



Trabajo ingeniería de datos ETL

24 - 09 - 2024

Trabajo  
Integrador  
ETL

Simon Londoño Londoño

Dennis Patrick Juilland Prada

Juan Felipe Muñoz



## Tabla de contenido

<b>1) Introducción</b>	<b>3</b>
<b>2) Tipo de fuente de datos</b>	<b>4</b>
<b>3) Mecanismos de transformación</b>	<b>5</b>
<b>4) Objetivo de la llegada de datos</b>	<b>7</b>
<b>5) Esquema visual del ETL</b>	<b>9</b>
<b>6) Modelo relacional de la ETL</b>	<b>9</b>
<b>7) Conclusiones</b>	<b>10</b>

## Tabla de figuras

Figura 1. Esquema visual ETL.....	9
Figura 2. Modelo relacional ETL.....	10

## 1) Introducción

El presente trabajo tiene como objetivo desarrollar un proceso de extracción, transformación y carga (ETL) de datos para la empresa Northwind Traders, utilizando un archivo SQL cargado en PostgreSQL. A través de este modelo analítico, se busca optimizar la estructura de los datos, permitiendo realizar consultas analíticas detalladas que faciliten la toma de decisiones empresariales. El enfoque principal es la integración de diversas fuentes de datos relacionadas con las operaciones comerciales de la empresa, proporcionando una base sólida para la evaluación de su rendimiento y crecimiento.

Las fuentes de datos utilizadas en este modelo incluyen tablas clave como clientes, productos, pedidos, proveedores, empleados y detalles de pedidos. Cada una de estas tablas contiene información relevante que se transforma para obtener insights valiosos. Por ejemplo, la tabla de clientes almacena nombres, direcciones y contactos, mientras que la tabla de productos incluye descripciones, precios y categorías. Esta diversidad de datos proporciona una visión integral del funcionamiento interno de la empresa y su cadena de suministro.

El proceso de transformación aplicado a los datos extraídos implica varias etapas clave, como la limpieza de datos, la eliminación de duplicados, la derivación de nuevas columnas a partir de los datos existentes y la agregación de información. Estos mecanismos permiten organizar los datos de manera eficiente y asegurar que solo se conserve la información relevante para el análisis. Además, se aplican uniones y filtrados para integrar diferentes tablas en una estructura unificada que facilite el análisis multidimensional.

El objetivo principal de este proceso es construir un modelo estrella optimizado para consultas analíticas. Esta estructura permitirá a la empresa Northwind Traders monitorear y analizar sus ventas de manera detallada, evaluar el desempeño de sus productos, segmentar a sus clientes por región, y analizar la efectividad de su personal de ventas. Además, el modelo está diseñado para detectar patrones temporales mediante el análisis por meses, trimestres o días de la semana, lo que facilita la identificación de tendencias y estacionalidades.

Finalmente, este modelo ETL proporciona una base robusta para la creación de cubos OLAP, los cuales permiten un análisis dinámico de los datos. Con esta capacidad, los usuarios podrán profundizar en los detalles de las operaciones comerciales o resumir la información a niveles más altos, logrando así insights estratégicos. El enfoque del trabajo en la integración, transformación y agregación de datos busca mejorar significativamente la toma de decisiones en Northwind Traders.

## 2) Tipo de fuente de datos

Las **fuentes de datos** constituyen el corazón de cualquier modelo analítico. En este caso, se extraen datos de tablas clave relacionadas con clientes, productos, empleados, proveedores y pedidos, lo que proporciona una visión integral del negocio. La diversidad de estas fuentes permite analizar cada aspecto de las operaciones comerciales y facilita la identificación de patrones o problemas que puedan afectar la rentabilidad o eficiencia operativa. Al integrar diferentes tipos de datos, Northwind Traders puede realizar análisis exhaustivos y tomar decisiones basadas en información completa y precisa.

