



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CONTADURÍA Y  
ADMINISTRACIÓN

*"LA MINERÍA DE DATOS UNA HERRAMIENTA  
PARA LA TOMA DE DECISIONES"*

TESIS PROFESIONAL  
QUE PARA OBTENER EL TÍTULO DE :

***LICENCIADO EN INFORMÁTICA***

Presenta:

ESMERALDA MARTÍNEZ MONTES

Asesor:

Dr. RICARDO RIVERA SOLER

Ciudad Universitaria

287820

200





Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



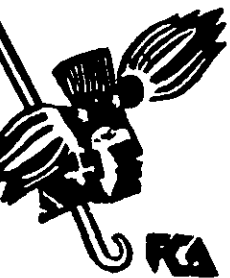
# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CONTADURIA Y ADMINISTRACION

"LA MINERIA DE DATOS UNA HERRAMIENTA  
PARA LA TOMA DE DECISIONES"

## TESIS PROFESIONAL INFORMATICA

ESMERALDA MARTINEZ MONTES



CIUDAD UNIVERSITARIA

2000



---

---

*A Dios :*

*Por permitirme tener la vida.*

*Por tener los padres que tengo.*

*Por enseñarme a valorar lo que tengo, y lo que tuve.*

*Por regalarme una estrella cada noche y un rayo  
de sol cada mañana.*

*Por dejarme soñar, y alcanzar mi sueño...*

*A mis padres :*

*Por darme lo que me han dado.*

*Por guiarme en la vida.*

*Por enseñarme a ganarme lo que tengo con esfuerzo.*

*Por enseñarme que con AMOR todo es posible.*

*Y simplemente por apoyarme a alcanzar un sueño.*

*A mis abuelos :*

*Victor Montes*

*Santos García*

*Virginia Alvarado*

*Por ser siempre para mí un ejemplo de AMOR,  
paciencia, tenacidad, perseverancia, coraje,  
fuerza y sobre todo aprender que cada día  
es un hermoso regalo de Dios.*

---

---

---

*A mi asesor:*

*Dr. Ricardo Rivera Soler*

*Por ser más que un maestro para mí, ser un  
AMIGO en quien confiar.*

*Por enseñarme que con tenacidad y constancia  
todo es posible.*

*Por guiarme por el camino correcto para llegar a  
alcanzar una pequeña estrella.*

*A la Universidad Nacional Autónoma de México  
UNAM:*

*Por ser una verdadera casa para mí.*

*Por dejar en mí un sello inolvidable.*

*Por compartir conmigo alegrías y tristezas.*

*Por permitirme llegar a ser una verdadera  
profesionista universitaria.*

---



---

## INTRODUCCIÓN

Debido a que nuestro País, se encuentra en un proceso de desarrollo científico y tecnológico, han surgido nuevas técnicas y herramientas que ayuden a los administradores, gerentes y en general a los tomadores de decisiones de las empresas.

Muchas de las técnicas y herramientas ya son utilizadas por muchos empresarios de otros países, principalmente de estados unidos, donde su manejo de la información se da a través de bases de datos; es por eso que ahora en esta investigación se da una visión general de la herramienta conocida como "minería de datos", para los tomadores de decisiones de nuestro País.

Esta herramienta tiene como principal objetivo el buscar relaciones ocultas entre la información almacenada en una base de datos para dar un mayor conocimiento de la organización a los tomadores de decisiones.

Por todos es conocido que el tomar una buena decisión es una parte vital del conocimiento e incremento o decremento de la productividad de la empresa, es por eso que en la presente investigación se darán a conocer los beneficios que puede dar el utilizar una técnica y / o herramienta de "minería de datos" para los tomadores de decisiones.

Esta investigación va dirigida a estudiantes y profesionales del área de informática que se encuentren con la inquietud de conocer una nueva herramienta que coadyuve a la toma de decisiones y sobre todo que deseen saber qué beneficios puede acarrear aplicarla en una organización.

