

# Estado del arte modelo de bodegas de datos para MiPymes de ventas al detal

Ingrid Paola Solano Benítez  
Docente Investigadora  
Institución Universitaria Tecnológica de Comfacauca  
isolano@unicomfacauca.edu.co

## Resumen

En este artículo se presenta una revisión del estado del arte de algunos modelos de Bodegas de Datos con el fin de evidenciar la aplicación de éstos en diversos modelos de negocio, especialmente poder determinar hasta el momento el uso de esta tecnología en las medianas y pequeñas empresas.

**Palabras clave:** bodegas de datos, modelos de bodegas de datos, MiPymes

## I. Introducción

Las Bodegas de datos (DW, por sus siglas en inglés) se han convertido en una herramienta de soporte que le permite a las organizaciones contar con información oportuna y útil para la toma de decisiones estratégicas, [1] reduciendo el riesgo e incertidumbre, dando un alto valor a la inteligencia de negocios, incrementando la calidad y flexibilidad, garantizando exactitud y confiabilidad en la información [2,3]. Una de las claves de éxito en las organizaciones es tomar las mejores decisiones en los negocios rápidamente, puesto que tener la información correcta, completa y oportuna les ofrece una ventaja competitiva [4]. Las grandes empresas de

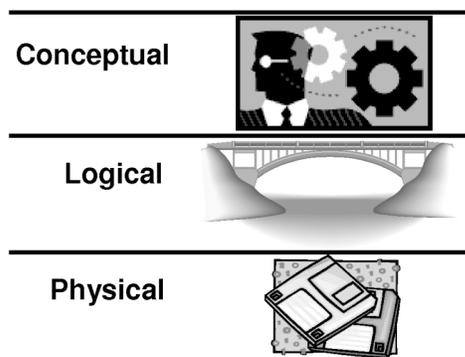
ventas al detal, desde hace varios años han encontrado en los proyectos de bodegas de datos un apoyo para la toma de decisiones estratégicas, que les permite obtener la información que necesitan para identificar y responder estratégicamente a las necesidades de su actividad comercial. [4,5,6]. El desarrollo de este tipo de proyectos se ha convertido en una buena alternativa para que estas empresas puedan ser más competitivas. Uno de los factores para que un proyecto de DW pueda tener éxito es elegir el modelo de representación dimensional conceptual [7], que permita modelar las necesidades analíticas de las organizaciones [8]. Es por eso que en este documento se presenta una revisión bibliográfica de diferentes modelos de DW con el objetivo de conocer la aplicación de esta tecnología en las organizaciones y los diferentes

modelados de representación dimensional a nivel conceptual que existen, para apreciar el nivel de aplicación en las medianas y pequeñas empresas MiPymes de ventas al detal. El contenido del artículo es el siguiente: en la sección 2 se presentan conceptos básicos, en la sección 3 se relacionan diferentes modelos de representación dimensional, en la sección 4 modelos de DW aplicados a empresas al detal, en la sección 5 modelos de DW aplicados a otras empresas y finalmente las conclusiones.

## 2. Bodegas de Datos

En la literatura existen diversas definiciones para una Bodega de Datos, de las cuales se mencionan las siguientes: Inmon, "Un DW es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales" [9]. Kimball, como una colección integrada de información corporativa diseñada para la recuperación y el análisis en apoyo a los procesos de toma de decisiones [10]

**Niveles de modelado de datos [11]:** "El modelado de datos es una técnica que registra el inventario, forma, tamaño, contenido, y las normas de elementos de datos utilizados en el ámbito de aplicación de un proceso de negocio". El resultado del modelado de datos es un tipo de mapa (el modelo) que describe los datos utilizados en un proceso. Tradicionalmente hay tres niveles de modelado de datos en bases de datos y DW: conceptual, lógico y físico.



**Figura 1.** Nivel conceptual, lógico y físico  
Fuente. *Data Warehouse Design with UML*

En la figura 1, se muestra la relación entre los tres niveles de modelado de datos. Considerando que el nivel conceptual está más cerca del dominio del usuario y el nivel físico está más cerca del equipo, el nivel lógico sirve como un puente entre lo conceptual y el nivel físico.

**Modelo conceptual de datos:** representa las entidades importantes y las relaciones entre ellas. El objetivo del modelado de datos conceptual es describir los datos de una manera que no se rige por los temas a nivel de implementación y detalle. Está más cerca del espacio del problema (el mundo real) que del espacio de la solución (la aplicación).

