

Universidad Autónoma del Estado de México

Centro Universitario UAEM Valle de Chalco

**DATA WAREHOUSE COMO PARADIGMA DE
EFICIENCIA EN UNA EMPRESA**

**ARTÍCULO ESPECIALIZADO PARA
PUBLICAR EN REVISTA INDIZADA**

QUE PARA OBTENER EL TÍTULO DE

INGENIERO EN COMPUTACIÓN

P R E S E N T A

CARLOS GARCÍA CORTEZ

ASESORA:

DRA. CRISTINA JUÁREZ LANDÍN

VALLE DE CHALCO SOLIDARIDAD, MÉXICO

OCTUBRE 2021.



CUVCH

**DATA WAREHOUSE COMO PARADIGMA DE
EFICIENCIA EN UNA EMPRESA**

ÍNDICE

| | |
|---------------------------------------|----|
| I. Resumen | 8 |
| II. Introducción | 9 |
| III. Metodología | 14 |
| IV. Resultados | 15 |
| V. Discusión | 17 |
| VI. Referencias bibliográficas | 18 |
| VII. Anexos | 21 |

I. RESUMEN

En este trabajo se presenta el paradigma del data warehouse (Almacén de datos) y su uso con el fin de mostrar crecimiento, evolución y optimización de los diferentes datos que con llevan a los sistemas complejos, tomando en cuenta que en las empresas trasnacionales recopilan grandes cantidades de datos de diversos sistemas, por tanto el data warehouse tradicional no puede seguir con el proceso usual ya que ahora los datos tienen una demanda de volúmenes que aumentan rápidamente.

Este articulo está enfocado a que las empresas y las personas conozcan una herramienta para agregar orígenes en los datos desde una ubicación central y así poder dar soporte a los negocios y poder crear informes ya que por medio del data warehouse se brindará la capacidad para procesar grandes cantidades de datos, hasta lograr almacenar petabytes de información.

La implementación de data warehouse en la mayoría de los casos es el primer paso para una solución completa y fiable de inteligencia empresarial (Business Intelligence) para obtener mejores resultados en los problemas de la organización, cabe recalcar que facilita el proceso para la gestión y es una gran solución para empresas que consideran eliminar datos que interfieren con el análisis de la información y su entrega, ya que analiza todos los tipos de datos que son útiles a sus usuarios, los estructura para su fácil acceso y operación.

Palabras Clave: DATA WAREHOUSE, PETABYTES, BUSINESS INTELLIGENCE

II. INTRODUCCIÓN

Los almacenes de datos aparecieron por primera vez en 1980 pero el data warehouse surge en 1985, William H. Inmon (Bill Inmon) construyó el primer almacén para un banco en Colorado, el propósito era ayudar a los sistemas operativos con el flujo de datos y soporte a la toma de decisiones (Decision Support System - DSS), ayudar a integrar y organizar los datos que estaban dispersos en múltiples bases de datos, los primeros almacenes requerían una enorme cantidad de redundancia ya que la mayoría de las empresas tenían múltiples entornos DSS que servían para sus usuarios. Los DSS usaban gran parte de datos, recopilación, limpieza e integración de datos y a menudo se replicaban para cada sistema o entorno (Bismart, 2020).

En el periodo de años de 1990 a 2000 el concepto de data warehouse se ha mejorado, lo cual conlleva más capacidad de almacenamiento, nuevas herramientas de consultas de datos y conexión a internet o una intranet.

El data warehouse es un tipo de gestión de datos o repositorio unificado, normalmente se aloja en el servidor de una empresa o en la nube el cual está hecho o diseñado para poder resolver y solventar actividades de inteligencia empresarial para poder integrar y depurar información de una o más fuentes distintas, así poder procesarlas para permitir el análisis desde diferentes perspectivas y con grandes velocidades de respuestas, se nutre de diferentes fuentes entre ellas los sistemas operacionales internos a través de los ETL (Extract, Transform and Load)

La ventaja principal radica en las estructuras de como almacena la información ya que lo puede almacenar por modelos de tablas en estrella, copo de nieve, cubos relacionales, entre otros. Mostramos algunas de las otras ventajas que tiene el data warehouse (Sinergia, 2020).

