



# Cómo crear su propia Bodega de Datos Espaciales

María Ximena Dueñas Reyes\*, Juan Eulises Bohórquez Carvajal\*\*  
Email: ximenadue@gmail.com, juan.bohorquez@gmail.com

## Resumen

Hoy, la información se ha convertido en el núcleo central de cualquier organización, a partir de ella se toman diversas decisiones que pueden significar mejoras en los procesos y por ende mejorar los niveles de productividad y competitividad.

Debido al crecimiento acelerado de la misma, los métodos tradicionales eran incapaces de extraer información útil de una forma rápida y eficiente, quedando gran cantidad de datos disgregados.

A partir de esto, surge la necesidad de crear herramientas que sean capaces de manejar en forma óptima los grandes volúmenes de datos, recolectarlos y agruparlos con base en las actividades del negocio, a esta herramienta se le denominó Datawarehouse, la cual sirve como apoyo para el análisis en búsqueda determinar tendencias y patrones, que logren convertir la información en conocimiento. [1]

Una información completa con un alto nivel de precisión y efectividad es sinónimo de un alto nivel de productividad, por tal motivo las empresas han fijado su atención en los datos de tipo espacial, los cuales anteriormente pasaban desapercibidos en la búsqueda de conocimiento, dando origen a los Spatial Datawarehouse.

En este artículo se guiará al lector a través de diversos pasos: extracción, selección, limpieza, integración y mantenimiento de datos mediante un ejemplo que permita la creación de una bodega de datos espaciales en el sector eléctrico.

## Palabras clave:

Spatial Data warehouse, Data warehouse, esquema multidimensional, OLAP, ROLAP, MOLAP, HOLAPç

\*Ximena Dueñas Reyes, Ingeniera de Sistemas egresada de la Fundación Universitaria Juan de Castellanos (Boyacá, Colombia) y candidata a Magíster en Ciencias de la Información y las Telecomunicaciones - Universidad Distrital "Francisco José de Caldas". Integrante del Grupo Internacional de Investigación en Informática, Comunicaciones y Gestión del Conocimiento (GICOGGE) - Universidad Distrital "Francisco José de Caldas".

\*\*Juan Eulises Bohórquez Ingeniera Catastral de la Universidad Universidad Distrital "Francisco José de Caldas". Magíster en Ingeniería de Sistemas, Universidad de los Andes.

## Abstract

Nowadays, information has become the core of any organization, from it taken various decisions that can mean improvements in processes and thus improve productivity levels and competitiveness. Due to the rapid growth of the same, traditional methods were unable to extract useful information in a fast and efficient large amount of data being fractured.

From this arises the need to create tools that are able to optimally handle large volumes of data, collected and grouped based on business activities, this tool is called Data Warehouse, which serves as support for searching analysis to identify trends and patterns, which turned the information into knowledge. [1]

Full information with a high level of accuracy and effectiveness is synon-

ymous with a high level of productivity, as such companies have focused on spatial data type, which previously went unnoticed in the pursuit of knowledge, giving rise to the Spatial Data Warehouse.

This article will guide you through several steps: collection, sorting, cleaning, data integration and maintenance through an example that allows the creation of a spatial data warehouse in the electricity Sector

### Keywords:

Spatial Data warehouse, Data warehouse, esquema multidimensional, OLAP, ROLAP, MOLAP, HOLAP

## 1. Introducción

Hoy un buen manejo de la información es sinónimo de efectividad y éxito empresarial, catalogándose como materia prima de cualquier proceso. A raíz de esto, se han creado diversas herramientas que permitan su manipulación como es el caso de las bodegas de datos, las cuales, han sido definidas como:

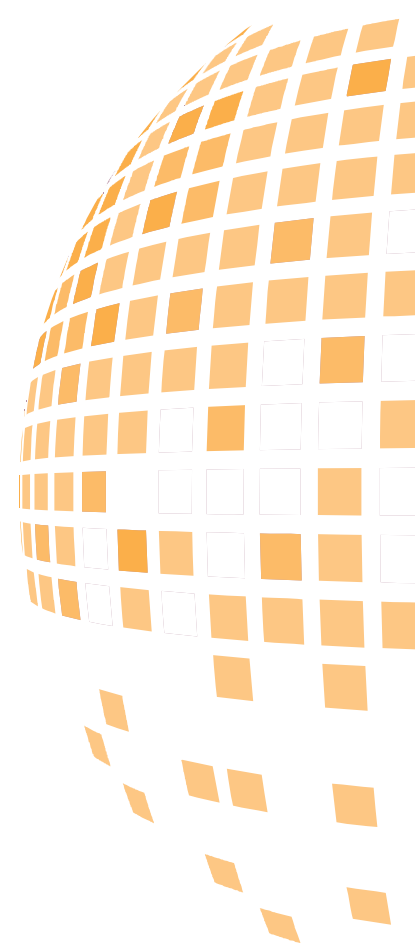
“Una colección de datos orientados al tema, integrados, no volátiles, historiadados, organizados para el apoyo de un proceso de ayuda a la decisión”.<sup>1</sup> Entre los diversos tipos de datos que se maneja dentro de una organización, la información de tipo geográfico está tomando gran relevancia, dando origen a las bodegas de datos espaciales. En este artículo se pretende abordar la importancia de las bodegas de datos espaciales que se utilizan como apoyo en las tomas de decisiones organizacionales

Este documento está estructurado de la siguiente manera: en la sección 2 se realiza una introducción a las bodegas de datos espaciales resaltando su importancia y características. En la sección 3 se describen los pasos necesarios para su construcción: en la parte 3.1 se describe el primer paso de recolección introduciendo al lector en el objetivo de la bodega. En la sección 3.2 se da una descripción del segundo paso: limpieza, destacando la importancia de la calidad de los datos. En la sección 3.3 y 3.4 se describen los procesos de integración y mantenimiento de la bodega. En la sección 4 mediante un ejemplo en el sector eléctrico se explica la creación de una bodega de datos espaciales.

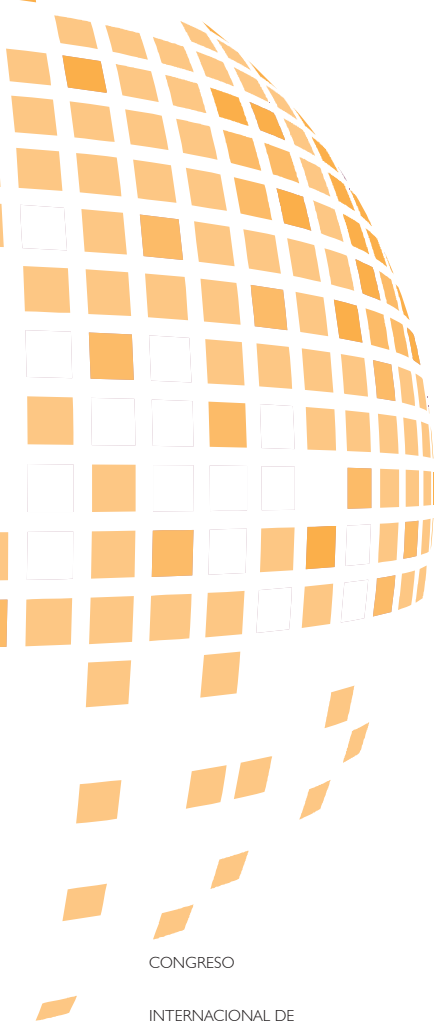
## 2. Bodega de datos espaciales

Debido a la creciente información que se maneja actualmente dentro de las empresas, ha sido necesario la creación de nuevas herramientas que brinden un manejo ágil, eficaz y eficiente a la misma, a raíz de esto surgen las bodegas de datos, las cuales, hacen parte un proceso conocido como descubrimiento de conocimiento (Knowledge Discovery from Databases, KDD), el cual, lo define Fayyad como: “Un proceso no trivial de identificar

