

# Prototipo Fachada

Natalia Angarita  
Johnny Castañeda  
Carolina Osorio  
Marcell Piraquive

# Usuario Final del Artefacto

**Analista de Pérdidas.**  
Monitorear el consumo de energía de los clientes, identificar posibles anomalías y tomar medidas para reducir las pérdidas de energía.

# Requerimientos que relacionan la necesidad de usuario con la construcción de un artefacto analítico

## Negocio

- Conocer mejor los patrones de comportamiento de clientes no regulados o libres.
  - **Prueba:** Análisis descriptivo detallado de los clientes.
  - **Métrica:** 100% de los clientes deben estar cargados con su respectivo resumen descriptivo del comportamiento histórico.
- Identificar anomalías asociadas a pérdidas no técnicas.
  - **Prueba:** Análisis de clústeres.
  - **Métrica:** Reconocer clústeres con comportamientos anómalos asociados a pérdidas no técnicas.

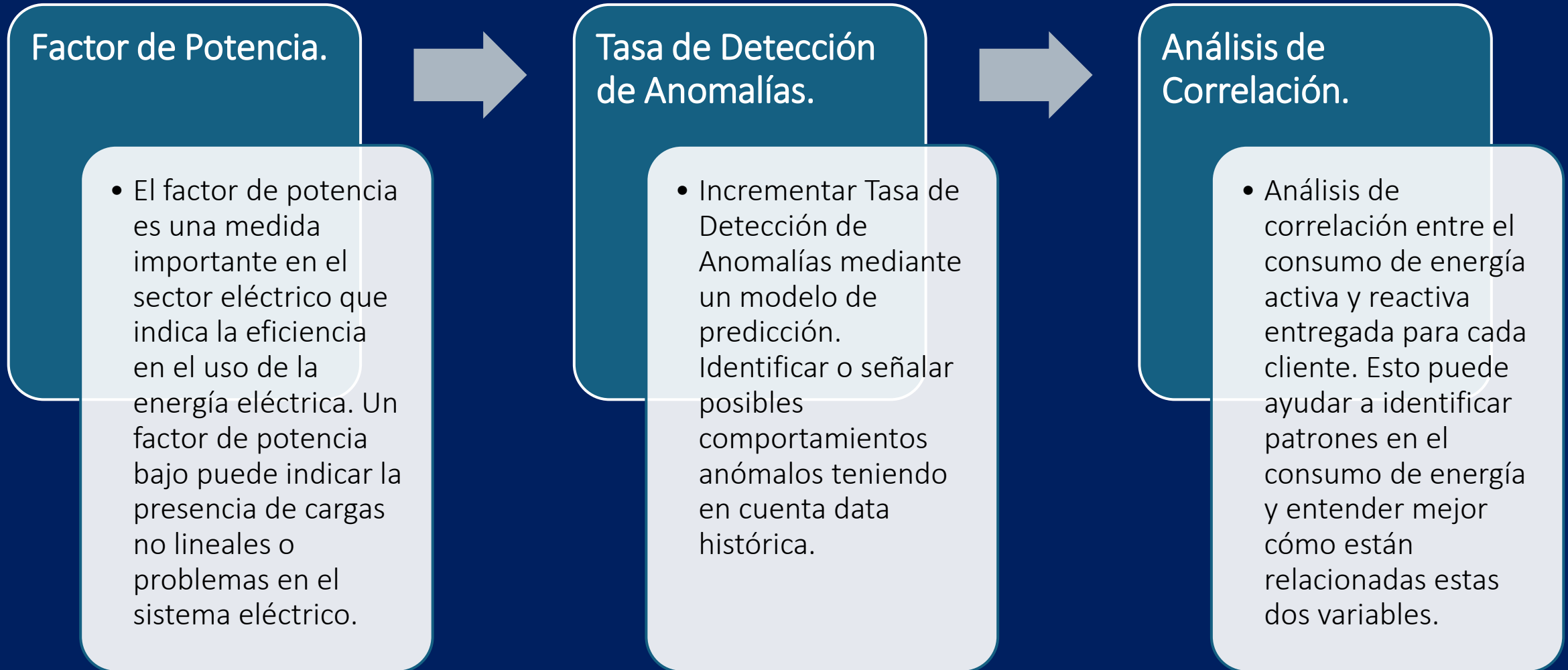
## Desempeño

- Análisis de clústeres para identificar tendencias, Coeficiente de Silhouette.
  - **Prueba:** Métrica de desempeño análisis de clústeres.
  - **Métrica:** Coeficiente de Silhouette  $> 0.4$ . Esto sugiere que los clústeres están relativamente bien definidos y separados entre sí.