**Modelo de datos lógico:** el objetivo es describir los datos con tanto detalle como sea posible, sin tener en cuenta la forma en que se ejecuta físicamente en la base de datos. El modelo lógico de datos por lo general incluye: todas las entidades y relaciones entre ellos, todos los atributos y los tipos de datos correspondientes a cada entidad, la llave principal para cada entidad especificada, llaves foráneas.

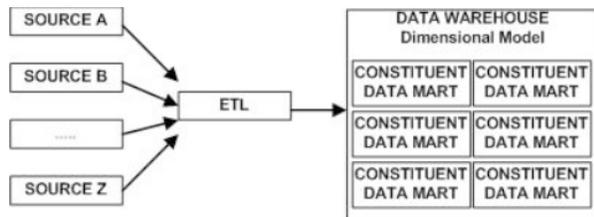
**Modelo de datos físico:** normalmente incluye la especificación, conjunto de todas las tablas y columnas siguiendo las reglas de la plataforma de implementación. Determina el diseño real de una base de datos. Es la base del código escrito para crear tablas, vistas y restricciones de integridad.

## 3. Modelo de representación dimensional

Para que un proyecto de DW pueda tener éxito es necesario elegir un modelo que se adecue a las necesidades de la organización [8], por lo que a continuación se mencionan diferentes tipos de modelos de DW que se encuentran en la literatura.

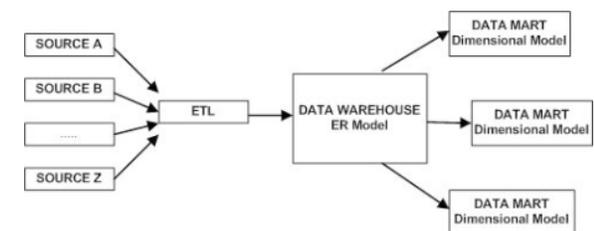
El Modelo Inmon Entidad Relación (ER) [8]. Propuesto por primera vez por Inmon [12], donde contempla una bodega de datos como un modelo de bases de datos integrado usando la técnica tradicional de modelado de bases de datos (ER modelo), como se observa en la Figura 2. Este modelo se fundamenta en que el centro de almacenamiento físico de la bodega de datos está

modelado con el modelo ER. La integración de todas las fuentes de datos operacionales se produce dentro de una bodega de datos central que es el modelo ER. El proceso de integración y consolidación de los datos de las bases de datos operacionales en la bodega de datos es conocido como el proceso de Extracción, Transformación y Carga (ETL, por sus siglas en inglés).



**Figura 2.** Modelo de bodegas de datos dimensional  
 Fuente. *Modeling strategies and alternatives for data warehousing projects. En: Communications of the ACM. Abril, 2006.*

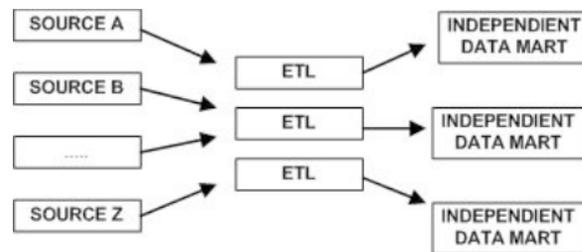
**El Modelo Dimensional de Kimball [8].** Contempla la bodega de datos como una colección de data marts dimensionales como se observa en la figura 3, este enfoque es semejante al planteado anteriormente en lo que se refiere a la utilización de fuentes de datos operacionales y el proceso ETL. La diferencia es la técnica de modelado utilizado para el modelo de la DW, este enfoque establece dimensiones usadas comúnmente (por ejemplo el calendario) diseñando en primer lugar las dimensiones asociadas, y luego se añaden las tablas de hechos correspondientes a los temas de negocio. Por lo tanto, un conjunto de modelos dimensionales es donde se crea cada tabla de hechos conectada a múltiples dimensiones y algunas de las dimensiones son compartidas por más de una tabla de hechos, además del conjunto de dimensiones creadas inicialmente, se incluyen dimensiones adicionales según sea necesario. El resultado es una bodega de datos que es una colección de modelos dimensionales de data marts (DM) entrelazados.