Facilitan el acceso a una gran variedad de datos (Fintechschooll, 2020).

Proporciona información clave para la toma de decisiones en la empresa como informes de tendencia o de excepción y estos datos son útiles a mediano y largo plazo.

Permite trabajar en conjunto y aumentar el valor operacional de las aplicaciones ya que si las fuentes de datos están claros son sencillos de instalar, contribuye con mayor flexibilidad y rapidez en el acceso a la información

Es útil para el almacenamiento de análisis y consultas históricas

Desventajas del data warehouse es el alto precio que puede llegar a costar a lo largo de su instancia o vida útil debido a que necesita un mantenimiento constante. La historia de los data warehouse comienza con la evolución de los sistemas de información y sistemas para el soporte de toma de decisiones.

Se considera que la tecnología basada en data warehouse está siendo relevante en últimos tiempos, sin embargo, no se cuenta con la información suficiente para aprovechar todas las potencialidades de dicha tecnología, por tanto en este trabajo se pretende la publicación del impacto de esta tecnología en la empresa, como se presenta en el anexo.

EVOLUCIÓN DE LOS SISTEMAS DE TOMA DE DECISIÓN

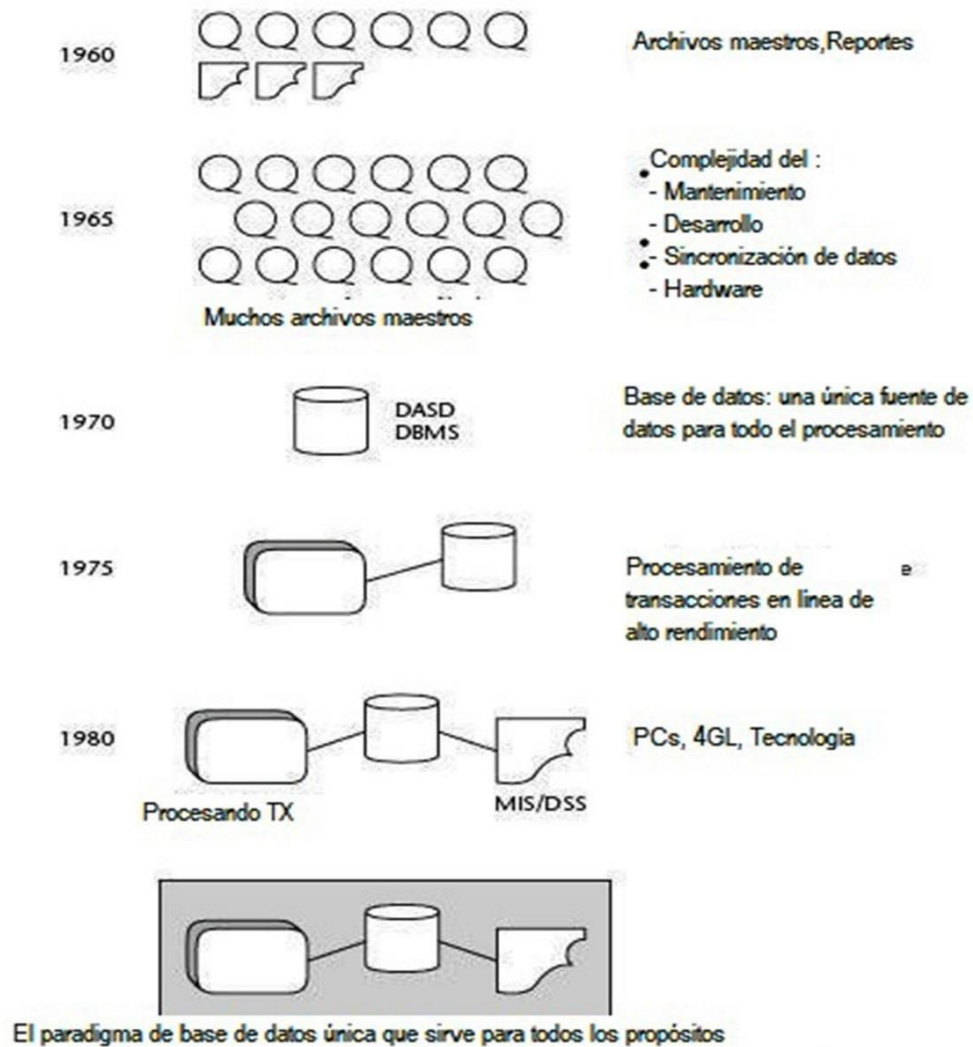


Figura 1. Las primeras etapas evolutivas del entorno (Carlosproal, 2020) .

Se puede visualizar la imagen la cual muestra una evolución de los sistemas de apoyo a la toma de decisiones, desde los reportes, base de datos hasta 4g technology que es dba virtual.

La arquitectura de un data warehouse se base en cinco niveles:

Capa de fuentes de datos – Es todas las fuentes que provienen los datos para ingresar en el Data warehouse.

Capa ETL(extract-transform-load)– Es todo el proceso que con lleva el data warehouse como la extracción de datos de las fuentes ya obtenidas, transformar los datos para cargarlos al data warehouse.

Capa de almacenamiento- Después de obtenerlos, transformarlos y extraerlos es donde se almacenan.

Capa de data marts – Son los almacenes más pequeños ya que permiten mayor accesibilidad a la información del data warehouse.

Capa de explotación de los datos – Es la capa del cliente front-end que presenta los resultados de los análisis mediante herramientas de visualización de datos, permiten realizar consultas de los datos como al data marts y al data warehouse.

