



## Introducción a la unidad

Las necesidades de información del ser humano y de las organizaciones evolucionan con el tiempo. Con el surgimiento de las bases de datos se pudo resolver el problema de registrar información útil para su futura recuperación; fueron posibles cálculos más rápidos y se le dio confiabilidad al procesamiento de grandes cantidades de transacciones. Hoy en día, podemos decir que estas necesidades están prácticamente cubiertas.

De esta forma, las necesidades de información se han vuelto más complejas. Las cantidades de información que guardan las bases de datos empresariales son enormes y su análisis se ha vuelto complicado. La primera necesidad ya no consiste en el almacenamiento y procesamiento, sino en el análisis. Por tanto, las nuevas tecnologías de bases de datos se están orientando al análisis automático de información para brindar soporte a las decisiones y generar conocimiento.

Antes, las necesidades de información eran del estilo: ¿cuánto fue vendido en Nuevo León en septiembre del año pasado?, cuya solución era aplicada con una consulta simple de SQL (*query*). Hoy en día, éstas son: ¿cuántas unidades se vendieron en Nuevo León con respecto a Guadalajara el año pasado, con respecto a este año? La solución ya no es un solo *query* de SQL; o por ejemplo, con base en las ventas de los últimos cinco años, ¿cuántas podrían ser las unidades vendidas este año? Esta situación obliga a usar métodos que permitan predecir el comportamiento de los datos.

Por otro lado, no es raro encontrar organizaciones que utilicen las bases de datos como centros de almacenaje del acontecer diario de la organización, ya que por



## Unidad VII. Nuevas tecnologías



ejemplo, esto les permite facturar y llevar control de stocks, pero pocas veces las usan para producir un análisis.

Entonces, ¿qué necesitan las organizaciones para utilizar este análisis automático de información y predecir el comportamiento? Primero, un almacén de datos construido con ese fin, un *Data Warehouse*, y después, un conjunto de estrategias y métodos de análisis, es decir, la minería de datos. Este tema presenta las características generales de ambas tecnologías.

### Objetivo particular de la unidad

Identificar los objetivos, estrategias, técnicas, enfoques y componentes de las nuevas tecnologías en bases de datos: minería de datos y data warehousing



## Unidad VII. Nuevas tecnologías



### Lo que sé:

Consulta las dos direcciones electrónicas siguientes y elabora un resumen de tres cuartillas en donde se haga énfasis en las Nuevas Tecnologías relativas a los RDBMS. Entregarlo en formato pdf.

<http://www.itchihuahua.edu.mx/it/sw/dbms/criteriosdbms.html>

“Evaluación de Soluciones de Bases de Datos” De Criterios utilizados por el Comité Técnico Informático de DGIT.

[http://www.eqsoft.net/presentas/sistemas\\_administradores\\_de\\_bases\\_de\\_datos\\_li\\_bres\\_para\\_el\\_entorno\\_empresarial.pdf](http://www.eqsoft.net/presentas/sistemas_administradores_de_bases_de_datos_li_bres_para_el_entorno_empresarial.pdf)

DE SOCIEDAD NACIONAL DE INDUSTRIAS

Realiza tu actividad en un procesador de textos, guárdala en tu computadora en formato .pdf y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.



## Unidad VII. Nuevas tecnologías



### Temas de la unidad VII

#### 1. Minería de datos

- 1.1 Definición
- 1.2 Pasos generales de la minería de datos
- 1.3 Proceso de minería de datos
- 1.4 Estrategias de minería de datos
- 1.5 Técnicas de minería de datos
- 1.6 Aplicaciones de la minería de datos

#### 2. *Data Warehousing*

- 1.1 Conceptos básicos
- 1.2 Características de un *data warehouse*
- 1.3 Componentes de un *data warehouse*
- 1.4 Áreas que usan un *data warehouse*



## Unidad VII. Nuevas tecnologías



### Resumen de la unidad

En esta unidad vimos que las Organizaciones requieren contar con información inmediata que ayude a la toma de decisiones, a fin de incrementar su productividad, encontrar hábitos de sus clientes y proveedores, planear las condiciones del mercado, responder a la competencia, etc.

Parte de esto se logra mediante el descubrimiento de ciertos patrones en los datos que contiene la Organización en sus bases de datos, a través de la aplicación de pasos y siguiendo estrategias, lo cual se denomina como Minería de Datos.

Los avances en la tecnología de bases de datos y el desarrollo de un conjunto variado de sistemas manejadores han brindado invaluable beneficios al procesamiento automatizado de información. Así, las organizaciones se han convertido en continuos desarrolladores de bases de datos transaccionales para apoyar sus actividades económico-administrativas.

Con el afán de responder a estas problemáticas, se han propuesto nuevos modelos de bases de datos que permiten contar con un gran almacén consolidado de datos listo para aplicarle procesos de análisis automático. A esta nueva tecnología se le llamó *data Warehousing*, que consiste en una colección de tecnologías que permiten mejores y más rápidas decisiones.



## Unidad VII. Nuevas tecnologías



### Tema 1. Minería de datos

#### Objetivo del tema

Identificar cómo se integra la minería de datos y su importancia en los procesos económicos y financieros.

