

# UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

## TECNOLOGÍAS DE LA INFORMACIÓN



### EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS

#### DISEÑO DE ESQUEMAS PARA DATA WAREHOUSE – CASO DE ESTUDIO: AGENCIA DE VIAJES “TU AVENTURA”

***IDGS91N***

PRESENTA EQUIPO 7:

SEBASTIÁN ACOSTA ORTIZ

IAN CARLOS CHÁVEZ ROJO

REGINA CHÁVEZ TAMAYO

IVÁN EDUARDO MARTÍNEZ MARTÍNEZ

ERICK FABIÁN TERRAZAS HERNÁNDEZ

DOCENTE:

LUIS ENRIQUE MASCOTE CANO

Chihuahua, Chih., 09 de octubre de 2025

## Contents

Introducción .....	3
Identificación de hechos y dimensiones .....	3
Procesos analíticos clave (granos y hechos) .....	3
Dimensiones (con atributos sugeridos) .....	4
Justificación del modelo de esquema seleccionado .....	5
Descripción narrativa del flujo de datos (fuentes > DW) .....	5
Fuentes operacionales (ejemplos) .....	5
Proceso ETL/ELT (resumen) .....	5
Diagramas del esquema propuesto .....	6
Fundamento técnico de la lógica de diseño .....	6
Respuesta a requerimientos del caso .....	7
Conclusiones .....	7
Referencias .....	8

## Introducción

El presente documento describe el diseño de un sistema de Data Warehouse (DW) para la agencia de viajes “Tu Aventura”, empresa que organiza paquetes turísticos con proveedores de transporte, hoteles y guías en tres países. El objetivo es centralizar reservas y cancelaciones, medir la satisfacción y costos por paquete, y optimizar los esfuerzos de marketing mediante análisis integrados.

El trabajo presenta la identificación de hechos y dimensiones, la justificación del modelo de esquema, la descripción del flujo de datos, los diagramas conceptuales y el fundamento técnico que sustenta las decisiones de diseño.

## Identificación de hechos y dimensiones

### Procesos analíticos clave (granos y hechos)

- **HechoReservas** (grano: una fila por reserva confirmada de un paquete en una fecha):
  - Métricas (aditivas/semiaditivas): importeBruto, descuento, importeNeto, numPersonas, costoTotal, margen = importeNeto - costoTotal.
  - Indicadores de marketing: idCampania, canalPromocion, utmSource, utmMedium, utmCampaign (si aplica).
- **HechoCancelaciones** (grano: una fila por cancelación registrada):
  - Métricas: importeDevuelto, penalizacion, diasAntiguedadReserva (días entre la fecha de reserva y la cancelación).
  - Atributos: motivoCancelacion (codificado), tipoCliente, politicaAplicable.
- **HechoSatisfaccion** (grano: una fila por valoración emitida por cliente y proveedor/paquete):
  - Métricas: scoreGlobal (1–5), scoreHotel, scoreTransporte, scoreGuia, numRespuestas, npsBucket (Detractor, Neutral o Promotor).
- **HechoCosteoProveedor** (grano: una fila por servicio componente del paquete facturado por un proveedor):
  - Métricas: costoUnitario, costoTotal, noches, trayectos, horasServicio.

- **HechoMarketingConversiones** (grano: una fila por interacción de campaña atribuida a una reserva):
  - Métricas: clics, impresiones, costoCampania, reservasAtribuidas, ingresosAtribuidos, roi, cac.

#### Notas:

- Se separan Reservas y Cancelaciones para analizar tasas, motivos y períodos sin afectar las métricas de ventas.
- Satisfacción se desacopla para permitir múltiples encuestas por reserva y comparativas entre proveedores.
- CosteoProveedor permite medir la rentabilidad individual de cada componente del paquete.
- MarketingConversiones habilita el cálculo de ROI y atribución por canal o campaña.