El ciclo de vida de un data warehouse tiene varias etapas las cual empieza con su planificación esto se basa en ver si tienen la solides o necesidad de un data warehouse dentro de la empresa verán los alcances y mejoras que podrá brindar dicho proceso y ver recursos económicos, técnicos y humano (Calzada, 2020).

El segundo paso para el ciclo es definir los requisitos que se llevan en el negocio para con el uso del data warehouse, esto se puede ir validando con la ayuda de algunos usuarios y ver la satisfacción o carencias que se vayan teniendo en el trascurso del proceso.Después tenemos la parte de la tecnología, la de datos y la de aplicación BI.

La tecnología tiene que ver con las herramientas que usaran software, hardware, etc, la de datos es la creación de diagrama de donde se almacenara el data warehouse, empieza con el modelado donde se plasman los requerimientos, teniendo el diagrama se crearan los datos en la base de datos, tablas, vistas, etc. Se creara después el diseño y estructura de los procesos ETL, son los que se encargaran de cargar los datos del data warehouse.El proceso final es el de despliegue y puesta en producción.



Figura 2. El ciclo de vida del datawarehouse (Calzada, 2020)

III. METODOLOGÍA

En este trabajo se considera el uso de data warehouse para llevar a cabo un análisis y diseño el cual se realizara el proceso de investigación que a continuación se indica.

En la investigación de tipo documental considerada como etapa de investigación, se dedicara al menos tres meses para revisar los procesos que expresan autores, desarrolladores y empresas reconocidas, así como artículos que traten del tema, para identificar la aportación de los data warehouse en la actualidad.

El análisis y diseño de sistemas se refiere al proceso de examinar la situación de un problema con el propósito de mejorar con métodos y procedimientos más adecuados. El método para implementar un data warehouse consta de varios pasos como comprender el problema, análisis. Identificar la fuente de los datos. Diseño del tipo de modelo que se llevara en la primera etapa, crear el modelo del data warehouse que es la carga de las tablas de dimensiones y de hechos, seleccionar la herramientas ETL que se necesitarán.

IV. RESULTADOS

Existen dos metodologías del DW que se imponen que son las de Kimball y la de Inmon.

