Balance de Energía

**Descripción del problema**

En el campo de la industria eléctrica, las compañías que se encargan de la generación de energía, se enfrentan con un gran reto y es que no pueden generar más ni menos energía de la que requieren debido a que esta no se puede almacenar, por esta razón los balances de energía deben ser supremamente precisos sin espacio para errores.

En la empresa de Juan, todos los días se lleva a cabo un proceso manual para establecer cuando, como y cuanta energía deben generar las plantas existentes. Juan, es el analista de mercado designado para hacer no solo el proceso de extracción, organización, comparación y análisis de la energía diaria, sino también de toda el área de planeación, proyecciones y contratación. La combinación de querer cumplir con todas estas tareas y teniendo la presión de saber que un error mínimo en el balance puede significar multas que llegan a sumas como los miles de millones de pesos, es para Juan razón suficiente para sentirse colapsado y llevarlo a comerte algún error humano.

Históricamente, los errores humanos en balances de energía han sido una problemática. Tenemos como un ejemplo para la industria colombiana algunas empresas de energía como son ENELAR en Arauca, Cedenar en Nariño, EMGESA, CODENSA,[1] entre otras las cuales fueron duramente sancionadas en el año 2023 por la entidad colombiana de regulación por sus incumplimientos en el año 2020 con sumas que rondan los 4,700.000 de pesos.

¿En que consiste el balance de energía y por qué representa un problema?

Para Juan realizar este proceso manualmente es complejo. El debe extraer la información de tres fuentes diferentes:

1. Un reporte diario de la energía generada en la empresa llamado DespachoGeneradoraX en formato Excel.
2. La cantidad de energía establecida por el ente regulador colombiano para los distintos tipos de plantas generadoras.
3. Los valores de la energía en la bolsa establecidos por el ente regulador colombiano.

Juan toma todos estos valores, los revisa, organiza y los lleva al mismo formato para poderlos comparar. Se hacen transformación de tipo:

* Formatos de fecha: Cada una de las fuentes entrega la información relaciona a la fecha en un formato diferente. El reporte diario entrega una fecha tipo DD/MM/AA HH: MM. Por otro lado, la información de la energía a proveer esta clasificada por fecha tipo AAAA-MM-DD HH:MM: SS y finalmente los valores de energía solo indican AAAA-MM-DD. Para poder realizar la comparación entre ellas, Juan manualmente las lleva al mismo formato.
* El reporte es entregado en un archivo Excel el cual tiene información adicional, esta debe ser leída pero no tenerla en cuenta en el análisis final. La información por parte del ente regulador se obtiene por medio de una API que retorna datos semiestructurados en formato JSON y estos deben ser filtrados y limpiados par obtener solo la información de interés.

