

DATALAKE TO DATAMART (AEMET)

Desarrollo de Aplicaciones para la Ciencia de Datos GCID 2022-2023

ULPGC EII

Jorge Cubero Toribio 12/01/2023

RESUMEN

En esta entrega se ha creado un proyecto multimodular, en el cual cada modulo cumple una función diferente.

En el primero, llamado datalake, se toman los datos de temperatura de la isla de Gran Canaria a través de la página de la Asociación española de meteorología (aemet), tras ello se seleccionan la fecha y la hora en la cual se tomaron las medidas; el nombre del lugar de la estación en la cual se tomaron dichas medidas y su código; la temperatura y las temperaturas máxima y mínimas, y cada hora se rehace con los nuevos datos, con ello se crea un fichero de tipo events, llamado datalakes, escrito al estilo json cada día.

En el segundo modulo, llamado datamart, se leen los ficheros y se apuntan las temperaturas máximas y mínimas y se crean dos tablas una con las temperaturas máximas y otra con las mínimas, guardando las bases de datos en el directorio datamarts.

Finalmente, el tercer módulo, llamado api_aemet, crea un api rest con tres posibles respuestas, v1/places devuelve todos los lugares que están en las tablas, v1/places/with-max-temperature?from={fecha1}to={fecha2} que devuelve los lugares con mayores temperaturas entre dos fechas y v1/places/with-min-temperature?from={fecha1}to={fecha2} que hace lo mismo que la anterior, pero con las temperaturas mínimas.

Para iniciar la función main es necesario pasar como argumento la apikey obtenida en la web de la aemet, puede ser que Datacrawler falle al principio en ese caso se debe volver iniciar unas 2 veces más.

En la api la url se debe cambiar al lugar en el que se genere el directorio “datamarts”.

ÍNDICE

Página 2 Resumen

Página 4 Recursos Utilizados

Página 5 Diseño y Diagrama de clases

Página 6 Diagrama de clases extendido

Página 7 Conclusión y Líneas futuras

Página 8 Webgrafía

RECURSOS UTILIZADOS

El entorno de desarrollo usado es IntelliJ IDEA 2022.2.1 en su versión premium, con JDK 19, la versión de Gson es 2.10, org.example versión 1.0-SNAPSHOT, versión de Jsoup 1.11.3, versión de Spark 2.9.4, versión de SQLite 3.23.1, versión de Jakarta Enterprise y WS-RS 3.0.0, versión de Jakarta Servlet 5.0.0 y JUnit `{junit.version}`.

Plugins instalados: Big Data Tools, Jakarta EE Servlet Faces(JSF), Jakarta EE Web Services (JAX-WS), Android, Android Design Tools, Smali Support, Ant, Gradle, Maven, Code Coverage for Java, Database Tools and SQL, Docker, FTP/SFTP/WebDAV Connectivity, GlassFish, Jakarta EE Application Servers, Tomcat and TomEE, WildFly, HTML Tools, Refactor-X, XPathView + XSLT, XLT Debugger, Settings Sync, Angular and AngularJS, JavaScript and TypeScript, JavaScript Debugger, JavaScript Intention Power Pack, Karma, Next.js Support, Node.js, Node.js Remote Interpreter, Polymer & Web Components, Prettier, TSLint, Vite, Vue.js, Webpack, AOP Pointcut Language, Hibernate, Jakarta EE Platform, Jakarta EE Bean Validation, Jakarta EE Contexts and Dependency Injection, Jakarta EE Expression Language, Jakarta EE Messaging JMS, Jakarta EE Persistence JPA, Jakarta EE RESTful Web Services JAX-RS, Jakarta EE Web/Servlets, JavaFX, JVM Microservices Frameworks, Lombok, Micronaut, Persistence Frameworks, Quarkus, Reactive Streams, Eclipse Keymap, NetBeans Keymap, Visual Studio Keymap, Markdown, Properties, Endpoints, gRPC, OpenAPI Specifications, Protocol Buffers, Plugin DevKit, Jakarta EE Server Pages JSP.

