

Unidad # 3

Almacenes de Datos y Minería de Datos

Conceptualización y estructura lógica de almacenes de datos

Administración de Base de Datos II

Nombre del docente:



**UNIVERSIDAD
GERARDO BARRIOS**
Líderes en Gestión del Conocimiento



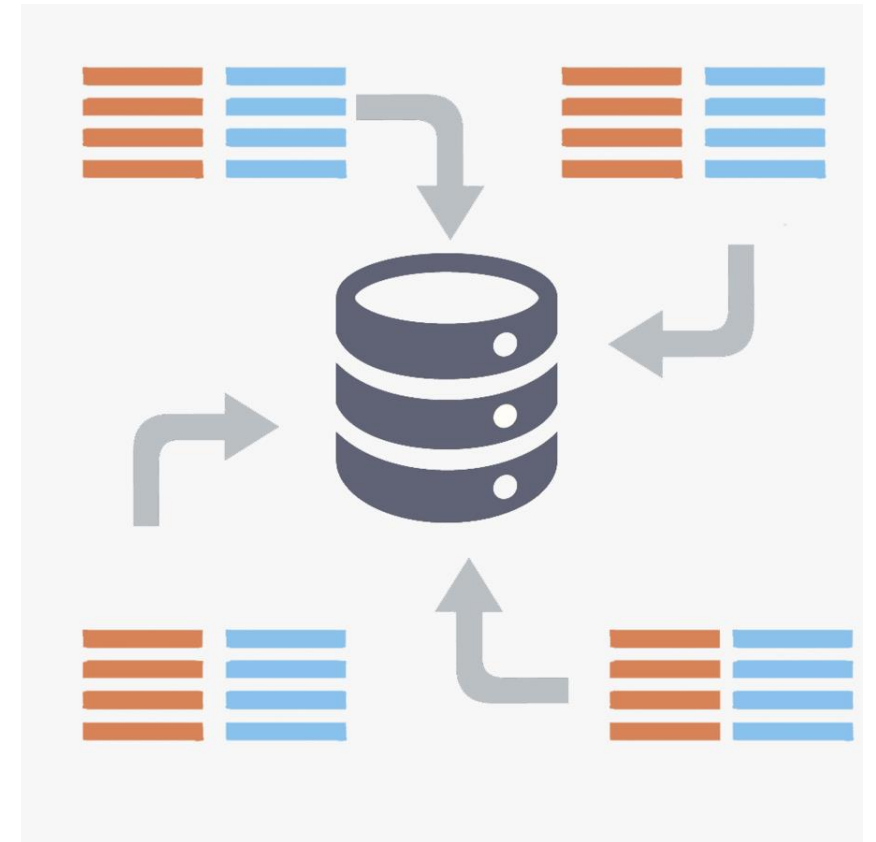
Contenidos a desarrollar (Parte I)

1. Conceptualización de almacén de datos
2. Elementos de un almacén de datos
3. Diferencias de un almacén de datos con una base de datos normal
4. Ventajas, usos y beneficios



Conceptualización de almacén de datos (1)

- Es un tipo de sistema de gestión de datos que está diseñado para habilitar y respaldar las actividades de inteligencia empresarial (BI), especialmente las analíticas. Los almacenes de datos están destinados únicamente a realizar consultas y análisis y, a menudo, contienen grandes cantidades de datos históricos. Los datos dentro de un almacén de datos generalmente se derivan de una amplia gama de fuentes, como archivos de registro de aplicaciones y aplicaciones de transacciones.



Conceptualización de almacén de datos (2)

- Un almacén de datos centraliza y consolida grandes cantidades de datos de múltiples fuentes. Sus capacidades analíticas permiten a las organizaciones obtener valiosos conocimientos comerciales a partir de sus datos para mejorar la toma de decisiones. Con el tiempo, crea un registro histórico que puede ser invaluable para los científicos de datos y los analistas comerciales. Debido a estas capacidades, un almacén de datos puede considerarse la "única fuente de información" de una organización.

Conceptualización de almacén de datos (3)

- Es un proceso para recopilar y administrar datos de diversas fuentes para proporcionar información empresarial significativa. Este se utiliza normalmente para conectar y analizar datos comerciales de fuentes heterogéneas. El almacén de datos es el núcleo del sistema de BI que está diseñado para el análisis y la generación de informes de datos.
- También, es una combinación de tecnologías y componentes que ayuda al uso estratégico de los datos. Siendo un almacenamiento electrónico de una gran cantidad de información por parte de una empresa, está diseñado para consultas y análisis en lugar de procesamiento de transacciones. Esto es un proceso de transformar datos en información y ponerlos a disposición de los usuarios de manera oportuna para marcar la diferencia.

Algo más acerca de un almacén de datos...

- La base de datos de apoyo a la toma de decisiones (almacén de datos) se mantiene separada de la base de datos operativa de la organización. Sin embargo, el almacén de datos no es un producto sino un entorno. Es una construcción arquitectónica de un sistema de información que proporciona a los usuarios información actual e histórica de apoyo a la toma de decisiones a la que es difícil acceder o está presente en el almacén de datos operativo tradicional.



