

## **Metodologías para desarrollar Almacén de Datos.**

*Methodologies to develop Data Warehouse.*



**Ing. Eric Ismael Leonard Brizuela**

Profesor Instructor

Universidad de Granma, Bayamo, Granma, Cuba,

Teléfono: 48 16 49 Email: [eleonardb@udg.co.cu](mailto:eleonardb@udg.co.cu)



**MSc. Yudi Castro Blanco**

Profesora Asistente

Universidad de Granma, Bayamo, Granma, Cuba.

Teléfono: 48 10 15 ext.: 208 Email: [ycastrob@udg.co.cu](mailto:ycastrob@udg.co.cu)

Recibido: 11-09-13

Aceptado: 04-11-13

### **Resumen:**

El desarrollo de un almacén de datos no es tarea fácil, para llevar a cabo su implementación es necesario disponer de la metodología adecuada; se requiere el diseño de un modelo conceptual que incluye tanto los requisitos de información de los usuarios así como las fuentes de datos operacionales, a partir del cual se obtiene un modelo lógico basado en una tecnología de base de datos específica que guía la implementación. Actualmente muchas de las metodologías existentes no definen mecanismos que abarquen las características particulares del desarrollo de un almacén de datos, convirtiéndolo en una tarea compleja y artesanal. Para dar solución a este problema, en esta investigación se realiza un estudio de varias metodologías para el desarrollo de almacenes de datos, realizando un análisis de sus principales características para determinar la más apropiada.

**Palabras clave:** Metodología, Almacén de Datos.

### **Abstract:**

The development of a data warehouse is not an easy task, in order to carry out its implementation it is necessary to have the adequate methodology; the design of a conceptual model takes that as much includes the requirements of information of the users as well as the operational data sources, from which obtains a logical model based in a data base specific technology that directs the implementation itself. Many of the existing methodologies do not define mechanisms that cover the particular characteristics for development of a data warehouse, turning it into a complex and craft task. In order to give solution to this problem, in this investigation a study of several developmental methodologies of data warehouses comes true, conducting an analysis of its main features to determine the most appropriate.

**Keywords:** Methodology, Data Warehouse.

### **Introducción:**

Los Almacenes de Datos o Data Warehouse (DW por sus siglas en Inglés), surgieron en la década del 90 del siglo pasado, conocidos como “una colección de datos orientados a un ámbito (empresa, organización), integrada, no volátil y variante en el tiempo, que ayuda al proceso de los sistemas de soporte a la toma de decisiones”(1). El diseño y construcción de los almacenes de datos están ganando cada vez mayor popularidad en las organizaciones, al considerar las ventajas que involucra el análisis de los datos históricos de forma multidimensional para apoyar el proceso de toma de decisiones (2)., resultando complejo en este proceso la recolección de requerimientos, el análisis y el diseño porque no siempre se emplea la metodología adecuada.

La definición de metodología para algunos autores es: “una colección de procedimientos, técnicas, herramientas y documentos auxiliares que ayudan a los desarrolladores de software en sus esfuerzos por implementar nuevos sistemas de información”. Una metodología está formada por fases, cada una de las cuales se puede dividir en sub-fases, que guiarán a los desarrolladores de sistemas a elegir las técnicas más apropiadas en cada momento del proyecto y también a planificarlo, gestionarlo, controlarlo y evaluarlo (3-6).

La finalidad de una metodología de desarrollo es garantizar la eficacia (p.ej. cumplir los requisitos iniciales) y la eficiencia (p.ej. minimizar las pérdidas de tiempo) en el proceso de generación de software, pero no siempre se han distinguido por ser muy exitosas, aún menos por su popularidad. La crítica más frecuente que sufren las metodologías es que son burocráticas, reiterativas y ambiguas. Hay tanto que hacer para seguir la metodología que el ritmo entero del desarrollo se retarda.

Hoy en día existen varias propuestas metodológicas que inciden en distintas dimensiones del proceso de desarrollo de un almacén de datos para facilitar la realización de nuevos proyectos y reducir su complejidad. Cada una de estas metodologías están marcadas por características particulares que las hacen diferentes, la elección de una o varias de ella para la construcción de un almacén de datos debe ser estudiado cuidadosamente.

### **Desarrollo:**

Varias personas e instituciones se han dado la tarea de estudiar el tema sobre la metodología correcta para la construcción de un almacén de datos (7), algunos han creado sus propias metodologías, otros han modificado las existentes, entre ellas se pueden mencionar: Ciclo de vida Kimball (8), DWEP (9), la propuesta de Trujillo (10), Rapid Warehousing Methodology (11) y HEFESTO (12).

