

Datalake To Datamart

Desarrollo Aplicaciones para Ciencia de Datos

2º curso

Grado en Ciencia e Ingeniería de Datos

Escuela de Ingeniería Informática y Matemáticas

Universidad de Las Palmas de Gran Canaria

Resumen

En este proyecto desarrollamos un código que se divide en tres módulos.

El primero llamado Feeder se encarga de extraer los datos diarios de la página de AEMET, serializándolos y creando un fichero ".events" para cada día en el datalake.

El segundo, DataBase, es el responsable de extraer esos datos del fichero del datalake, filtrarlos para solo obtener la máxima y mínima temperatura del día y, por último, crear una base de datos e introducirlos en ella.

Finalmente el tercero, denominado API, se dedica al control de la misma, otorgando la posibilidad al usuario de solicitar las máximas o mínimas temperaturas en un intervalo de tiempo determinado. Con estos datos, extraemos y filtramos por las fechas los datos almacenados en el datamart, es decir, la base de datos, y se lo mostramos al usuario.

Índice

1. Recursos utilizados	4
2. Diseño	4
3. Conclusiones	5
4. Líneas futuras	6
5. Bibliografía	6

1. Recursos utilizados

Para el desarrollo de este proyecto se ha utilizado como entorno de programación:

- IntelliJ

Herramienta de control de versiones:

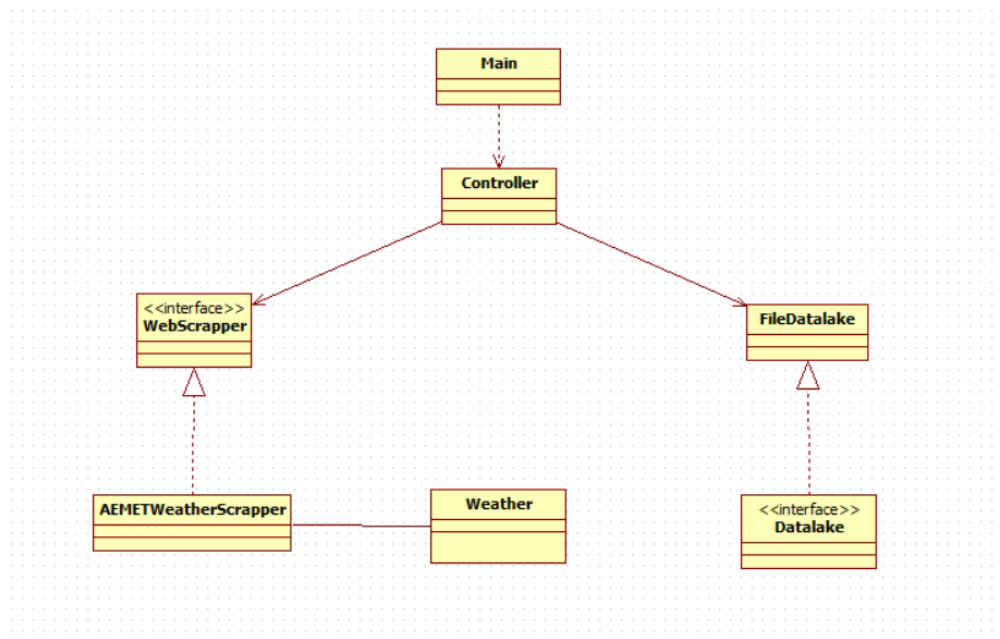
- Git
- GitHub

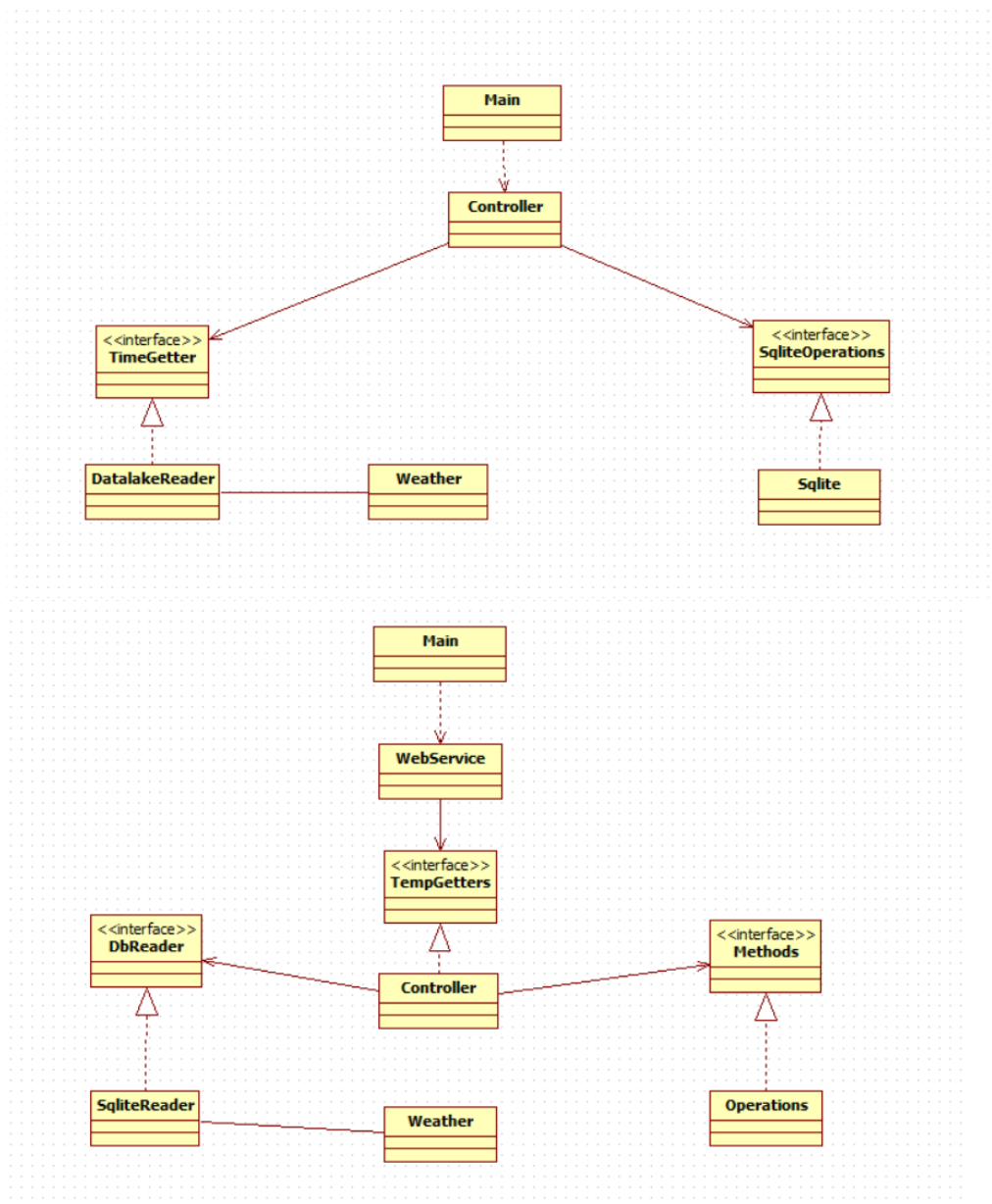
Herramienta de documentación:

- LaTeX

2. Diseño

Para el diseño de este proyecto se ha intentado seguir los principios SOLID y una arquitectura MVC; haciendo que cada clase solo sea responsable de una cosa, el uso de interfaces para encapsular métodos genéricos, modularidad de las clases y abstracción de todo lo que no fuese necesario para un algoritmo genérico.





3. Conclusiones

Como lecciones aprendidas tenemos el uso de módulos para la división de clases con respecto a su uso, la aparición del datalake y datamart, mejor uso de interfaces, aunque todavía me cuesta poner en práctica las clases abstractas, y el método "select".^{en} el apartado de las bases de datos.

4. Líneas futuras

Para el futuro uso de este proyecto se debería mejorar la obtención de datos de la AEMET para la creación de eventos en el datalake. Además de seleccionar otros servicios como este para aumentar la precisión de datos.

Por otra parte, habría que optimizar el código, tener un mejor control de excepciones y extraer métodos genéricos, como por ejemplo el uso en otros tipos de bases de datos.

5. Bibliografía

<https://aep2223.ulpgc.es/course/view.php?id=2098>

<https://www.aemet.es/>

<https://www.sqlitetutorial.net/>

Jesús María Matos Torres

12 de enero de 2023