#### Dimensiones (con atributos sugeridos)

- **DimTiempo** (idTiempo): fecha, diaSemana, semana, mes, trimestre, anio, esTemporadaAlta, esFeriado, semanaAnio, finDeSemana.
- **DimDestino** (idDestino): pais, region, ciudad, tipoDestino (playa, ciudad, naturaleza), popularidad, mesTemporadaAlta.
- **DimPaquete** (idPaquete): nombrePaquete, categoria (familiar, aventura, lujo), duracionDias, incluyeHotel, incluyeTransporte, incluyeGuia.
- **DimCliente** (idCliente): segmento (nuevo, recurrente, VIP), pais, edadRango, canalCaptacion, tipoCliente (individual, grupo, empresa).
- **DimProveedor** (idProveedor): tipoProveedor (hotel, transporte, guia), nombreProveedor, cadenaHotelera, categoriaHotel (estrellas), certificaciones, pais, ciudad.
- **DimCampania** (idCampania): canal (SEO, SEM, Redes, Email, Agente), nombreCampania, objetivo (awareness o performance), presupuesto, fechaInicio, fechaFin.
- **DimAgente** (idAgente): nombre, sede, experienciaAnios, certificaciones, idioma.

- **DimMotivoCancelacion** (idMotivoCancelacion): motivo (precio, disponibilidad, visado, enfermedad, otro), categoriaMotivo (controlable o no controlable).
- **DimMoneda** (idMoneda): codigo, simbolo, tipoCambioDia.
- **DimGeografia** (si se separa de Destino): pais, region, ciudad, latitud, longitud.

#### **Surrogate keys en todas las dimensiones. SCD:**

- Tipo 2 en Proveedor, Paquete y Canal/Campaña (cambios históricos relevantes).
- Tipo 1 en correcciones menores (p. ej., ortografía).

## **Justificación del modelo de esquema seleccionado**

Se propone un modelo en estrella (Star Schema) con hechos granulares y dimensiones desnormalizadas para mejorar la velocidad de consulta y facilitar el análisis OLAP.

El modelo es simple, flexible y alineado con los objetivos de negocio, permitiendo comparar temporadas, medir tasas de cancelación, evaluar proveedores y calcular ROI.

En casos jerárquicos, como proveedores o ubicaciones, puede aplicarse una forma en copo de nieve (Snowflake) para normalizar información sin afectar el rendimiento.

## **Descripción narrativa del flujo de datos (fuentes > DW)**

### **Fuentes operacionales (ejemplos)**

- **Sistema de reservaciones (OLTP):** reservas, detalles de pasajeros, pagos, cancelaciones, políticas.
- **Sistemas de proveedores:** facturación de hotel, rutas de transporte, horas de guía, calificaciones.
- **Plataformas de marketing:** campañas, costos diarios, clics/impresiones, parámetros UTM.
- **Encuestas/NPS:** formularios web o herramientas de CSAT/NPS.

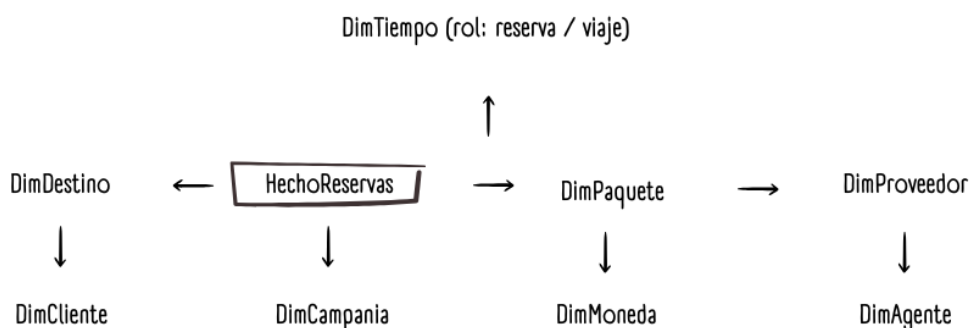
### **Proceso ETL/ELT (resumen)**

1. **Ingesta:** conectores REST/JDBC/SFTP; cargas diarias o intradía.
2. **Staging:** almacenamiento temporal con validaciones y control de duplicados.

