

LA INTEGRACIÓN DE DATAMART CON DATAWAREHOUSE

AUTORES: Carlos Ariel Avila Cruz ¹
Jair Josue Chiquito Muñiz ²



DIRECCIÓN PARA CORRESPONDENCIA: (avila-carlos@unesum.edu.ec)

Fecha de recepción: 12/03/2021

Fecha de aceptación: 25/12/2021

RESUMEN

Este trabajo de investigación se basó sobre la integración de datamart con datawarehouse como su posible implementación ya que en nuestra documentación se encontró evidencia que la integración del base de datos en este caso las datamart deben ser realizadas de una forma correcta y aquí así se tiene una mayor integración de datos para los usuarios que la utilice y se ha difícil la pérdida información ya que un datawarehouse es un repositorios de información, el actualidad existen muchos programas que nos permite realizar la sincronización información de un datamart con el datawarehouse de una formas rápida y sencilla. Aunque no todas la base datos nos brindar seguridad al momento de realizar la integración de datos por este motivo hay que poner en cuenta varios factores Para realizar este artículo científico se realizó algunos métodos metodológicos como son los siguientes: Histórico-lógico, Análisis-Síntesis entre otros ya ayudaron mucho en la búsqueda de información y la vez tener un mejor conocimiento del tema cómo es la integración de datamart con datawarehouse.

PALABRAS CLAVE: Datawarehouse; Bases de datos; Integración.

THE INTEGRATION OF DATAMART WITH DATAWAREHOUSE**ABSTRACT**

This research work was based on the integration of datamart with datawarehouse as its possible implementation since in our documentation evidence was found that the integration of the database in this case the datamart must be carried out in a correct way and here we have a greater integration of data for users who use it and it has been difficult to lose information since a datawarehouse is a repository of information, nowadays there are many programs that allow us to synchronize information from a datamart with the datawarehouse in a fast and simple. Although not all the databases provide security when performing data integration for this reason, several factors must be taken into account. To carry out this scientific article, some methodological methods were carried out, such as the following: Historical-logical, Analysis-Synthesis between

¹ Estudiante de la carrera de Tecnología de la información. Universidad Estatal Del Sur De Manabí. Jipijapa, Manabí Ecuador E-mail: jairchiquito10@gmail.com

² Estudiante de la carrera de Tecnología de la información. Universidad Estatal Del Sur De Manabí. Jipijapa, Manabí Ecuador. E-mail: rjcastrocastro@hotmail.com

others have already helped a lot in the search for information and at the same time have a better understanding of the subject of how datamart integrates with datawarehouse

KEYWORDS: Datawarehouse; Databases; Integration.

INTRODUCCIÓN

El análisis de los estudios de la integración de datamart con datawarehouse ya que es muy importante por lo única razón de que esté mundo globalizado el activo más preciado es la información ya que como ella se puedes ver tanto el progreso de la empresa como sus ingresos y aporte en la sociedad por esta razón las bases de datos debe estar integradas o asociadas de una manera correcta para que la información no sé filtre por ello estas herramientas por lo general muestra la información en forma de reportes específicos que se puede crear en los datos obtenido ERP llamado por su nombre “empresa utiliza su gestión” de tal forma es presentada al usuario de manera ágil. Por ejemplo, cabe mencionar que estas herramientas son de gran utilidad para distintas áreas de la organización como mercadotecnia, ventas, finanzas, logística, producción.

DESARROLLO

Metodología

A partir de la investigación realizada, se determinó el uso de los siguientes métodos:

Los métodos utilizados en nuestra investigación científica fueron los métodos de inducción-deducción y análisis de síntesis, que ayudaron a tener más conocimiento y conocimiento sobre el desarrollo de la seguridad de las redes telemáticas en el desempeño de las mujeres. El método de análisis-síntesis se realizó para mujeres de la Universidad Estatal del Sur de Manabí, diagnosticando Carreras en Tecnología de la Información, sobre el desempeño en la seguridad de las redes telemáticas, donde permite saber cuán importante es la seguridad en las redes, una vez que es un conjunto de técnicas que intenta minimizar la vulnerabilidad de los sistemas y es esencial en todas las instituciones, organizaciones, empresas, entre otras, en las que una mujer puede tener la capacidad de administrar o ejecutar con nuevas herramientas para la seguridad de las redes telemáticas. El método de deducción-inducción fue importante para comprender la realidad del problema que existe no solo en las universidades, sino en otras partes del mundo, que las mujeres dicen que no saben cuán importante es la seguridad de las redes telemáticas y lo mejor que pueden hacer. Para ejecutarse en cualquier institución que gestione la seguridad de la red.