Existen otras metodologías que pudieran aplicarse al desarrollo de un almacén de datos y que la comunidad científica considera como probadas (13-15) pero se enfocan hacia la minería de datos, ellas son: CRISP-DM, SEMMA y P<sup>3</sup>TQ. Existe además una metodología llamada KM-IRIS (16) que pudiera

aplicarse de cierto modo aunque fue concebida para dirigir proyectos de desarrollo de Sistemas de Gestión del Conocimiento.

### Arquitectura de un Almacén de Datos

Antes de analizar las características particulares de cada metodología es necesario conocer la arquitectura de un almacén de datos, la que incluye: datos operacionales, extracción, transformación y carga de los datos, almacén y herramienta de acceso al depósito; como se muestra en la Figura 1, sus componentes según (12) son:

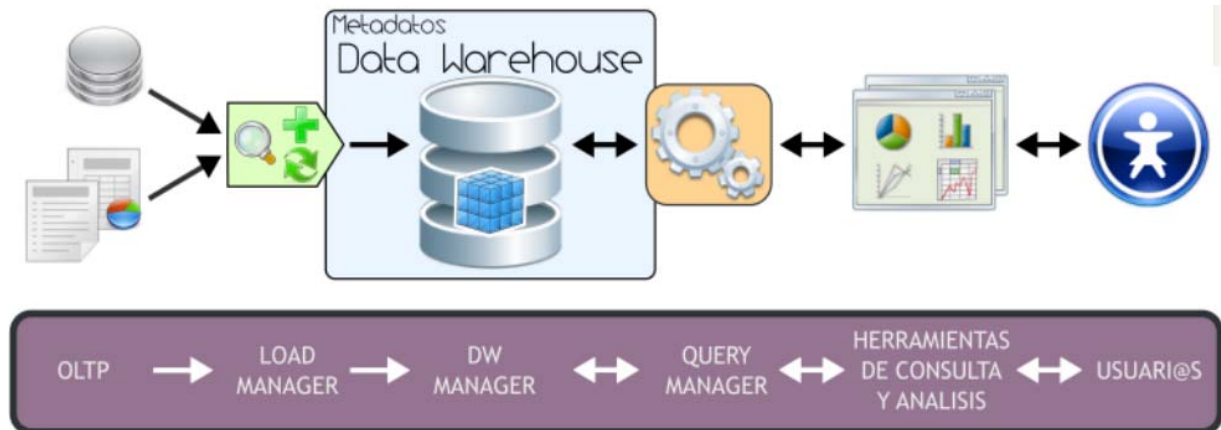


Figura 1: Arquitectura de un Almacén de Datos.

OLTP (On-line Transaction Processing / Procesamiento de Transacciones en línea), representa toda aquella información transaccional que genera la organización diariamente y las fuentes externas.

LOAD MANAGER. Los ETL (Extract-Transform-Load / Extracción, Transformación y Carga) se encargan de extraer los datos desde los OLTP para manipularlos, integrarlos, transformarlos y posteriormente cargar los resultados obtenidos en el almacén de datos, es necesario contar con un sistema que se encargue de ello.

DW MANAGER. Su finalidad es transformar e integrar los datos fuentes y de almacenamiento intermedio en un modelo adecuado para la toma de decisiones. Permitiendo realizar todas las funciones de definición y manipulación del depósito de datos, para poder soportar todos los procesos de gestión del mismo.

QUERY MANAGER. Este componente realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, propias del análisis de datos, recibe las consultas del usuario, las aplica a la estructura de datos correspondiente y devuelve los resultados obtenidos.

HERRAMIENTAS Y CONSULTAS DE DATOS. Son los sistemas que permiten al usuario realizar la exploración de datos del almacén de Datos. Básicamente constituyen el nexo entre el depósito de datos y los usuarios.

USUARIOS. Son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio.

Esta arquitectura opera de la siguiente manera:

Los datos se extraen de aplicaciones, bases de datos, archivos, entre otros. Esta información generalmente reside en diferentes tipos de sistemas, orígenes y arquitecturas con diferentes formatos, los datos son integrados, transformados y limpiados, para luego ser cargados en el almacén de Datos, la información se estructura en cubos multidimensionales para responder a consultas dinámicas con una buena presentación.

Los usuarios acceden a los cubos, utilizando diversas herramientas de consulta: Query & Reporting, On-line analytical processing (OLAP), Executive Information System (EIS) ó Información de gestión, Decision Support Systems (DSS), Visualización de la información, Data Mining, etc (17).

### Características de las metodologías

El **Ciclo de vida Kimball** (18) presenta un marco de trabajo como el ilustrado en la Figura 2, en la cual se muestran las diferentes etapas durante todo el proceso de creación del almacén de datos.

