

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA
TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN



**DISEÑO DE ESQUEMAS PARA DATA WAREHOUSE CASO DE
ESTUDIO ASIGNADO**

MATERIA: Extracción de Conocimiento en Bases de Datos

MAESTR@: Enrique Mascote

ALUMNOS: Kevin Iván Aguirre Silva

Jatzel Israel Cruz Castruita

Jorge Alejandro Hernández Contreras

Carlos Adrián Mata Nevárez

Erick Adrián Sánchez Cervantes

GRUPO: IDGS91N

FECHA:09/10/2025

ÍNDICE

| | |
|--|----|
| INTRODUCCIÓN..... | 1 |
| IDENTIFICACIÓN DE HECHOS Y DIMENSIONES | 2 |
| 1. Hechos..... | 2 |
| a) Hecho Siniestro | 2 |
| b) Hecho Póliza | 2 |
| c) Hecho Pago | 3 |
| 2. Dimensiones | 3 |
| 3. Síntesis | 4 |
| JUSTIFICACIÓN DEL MODELO DE ESQUEMA SELECCIONADO | 5 |
| DESCRIPCIÓN NARRATIVA DEL FLUJO DE DATOS..... | 7 |
| Fase 1: Recolección de Datos (Extracción)..... | 7 |
| Fase 2: Puesta a Punto (Transformación)..... | 8 |
| Fase 3: Almacenamiento Final (Carga) | 8 |
| DIAGRAMAS CONCEPTUALES Y FÍSICOS DEL ESQUEMA | 9 |
| FUNDAMENTO TÉCNICO SOBRE LA LÓGICA DE DISEÑO | 10 |
| CONCLUSIÓN..... | 12 |
| REFERENCIAS | 13 |

INTRODUCCIÓN

El documento desarrolla el diseño conceptual de un Data Warehouse (DW) para la compañía “ProtecSeguros”, con el fin de soportar la toma de decisiones estratégicas relacionadas con el análisis de siniestros, detección de fraudes y optimización de primas.

El objetivo del diseño es consolidar los datos operacionales en un entorno analítico que permita detectar patrones de riesgo, analizar costos de indemnización y evaluar el desempeño de clientes y pólizas.

IDENTIFICACIÓN DE HECHOS Y DIMENSIONES

El diseño del modelo multidimensional del Data Warehouse para la compañía ProtecSeguros parte del análisis de sus principales procesos de negocio: emisión de pólizas, registro de siniestros e indemnizaciones a clientes o terceros. A partir de estos procesos, se identifican los hechos (eventos medibles) y las dimensiones (atributos descriptivos) que permitirán realizar análisis estratégicos relacionados con riesgo, fraude, eficiencia operativa y comportamiento de los clientes.

1. Hechos

Los hechos representan los eventos cuantificables del negocio que serán objeto de análisis. En el caso de ProtecSeguros, se identifican los siguientes hechos principales:

a) Hecho Siniestro

Registra los reclamos o siniestros reportados por los clientes y constituye una de las principales fuentes de análisis del DW.

Medidas (indicadores):

- Número de siniestros registrados.
- Monto total del reclamo.
- Tiempo de resolución del siniestro (en días).
- Monto pagado al cliente o proveedor.
- Clasificación del siniestro como legítimo o potencial fraude.

Dimensiones relacionadas:

Tiempo, Cliente, Póliza, Tipo de Siniestro, Región, Agente, Proveedor.

b) Hecho Póliza

Representa las operaciones relacionadas con la emisión, renovación o cancelación de pólizas de seguros.

Medidas (indicadores):

- Prima cobrada.
- Duración o antigüedad de la póliza (en años).
- Estado de la póliza (activa, vencida, cancelada).
- Número de siniestros asociados.

Dimensiones relacionadas:

Tiempo, Cliente, Tipo de Seguro, Región, Agente.

c) Hecho Pago

Registra los desembolsos realizados por concepto de indemnizaciones o pagos a terceros vinculados a un siniestro.

Medidas (indicadores):

- Monto del pago.
- Fecha del pago.
- Número de pagos por siniestro.
- Tipo de destinatario (cliente, taller, hospital, etc.).