3. **Transformación:** estandarización de monedas, catálogos y formatos.
4. **Carga de dimensiones:** generación de surrogate keys y manejo de SCD.
5. **Carga de hechos:** cálculo de métricas (margen, roi, diasAntiguedadReserva).
6. **Validación:** comparación con totales OLTP y publicación en el modelo semántico.
7. **Gobierno y seguridad:** control de accesos, enmascaramiento de datos y trazabilidad.

## Diagramas del esquema propuesto

El diseño del Data Warehouse se centra en el hecho HechoReservas, como núcleo del modelo analítico, al ser el punto donde convergen los procesos de ventas, clientes, destinos y campañas.



Los hechos Cancelaciones, Satisfaccion, CosteoProveedor y MarketingConversiones se modelan como estrellas independientes, conectadas a las mismas dimensiones para mantener coherencia y comparabilidad entre áreas de negocio.

## Fundamento técnico de la lógica de diseño

- Modelado dimensional (Kimball): separación entre hechos (eventos) y dimensiones (contexto).
- Granularidad definida: una fila por evento o transacción.
- Métricas aditivas y derivadas: importeNeto, costoTotal (aditivas); margen, roi (derivadas).
- Dimensiones conformadas: Tiempo, Destino, Paquete, Cliente y Campania.
- SCD Tipo 2: en Proveedor y Campania para conservar históricos.
- Data Quality: catálogos únicos, integridad y validaciones automáticas.
- Escalabilidad: particionamiento por fecha o país, índices por claves foráneas.
- BI Semántico: indicadores certificados (KPI) como tasa de cancelación, NPS y ROI.

## Respuesta a requerimientos del caso

- Reservas por destino (temporadas altas/bajas):  $\text{HechoReservas} + \text{DimTiempo (mes, esTemporadaAlta)} + \text{DimDestino} = \text{comparativos YoY/YTD}$ .
- Tasa de cancelación y motivos:  $\text{HechoCancelaciones} / \text{HechoReservas}$  por periodo; desgloses por  $\text{DimMotivoCancelacion}$ ,  $\text{DimDestino}$  y  $\text{DimPaquete}$ .
- Evaluación de proveedores:  $\text{HechoSatisfaccion} + \text{HechoCosteoProveedor} = \text{calidad/costo por hotel, transporte o guía}$ .
- ROI de campañas:  $\text{HechoMarketingConversiones}$  unido a  $\text{HechoReservas}$  por  $\text{DimCampania} = \text{ROAS, ROI, CAC y funnel por canal}$ .

## Conclusiones

El modelo propuesto para el Data Warehouse de “Tu Aventura” ofrece una estructura sólida y eficiente para consolidar información de reservas, cancelaciones, satisfacción, costos y marketing.

Su diseño en estrella permite un análisis claro y escalable, con dimensiones conformadas y control de versiones históricas.

Esta solución facilitará la toma de decisiones estratégicas basadas en datos reales, mejorando la rentabilidad de campañas, la gestión de proveedores y la satisfacción de los clientes, fortaleciendo el posicionamiento competitivo de la agencia.

## Referencias

ChatGPT. (2025). *Consulta de información sobre diseño de esquemas de Data Warehouse*.

OpenAI.

IBM. (2024). *Best practices for building a data warehouse*. Recuperado de

<https://www.ibm.com/topics/data-warehouse>

Google Cloud. (2023). *Data warehouse concepts overview*. Recuperado de

<https://cloud.google.com/learn/what-is-a-data-warehouse>

Microsoft Learn. (2024). *Diseño de modelos dimensionales para inteligencia empresarial*.

Recuperado de <https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>

Oracle. (2024). *Data warehouse architecture and design best practices*. Recuperado de

<https://docs.oracle.com/en/solutions/data-warehouse-design/>

SAP. (2023). *Dimensional modeling in data warehouse cloud*. Recuperado de

<https://help.sap.com/docs/>

Tableau. (2024). *What is a star schema?* Recuperado de

<https://www.tableau.com/learn/articles/star-schema>