

**II Conferencia Internacional de Procesamiento de la Información
(CIPI2019)**

**Selección de la metodología apropiada para la solución de un almacén
de datos**

*Selection of the appropriate methodology for the solution of a data
warehouse*

Luis Izquierdo Rivero¹, Rosendo de Jesús Moreno Rodríguez²

1-Luis Izquierdo Rivero. Empresa Electroquímica de Sagua, Cuba,
lirivero92@gmail.com

2-Rosendo de Jesús Moreno Rodríguez, Universidad Central “Marta Abreu de las
Villas”, Cuba, rmoreno@uclv.edu.cu

Resumen: Los almacenes de datos [acrónimo del inglés data warehouses (DWH)] son cada vez más importantes para las organizaciones que manejan un gran volumen de información histórica. Actualmente hay varias metodologías que ayudan a simplificar el proceso de diseño y desarrollo del DWH, pero cada una de ellas tiene diferentes enfoques que al conocerlos hace que sea más fácil elegir cuál es el ideal para su empresa. Este artículo de revisión bibliográfica muestra cuán importante puede ser la construcción de un DWH utilizando la metodología apropiada para su entorno empresarial.

Abstract: *Data warehouses [acronym for English data warehouses (DWH)] are becoming increasingly important for organizations that handle a large historical volume of information. Currently there are several methodologies that help streamline the process of design and development of the DWH, but each of them have different approaches that knowing them makes it easier to choose which is the ideal for your company. This bibliographic review article shows how important the construction of a DWH can be using the appropriate methodology for its business environment.*

Palabras Clave: Almacenes de datos; Metodologías

Keywords: *Data Warehouses - Methodologies*

1. Introducción

La mayoría de las decisiones de empresas, organizaciones e instituciones se basan en información de experiencias pasadas. Generalmente, la información que es necesaria para investigar sobre un cierto dominio de la organización se encuentra en sistemas operacionales de bases de datos, tanto internas como externas, y otras fuentes muy diversas, no necesariamente bases de datos (Gómez 2015). Desde hace varias décadas las organizaciones empresariales han buscado en el almacenamiento de datos de sus sistemas operacionales soluciones que les ayuden a atender sus necesidades a la hora de tomar decisiones de negocio (Sanz 2010). Los usuarios que toman estas decisiones y planifican día a día, a mediano o a largo plazo, la calidad, disponibilidad y presentación de la información juegan un papel categórico, pues este tipo de usuarios necesitan disponer de información tanto consolidada y detallada de cómo marchan las actividades ya cumplidas, predecir tendencias y comportamientos para tomar decisiones proactivas (Marroquín & Tejada 2007).

Por otra parte, las necesidades de información hoy en día han variado. La disponibilidad de gran cantidad de información es de vital importancia para los negocios, ya que las decisiones de futuro se suelen tomar sobre la base de dicha información. Este tipo de necesidades para reflejar tendencias, evoluciones, hechos históricos en el negocio y posibilidades futuras son temas que la alta dirección de las instituciones y empresas deben manejar y manejan de una forma habitual y son las causantes de que hayan aparecido en el mercado herramientas denominadas "ayudas a la toma de decisiones" (Sanz 2010).

El concepto de DWH, cuya traducción literal sería almacén o repositorio de datos, surge alrededor del año 1990 con la necesidad de recopilar información de datos acumulados durante años por los sistemas de gestión. Este concepto nace como producto de la evolución de los sistemas para dar soporte a la toma de decisiones.

