

# Prueba técnica Ingeniero de datos

Hola, ¡Felicidades por haber llegado hasta aquí y gracias por dedicar parte de tu tiempo para poder conocer tu talento!

La siguiente fase del proceso de selección consiste en realizar una prueba técnica, es una forma para evaluar inicialmente. Con esto tendremos una herramienta para poder calificar el conocimiento de los candidatos.

La prueba consta de 2 partes: La primera es de consultas SQL sobre una base de datos basada en el estándar de información de XM ([se adjunta mapeo de datos](#)) y la otra es realizar una ETL usando Apache Airflow.

Si te sientes bloqueado en alguna de las tareas, haznos saber hasta qué punto puedes avanzar.

## Prueba SQL

1. Escribir un query para consultar cada uno de los siguientes requerimientos de información: Aportes diarios en metros cúbicos de los ríos de la región Antioquia y aportes totales del sistema durante el último año.
2. Reservas del SIN en porcentaje durante el último año.
3. Máximo diario del precio de bolsa de los últimos tres meses.
4. Máximo diario del precio de bolsa de los últimos tres meses.
5. Promedio diario del precio de bolsa de los últimos tres meses.
6. Demanda mensual del SIN durante los últimos tres años.
7. Precios de oferta promedio por tecnología de generación durante el último año.
8. Compras y ventas en bolsa por agente durante el último mes.
9. Demanda de la última semana y de la misma semana del año anterior.

Los siguientes son las credenciales de acceso a la base de datos de PostgreSQL:

host [liquidacionxm.ccealdyo93sp.sa-east-1.rds.amazonaws.com](https://liquidacionxm.ccealdyo93sp.sa-east-1.rds.amazonaws.com)

port 5433

database liquidacionxm

user read\_only\_esinc

password X\$~4re7H%ahAKSxRh8H\*\*eB2V<2hh|M

Para realizar las consultas anteriores se agrega un listado de las tablas donde se puede encontrar:

public.grip,public.trsd,public.ddec,public.ddec\_nal,public.aportesproduccion,public.aportes\_ido,public.aportes\_bi,public.reservasproduccion,public.reservas\_ido,public.reservas\_bi,public.afac,public.ofei,public.capains,public.maestra\_recurso,public.tgrl,public.arrpas

## Prueba ETL

Realiza una ETL con las siguientes indicaciones

1. Por medio de la herramienta Apache Airflow (en su defecto en un script de Python) crear un DAG que conste de por lo menos las siguientes tareas.
  - a. **Extract:** utilizando la API de Twitter <https://developer.twitter.com/en/docs/twitter-api> extraer datos de su preferencia (Tweets, Perfiles de usuario, Tendencias), se deben realizar las peticiones necesarias para conformar un dataset de mínimo 100 filas.
  - b. **Transform:** Realizar el proceso de transformación, limpieza y normalización que considere necesario sobre los datos utilizando la librería Pandas. Tener en cuenta la calidad de los datos (Integridad, Consistencia, Relevancia) para garantizar un análisis fiable.
  - c. **Load:** Cargar o almacenar los datos en una base de datos relacional (Postgresql, SQL Server) o en archivos planos en formato .csv
  - d. **Notify (opcional):** Enviar un correo notificando el resultado de la ejecución del DAG.
2. Crear un tablero en alguna herramienta de visualización (puede ser Tableau, PowerBI, Jupyter Notebook con librería Matplotlib) que consuma los datos almacenados e incluya mínimo 3 visualizaciones (métricas, gráficos, mapas, esquemas, diagramas) de su preferencia. Por ejemplo, gráfico de líneas múltiples de tweets y retweets vs tiempo para 5 usuarios distintos.
3. Se debe subir el código fuente a un repositorio de Github donde se explique en el README el tipo de información que se ha recopilado desde la API, de igual forma debe incluir los .csv generados (si se ha elegido esta opción) y la visualización ya sea en formato PDF o enlace directo hacia el Notebook (Google Colab).
4. Tener en cuenta buenas prácticas en el código como documentación, estructura, seguridad, reusabilidad.

**Cualquier duda puedes escribir al email: [ecardona@enersinc.com](mailto:ecardona@enersinc.com)**