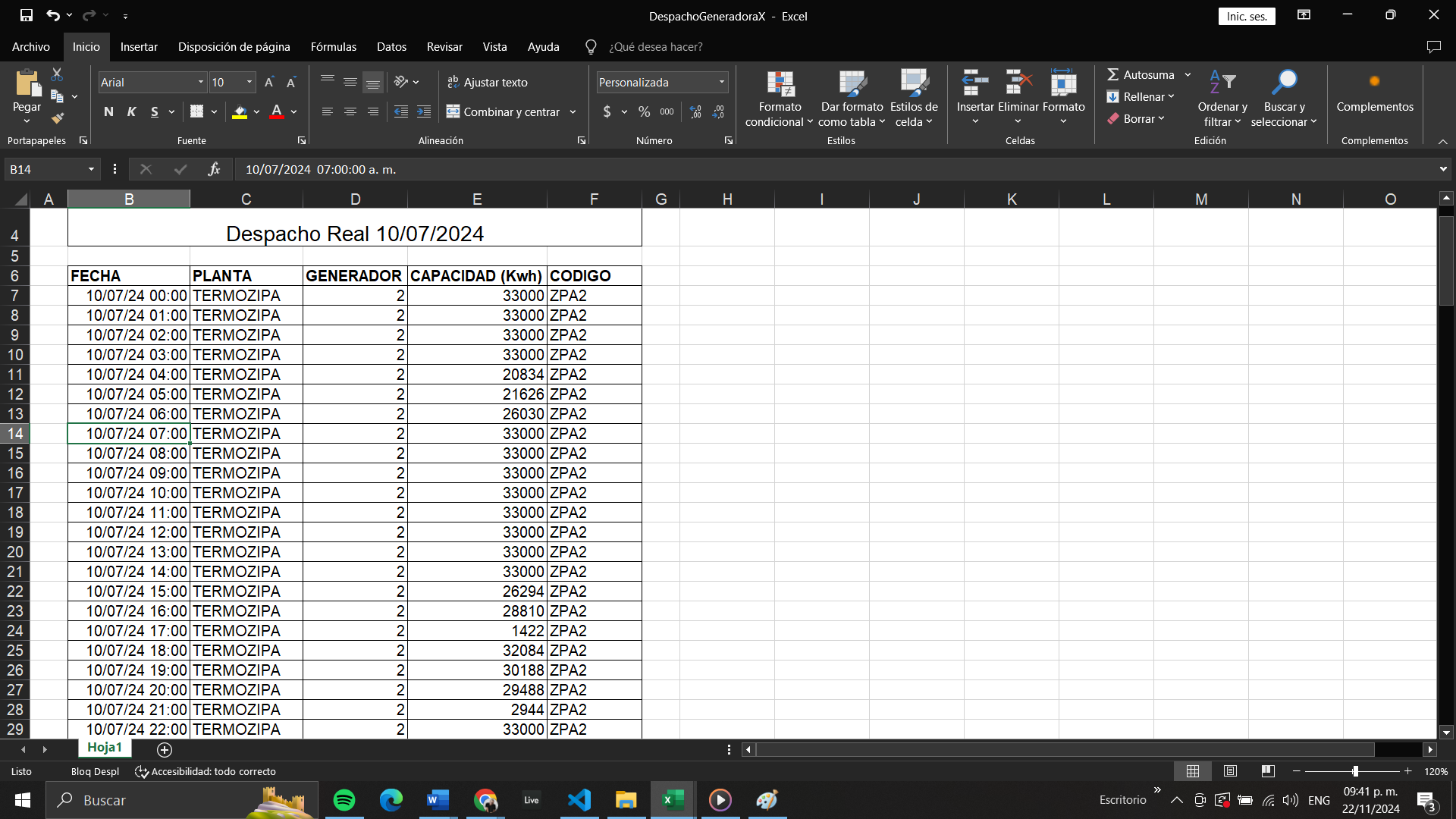
Una vez Juan tiene todos los datos en el formato deseado, comienza con el calculo manual por hora de la energía generada vs la que debe proveer, cuando completa las 24 horas a revisar, es capaz de determinar si el total de energía generado fue suficiente, falta o sobra y reporta esta información.

*¿Que se busca solucionar?*

Como se menciono anteriormente se busca reducir el porcentaje de error en los balances diarios de energía, evitando el cálculo manual y disminuyendo al máximo la intervención humana en estos procesos que representan un riesgo critico para la empresa y para los usuarios que dependen del servicio.