#### Desarrollo

El volumen de datos que guardan actualmente las bases de datos se ha convertido en un recurso importante que debe analizarse. Este análisis permitiría describir y entender los procesos económicos y financieros de las organizaciones. Además, el ser humano cuenta hoy en día con enormes bases de datos científicas como las relacionadas al genoma o la astronomía, que encierran conocimiento que debe ser descubierto.

Una manera de obtener este análisis es mediante el descubrimiento de patrones en los datos. Un patrón es una serie de características o de eventos que presenta alguna regularidad, suceden cada cierto tiempo, en las mismas circunstancias o con los mismos efectos. Si vemos una base de datos, será difícil percibir a simple vista si existe un patrón, de aquí que necesitemos utilizar técnicas especializadas llamadas en conjunto: minería de datos.

#### Definición

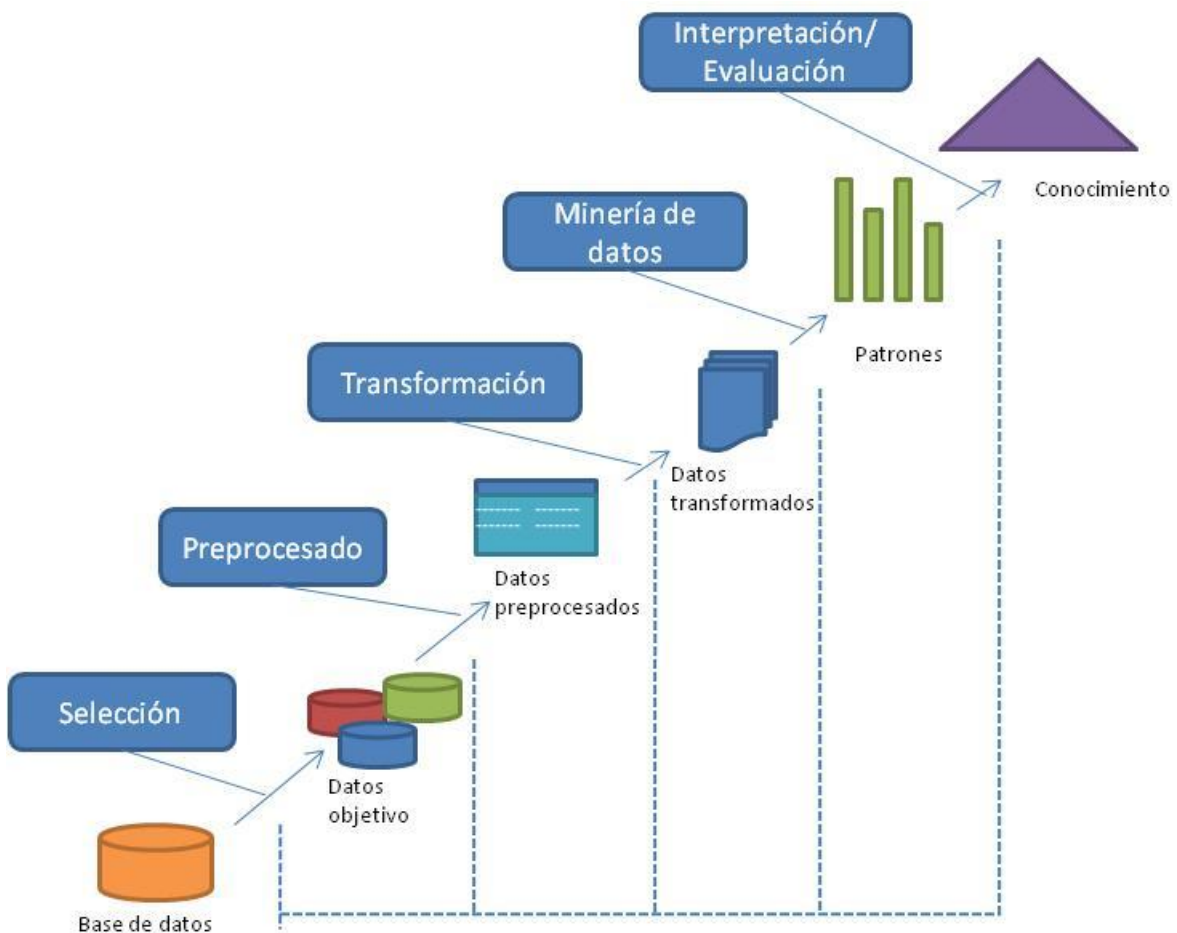
La minería de datos (MD) es también conocida como Descubrimiento de Conocimiento en Bases de Datos (*Knowledge Discovery in Databases*, KDD). Podemos definirla como la aplicación de técnicas estadísticas y de aprendizaje automático para encontrar patrones no triviales en bases de datos, que resulten de interés para el experto de un dominio determinado.



## Unidad VII. Nuevas tecnologías



La MD tiene dos enfoques. El primero es descriptivo y consiste en encontrar patrones que nos permitan describir la situación actual de la organización. El segundo es predictivo y trata de obtener modelos que pronostiquen, a partir de los patrones, algún comportamiento interesante en el futuro.