- a) **Tablas de Clientes:** La información sobre los clientes es fundamental para comprender el comportamiento de compra y segmentar adecuadamente el mercado. Estas tablas contienen detalles críticos como nombres, direcciones, y contactos que permiten identificar patrones de compra por región o sector, así como personalizar las estrategias de marketing y ventas. Esto permite a Northwind Traders realizar un análisis más preciso sobre las necesidades de sus clientes y ajustar sus estrategias comerciales en función de estos datos.
- b) **Tablas de Productos:** Los datos sobre los productos disponibles son esenciales para gestionar el inventario y optimizar la oferta de la empresa. Estas tablas incluyen información detallada sobre los productos, como descripciones, precios y categorías, lo que permite identificar qué productos tienen mayor demanda, cuáles necesitan promociones o ajustes de precio, y cómo fluctúan las ventas según la categoría. Con esta información, Northwind Traders puede ajustar su inventario y políticas de precios de manera más eficiente.
- c) **Tablas de Pedidos:** Las tablas de pedidos recogen todas las transacciones que se han realizado, ofreciendo una visión completa del historial de compras. Estos datos incluyen fechas, estados de los pedidos y relaciones con los clientes. Analizar esta información es crucial para evaluar el flujo de ventas, identificar periodos de alta o baja demanda, y rastrear la eficiencia en el procesamiento y entrega de los pedidos. Esto facilita a la empresa mejorar la experiencia del cliente y optimizar el ciclo de ventas.
- d) **Tablas de Proveedores:** Estas tablas contienen información sobre los proveedores, como nombres y datos de contacto, lo que es esencial para gestionar relaciones con los proveedores y garantizar un suministro constante de productos. Evaluar la efectividad de los proveedores y los tiempos de entrega puede ayudar a mejorar la cadena de suministro, reducir costos y asegurar que los productos estén disponibles cuando los clientes los necesitan. Con estos datos, Northwind Traders puede renegociar contratos o cambiar de proveedores en caso de problemas con el abastecimiento.
- e) **Tablas de Empleados:** Los empleados juegan un papel clave en la operación diaria de la empresa, y las tablas relacionadas proporcionan información sobre

sus roles, contacto y desempeño. Estos datos permiten a la gerencia evaluar la productividad y eficiencia de los empleados, facilitando la toma de decisiones en cuanto a ascensos, capacitaciones o ajustes en la estructura operativa. Además, el análisis de estas tablas permite vincular el desempeño individual con el éxito de las ventas y las operaciones.

- f) **Tabla de Pedidos Detallados:** Esta tabla desglosa cada pedido en sus elementos más pequeños, proporcionando información detallada sobre cantidades, precios y descuentos aplicados a cada producto. Este nivel de granularidad es esencial para calcular los ingresos exactos y el rendimiento de ventas a nivel de producto. Además, facilita la identificación de patrones de compra y permite realizar análisis más detallados sobre cómo afectan los descuentos al comportamiento de los clientes y las ventas totales.
- g) **Tablas de Detalles de Pedidos:** Estas tablas registran cada transacción al nivel más detallado, incluyendo el precio unitario, las cantidades de producto y los descuentos aplicados. Esta información es crucial para analizar la rentabilidad de las ventas, evaluar la efectividad de las promociones y detectar posibles ineficiencias o errores en el procesamiento de órdenes. Northwind Traders puede utilizar estos datos para optimizar sus márgenes y ajustar sus estrategias de ventas y descuentos.

### 3) Mecanismos de transformación

Los **mecanismos de transformación** aplicados durante el proceso ETL son fundamentales para asegurar la calidad, coherencia y utilidad de los datos. Estos mecanismos incluyen la limpieza de datos, la eliminación de duplicados, la derivación de nuevas variables, la unión de diferentes tablas y la agregación de datos. Cada uno de estos pasos garantiza que los datos sean consistentes y que reflejen con exactitud la realidad operativa de la empresa. Además, estos mecanismos permiten estructurar los datos de forma que sean fácilmente consultables y manejables en el modelo estrella, optimizando así la velocidad y eficiencia de las consultas analíticas.