**Fuentes de los Datos:**

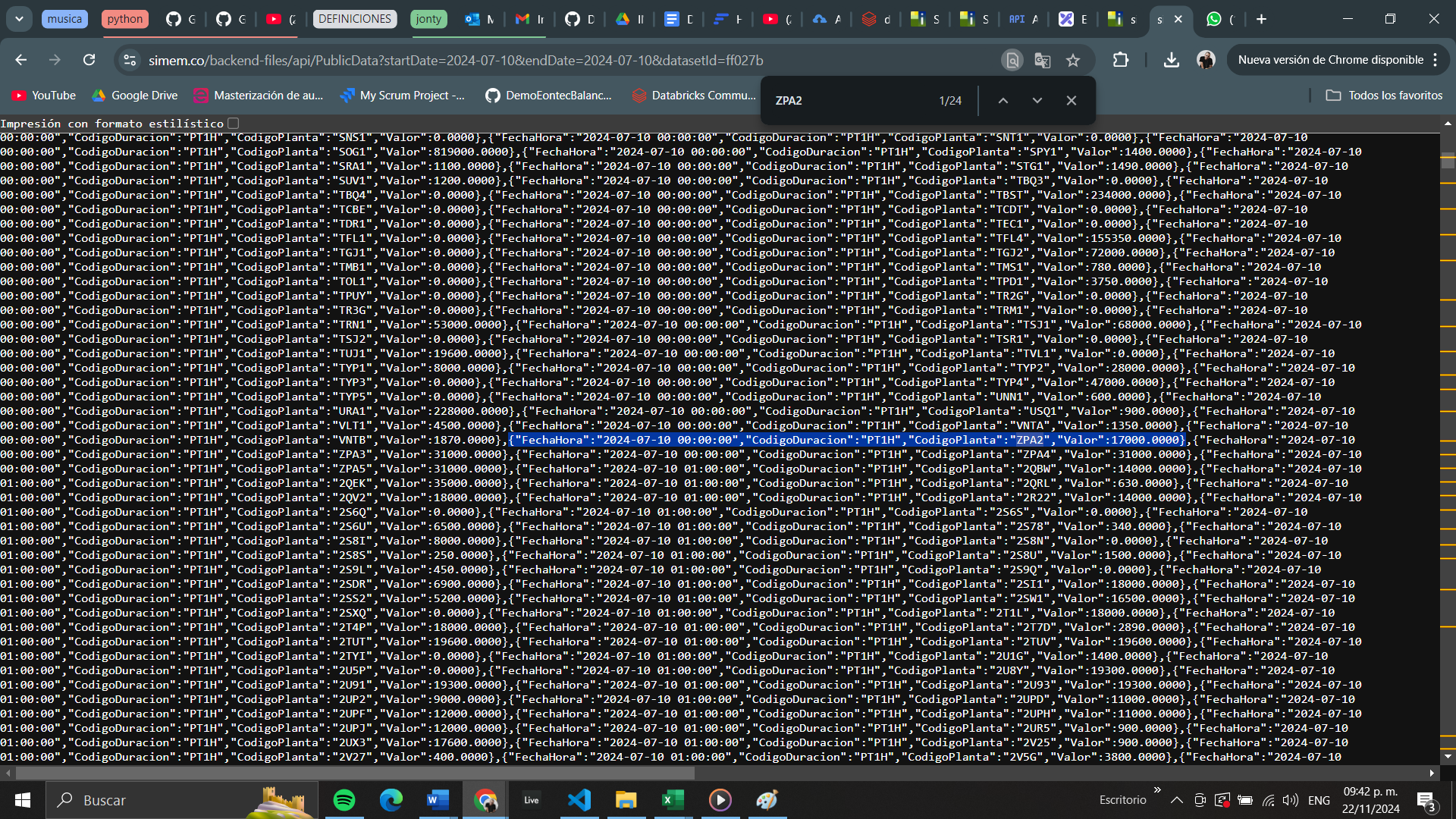
*Fuentes internas:* Reporte de cantidad de energía generada en una fecha y hora especifica, clasificada por plantas y generadores.



*Fuentes externas:* Entidad Colombia de regulación XM - Esta entidad cuenta con una API con diversos data sets. En este caso nos interesan dos:

* *Despacho programado recursos de generación (idDataset:"ff027b")*

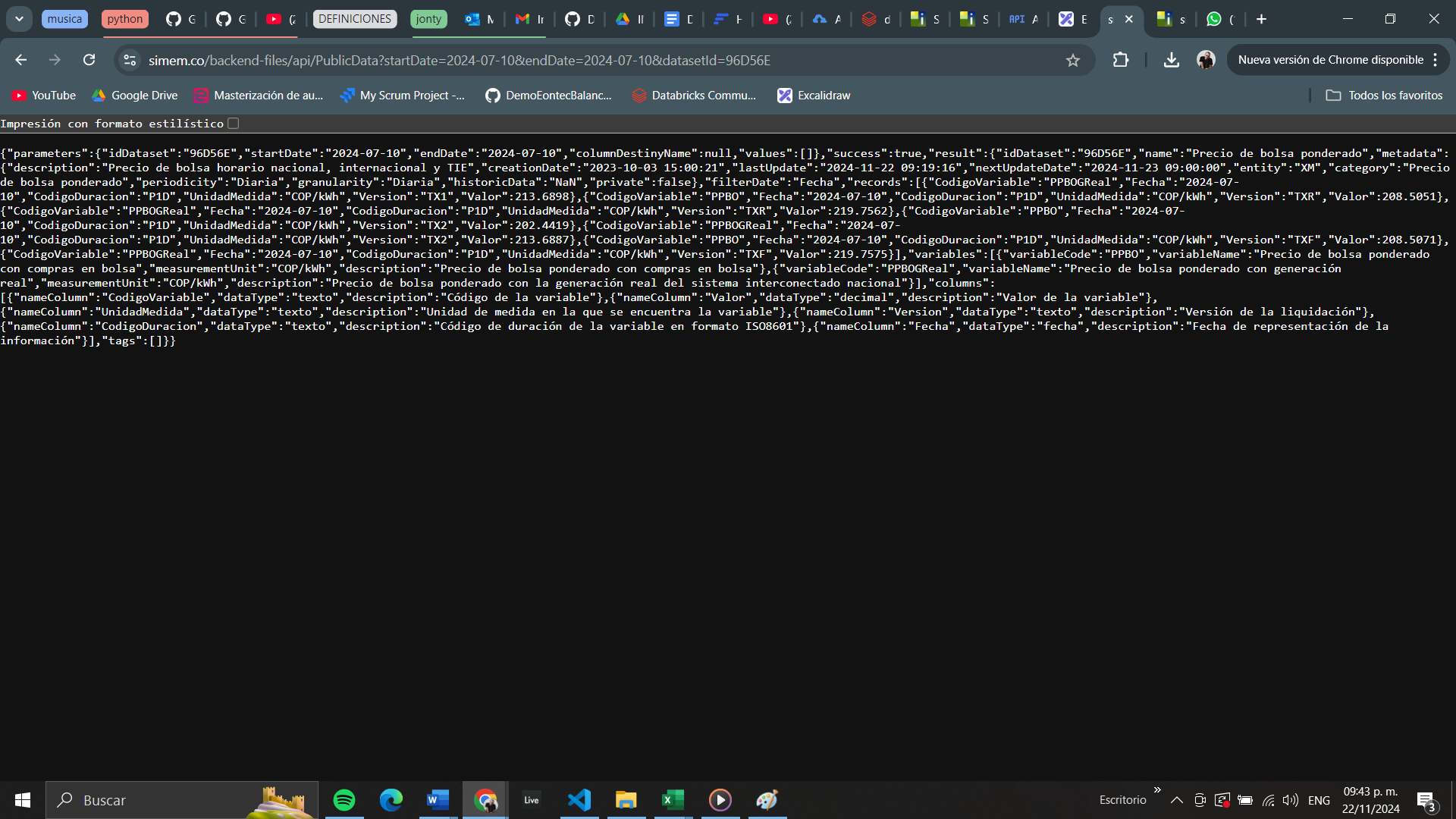
Con la fecha y hora de interés podemos consultar y filtrar la cantidad de energía que se debe proveer según los códigos identificadores de las diversas plantas a disposición dentro de la compañía.



Sample Api "idDataset":"ff027b"

* *Precio de bolsa ponderado (idDataset:"96D56E")*

Con el tipo de duración que tienen las plantas se determina el valor del kW.



Sample API "idDataset":"96D56E"

# Diseño tecnico o Two Pager

## (Declaración de Trabajo)

Construir un sistema suficientemente robusto y automatizado capaz de consultar las 3 fuentes de datos entregadas por la empresa, garantizando la lectura de estos datos que pasaran por un proceso de ETL para posteriormente analizarlos con el fin de generar un balance de energía que deberá ser entregado diariamente a las 4 pm por en analista de mercados.

## Análisis de Arquitectura

Dado que los inputs al proceso son archivos Excel y resultados de consulta a diversas API’s en formato JSON, los datos de entrada con los que se van a trabajar son datos semi estructurados.