Los DWH surgen por la necesidad de resolver problemas de análisis de grandes masas de información para la creación y mantenimiento de estructuras destinadas al análisis de datos (Villanueva 2008), estos se subdividen en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sea necesario; dichas unidades lógicas se denominan Mercado de Datos [acrónimo del inglés Data Mart (DM)], los cuales resuelven estudios a nivel de departamento. Un DM es una

versión del almacén de datos, la diferencia principal es que la creación de un DM es específica para una necesidad de datos seleccionados, enfatizando el fácil acceso a una información relevante y el análisis de datos estadísticos que apoyan la toma de decisiones (Rodríguez et al. 2013). Soportan varios tipos de aplicaciones, como OLAP, DSS y aplicaciones de minería de datos. (Villanueva 2008)

Para abordar un proyecto de DWH es necesario hacer un estudio de algunos temas generales de la organización o empresa, los cuales se describen a continuación:

- ✓ **Situación actual de partida:** Cualquier solución propuesta de DWH debe estar muy orientada por las necesidades del negocio y debe ser compatible con la arquitectura técnica existente y planeada de la compañía.
- ✓ **Tipo y características del negocio:** Es indispensable tener el conocimiento exacto sobre el tipo de negocios de la organización y el soporte que representa la información dentro de todo su proceso de toma de decisiones.
- ✓ **Entorno técnico:** Se debe incluir tanto el aspecto del hardware (mainframes, servidores, redes) así como aplicaciones y herramientas. Se dará énfasis a los Sistemas de Soporte a Decisiones (DSS).
- ✓ **Expectativas de los usuarios:** Un proyecto de DWH no es únicamente un proyecto tecnológico, es una forma de vida de las organizaciones y como tal, tiene que contar con el apoyo de todos los usuarios y su convencimiento sobre su bondad.
- ✓ **Etapas de desarrollo:** Con el conocimiento previo, ya se entra en el desarrollo de un modelo conceptual para la construcción del DWH.
- ✓ **Prototipo:** Un prototipo es un esfuerzo designado a simular tanto como sea posible el producto final que será entregado a los usuarios.
- ✓ **Piloto:** El piloto de un DWH es el primero, o cada uno de los primeros resultados generados de forma iterativa que se harán para llegar a la construcción del producto final deseado.
- ✓ **Prueba del concepto tecnológico:** Es un paso opcional que se puede necesitar para determinar si la arquitectura especificada del DWH funcionará finalmente como se espera.

Las metodologías más usadas en la literatura para el desarrollo de un DWH están estructuradas en dos bloques: top-down y bottom-up, que se corresponden con las

metodologías propuestas por (Inmon 1992) y (Kimball & Ross 2002), respectivamente. Estos autores, en muchos aspectos, se consideran los precursores del DWH (Rodríguez 2010).

2. Metodología

Una metodología no es más que aquella guía que se sigue con el fin de realizar las acciones propias de una investigación. En términos más sencillos se trata de la guía que nos va indicando qué hacer y cómo actuar cuando se quiere obtener algún tipo de investigación. Es posible definir una metodología como aquel enfoque que permite observar un problema de una forma total, sistemática y disciplinada. Esta investigación tiene como objetivo realizar una breve descripción de las metodologías más utilizadas en la construcción de un DWH ideal en un entorno empresarial determinado.

2.1 Metodología top-down

La metodología descendente (top-down) se define en el año 1992 por Bill Inmon (Inmon 1992), se actualiza en el 2002 (Inmon 2002) y se basa en conceptos bien conocidos del diseño de bases de datos relacionales (Inmon 2002; Imhoff et al. 2003), en esta metodología los DM se crearán después de haber terminado el DWH completo de la organización.

La metodología top-down tiene un enfoque a modo de explosión porque en cierta forma no viene acompañada del ciclo de vida normal de las aplicaciones, sino que los requisitos van acompañando al proyecto según vaya comprobándose su necesidad. Esto ocasiona riesgos a la compañía, que invierte grandes esfuerzos en el desarrollo del DWH y no explota la inversión y obtiene los beneficios hasta la aparición de los DM (García 2014). Este enfoque tiende a presentar problemas en organizaciones que presenten cambios fundamentales en su proceso de negocio.