**Figura 3.** Modelo ER de Bodegas de Datos.

*Fuente. Modeling strategies and alternatives for data warehousing projects. En: Communications of the ACM. Abril, 2006*

Existe un tercer modelo que involucra la creación de DM independientes como se muestra en la figura 4. En este método, los DM son creados independientemente de otros data marts en la organización. Por consiguiente, múltiples sistemas de ETL son creados y mantenidos. Hay un consenso de los miembros de la comunidad de bodegas de datos, sobre el inapropiado uso de este enfoque de los DM independientes como estrategia de modelado de una bodega de datos. Hay razones para que se consideren los data marts independientes como una estrategia de calidad inferior. Dos de las principales deficiencias son la repetición innecesaria del esfuerzo de ETL y la inhabilidad de análisis y comunicación entre los DM [12].



**Figura 4.** Data Marts independientes  
 Fuente. *Modeling strategies and alternatives for data warehousing projects. En: Communications of the ACM. Abril, 2006*

Por lo anterior, se puede establecer que los dos primeros modelos, definidos en este artículo, son viables para el modelado y creación de DW. La decisión sobre el enfoque de modelado debe basarse en cual enfoque se ajusta mejor de acuerdo a sus necesidades de negocio.

En [13] se hace mención de los siguientes modelos conceptuales: el Modelo Dimensional-Fact (DF) de Golfarelli [14,15,16] que es un modelo conceptual gráfico donde la representación de la realidad con DFM (Dimensional Fact Model) se denomina esquema dimensional y consiste en un conjunto de esquemas de hechos, cuyos elementos básicos son los hechos, las dimensiones y las jerarquías. El autor parte de que la mayoría de los

sistemas de información implementados en las empresas son sistemas relacionales y en muchos casos el análisis se basa en sistemas ER, por lo tanto, propone también un método semi-automatizado para llevar a cabo la modelización conceptual de una empresa, a partir de la pre-existencia de un sistema de E/R que describe el sistema de información operacional. El DFM es independiente del modelo lógico (multidimensional o relacional).

El Modelo Multidimensional/ER (M/ER) de Sapia [17,18] es una especialización del modelo E/R llamado modelo multidimensional entidad-relación, donde para poder expresar la estructura multidimensional de los datos se definen dos conjuntos de relación especializado y un conjunto de entidad especializada, permitiendo la representación conceptual adecuada de los datos multidimensionales inherentes vistos en OLAP. Esta propuesta se hace porque los autores sostienen que el modelo E/R no es adecuado para el modelado multidimensional conceptual, porque la semántica de las principales características no puede ser adecuadamente representada.

El modelo StarER que propone Tryfona [19], combina la estructura estrella que es dominante en las bodegas de datos, con las construcciones semánticas del modelo E/R, partiendo de que el modelo E/R se ha probado durante años y ha demostrado ser lo suficientemente potente para modelar aplicaciones complejas como de espacio-temporal y multimedia.

El modelo de Hu Seman [20] propone una forma normal multidimensional (MNF), en la cual se divide la fase de proceso de diseño conceptual de bodegas de datos, en tres fases secuenciales: la primera es la definición de contexto de las medidas, la segunda la jerarquía de diseño dimensional, y la tercera la definición de las restricciones de sumarización.

Otro de los modelos multidimensionales es el de Abello (YAM2) [21], que permite el uso de las diferentes relaciones Orientadas a Objetos (OO), este modelo ha sido definido como una extensión de UML. Argumentando que el modelado

multidimensional carece de una buena representación semántica que se puede obtener utilizando el paradigma OO, que permite aplicar seis dimensiones OO (clasificación/instanciación, generalización/especialización, agregación/descomposición, llamante/llamado, derivabilidad y dinamismo) y es el modelo que más se ajusta a la realidad ya que utiliza el paradigma OO que trata de modelar los programas de una forma parecida a como se perciben en la realidad.

Recientemente, Luján, Trujillo y Song [22, 23], han propuesto un perfil del lenguaje de modelado unificado (UML), que utiliza los mecanismos de extensión estándar (estereotipos, valores etiquetados y restricciones) proporcionada por el UML, basado en dos razones principales: que el UML es un lenguaje de modelado estándar, muy conocido por los diseñadores de bases de datos y que UML se puede extender fácilmente, de modo que se pueda adaptar a un dominio específico con unas particularidades concretas, tales como el modelado multidimensional de bodegas de datos.