<sup>1</sup> Bill Inmon



“ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL”



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

patrones validos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos". Las bodegas de datos buscan transformar los datos unificándolos en un formato común que facilite la navegación entre ellos para posteriormente discernir cuáles de ellos son necesarios en el análisis de los datos. [1]

Con el aumento de información también aumentan los diversos tipos de datos y uno de los que está tomando relevancia en la actualidad es el tipo de dato geográfico, el cual, se presenta a través de direcciones de clientes, proveedores, sucursales, entre otros, [2] es por esto, que surgen los Spatial Data Warehouse, los cuales, combinan las bases de datos espaciales y las tecnologías de las bodegas de datos, logrando un mayor apoyo para el manejo de datos de tipo espacial en donde la parte geográfica se toma como dimensión no como atributo. [3]

Una bodega de datos a diferencia de una base de datos transaccional, permite: [4]

- Facilitar el análisis de datos en tiempo real mediante el procesamiento analítico en línea (OLAP)
- Aplicar diversas reglas que permitan describir y analizar los datos con el fin de brindar un apoyo a la toma de decisiones organizacionales.

### Caracterización de una bodega de datos espaciales

Una bodega de datos espacial se caracteriza porque es: orientado a tema, no volátil, integrado, de tiempo variante y geografía del dato. [5]

#### **Orientado a tema**

La orientación a tema es una característica que permite clasificar la información de acuerdo con los aspectos relevantes de la organización que sean necesarios para la toma de decisiones. [6]

#### **Integrado**

Los Data Warehouse se caracterizan por tener su información de forma integrada y estructurada brindando: confiabilidad, consistencia y estandarización a los datos. [3][4]

#### **Tiempo variante**

Las bodegas de datos trabajan con información redundante al igual que con duplicidad en los datos, permitiendo un manejo histórico que facilite la identificación de patrones entre los mismos. [4]

#### **No volátil**

En las bodegas, a diferencia de las bases, de datos se manejan solo dos tipos de operaciones: carga inicial de datos y acceso a estos, logrando

un mejor análisis de la información y por ende un apoyo para la toma de decisiones que signifiquen efectividad para la empresa. [7]

## Geografía del dato

Las bodegas de datos han evolucionado constantemente con el fin de obtener una mejor y completa recolección de información, es por esto que se han podido integrar nuevos tipos de datos, como son: imágenes de satélite e información geográfica. “Los datos geográficos, presentan la información en representaciones subjetivas a través de mapas y símbolos, que representan la geografía como formas geométricas, redes, superficies, ubicaciones e imágenes, a los cuales se les asignan sus respectivos atributos que los definen y describen”<sup>2</sup>.

## 3. Construcción de la bodega de datos espaciales

La metodología que se empleará para la construcción de bodegas de datos espaciales será la misma de las bodegas de datos tradicionales integrando la parte espacial. Para llevar a cabo con éxito esta metodología el proceso debe ser iterativo e interactivo. Iterativo en la medida de que algunas salidas de las fases pueden hacer que se vuelven a pasos anteriores con el fin de obtener un mejor resultado e interactivo ya que es necesario contar con la participación activa de aquellas personas parte de la organización involucradas en los procesos. [8]

A partir de un modelo y diseño se debe seguir con una serie de pasos necesarios para llevar a cabo la construcción de la bodega.

### a. Recopilación

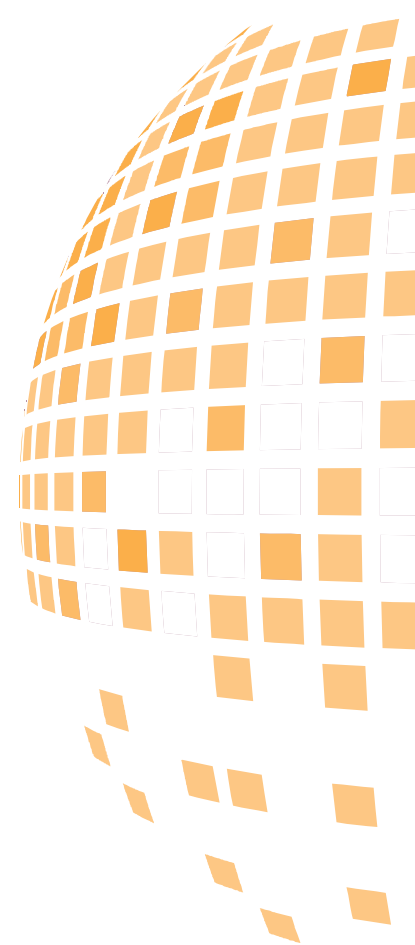
Teniendo en cuenta el volumen de información que se maneja dentro de una empresa, las organizaciones han optado por trabajar de forma independiente en sus departamentos, por ejemplo, se maneja una base de datos para los procesos de facturación, otra para el manejo de productos, logrando acuñar la frase “Divide y vencerás”, esto es muy útil en muchos procesos sin embargo no para la construcción de la bodega.

Generalmente, el almacenamiento de datos se realiza: en hojas de cálculo, documentos, textos planos o incluso no se encuentran en forma electrónica sino manual por lo que es necesario determinar cual información va acorde con la necesidad de la empresa para posteriormente realizar el proceso de extracción.

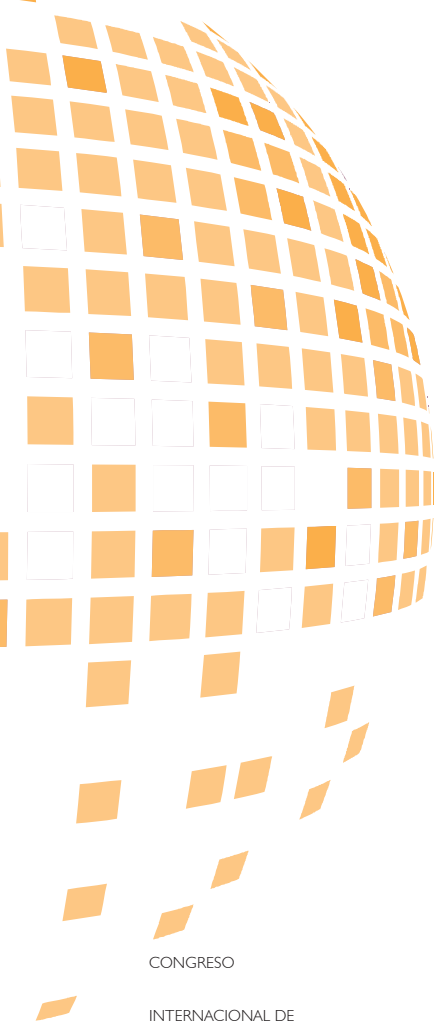
### b. Limpieza

El éxito de que los datos puedan ser útiles para ser analizados y posteriormente sirvan de apoyo en las tomas de decisiones depende en gran

<sup>2</sup> Bohórquez, Juan Eulises. Aproximación Metodológica de un Spatial Data Warehouse.



“ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL”



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

parte de su selección pero también de que estos sean íntegros y consistentes [9]. Estas características son dadas en la fase de limpieza.

Es muy común que en el almacenamiento diario de los datos, algunos de éstos presenten inconsistencias, como: un dato vacío, incompleto, entre otros. Este tipo de problemas se descubre en forma reiterada en el momento de integrar las distintas fuentes para la construcción de la bodega. Las diversas anomalías que pueden ser descubiertas, afectan la calidad de los datos dificultando los procesos de análisis y descubrimiento de conocimiento. De igual forma, se debe tener cuidado con aquellos datos faltantes, los cuales conducen a resultados con baja precisión. En algunos casos los datos faltantes pueden ser: ignorados, eliminados o reemplazados en forma manual, por un valor de acuerdo con la media o moda dependiendo del tipo de valor: numérico o nominal. [10]

### c. Integración

Como su nombre lo indica es recolectar los datos de las diversas fuentes que maneja la organización e integrarlos en la bodega. Este proceso se realiza mediante la homogenización de la información y la unificación.

Con la homogenización se busca identificar datos que correspondan al mismo objeto para posteriormente unificarlos mientras que los diversos se mantienen separados. Cuando la integración entre diversas fuentes se realiza con éxito suelen aparecer aquellos datos faltantes o inconsistentes. [11]

En una bodega la carga de datos puede ser de dos tipos: carga total e incremental. En la carga total los datos son borrados de las tablas cada vez que se ejecuta en la bodega de datos mientras que la carga incremental va añadiendo los datos haciendo procesos para cambios de datos históricos. [12]

Un almacén de datos, permite no solo integrar sino organizar la información de tal forma que se puedan aplicar herramientas que permitan describir y analizar los datos con el fin de apoyar las decisiones estratégicas de la organización.

### d. Mantenimiento de los datos

Para llevar a cabo un buen mantenimiento de los datos contenidos en la bodega, pueden realizarse dos procesos: llevar los datos a las bodegas minutos después de que fueron actualizadas las fuentes o acumular los datos en una bodega de datos transitoria en la cual sean almacenados y periódicamente llevados a la bodega. [13] De igual forma, se debe recordar que es necesario refrescar la bodega en forma periódica dependiendo de los datos o cuando sea solicitado por el administrador.

A partir de los pasos anteriores se comienza analizar la información en forma geográfica. Se debe recordar que los datos de tipo espacial pueden considerarse como: estáticos o versátiles. Los datos espaciales que se consideran estáticos son aquellos en donde la probabilidad de cambio es

mínimo a diferencia de los versátiles que tienen una alta probabilidad de variación. [4]

En este punto, todos los datos que puedan ser georeferenciados, como por ejemplo: direcciones de clientes, puntos de venta, entre otros, son referenciados mediante herramientas SIG, las cuales, son necesarias para que la bodega de datos sirva no solo como un repositorio de almacenamiento, sino como una herramienta que facilite el análisis y la visualización de los resultados obtenidos.[3]

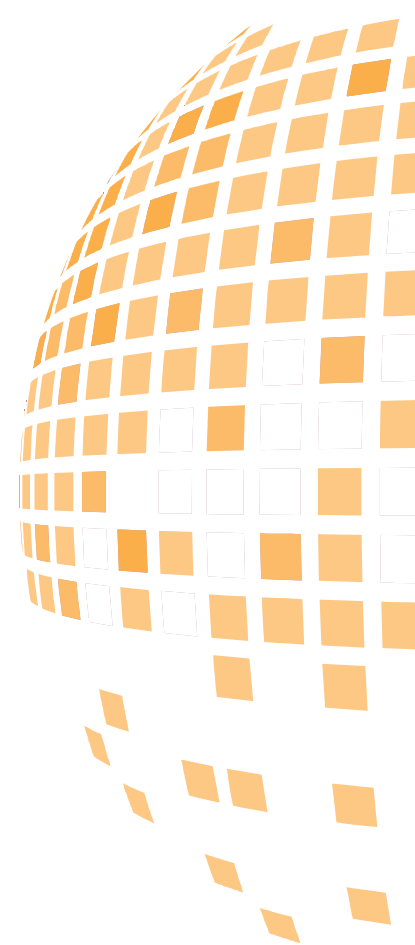
Una vez que la información ya se encuentra dentro de la bodega de datos transitoria se debe seleccionar las herramientas que se van a utilizar para el análisis de la misma, como es el caso de OLAP. El procesamiento analítico en línea (OLAP) es una herramienta que permite obtener información sobre el comportamiento del negocio a partir de los datos almacenados dentro de la bodega en forma multidimensional. Los conceptos multidimensionales que se manejan en OLAP incluyen: dimensiones, atributos de medidas, miembros, hechos y cubos de datos. Las dimensiones son tablas desnormalizadas que se relacionan al menos con una medida, la cual, se encuentra en la tabla de hechos. [11]

Las medidas son los datos o hechos que relacionan los temas de negocio reflejando un desempeño de éste. Éstas pueden ser: aditivas, las cuales se combinan en cualquier dimensión, semi-aditivas las que se combinan en ciertas dimensiones y no aditivas las que no se combinan. [15]

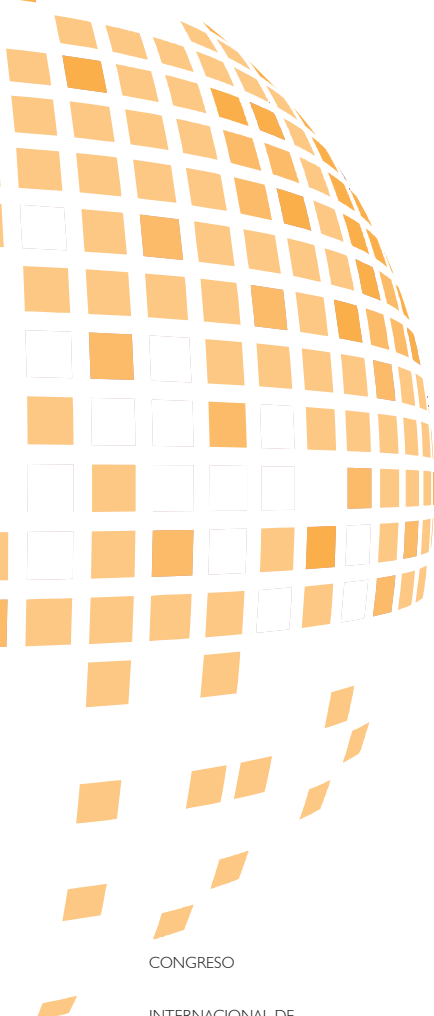
Los cubos de datos están compuestos por un conjunto de medidas agregadas de acuerdo al conjunto de dimensiones. Se pueden agregar múltiples dimensiones para realizar cruces que permitan extraer, en forma rápida y eficiente, la información que se requiere. Con el uso de cubos OLAP, se pueden realizar análisis de tipo multidimensional de un modo ágil y eficiente. [14]

El modelamiento de las bases de datos multidimensionales pueden ser en forma: estrella, copo de nieve o constelación. El modelo estrella maneja una tabla central, la cual es la tabla de hechos y en sus extremos un conjunto de tablas, denominadas dimensiones. [15] El esquema copo de nieve se deriva del esquema estrella en el cual, las dimensiones se normalizan en múltiples tablas. En cada dimensión se almacenan jerarquías de atributos o simplemente estos se separan en otra entidad para así mejorar el desempeño y utilización del espacio. [16]

Este tipo de modelos puede ser implementado mediante ROLAP, MOLAP o su híbrido HOLAP. [17] ROLAP es una forma de almacenamiento en donde toda la información del cubo: datos, agregaciones, entre otros, son almacenados en una base de datos relacional a diferencia de MOLAP en donde la información del cubo se almacena en una estructura multidimensional. HOLAP es un híbrido en el cual, las agregaciones se almacenan en forma multidimensional mientras que los datos a nivel detalle se recopilan en una base de datos relacional. [18] En la siguiente tabla se verán algunas ventajas y desventajas de las formas de almacenamiento.