- a) **Limpieza de Datos (Data Cleansing):** La calidad de los datos es un aspecto crítico en cualquier análisis. Para garantizar que los resultados sean precisos, se eliminan columnas innecesarias y se gestionan los valores nulos. Por ejemplo, en la tabla `dim_employees`, los valores faltantes son reemplazados por `-1` para evitar problemas al momento de realizar cálculos o análisis. Este paso asegura que los datos utilizados en el modelo estén completos y listos para su análisis.
- b) **Eliminación de Duplicados (Data Deduplication):** La duplicación de datos puede llevar a resultados incorrectos o inflados durante el análisis. Por eso, en el caso de las tablas de pedidos, se eliminan los registros duplicados para

garantizar que cada transacción sea única. Al hacer esto, Northwind Traders se asegura de que las métricas de ventas y los análisis de rendimiento se basen en datos confiables, lo que permite tomar decisiones más precisas.

- c) **Derivación de Datos (Data Derivation):** Derivar datos a partir de los campos existentes enriquece el análisis sin la necesidad de recolectar información adicional. Por ejemplo, a partir de la fecha de pedido, se generan nuevas columnas que detallan el mes, día, año y trimestre, lo que facilita la evaluación de patrones estacionales o temporales. Además, se crean campos como `full_name` en la tabla de empleados o se calculan montos totales de los pedidos a partir de precios, cantidades y descuentos, permitiendo un análisis más profundo y detallado de las ventas.
- d) **Unión de Datos (Data Joining):** En el proceso ETL, unir datos de varias tablas es esencial para formar una estructura coherente y unificada que facilite el análisis. Al combinar tablas como órdenes, productos y detalles de pedidos, se crea una tabla de hechos que centraliza toda la información necesaria para el análisis de ventas. Esta unificación permite a Northwind Traders realizar análisis multidimensionales que incluyan diversas perspectivas, como la combinación de datos de clientes con productos comprados, lo que mejora la capacidad de generar insights valiosos.
- e) **Integración de Datos (Data Integration):** La integración de distintas fuentes de datos es clave para proporcionar una visión holística de las operaciones comerciales. Este proceso asegura que los datos de empleados, productos, pedidos y detalles de pedidos se combinen en una sola tabla unificada. Al tener toda la información en un solo lugar, Northwind Traders puede realizar análisis más robustos y precisos, facilitando la toma de decisiones a nivel gerencial y operativo.
- f) **Agregación de Datos (Data Aggregation):** El proceso de agregación permite resumir grandes volúmenes de datos y transformarlos en métricas significativas. Por ejemplo, se calcula el monto total de cada pedido sumando precios, cantidades y aplicando descuentos. Este tipo de agregación facilita la creación de reportes y análisis de rentabilidad a nivel de producto, cliente o periodo, lo que ayuda a la empresa a enfocar sus esfuerzos en las áreas más rentables.
- g) **Filtrado de Datos (Data Filtering):** El filtrado de datos es esencial para reducir el volumen de información a lo realmente relevante para el análisis. Por ejemplo, en la tabla de clientes, se seleccionan columnas como `customer_id`, `company_name`, `contact_name`, `city` y `region` para concentrarse en la información necesaria para un análisis geográfico o de segmentación. Este proceso optimiza el uso de los datos y mejora la eficiencia en el análisis.

#### 4) Objetivo de la llegada de datos

El **objetivo de la llegada de los datos** en este proceso ETL es crear un **modelo estrella** optimizado que facilite la toma de decisiones empresariales. Este modelo permite realizar consultas multidimensionales sobre las operaciones de Northwind Traders, evaluando ventas, clientes, productos y empleados de manera segmentada o en conjunto. Gracias a este modelo, la empresa puede obtener insights estratégicos y operativos en tiempo real, ajustando sus estrategias de ventas, marketing y logística según los resultados obtenidos. El modelo estrella, por tanto, es una herramienta clave para la agilidad y precisión en la gestión empresarial.

El objetivo principal del proceso de ETL implementado en este trabajo es construir un **modelo estrella** optimizado para consultas analíticas en la empresa Northwind Traders. Este modelo está diseñado para organizar los datos de manera que sea fácil acceder a información clave sobre las ventas, clientes, productos, empleados y proveedores. Al estructurar los datos en un modelo estrella, se busca facilitar el análisis rápido y eficiente, proporcionando insights que apoyen la toma de decisiones estratégicas en la empresa.