Inmon describió un data warehouse como una colección de datos orientada a un tema específico, integrado, variante en el tiempo y no volátil, que soporta el proceso de toma de decisiones. (PowerData, 2020)

Kimball describió que el data warehouse es la copia de un data mart(Almacén de datos orientado a un área específica) que describen un evento, "el cambio como resultado de una transacción".

Mencionando en las dos principales metodologías que se presenta se distinguen las dos partes: la de la base de datos relacional la cual contiene todos los datos para conformar el data warehouse.

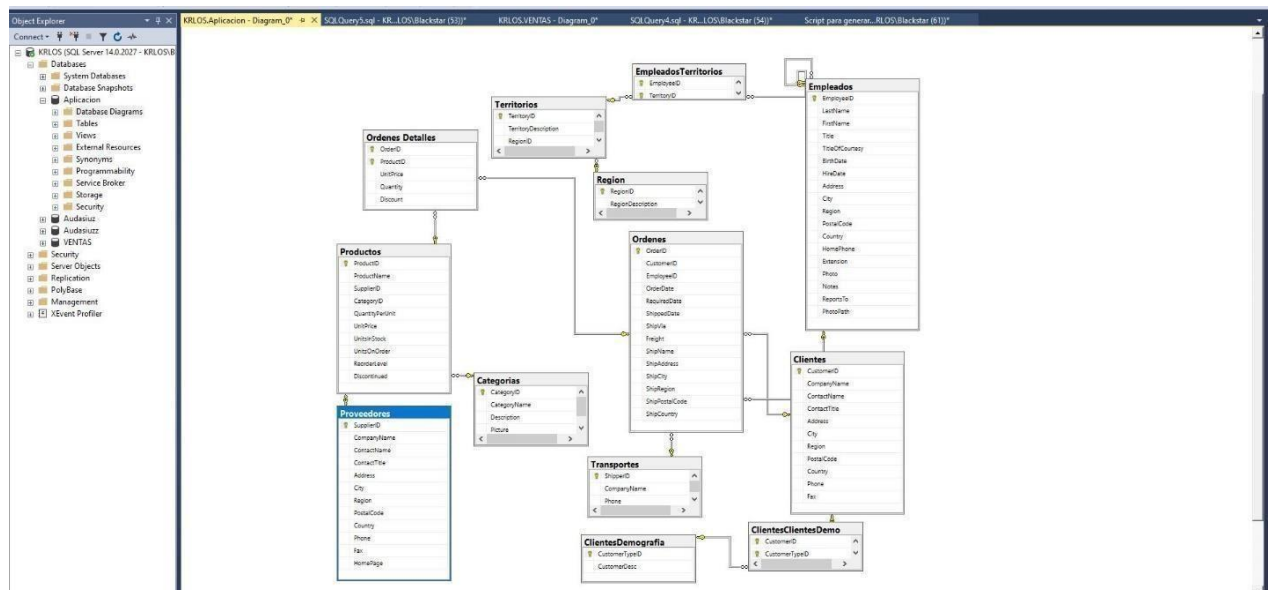


Figura 3. Diagrama Modelo Relacional de la Base de Datos.

De acuerdo al data warehouse se establece en un sistema para una empresa la cual se ve que disminuyen las tablas y es más rápido el acceso a las consultas, ya que puede acceder a los datos de manera fluida, rápida y remota, es una gestión exitosa para cualquier compañía de grandes masa de información tanto por motivos de producción y gestión para las respuestas eficientes a clientes que usen en este caso dicha aplicación.