#### 4. Modelo de bodegas de datos aplicados a empresas al detal.

En [24], se presenta el caso de una empresa grande de Estados Unidos de ventas al detal, donde eligió el modelo de Inmon. En este artículo se mencionan los diferentes factores que se tuvo en cuenta en la decisión de elegir el modelo de Inmon (Tercera forma normal -3FN), y la comparación con el de Kimball. La bodega de datos ha permitido a la empresa fortalecer la gestión de relaciones con los clientes, las capacidades básicas y las asociaciones empresariales.

La selección del modelo a elegir, para implementar una bodega de datos, depende de muchos factores y consideraciones que permitan resolver problemas y aportar beneficios a la organización, sin embargo, cualquiera que sea el modelo elegido debe cumplir con los requisitos de la organización, ser flexible y escalable. La empresa seleccionó el modelo de Inmon porque maneja una combinación singular de sus entornos empresariales y técnicos, lo que le ha permitido

crear y proporcionar el entorno de análisis necesario para satisfacer las necesidades cambiantes del negocio.

En [25] se presenta el caso de una empresa al detal de medicamentos, que propone un modelo sencillo de ventas llamado cubo de datos y aplica a las ventas de medicamentos, el análisis de datos OLAP. Utilizan la tecnología OLAP, sobre la bodega de datos para analizar las ventas de la empresa y soporte para la toma de decisiones. En este caso, se utiliza un modelo orientado a objetos [26], que representa principalmente la relación entre las clases de objetos de ventas de la empresa y refleja las necesidades de datos de las ventas de departamentos y empleados.

En este modelo dimensional (MD), el modelo se refiere a los datos como los datos del cubo, donde la información se estructura en torno a los hechos y las dimensiones, en el cubo de datos de ventas de drogas hay tres dimensiones: el tiempo, los lugares y medicamentos. La dimensión temporal almacena los datos de ventas por mes, trimestre o año. La dimensión de lugar almacena la información sobre la ubicación (por ejemplo, ciudad o región), en la que se venden los medicamentos. La dimensión medicamentos, almacena la clasificación e información de los medicamentos.

En [27], se presenta el caso de la empresa Vitivinícola Miguel Torres Chile, donde se diseña e implementa una bodega de datos para la gestión de ventas. El modelo empleado es el Conceptual Multidimensional Data Model (CMDM) propuesto por Carpani [28], este modelo está orientado al diseño de estructuras multidimensionales, que propone estructuras de datos y un mecanismo de especificación de restricciones de integridad. Este modelo distingue entre dimensiones que identifican objetos de la realidad y relaciones dimensionales, que representan las relaciones multidimensionales existentes entre dichos objetos.

La estructura básica del CMDM está conformada por: niveles, dimensiones, relaciones dimensionales. Las dimensiones definidas son: tiempo, producto, cliente, vendedor, cantidad, la

relaciones dimensionales, se tiene una única relación dimensional, ventas. Obteniendo como resultado en la empresa disminución en los tiempos de respuesta, mayor flexibilidad ante el entorno, visión global de la información, mejor calidad y flexibilidad en el análisis del mercado, mejor servicio al cliente y mejora en la calidad de gestión de ventas para la empresa.

## 5. Modelo de bodegas de datos aplicados a otras empresas.

En [29], se presenta el caso de la construcción de una bodega de datos para la atención integral de salud, donde emplean el esquema de estrella o modelo dimensional. Se caracteriza por una tabla de hechos central de donde se desprenden tablas de dimensiones. Esta estructura proporciona un cubo de datos lógicos, con dimensiones tales como tiempo y la identificación de la ubicación, logrando proporcionar evaluaciones más detalladas de los diferentes distritos dentro del condado, permitiendo estos análisis con rapidez y eficacia.

En [30], se presenta el caso de aplicación de bodegas de datos en la gestión de la construcción en Hong Kong, donde se empleó el modelo multidimensional, adoptando el esquema estrella por su claridad, comodidad y capacidad de indexación rápida, donde se definieron tablas de hechos y tablas de dimensiones. Las tablas de hechos contienen los datos cuantitativos sobre una entidad de gestión de la construcción, las tablas de dimensiones son más pequeñas y tienen los datos descriptivos que reflejan las dimensiones de una entidad. Lo que se podrá obtener con esta DW es: consultas sobre los factores que tienen impacto en las actividades de gestión de las obras, que ayudarán a los administradores en la toma de decisiones para mejorar la administración, respuestas extremadamente rápidas a las consultas, información exacta y oportuna para la toma de decisiones en la construcción.