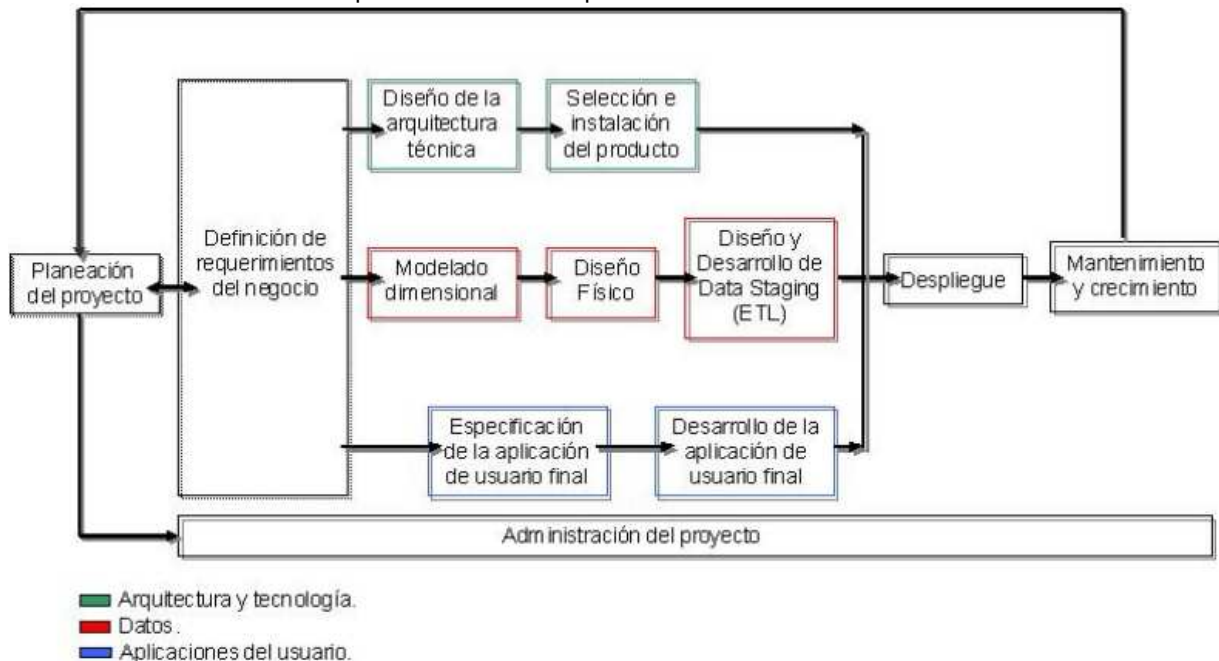


Figura 2: Ciclo de vida para la construcción de un almacén de datos según Ralph Kimball.

La fase de *planeación del proyecto*, pretende establecer la definición y el alcance del proyecto de la bodega de datos, incluyendo la valoración y justificación del negocio. La fase de *definición del proyecto* es donde se establece la base relacionada con la tecnología, los datos y las aplicaciones del usuario.

La ruta de mayor importancia es la relacionada con los datos, en la cual se realiza el *modelado dimensional*, partiendo de los requerimientos obtenidos y de las necesidades de análisis de los usuarios; el *diseño físico*, el cual se enfoca en definir las estructuras físicas necesarias para soportar el modelado dimensional; y la etapa *ETL* en la cual se diseña y desarrollan procesos para extraer, transformar y cargar datos.

A lo largo de todo el ciclo de vida se debe seguir una *administración general* del proyecto la cual asegura que todas las actividades del ciclo de vida se alcancen y se sincronicen.

**DWEP (Data Warehouse Engineering Process)** está basada en el proceso unificado (en inglés: *Unified Process UP*) estándar aceptado en el ámbito científico e industrial para el desarrollo de software; entre sus principales características se encuentra que es iterativo e incremental, se basa en cuatro fases de desarrollo y siete flujos de trabajo, en la Figura 3 se presentan gráficamente la relación existente entre los flujos de trabajo y las fases tanto del UP como de DWEP, está basado en componentes, utiliza el UML (*Unified Modeling Language* - Lenguaje Unificado de Modelado) como lenguaje para modelado gráfico (19-21), es orientada a objetos, independiente de cualquier implementación específica, ya sea relacional o multidimensional y permite la representación de todas las etapas del diseño de un almacén de datos (9, 22).