- Desarrollar un modelo que permita identificar anomalías.
  - **Prueba:** Métricas basadas en la densidad.
  - **Métrica:** Métrica de Densidad Local y Densidad Local Promedio.

## Funcional

- Construir un dashboard amigable con el usuario, con una interfaz clara y comprensible.
  - **Prueba:** Demo a usuarios.
  - **Métrica:** Dashboard que cumple con requerimientos.
- Cargar el dashboard completamente en menos de 10 segundos.
  - **Prueba:** Pruebas de Performance.
  - **Métrica:** Cargar front end en menos de 10 segundos.

# Métricas clave del negocio o sector de contexto



# ¿Qué hace el artefacto? BACKEND



## Data.

- DatosClientes.csv
- SectorEconómico Clientes.csv



## Extracción y Disponibilización

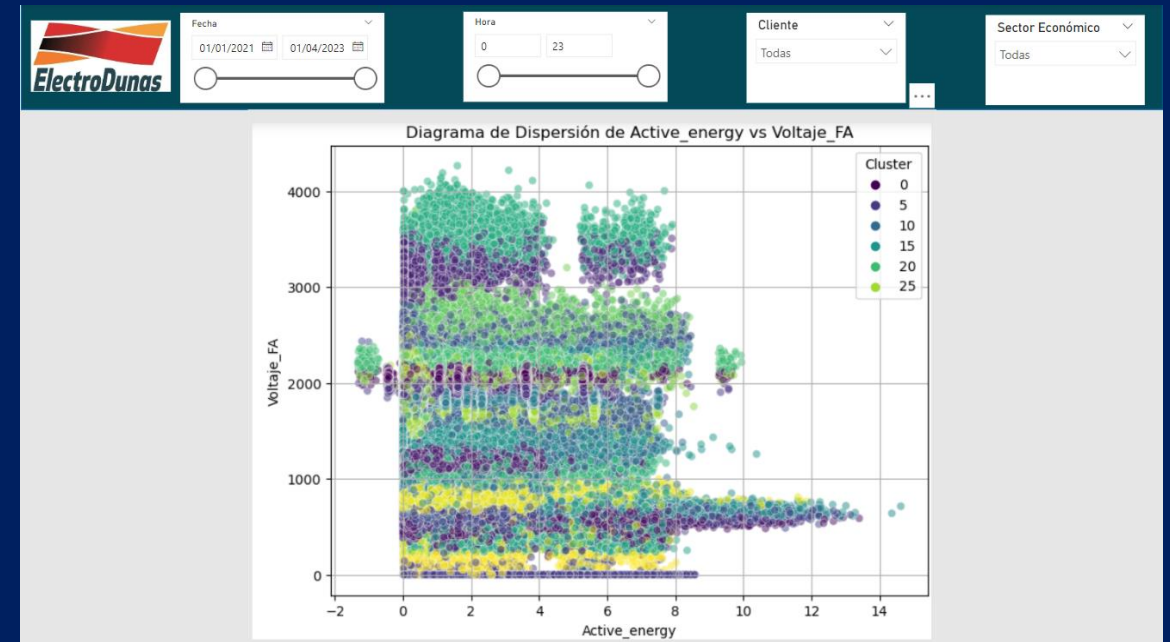
- Extraer archivos CSV desde Github o desde una base de datos si lo dispone la empresa.
- Versionamiento de datos utilizando git y procesamiento en área de staging de Python



## Transformación y Carga

- Transformaciones de datos utilizando Python, incluyendo técnicas de limpieza de datos (datos faltantes, inconsistencias, valores duplicados y datos atípicos). Además, agregar columna de mes y hora.
- Almacenamiento de datos en Bodega Snowflake.