La metodología que Bill Inmon propone es iterativa la cual sigue un esquema contrario al clásico de desarrollo de sistemas ya que lo primero con lo que se trabaja son datos, estos se integran para ser probados y programar de acuerdo a ellos para analizar los resultados y de esta manera comprender los requerimientos.

2.2 Metodología bottom-up

La metodología bottom-up se convierte en el estándar para el apoyo a las decisiones empresariales (Rodríguez 2010). La misma se basa en lo que (Kimball & Ross 2002) denominan Ciclo de Vida Dimensional del Negocio que utiliza el modelo dimensional (Kimball et al. 1998; Mundy et al. 2006; Kimball et al. 2008; Rivadera 2010). La metodología de Kimball conduce a una solución completa en una cantidad de tiempo relativamente pequeña.

Además, debido a la gran cantidad de documentación que se puede encontrar y a los numerosos ejemplos aportados en diferentes entornos, permite hallar una respuesta a casi todas las preguntas que puedan surgir, sobre todo cuando no se dispone de la experiencia previa necesaria. Es ideal para los primeros pasos de la implantación de BI a un cliente, cuando la complejidad del almacenamiento de datos no es demasiado grande y donde la infraestructura del BI se encarga de los datos procedentes de un número limitado de fuentes (Sanz 2010).

2.3 Metodología de HEFESTO

HEFESTO es una metodología desarrollada por el Ing. Bernabeu Ricardo Dario cuya propuesta se fundamenta en una amplia investigación, comparación de metodologías existentes y la experiencia en la elaboración de almacenes de datos. Cabe destacar que está en continua evolución, y se han tenido en cuenta, como gran valor agregado, todos los feedbacks que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines (Dario and Mattío 2017). La ventaja principal de esta metodología es que especifica puntualmente los pasos a seguir en cada fase a diferencia de otras metodologías que mencionan los procesos, pero no explican cómo realizarlos. Según (Dario and Mattío 2017) esta metodología puede ser embebida en cualquier ciclo de vida que cumpla la siguiente condición:

- ✓ No entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas.

Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DWH y motivar a los usuarios.

3. Resultados y discusión

Al plantear estas metodologías en el proceso de creación de un DWH es importante tener en cuenta la diferencia de visiones entre los autores citados. Para Bill Inmon el DWH es una parte de un sistema de la inteligencia de negocio [acrónimo del inglés business intelligent (BI)] dentro de una empresa, que tiene un DWH y los DM obtienen su información a partir de este DWH. Por otro lado, Ralph Kimball plantea que el DWH es un conglomerado de todos los DM dentro de una empresa y la información siempre se almacena de acuerdo al Modelo Dimensional. Esto implica que, para Bill Inmon el desarrollo del DWH debe ser completo para su correcto funcionamiento mientras que la visión de Ralph Kimball nos permite desarrollar DM particulares que contengan la lógica de negocio en la que se está interesado en profundizar. Por otra parte, HEFESTO se basa más en el estudio y comparación de las metodologías existentes, interpreta de forma óptima las necesidades de información de la empresa satisfaciendo la necesidad del negocio lo que implica una mejor construcción y adaptación del almacén de datos. En la siguiente tabla se comparan las metodologías por diferentes factores de análisis dentro de los cuales se destacan: complejidad de implementación, tiempo de desarrollo, tiempo en construcción, adaptabilidad sobre cualquier tecnología, tamaño del proyecto, más usado en el mundo y documentación técnica.

