**Preguntas prueba técnica:**

1. **Explique cuál sería el procedimiento para cargar los archivos CSV generados en el ejercicio anterior en una base de datos mediante el uso de alguna herramienta ETL que usted conozca:**

La herramieta con la que más he trabajado es **Matillion**.  
El procedimiento para cargar los archivos CSV generados en la base de datos mediante **Matillion** como herramienta ETL sigue estas fases:

* **Extracción (Extract):** En Matillion se configura un *File Input Component* apuntando a los **archivos CSV** (top10\_avance.csv y bottom10\_avance.csv). Estos archivos pueden residir en servicios de almacenamiento como **Amazon S3**, **Google Cloud Storage**, **Azure Blob** o en un directorio accesible.
* **Transformación (Transform):** Con los componentes de transformación de Matillion se puede limpiar y validar los datos. Por ejemplo, normalizar los nombres de alumnos, eliminar duplicados o agregar un campo adicional que identifique si cada registro corresponde al **Top 10** o **Bottom 10**.
* **Carga (Load):** Finalmente, se utiliza un componente de carga como **Database Output Table** para escribir los registros en la tabla destino de **PostgreSQL** (por ejemplo, reporte\_avance\_estudiantes). De esta forma, los datos procedentes de los CSV quedan integrados y disponibles directamente en la base de datos para posteriores consultas o análisis.

1. **Describa que pasos usted usaría para retirar campos vacíos y caracteres especiales (comas, saltos de línea, diagonales, pipelines, etc.) en un archivo CSV y depositarlo en un repositorio utilizando una herramienta de flujos analíticos que usted conozca (Alteryx, Tableau Prep, etc.):**

Para retirar campos vacíos y caracteres especiales en un archivo CSV y depositarlo en un repositorio utilizando **Power BI Dataflows**, yo seguiría los siguientes pasos:

1. **Extracción del archivo CSV:**
   * Ingresar a Power BI Service y crear un nuevo **Dataflow**.
   * Seleccionar como origen de datos el archivo **CSV** ubicado en el sistema de archivos, SharePoint, OneDrive o en un almacenamiento en la nube.
   * Power Query detectará automáticamente las columnas y filas del archivo.
2. **Limpieza de campos vacíos:**
   * Usar las transformaciones de Power Query para **quitar filas vacías** o **reemplazar valores nulos** con reglas específicas.
   * Eliminar columnas que no contienen información relevante.
3. **Eliminación de caracteres especiales:**
   * Aplicar la transformación **Reemplazar valores** para eliminar o sustituir caracteres como comas, saltos de línea, diagonales, tuberías (|), etc.
   * Ejemplo: Usar la función Text.Replace() en una columna para limpiar caracteres indeseados.
   * Aplicar transformaciones de tipo **Limpiar** o **Espacios en blanco** para estandarizar los textos.
4. **Validación de datos:**
   * Verificar que después de la limpieza los datos tengan el formato correcto (por ejemplo, sin espacios extra en nombres de alumnos ni símbolos especiales en textos).
5. **Carga en repositorio destino:**
   * Guardar el flujo y configurar la **entidad** del Dataflow para que se guarde en el **Data Lake de Power BI** (Azure Data Lake Storage).
   * De esta manera, el CSV procesado y limpio queda disponible como **fuente estructurada en la nube**, listo para ser consumido en **Power BI Desktop** y en otros procesos analíticos.
6. **Sobre el ejercicio, ¿tienes alguna sugerencia de accionable para algún universo de alumnos?**

Con base en la clasificación por banderas, dar seguimiento prioritario a los alumnos con bajo avance (banderas 1 y 2) mediante tutorías y orientación, mientras que a los de mayor avance (banderas 4 y 5) se les podría ofrecer incentivos o vinculación con prácticas profesionales.

1. **Describe la expectativa que tienes de tu día a día y la manera en la que tu propones sumar al equipo actual**

Espero que mi día a día sea muy enriquecedor, lleno de aprendizaje y retos que me permitan crecer. Por mi parte, puedo aportar muchas ganas, disposición y compromiso para sumar al equipo y apoyar en lo que se necesite.

1. **Tiempo de duración:**

3 horas aproximadamente.