El archivo de salida registra el balance entre la energía generada y la que se debe proveer.

Sabemos que si no se cumplen los niveles necesarios se pueden generar multas millonarias, por esto se considera importante poder generar un histórico de los balances de energía, tanto los archivos de entrada, como los de salida.

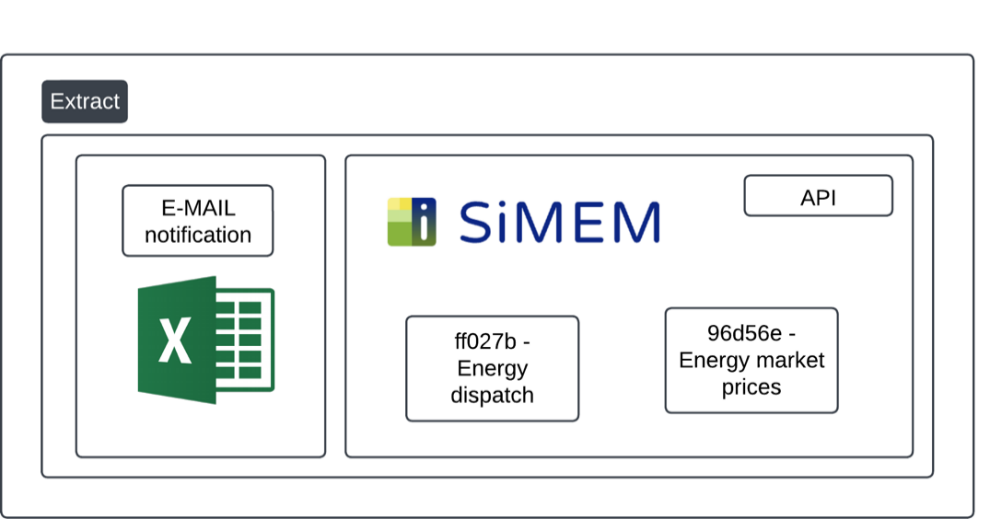
La Calidad de los datos es fundamental como parte de este proceso de ETL, especialmente en este caso para evitar la pérdida de datos, garantizar precisión y completitud de la información.

**E x t r a c c i o n:**

**Fuente 1:** Buscando la menor intervención humana posible en el proceso, se propone extraer el archivo con el reporte de energía generada real por medio de un email esperado todos los días a las 7am con un encabezado y un correo establecido. Para cumplir esto es necesario el compromiso por parte de la empresa de Juan, para no solo enviar el correo y el documento con el formato adecuado sino también la capacitación de sus empleados para llevar cabo esta tarea de la manera correcta.

Se agregará un proceso de validación de formato en el correo y archivo antes de proceder a filtrar los datos y se enviará una alerta por correo en caso de que no se ajuste al esquema establecido, evitando así el desperdicio de recursos, incumplimiento y falsas alarmas en el balance.

**Fuente 2:** API SiMEM – Se va a hacer una petición (get) a la API en los dos endpoints establecidos diariamente para obtener la información requerida. Estos datos son retornados en formato JSON. Se enviará igualmente un correo de alerta en caso de que la petición no sea exitosa o no se registren datos para los filtros deseados.