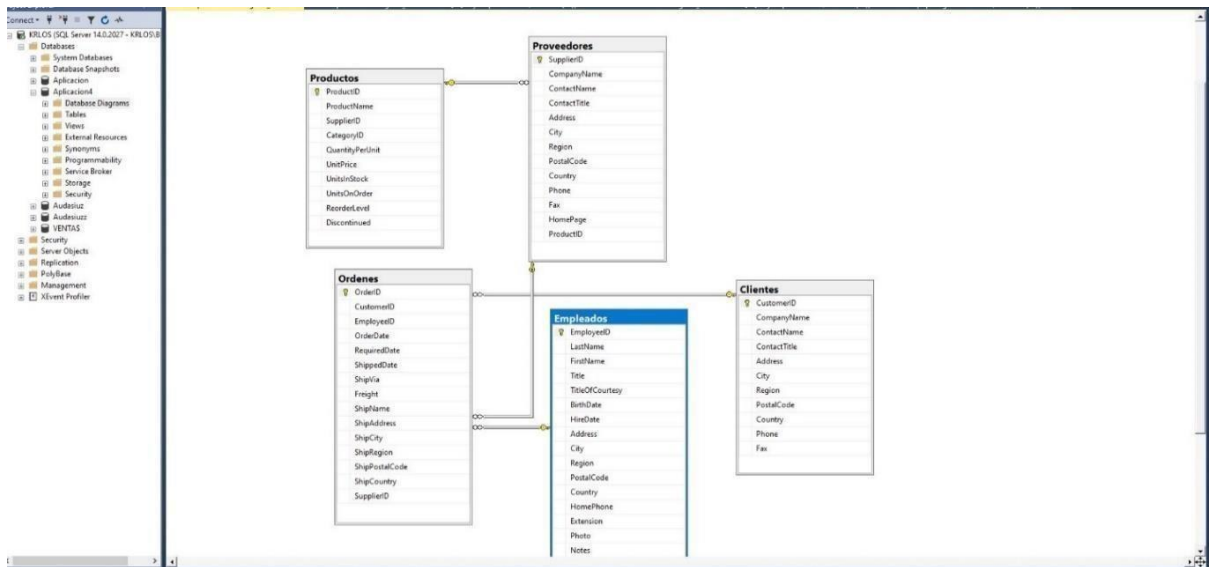


Figura 4. Diagrama simplificado de la Base de Datos.

V. DISCUSIÓN

También tenemos un mayor reto al hacer una migración de datos a un data warehouse es la de preservar la misma inversión con la que se tiene pensado desde un principio, la de almacenar estos mismos y la de maximizar la capacidad con un mínimo de interrupción al hacer esta migración para el negocio.

En este apartado se llevó a cabo el data warehouse dentro de las instalaciones para evitar latencias y reducir riesgos en la complejidad de los datos, ya que también se tienen costos variables en la red en diferentes regiones que afectaría al tenerlo en líneas de red, se mejoró el rendimiento ya que al elegir o procesar datos muy robustos se pudo hacer con base de datos Analytics y aparte se pudo hacer uso de nuestro propio hardware.

La ventaja principal radica en las estructuras de como almacena la información ya que lo puede almacenar por modelos de tablas en estrella, copo de nieve, cubos relacionales, entre otros. Mostramos algunas de las otras ventajas que tiene el data warehouse.

- Facilitan el acceso a una gran variedad de datos

- Proporciona información clave para la toma de decisiones en la empresa como informes de tendencia o de excepción y estos datos son útiles a mediano y largo plazo

- Permite trabajar en conjunto y aumentar el valor operacional de las aplicaciones ya que si las fuentes de datos están claros son sencillos de instalar, contribuye con mayor flexibilidad y rapidez en el acceso a la información

- Es útil para el almacenamiento de análisis y consultas históricas

Desventajas del data warehouse es el alto precio que puede llegar a costar a lo largo de su instancia o vida útil debido a que necesita un mantenimiento constante.

VI.REFERENCIAS BIBLIOGRÁFICAS

Bismart. (01 de 12 de 2020). *bismart*. Obtenido de bismart:

<https://blog.bismart.com/es/cuando-elegir-data-warehouse-para-empresa>

Calzada, J. M. (01 de 12 de 2020). *certia*. Obtenido de certia:

<https://www.certia.net/ciclo-de-vida-del-data-warehouse/#:~:text=Como%20cualquier%20otro%20proyecto%2C%20el,d e%20vida%20del%20data%20warehouse.>

Fintechschooll. (2020). Fintechschooll. Obtenido de Data Warehouse: una forma de almacenar información: <https://escuelafintech.com/data-warehouse/>

Gómez Laura. (2020). *Prezi*. Obtenido de Historia Data Warehouse:

<https://prezi.com/p/rioy6i5yq69g/historia-data-warehouse/>

Holloway Christopher. (12 de 08 de Consultado el 2020). *ITMASTERS MAG*.