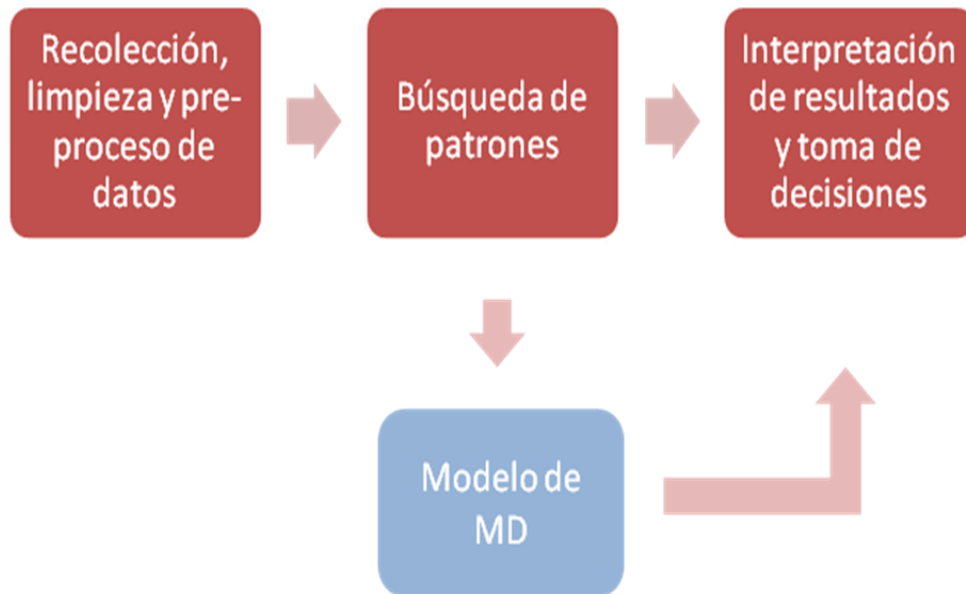
**Pasos generales de la minería de datos**



## Unidad VII. Nuevas tecnologías



La MD se lleva a cabo siempre bajo la idea de cooperación entre el experto del dominio (finanzas, mercadotecnia, ventas) y el personal de informática. A continuación los pasos generales de la MD:



*Pasos generales de la Minería de Datos*

El proceso de MD produce un modelo de MD, que puede ser descriptivo o predictivo, según los patrones identificados en los datos. Con él se analizan posibles cursos de acción para tomar decisiones. Este modelo puede ser desde una gráfica hasta una red neuronal. La recolección de datos se realiza a partir de las bases de datos transaccionales o de un *data warehouse*. Finalmente, debes saber que el modelo de MD puede ser estadístico o de aprendizaje automático.





## Unidad VII. Nuevas tecnologías



### Proceso de minería de datos

#### Primer paso

El proceso de MD comienza con la definición del objetivo de la minería. Esto se hace entre el experto del dominio y el experto en minería de datos. El objetivo puede ser explorar los datos de forma general (conocer las características de los clientes), obtener un modelo descriptivo o clasificador (obtener un modelo que clasifique a los clientes en sujetos de crédito o no), o demostrar una hipótesis (¿es cierto que los cursos más rentables se dan en Guadalajara a personas mayores de 40 años con puestos gerenciales y con ingresos superiores a \$20,000?).

#### Segundo paso

Una vez identificado el objetivo, será necesario definir las fuentes de datos y los datos relevantes para el análisis. Estos datos pueden estar resumidos o acumulados.

#### Tercer paso

En seguida, se puede hacer un acercamiento preliminar a los datos. Por lo general se tratan de aplicar medidas estadísticas y obtener gráficas representativas de los datos, por ejemplo: promedios, máximos, mínimos, histogramas, diagramas de barras, pareto o análisis de correlación.

#### Cuarto paso

En esta etapa se deciden los métodos estadísticos o de aprendizaje automático que serán aplicados a los datos. Estos se verán en los apartados de estrategias y técnicas de minería.



## Unidad VII. Nuevas tecnologías



### Quinto paso

Este paso consiste en obtener los datos y adaptarlos al formato de entrada de nuestro método. Es común que se llegue a una sola tabla con datos no normalizados como insumo de la minería. Esta tabla suele llamarse “vista minable”.

### Sexto paso

Se corren los procesos de aprendizaje automático para obtener un modelo a partir de los datos. Por lo general se utiliza un 70% de los datos originales, como datos de entrenamiento, y el resto se emplea para evaluar si el modelo obtenido es bueno.

### Séptimo paso

Evaluación del modelo obtenido mediante medidas especiales (*precision* y *recall*). Esto nos permitirá saber si el modelo obtenido es bueno.

### Octavo paso

Evaluar los resultados del proceso de MD con el experto del dominio para saber si son útiles y no triviales.

### Estrategias de minería de datos

Las estrategias de minería de datos pueden clasificarse en: supervisadas (clasificación y estimación), no supervisadas (agrupamiento o *clustering*) y análisis de canasta.