El criterio que se toma para la elaboración de esta investigación no es el proporcionar una solución genérica a los problemas de toma de decisiones; sino dar una alternativa de apoyo a una mejor toma de decisiones.



---

La presente investigación se encuentra dividida en 6 capítulos como lo son :

1. *Marco Problemático*: en este capítulo se da una explicación del por qué de la elaboración de esta investigación, es decir se dan los objetivos generales y específicos de la misma.
2. *Marco Teórico*: en este capítulo se dan a los lectores las referencias bibliográficas, y de internet que fueron utilizadas para la elaboración de esta investigación.
3. *Marco Conceptual*: dentro de este capítulo se desarrolla el detalle pormenorizado del tema de "minería de datos", es decir se da una introducción, las definiciones de diversos autores, así como la propia, la historia, evolución, técnicas, y las últimas tendencias y / o futuro de este tema.
4. *Marco Metodológico*: aquí se muestran los pasos que se siguieron para comprobar que en nuestro País no son muy conocidas las herramientas de "minería de datos" y cuáles son los posibles beneficios que se pueden tener al aplicar algunas de estas herramientas. En general se prueba la hipótesis planteada desde el capítulo número uno.
5. *Marco Instrumental*: dentro de este capítulo se describen las actividades que se llevan y se llevarán a cabo como resultado de esta investigación.
6. *Conclusiones*: por último se da un resumen del resultado de la investigación.



# Índice

---

---

# ÍNDICE

<b>1. Marco Problemático</b>	<b>1</b>
1.1 Antecedentes	2
1.2 Identificación del Problema	2
1.3 Demarcación del fenómeno	2
1.4 Conocimiento empírico del medio	3
1.5 Opiniones Profesionales	6
1.6 Hipótesis Preliminar	6
1.7 Objetivos Personales y Generales	8
<b>2. Marco Teórico</b>	<b>9</b>
2.1 Acopio Bibliográfico	10
2.1.1 Lectura Total o libros de estudio	10
2.1.2 Lectura ligera	13
2.1.3 Lectura Rápida	14
2.1.4 Lectura Superficial	14
2.2 Tesis	14
2.3 Folletos	15
2.4 Direcciones de Internet	16
2.5 Congresos y Seminarios	22
2.6 Desarrollo de Software especializado	24
<b>3. Marco Conceptual</b>	<b>27</b>
3.1 Antecedentes	28
3.2 Definiciones	36
3.3 Evolución	40
3.4 Clasificación	59
3.5 Técnicas	60
3.6 Ingredientes de la minería de datos	64
<b>4. Marco Metodológico</b>	<b>70</b>
4.2 Variables	71
4.3 Hipótesis definitiva	71
4.4 Definición del Universo	71
4.5 Definición de la muestra	72

---

4.6 Definición del Método de la investigación	73
4.7 Costo de la Investigación	74
4.8 Construcción del Cuestionario	75
4.10 Cuestionario Piloto	79
4.11 Conclusión y aprobación de la hipótesis	85
<b>5. Marco Instrumental</b>	<b>87</b>
<b>6. Conclusiones</b>	<b>89</b>
<b>ANEXOS</b>	<b>90</b>
Anexo 1 - Propuesta de temario	91
Anexo 2 - Artículo de revista	95
Anexo 3 - Página Web	98
Anexo 4 - Presentación	103
<b>GLOSARIO</b>	<b>109</b>



# Marco Problemático

XX

---

# 1. MARCO PROBLEMÁTICO

## 1.1 ANTECEDENTES

El tema Minería de datos, ó como se conoce en Estados Unidos "Data Mining" me interesó principalmente por que es un tema de actualidad, es decir, que no es un tema muy conocido dentro del ámbito informático dentro de nuestro país.

En la actualidad se cuenta con muchas herramientas para la explotación de una base de datos, pero aún así considero que no es suficiente porque he logrado observar que todavía se tiene la oportunidad de poder sacarle más provecho a toda esa información que se tiene almacenada, y sobre todo que pueda coadyuvar a una mejor toma de decisiones y esto se puede lograr con la Minería de Datos.