Datamart

Un datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información (Konikov et al., 2018), (Visscher et al., 2017).

Por tanto, para crear el datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio datawarehouse, o sobre una base de datos OLAP. La

designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. De esta forma se pueden plantear dos tipos de datamart (Homayouni et al., 2019):

Datamart OLAP: Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.

Datamart OLTP: Pueden basarse en un simple extracto del datawarehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas report, que vienen a ser factibles reducidas (que agregan las dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de queries (aunque sólo es posibles en algunos SGBD avanzados, como Oracle).

Los datamarts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información
- Facilidad para la historización de los datos (Datamart, 2009)

Datawarehouse

Una data warehouse o datawarehouse empresarial guarda grandes cantidades de datos que han sido recopilados e integrados de múltiples fuentes. Como las organizaciones dependen de estos datos para realizar análisis o reportes, los datos necesitan tener un formato coherente y ser de fácil acceso – dos cualidades que definen a los datawarehousing y los hacen esenciales para las empresas de hoy.

Los datos son esenciales para las organizaciones que toman decisiones informadas, de modo que es razonable que los datawarehouses sean tan importantes porque almacenan todos esos datos. Los datawarehouses pueden:

- Almacenar grandes cantidades de datos en una base de datos central – y en un formato estándar.
- Integrar datos de diferentes fuentes y estandarizarlos, de modo que se faciliten el análisis y los reportes.
- Conservar registros históricos, ya que pueden almacenar meses o incluso años de datos.
- Mantener sus datos seguros almacenándolos en un sitio único. Se puede otorgar acceso sólo a aquellas personas que necesitan datos específicos.

- Proporcionar acceso fácil y rápido a datos para hacer posibles decisiones de negocios más ágiles (Homayouni et al., 2019)

Integración de datos

En una primera aproximación, podemos definir la integración de datos como un proceso de transformación y conciliación de datos que permita una mayor agilidad en la gestión, proporcionando datos conectados, seguros y de calidad.

Integrar significa combinar datos que se encuentran en diferentes fuentes para permitirle al usuario final tener una vista unificada de los mismos para una accesibilidad idónea, que sirva a las necesidades de negocio.

Actualmente, la aparición de nuevas tecnologías y la explosión de datos plantean un gran desafío en este aspecto, y ello se traduce en una mayor complejidad técnica a la hora de implementar un plan de integración de datos, si bien el objetivo sigue siendo el mismo: evitar su fragmentación mediante el desarrollo de soluciones ad hoc.

Son soluciones para variedad de tecnologías, incluyendo Datawarehouse, aplicaciones MDM, aplicaciones de BI, arquitecturas SOA, etc. y para todas ellas la integración de datos representa un elemento esencial.

Al margen de la casuística las políticas de integración de datos están orientadas a lograr una óptima accesibilidad para aprovechar al máximo el potencial de la información empresarial gracias a su multifuncionalidad, que puede servir a distintas utilidades:

- Sincronizar bases de datos.
- Consolidación de sistemas.
- Proveer datos para proyectos de BI.

Migraciones que requieren transformar los modelos de datos.

Estas diversas necesidades de integración de datos, por otra parte, han relegado a los ETL tradicionales a entornos analíticos y aparecen necesidades de integración de datos para otro tipo de proyectos, como los mencionados, desde las externalización, integración de aplicaciones o migraciones hasta la BBDD, sincronización, y un largo etcétera.

Las tecnologías o soluciones de integración de datos, por lo tanto, han de permitir su movimiento y actualización de manera rápida y fiable pues, al fin y al cabo, solo con cuando una integración es ágil se consigue un negocio que también lo sea (Gunadi & Widianto, 2020).