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

Tabla 1. Ventajas y desventajas de las formas de almacenamiento [14]

	VENTAJAS	DESVENTAJAS
ROLAP	Los espacios de almacenamiento son menores.	Las consultas tienen un mayor tiempo de respuesta
MOLAP	Mejores tiempos de respuesta.	Espacio de almacenamiento de datos mayor
HOLAP	Tiempos de respuesta mejores para información sumariada	Los datos en las bases de datos relacionales son de gran volumen.

Una bodega de tipo espacial debe manejar obligatoriamente dimensiones de tiempo y espacio. La dimensión de tiempo es imprescindible en cualquier organización para conocer: el día de venta de un producto, compra a un proveedor, ingreso de un nuevo cliente, pagos de quincena, en fin, es obligatorio tener un control sobre el mismo.

Para la creación de esta dimensión es importante analizar la temporabilidad de los datos, la precisión que debe manejar y su flexibilidad, logrando realizar un modelo de dimensión que se adecue a la necesidad del negocio. [18]

Antes de entender la importancia de la dimensión espacial, se debe recordar que los datos de este tipo reflejan la primera ley de la geografía enunciada por Tobler (1979) en donde afirma que en el análisis geográfico todo está relacionado con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas. [19][20]

Los objetos de tipo espacial manejan un componente descriptivo, por ejemplo: nombres, datos de población, cantidades de ventas y un componente espacial en donde se incluye su geometría, la cual se representa por: puntos, líneas y polígonos, como se aprecia en la figura 1. Los puntos representan espacialmente una geometría de cero dimensiones, el cual denota una localización en el espacio, las líneas representan un objeto de una dimensión, denotando un conjunto conectado de puntos mientras que los polígonos son formados por composición de líneas ordenadas. [21]

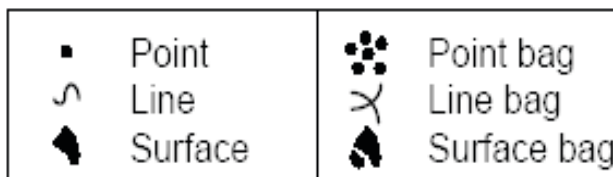


Figura 1. Tipos Datos Espaciales. [3]

Las relaciones entre los objetos de tipo tradicionales son aquellas explícitas en la entrada de los datos como: las relaciones aritméticas, de orden, subclases, miembros. A diferencia de las relaciones en objetos con características espaciales, las cuales son implícitas en la entrada de los

datos, como se aprecia en la tabla 2. [21], mostrando las relaciones entre los mismos.

Tabla 2. Relaciones entre objetos con y sin características espaciales. [21]

Non-spatial Relationship (Explicit)	Spatial Relationship (Often Implicit)
Arithmetic	Set-oriented: union, intersection, membership, ...
Ordering	Topological: meet, within, overlap, ...
Is_instance_of	Directional: North, NE, left, above, behind, ...
Subclass_of	Metric: e.g., distance, area, perimeter, ...
Part_of	Dynamic: update, create, destroy, ...
Membership_of	Shape-based and visibility

A partir de herramientas analíticas se busca encontrar relaciones entre objetos de tipo espacial y no espacial a través de relaciones: topológicas, de orientación espacial y distancia de información. [22].

Para poder entender la importancia del análisis de datos espacial, imagine ser un analista de una empresa de energía en la cual se han detectado pérdidas de tipo no técnico dentro de la ciudad. Encontrar la relación entre el estrato socio-económico del usuario y la pérdida genera patrones no esperados con el análisis tradicional. Uno de las formas utilizadas para la creación de Spatial Data Warehouses, es propuesta por Elzbieta Malinowski y Esteban Zimányi, la cual se basa en tres principios básicos:

- Aproximación de análisis
- Aproximación de fuentes
- Aproximación de análisis y fuentes

Aproximación de Análisis. En esta fase los usuarios deben manejar conceptos de espacialidad, identificando objetos de tipo espacial y relaciones topológicas entre los mismos, logrando obtener un mejor provecho en la creación del modelo multidimensional. En esta fase se identifican los objetivos de la organización, los requerimientos de los procesos y la disponibilidad de datos y fuentes que permitan generar con éxito una bodega de datos de tipo espacial, como se aprecia en la figura 2.

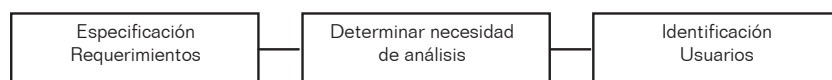
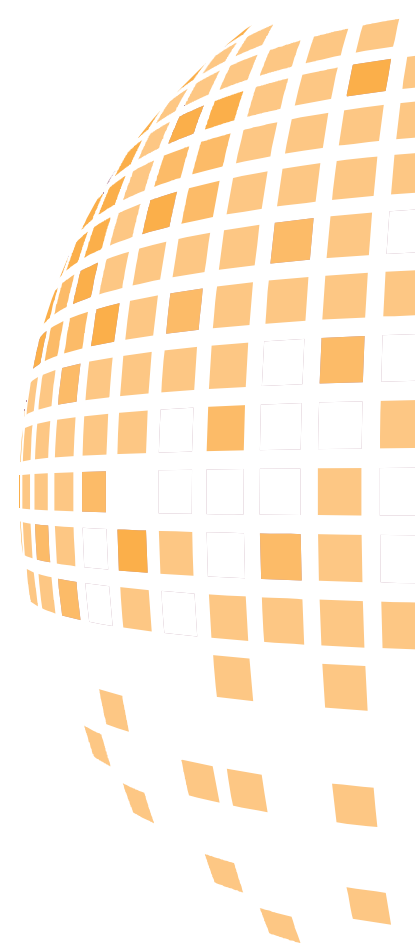


