**Reto técnico Data Engineering**

**Instrucciones:**

* Al recibir el reto, tendrás tres días para resolverlo y entregarlo.
* Siéntete libre de proponer la solución que consideres la más adecuada.
* Crea un archivo “README” en el cual expliques el razonamiento de la solución que propongas, así como los puntos fuertes y débiles que detectes en tu propuesta. También utiliza este archivo para responder las preguntas adicionales al ejercicio.
* Cualquier duda o ayuda que requieras, comunícate con José, el Lead de Data, al 5544509009.

**Ejercicio**

El equipo de Ciencia de Datos se ha acercado a ti ya que necesitan incorporar información sobre clima dentro de sus modelos predictivos. Para ello, han encontrado el siguiente servicio web: <https://smn.conagua.gob.mx/es/web-service-api>. Te piden que los apoyes en lo siguiente:

1. Cada hora debes consumir el último registro devuelto por el servicio de pronóstico por municipio y por hora.
2. A partir de los datos extraídos en el punto 1, generar una tabla a nivel municipio en la que cada registro contenga el promedio de temperatura y precipitación de las últimas dos horas.
3. Hay una carpeta “data\_municipios” que contiene datos a nivel municipio organizados por fecha, cada vez que tu proceso se ejecute debe generar una tabla en la cual se crucen los datos más recientes de esta carpeta con los datos generados en el punto 2.
4. Versiona todas tus tablas de acuerdo a la fecha y hora en la que se ejecutó el proceso, en el caso del entregable del punto 3, además, genera una versión “current” que siempre contenga una copia de los datos más recientes.

**Preguntas adicionales**

* ¿Qué mejoras propondrías a tu solución para siguientes versiones?

*R= Agregaría algunas lineas de código para mejorar el performance de la base de datos en el proceso, pro ejemplo indexado de datos, creación de llaves primarias, etc.*

*Igual incluirría alertas cuando valores estén fuera de rangos o cuando el proceso falle, que lleguen por diversos medios al desarrollador (slack, email, whatsapp, etc...)*

* Tu solución le ha gustado al equipo y deciden incorporar otros procesos, habrá nuevas personas colaborando contigo, ¿Qué aspectos o herramientas considerarías para escalar, organizar y automatizar tu solución?

*R= Primero versionar el código con una herramienta tipo git, adicional considerar la implemantación de CI/CD para un despliegue mas practico. Los proceso los ejecutaría en un ambiente en la nube, una base de datos autogestionada bajo demanda podría ser Aurora db o Mysql, postgress, etc si muestra tabla será de uso transaccional, si el uso será mas de análisis de datos y ciencias de datos sugiero usar bases mas enfocadas para ese uso como Big Query o Athena por ejemplo. Incluiría el proceso en un orquestador del tipo Airflow, o dagster, incluso el orquestador de Cloud Composer de GCP es muy eficiente para estas tareas.*

**Requerimientos técnicos del ejercicio:**

* Link al repositorio que contenga los recursos que resuelven el reto.
* Organización o estructura de carpetas por zonas de acuerdo al nivel de procesamiento en los datos.
* La solución debe generar logs.
* Bonus: Entregar la solución en Docker, Docker Compose o Airflow.