Tipos de Datamart

A diferencia de un almacén de datos que almacena datos de toda la empresa, un mercado de datos incluye información relacionada con un departamento o área temática en particular. Por ejemplo, un centro de ventas puede contener datos relacionados con productos, clientes y ventas solamente.

Los datamarts a menudo son contruidos y administrados por un solo departamento comercial. Como están orientados a temas, los datamarts generalmente toman datos de solo un pequeño número de fuentes, que podrían ser sistemas operativos internos, un repositorio de datos

centralizado o fuentes externas. Los datamarts suelen ser de menor tamaño y menos complejos que los data warehouse, lo que los hace más fáciles de construir y mantener.

Tipos de Datamarts

La datamarts se pueden clasificar en tres tipos principales:

Dependiente: Un centro de datos dependiente le permite combinar todos sus datos comerciales en un solo almacenamiento de datos, ofreciéndole los beneficios típicos de la centralización. En caso de que se necesiten uno o varios marts de datos físicos, deberá construirlos como marts de datos dependientes para garantizar la coherencia y la integración en todos los sistemas de almacenamiento de datos.

Los marts de datos dependientes se pueden construir utilizando dos enfoques diferentes. En el primer enfoque, el almacén de datos empresariales, así como los datamarts, se crean para que el operador pueda acceder a ambos, siempre que sea necesario. En el segundo enfoque, también conocido como enfoque federado, los resultados del proceso ETL se almacenan en un área de almacenamiento temporal, como un bus de datos común, en lugar de una base de datos física, por lo que el operador solo puede acceder al datamart. La última metodología no es ideal ya que ocasionalmente produce un depósito de datos en el que todos los datos se originan en una fuente compartida, pero se descartan en su mayoría.

Independiente: Se puede crear una despensa de datos independiente sin usar el almacén de datos central. Se recomienda principalmente para unidades o grupos más pequeños dentro de una organización. Como su nombre indica, este tipo de datamart no está relacionado con la data warehouse empresarial ni con ningún otro datamart. Introduce datos por separado y los análisis también se ejecutan de forma independiente.

A medida que se construyen más y más marts de datos independientes, la redundancia de datos también aumenta en toda la organización. Esto se debe a que cada centro de datos independiente necesita el suyo, generalmente una copia duplicada de la información comercial integral. Como estos datamarts acceden directamente a los archivos y / o tablas del sistema operativo, se reducen considerablemente limitar la escalabilidad de los sistemas de soporte de decisiones (DSS).

Híbrido: Al usar un centro de datos híbrido, puede combinar datos de varios sistemas fuente operativa además de un almacén de datos. Estos conjuntos de datos son particularmente útiles cuando necesita una integración ad hoc, por ejemplo, después de agregar un nuevo grupo o productos a la empresa.

Como su nombre indica, un datamart híbrido es una mezcla de datamarts dependientes e independientes. Es adecuado para empresas que tienen múltiples bases de datos y necesitan un cambio rápido. Un centro de datos híbrido necesita una ligera limpieza de datos, admite enormes estructuras de almacenamiento y es flexible, ya que combina los beneficios de los mercados de datos dependientes e independientes.

Estructuras de bases de datos y nombres de elementos en expresiones corporativas para que los operadores no técnicos puedan usar fácilmente la datamart. Si es necesario, también puede configurar API e interfaces para simplificar el acceso a los datos (Girsang et al., 2017).

Integración de Datamart - Datawarehouse

Para ello actualmente este repositorio cuenta con el soporte tecnológico para controlar las áreas críticas de este sector empresarial cubriendo los siguientes puntos:

- Procesos Extracción de Datos
- Carga y Transformación de Datos

Procesos Extracción de Datos: Este proceso se encarga de extraer información de todas las áreas de la empresa, actualmente da soporte las siguientes fuentes de datos:

1. Excel, Acceses, archivos de texto
2. Sistemas SAP
3. Sistemas Internos (RRHH, Logistico, Ventas, Comercial, entre Otros)

Carga y Transformación de Datos: Este proceso se encarga de transformar los formatos de los datos, un único formato estándar para ser enviado a los repositorios de datos ya estructurados (Zambrano Matamala et al., 2017).