1. **Monitoreo y análisis detallado de ventas:** Una vez que los datos han sido transformados e integrados, el modelo permite a Northwind Traders realizar un monitoreo exhaustivo de sus ventas. Esto incluye información clave como el número de orden, precio unitario, cantidad de productos, flete, descuentos aplicados y el monto total de cada transacción. Tener acceso a estos datos organizados en un modelo estrella facilita el análisis detallado y la generación de reportes sobre el rendimiento comercial.
2. **Segmentación de clientes, productos, empleados y proveedores:** El modelo permite segmentar la información en categorías clave como clientes, productos y empleados, lo que facilita un análisis más profundo. Por ejemplo, se puede analizar el comportamiento de compra de los clientes por región, el rendimiento de los productos en diferentes categorías o la efectividad de los empleados en la gestión de ventas. Esta segmentación permite a Northwind Traders tomar decisiones más precisas sobre marketing, ventas y operaciones.
3. **Análisis de patrones temporales:** El modelo estrella también permite analizar los datos desde una perspectiva temporal, lo que es esencial para identificar patrones y tendencias. Gracias a la dimensión de tiempo, Northwind Traders



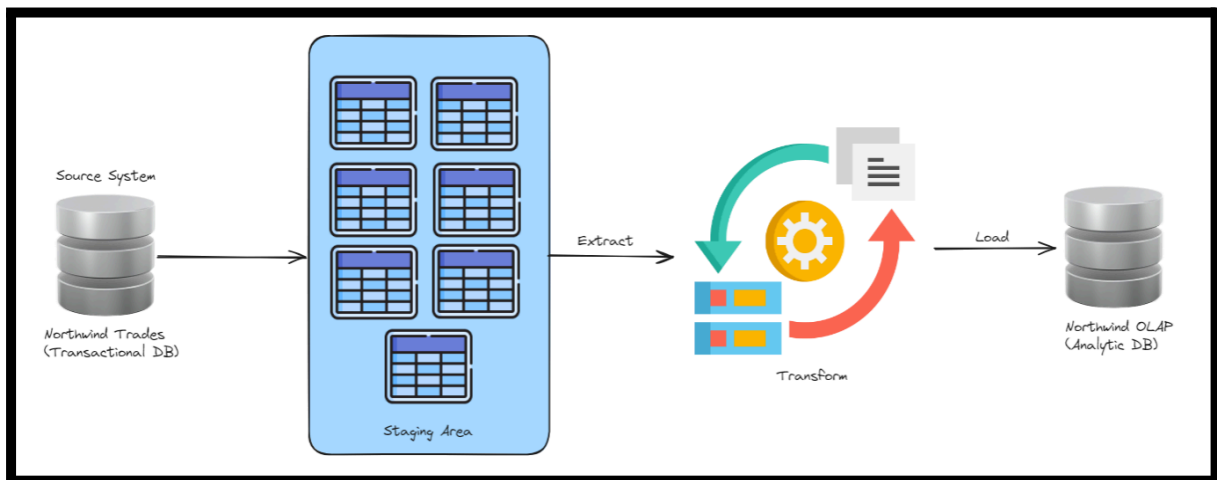
puede evaluar las ventas por meses, trimestres y días de la semana, facilitando la identificación de tendencias estacionales o fluctuaciones periódicas. Este análisis es vital para ajustar estrategias de inventario, promociones y personal, optimizando los recursos de la empresa.

4. **Consultas multidimensionales:** Una de las principales ventajas del modelo estrella es que permite realizar consultas multidimensionales sobre los datos. Esto significa que Northwind Traders puede analizar sus ventas desde diferentes perspectivas, como por categoría de producto, región, empleado o periodo de tiempo. Este enfoque multidimensional proporciona una visión más amplia y profunda de las operaciones comerciales, permitiendo a la empresa identificar oportunidades de mejora y optimización.
5. **Agregación rápida de datos:** El modelo está diseñado para permitir la agregación rápida de datos, lo que facilita la generación de reportes y métricas clave para la toma de decisiones. Northwind Traders puede utilizar estas capacidades para generar reportes sobre ventas totales, ingresos por producto o cliente, análisis de descuentos y evaluaciones del rendimiento de empleados. La capacidad de agregar y resumir datos rápidamente es esencial para ofrecer insights en tiempo real que apoyen la gestión operativa y estratégica.
6. **Construcción de cubos OLAP:** Finalmente, el modelo estrella también está optimizado para la construcción de cubos OLAP (Online Analytical Processing), que permiten realizar análisis dinámicos de los datos. Los cubos OLAP proporcionan a los usuarios la capacidad de profundizar en los detalles de las operaciones comerciales o resumir la información a niveles más altos, lo que facilita la obtención de insights estratégicos y la generación de reportes ejecutivos. Esto permite a Northwind Traders realizar análisis más interactivos y profundos, apoyando la toma de decisiones basadas en datos.