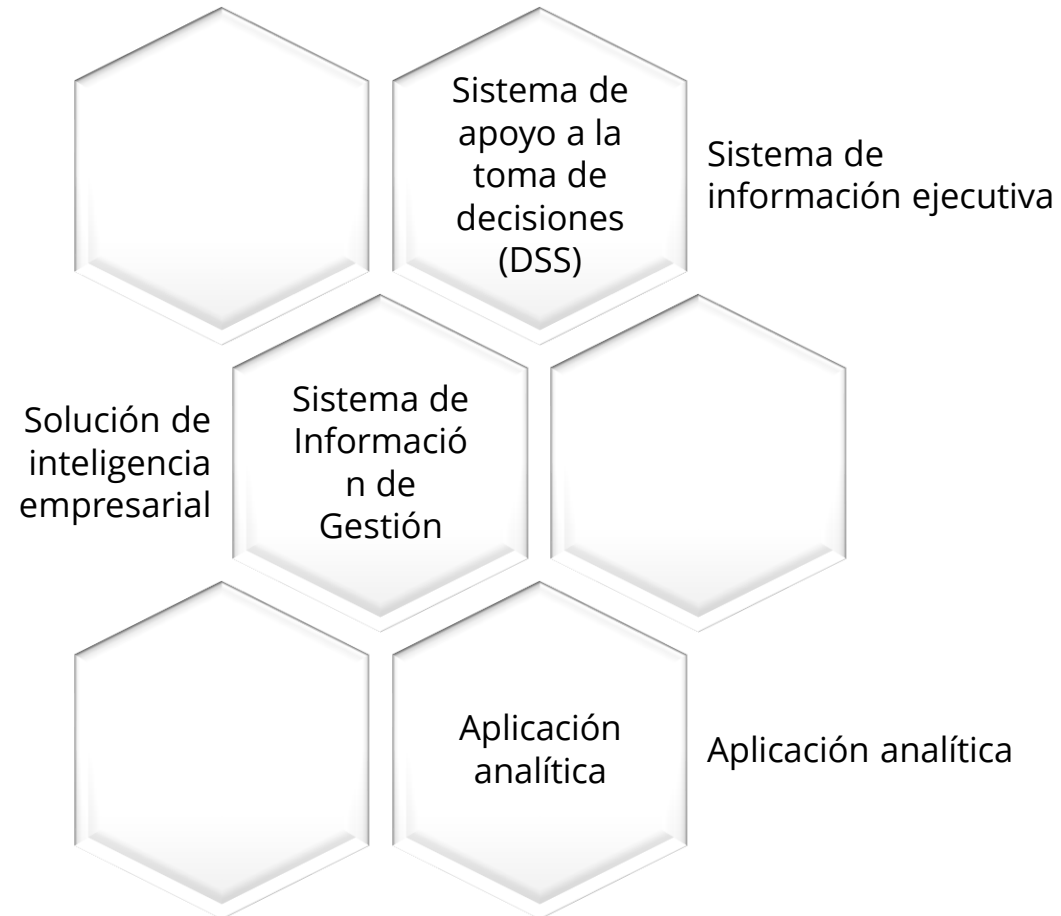
¿Cómo funciona Data Warehouse?

- Funciona como un depósito central donde la información llega de una o más fuentes de datos. Los datos fluyen hacia un almacén de datos desde el sistema transaccional y otras bases de datos relacionales.
- Los datos pueden ser:
 - Estructurado
 - Semiestructurada
 - Datos no estructurados

¿Cómo funciona Data Warehouse?

- Los datos se procesan, transforman e ingieren para que los usuarios puedan acceder a los datos procesados en el almacén de datos a través de herramientas de inteligencia empresarial, clientes SQL y hojas de cálculo. Un almacén de datos fusiona información procedente de diferentes fuentes en una base de datos completa.
- Al fusionar toda esta información en un solo lugar, una organización puede analizar a sus clientes de manera más integral. Esto ayuda a garantizar que se ha tenido en cuenta toda la información disponible. El almacenamiento de datos hace posible la minería de datos. La minería de datos busca patrones en los datos que puedan generar mayores ventas y ganancias.

Otros nombres de un almacén de datos



¿Qué elementos incluye un almacén de datos típico?

- 1** Una base de datos relacional para almacenar y administrar datos
- 2** Una solución de extracción, carga y transformación (ELT) para preparar los datos para el análisis
- 3** Capacidades de análisis estadístico, informes y minería de datos
- 4** Herramientas de análisis de clientes para visualizar y presentar datos a usuarios comerciales
- 5** Otras aplicaciones analíticas más sofisticadas que generan información procesable mediante la aplicación de algoritmos de ciencia de datos e inteligencia artificial (IA)

Diferencias de un almacén de datos con una base de datos normal



Los almacenes de datos utilizan un diseño diferente al de las bases de datos operativas estándar.



Las bases de datos estándar están optimizadas para mantener una precisión estricta de los datos en el momento, actualizando rápidamente los datos en tiempo real.



Los almacenes de datos, por el contrario, están diseñados para brindar una visión de largo alcance de los datos a lo largo del tiempo.



Intercambian el volumen de transacciones y en cambio se especializan en la agregación de datos.

Ventajas de los almacenes de datos

- A pesar que los almacenes de datos son costosos de escalar y no sobresalen en el manejo de datos sin procesar, no estructurados o complejos, estos siguen siendo una herramienta importante en la era del Big Data.
- Los beneficios de un almacén de datos incluyen análisis de datos mejorados, mayores ingresos y la capacidad de competir de manera más estratégica en el mercado.
- Al alimentar de manera eficiente datos contextuales estandarizados al software de inteligencia empresarial de una organización, un almacén de datos impulsa una estrategia de datos más efectiva.
- A continuación, se muestran los beneficios que un almacén de datos puede tener:

1. Habilita el conocimiento histórico

- Ninguna empresa puede sobrevivir sin un almacén grande y preciso de datos históricos, desde datos de ventas e inventario hasta registros de propiedad intelectual y de personal. Si un ejecutivo de negocios de repente necesita conocer las ventas de un producto clave hace 24 meses, los datos históricos proporcionados por un almacén de datos lo hacen posible.
- También es importante que un almacén de datos pueda agregar contexto a estos datos históricos al enumerar todas las tendencias clave de rendimiento que rodean esta investigación retrospectiva. Este tipo de eficiencia no se puede igualar con una base de datos heredada.