DISEÑO

Principios de diseño: Principio de responsabilidad simple, cada clase tiene una única responsabilidad

DIAGRAMA DE CLASES Y RELACIONES

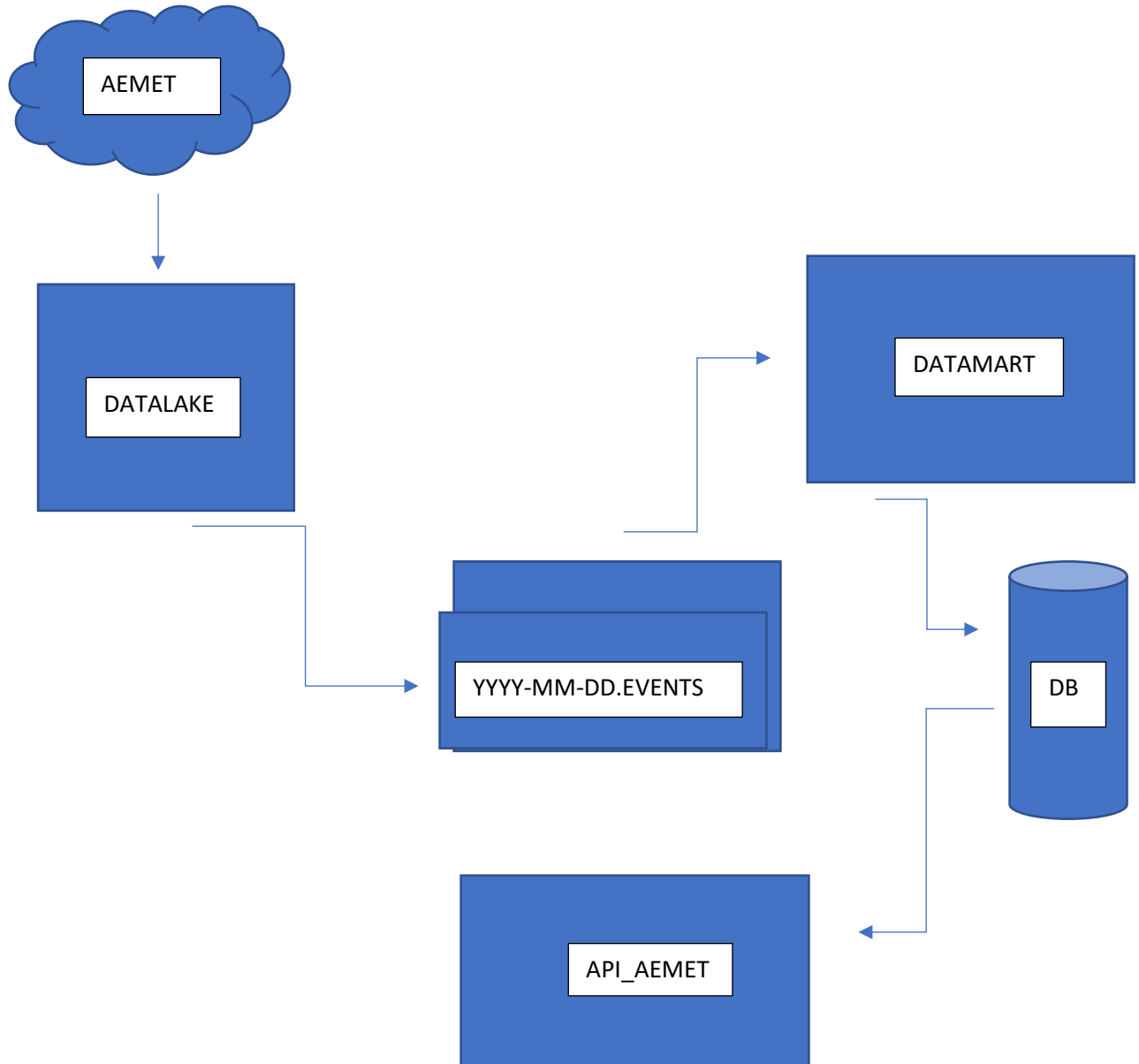
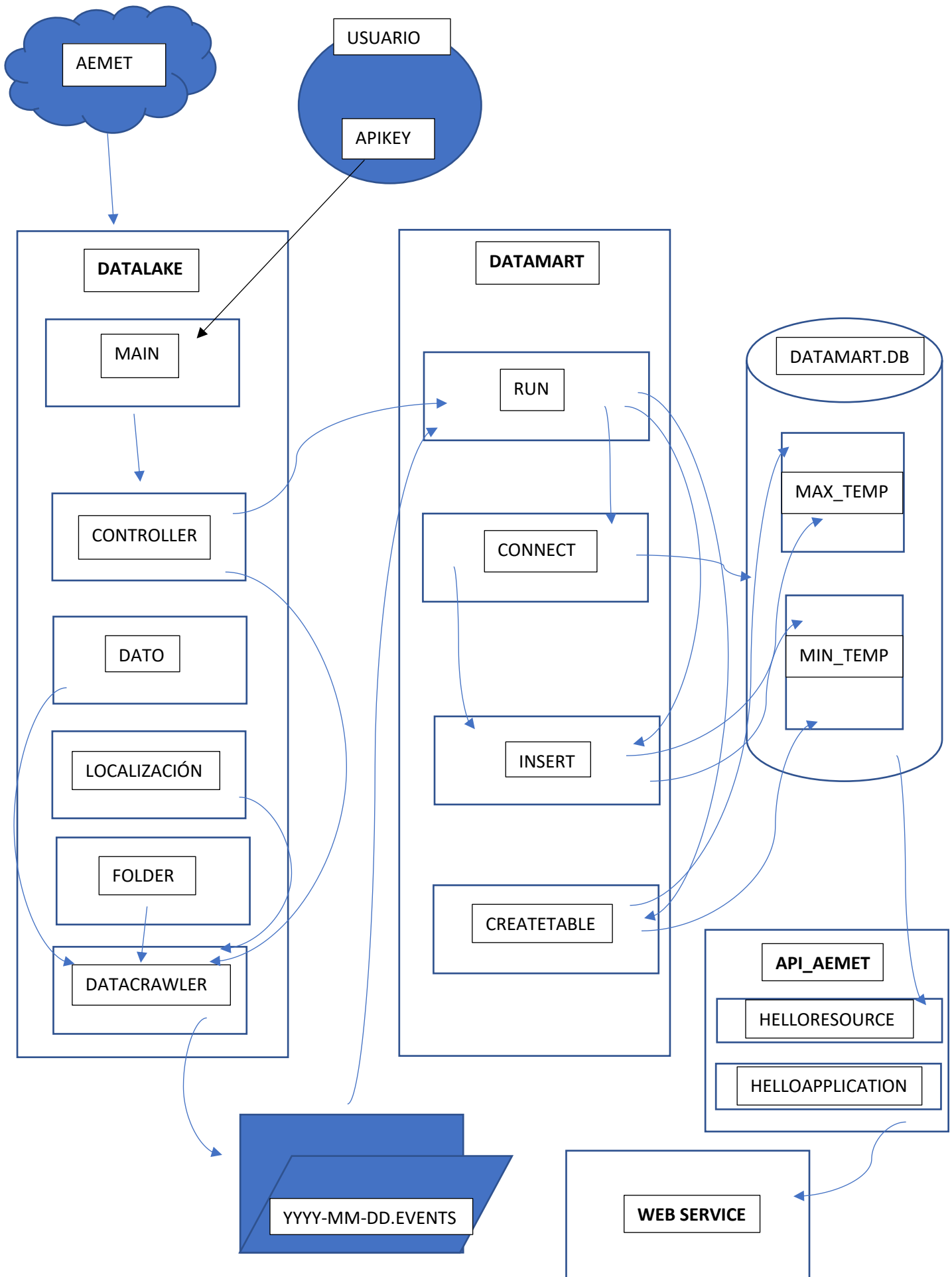


DIAGRAMA DE CLASES EXTENDIDO



CONCLUSION

Como conclusión que saco de este trabajo, me quedo con ser capaz de esforzarme lo suficiente y no sobrepasar mis límites, si empezara de nuevo lo tomaría con más calma y pararía cuando me empiece a agobiar y no seguir hasta quedarme sin aire del ataque. Además de poder preguntar cuando dudo y poder empezar de nuevo cuando tengo un fallo sin ponerme nervioso.

LINEAS FUTURAS

A futuro se podría añadir las temperaturas medias en diferentes momentos del año, de esa manera se podría vender el producto a compañías de viajes para que los turistas fueran a un sitio o a otro según el mes o para aquellos que quieran comprar una segunda vivienda sepan cual zona es mejor para vivir o alquilar.

WEBGRAFÍA

<https://opendata.aemet.es/centrodedescargas/altaUsuario>

<https://maven.apache.org>

<https://www.jetbrains.com/idea/download/#section=windows>

<https://jakarta.ee>

<https://chat.openai.com/chat>