

Actividad 3 Pokémon.

Big Data Aplicado.

Álvaro Martínez Lineros.

Instalación y funcionamiento.

Para ejecutar los scripts "volcarDatos.js" y "scrapData.py" es necesario tener instalado python y node. Realizar **npm update** y posteriormente ejecutar scrapData.py con **py scrapData.py** en la consola. Antes de insertar los datos en MongoDB se debe añadir el validador (validator.txt) de pokemons en la pestaña pokemons/Validation de MongoDB compass. Finalmente, tras haberse generado 2 archivos JSON (entrenadores y gimnasios) se ejecuta el otro fichero con **node volcarDatos.js**.

Adicionalmente se incluyen los JSON con toda la BBDD generada en caso de que no se genere correctamente con los scripts.

Web Scrapping con Python.

Para encontrar los entrenadores y los gimnasios se ha requerido hacer Web Scrapping de la Wiki de Pokémon. Para llevar a cabo esto, se realizó una petición a las url y se almacenó el texto HTML de la página. Se analizó el código y, buscando un patrón de aparición de los datos, se recogieron los datos encontrados y se filtraron para eliminar duplicados o datos inválidos. Finalmente los datos se almacenaron en un JSON asociado.

Recuperación de datos de la PokeAPI y subida a MongoDB.

Gracias a la documentación proporcionada en la PokeAPI se realizó un script con JavaScript para realizar la petición, realizar las relaciones entre los datos y, posteriormente, almacenarlos en MongoDB. Las peticiones se realizan de forma asíncrona y con promesas para poder agilizar lo máximo el proceso.