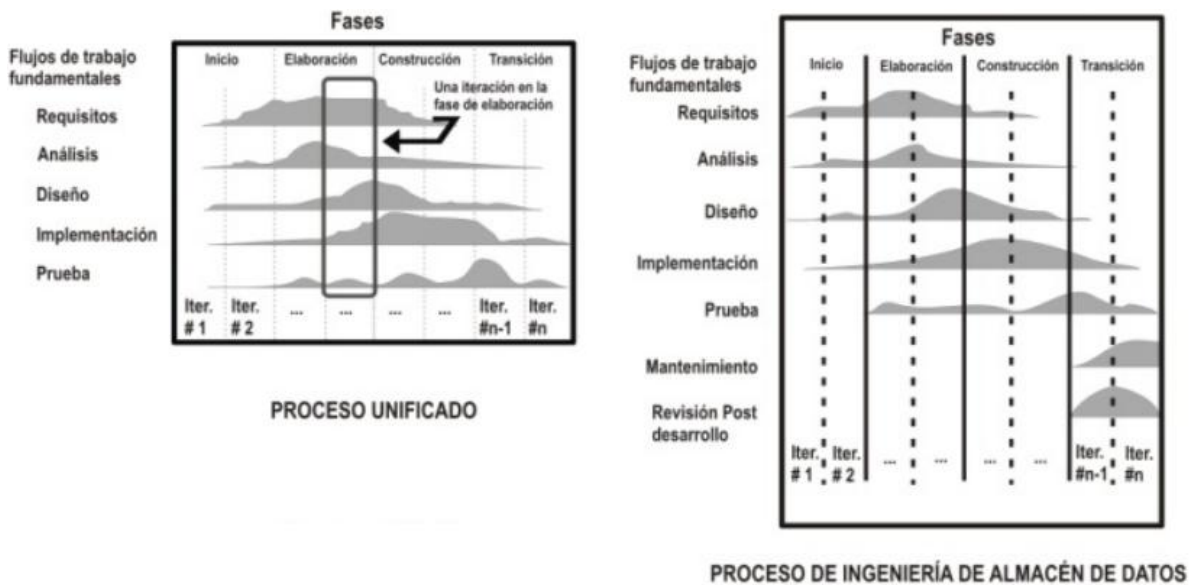


Figura 3 (a) Proceso Unificado y (b) DWEP

#### Fases de desarrollo:

**Fase de inicio:** El objetivo de esta fase es analizar el proyecto para justificar su puesta en marcha, para lograrlo se realiza una descripción general del proyecto, se detectan los riesgos críticos y se establecen la funcionalidad básica del *software* con una descripción de la arquitectura candidata.

**Fase de elaboración:** Una vez finalizada la fase de inicio, se pretende formar una arquitectura sólida para la construcción del *software*. En esta fase se busca establecer la base lógica de la aplicación con los casos de uso definitivos y los artefactos del sistema que lo componen.

**Fase de construcción:** Se inicia a partir de la línea base de arquitectura que se especificó en la fase de elaboración y su finalidad es desarrollar un producto listo para la operación inicial en el entorno del usuario final.

**Fase de transición:** Una vez que el proyecto entra en la fase de transición, el sistema ha alcanzado la capacidad operativa inicial. Esta fase busca implantar el producto en su entorno de operación.

#### Flujos de trabajo:

En términos generales para el proceso unificado y el DWEP un flujo de trabajo es un conjunto de actividades realizadas en un área determinada cuyo resultado es la construcción de artefactos (un texto, un diagrama, una página Web, código en lenguaje de programación, etc.).

**Requerimientos:** Durante este flujo de trabajo, los usuarios especifican las medidas y agregaciones más interesantes, el análisis dimensional, consultas usadas para la generación de reportes periódicos y frecuencia de la actualización de los datos. El proceso unificado sugiere el uso de casos de uso. Esto ayuda a comprender el sistema y obtener los requisitos y funciones para la solución. Además establece como deben ser las interacciones del sistema.

**Análisis:** Tiene como objetivo mejorar la estructura y los requisitos obtenidos en la etapa de requerimientos. En esta etapa se documentan los sistemas operacionales preexistentes que alimentaran el almacén de datos.

**Diseño:** Al final de este flujo de trabajo, está definida la estructura del almacén de datos. El principal resultado de este flujo de trabajo es el modelo conceptual del almacén de datos. Además las transformaciones necesarias de los datos para extraerlos del origen hacia el destino quedarán también definidas a nivel conceptual.

**Implementación:** Durante este flujo de trabajo, el almacén de datos es construido y se empiezan a recibir datos de los sistemas operaciones, se afina para un funcionamiento optimizado, entre otras tareas.

