

Proyecto Final

Para el proyecto final del curso de Data Products se busca evaluar todos los conocimientos adquiridos por los estudiantes a lo largo del semestre, por lo que el proyecto consistirá en lo siguiente:

Covid 19 Dashboard

Utilizando los datos provistos, los cuales contienen información sobre los contagios, las recuperaciones y las muertes, dados por país y región. Se solicita que usted construya un data pipeline que procese los 3 archivos csv, los inserte a una base de datos, y luego basado en los datos procesados, debe construir un dashboard que permite analizar las estadísticas de cada uno de los archivos.

Ingesta de Datos

Con los 3 archivos csv debe construir uno o varios DAGs utilizando Airflow que permite leer, transformar e insertar los archivos en una base de datos, la cual puede ser un MySQL o cualquier otra de DB de su preferencia. Para la construcción de los DAGs pueden utilizar cualquier tipo de task y operadores que considere apropiados, sólo debe tomar en cuenta que al momento de la calificación debe poder ejecutar el archivo y que este actualice la base de datos.

Dashboard

Una vez la data ha sido procesada, debe construir un Dashboard sobre la data, el cual debe contener por lo menos los siguientes criterios:

- Un mapa que muestre por código de colores, burbujas o como mejor considere la totalidad de los casos, recuperaciones y muertes para una fecha dada y un país determinado.
- Mostrar estadísticas de los incrementos de los casos, recuperaciones y muertes por varios países.
- Cualquier otro tipo de estadística o gráfica que considere necesario.
- Para el dashboard puede utilizar cualquier herramienta vista en clase, Shiny Apps, Stream List, etc.

Docker

Todo el proyecto debe correr utilizando docker, por lo que se espera que usted pueda crear en docker una instancia de airflow (no necesariamente distribuida) que ejecute los DAGs creados, la base de datos que decida utilizar, también debe estar corriendo en Docker y finalmente la herramienta que utilice para ejecutar el dashboard debiera tambien correr en Docker y debe utilizar la base de datos con la data ingresada desde airflow. Si la herramienta de dashboard utiliza los csv directamente, será penalizado con 0 en esta rúbrica.

Entregable

Para el proyecto se evaluará el uso de Docker y Docker Compose, por lo que deberá entregar en un repositorio de Github que contenga lo siguiente:

- Uno o todos los dockerfiles que haya utilizado para su proyecto, tome en cuenta que puede usar imágenes que encuentre en internet, pero no puede utilizar imágenes hechas por sus compañeros, de ser así, será penalizado con 0.
- El código fuente de los dags que creo
- El código fuente del dashboard que creo
- Un docker-compose.yml que contenga lo siguiente:
 - La configuración para ejecutar Airflow.
 - La base de datos que utilizará para el proyecto, diferente de la base de datos de airflow.
 - La configuración para la ejecución del dashboard, si es un shiny app, debe ser un container que contenga su shiny app y la ejecute, o bien si utiliza streamlight.
- Para la calificación se correrá el docker-compose y este debe generar los containers y al activar los DAGs, la data sea ingestada y luego se puede ver en el dashboard.
- Documentación de todo el proceso que realizó para el proyecto, esta puede ser en github, wordpress o cualquier otra herramienta.

Calificación

Las Rúbricas a calificar serán:

- Documentacion: 20%
- DAGS en Airflow: 40%
- Dashboard:
 - Mapa Interactivo: 20%
 - Gráficas y tablas: 10%
 - Creatividad en el Dashboard: 10%