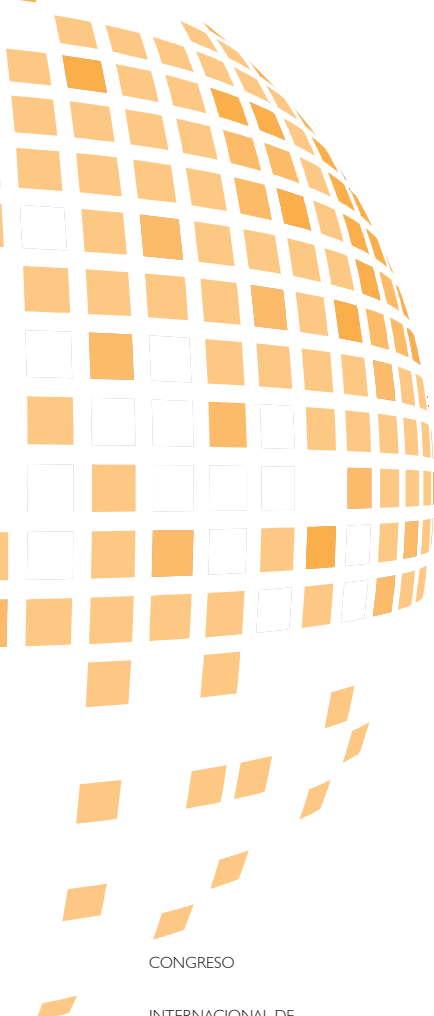
Figura 2. Aproximación de análisis [3]

A continuación se crea un esquema inicial de modelo multidimensional. En esta parte el diseñador debe manejar conceptos de tipo espacial para determinar qué atributos o medidas y entidades se catalogarán espaciales,



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"





CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

jerarquías espaciales y determinación de dimensiones que manejen relaciones topológicas.

Aproximación de fuentes. En esta fase se identifican las fuentes internas y externas de los datos que serán integrados en la bodega, con el fin de descubrir elementos que sean necesarios para el esquema multidimensional. Los procesos de derivación se originan de los objetivos del negocio y de conceptos espaciales que permita la identificación de elementos del esquema multidimensional, como se aprecia en la figura 3.

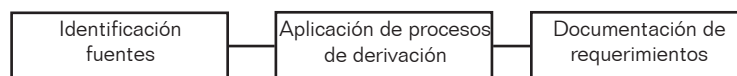


Figura 3. Aproximación de fuentes [3]

En cada una de las fases se desarrolla un modelo conceptual en donde se realiza un esquema inicial sin tener en cuenta la parte espacial, posteriormente al análisis de requerimientos se determina la disponibilidad de datos a través de las fuentes internas y externas y por último se adicionan elementos y relaciones espaciales que nos permita diseñar un completo y eficiente esquema multidimensional, como se aprecia en la figura 4.



Figura 4. Diseño Conceptual [21]

## 4. Ejemplo práctico

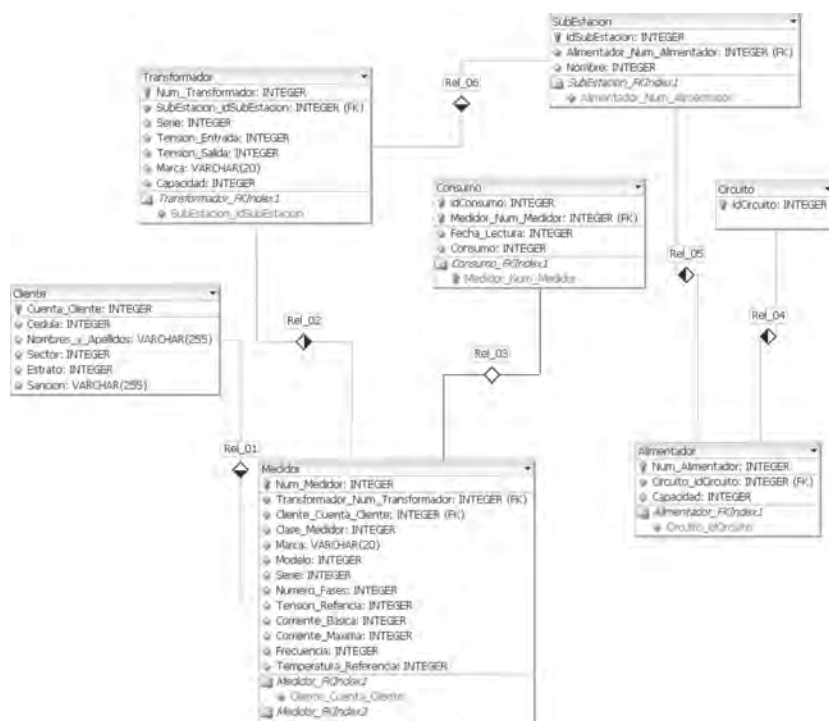
Las bodegas de datos se hacen más necesarias a medida que aumenta la información dentro de las organizaciones. A continuación se tratará un caso práctico en el sector eléctrico que guiará el resto del artículo.

En el sector eléctrico, las pérdidas de energía para las empresas distribuidoras y comercializadoras han significado un freno en la viabilidad y productividad de las empresas. Este tipo de pérdidas pueden clasificarse en: pérdidas técnicas y no técnicas. Estas últimas son consecuencia de: hurtos, fallas en la administración, conexiones ilegales, fraudes por parte de los usuarios en general.

A partir de lo anterior, lo primero que debe llegar a la mente es conocer más acerca del negocio y lo que se quiere lograr. Imagine que desea crear un modelo de predicción de pérdidas no técnicas en el sector eléctrico, algunos de los datos primordiales que necesitaría son:

- Identificación de los puntos que generaron pérdidas de tipo no técnico para la empresa en años anteriores.
- Caracterización de la zona socioeconómica donde han sido identificados los puntos de pérdida.
- Cantidad de energía perdida.
- Clase de servicio prestado ha dicho punto.
- Tipo de instalación

Però identificar datos a simple vista no es suficiente, es necesario conocer como es el proceso de distribución de energía dentro de la ciudad para así comprender mejor el problema y encontrar posibles nuevos datos que sean relevantes para lo que se quiere lograr. En la figura 5 se aprecia el modelo operacional del sector eléctrico en su proceso de distribución desde la subestación hasta el usuario.