FACTORES DE ANÁLISIS PUNTAJE	INMON	KIMBALL	HEFESTO
Bueno/Simple/Alto/Pequeño=3, Medio/Mediano=2, Compleja/Costoso/Bajo/Grande=1 TODOS/SI=1 Nulo/NO=0,			
Generalización	General a detalle	Detalle a General	Detalle a General
Arquitectura Orientada a	DWH	DM	DWH y DM
Complejidad de Implementación	Compleja (1)	Simple (3)	Simple (3)
Usabilidad para el usuario	Bajo (1)	Alto (3)	Alto (3)

**II CONVENCION CIENTIFICA INTERNACIONAL
"II CCI UCLV 2019"**



**2019
UCLV**

Orientado a	Orientado a temas	Orientado a proceso	Orientado a temas
Modelamiento	Tradicional	Dimensional	Dimensional
Esquemas de modelamiento	Normalizado	Desnormalizado	Desnormalizado o semidesnormalizado
Manejo de Cambios en Dimensiones	Continuo y discreto	Dimensiones Cambiantes	Dimensiones Cambiantes
Dirigido a	IT	Usuarios finales	Usuarios finales
Tiempo de desarrollo	Largo plazo	Corto y mediano plazo	Corto y mediano plazo
Ayuda a toma de decisiones	Estratégicas	Tácticas	Estratégicas y tácticas
Flexibilidad	Bajo (1)	Alto (3)	Alto (3)
Costo de Implementación	Costoso	Medio (2)	Medio (2)
Equipo de Desarrollo	Especialistas	Generalistas	Especialistas
Adaptable sobre cualquier tecnología	Si (1)	Si (1)	Si (1)
Tamaño del Proyecto	Grande (1)	Grande (1)	Pequeño o mediano (2)
Tiempo en el análisis y diseño	Costoso (1) por ser iterativo	Medio (2) una sola vez	Medio (2) una sola vez
Tiempo en construcción	Costoso (1)	Medio (2)	Medio (2)
Perspectiva	Relacional	Estrella	Estrella/Copo Nieve
Más usada en el mundo	Alto (3)	Medio (2)	Alto (3)
Documentación Técnica	Si (1) Extensa	Si (1) Extensa	Si (1) Extensa
TOTAL	11	20	22

Tabla 1. Comparación de las metodologías de Inmon, Kimball y HEFESTO

Como se aprecia en la *Tabla 1* las metodologías de Kimball y HEFESTO son las mejores puntuadas destacando factores de análisis como: complejidad de implementación,

usabilidad para el usuario, tiempo de desarrollo, flexibilidad, costo de implementación, tiempo en el análisis y diseño, tiempo en construcción.

A pesar de que ambas metodologías no contienen un estudio previo relacionado con las ventajas y desventajas que el DWH proporcionaría a la empresa, estas realizan requerimientos amplios para lograr una correcta implementación del DWH ya que recolectan datos mediante entrevistas, análisis de documentos y sistemas fuentes, destacando a HEFESTO que adicionalmente realiza un análisis del modelo conceptual de los sistemas fuente.

Cabe destacar que la metodología de HEFESTO es una de las más completas en la actualidad porque demuestra de manera metódica y sencilla como podemos comprender cada paso que se ejecuta para entender el porqué del proceso y no caer en el error de seguir un método sin saber que se está haciendo, su arquitectura es orientada tanto a los DWH como a los DM, lo que implica que sea una metodología muy poderosa para la construcción de herramientas para la toma de decisiones.

4. Conclusiones

La metodología de Kimball está dirigida principalmente a pequeños y medianos desarrollos, en donde las características esenciales son la usabilidad, facilidad para el usuario y el rápido desarrollo de la solución, es ideal para los primeros pasos de la implantación del BI a un cliente, cuando la complejidad del almacenamiento de datos no es demasiado grande y donde la infraestructura del BI se encarga de los datos procedentes de un número limitado de fuentes, es mejor utilizarla cuando: se quiere enfatizar la planificación y el conocimiento de una parte del proceso de negocio, se desee obtener información histórica de algún departamento en específico del negocio y no se tenga suficiente tiempo para implementar el DWH y se desee construir por departamentos el almacén. Sin embargo, cuando el almacén de datos adquiere complejidad el desarrollo de esta metodología se complica. En el mundo del BI, cuando las cosas adquieren gran complejidad, es el momento de introducir nuevos enfoques al problema, como el propuesto por Inmon.