Dimensiones relacionadas:

Tiempo, Cliente, Tipo de Siniestro, Proveedor, Región.

2. Dimensiones

Las dimensiones describen los diferentes contextos desde los cuales se pueden analizar los hechos, permitiendo explorar la información mediante filtros, jerarquías y agrupaciones.

| Dimensión | Descripción | Atributos principales |
|----------------|---|--|
| Tiempo | Permite analizar la evolución de los hechos en diferentes períodos. | Año, Trimestre, Mes, Día, Semana |
| Cliente | Contiene información demográfica y de comportamiento del asegurado. | ID Cliente, Nombre, Edad, Género, Estado civil, Ocupación, Ingreso estimado, Región, Historial de siniestros. |
| Póliza | Describe las características de las pólizas contratadas. | D Póliza, Tipo de seguro (Auto, Hogar, Vida), Fecha de inicio, Fecha de vencimiento, Estado, Cobertura, Monto asegurado. |

| | | |
|----------------------------|---|---|
| Tipo de Siniestro | Clasifica los eventos que generan reclamos. | ID Tipo, Categoría (Accidente, Robo, Daño, Fallecimiento), Descripción. |
| Región | Permite analizar los hechos según la localización geográfica. | ID Región, Ciudad, Estado, País |
| Agente | Describe al personal responsable de la gestión comercial. | ID Agente, Nombre, Oficina, Zona, Años de experiencia. |
| Proveedor / Tercero | Identifica a los proveedores o terceros que reciben pagos por servicios asociados a un siniestro. | ID Proveedor, Tipo (Taller, Hospital, etc.), Nombre, Ciudad, Estado |

3. Síntesis

El modelo propuesto permite que los hechos principales —Siniestro, Póliza y Pago— se analicen desde múltiples perspectivas proporcionadas por las dimensiones Tiempo, Cliente, Región, Agente, Póliza, Tipo de Siniestro y Proveedor.

Este diseño facilitará la generación de indicadores clave de desempeño (KPIs) como:

- Tasa de siniestros por tipo de seguro o región.
- Costo promedio de indemnización.
- Frecuencia de reclamos por cliente.
- Evaluación del perfil de riesgo y detección de posibles fraudes.

De esta manera, el Data Warehouse propuesto proporcionará una base sólida para la toma de decisiones estratégicas en ProtecSeguros.

JUSTIFICACIÓN DEL MODELO DE ESQUEMA SELECCIONADO

Para el diseño del Data Warehouse de “ProtecSeguros” decidimos utilizar el modelo de esquema en estrella, ya que es el que mejor se adapta a las necesidades del caso y facilita mucho el análisis de información. Este modelo se caracteriza por tener una estructura sencilla pero muy funcional, lo cual es ideal para una empresa como “ProtecSeguros”, que maneja grandes volúmenes de datos sobre pólizas, clientes, siniestros y pagos.

El modelo en estrella tiene una tabla de hechos central, que en este caso sería la tabla Siniestros. En ella se registran los datos principales del negocio, como los montos de indemnización, el número de reclamos, los tiempos de resolución y los costos involucrados. Alrededor de esta tabla se encuentran las tablas de dimensiones, que en nuestro diseño son: Cliente, Póliza, Tipo de Siniestro, Tiempo, Región y Tercero. Cada una aporta una perspectiva diferente del mismo hecho, permitiendo hacer análisis más profundos y detallados.

Por ejemplo, el área de análisis de “ProtecSeguros” podría consultar cuántos siniestros hubo en una región específica, cuánto se pagó por cada tipo de seguro o qué perfiles de clientes tienden a tener más reclamos. Gracias a la estructura del modelo en estrella, estos análisis se pueden realizar de manera más rápida y sin complicaciones, ya que la información está organizada de forma clara y directa.

