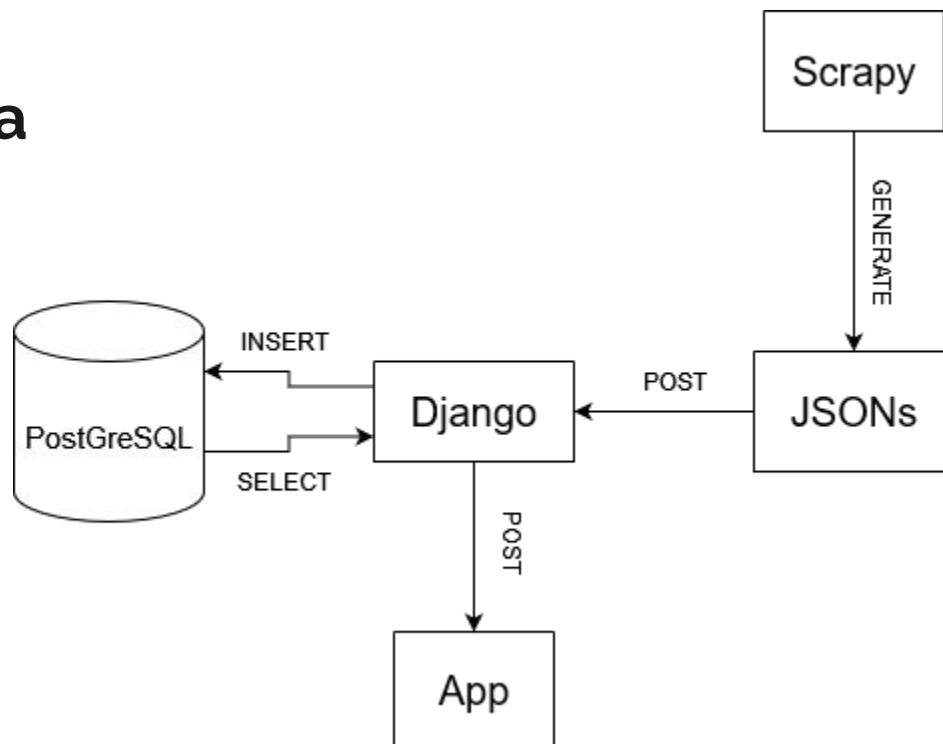
Propuesta

Crear una plataforma en la que centralizar datos fluviales y pluviométricos

¿Datos de donde?

- Conferencia hidrográfica del cantábrico (CH Cantábrico)
- Agencia estatal de meteorología (AEMET)
- Web de meteorología y climatología de Navarra (Meteo Navarra)
- Web de medio ambiente y meteorología de Navarra (El agua en Navarra)