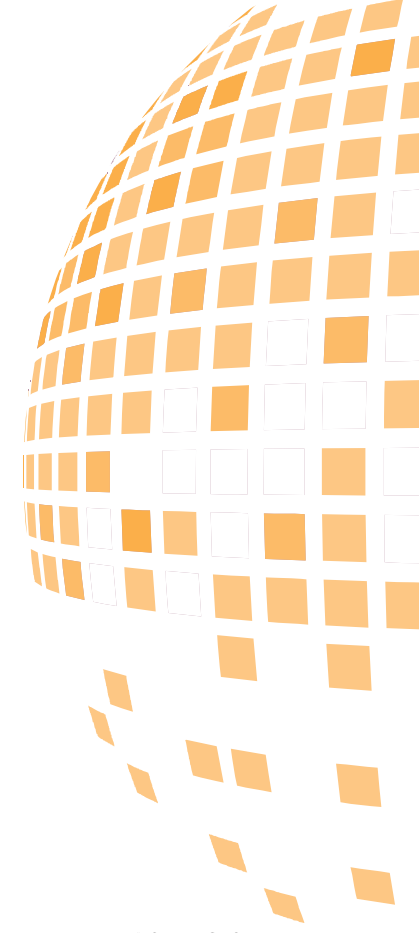


Fuente: Autores investigadores

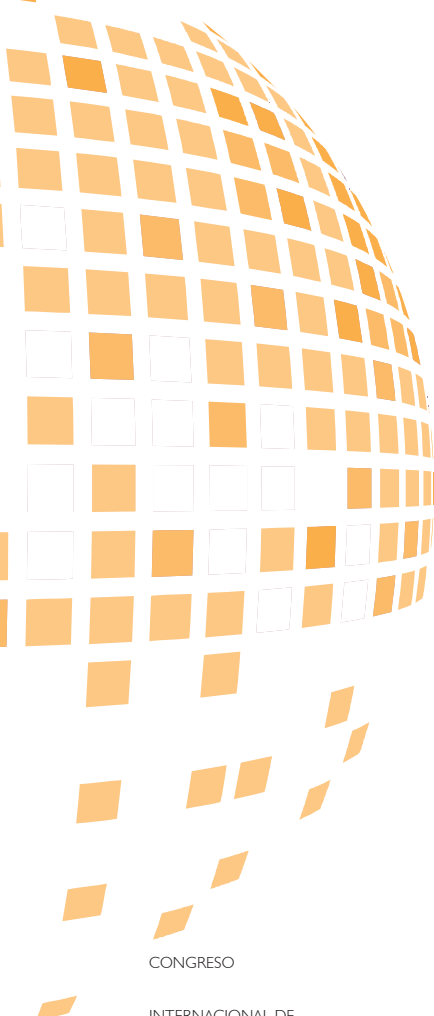
Figura 5. Modelo Operacional de distribución en el sector eléctrico.

Después de conocer el fenómeno y haber hablado con aquellas personas que hacen parte del proceso de distribución ya se deben entender: las reglas de negocio, la temporabilidad de algunos procesos, el alcance de los mismos, entre otros, necesarios para comenzar con el modelo multidimensional.

Algunas personas podrían preguntarse si es necesario contar con el modelo operativo para la construcción de la bodega de datos, la respuesta es no, sin embargo este modelo permite entender más fácil el fenómeno y abarcar el modelo lógico de los sistemas operacionales. [4]



"ESTRATEGIAS DE  
LA TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

**COMTEL 2009**

Para comenzar con el modelo multidimensional es importante determinar cuál va a ser la tabla de hecho. En este caso, se debe recordar que ésta contiene los valores que se quieren analizar, en este caso: detectar futuras pérdidas de tipo no técnico, por lo tanto la tabla de hecho sería el consumo de los usuarios. Para la selección de las medidas, se debe tener en cuenta lo que se busca a partir del modelo, en este caso es importante conocer la energía consumida por el usuario. Para determinar las dimensiones, recuerde que cuenta con dos dimensiones obligatorias: espacio y tiempo, las demás son aquellas que se relacionan con al menos una medida contenida en la tabla de hechos, para este modelo la dimensión que se relaciona con energía consumida es facturación. Se debe recordar que las tablas de dimensiones se unen a la tabla de hechos mediante referencias de una llave foránea

A partir de lo anterior se podría clasificar algunas de las dimensiones, medidas y tabla de hechos de la siguiente forma:

Tabla 3. Dimensiones, Tabla de hechos y medidas en el sector eléctrico

Dimensiones	Facturación Cliente Ubicación Tiempo
Tabla de hechos	Consumo
Medidas	Energía consumida Energía facturada

Fuente: Autores investigadores

Como se había mencionado anteriormente, la dimensión tiempo se modela de acuerdo con la temporabilidad que se requiera en la organización. Para este caso, quedaría de la siguiente forma:



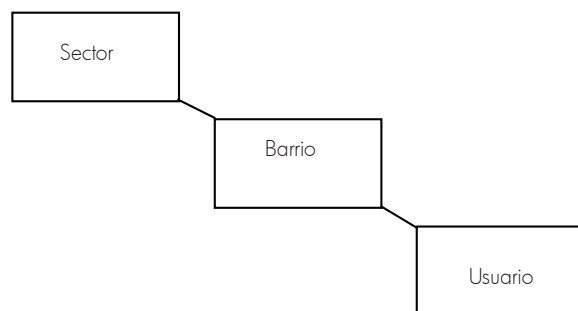
Fuente: Autores investigadores

Figura 6. Dimensión tiempo

Ahora se dará un vistazo al modelamiento de la dimensión espacial. En esta parte se debe tener presente las relaciones topológicas mencionadas anteriormente.

Teniendo en cuenta que el objetivo del modelamiento es predecir pérdidas de tipo no técnico en el sector eléctrico, se debe tener presente que

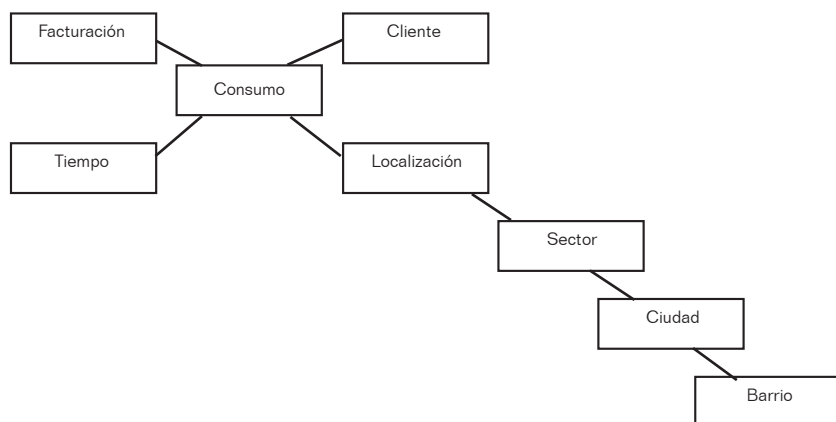
éstas pueden ser identificadas en un sector o barrio específico, por lo tanto el modelamiento preliminar de la dimensión espacial sería de la siguiente forma:



Fuente: Autores investigadores

Figura 7. Dimensión espacial inicial

A continuación se dará un vistazo al modelo preliminar identificando algunas de las dimensiones que hacen parte del esquema multidimensional.