Obtenido de Obtenido de ITMASTERS MAG:

<https://itmastersmag.com/analytics-big-data/data-lake-vs-data-warehouse-vs-data-mart-caracteristicas-y-beneficios/>

IBM. (2020). *IBM Data warehouse*. Obtenido de IBM Data warehouse:

https://www.ibm.com/mx-es/analytics/data-warehouse?p1=Search&p4=43700052831757837&p5=b&cm_mmc=Search_Google-_-1S_1S-_-LA_MX-_-%2Bdata%20%2Bwarehouse_b&cm_mmca7=71700000065287838&cm_mmca8=aud-311016886972:kwd-19689304787&cm_mmca9=Cj0KCQjw6575BRCQARIsAMp-ks

- IMF Business School. (2020). *IMF Business School*. Obtenido de Data Warehouse, la nueva columna vertebral de las organizaciones: <https://blogs.imf-formacion.com/blog/tecnologia/data-warehouse-organizaciones-201802/>
- Informatica LLC. (2020). *Informatica LLC*. Obtenido de La guía definitiva para la modernización del data warehouse: <https://www.informatica.com/mx/lp/the-definitive-guide-to-data-warehouse-modernization.html>
- Oracle México. (2020). *Oracle México, Base de datos*. Obtenido de ¿Qué es un almacén de datos?: <https://www.oracle.com/mx/database/what-is-a-data-warehouse/>
- Pearlman, S. (2020). *talend*. Obtenido de ¿Qué es un almacén de datos y por qué es relevante para su empresa?: <https://es.talend.com/resources/what-is-data-warehouse/>
- PowerData. (2020). *PowerData*. Obtenido de ¿Qué es un Data Warehouse y cómo saber cuándo lo necesitas implementar?: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/que-es-un-data-warehouse-y-como-saber-cuando-lo-necesitas-implementar>
- PowerData. (2020). *PowerData*. Obtenido de Data Warehouse: todo lo que necesitas saber sobre almacenamiento de datos: <https://www.powerdata.es/data-warehouse>
- Proal Carlos. (25 de 11 de 2020). *Evolución de los Decision Support Systems*. Obtenido de Evolución de los Decision Support Systems: <http://carlosproal.com/dw/dw01.html#:~:text=La%20definici%C3%B3n%20universal%20de%20un,estrategias%20de%20toma%20de%20decisiones%E2%80%9D>.