## Unidad VII. Nuevas tecnologías



### 1. Supervisadas

<b>Clasificación</b>	<p>Construye modelos que puedan asignar o clasificar nuevos ejemplos a un conjunto de clases, definidas previamente. Los modelos de clasificación pueden servirnos para:</p> <ul style="list-style-type: none"><li>A) Clasificar personas con riesgo de sufrir paro cardiaco.</li><li>B) Clasificar aspirantes con bajo o alto riesgo crediticio.</li></ul>
<b>Estimación</b>	<p>Consiste en determinar el valor de un atributo numérico. Ejemplos:</p> <ul style="list-style-type: none"><li>A) Estimar las ventas de próximo mes.</li><li>B) Estimar la probabilidad de que una tarjeta de crédito sea usada de manera fraudulenta.</li></ul>

### 2. No supervisadas

#### *Agrupamiento*

En este caso, no existen categorías predefinidas para clasificar ejemplos o instancias. Estas clases (*clusters*) se proponen como resultado del proceso de minería de datos. A este modelo de agrupamiento le acompañan medidas de cercanía entre los ejemplos o instancias agrupadas. Algunos ejemplos son:

- A) Clustering de clientes de acuerdo con su historial de ventas.
- B) Agrupamiento de nuevas especies de flores.

### 3. Análisis de canasta

Este análisis nos permite encontrar relaciones entre productos de acuerdo con sus ventas. Para este análisis suelen utilizarse algoritmos que producen reglas de



## Unidad VII. Nuevas tecnologías



asociación. Una regla sería la siguiente: “Cuando un cliente compra pan marca ‘Bready’ también compra leche ‘Milky’”.

### Técnicas de minería de datos

Las técnicas de MD se usan para aplicar una estrategia determinada a un conjunto de datos. Éstas cuentan generalmente con un algoritmo y una estructura de conocimiento. Las principales técnicas de minería son:

#### 1. Reglas de Producción.

Son reglas con la forma:

IF antecedente

THEN consecuencia

#### 2. Regresión Lineal.

Permite crear ecuaciones matemáticas con más de una variable independiente.

#### 3. Árboles de Decisión.

Es un conjunto de nodos que representan preguntas con las que se clasifica un ejemplo o instancia en una categoría predefinida.

#### 4. Clustering.

Como lo habíamos mencionado, consiste en encontrar *clusters*, también llamados cúmulos, nubes, agrupamientos o categorías en conjuntos de datos. Lo empleamos cuando no estamos seguros de las categorías que existen en ellos.



## Unidad VII. Nuevas tecnologías



### 5. Reglas de asociación (análisis de canasta).

Representan asociaciones entre atributos contenidos en las bases de datos.

### Aplicaciones de la minería de datos

Algunas de las actividades en las que resulta útil el uso de la minería de datos:

- ❖ Identificar patrones de fraudes en el uso de tarjetas de crédito.
- ❖ Identificar buenos y malos clientes.
- ❖ Predecir los clientes que podrían cambiar de banco.
- ❖ Encontrar relaciones entre indicadores financieros.
- ❖ Desarrollar modelos para diagnóstico o identificación de enfermedades.
- ❖ Predecir el perfil de los clientes que podrían adquirir nuevos productos.



### ACTIVIDAD 1

De la siguiente tabla, relaciona el campo sucursal con producto y cantidad.

Determina:

1. Qué sucursal tiene más artículos
2. Qué artículo hay en más cantidad.

Entregar los resultados en un archivo .pdf

Código	Producto	Cantidad	Sucursal	Código
4500	Regulador	45	Norte	4500
4700	No-Breker	80	Poniente	4700
4900	Pantalla	50	Centro	4900
5200	Mouse	250	Norte	5200
4500	Regulador	55	Poniente	4500
4900	Pantalla	80	Norte	4900
5200	Mouse	90	Centro	5200
4500	Regulador	55	Poniente	4500
4900	Pantalla	80	Norte	4900
5200	Mouse	90	Centro	5200

Realiza tu actividad en un procesador de textos, guárdala en tu computadora en formato .pdf y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.



### ACTIVIDAD 2

De la siguiente tabla, relaciona el campo Producto con Proveedor y Cantidad.

Posteriormente, determina:

1. ¿Qué Sucursal tiene la mayor cantidad de Productos?
2. ¿De qué Proveedor proviene la mayor cantidad de Productos?

Entrega los resultados en un archivo .pdf

Código	Producto	Cantidad	Sucursal	Proveedor
4500	Regulador	45	Norte	Emitech
4700	No-Breker	80	Poniente	Viceform
4900	Pantalla	50	Centro	Bytecom
5200	Mouse	250	Norte	Emitech
4500	Regulador	55	Poniente	Viceform
4900	Pantalla	80	Norte	Bytecom
5200	Mouse	90	Centro	Emitech
4500	Regulador	55	Poniente	Viceform
4900	Pantalla	80	Norte	Bytecom
5200	Mouse	90	Centro	Viceform

Realiza tu actividad en un procesador de textos, guárdala en tu computadora en formato .pdf y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.



### ACTIVIDAD 3

Contesta las siguientes preguntas.

1. ¿En qué consiste la minería de datos?
2. ¿A qué nos referimos al hablar de patrones en la minería de datos?
3. Enlista los pasos generales de la MD.
4. Indica, dentro del proceso de minería de datos, ¿en qué consiste el tercer paso?
5. Explica en qué consiste la estrategia de MD que se denomina Análisis de Canasta.