En [31], se presenta el caso del diseño de un bodega de datos, denominada Agrícola INARIS, un proyecto del gobierno de la India, donde se empleó el esquema estrella, en el cual la tabla de

hechos se conecta a las tablas de dimensiones que contienen datos descriptivos sobre los temas de negocio y la tabla de hechos contiene los datos cuantitativos como: el número de unidades vendidas y el precio. Esta bodega de datos está siendo utilizada para dar respuestas a las preguntas planteadas, por los legisladores en el Parlamento de la India, sobre cuestiones agrícolas (Gobierno de la India, 2003).

Como se observa en la literatura, existe gran cantidad de información relacionada con modelos de bodegas de datos aplicados a diversas organizaciones, algunas veces considerando las características y necesidades de cada una, otras aplicando el modelo que creen es mejor para dicha empresa, obteniendo en algunos casos éxito y en otros, fracaso en el momento de incursionar en este tipo de proyectos.

## 6. Conclusiones

- En la literatura se evidencia estudios acerca de la implementación de Bodegas de Datos en empresas grandes de diversas actividades económicas.
- El modelo de datos a utilizar para la implementación de una Bodega de Datos dependerá del estudio o investigación que se haga de las necesidades y características de cada empresa.
- Como se aprecia en la literatura, existe una motivación principal por definir un modelo conceptual que permita una mayor representación semántica del mundo real, pero hasta el momento no existe en la literatura un estándar de representación para el modelado conceptual de las DW, ni tampoco para las empresas MiPymes de ventas al detal.
- No existe un modelo dimensional de DW para empresas MiPymes de ventas al detal y hasta el momento no se ha realizado un estudio que permita modelar las necesidades analíticas de estas empresas. Por lo tanto, se ve la necesidad de hacer estudios o trabajos en cuanto a modelos de Bodegas de Datos en las MiPymes.

## 7. Bibliografía

- [1] G. S, Esperanza, "Data Warehouse Antecedentes, Situación Actual y Tendencias," Documentos de trabajo. Serie Estudios 2001/44: Septiembre 2001, vol 44, p. 2 – 23
- [2] P. Alan, "A Strategic Approach to Data Warehouse Development," Visible Solution Paper: Visible Systems Corp., 1997.
- [3] C. C., Leopoldo, "Guía para obtener el retorno a la inversión en proyectos de Data Warehouse," Tesis Maestría en Administración de Tecnologías de la Información, Instituto Tecnológico y de estudios Superiores de Monterrey, División de graduados en electrónica, computación, Información y comunicaciones. 2001. 103 p.
- [4] T. G. Ernesto, "Nivel de implementación de la tecnología Data Warehouse en grandes empresas," Tesis Maestría en Administración de Tecnologías de la Información, Instituto Tecnológico y de estudios Superiores de Monterrey, División de graduados en electrónica, computación, Información y comunicaciones. 2002. 87 p.
- [5] A. S. Elizabeth, "La implantación de Data Warehouse en grandes empresas: una ventaja competitiva," Tesis Maestría en Administración de Tecnologías de la Información, Instituto Tecnológico y de estudios Superiores de Monterrey, División de graduados en electrónica, computación, Información y comunicaciones. 2001. 127 p.
- [6] H. T., Visitación, A. G. Mar. H. B. Miguel y M. S. Rosana, "Bases de datos y data warehouse: herramientas estratégicas para la eficacia comercial," presentado en el IV Encuentro Iberoamericano de Finanzas y Sistemas de la Información, Febrero, 2004: Sevilla (España)
- [7] L. M. Sergio "Data Warehouse Design With UML," Tesis de doctorado, Departamento de Software y Sistemas de Computación, Universidad de Alicante, Junio de 2005, p. 291
- [8] N. Jukic, "Modeling strategies and alternatives for data warehousing projects," Communications of the ACM, Abril, 2006, vol. 49 no. 4, p. 83 – 88
- [9] I. William, Building the data warehouse New York: Wiley Computer Pub., 1996, 2nd ed
- [10] K. Ralph, The data warehouse toolkit: practical techniques for building dimensional data warehouses. New York: John Wiley & Sons, 1996
- [11] S. Lujan, "Data Warehouse Design with UML," Tesis doctorado, Departamento de Software y Sistemas de computación, Universidad de Alicante. 2005. 291 p