**Pruebas:** El objetivo de este flujo de trabajo es verificar que la aplicación funcione correctamente, realizar las pruebas y analizando los resultados de cada prueba. Concretamente, los efectos de las pruebas son los siguientes: planificar las pruebas necesarias, diseñar y aplicar las pruebas mediante la creación de casos de prueba y realizar las pruebas y analizar los resultados de cada prueba (23).

**Mantenimiento:** Un almacén de datos es un sistema que se retroalimenta constantemente. El objetivo de este flujo de trabajo es definir la actualización y carga de los procesos necesarios para mantener el almacén de datos. Este flujo de trabajo comienza cuando se construye la bodega de datos y es entregada a los usuarios finales, pero no tiene una fecha de finalización. Durante este trabajo, los usuarios finales pueden tener nuevas necesidades, lo que desencadena el comienzo de una nueva iteración con los requisitos de flujo de trabajo.

**Revisiones post desarrollo:** Esto no es un flujo de trabajo de las actividades de desarrollo, sino un proceso de revisión para la mejora de proyectos a futuro. Si hacemos un seguimiento del tiempo y esfuerzo invertido en cada fase es útil en la estimación de tiempo y de las necesidades para generar los requisitos para desarrollos futuros.

**Trujillo** propone una metodología de modelado basada en el proceso unificado. En esta propuesta se emplean los perfiles de UML como mecanismo para especializar este lenguaje al dominio de los almacenes de datos. Contemplando cada una de las fases de su desarrollo, se han diseñado perfiles para el modelado multidimensional del repositorio del almacén de datos y los procesos ETL (10).

**Rapid Warehousing Methodology** es una metodología propuesta por SAS Institute. Esta metodología es iterativa, y está basada en el desarrollo incremental de almacén de datos dividido en cinco fases como se puede apreciar en la figura 4.

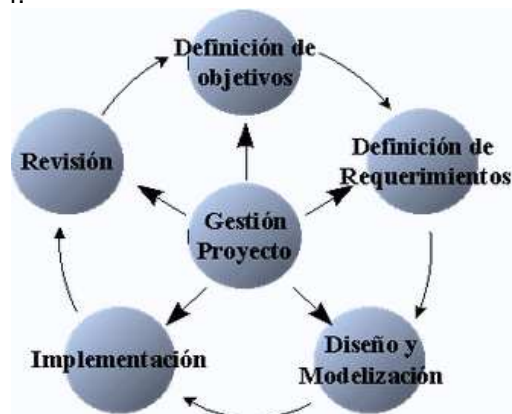


Figura 4: Fases de la metodología Rapid Warehousing Methodology.

**Definición de los objetivos:** Se definirá el equipo de proyecto, el alcance del sistema y cuáles son las funciones que el almacén de datos realizará como suministrador de información de negocio estratégica. Se definirán así mismo, los parámetros que permitan evaluar el éxito del proyecto.

**Definición de los requerimientos de información:** Se analizará las necesidades y se definirán los requerimientos de información.

**Diseño y modelización:** se identificarán las fuentes de los datos (sistema operacional, fuentes externas,...) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico de datos del almacén de datos. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades de negocio de la organización.

**Implementación:** Se realizará la extracción y carga de los datos, así como la explotación del almacén de datos para su uso por parte de los usuarios finales.

**Revisión:** Después de implantarse, se debe realizar una revisión del almacén de datos planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.



**HEFESTO** es una metodología creada por el Ing. Bernabeu Ricardo Darío, su última actualización es la versión 1.1 en abril del 2009 y disponible bajo licencia GNU FDL, se fundamenta en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. Consta de cuatro fases: análisis de requerimientos, análisis de los OLTP, modelo lógico del almacén de datos y proceso ETL. Puede ser utilizada en cualquier ciclo de vida que no requiera fases extensas de requerimientos y análisis, con el fin de entregar una implementación que cumpla con una parte de las necesidades proporcionadas por el usuario (12, 17), en la figura 5 se puede apreciar los pasos que se realizan en cada fase de la metodología.



Figura 5: Metodología HEFESTO, pasos.

**Análisis de Requerimientos:** Se identifican los requerimientos del usuario con el fin de entender los objetivos de la organización, haciendo uso de técnicas y herramientas, como la entrevista, la encuesta, el cuestionario, la observación, el diagrama de flujo y el diccionario de datos, obteniendo como resultado una serie de preguntas que se deberán analizar con el fin de establecer cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del almacén de datos. Finalmente se realizará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso.

**Análisis de los OLTP:** Tomando en cuenta el resultado obtenido en el paso anterior se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores con el objetivo de establecer las respectivas correspondencias entre el modelo conceptual y las fuentes de datos. Luego, se definirán qué campos se incluirán en cada perspectiva y finalmente, se ampliará el modelo conceptual con la información obtenida en este paso.