# ¿Qué hace el artefacto? FRONTEND



La herramienta permitirá la visualización de información de manera dinámica, generando datos históricos segmentados según requerimiento del usuario, cálculos de KPI-Mediciones que facilitarán la toma de decisiones y, además, identificación de anomalías mediante medidas estadísticas en un boxplot. También mostrará un diagrama donde podremos evidenciar la distribución de clústeres que el algoritmo asignó.

# ¿Qué hace el artefacto? FRONTEND

**Filtros interactivos:** El dashboard proporciona filtros interactivos que permiten al analista ajustar los datos según sus necesidades. En este MVP1: Fecha, Hora, Cliente, Sector Económico.

**Selección de variables:** Permite al analista seleccionar qué variables desea incluir en la visualización. En este caso, seleccionar entre energía activa, energía reactiva y voltaje.

**Agrupaciones y desagregaciones:** Facilita la agrupación y desagregación de datos para proporcionar una vista más detallada o resumida según sea necesario. Por ejemplo, el analista puede agrupar los datos por día, semana, mes o año, o desagregarlos por franjas horarias o cliente.

**Interactividad en gráficos:** Proporciona detalles adicionales, o resalta áreas de interés, esto facilita la exploración de los datos y la identificación de patrones.

**Investigación de anomalías:** Cuando son detectadas anomalías en el suministro eléctrico, los analistas utilizarán el Dashboard para investigar y comprender la causa de las mismas. Esto puede implicar revisar los datos históricos, comparar con patrones anteriores y buscar correlaciones con otros factores.

# Tipos de análisis, modelos o algoritmos

## Análisis Descriptivo de la Data.

- **Descripción de estadísticas básicas:** Características de los datos mediante análisis de las estadísticas básicas de la data para entender el comportamiento y distribución de la misma.
- **Análisis de correlación entre energía activa y reactiva:** Calcula la correlación entre la energía activa y reactiva entregada para cada cliente. Esto puede ayudar a identificar patrones en el consumo de energía y entender mejor cómo están relacionadas estas dos variables:
  - a) Identificación de anomalías: Una correlación anormalmente alta o baja entre el consumo de energía activa y reactiva podría indicar la presencia de anomalías en el suministro eléctrico de un cliente. Por ejemplo, una correlación baja podría sugerir la presencia de cargas no lineales que generan un consumo desproporcionado de energía reactiva en relación con la energía activa consumida.
  - b) Detección de fraudes: Posibles fraudes o manipulaciones en el suministro eléctrico. Por ejemplo, una correlación inusualmente alta o baja podría ser indicativa de manipulaciones en el medidor, como la instalación de dispositivos para reducir la medición de energía reactiva.
  - c) Optimización de la red eléctrica: Comprender la relación entre el consumo de energía activa y reactiva de los clientes puede ayudar a identificar áreas de la red eléctrica que podrían beneficiarse de mejoras en la compensación de energía reactiva, como la instalación de capacitores o la reconfiguración de la red.
  - d) Gestión de la carga: El análisis de correlación puede proporcionar información sobre los patrones de consumo de energía de los clientes y ayudar a planificar la gestión de la carga de manera más efectiva.

## Clustering.

- **Funcionamiento:** El algoritmo de clustering busca identificar patrones inherentes en los datos, donde los elementos dentro de un mismo grupo comparten características comunes y, al mismo tiempo, son diferentes a los elementos en otros grupos. El objetivo es organizar los datos de tal manera que los elementos dentro de un mismo grupo sean más similares entre sí que con aquellos en otros grupos.
- **Implementación:** Los métodos de clustering están divididos en dos enfoques principales:
  - a) Partitioning Clustering: Algoritmos que dividen el conjunto de datos en un número específico de clusters. Ejemplos incluyen K-Means y K-Medoids.
  - b) Hierarchical Clustering: Algoritmos que construyen una jerarquía de clusters. Pueden ser aglomerativos (comienzan con clústeres individuales y los fusionan) o divisivos (comienzan con un clúster único y lo dividen iterativamente).
- **Resultados esperados:** Capacidad para predecir si un patrón de consumo es normal o anómalo, con una alta precisión y robustez frente a datos no lineales y complejos. Aunque estos métodos varían en su enfoque y forma de operar, todos comparten el objetivo de agrupar datos de manera significativa y útil para revelar patrones, relaciones o estructuras subyacentes en los conjuntos de datos.