Una de las principales razones por las que elegimos este modelo es porque ofrece un excelente rendimiento en consultas. A diferencia de otros modelos como el de copo de nieve, el esquema en estrella no necesita hacer tantas uniones entre tablas, lo cual hace que los reportes y consultas sean más ágiles. Además, al tener las dimensiones desnormalizadas, los analistas pueden entender fácilmente la estructura de los datos y trabajar con ellos sin necesidad de conocimientos técnicos avanzados. Otra ventaja importante es su flexibilidad y escalabilidad. Si en el futuro “ProtecSeguros” decide ampliar sus servicios, por ejemplo, incluyendo seguros médicos o empresariales, se podrían agregar nuevas dimensiones o métricas sin necesidad de modificar por completo el modelo

existente. Esto permite que el Data Warehouse crezca junto con la empresa y siga siendo útil a largo plazo.

También consideramos que este modelo ayuda a mejorar la toma de decisiones estratégicas, ya que facilita la identificación de patrones de riesgo, posibles fraudes o comportamientos repetitivos en los siniestros. Con la información bien organizada y accesible, la dirección de la empresa puede detectar tendencias, ajustar las primas según el perfil del cliente y optimizar los tiempos de respuesta ante reclamos.

DESCRIPCIÓN NARRATIVA DEL FLUJO DE DATOS

El proceso para llevar la información al Data Warehouse (DW) de “ProtecSeguros” se organiza a través del conocido ciclo de Extracción, Transformación y Carga (ETL) (aws amazon, 2025).

La meta principal de este sistema es sencilla: tomar datos de todas las fuentes operacionales dispersas y unificarlas en una base analítica coherente y pulcra, lista para que la dirección pueda tomar decisiones estratégicas.

Fase 1: Recolección de Datos (Extracción)

Todo empieza extrayendo la información directamente de los sistemas transaccionales y de algunas fuentes externas cruciales. ¿Qué sistemas son los principales?

- El sistema de gestión de pólizas, que registra absolutamente todo sobre los seguros: contratos, renovaciones, fechas clave y, por supuesto, los montos de las primas.
- El sistema de siniestros y reclamos, que es donde se registran todos los eventos reportados por los clientes, incluyendo el tipo de incidente, los costos reales y cómo va el proceso de resolución.
- El sistema financiero-contable, encargado de manejar los pagos de primas, las indemnizaciones y toda la relación con terceros (como talleres, hospitales o aseguradoras asociadas).

Además, no nos olvidamos de las fuentes externas, que son vitales. Se usan bases de datos de zonas de riesgo, estadísticas demográficas y reportes de siniestralidad por región para enriquecer el análisis de perfiles y la posible detección de fraude.

Fase 2: Puesta a Punto (Transformación)

Una vez que tenemos los datos, hay que "limpiarlos" y darles forma. En la fase de transformación, aplicamos procesos intensivos de depuración, estandarización y enriquecimiento. Esto significa eliminar los duplicados, corregir inconsistencias (nombres, fechas, formatos de moneda) y calcular nuevos indicadores.

Por ejemplo, se podría generar el índice de frecuencia de siniestros por cliente, la antigüedad de una póliza o la tasa promedio de indemnización por zona.

Es aquí también donde se asignan las claves sustitutas (o surrogate keys) a las dimensiones, asegurando que los registros mantengan su consistencia histórica.

Fase 3: Almacenamiento Final (Carga)

Finalmente, los datos ya transformados se integran en las estructuras del DW. La información se organiza típicamente en la tabla de hechos "Siniestros", que es el núcleo y concentra las métricas principales (como el monto de indemnización, número de reclamos, días de resolución, etc.).

Esta tabla se vincula a sus dimensiones correspondientes: Cliente, Póliza, Tipo de Siniestro, Tiempo, Región y Tercero.

