# Anexo 2.6 — Modelo de Datos (ComparaYa)

Versión 1.0 · Fecha: 2025-10-13 · Autores: Equipo ComparaYa

## 1. Objetivo

Describir el modelo de datos implementado en la base de datos PostgreSQL de ComparaYa, que soporta la ingesta, almacenamiento y análisis de precios mayoristas y de consumidor, incluyendo los procesos ETL y la trazabilidad de los datos.

## 2. Alcance

El modelo abarca las tablas de hechos, dimensiones, catálogos de referencia y las estructuras de control asociadas a los procesos ETL. No incluye aspectos de frontend ni la API REST.

## 3. Arquitectura General del Modelo

El modelo de datos de ComparaYa adopta un enfoque de esquema en estrella (Star Schema) para optimizar el análisis de precios. Existen dos tablas de hechos principales: PrecioConsumidor y PrecioMayorista, que se relacionan con diversas dimensiones como Región, Producto, Unidad, Mercado y Tiempo. Además, se incluyen tablas de control para los procesos ETL, garantizando trazabilidad y calidad de datos.

## 4. Diagramas ER

A continuación se describe la estructura conceptual del modelo. El diseño sigue un enfoque modular, con submodelos para cada dominio.

┌────────────────────────┐  
 │ RegionRef │  
 └────────────┬───────────┘  
 │  
 ┌────────────┴────────────┐  
 │ PrecioConsumidor │◄────────┐  
 └────────────┬────────────┘ │  
 │ │  
 ┌────────────┴────────────┐ │  
 │ PrecioMayorista │◄────────┘  
 └────────────┬────────────┘  
 │  
 ┌───────────────────┼───────────────────┐  
 │ │ │  
 UnidadRef\* ProductoRef\* SemanaRef/MercadoRef  
 (por dataset) (por dataset) (Tiempo/Ubicación)

## 5. Definición de Tablas Principales

### 5.1 PrecioConsumidor

Tabla de hechos que almacena los precios semanales del mercado consumidor. Cada registro representa un producto en una región y semana determinadas.

Campos principales:

- semana\_id (FK → SemanaRef)  
- region\_id (FK → RegionRef)  
- sector\_id (FK → SectorRef)  
- tipo\_punto\_id (FK → TipoPuntoRef)  
- grupo\_id (FK → GrupoRef)  
- producto\_id (FK → ProductoRefConsumidor)  
- unidad\_id (FK → UnidadRefConsumidor)  
- precio\_minimo, precio\_maximo, precio\_promedio (Decimal)

### 5.2 PrecioMayorista

Tabla de hechos que almacena los precios diarios del mercado mayorista. Cada registro vincula producto, mercado, región y unidad de comercialización.

Campos principales:

- fecha (Date)  
- region\_id (FK → RegionRef)  
- mercado\_id (FK → MercadoRef)  
- subsector\_id (FK → SubsectorRef)  
- producto\_id (FK → ProductoRefMayorista)  
- variedad\_id (FK → VariedadRef)  
- calidad\_id (FK → CalidadRef)  
- unidad\_id (FK → UnidadRefMayorista)  
- origen\_id (FK → OrigenRef)  
- volumen, precio\_minimo, precio\_maximo, precio\_promedio\_ponderado

### 5.3 EtlFileManifest / EtlRunLog

Registros de control que documentan la descarga, validación y carga de archivos fuente. Permiten mantener trazabilidad sobre el dataset procesado y su estado.

Campos clave de EtlFileManifest: dataset, anio, url, sha256, fecha\_descarga, estado.

Campos clave de EtlRunLog: dataset, etapa, anio, estado, inicio, fin, metricas.

## 6. Reglas de Integridad y Normalización

• Toda relación entre hechos y dimensiones se implementa mediante claves foráneas (FK).  
• Se normalizan productos y unidades en catálogos separados por dataset.  
• Se validan fechas y valores de precios para evitar duplicados y datos inconsistentes.  
• Los índices en campos de búsqueda (producto, región, unidad, fecha) optimizan el rendimiento.

## 7. Consideraciones de Rendimiento

El modelo utiliza índices en las claves primarias y foráneas más consultadas, como fecha, región, producto y unidad. Se contempla la posibilidad de particionar las tablas de hechos por año o dataset en futuras versiones, manteniendo integridad referencial.

## 8. Trazabilidad con el ERS

Este modelo implementa los requisitos funcionales RF04 (Procesamiento ETL), RF05 (Registro de metadatos) y RF06 (Exposición de API), además de los no funcionales RNF01–RNF05 (rendimiento, portabilidad y mantenibilidad).