2. Mejora la calidad de los datos

- Una empresa genera datos en innumerables formas y diferentes, incluidos datos estructurados y no estructurados, datos de redes sociales y datos de campañas de ventas. Un almacén de datos convierte estos datos en los formatos consistentes que requieren sus plataformas de análisis. Además, al garantizar esta conformidad, un almacén de datos garantiza que los datos producidos por diferentes divisiones comerciales tengan la misma calidad y estándar, lo que permite una alimentación más eficiente para el análisis.



3. Aumenta la eficiencia

- Es mucho tiempo para un usuario empresarial o un científico de datos tener que recopilar datos de múltiples fuentes. Es mucho más ventajoso que estos datos se recopilen en un solo lugar, de ahí el beneficio de un almacén de datos.
- Además, si, por ejemplo, un científico de datos necesita datos para ejecutar un informe rápido, no necesita la asistencia del soporte técnico para realizar esta tarea. Un almacén de datos hace que estos datos estén fácilmente disponibles, en el formato correcto, mejorando la eficiencia de todo el proceso.

4. Aumenta la potencia y la velocidad del análisis de datos

- La inteligencia empresarial y el análisis de datos son lo opuesto al instinto y la intuición. BI y la analítica requieren datos estandarizados de alta calidad, a tiempo y disponibles para una rápida extracción de datos. Un almacén de datos habilita esta potencia y velocidad, lo que permite una ventaja competitiva en sectores comerciales clave, que van desde CRM hasta RR.HH., éxito de ventas e informes trimestrales.



5. Genera ingresos

- “Los datos son el nuevo petróleo”, o el alto valor en dólares de los datos en el mundo actual. Esto permite que la creación de datos más estandarizados y de mejor calidad, fortalezca un almacén de datos traducido en importantes ganancias. La fórmula del almacén de datos funciona así: una mejor inteligencia empresarial ayuda a tomar mejores decisiones y, a su vez, las mejores decisiones crean un mayor retorno de la inversión en cualquier sector de un negocio.
- Lo que es más importante, estas ganancias se basan en sí mismas con el tiempo, ya que mejores decisiones fortalecen el negocio.
- En resumen, un almacén de datos de alta calidad y totalmente escalable puede verse como un costo menor y más como una inversión, uno que agrega valor exponencial como pocas otras inversiones que hacen las empresas.

6. Es escalable

- La palabra clave principal en la era de la nube es "escalable" y un almacén de datos es un componente crítico para impulsar esta escala. Un almacén de datos de primer nivel es escalable en sí mismo y también permite una mayor escalabilidad en el negocio en general.
- Es decir, el almacén de datos sofisticado de hoy en día está construido para escalar, manejando cada vez más consultas a medida que crece el negocio (aunque esto requerirá más hardware de soporte). Además, la eficiencia en el flujo de datos que permite un almacén de datos impulsa en gran medida el crecimiento de una empresa; este crecimiento es el núcleo de la escalabilidad empresarial.

7. Interopera con Cloud

- A diferencia de las bases de datos heredadas de antaño, los almacenes de datos de hoy se crean teniendo en cuenta la nube híbrida y la nube múltiple. Muchos almacenes de datos ahora están completamente basados en la nube, e incluso aquellos que están contruidos en las instalaciones normalmente interactuarán bien con la parte basada en la nube de la infraestructura de una empresa. Como un punto lateral adicional importante: este enfoque basado en la nube también significa que los usuarios móviles pueden acceder mejor al almacén de datos; esto es beneficioso para los representantes de ventas en particular.



8. Seguridad de los datos

- Varios avances clave en el almacenamiento de datos han mejorado su seguridad, lo que mejora la seguridad general de los datos de la empresa. Entre estos avances se encuentran técnicas como una configuración de "solo lectura esclava", que bloquea el código SQL malicioso, y columnas cifradas, que protege los datos confidenciales.
- Algunas empresas configuran grupos de usuarios personalizados en sus almacenes de datos, que pueden incluir o excluir varios grupos de datos e incluso otorgar permisos fila por fila.



9. Rendimiento de consultas e información

- Las consultas constantes de inteligencia empresarial que forman parte del negocio actual pueden ejercer una gran presión sobre la infraestructura de análisis, desde las bases de datos heredadas hasta los mercados de datos. Tener un almacén de datos para manejar consultas de manera más efectiva elimina parte de la presión sobre el sistema.
- Además, dado que un almacén de datos está diseñado específicamente para manejar niveles masivos de datos y una miríada de consultas complejas, es el núcleo de alto funcionamiento de la práctica de análisis de datos de cualquier empresa.

10. Ventaja competitiva

- Este es absolutamente el beneficio final de un almacén de datos: permite a una empresa diseñar estrategias y ejecutar de manera más efectiva contra otros proveedores en su sector.
- Con la calidad, la velocidad y el contexto histórico que proporciona un almacén de datos, la mayor comprensión de la minería de datos puede impulsar decisiones que generen más ventas, productos más específicos y tiempos de respuesta más rápidos.
- En resumen, un almacén de datos mejora la toma de decisiones comerciales, lo que a su vez brinda a cualquier empresa una ventaja competitiva clave.

Desventajas de los almacenes de datos

- No es una opción ideal para datos no estructurados.
- La creación e implementación de Data Warehouse es sin duda un asunto confuso.
- El almacén de datos puede quedar obsoleto con relativa rapidez.
- Es difícil realizar cambios en los tipos y rangos de datos, el esquema de la fuente de datos, los índices y las consultas.
- El almacén de datos puede parecer fácil, pero en realidad es demasiado complejo para los usuarios promedio.
- A pesar de los mejores esfuerzos en la gestión de proyectos, el alcance del proyecto de almacenamiento de datos siempre aumentará.
- En algún momento, los usuarios del almacén desarrollarán diferentes reglas comerciales.
- Las organizaciones necesitan gastar gran parte de sus recursos para la capacitación y el propósito de implementación.