Para enviar tu respuesta, pulsa el botón **Editar mi envío**; se mostrará un editor de texto en el cual puedes redactar tu información; una vez que hayas concluido, salva tu actividad pulsando el botón **Guardar cambios**.

### ACTIVIDAD 4

Investiga qué tipo de Bases de Datos son importantes en las empresas, en las siguientes áreas: Ventas, Compras, Finanzas, Mercadotecnia y Recursos Humanos.

Entregar sus resultados impresos en formato pdf.

Realiza tu actividad en un procesador de textos, guárdala en tu computadora en formato .pdf y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.





## Unidad VII. Nuevas tecnologías



### ACTIVIDAD 5

Investiga cuáles Bases de Datos a nivel Federal son importantes en material electoral, demográfico, contribuciones, empresarial y de vehículos. Entregar sus resultados impresos en formato pdf.

Realiza tu actividad en un procesador de textos, guárdala en tu computadora en formato .pdf y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.

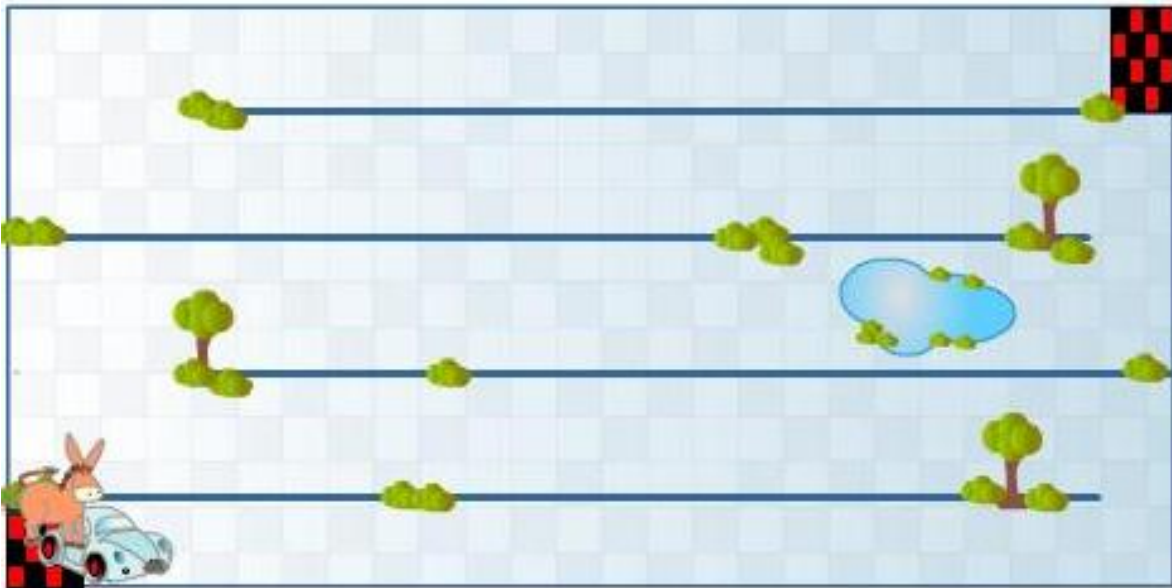


## Unidad VII. Nuevas tecnologías



### Autoevaluación

Participa en el siguiente Rally. Responde correctamente para que puedas avanzar.



1. Los patrones en los datos generan nuevos conocimientos.

V ( ) F ( )

2. Con la existencia de Bases de Datos se crea la Minería de Datos.

V ( ) F ( )

3. En el segundo paso del proceso de minería de datos se crean gráficas.

V ( ) F ( )

4. En el sexto paso del proceso de minería de datos se aplican métodos estadísticos.

V ( ) F ( )

5. El agrupamiento es una estrategia supervisada.

V ( ) F ( )



## Unidad VII. Nuevas tecnologías



6. Los dos enfoques de la minería de datos son: clasificador y estimador.

V ( ) F ( )

7. La minería de datos busca descubrir patrones no triviales en los datos.

V ( ) F ( )

8. La minería de datos se realiza únicamente por un experto en cómputo.

V ( ) F ( )

9. La estrategia de clasificación permite obtener grupos o clusters de datos.

V ( ) F ( )

10. El análisis de canasta permite obtener valores numéricos estimados.

V ( ) F ( )

11. Las reglas de producción consisten de un grupo de nodos que clasifican un ejemplo.

V ( ) F ( )

12. El registro nacional de vehículos es ejemplo de una Base de Datos a nivel Estatal.

V ( ) F ( )



## Unidad VII. Nuevas tecnologías



### Bibliografía básica

Autor	Capítulo	Páginas

### Sitios electrónicos

Sitio	Descripción



## Unidad VII. Nuevas tecnologías



### Tema 2. Data Warehousing

#### Objetivo del tema

Identificar los nuevos aspectos de la Tecnología de Bases de Datos y el desarrollo en los Sistemas Manejadores de Bases de Datos para crear recursos informativos e informáticos en las organizaciones.

#### Desarrollo

Los avances en la tecnología de bases de datos y el desarrollo de un conjunto variado de sistemas manejadores han brindado invaluables beneficios al procesamiento automatizado de información. Así, las organizaciones se han convertido en continuos desarrolladores de bases de datos transaccionales para apoyar sus actividades económico-administrativas.