## Detección de Anomalías.

- **PCA (Análisis de Componentes Principales):** Identifica anomalías mediante la detección de puntos que se desvían significativamente de la estructura general de los datos en un espacio de dimensionalidad reducida. Las anomalías se pueden detectar observando los puntos que están lejos del espacio proyectado de los datos normales.
- **LOF (Factor Local de Anomalía):** Evalúa la densidad local de los puntos y compara esta densidad con la densidad de sus vecinos. Los puntos que tienen una densidad significativamente más baja que la de sus vecinos se consideran anomalías. Esto permite detectar puntos que son inusuales en su vecindario local.
- **DBSCAN (Agrupamiento Espacial Basado en Densidad):** Detecta anomalías al identificar puntos que no pueden ser agrupados en ningún cluster o que están en regiones de baja densidad. Los puntos que quedan como ruido o que no pueden ser asignados a ningún cluster se consideran anomalías.
- Los resultados esperados de cada modelo incluyen la identificación de puntos que se desvían de la estructura general de los datos o que tienen una densidad local inusualmente baja en relación con sus vecinos. Estas anomalías pueden indicar la presencia de comportamientos inusuales o anómalos en los datos, como el fraude, errores de medición o eventos inesperados.

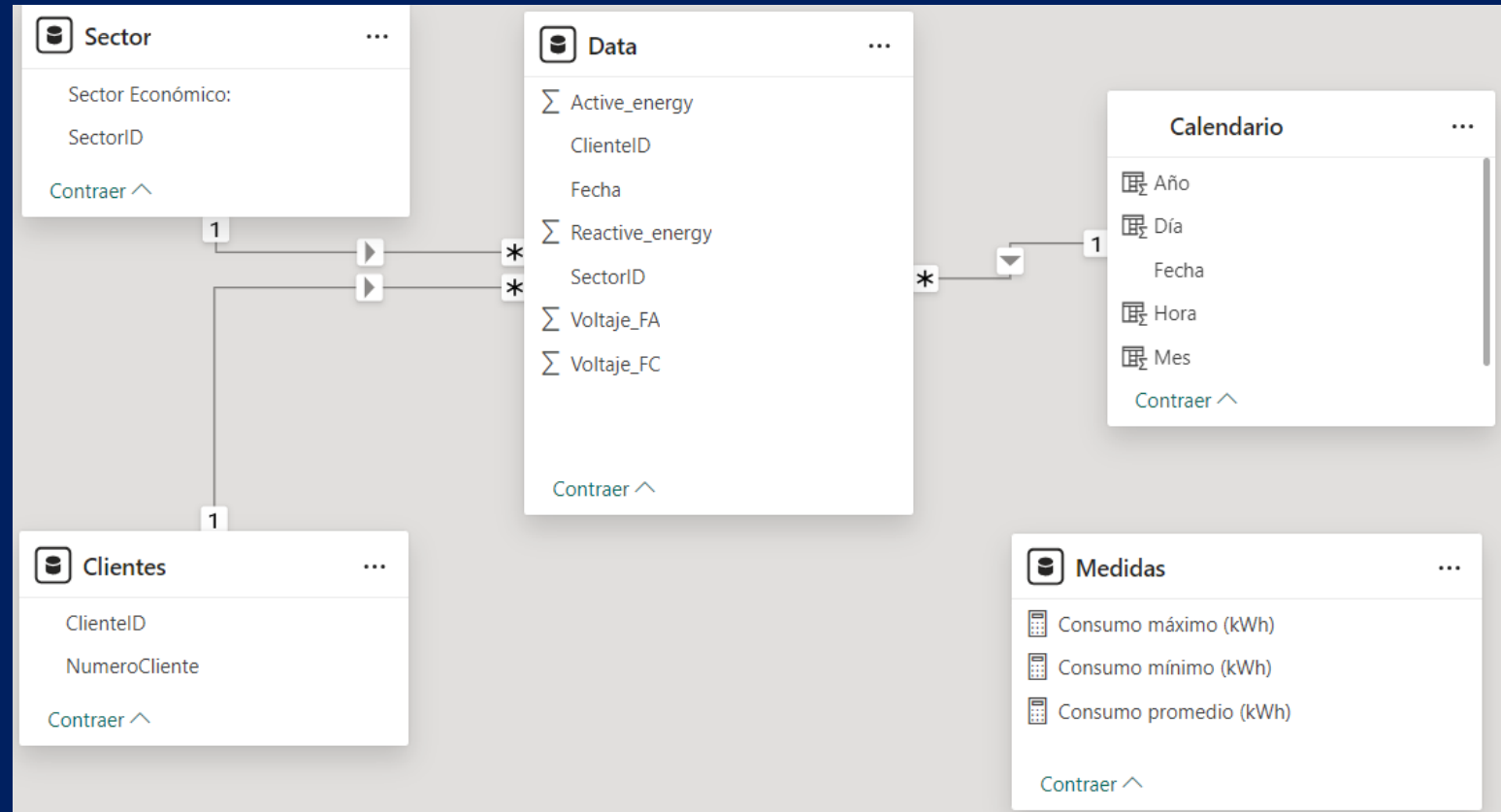


# Conjuntos de datos necesarios para el funcionamiento del artefacto

DatosClientes.csv



SectorEconómico  
Clientes.csv



Para empezar, la herramienta considera el desarrollo de un modelo estrella, donde la tabla de hechos es el concatenado de los datos de todos los clientes y las dimensiones son las características de los clientes. Para esto, también fue integrada una tabla calendario y una tabla de medidas