¿Cuál podría ser el uso de los almacenes de datos?

- Cada instalación de base de datos viene con una serie de paquetes o módulos adicionales de distintas formas y tamaños que en muy pocas ocasiones todos ellos son utilizadas por las compañías, lo que las convierten en una posible puerta de entrada para sufrir algún tipo de ataque si en esos paquetes se descubre cualquier problema de seguridad.
- Para reducir riesgos, es recomendable que los usuarios detecten esos paquetes que no se utilizan y se desactiven del servidor donde estén instalados.
- Esto no sólo reduce los riesgos de ataques, sino que también simplifica la gestión de parches ya que únicamente será de máxima urgencia actualizar aquellos que hagan referencia a un módulo que estemos utilizando.

Sectores que los utilizan:

- **Aerolínea:** Se utiliza para fines operativos como asignación de tripulaciones, análisis de rentabilidad de rutas, promociones de programas de viajero frecuente, etc.
- **Bancario:** Se usa ampliamente en este sector para administrar los recursos disponibles en el escritorio de manera efectiva. Pocos bancos lo utilizan para la investigación de mercado, el análisis de rendimiento del producto y las operaciones.
- **Cuidado de la salud:** Este sector ha utilizado el almacén de datos para diseñar estrategias y predecir resultados, generar informes de tratamiento de pacientes, compartir datos con compañías de seguros vinculadas, servicios de asistencia médica, etc.
- **Sector público:** En el sector público, el almacén de datos se utiliza para la recopilación de inteligencia. Ayuda a las agencias gubernamentales a mantener y analizar registros de impuestos y registros de políticas de salud para cada individuo.

Sectores que los utilizan:

- **Sector de inversiones y seguros:** Aquí se utilizan principalmente para analizar patrones de datos, tendencias de los clientes y realizar un seguimiento de los movimientos del mercado.
- **Retener cadena:** En las cadenas minoristas, el almacén de datos se utiliza ampliamente para distribución y marketing. También ayuda a realizar un seguimiento de los artículos, el patrón de compra del cliente, las promociones y también se utiliza para determinar la política de precios.
- **Telecomunicación:** En este sector se utiliza para promociones de productos, decisiones de ventas y para tomar decisiones de distribución.
- **Industria de la hospitalidad:** Esta industria lo utiliza para diseñar y estimar sus campañas de publicidad y promoción donde quieren dirigirse a los clientes, en función de sus comentarios y patrones de viaje.

Tipos de almacén de datos

1. Almacén de datos empresariales (EDW):

- Enterprise Data Warehouse (EDW) es un almacén centralizado. Proporciona un servicio de soporte de decisiones en toda la empresa. Ofrece un enfoque unificado para organizar y representar datos. También proporciona la capacidad de clasificar datos según el tema y dar acceso según esas divisiones.

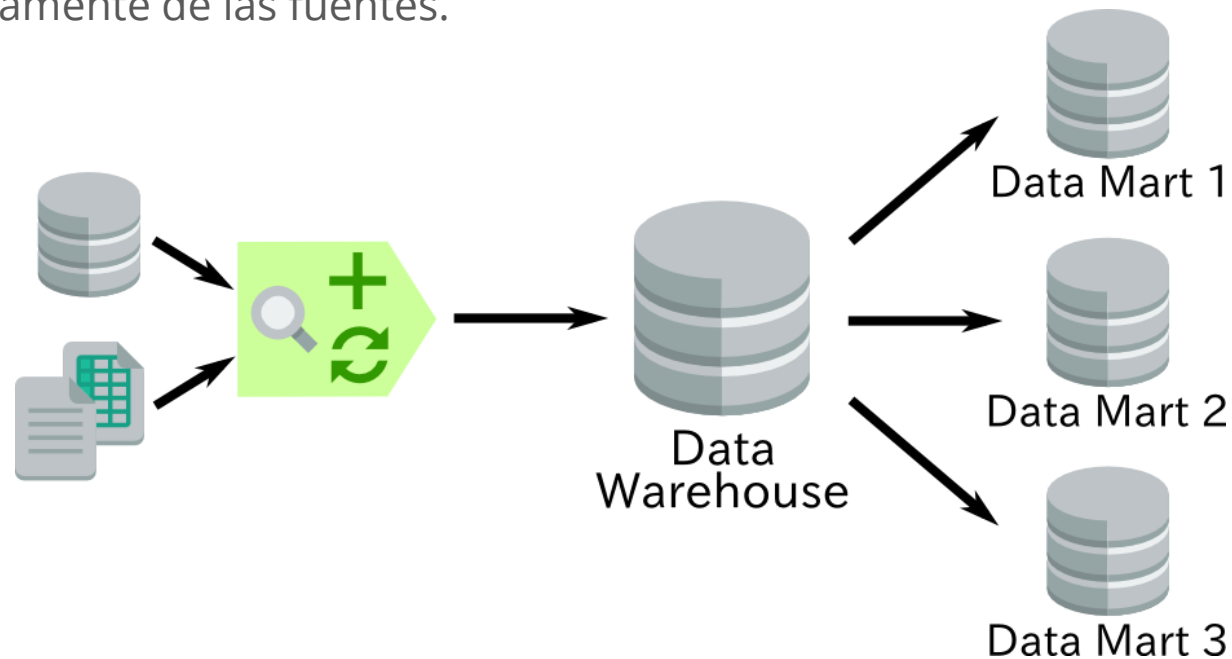
2. Almacén de datos operativos:

- Los almacenes de datos operativos, que también se denominan ODS, no son más que un almacén de datos necesario cuando ni el almacén de datos ni los sistemas OLTP respaldan las necesidades de informes de las organizaciones. En ODS, el almacén de datos se actualiza en tiempo real. Por lo tanto, se prefiere ampliamente para actividades de rutina como almacenar registros de los Empleados.