El proceso de ETL y el modelo estrella resultante no solo optimizan la estructura de los datos, sino que también proporcionan una base sólida para el análisis de la empresa, mejorando la eficiencia en la toma de decisiones y permitiendo un análisis detallado y preciso de las operaciones comerciales.

## 5) Esquema visual del ETL

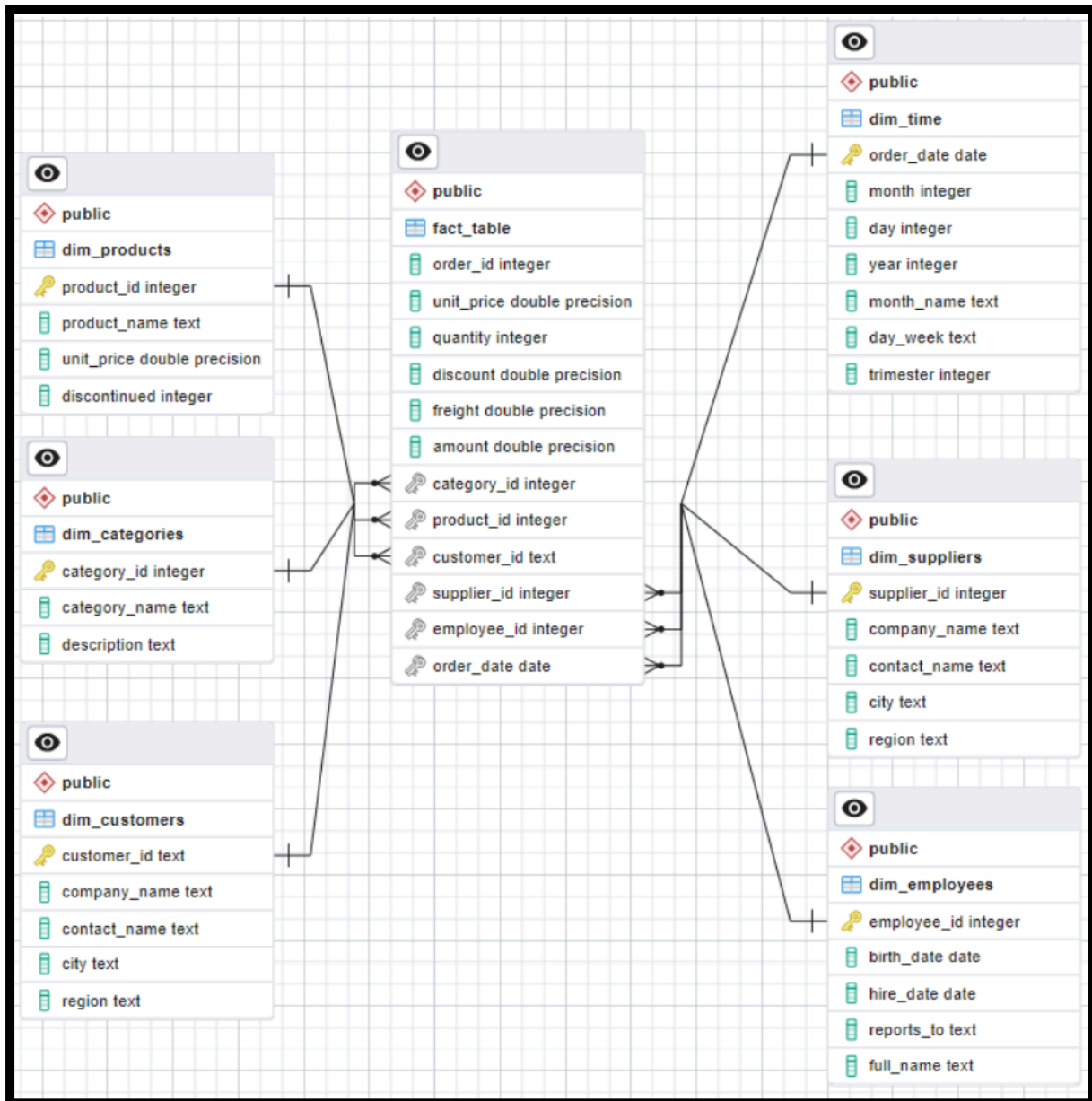
El **esquema visual del ETL** proporciona una representación gráfica clara de todo el proceso, facilitando la comprensión de cómo fluyen los datos desde sus fuentes originales hasta su transformación en un modelo estrella optimizado. Este esquema ayuda a visualizar las relaciones entre los diferentes pasos del proceso ETL, las transformaciones aplicadas y cómo los datos se integran en una estructura final. Tener una representación visual facilita la comunicación del proceso a diferentes equipos dentro de la empresa y garantiza que todos los involucrados comprendan el flujo de trabajo y la importancia de cada paso.