La metodología de Inmon está orientada al desarrollo de soluciones complejas y de gran alcance brindando perdurabilidad y robustez a la solución, es mejor utilizarla cuando: se quiere enfatizar la planificación y el conocimiento completo del proceso de negocio,

se desee obtener información histórica de todos los departamentos del negocio, se tiene tiempo suficiente para implementar el DWH y no se realicen cambios fundamentales en el proceso de negocio.

La metodología de HEFESTO especifica puntualmente los pasos a seguir en cada fase a diferencia de otras metodologías que mencionan los procesos y no explican cómo realizarlos, esto hace una total comprensión por los que la escogen, ayudando a comprender mejor el proceso de negocio, se puede utilizar en cualquier entorno empresarial siempre y cuando se mantenga su buena práctica de desarrollo.

5. Referencias bibliográficas

1. DARIO, B. R. & MATTÍO, M. G. 2017. HEFESTO DATA WAREHOUSING. *Guía completa de aplicación teórico-práctica; metodología Data Warehouse.*
2. INMON, B. 1992. *Building the Data Warehouse*, Technical Publishing Group.
3. INMON, W. 2002. *Building the Data Warehouse*, John Wiley & Sons, Inc.
4. KIMBALL, R. & ROSS, M. 2002. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, New York, John Wiley and Sons, Inc.
5. RODRÍGUEZ, M. 2010. *Análisis y Diseño de un Data Mart para el seguimiento académico de alumnos en un entorno universitario*. Universidad Carlos III de Madrid.
6. VILLANUEVA, W. D. 2008. *Almacenes de datos*.
7. Gómez, O.J.R., 2015. *Sistema ETL para el Almacén de Datos de Aspirantes de la UCLV*.
8. Marroquín, R.M.R. & Tejada, R.P.A., 2007. *Desarrollo de un datamart de información*.
9. Sanz, M.R., 2010. *Análisis y diseño de un data mart para el seguimiento académico de alumnos en un entorno universitario*.
10. Rodríguez, Y.V. et al., 2013. *Mercado de datos para la dirección de cuadros de la Administración Provincial de Artemisa*, 7(1), pp.1–8.
11. Dario, B.R., 2010. HEFESTO: *Investigación y Sistematización de Conceptos y Metodología para la Construcción de un Data Warehouse*.
12. García, J.L., 2014. *Automatización de los procesos de carga en el mercado de datos Recursos Humanos de la UCLV*. Universidad Central “Marta Abreu” de Las Villas.
13. Infante, L. (2009). *Business Intelligence Latin America*.
14. Gutiérrez, V., 2010. *Cómo abordar un proyecto de business intelligence*

II CONVENCION CIENTIFICA INTERNACIONAL
“II CCI UCLV 2019”



2019
UCLV

15. Zaldívar, A.R., 2014. *Implementación de un data mart como solución de inteligencia de negocios, bajo la metodología de ralph kimball para optimizar la toma de decisiones en el departamento de finanzas de la contraloría general de la república.*
16. ROJAS, W. C., QUISPE, F. M. & MOLINA, F. F. 2018. *Una Metodología para Procesos Data WareHousing Basada en la Experiencia.*
17. BLANCO, Y. C., BRIZUELA, E. I. L. & PELLICER, Y. S. 2014. *Almacén de Datos para la Gestión Energética en el Ministerio de la Agricultura.*
18. PEÑAFIEL, G. E. S. 2018. *Análisis de metodologías para la implementación de un data warehouse aplicado a la toma de decisiones del instituto nacional de patrimonio cultural regional 3.*
19. TORRES, V. A. T., TORRES, E. N., HERNÁNDEZ, Y. M., FERIA, D. C., GONZÁLEZ, Y. P., ROBLEJO, M. G. & BLANCO, Y. A. 2015. *Inteligencia de negocios para la empresa de servicios de la Unión del Níquel.*