Tipos de almacén de datos

3. Data Mart:

Un Data Mart es un subconjunto del almacén de datos. Está especialmente diseñado para una línea de negocio en particular, como ventas o finanzas. En una despesa de datos independiente, los datos se pueden recopilar directamente de las fuentes.



Etapas generales del almacén de datos

1. Base de datos operativa sin conexión:

En esta etapa, los datos simplemente se copian de un sistema operativo a otro servidor. De esta manera, la carga, el procesamiento y la generación de informes de los datos copiados no afectan el rendimiento del sistema operativo.

2. Almacén de datos sin conexión:

Los datos del Data Warehouse se actualizan periódicamente desde la base de datos operativa. Los datos en Data Warehouse se mapean y transforman para cumplir con los objetivos de esta.

Etapas generales del almacén de datos

3. Almacén de datos en tiempo real:

En esta etapa, los almacenes de datos se actualizan cada vez que se realiza alguna transacción en la base de datos operativa. Por ejemplo, un sistema de reserva de líneas aéreas o ferrocarriles.

4. Almacén de datos integrado:

En esta etapa, los almacenes de datos se actualizan continuamente cuando el sistema operativo realiza una transacción. El Data Warehouse luego genera transacciones que se devuelven al sistema operativo.

Contenidos a desarrollar (Parte II)

1. Características del almacén de datos
2. Arquitectura de un almacén de datos
3. Estructura lógica de un almacén de datos



Características del almacén de datos

Los conceptos de almacenamiento de datos tienen las siguientes características:

- Orientado al sujeto
- Integrado
- Variante de tiempo
- No volátil

Características del almacén de datos (2)

1. Orientado al sujeto. Un almacén de datos está orientado al tema, ya que ofrece información sobre un tema en lugar de las operaciones en curso de las empresas. Estos temas pueden ser ventas, marketing, distribuciones, etc.

Un DW Nunca se centra en las operaciones en curso, pero pone énfasis en el modelado y análisis de datos para la toma de decisiones y proporciona una vista simple y concisa sobre el tema específico al excluir datos que no son útiles para respaldar el proceso de decisión.

2. Integrado. La integración aquí significa el establecimiento de una unidad de medida común para todos los datos similares de la base de datos diferente. Los datos también deben almacenarse en el Data Warehouse de manera común y universalmente aceptable.

Un almacén de datos se desarrolla integrando datos de diversas fuentes como un mainframe, bases de datos relacionales, archivos planos, etc. Manteniendo convenciones de nomenclatura, formato y codificación consistentes.

Características del almacén de datos (3)

Esta integración ayuda a un análisis eficaz de los datos. Debe garantizarse la coherencia en las convenciones de nomenclatura, las medidas de atributos, la estructura de codificación, etc. Considere el siguiente ejemplo:

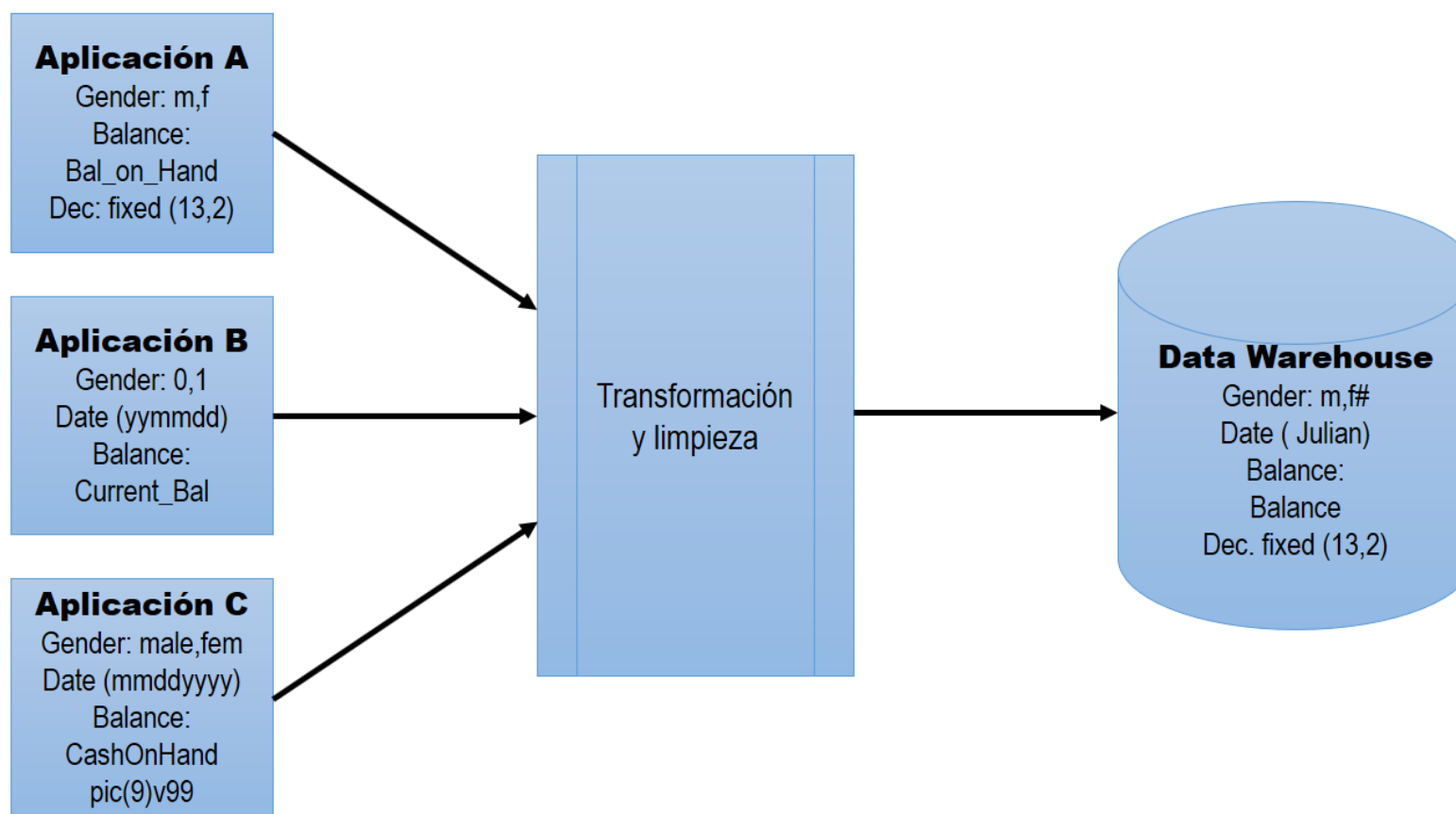


Figura 1. Problemas de integración de datos. Imagen basada en Fuente: https://www.guru99.com/images/1/022218_0735_DataWarehou1.png

Características del almacén de datos (4)

En el ejemplo anterior, hay tres aplicaciones diferentes etiquetadas como A, B y C. La información almacenada en estas aplicaciones es Sexo, Fecha y Saldo. Sin embargo, los datos de cada aplicación se almacenan de forma diferente.

- En la aplicación A, un campo de género almacena valores lógicos como M o F
- En la aplicación B, el campo de género es un valor numérico,
- En la aplicación C, el campo de género se almacena en forma de valor de carácter.
- Lo mismo ocurre con la fecha y el saldo.

Sin embargo, después del proceso de transformación y limpieza, todos estos datos se almacenan en un formato común en el Data Warehouse.

Características del almacén de datos (5)

3. Variante de tiempo. El horizonte temporal del almacén de datos es bastante amplio en comparación con los sistemas operativos. Los datos recopilados en un almacén de datos se reconocen con un período particular y ofrecen información desde el punto de vista histórico. Contiene un elemento de tiempo, explícita o implícitamente.

Uno de esos lugares donde los datos de Data Warehouse muestran la variación del tiempo es en la estructura de la clave de registro. Cada clave primaria contenida con el DW debe tener implícita o explícitamente un elemento de tiempo. Como el día, la semana, el mes, etc.

Otro aspecto de la variación de tiempo es que una vez que los datos se insertan en el almacén, no se pueden actualizar ni modificar.

Características del almacén de datos (6)

4. No volátil. El almacén de datos también es no volátil, lo que significa que los datos anteriores no se borran cuando se ingresan nuevos datos en él.

Los datos son de solo lectura y se actualizan periódicamente. Esto también ayuda a analizar datos históricos y comprender qué y cuándo sucedió. No requiere procesos de transacción, recuperación y mecanismos de control de concurrencia.

Las actividades como eliminar, actualizar e insertar que se realizan en un entorno de aplicación operativa se omiten en el entorno de almacenamiento de datos. Solo dos tipos de operaciones de datos realizadas en el almacenamiento de datos son

1. Carga de datos
2. Acceso a los datos

Arquitectura de un almacén de datos

La arquitectura del almacén de datos es compleja ya que es un sistema de información que contiene datos históricos y conmutativos de múltiples fuentes. Hay 3 enfoques para construir capas de almacenamiento de datos: nivel único, dos niveles y tres niveles. Esta arquitectura se explica a continuación.

Arquitectura de un solo nivel

El objetivo de una sola capa es minimizar la cantidad de datos almacenados. Este objetivo es eliminar la redundancia de datos. En la práctica, esta arquitectura no se utiliza con frecuencia.

Arquitectura de dos niveles

La arquitectura de dos capas, separa las fuentes disponibles físicamente y el almacén de datos. Esta arquitectura no es ampliable y tampoco es compatible con una gran cantidad de usuarios finales. También tiene problemas de conectividad debido a las limitaciones de la red.

Arquitectura de un almacén de datos (2)

Arquitectura de almacenamiento de datos de tres niveles

Esta es la arquitectura de almacenamiento de datos más utilizada. Consiste en los niveles superior, medio e inferior.

1. **Nivel inferior:** la base de datos de los servidores de Data Warehouse como nivel inferior, suele ser un sistema de base de datos relacional. Los datos se limpian, transforman y cargan en esta capa utilizando herramientas de back-end.
2. **Nivel intermedio:** el nivel intermedio en el almacén de datos es un servidor OLAP que se implementa utilizando el modelo ROLAP o MOLAP. Para un usuario, este nivel de aplicación presenta una vista resumida de la base de datos. Esta capa también actúa como mediadora entre el usuario final y la base de datos.
3. **Nivel superior:** este nivel es una capa de cliente de front-end. Son las herramientas y la API que conecta y obtiene datos del almacén de datos. Pueden ser herramientas de consulta, de informes, de consulta gestionada, de análisis y de minería de datos.