**Modelo lógico del Almacén de Datos:** Como tercer paso, se realizará el modelo lógico de la estructura del almacén de datos, teniendo como base el modelo conceptual. Para esto, se debe definir el tipo de representación de un almacén de datos que será utilizado, posteriormente se llevarán a cabo las acciones propias al proceso, para diseñar las tablas de dimensiones y de hechos. Por último, se realizarán las uniones pertinentes entre estas tablas.

**Procesos ETL:** Se prueban los datos a través de procesos ETL. Para realizar la compleja actividad de extraer datos de diferentes fuentes, luego integrarlos, filtrarlos y depurarlos, se podrá hacer uso de

*software* que facilita dichas tareas, por lo cual este paso se centrará solo en la generación de las sentencias SQL que contendrán los datos que serán de interés.

Esta metodología cuenta con las siguientes características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del almacén de datos.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el almacén de datos y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para almacén de datos como para Data Mart.

**CRISP-DM (CROSS Industry Standard Process for Data Mining)**, es un estándar industrial, utilizado por más de 160 empresas e instituciones de todo el mundo, que surge en respuesta a la falta de estandarización. Interrelaciona las diferentes fases del proceso entre sí, de tal manera que se consolida un proceso iterativo y recíproco. Es planteada como una metodología imparcial o neutra respecto a la herramienta que se utilice para el desarrollo de almacén de datos y Data Mining siendo su distribución libre y gratuita (24, 25).

CRISP-DM cuenta con una herramienta CASE (*Computer Aided Software Engineering*) integrada (que soporta todas las fases del proceso) para satisfacer el desarrollo de proyectos de minería de datos llamada CMIN, la cual incluye la gestión de procesos, plantillas y proyectos, y permite realizar el seguimiento de los proyectos de una forma fácil e intuitiva. CMIN permite también enlazar en tiempo de ejecución (sin necesidad de volver a compilar la herramienta) nuevos algoritmos de minería de datos que apoyen la labor de modelado (basada en un flujo de trabajo) (26).

**SEMMA** se define como el proceso de selección, exploración y modelado de grandes cantidades de datos para descubrir patrones de negocio desconocidos. Su nombre es el acrónimo correspondiente a las cinco fases básicas del proceso (muestro (*sample*), explotación (*explore*), modificación (*modify*), modelado (*model*), valoración (*assess*)) (28).

**P<sup>3</sup>TQ** (*Product, Place, Price, Time, Quantity*) está compuesta por dos modelos, el Modelo de Negocio y el Modelo de Explotación de Información. El Modelo de Negocio proporciona una guía de pasos para identificar un problema de negocio o la oportunidad del mismo. El Modelo de Explotación de Información proporciona una guía de pasos para la ejecución de los modelos de explotación de información de acuerdo al modelo identificado en Modelo del Negocio (14).

**KM-IRIS** fue elaborado por el grupo de Integración y Re-Ingeniería de Sistemas (IRIS) de la Universidad Jaume. Se crea con el objetivo de dirigir el proyecto de desarrollo de un sistema de gestión del conocimiento, consta de cinco fases: identificar, extraer, procesar, almacenar y compartir. Esta metodología pretende cubrir el ciclo completo en el desarrollo de un sistema de gestión del conocimiento. Es una metodología poco difundida y con escasa documentación (16, 27). En la tabla 1 se muestra una breve descripción de sus fases.



Tabla 1. Metodología KM-IRIS.

Fases	Objetivo	Técnica/Herramienta
Identificar	Seleccionar las bases de datos que puedan aportar la información necesaria para obtener el conocimiento.	Experiencia de los expertos.
Extraer	Ensamblar datos desde fuentes dispares, enriqueciéndolos de manera que cree información valiosa.	Herramientas ETL/ almacén de datos.
Procesar	Construir por medio de algoritmos de Minería de Datos, modelos de comportamiento	Minería de Datos
Almacenar	Validar, seleccionar y mantener los modelos de comportamiento.	Minería de Datos/ Experiencia del Ingeniero del Conocimiento
Compartir	Poner a disposición de la Organización el conocimiento descubierto	Portal del Conocimiento

### Discusión:

Elegir una u otra metodología a la hora de desarrollar un almacén de datos dependerá de las características particulares de cada metodología y del almacén de datos, pero si de las metodologías se trata se puede analizar lo siguiente:

El **Ciclo de vida Kimball** es muy amplia la manera de abordar los elementos para las etapas de desarrollo, y deja claro qué se debe hacer, pero no cómo lograrlo, lo que provoca demoras en los resultados. Esta metodología no detalla la manera en que se deben diseñar los modelos de datos ni la forma de obtener las variables para lograr la correspondencia con los datos fuentes. Además, no es contemplada como un ciclo completo de desarrollo sino que abarca solo la etapa de diseño del almacén de datos (8, 18).