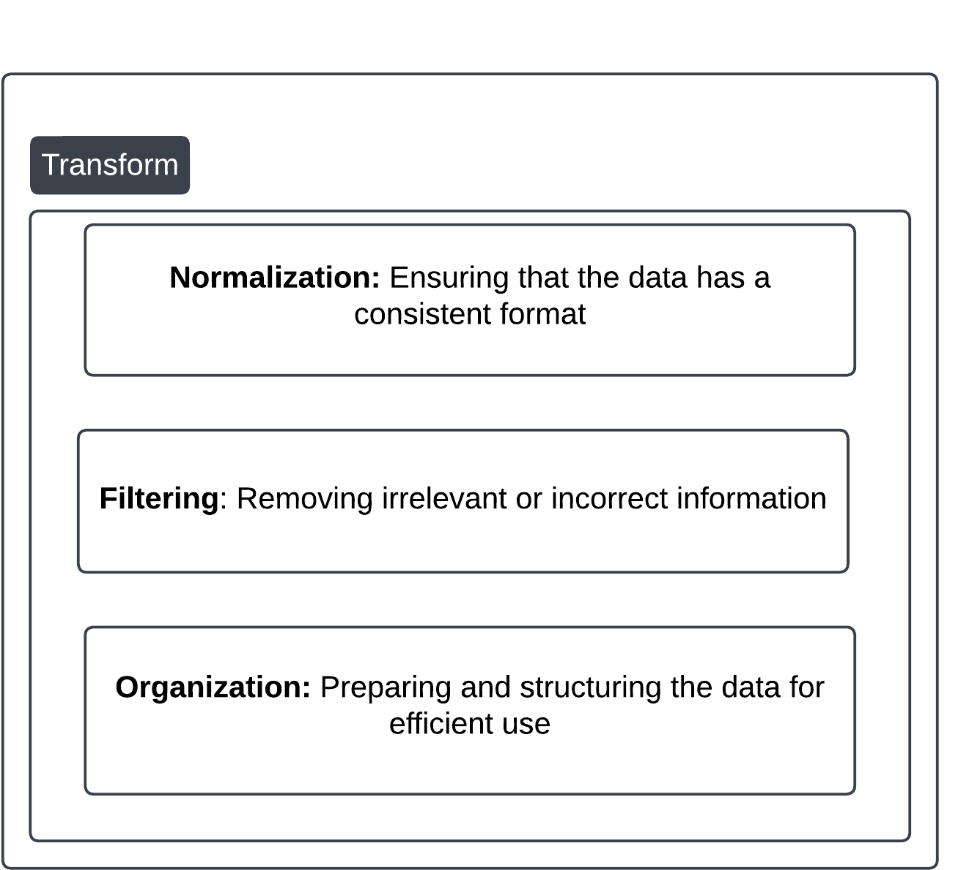


**T r a n s f o r m a c i o n:**

Como se mencionó anteriormente es importante hacer un proceso de normalización para contar con una versión de la información útil, uniforme y comparable sin importar cual fue la fuente de datos.

También se filtrará o se eliminaran todos aquellos registros que no brinden información valiosa para el balance de energía diario.

Finalmente, se prepararán los datos con estructura para ser usados eficientemente.



**C a r g a:**

Todos los datos transformados seguirán un formato establecido y esto nos permite hablar de datos estructurados.

Como el objetivo principal de este proyecto es realizar un análisis a estos datos, resaltando que se necesitan consultas rápidas en conjuntos de datos pequeños tenemos la posibilidad de usar lenguaje SQL para realizarlas. Se hará la carga de la información a un DataWarehouse, esta alternativa, puede verse un poco más costosa que un “DataLake” sin embargo garantizamos la velocidad, uniformidad, confiabilidad y estructura de los datos para este proceso crítico a diferencia del DataLake que puede ser menos eficiente para consultas rápidas en conjuntos pequeños, especialmente si los datos están en formatos no optimizados (como Excel o JSON).

Y el tiempo de "puesta en marcha" (cold start) de algunas herramientas puede ser más largo en comparación con un DataWarehouse.

Finalmente, como parte del proceso de análisis se entregará no solo una tabla con el resultado del balance, sino también un archivo Excel con los resultados buscando que otro analista de mercados pudiera acceder y entender el balance de energía. Esto particularmente si Juan (el analista encargado) no está disponible para realizar la lectura (se encuentra en permiso o ausente) y un suplente deba asumir su posición por un determinado tiempo.

## Seguridad y Privacidad

Debido a que esta información es sensible ya que muestra cuanta energía genera cada planta, su horario de operación y su comportamiento, no deberían ser de acceso libre para cualquier persona de la empresa. Se sugiere sea entregada solo al analista y encargado de los balances de esta o un suplente autorizado por la empresa si es el caso.

Por otro lado, se propone que la compañía entregue un correo electrónico de uso exclusivo para el proceso de generación de balance, este correo debería ser generado por la empresa con el propósito de tener un registro de cada movimiento en relación al sistema y así emitir los reportes requeridos en el desarrollo.

## Modelo