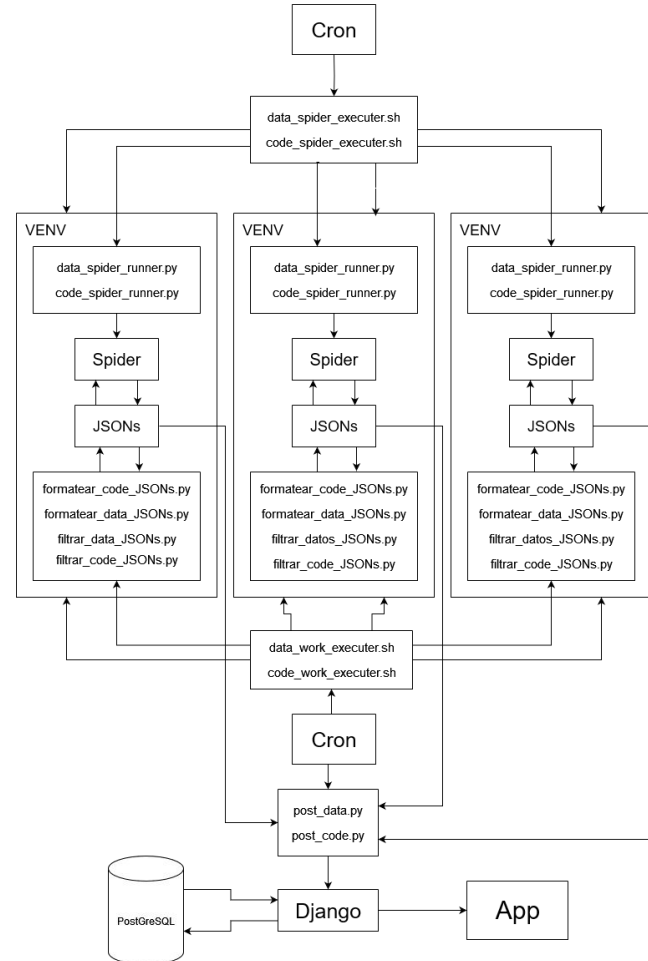
Plataforma



Arquitectura

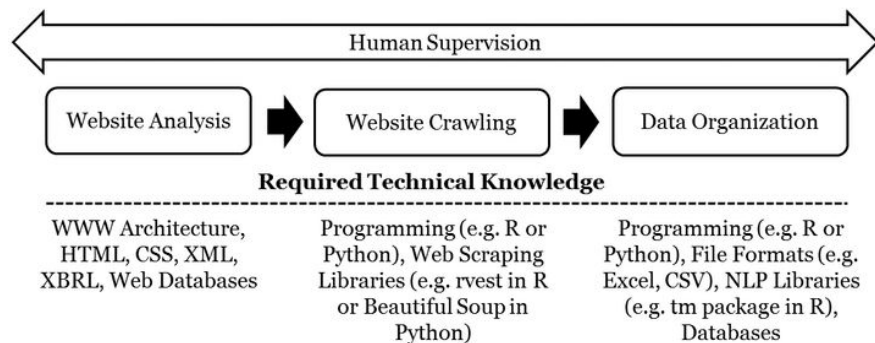
Dos partes:

- Obtención de datos
- Tratamiento de datos



Web Scraping

Técnica de extracción de datos desestructurados de la World Wide web (WWW) y guardarlos de forma estructurada en una base de datos o en un sistema de ficheros en formato XML, JSON o CSV para su posterior recuperación o análisis

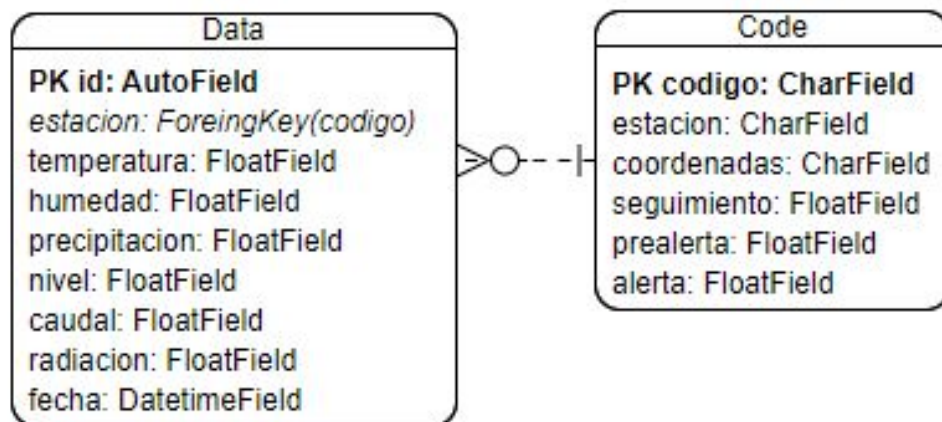




Datos obtenidos

- Nivel del río (m)
- Caudal del río (m^3/s)
- Precipitación (mm)
- Temperatura ($^{\circ}\text{C}$)
- Humedad (%)
- Radiación (W/m^2)
- Nombre estación
- Código estación
- Coordenadas
- Seguimiento del río
- Pre-alerta del río
- Alerta del río

Tablas de la base de datos resultantes





Flujo de datos

- Pedir datos
- Formatear datos
- Mandar datos a la base de datos
- Marcar datos como old
- While (verdadero)
 - Pedir datos
 - Formatear datos
 - Filtrar datos
 - Mandar datos a la base de datos
 - Marcar datos como old



Cron

Administrador de procesos en segundo plano presente en los sistemas basados en Unix.

Permite la automatización de la plataforma

Uso de Cron Jobs

```
.----- minuto (0-59)
| .----- hora (0-23)
| | .----- día del mes (1-31)
| | | .----- mes (1-12) o meses en inglés
| | | | .--- día de la semana (0-6) (domingo=0 o 7) o días en inglés
| | | | |
* * * * * comando a ejecutar
```



Spider

Objeto de Scrapy para la obtención de datos web mediante web scraping

Un proyecto Scrapy por web - Múltiples Spider por web

Proceso:

- Realizar petición web
- Filtrar y tratar los datos deseados
- Guardar datos en fichero JSON



API Django

Aplicación realizada para la recepción de datos y posterior subida a la base de datos

Intermediario entre la plataforma de adquisición de datos y el servidor base de datos



Executer

Script bash

Proceso:

- Activa entorno virtual
- Cambia a directorio definido
- Ejecuta Script Python indicado



Runner

Script Python para la ejecución automática de las Spider

Permitir ejecución de múltiples Spider mediante un único comando por terminal

Proceso:

- Tomar configuración de las Spider
- Iniciar entorno para web scraping
- Ejecutar Spider indicadas



Formatear Datos

Script Python

Tomar los datos recibidos de las webs e imponerles un formato estándar

Objetivo: Facilitar el trabajo sobre los datos, posterior tratamiento y subida a la base de datos



Filtrar Datos

Script Python

Eliminar la redundancia de datos antes de ser enviados a la base de datos

Comparación de datos recién recibidos frente a los ya disponibles

Objetivo: Reducir carga en la base de datos



Directorios JSON

Almacenar los JSON de forma jerárquica

Flujo realizado:

- Raw
- Parsed
- Refined
- Old



Post

Script Python

Enviar mediante POST request los datos restantes del filtrado a la API Django



Preguntas