Junto a la observación anterior es necesario rescatar dos aspectos que no han sido favorables para la misma organización:

**Primero**, se volvió práctica común que distintos departamentos de una misma organización decidieran trabajar con distintos sistemas manejadores de bases de datos, incluso no resulta raro encontrar en una misma área distinto software de base de datos; situación que ha generado problemas para el análisis y procesamiento consolidado de datos.

**Segundo**, las organizaciones se preocuparon mucho - y con justa razón - por crear sistemas robustos para captar las operaciones diarias de la organización, y no contemplaron la labor del análisis de los datos. De esta manera, es normal encontrar en muchas organizaciones enormes bases de datos transaccionales y procesos de análisis de datos manuales basados en reportes automatizados, eso si hay buena suerte.



## Unidad VII. Nuevas tecnologías



Con el afán de responder a estas problemáticas, se propusieron nuevos modelos de bases de datos que permiten contar con un gran almacén consolidado de datos listo para aplicarle procesos de análisis automático. A esta nueva tecnología se le llamó *data warehousing*.

### Conceptos básicos

Para Chaudhuri y Dayal, el *data warehousing* consiste en una colección de tecnologías que permiten mejores y más rápidas decisiones:

Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions.<sup>1</sup>

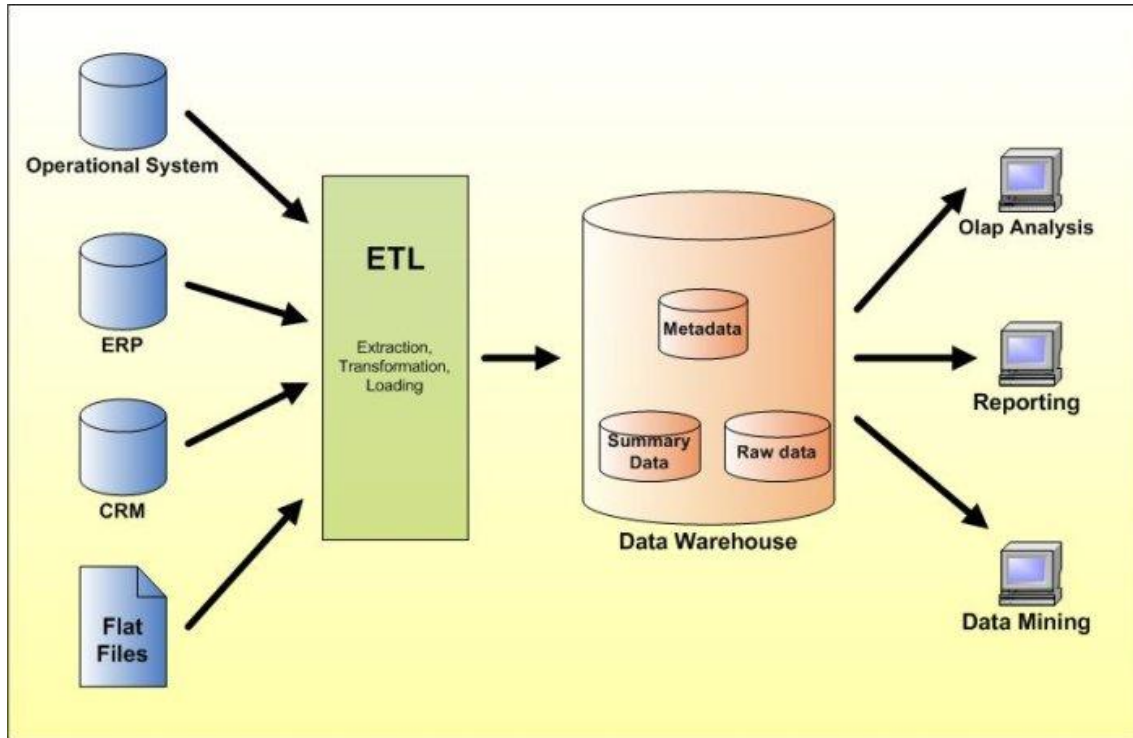
El objetivo del *data warehousing* es brindar la tecnología necesaria para obtener resúmenes complejos de información y conocimiento. Asimismo, empleando las respectivas tecnologías, herramientas y metodologías, se podría crear, usar y mantener un *data warehouse*.

---

<sup>1</sup> CHAUDHURI, S., DAYAL, U., An Overview of Data Warehousing and OLAP Technology, ACM SIGMOD Record, no. 26-1. 1997.



## Unidad VII. Nuevas tecnologías



Por su parte, el concepto de *data warehouse* (DW) se puede entender como un depósito centralizado de datos que ayuda al análisis del negocio. Es un almacén que unifica las bases de datos empresariales o departamentales sin importar el sistema manejador en el que se encuentren. Uno de los precursores de esta tecnología indica: “A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions”.<sup>2</sup>

Al respecto, también se suele escuchar el concepto de *Data Mart*. Vamos a entender un *data mart* como un *data warehouse* departamental, con las mismas características y componentes, pero a nivel de un departamento de la organización. Es más, podríamos decir que un *data warehouse* se forma de todos los *data mart* de la organización.

<sup>2</sup> William H. Inmon, *Building the data warehouse*, 3a ed., Nueva York, John Wiley & Sons. 2002.



