

¡Hola, artefactor!

Necesitamos un pipeline de datos para proveer información limpia proveniente del equipo de servicios galácticos.

Nuestro stack tecnológico es AWS, por lo que con S3, Lambda y Glue debería ser suficiente.

- La información del equipo de servicios galácticos la encuentras en: <https://swapi.info/people>
- El pipeline debe ejecutarse a demanda, tomar la información de las personas e ingestarla en AWS usando lambda
- La información debe ser guardada en S3 en un bucket de tu elección. Dentro del bucket crea 2 carpetas: raw y processed. Los jsons deben almacenarse en la carpeta raw con un timestamp como parte del nombre del archivo
- Un job de Glue Studio debe tomar el archivo más reciente ingestado en la carpeta raw del bucket S3 y realizar las siguientes transformaciones:
 - - a. Necesitamos los campos name, height, mass, hair color, skin color, eye color, birth year y gender de cada miembro del equipo
 - - b. Crear una nueva columna "normalized_birth_year" que convierta el año de nacimiento de la sintaxis BBY a sintaxis humana. Un simple 2000 menos el número que acompaña a BBY debería darnos el año de nacimiento. Si el campo birth_year no está disponible, deja este campo vacío
 - - c. Crear una nueva columna "mass_lb". Necesitamos la masa en libras y kg es la información por defecto
 - - d. Crear una nueva columna "gender_id". Necesitamos identificar el género, usando M, F respectivamente. Deja el campo con una N si no está disponible o si se encuentra cualquier otro género
 - - e. Elimina las filas si el campo mass excede los 1000 kg, no tenemos suficiente espacio en la nave
 - - f. Finalmente, si una o más filas tienen 3 o más campos en blanco, n/a o desconocidos, elimina la fila por completo
- La salida de dichas transformaciones debe convertirse a un archivo csv y guardarse en la carpeta processed del bucket

Para cerrar este tema, estamos esperando lo siguiente:

- Diagrama de tu solución: Un dibujo a mano, excalidraw o drawio
- El pipeline construido (lambda, glue, s3) y el código fuente de todos los scripts guardados
- Breve explicación de cómo podríamos automatizar esto para ejecutarlo diariamente