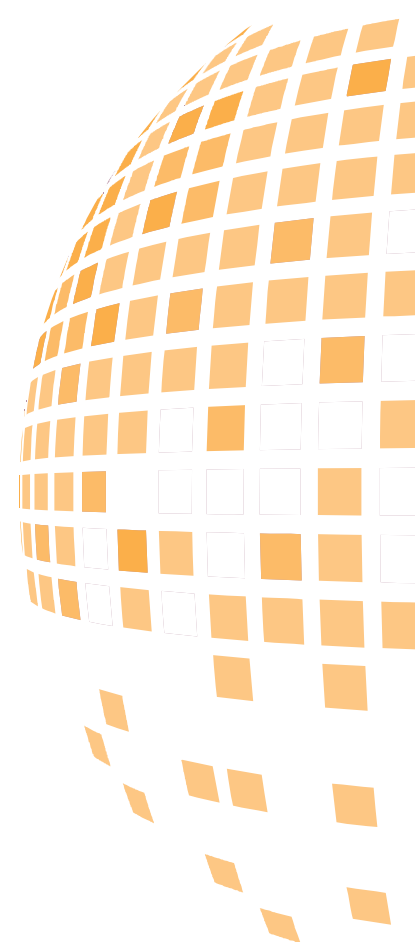


Fuente: Autores investigadores

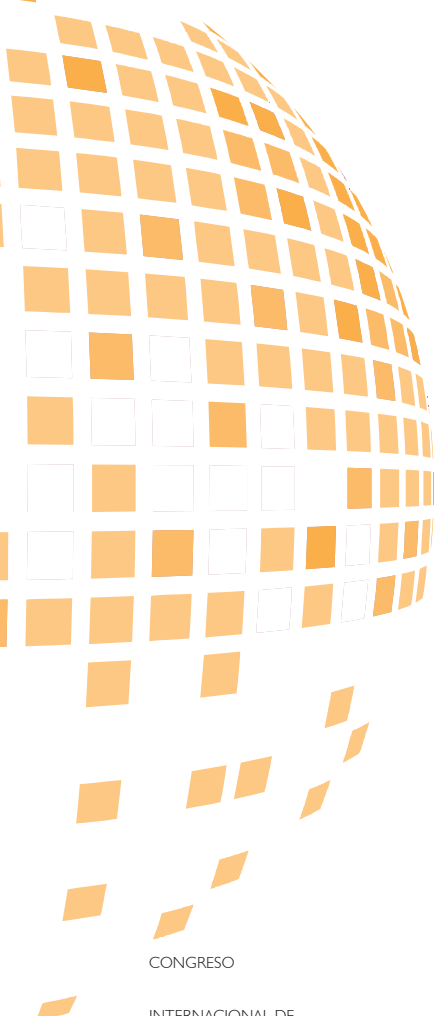
Figura 8. Modelo multidimensional inicial

Luego de contar con un modelo inicial se comienza a identificar los elementos de tipo espacial al igual que las relaciones entre los mismos. En la dimensión de localización se ve que se desprenden de ella otros elementos, como ciudad, sector y barrio identificando una jerarquía espacial.

En esta parte del modelamiento podría preguntarse, ¿cómo se relaciona en forma topológica sector con barrio? La respuesta es una relación de tipo Contains/inside, es decir, es una relación de tipo simétrico, en donde: a contiene b si y solamente si b está dentro de a. En este caso el sector contiene un barrio ya que éste está dentro del sector. En la siguiente figura se puede apreciar una jerarquía de tipo espacial al igual que las relaciones topológicas entre ella.



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

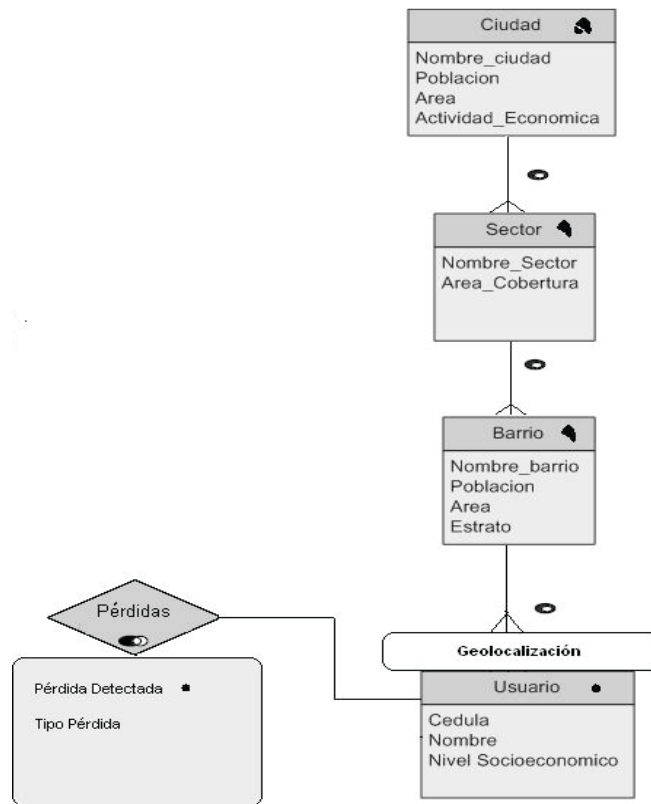
COMTEL 2009



Fuente: Autores investigadores

Figura 9. Jerarquía Espacial

Por último, se integraran los elementos espaciales y no espaciales al igual que las relaciones que existan entre ellos. Recuerde que el esquema debe ser funcional y operacional, el cual, debe ser revisado por los miembros de la organización. En la siguiente figura, se aprecia una parte del esquema multidimensional.



Fuente: Autores investigadores

Figura 10. Esquema multidimensional

## 5. Conclusiones

La inteligencia de negocio se ha definido como la transformación de datos en conocimiento brindando soporte a la toma de decisiones a nivel

estratégico y táctico en el momento y lugar oportuno generando una ventaja competitiva y de efectividad dentro de la empresa.

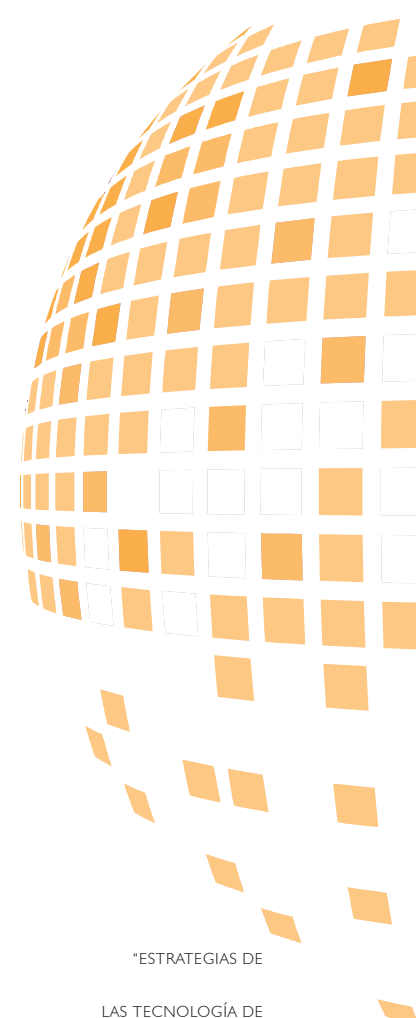
Las bodegas de datos son herramientas útiles dentro de las organizaciones ya que provee grandes ventajas en el manejo de la información, como lo son: la rapidez, confiabilidad y efectividad de la misma.

La construcción de estas herramientas dentro de las empresas ofrece grandes beneficios, como:

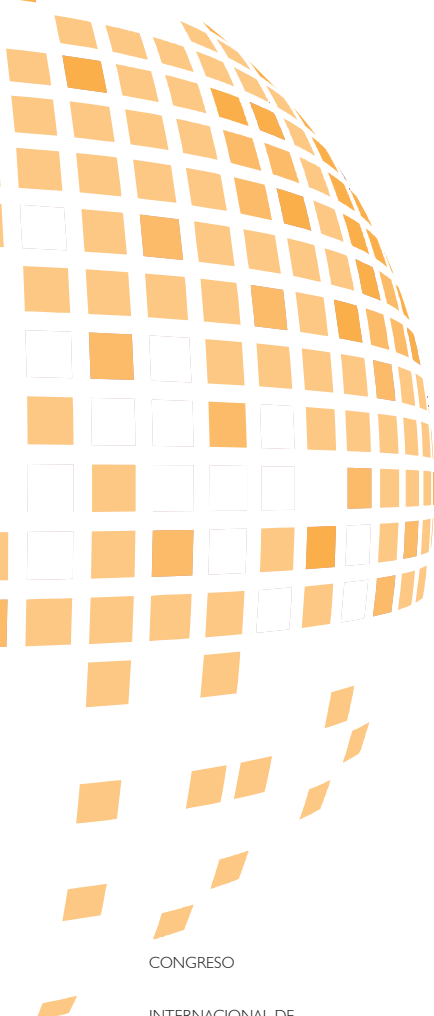
- El acceso a la múltiple información que se maneja en la organización es mucho más ágil debido a que todos los datos se encuentran consolidados e integrados en la bodega.
- Mantenimiento de la información histórica de los datos logrando así un mejor análisis en el momento de toma decisiones.
- Los tiempos de respuesta en el momento de solicitar la información es mucho más ágil y eficiente debido a que se tiene un mejor manejo de los datos gracias a la bodega.

