

## ETL Factable Sales

Para optimizar el almacenamiento de transacciones de ventas en un **Data Warehouse** utilizando **Pentaho Data Integration (PDI)**, seguiría los siguientes pasos:

---

### 1. Diseño de la Tabla de Hechos y Dimensiones

Las tablas de dimensiones y hechos en el Data Warehouse se diseñarían de la siguiente manera:

- **Dimensión Clientes (Dim\_Clientes):** ID\_Cliente (PK), Nombre, Email, Ubicación, etc.
  - **Dimensión Productos (Dim\_Productos):** ID\_Producto (PK), Nombre\_Producto, Categoría, Precio\_Unitario, etc.
  - **Dimensión Tiempo (Dim\_Tiempo):** ID\_Tiempo (PK), Fecha, Mes, Año, Trimestre, etc.
  - **Tabla de Hechos Ventas (Fact\_Ventas):**
    - ID\_Venta (PK)
    - ID\_Cliente (FK)
    - ID\_Producto (FK)
    - ID\_Tiempo (FK)
    - Unidades\_Vendidas
    - Precio\_Unitario
    - Monto\_Total
- 

### 2. Implementación en Pentaho Data Integration (PDI)

Para construir y poblar la **tabla de hechos**, utilizaría las siguientes herramientas y operadores de Pentaho:

#### a. Extracción de Datos (ETL - Extract)

- **Input Steps:**
  - *Table Input:* Para extraer datos de las fuentes de ventas, clientes y productos.
  - *CSV File Input* (si los datos vienen en archivos CSV o Excel).

#### b. Transformación de Datos (Transform)

##### 1. Obtener los ID de las Dimensiones:

- Utilizar el operador "**Stream Lookup**" para buscar el **ID\_Cliente**, **ID\_Producto** y **ID\_Tiempo** en las respectivas tablas de dimensiones.

- Si el registro no existe en la dimensión, utilizar "**Dimension Lookup/Update**" para insertarlo.

## 2. Cálculo de Indicadores Clave:

- *Unidades\_Vendidas*: Tomar de los datos de la transacción.
- *Monto\_Total*: Usar "**Calculator**" o "**Formula**" para calcular:  
$$\text{Monto\_Total} = \text{Unidades\_Vendidas} \times \text{Precio\_Unitario}$$

## 3. Generación de la Clave de Tiempo:

- Usar "**Date Dimension Lookup**" o generar un ID basado en la fecha.

## c. Carga en la Tabla de Hechos (Load)

- **Table Output**: Para insertar datos en la tabla de hechos **Fact\_Ventas**.
- Habilitar el modo **Batch Insert** para mejorar el rendimiento.

---

## 3. Flujo de Trabajo en Pentaho (ETL)

1. **Extracción** de datos de ventas, clientes y productos.
2. **Búsqueda de claves** de dimensiones usando *Stream Lookup*.
3. **Cálculo de métricas** como monto total y unidades vendidas.
4. **Carga de datos** en la tabla **Fact\_Ventas** mediante *Table Output*.

---

## Conclusión

Este enfoque en **Pentaho** permite una carga eficiente de datos en la **tabla de hechos**, asegurando integridad referencial con las dimensiones y optimizando consultas analíticas. Al usar herramientas como **Stream Lookup**, **Dimension Lookup/Update** y **Table Output**, garantizamos un procesamiento rápido y eficiente del ETL.