- [12] I. William, *Building the Data Warehouse*, 3rd Edition. Wiley, 2002.
- [13] S. Manuel, T. Juan, C. Coral, y P. Mario, "Metrics for data warehouse conceptual models understandability," *Information and Software Technology*, 2007, vol 49, p. 851 - 870
- [14] M. Golfarelli, D. Maio, S. Rizzi, "The Dimensional Fact Model: A Conceptual Model for Data Warehouses," *International Journal of Cooperative Information Systems (IJCIS)* 7 (1998) 215–247.
- [15] M. Golfarelli, y S. Rizzi, "A methodological framework for data warehouse design," en *1st International Workshop on Data Warehousing and OLAP (DOLAP '98)*, Maryland (USA), 1998, p. 3–9.
- [16] M. Golfarelli, D. Maio, y S. Rizzi, "Conceptual Design of Data Warehousing from E/R Schemes," en: *Proceedings of the 31st Hawaii International Conference on System Sciences*. IEEE Computer Society, 1998, p. 334–343
- [17] C. Sapia, "On modeling and predicting query behaviour in olap systems, *International Workshop on Design and Management of Data Warehouses*," (DMDW '99), Heidelberg (Germany), 1999, p. 1–10.
- [18] C. Sapia, M. Blaschka, G. Hoffling, y B. Dinter, "Extending the E/R model for the multidimensional paradigm," en: *1st International Workshop on Data Warehouse and Data Mining (DWDM' 98)*, Springer-Verlag, Singapore, 1998, p. 105–116.
- [19] N. Tryfona, F. Busborg, y J. Christiansen, "starER: A Conceptual Model for Data Warehouse Design," en *ACM 2nd International Workshop on Data Warehousing and OLAP (DOLAP' 99)*, ACM, Missouri (USA), 1999, p. 3–8.
- [20] B. Husemann, J. Lechtenbo, y G. Vossen, "Conceptual data warehouse design," en: *2nd, International Workshop on Design and Management of Data Warehouses (DMDW 2000)*, Stockholm (Sweeden), 2000, pp. 3–9.
- [21] A. Alberto, "YAM2: A multidimensional conceptual model," Tesis doctorado, Programa de Doctorado de Software, Departamento de Lenguajes y Sistemas Informáticos. Universidad Politécnica de Catalunya. 2002. 179 p
- [22] S. Lujan, J. Trujillo, Y. Song, "Extending UML for multidimensional modeling," en: *5th International Conference on the Unified Modeling Language (UML 2002)*, LNCS 2460, Dresden (Germany), 2002, p. 290–304.
- [23] S. Lujan, J. Trujillo, Y. Song, "A UML profile for multidimensional modeling in data warehouses," *Data & Knowledge Engineering*. 2006. vol 59, p 725–769
- [24] L. Jeff. C. Shamsul, "Best Practices in Data Warehousing to Support Business Initiatives and Needs," Presentado en la 37th Conferencia Internaonal en System Sciences, Haway 2004. p. 1–9
- [25] C. Ju, y M. Han, "Effectiveness of OLAP-based Sales Analysis in Retail Enterprises," en: *ISECS International Colloquium on Computing, Communication, Control, and Management*. May, 2008, p. 240 - 244
- [26] L. Yu, "Data Warehouse Principles and Practice," Post & Telecommunication Press, Beijing, China, 2003.
- [27] C. Jorge. y U. Luis, "Diseño e implementación de un data Warehouse para la gestión de ventas de la empresa vitivinícola Miguel Torres Chile," Tesis para optar al título de Ingeniero de Ejecución en computación e informática, Facultad de Ciencia de la Ingeniería, Escuela de Ingeniería Civil Informática, Talaca, Diciembre de 2004, p. 161
- [28] C. Fernando, "CMDM: Un Modelo Conceptual para la Especificación de Bases Multidimensionales," Tesis de maestría, Instituto de Computación, Facultad de Ingeniería, Universidad de la Republica, Agosto de 2000, p. 156
- [29] B. Donald, J. Hevner, R. James, "The Catch data warehouse: support for community health care decision-making," *Decision Support Systems*, 2003. Vol. 35 p. 367–384
- [30] K.W. Chau, Y. Cao. M. Anson, J. Zhang, "Application of data warehouse and Decision Support System in construction Management," *Automation in Construction*, 2002. vol 12, p. 213–224
- [31] N. Sree, S. Kevin, Rai. Anil, "Dimensional issues in agricultural data warehouse designs," *computers and electronics in agriculture*, 2008. vol 60. p. 263–278