## 6. Referencias

1. J. Hernández, M. J. Ramírez, C. Ferri,(2007) **Introducción a la minería de datos**. Mexico D.F. Pearson Prentice Hall, ch 2
2. J. Reinschmidt, and A. Franchoise (2000), "Business Intelligence Certification Guide", IBM Redbooks.[Online]. Available: <http://www.redbooks.ibm.com/abstracts/sg245747.html>
3. E. Malinowski, and E. Zimanyi, "Advanced Datawarehouse Design. From conventional to spatial and temporal applications". Springer, 2008, ch1.
4. J. Bohórquez, "Aproximación Metodológica de un Spatial Data Warehouse", ESRI.[Online]. Available: [http://proceedings.esri.com/library/userconf/latinproc00/colombia/spatial\\_data.pdf](http://proceedings.esri.com/library/userconf/latinproc00/colombia/spatial_data.pdf)
5. W. H. Inmon, "Building Data Warehouse ", 4 nd ed. Wiley.2005, ch 3.
6. C. Imhoff, N. Galemme, and J. G. Geiger. "Mastering Data Warehouse Design: Relational and Dimensional Techniques". Wiley. 2003, ch 2.
7. M. Bramer, "Principles of Data Mining", Verlag London: Springer, 2007, ch 1
8. S. Harms, D. Kozzal, and M. Culver (2004), "[Building knowledge dis-](#)



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

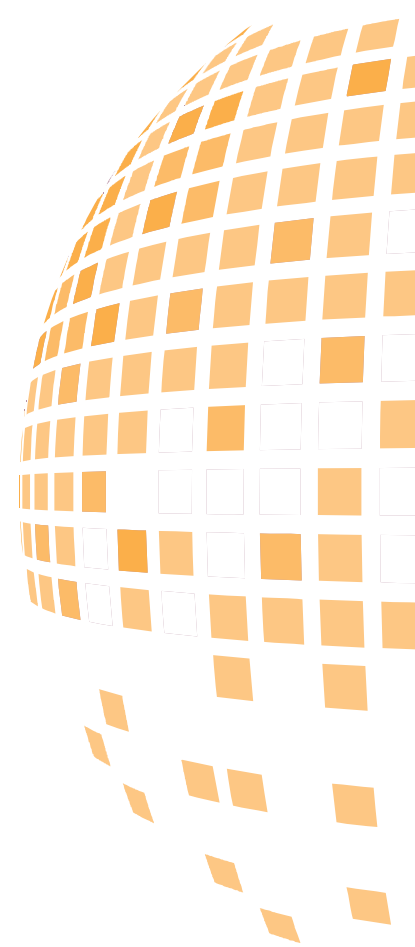
TELECOMUNICACIONES

COMTEL 2009

[covery into a geo-spatial decision support system](http://www.dgrc.org/dgo2004/disc/demos/mondemos/kozal.pdf)". [Online]. Available: <http://www.dgrc.org/dgo2004/disc/demos/mondemos/kozal.pdf>

9. R. Fang yan, Z. Long, Y. Xiu Lan, L. Ying, and C. Ying (2003), "Spatial hierarchy and OLAP-favored search in spatial data warehouse", [Online]. Available: <http://www.cis.drexel.edu/faculty/song/dolap/dolap03/paper/p48-rao.pdf>
10. D. Olson David and D. Denle. "Técnicas avanzadas de minería de datos", Berlín: Springer, 2008, ch2.
11. J. Curto Josep (2007), "Information Management. Diseño de un data Warehouse. Slowly changing dimensions", [Online]. Available: <http://informationmanagement.wordpress.com/2007/12/26/disenio-de-un-data-warehouse-slowly-changing-dimensions/>
12. Kan, Yao and Concordia University (2003), "Design issues in data warehousing: A case study 18". [Online]. Available: [http://www.lib.umi.com/dissertations/preview\\_page/MQ78001/18](http://www.lib.umi.com/dissertations/preview_page/MQ78001/18)
13. M. A. Ibarra (2006), Procesamiento Analítico en Línea. [Online]. Available: [http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/IBM\\_DB2\\_OLAP\\_SERVER\\_2006.pdf](http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/IBM_DB2_OLAP_SERVER_2006.pdf)
14. D. Awalt, and B. Lawton (2000), "Data Warehousing: Back to Basics". [Online]. Available: <http://www.winnetmag.com/SQLServer/Article/ArticleID/7833/7833.html>
15. R. Fang yan, Z. Long, Y. Xiu Lan, L. Ying and C. Ying (2003), "Spatial hierarchy and OLAP-favored search in spatial data warehouse". [Online]. Available: <http://www.cis.drexel.edu/faculty/song/dolap/dolap03/paper/p48-rao.pdf>
16. E. Soler, V. Stefanov, J. N. Mazon, J. Trujillo, E. Fernandez and M. Piattini (2008), "Towards Comprehensive Requirement Analysis for Data Warehouses: Considering Security Requirements". Third International Conference on Availability, Reliability and Security. Available: <http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/4529302/4529303/04529327.pdf?amumber=4529327>
17. S. Shekhar, P. Zhang, Y. Hung and R. R. Vatsavai. "Trends in Spatial Data mining". [Online]. Available: [http://www.spatial.cs.umn.edu/paper\\_ps/dmchap.pdf](http://www.spatial.cs.umn.edu/paper_ps/dmchap.pdf)
18. R. Matias, and J. Moura-Pires." Spatial On-Line Analytical Processing (SOLAP): A Tool the to Analyze the Emission of Pollutants in Industrial Installations". [Online]. Available: <http://centria.fct.unl.pt/~jmp/page11/page14/files/EPIA05-RM-JMP.pdf>

19. J. M. Chica Olmo, and T. Luque Martínez, Aplicación de la teoría de variables regionalizadas en la investigación de marketing. Vol. 1, pp.10-15. ISSN 1019-6838
20. W. Hsu, L. Mong and J. Wang. Temporal and spatio-temporal data mining. Hershey: IGI publishing. 2008. Ch. 1
21. E. Malinowski and E. Zimanyi (2006), "Representing Spatiality in a Conceptual Multidimensional Model". [Online]. Available: [http://www.cs.aau.dk/~simas/dat5\\_06/presentations/Presentation\\_Nermin.pdf](http://www.cs.aau.dk/~simas/dat5_06/presentations/Presentation_Nermin.pdf)
22. M. Ester, H. P. Kriegel and J. Sander (2001), "Algorithms and Applications for Spatial Data Mining". [Online]. Available: <http://www.dbs.informatik.unimuenchen.de/Publikationen/Papers/Chapter7.revised.pdf>
23. F. B. Zhan (2002), "A fuzzy set model of approximate linguistic terms in descriptions of binary topological relations between simple regions". [Online]. Available: <http://portal.acm.org/citation.cfm?id=644740>



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"