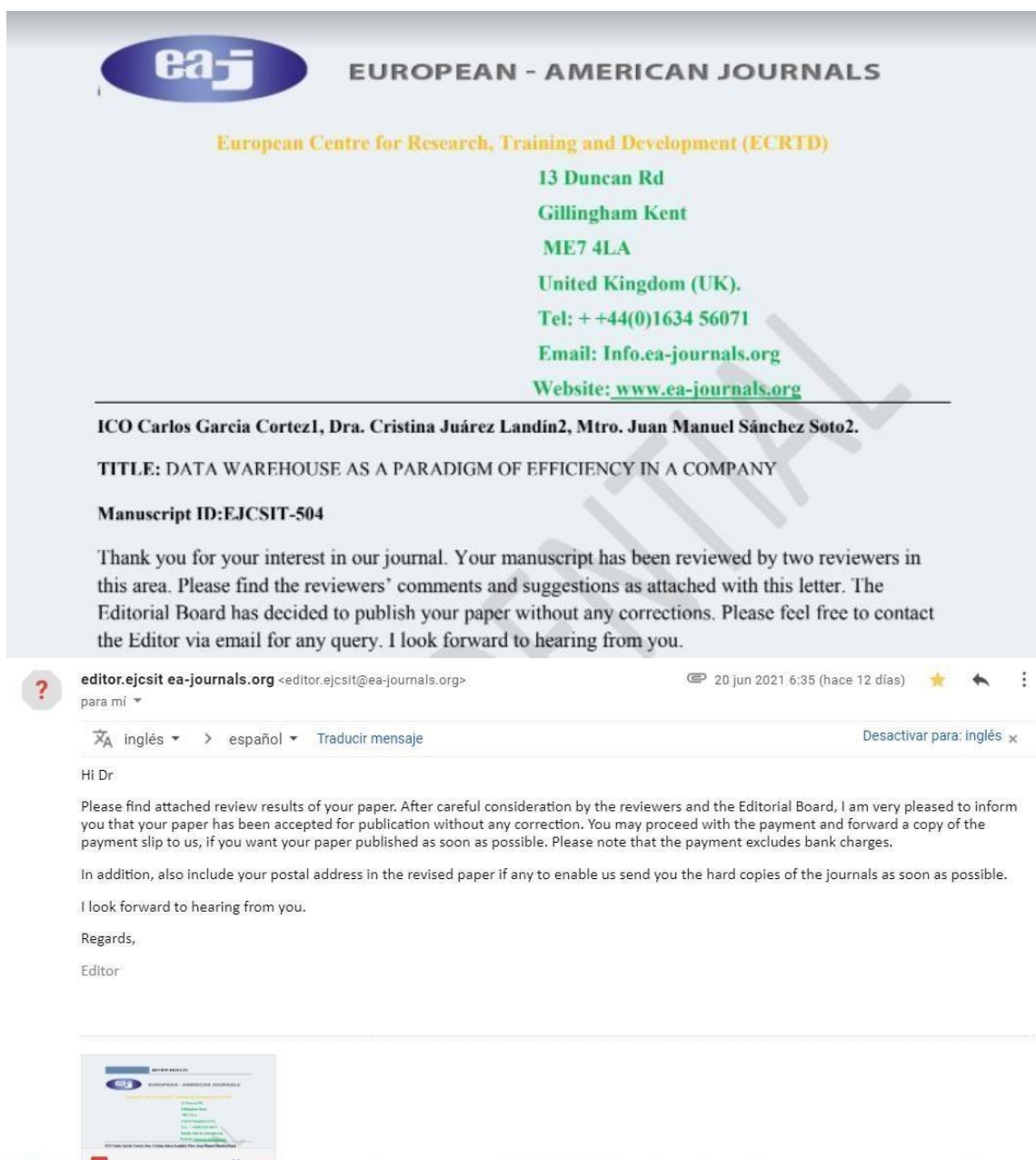
Rivadera, G. (10 de 11 de 2020). *La metodología de Kimball para el diseño de almacenes de*. Obtenido de Rivadera: La Metodología de Kimball para el Diseño de almacenes...:

<https://www.ucasal.edu.ar/hm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>

Sierra, Y. (25 de 11 de 2020). *MediaCloud*. Obtenido de MediaCloud ADN:
<https://blog.mdcloud.es/dwh-ejemplos-arquitectura-data-warehouse/>

Sinergia e Inteligencia de Negocio. (2020). *Sinnexus*. Obtenido de Manual de BI:
https://www.sinnexus.com/business_intelligence/datawarehouse.aspx

VII. ANEXOS



ea-j EUROPEAN - AMERICAN JOURNALS

European Centre for Research, Training and Development (ECRTD)

13 Duncan Rd
Gillingham Kent
ME7 4LA
United Kingdom (UK).
Tel: +44(0)1634 56071
Email: Info.ea-journals.org
Website: www.ea-journals.org

ICO Carlos Garcia Cortez1, Dra. Cristina Juárez Landín2, Mtro. Juan Manuel Sánchez Soto2.