Resultados y discusión

Como resultado se obtuvo que una datawarehouse utiliza un diseño diferente al de las bases de datos operacionales estándar. Estas últimas se optimizan manteniendo una estricta exactitud de los datos mediante la actualización de datos en tiempo real. Un datawarehouse, por el contrario, está diseñado para dar una visión de largo alcance de los datos a través del tiempo integrado con datamart se encarga de recolectar información de toda las áreas de la empresa de una manera rápida y organizada (castro, castro robert).

Se dio como resultado que la existencia de un datawarehouse no descarta la existencia de datamarts, ni viceversa. Es decir, puede haber organizaciones que tengan sólo un datawarehouse, que sólo tengan datamarts, o que tengan un datawarehouse y datamarts (Seneviratne et al., 2018).

Mejora el rendimiento de los reportes por medio del uso de una plataforma de BI, el cual permite diseñar los informes y reportes de una manera sencilla que muestre la información relevante a los usuarios finales. Mediante la integración combina datos que se encuentran en diferentes fuentes para permitirle al usuario final tener una vista unificada de los mismos para una accesibilidad idónea.

Datamart es una aplicación del datawarehouse local o departamental basados en conjuntos de información contenida en el almacén de datos maestros (datawarehouse). El datamart está básicamente indicado para líneas de negocio simples y responde a la estrategia de divide y vencerás, segmentando datos. El datawarehouse ayuda a minimizar el tiempo de análisis de la información (Seneviratne et al., 2018).

CONCLUSIONES

La datamarts es una base de datos de almacenamiento específico implementado con datawarehouse mejoran la flexibilidad para añadir rápidamente nuevas fuentes de datos.

Los datamarts se extienden fácilmente a la toma de decisiones estratégicas y la datawarehouse (DW) es una base de datos usada para generación de informes. Los datos son cargados desde los sistemas operacionales para su consulta.

RECOMENDACIONES

Se recomienda que para integrar datamart con datawarehouse se debe escoger el modelo que esté acorde a la implementación.

Para crear el Datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información.

REFERENCIAS BIBLIOGRÁFICAS

- Girsang, A. S., Satya, D., Isa, S. M., Al Fariz, S., Susilo, B., Ramdani, D., & Lian, M. (2017). Decision support system using data warehouse for hotel reservation system. 2017 International Conference on Sustainable Information Engineering and Technology (SIET),
- Gunadi, F., & Widiyanto, S. R. (2020). Perbandingan Data Warehouse Cloud Computing Menggunakan Konvensional Berbasis Kriptografi. Seminar Nasional Teknologi Komputer & Sains (SAINTEKS),
- Homayouni, H., Ghosh, S., & Ray, I. (2019). Data warehouse testing. In *Advances in Computers* (Vol. 112, pp. 223-273). Elsevier. <https://rayscyberlab.org/wp-content/uploads/2021/10/Chapter-Five-Data-Warehouse-Testing.pdf>
- Konikov, A., Kulikova, E., & Stifeeva, O. (2018). Research of the possibilities of application of the Data Warehouse in the construction area. MATEC Web of Conferences,
- Seneviratne, M. G., Seto, T., Blayney, D. W., Brooks, J. D., & Hernandez-Boussard, T. (2018). Architecture and implementation of a clinical research data warehouse for prostate cancer. *eGEMs*, 6(1). <https://www.ncbi.nlm.nih.gov/pmc/articles/pmc6078122/>
- Visscher, S. L., Naessens, J. M., Yawn, B. P., Reinalda, M. S., Anderson, S. S., & Borah, B. J. (2017). Developing a standardized healthcare cost data warehouse. *BMC health services research*, 17(1), 1-11.
- Zambrano Matamala, C., Urrutia Sepúlveda, A., & Varas Contreras, M. (2017). Análisis de rendimiento académico estudiantil usando Data Warehouse Difuso. *Ingeniare. Revista chilena de ingeniería*, 25(2), 242-254. https://scielo.conicyt.cl/scielo.php?pid=S0718-33052017000200242&script=sci_arttext