# ETL – Extraer, Transformar y Cargar

Asumimos que la información será suministrada vía CSV por la empresa, sin embargo, también es posible considerar algún tipo de conexión a las bases de datos del cliente. Una vez adquirida la data, esta es versionada (con la implementación de git) y después procesada en la staging área con Python. Algunas de las transformaciones requeridas están a continuación:

```
# Adquirir los archivos CSV y almacenarlos en la lista
for i in range(1, 31):
    file_name = f'DATOSCLIENTE{i}.csv'
    df = pd.read_csv(file_name)
    df['NumeroCliente'] = str("Cliente "+str(i))
    dataframes.append(df)

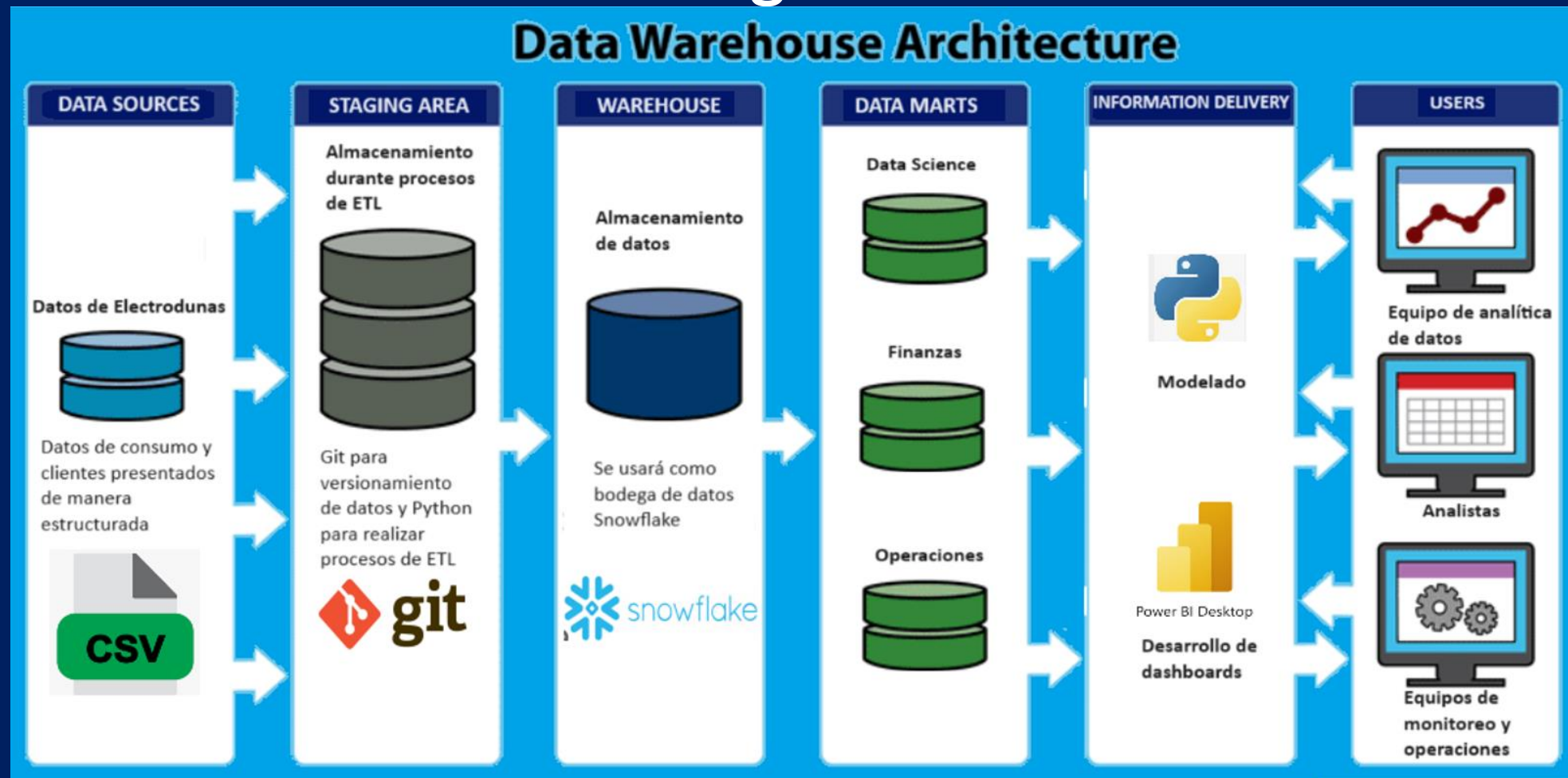
# Extraer Mes, Día y Hora
combined_df = pd.concat(dataframes)
combined_df['Mes'] = pd.to_datetime(combined_df['Fecha']).dt.month
combined_df['Día'] = pd.to_datetime(combined_df['Fecha']).dt.day
combined_df['Hora'] = pd.to_datetime(combined_df['Fecha']).dt.hour

# Adquirir el archivo del sector económico
archivo_excel = 'sector_economico_clientes (2).xlsx'
dataframe = pd.read_excel(archivo_excel)
dataframe = dataframe.rename(columns={'Cliente': 'NumeroCliente'})
dataframe['NumeroCliente'] = dataframe['NumeroCliente'].str.strip()

# Realizar la unión (join) basado en la columna "NumeroCliente"
combined_df = combined_df.merge(dataframe[['NumeroCliente', 'Sector Económico']], on='NumeroCliente', how='left')
```

Finalizada la etapa de transformación, es necesario proceder a separar la data según Data Mart y cargar en un servicio SaaS (Snowflake en nuestro caso) para ser usada en el desarrollo de Dashboards y modelamiento.

# Diagrama



El diagrama de la arquitectura está sujeto a cambios en cuanto a la manera de adquirir la data, teniendo en cuenta el alcance que pueda tener el proyecto en el MVP2.

# Mockup funcionamiento artefacto. Relación mockup y prototipo fachada



El funcionamiento de la herramienta permitirá la segmentación por cliente, por sector y por fecha, actualizando las medidas y KPIs seleccionados automáticamente. Además, mostrará la segmentación de clustering comentada anteriormente. En la visualización, es posible apreciar la aplicación de filtros, en este caso Sector Económico = Cultivo de Hortalizas.

# Gracias