Arquitectura de un almacén de datos (3)

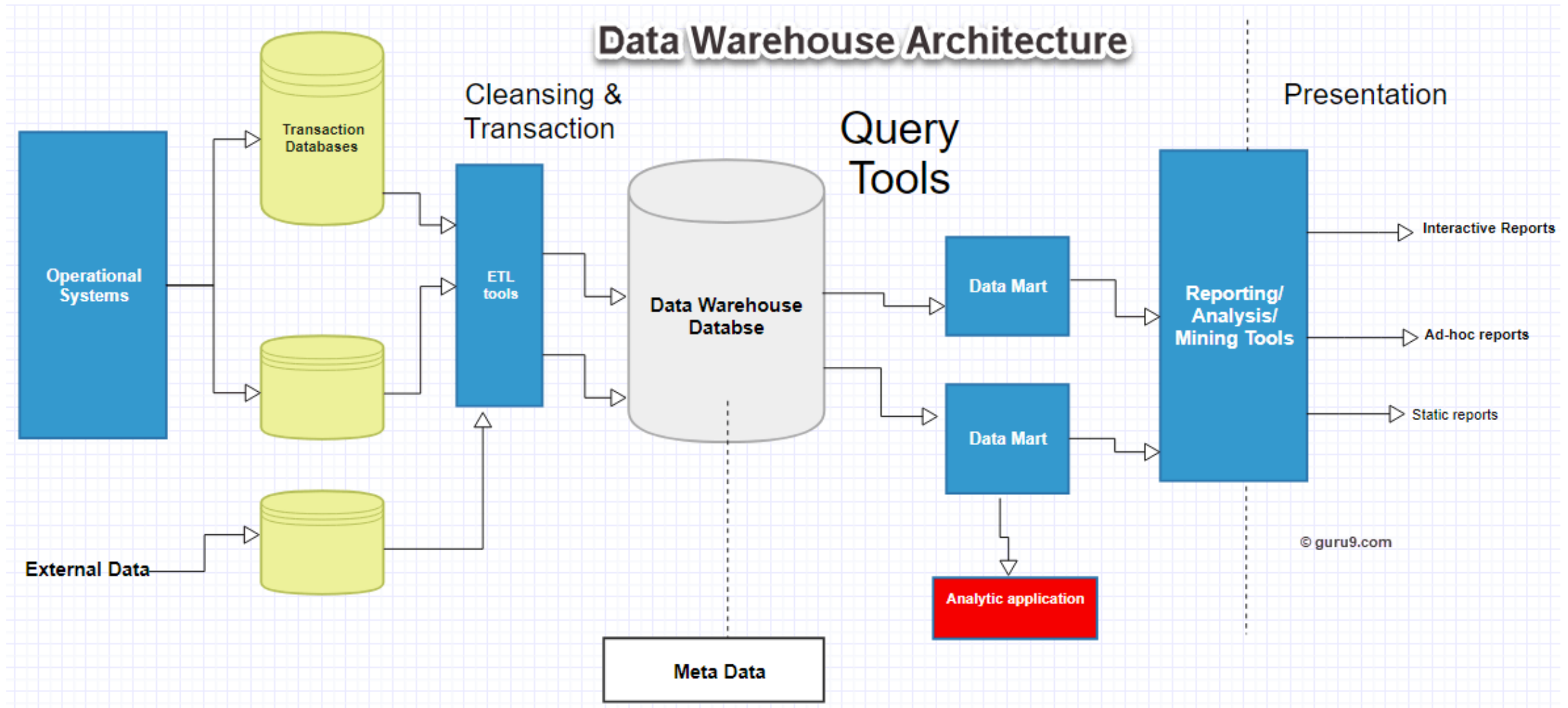


Figura 2. Arquitectura de un almacén de datos. Fuente: https://www.guru99.com/images/1/022218_0735_DataWarehou2.png

Componentes de un almacén de datos

El almacén de datos se basa en un servidor RDBMS que es un repositorio de información central que está rodeado por algunos componentes clave de almacenamiento de datos para hacer que todo el entorno sea funcional, manejable y accesible.

Esencialmente, hay 5 componentes:

1. Base de datos de almacenamiento de datos

La base de datos central es la base del entorno de almacén de datos. Esta se encuentra implementada en la tecnología RDBMS. Sin embargo, este tipo de implementación está limitado por el hecho de que el sistema RDBMS tradicional está optimizado para el procesamiento de bases de datos transaccionales y no para el almacenamiento de datos. Por ejemplo, las consultas ad-hoc, las combinaciones de varias tablas y los agregados consumen muchos recursos y ralentizan el rendimiento.

Por lo tanto, se utilizan enfoques alternativos a la base de datos que se enumeran a continuación:

- En un almacén de datos, las bases de datos relacionales se implementan en paralelo para permitir la escalabilidad. Las bases de datos relacionales paralelas también permiten la memoria compartida o el modelo de nada compartido, en varias configuraciones de multiprocesador o procesadores masivamente paralelos.
- Se utilizan nuevas estructuras de índice para omitir el escaneo de tablas relacionales y mejorar la velocidad.
- Uso de bases de datos multidimensionales (MDDDB) para superar las limitaciones que se imponen debido a los modelos de almacenamiento de datos relacionales. Ejemplo: Essbase de Oracle.

2. Herramientas de abastecimiento, adquisición, limpieza y transformación (ETL)

Las herramientas de origen, transformación y migración de datos se utilizan para realizar todas las conversiones, resúmenes y todos los cambios necesarios para transformar los datos en un formato unificado en el almacén de datos. También se denominan herramientas de extracción, transformación y carga (ETL).

Su funcionalidad incluye:

- Anonimizar los datos según las estipulaciones reglamentarias.
- Eliminar la carga de datos no deseados en bases de datos operativas en el almacén de datos.
- Buscar y reemplazar nombres y definiciones comunes para datos provenientes de diferentes fuentes.
- Calcular resúmenes y datos derivados.
- En caso de que falten datos, completarlos con los valores predeterminados.
- Datos repetidos o duplicados que llegan de múltiples fuentes de datos.

2. Herramientas de abastecimiento, adquisición, limpieza y transformación (ETL) (2)

Estas herramientas de extracción, transformación y carga pueden generar trabajos cron, trabajos en segundo plano, programas Cobol, scripts de shell, etc. que actualizan regularmente los datos en el almacén de datos. Estas herramientas también son útiles para mantener los metadatos.

Estas herramientas ETL tienen que hacer frente a los desafíos de la heterogeneidad de bases de datos y datos.