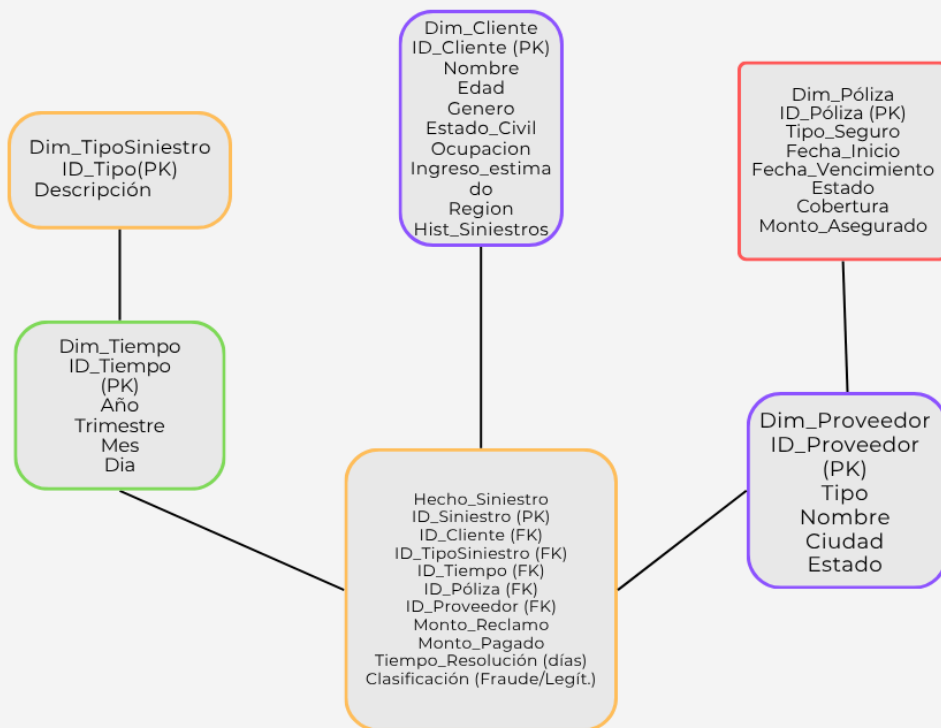
Este diseño en estrella facilita muchísimo el análisis multidimensional, permitiendo a los analistas explorar la información desde cualquier ángulo, ya sea por tipo de seguro, periodo o ubicación geográfica.

Este flujo de datos asegura que "ProtecSeguros" tenga trazabilidad completa desde las fuentes originales hasta los reportes finales. Así pueden ejecutar consultas OLAP, generar informes de desempeño detallados y aplicar modelos de minería de datos para encontrar patrones de riesgo o posibles fraudes.

El proceso ETL se programa de forma periódica (diaria o semanal, según lo necesiten) para garantizar que los directivos y analistas siempre trabajen con información actualizada.

DIAGRAMAS CONCEPTUALES Y FÍSICOS DEL ESQUEMA

DIAGRAMA CONCEPTUAL (VISIÓN GENERAL)



| <i>Tabla</i> | <i>Campo</i> | <i>Tipo de dato</i> | <i>Clave</i> | <i>Descripción</i> |
|------------------------|-------------------|---------------------|--------------|-----------------------------------|
| Hecho_Siniestro | ID_Siniestro | INT | PK | Identificador único del siniestro |
| | ID_Cliente | INT | FK | Relación con Dim_Cliente |
| | ID_Póliza | INT | FK | Relación con Dim_Póliza |
| | ID_TipoSiniestro | INT | FK | Relación con Dim_TipoSiniestro |
| | ID_Tiempo | INT | FK | Relación con Dim_Tiempo |
| | ID_Proveedor | INT | FK | Relación con Dim_Proveedor |
| | Monto_Reclamo | DECIMAL(10,2) | | Valor reclamado |
| | Monto_Pagado | DECIMAL(10,2) | | Valor indemnizado |
| | Tiempo_Resolucion | INT | | Días promedio |
| | Clasificación | VARCHAR(20) | | Tipo de caso (Fraude/Legítimo) |

FUNDAMENTO TÉCNICO SOBRE LA LÓGICA DE DISEÑO

El diseño del Data Warehouse (DW) para “ProtecSeguros” se basa en principios de integración de datos, optimización de consultas analíticas y soporte a la toma de decisiones estratégicas en seguros. La lógica de diseño se fundamenta en los siguientes aspectos:

- Orientación a análisis de negocio (OLAP)

Se definieron los hechos principales siniestros, primas cobradas y pagos a terceros que representan eventos cuantificables y medibles. Las dimensiones asociadas cliente, póliza, tipo de seguro, región, tiempo y perfil de riesgo permiten desglosar y agrupar los hechos según criterios de interés estratégico. Esto garantiza consultas rápidas y multidimensionales para detectar fraudes, analizar costos y ajustar primas.