TITLE: DATA WAREHOUSE AS A PARADIGM OF EFFICIENCY IN A COMPANY

Manuscript ID:EJCSIT-504

Thank you for your interest in our journal. Your manuscript has been reviewed by two reviewers in this area. Please find the reviewers' comments and suggestions as attached with this letter. The Editorial Board has decided to publish your paper without any corrections. Please feel free to contact the Editor via email for any query. I look forward to hearing from you.

editor.ejcsit ea-journals.org <editor.ejcsit@ea-journals.org> 20 jun 2021 6:35 (hace 12 días) ★ ↶ ⋮
para mí ▾

inglés > español Traducir mensaje Desactivar para: inglés x


Hi Dr,

Please find attached review results of your paper. After careful consideration by the reviewers and the Editorial Board, I am very pleased to inform you that your paper has been accepted for publication without any correction. You may proceed with the payment and forward a copy of the payment slip to us, if you want your paper published as soon as possible. Please note that the payment excludes bank charges.

In addition, also include your postal address in the revised paper if any to enable us send you the hard copies of the journals as soon as possible.

I look forward to hearing from you.

Regards,
Editor



REVIEW rRESULTS.pdf

Se anexa el documento de la revista

Hola Dr.

Por favor descargue su documento:

<https://www.eajournals.org/journals/european-journal-of-computer-science-and-information-technology-ejcsit/vol-9-issue-3-2021/>

Quiere 9, Número 3, 2021 - EA Journals

Tel: +44 (0) 1634 560711 Fax: +44 (0) 1582 515277 Correo electrónico: info@ea-journals.org Kemp House, 152-160 City Road, Londres, EC1V 2NX. Reino Unido

www.eajournals.org

Data Warehouse as a Paradigm of Efficiency in a Company

Article Author: ICO Carlos García Cortez, Dra. Cristina Juárez Landín, Mtro. Juan Manuel Sánchez Soto

Abstract

This paper presents the paradigm of the data warehouse and its use in order to show growth, evolution and optimization of the different data that lead to complex systems, taking into account that in transnational companies they collect large amounts of data from various systems, therefore the traditional data warehouse cannot continue with the usual process as the data now have a demand for rapidly increasing volumes. This article is focused on making companies and people know a tool to add data sources from a central location so that they can support businesses and create reports since through the data warehouse will provide the capacity to processing large quantities of data, until you can store petabytes of information. The implementation of data warehouse in most cases is the first step to a complete and reliable solution of business intelligence to obtain better results in the problems of the organization, it should be stressed that it facilitates the process for management and is a great solution for companies that consider deleting data that interferes with the analysis of information and its delivery, as it analyses all types of data that are useful to its users, structures for easy access and operation

Keywords: Business Intelligence, data warehouse, petabytes

Article Review Status: Published

ECRTD Members

| |
|--|
| Business and Management |
| Education and Research Methods |
| Engineering, Science and Technology |
| English Language and Inter-Continental Studies |
| Health and Agricultural Matters |
| Health and Food Sciences |
| Humanities and Social Science |
| Scientific Matters |

European Journal of Computer Science and Information Technology (EJCSIT)

Vol 9, Issue 3, 2021

European Journal of Computer Science and Information Technology is peer-reviewed journal published by the European Centre for Research, Training and Development (ECRTD), UK. EJCSIT welcomes research papers in Artificial Intelligence, Computer Architecture and Engineering, Computer Graphics and Visualisation, Software Engineering, Information Science, Computational Science, Computer Security and Cryptography, Semantic Web, Information Systems, Documentation Management, Human Computer Integration, Information Ethics, Knowledge Management, Personal Information Management, User Centre Design, Group Wise, Data Modelling, and other related areas. European Journal of Computer Science and Information Technology is published in both online and printed versions.

Published Status: Current

List of Articles

DATA WAREHOUSE AS A PARADIGM OF EFFICIENCY IN A COMPANY

ICO Carlos García Cortez, Dra. Cristina Juárez Landín, Mtro. Juan Manuel Sánchez Soto

Keywords: Business Intelligence, data warehouse, petabytes

VIEW PDF PAGES

1-7

ECRTD Members

| |
|--|
| Business and Management |
| Education and Research Methods |
| Engineering, Science and Technology |
| English Language and Inter-Continental Studies |
| Health and Agricultural Matters |
| Health and Food Sciences |
| Humanities and Social Science |
| Scientific Matters |