**Figura 1.** *Esquema visual ETL*

## 6) Modelo relacional de la ETL

El **modelo relacional de la ETL** es una pieza crucial para organizar los datos de manera lógica, manteniendo la integridad y relaciones entre diferentes tablas. Este modelo organiza los datos en un formato relacional, asegurando que las fuentes de datos estén correctamente interconectadas y que se mantenga la consistencia entre ellas. Además, este modelo facilita la implementación de consultas complejas, permitiendo realizar análisis que abarcan múltiples dimensiones y tablas. En el contexto de Northwind Traders, el modelo relacional permite aprovechar al máximo la información disponible para generar reportes y análisis profundos que apoyen la toma de decisiones estratégicas.



**Figura 2.** Modelo relacional ETL

## 7) Conclusiones

El desarrollo de este proceso ETL para Northwind Traders ha sido clave para la optimización del manejo de datos dentro de la empresa. En primer lugar, la integración de diversas fuentes de datos como clientes, productos, empleados y pedidos ha permitido una visión global del negocio, proporcionando una base sólida para análisis más precisos y completos. Esta integración mejora la capacidad de tomar decisiones fundamentadas en datos, optimizando tanto las operaciones comerciales como las estrategias de marketing y ventas.

Los mecanismos de transformación aplicados durante el proceso, como la limpieza de datos, eliminación de duplicados, derivación de nuevas columnas y la unión de diferentes tablas, aseguran que los datos sean consistentes, confiables y fácilmente utilizables. Estas transformaciones no solo mejoran la calidad de los datos, sino que también permiten estructurarlos de manera óptima para su consulta y análisis en un modelo estrella. La eficiencia y precisión de las consultas resultantes son fundamentales para la generación de insights estratégicos.

El modelo estrella implementado permite a Northwind Traders realizar consultas multidimensionales y segmentar clientes, productos y empleados, facilitando un análisis detallado de ventas y rendimiento. Además, el análisis de patrones temporales mediante dimensiones de tiempo ha proporcionado a la empresa una mejor comprensión de las tendencias estacionales y las fluctuaciones en el comportamiento del consumidor, lo que apoya la planificación estratégica y operativa.

La creación de un esquema visual del ETL ha sido crucial para facilitar la comprensión del flujo de datos y el proceso de transformación, lo que ha permitido una comunicación clara y eficiente entre los diferentes equipos de la empresa. Además, el modelo relacional asegura que los datos estén bien organizados y conectados, permitiendo consultas complejas que abarcan múltiples dimensiones.

En resumen, este proceso ETL ha proporcionado a Northwind Traders una estructura sólida para gestionar sus datos de manera eficiente, optimizando su capacidad de análisis y mejorando su toma de decisiones a nivel estratégico. El uso del modelo estrella, junto con la construcción de cubos OLAP, ha elevado la capacidad de análisis interactivo y profundo, permitiendo a la empresa identificar oportunidades de mejora y aumentar su competitividad en el mercado.