## Unidad VII. Nuevas tecnologías



### Características de un *data warehouse*

A partir de la definición del llamado padre del *data warehouse*, William H. Inmon, podemos identificar sus características:

- ◆ **Orientados a temas**

Los datos de un DW deben estar recolectados para proporcionar información sobre los temas importantes de la empresa y no sobre sus operaciones diarias.

- ◆ **Integrado**

Los datos están integrados a partir de una variedad de bases de datos transaccionales provenientes de los sistemas OLTP de la organización. Los datos deben ofrecer una imagen corporativa general de ella.

- ◆ **No volátil**

Los datos de un DW no son eliminados ni modificados, ya que se quiere representar la historia de la organización.

- ◆ **Variante en el tiempo**

Los datos están asociados a períodos de tiempo, específicos y bien identificados.

### Componentes de un *data warehouse*

Los principales componentes de un DW incluyen el *data warehouse* como tal y algunos elementos adicionales como las fuentes y destinos del mismo.





## Unidad VII. Nuevas tecnologías



### ◆ Bases de datos transaccionales y otros recursos externos

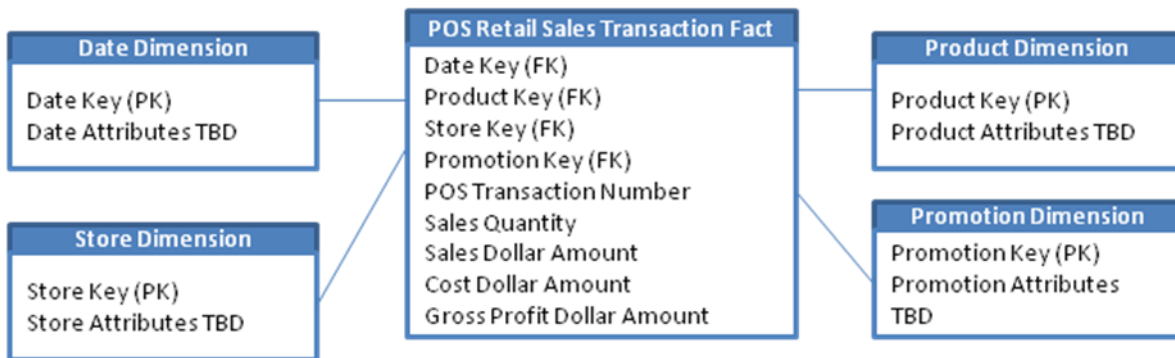
Son las fuentes de origen de los datos las que forman el DW. De ellas se extraen los datos y se cargan en las estructuras del DW. Es común que los datos se tengan que limpiar y transformar. Este proceso es conocido como ETL (extract, transform and load).

### ◆ Metadatos

Son datos que describen el contenido del *data warehouse*. Entre estos se incluyen el origen de los datos, el responsable y las descripciones sobre los tipos de resúmenes de datos.

### ◆ La base de datos del DW

Es la base de datos que almacena los datos del *data warehouse*. Es normal observar que el almacenamiento se haya realizado bajo un modelo dimensional y no relacional. Este modelo consiste en un grupo de tablas de dimensiones que guardan una relación de uno a muchos con una tabla principal o tabla de hechos.



Measured facts in the retail sales schema

### Ejemplo de un Modelo dimensional de almacenamiento

<sup>3</sup> Kimball (2002: 36).



## Unidad VII. Nuevas tecnologías



### ◆ Herramientas de consulta

La construcción de un *data warehouse* conlleva la creación de herramientas de procesamiento y análisis de los datos contenidos en él. Estas herramientas pueden ser parte de lo que se conoce como un sistema OLAP (*On line analytical processing*) o ser herramientas de minería de datos.

### ◆ Los usuarios

Son el último componente de un DW y son los que explotan sus beneficios, ya sea mediante el resultado de la minería de datos o de sistemas de consulta para soporte de las decisiones.

### Áreas que usan un *data warehouse*

A continuación se listan las diversas áreas que en la actualidad han adoptado la tecnología del *data warehousing* para mejorar sus negocios:

Servicios financieros	Comercio minorista
Análisis de rentabilidad	Análisis y planificación de distribución
Gestión de riesgo y prevención de fraude	Telecomunicaciones
Análisis de marketing	Gobierno
Seguros	Salud

El *data warehouse* se está convirtiendo en la base de los procesos analíticos de negocios y es importante que tú, como estudiante de informática, conozcas sus aspectos fundamentales, que hemos expuesto a lo largo de este tema.



## Unidad VII. Nuevas tecnologías



### ACTIVIDAD 1

Elabora un mapa conceptual del tema 7.1 Minería de datos, que abarque todos los aspectos expuestos en la lectura.

Realiza tu actividad en un procesador de textos, guárdala en tu computadora y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.

### ACTIVIDAD 2

Realiza un cuadro sinóptico de las estrategias y las técnicas de minería de datos.

Realiza tu actividad en un procesador de textos, guárdala en tu computadora y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.

### ACTIVIDAD 3

Plasma en un cuadro sinóptico, los componentes de un data warehouse.

Realiza tu actividad en un procesador de textos, guárdala en tu computadora y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.