- Modelo en estrella como base de diseño

Se opta por un modelo en estrella, donde la tabla central de hechos está directamente conectada con las dimensiones. Esta elección se justifica porque:

- Facilita la comprensión del modelo por los analistas de negocio.
- Optimiza el rendimiento de consultas agregadas, dado que las dimensiones están desnormalizadas y reducen los joins complejos.
- Permite flexibilidad para agregar nuevas métricas o dimensiones futuras sin afectar significativamente la estructura existente.

- Integración de datos de múltiples fuentes

Los datos operacionales provienen de sistemas de pólizas, pagos y siniestros. La lógica de diseño considera:

- Extracción: lectura periódica de los sistemas transaccionales.
- Transformación: limpieza, deduplicación, validación y categorización (por ejemplo, normalización de tipos de siniestros y perfiles de riesgo).
- Carga: inserción en las tablas del DW de manera incremental o por lotes, preservando consistencia histórica.

- Granularidad y agregación

La tabla de hechos registra información a nivel de siniestro individual, lo que permite medir con precisión frecuencia y costos. Se contemplan tablas de agregación para consultas rápidas por cliente, región o tipo de seguro, optimizando el desempeño analítico sin perder detalle operativo.

- Flexibilidad para análisis de riesgo y fraude

La inclusión de dimensiones como perfil de riesgo y antigüedad de póliza permite segmentar clientes y calcular indicadores de riesgo de manera eficiente. Esto facilita la detección de patrones anómalos (fraudes) y ajustes automáticos de primas según criterios de riesgo.

- Preservación histórica y trazabilidad

Se diseñan dimensiones con soporte tipo Slowly Changing Dimensions (SCD) Tipo 2, permitiendo mantener el historial de cambios en clientes o pólizas. Esto es crítico para auditorías, análisis longitudinal de siniestros y evolución de perfiles de riesgo.

CONCLUSIÓN

El proyecto Data Warehouse propuesto para ProtecSeguros representa una solución estratégica. Su función principal es integrar y consolidar información de diversas fuentes operativas, permitiendo así transformar grandes volúmenes de datos en conocimiento útil para la toma de decisiones.

Mediante un modelo de datos robusto y la implementación de procesos ETL efectivos, la organización tendrá la capacidad de analizar integralmente la relación entre siniestros, primas y pólizas. Esto facilitará la identificación de patrones de riesgo, tendencias de siniestros y posibles actividades fraudulentas.

Este enfoque analítico multidimensional aumenta significativamente la capacidad de la empresa para ajustar sus pólizas, mejorar la eficiencia en el procesamiento de reclamaciones y optimizar los recursos de gestión de riesgos. Adicionalmente, el DW establece una base sólida para la futura implementación de herramientas de inteligencia empresarial y minería de datos, respaldando una cultura organizacional basada en el análisis y la toma de decisiones informadas.

Sin duda, el Data Warehouse de ProtecSeguros es un recurso estratégico que asegura la competitividad, la transparencia y la eficiencia operativa de la empresa en el sector asegurador.

REFERENCIAS

Kimball vs. Inmon in Data Warehouse Architecture. (5 de junio de 2021). Zentut.

<https://www.zentut.com/data-warehouse/kimball-and-inmon-data-warehouse-architectures/>

Gasparini, S. (2025, enero 20). *Your conceptual guide to building a star schema data warehouse*. Medium. <https://medium.com/@sarahryliegaspardini/your-conceptual-guide-to-building-a-star-schema-data-warehouse-3ea25ccf0fce>

IBM. (2024). Esquemas de estrella - Modelado dimensional. Recuperado de <https://www.ibm.com/docs/es/ida/9.1.2?topic=entities-primary-foreign-keys>

Microsoft. (2024, diciembre 27). Descripción de un esquema de estrella e importancia para el modelado de datos. Microsoft. Recuperado de <https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>

Astera. (2024, marzo 25). ¿Qué es el esquema de estrella? Ventajas y desventajas. Astera. Recuperado de <https://www.astera.com/es/knowledge-center/star-schema/>