**DWEP** es una metodología que contempla el ciclo completo de desarrollo de un almacén de datos, permitiendo crear todos los aspectos fundamentales en los modelos de datos (lógico conceptual y físico). La captura de requerimientos es la base para el posterior análisis y diseño del almacén de datos, en este punto se considera que se manejan muchos artefactos.

**La propuesta de Trujillo** para el desarrollo de almacenes de datos dirigidos por modelos no detalla la manera en que se deben diseñar los modelos de datos ni la forma de obtener las variables para lograr la correspondencia con los datos fuentes. Además, no es contemplada como un ciclo completo de desarrollo sino que abarca solo la etapa de diseño del almacén de datos.

**Rapid Warehousing Methodology** es una metodología iterativa que está basada en el desarrollo incremental de un almacén de datos dividido en cinco fases. Esta metodología no incluye lo relativo a técnicas de análisis de la información, por lo que con su aplicación solo se obtendría el almacén de datos y no los multianálisis de los datos para apoyar la toma de decisión.

**HEFESTO** es una metodología cuya propuesta se fundamenta en una amplia investigación, comparación de metodologías existentes y la experiencia en la elaboración de almacenes de datos. 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, más no explican cómo realizarlos. Se debe señalar como elemento negativo que su última fase es el proceso ETL por lo que no permite la obtención del modelo conceptual, lógico y físico.

**CRISP-DM, SEMMA y P<sup>3</sup>TQ** se centran fuertemente en las técnicas de explotación de información y en la tipificación de los datos, además no determinan cómo las variables vinculadas a los datos modelan el

negocio, ni cuáles son los procesos de explotación de información, ni el modelo asociado, que a partir de aplicar las técnicas al conjunto de valores de las variables, permiten obtener una solución para cada problema de inteligencia de negocio (28, 29)

**KM-IRIS** describe, formal y estructuradamente, las actividades a llevar a cabo para extraer conocimiento a partir de los datos, así como determina las herramientas y técnicas necesarias para cada fase pero es una metodología poco difundida y con escasa documentación.

### **Resultados:**

Una vez analizadas las principales características de las metodologías que se emplean en el desarrollo de almacenes de datos se percibió que no todas contemplan el ciclo completo de desarrollo, en algunas no se crean artefactos que especifiquen como implementar la solución y la documentación en otras es escasa. Además el almacén de datos se basa en el diseño de un modelo conceptual y a partir de este se obtiene un modelo lógico, muchas de estas metodologías no definen mecanismos para estructurar de manera sistemática este proceso, convirtiéndolo en una tarea compleja y artesanal.

Después de realizado un análisis exhaustivo de cada metodología se considera apropiada la integración de la metodología HEFESTO y DWEP para el desarrollo de almacenes de datos. La metodología HEFESTO podrá definir la arquitectura de los datos, es decir recolectar los requerimientos y necesidades de información del usuario y DWEP podrá elaborar el modelo conceptual, lógico y físico pues permite controlar el alcance y la agilidad de forma disciplinada.

### **Conclusiones:**

Al elegir una metodología para desarrollar un almacén de datos no se deben utilizar 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. El objetivo de cada desarrollador debe ser entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del almacén de datos y motivar a los usuarios, es por eso que se debe elegir una metodología que cumpla con estos requisitos, pues el trabajo siempre debe estar dirigido a mejorar la calidad y aceptación del mismo por los usuarios que benefician.

### **Referencias Bibliográficas:**

1. Inmon WH. Building the Data Warehouse. Cuarta ed. NuevaYork: John Wiley & Sons; 2005.
2. Fuentes L, Valdivia R. Incorporación de elementos de inteligencia de negocios en el proceso de admisión y matrícula de una Universidad Chilena. Revista Chilena de Ingeniería 2010;18(11/2010;18(3):383-94.
3. Tinoco O, Rosale PP, Salas J. Criterios de selección de metodologías de desarrollo de software. Industrial Data, Revista de Investigación 2010;13(2):70-4.
4. Jacobson I, Boock G, Rumbaugh J. El Proceso Unificado de Desarrollo de Software. La Habana: Félix Varela; 2005.
5. Pressman R. Ingeniería del Software. Un enfoque práctico. Tercera ed 2007.
6. Pressman R. SOFTWARE ENGINEERING. A practitioner's Approach. Fourth ed 2008.