## Unidad VII. Nuevas tecnologías



### ACTIVIDAD 4

Realiza un cuadro con las características de las nuevas tecnologías aplicadas al data warehousing y del data mart.

Realiza tu actividad en un procesador de textos, guárdala en tu computadora y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.

### Cuestionario de autoevaluación

Contesta las siguientes preguntas.

1. Define el data warehousing.
2. ¿Qué es un data warehouse?
3. ¿Cuáles son las características de un data warehouse?
4. Describe los componentes de un data warehouse.

Para enviar tu respuesta, pulsa el botón **Editar mi envío**; se mostrará un editor de texto en el cual puedes redactar tu información; una vez que hayas concluido, salva tu actividad pulsando el botón **Guardar cambios**.



## Unidad VII. Nuevas tecnologías



### Autoevaluación

Selecciona falso (F) o verdadero (V) de acuerdo con las siguientes afirmaciones. Al concluir tendrás tu calificación de manera automática.

	Verdadera	Falsa
1. Para C. Date y Dayal, el data warehousing es una serie de tecnologías aplicadas a las bases de datos.	( )	( )
2. Un data warehouse refleja la actividad diaria de la empresa, es actualizable y sin referencia a un tiempo específico.	( )	( )
3. Un data mart se forma de varios data warehouse.	( )	( )
4. Los metadatos y las bases de datos transaccionales son componentes de un data warehouse.	( )	( )
5. El modelo dimensional se basa en tablas de muchos a muchos entre diversos catálogos dimensionales.	( )	( )

### Bibliografía básica

Autor	Capítulo	Páginas

### Sitios electrónicos

Sitio	Descripción



## Unidad VII. Nuevas tecnologías



### Lo que aprendí

De las siguientes Tablas crea la Base de Datos “Recursos” y realiza lo siguiente.

1. Aplica el Modelo Entidad-Relación para definir las relaciones, las Llaves Primarias y Secundarias y Compuestas en su caso.
2. Realiza el Proceso de Normalización con sus etapas para optimizar la funcionalidad de la Base de Datos.
3. Genera la Minería de Datos para que conteste a las preguntas de que Clientes consumen más, que artículos se compran más, a que proveedor se le compra más y la sumatoria de las Ventas en un mes determinado.

Imprime cada una de las actividades en formato pdf.

ARTICULO							
Código	Artículo	Cantidad	Precio unitario	Sucursal	Proveedor		
CLIENTES							
Numero	Nombre	Rfc	Dirección	Factura	Saldo	Teléfono	Delegación
VENTAS							
Factura	Importe	Rfc	Código	Numero	Fecha	Proveedor	Forma pago
PROVEEDOR							
IdProve	Nom_Prov	RFC_Prov	Direc_Prov	Tel_Prov	Código	Pago_Prov	

Realiza tu actividad en un procesador de textos, guárdala en tu computadora en formato .pdf y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.



## Unidad VII. Nuevas tecnologías



### GLOSARIO

#### Árboles de decisión

Es un conjunto de nodos que representan preguntas con las que se clasifica un ejemplo o instancia en una categoría predefinida.

#### Base de datos del DW

Es la base de datos que almacena los datos del *data warehouse*. Es normal observar que el almacenamiento se haya realizado bajo un modelo dimensional y no relacional. Este modelo consiste en un grupo de tablas de dimensiones que guardan una relación de uno a muchos con una tabla principal o tabla de hechos.

#### Bases de datos transaccionales y otros recursos externos

Son las fuentes de origen de los datos las que forman el DW. De ellas se extraen los datos y se cargan en las estructuras del DW.

#### Cluster

También llamados cúmulos, nubes, agrupamientos o categorías en conjuntos de datos.

#### Data mart

Es un *data warehouse* departamental, con las mismas características y componentes, pero a nivel de un departamento de la organización. Es más, podríamos decir que un *data warehouse* se forma de todos los *data mart* de la organización.



## Unidad VII. Nuevas tecnologías



### **Data Warehousing**

Son nuevos modelos de bases de datos que permiten contar con un gran almacén consolidado de datos listo para aplicarle procesos de análisis automático. Consiste en una colección de tecnologías que permiten mejores y más rápidas decisiones.

### **Enfoque descriptivo de MD**

Consiste en encontrar patrones que permitan describir la situación actual de las organizaciones.

### **Enfoque predictivo de MD**

Consiste en encontrar modelos que pronostiquen, a partir de los patrones, algún comportamiento interesante en el futuro.

### **Metadatos**

Son datos que describen el contenido del data warehouse. Entre estos se incluye el origen de los datos, el responsable y las descripciones sobre los tipos de resúmenes de datos.

### **Minería de datos (MD)**

Es la aplicación de técnicas estadísticas y de aprendizaje automático para encontrar patrones no triviales en bases de datos, que resulten de interés para el experto de un dominio determinado. Es también conocida como descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases, KDD).

### **Patrón**

Es una serie de características o de eventos que presentan alguna regularidad; suceden cada cierto tiempo, en las mismas circunstancias o con los mismos efectos.





## Unidad VII. Nuevas tecnologías



### **MESOGRAFÍA**

#### **Bibliografía básica**

#### **Bibliografía complementaria**

#### **Sitios electrónicos**