3. Metadatos

El nombre Meta Data sugiere algunos conceptos de almacenamiento de datos tecnológicos de alto nivel. Sin embargo, es bastante sencillo. Los metadatos son datos sobre los datos que definen el Data Warehouse. Se utiliza para construir, mantener y administrar el almacén de datos.

En la arquitectura de almacenamiento de datos, los metadatos desempeñan un papel importante, ya que especifican la fuente, el uso, los valores y las características de los datos del almacenamiento de datos. También define cómo se pueden modificar y procesar los datos. Está estrechamente relacionado con el almacén de datos.

Por ejemplo, una línea en la base de datos de ventas puede contener:

4030 KJ732 299,90

3. Metadatos (2)

Este es un dato sin sentido hasta que consultamos el Meta que nos dice que fue:

- Número de modelo: 4030
- ID de agente de ventas: KJ732
- Monto total de ventas de \$ 299.90

Por lo tanto, los metadatos son ingredientes esenciales en la transformación de datos en conocimiento. Los metadatos ayudan a responder las siguientes preguntas:

- ¿Qué tablas, atributos y claves contiene el almacén de datos?
- ¿De dónde provienen los datos?
- ¿Cuántas veces se recargan los datos?
- ¿Qué transformaciones se aplicaron con la limpieza?

3. Metadatos (3)

Los metadatos se pueden clasificar en las siguientes categorías:

- **Metadatos técnicos:** este tipo de metadatos contiene información sobre el almacén que utilizan los diseñadores y administradores del almacén de datos.
- **Metadatos comerciales:** este tipo contiene detalles que brindan a los usuarios finales una forma fácil de comprender la información almacenada en el almacén de datos.

4. Herramientas de consulta

Uno de los principales objetos del almacenamiento de datos es proporcionar información a las empresas para que tomen decisiones estratégicas. Las herramientas de consulta permiten a los usuarios interactuar con el sistema de almacenamiento de datos.

Estas herramientas se dividen en cuatro categorías diferentes:

- Herramientas de consulta e informes
- Herramientas de desarrollo de aplicaciones
- Herramientas de minería de datos
- Herramientas OLAP

4. Herramientas de consulta: - Consulta e informes

Las herramientas de consulta e informes se pueden dividir en:

- **Herramientas de informes:** se pueden sub dividir en:
 - **Redactores de informes:** este tipo de herramienta de informes son herramientas diseñadas para los usuarios finales para su análisis.
 - **Informes de producción:** permite a las organizaciones generar informes operativos regulares. También admite trabajos por lotes de gran volumen, como la impresión y el cálculo. Algunas herramientas de informes populares son Brio, Business Objects, Oracle, PowerSoft, SAS Institute.
- **Herramientas de consulta gestionadas:**

Este tipo de herramientas de acceso ayuda a los usuarios finales a resolver problemas en la base de datos, SQL y la estructura de la base de datos insertando una metacapa entre los usuarios y la base de datos.

4. Herramientas de consulta: - desarrollo de aplicaciones - minería de datos - OLAP

2. Herramientas de desarrollo de aplicaciones: A veces, las herramientas gráficas y analíticas integradas no satisfacen las necesidades analíticas de una organización. En tales casos, los informes personalizados se desarrollan utilizando herramientas de desarrollo de aplicaciones.

3. Herramientas de minería de datos: La minería de datos es un proceso para descubrir nuevas correlaciones, patrones y tendencias significativas mediante la extracción de grandes cantidades de datos. Se utilizan herramientas de minería de datos para que este proceso sea automático.

4. Herramientas OLAP: Estas herramientas se basan en conceptos de una base de datos multidimensional. Permite a los usuarios analizar los datos utilizando vistas multidimensionales complejas y elaboradas.

5. Arquitectura de bus de almacenamiento de datos

El bus del almacén de datos determina el flujo de datos en su almacén. Este se puede clasificar como flujo de entrada, flujo ascendente, flujo descendente, flujo de salida y metaflujo.

Al diseñar un bus de datos, es necesario tener en cuenta las dimensiones compartidas, los hechos en los mercados de datos.

Data marts.

Un Data Mart es una capa de acceso que se utiliza para enviar datos a los usuarios. Se presenta como una opción para almacenes de datos de gran tamaño, ya que su construcción requiere menos tiempo y dinero. Sin embargo, no existe una definición estándar de un mercado de datos que difiera de una persona a otra.

Data Mart es una subsidiaria de un almacén de datos. Los mercados de datos se pueden crear en la misma base de datos que el Data Warehouse o en una base de datos físicamente separada.

Tabla de actividades

Nombre de la actividad	Conociendo los almacenes de datos (Data Warehouse)
Tipo de actividad	Tríos
Competencias específica de la asignatura	Manipular bases de datos para asegurar la disponibilidad y seguridad de los datos, utilizando entornos web o locales, implementando Data Warehouse y minería de datos, trabajando de manera individual o colaborativa.
Instrucciones	Luego de haber dado lectura al contenido de la semana, elaboren un mapa mental sobre los almacenes de datos lo suficientemente amplio y bien fundamentado, en el cual se incluyan los contenidos más relevantes de este. La herramienta para realizarlo es libre.
Fecha de entrega	Domingo al finalizar la semana, a las 11:55 pm
Instrumento de evaluación	Buzón de Tareas
Criterios de evaluación	Evaluación formativa
Ponderación	Evaluación formativa

Recursos Complementarios

Recurso	Título	Cita Referencial
Sitio Web	Almacenes de Datos (Data Warehouse)	https://sites.google.com/site/mercadotecniaige2017/5-3-1-almacenes-de-datos-data-warehouse
Video	¿Qué es un Data Warehouse? Business Intelligence?	https://www.youtube.com/watch?v=jFsRdTcljeU

¿Preguntas?

¡Muchas gracias!