7. Hernández MD. Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular. Revista Cubana de Información en Ciencia de la salud ACIMED2011;22(4).
8. Kimball R, Ross M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Second ed: John Wiley & Sons; 2002.
9. Luján S. Data Warehouse Design with UML [Doctor]. Alicante, España: Departamento de software y sistemas computacionales. Universidad de Alicante; 2005.
10. Trujillo J, Soler E, Zubcoff J, Mazón JN, Glorio O, Pardillo J. Desarrollo de almacenes de datos dirigido por modelos. 2007.
11. Espinosa R. Fases en la implantación de un sistema DW. Metodología para la construcción de un DW. El Rincón del BI. 2009 [Consultado: 01/03/2012]; Disponible en: <http://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw/>.
12. Dario BR. HEFESTO: Metodología propia para la Construcción de un Data Warehouse. Córdoba, Argentina2009.
13. Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, et al. Metodología CRISP-DM 1.0. 2007 [Consultado: 04/01/2012]; Disponible en: [http://www.dataprix.com/files/Metodologia\\_CRISP\\_DM.pdf](http://www.dataprix.com/files/Metodologia_CRISP_DM.pdf).
14. Britos P. Procesos de Explotación de Información basados en Sistemas Inteligentes [Doctor en Ciencias Informáticas]. Buenos Aires, Argentina: Universidad Nacional de la Plata; 2008.
15. García R, Lelli R, Merlino H, Cornachia L, Rodríguez D, Pytel P, et al., editors. Ingeniería de proyectos de explotación de información para PYMES. XIII Workshop de Investigadores en Ciencias de la Computación; 2011; Universidad Nacional de Rosario. Rosario. Santa Fe. Argentina.
16. Matos G, Chalmeta R, Coltell O. Metodología para la Extracción del Conocimiento Empresarial a partir de los Datos. Información Tecnológica2006;17(2):81-8.
17. Castelán L, Ocharán JO. Diseño de un Almacén de datos basado en Data Warehouse Engineering Process (DWEPE) y HEFESTO. 2010.
18. Muñoz J, Acosta L, Mendoza M. Bodega de datos para Unicauca Virtual: Una primera visión. Enlace Informático2005;4(1).
19. Luján S, Trujillo J, Song IY. A UML profile for multidimensional modeling in data warehouses. Data Knowledge Engineering2006;59(3):725-69.
20. Alaskar K, Shaik A. Object Oriented Data Modeling for Data Warehousing (An Extension of UML approach to study Hajj pilgrim's private tour as a Case Study). International Arab Journal of e-Technology2009;1(2):37-41.
21. Dhawan B, Gosain A. Extending Uml for Multidimensional Modeling in Data Warehouse. International Journal of Computer & Communication Technology2009;2:59-64.
22. Herrera EJ, editor. Data Warehouse Engineering Process (DWEPE) with U.M.L. 2.1.1. 5ta Conferencia de Euro-American Association On Telematics And Information Systems Eatis; 2010; Universidad Tecnológica de Panama.
23. Alarcón R, Basurto C, Dávila A. Infraestructura de pruebas para una plataforma de inteligencia de negocios: lecciones aprendidas de una experiencia académica. REICIS Revista Española de Innovación, Calidad e Ingeniería del Software2008;4(2):83-95.
24. Salcedo OJ, Galeano RM, Rodríguez LG. Metodología crisp para la implementación Data Warehouse. Tecnura: Tecnología y Cultura Afirmando el Conocimiento2010;14(26):35-48.
25. Shearer C. The CRISP-DM Model: The New Blueprint for DataMining". Revista de Data Warehousing2009;5(4):13-22.

*Ing. Eric Ismael Leonard Brizuela, MSc. Yudi Castro Blanco. Metodologías para desarrollar Almacén de Datos.*

26. Cobos C, Zuñiga J, Guarín J, León E, Mendoza M. CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *Revista Ingeniería e Investigación* 2010 diciembre, 2010;30(3):45-56.
27. Matilla MM, Chalmeta R. Metodología para la Implantación de un Sistema de Medición del Rendimiento Empresarial. *Información Tecnológica* 2007;18(1):119-26.
28. Azevedo A, Santos MF, editors. KDD, SEMMA AND CRISP-DM: A Parallel overview. *IADIS European Conference Data Mining*; 2008; S. M. de Infesta - Portugal.
29. Pollo F, Britos P, Pesado P, García R, editors. Metodología para especificación de requisitos en proyectos de explotación de información. *XI Workshop de Investigadores en Ciencias de la Computación*; 2009; Universidad Nacional de Rosario. Rosario. Santa Fe. Argentina.