Mi interés sobre el tema es el de dar a conocer a la comunidad informática en general una nueva alternativa para explotar la información contenida en una base de datos.

## 1.2 IDENTIFICACIÓN DEL PROBLEMA.

Dentro de un entorno informático, ya sea en el ámbito empresarial o nivel de educativo; no se ha considerado a la minería de datos como una de las más completa y principales herramientas de aprovechamiento y explotación de los datos almacenados en una base de datos. Todo esto trae como consecuencia un desaprovechamiento de los datos y un desperdicio de recursos, por así decirlo al no poder explotar al 100% una base de datos con herramientas que tenemos a nuestro alcance.

## 1.3 DEMARCACIÓN DEL FENÓMENO

Esta investigación estará enfocada principalmente a empresas del sector privado, las cuales se dediquen a desarrollar sistemas o simplemente a dar consultorías a sistemas de empresas, donde estos hagan uso de bases de datos relacionales.

## 1.4 CONOCIMIENTO EMPÍRICO DEL MEDIO

### CUESTIONARIO

Pregunta	Razón de ser	Respuesta Esperada
1. <i>¿En su ámbito laboral cuentan con alguna metodología para el análisis de información?</i> SÍ ( ) NO ( )	Saber si en distintas empresas del país tienen alguna metodología definida para el análisis de la información contenida en una base de datos.	Sí
2. Si <i>contesto afirmativamente la pregunta anterior. Diga que metodología.</i>	Saber si se cuenta con una metodología propia de la empresa o alguna de las conocidas, como la del modelo jerárquico, de red, etc.	Si, utilizamos una combinación del modelo jerárquico y el modelo distribuido para analizar la información.
3. <i>Describa brevemente el procedimiento que sigue para el análisis de la información</i>	Llegar a conocer la(s) técnica(s) que se utilizan para la obtención de información	1. análisis de del manejo de los datos 2. Identificación de factores importantes 3. Detección de excepciones.
4. <i>¿Considera que las metodologías utilizadas son eficientes para la obtención y manejo de información histórica?</i>	Saber si han podido obtener y manipular información histórica con los resultados esperados.	Si hemos podido obtener información histórica, pero no con la rapidez y certidumbre que se desearía.
5. <i>¿Utiliza alguna herramienta automatizada para el manejo de la información histórica?</i>	Conocer si utilizan herramientas de manejo de información histórica, para saber si alguna de estas puede ser de minería de datos	Si, la herramienta que utilizamos es el DBMine
6. Si <i>contesto afirmativamente a la pregunta anterior. ¿Considera que la</i>	Saber si los resultados obtenidos de estas herramientas son	Si, nos ayuda tanto a tomar decisiones y a conocer las posibles

<i>información que obtiene por medio de estas herramientas es suficiente para la toma de decisiones, y/o conocer de forma detallada a su empresa?</i>	suficiente para tomar decisiones y conocer de forma detallada a la empresa donde trabaja.	desviaciones o aciertos que tiene la empresa en diversos ámbitos.
<i>7. Conoce el término Minería de Datos. Si es así, de una definición de Minería de Datos</i>	Conocer la definición personal de Minería de Datos	Es una combinación de técnicas, por medio de las cuales podemos encontrar información muy valiosa
<i>8. ¿Conoce alguna persona que se dedique a la investigación y promoción de la Minería de Datos en nuestro país?</i>	Saber si tienen el conocimiento de alguna persona que se dedique a está área del conocimiento	

#### Cuestionario

1. *¿En su ámbito laboral cuentan con alguna metodología para el análisis de información?  
SÍ ( ) NO ( )*
2. *Si contesto afirmativamente la pregunta anterior. Diga que metodología.*
3. *Describa brevemente el procedimiento que sigue para el análisis de la información*
4. *¿Considera que las metodologías utilizadas son eficientes para la obtención y manejo de información histórica?*
5. *¿Utiliza alguna herramienta automatizada para el manejo de la información histórica?*
6. *Si contesto afirmativamente a la pregunta anterior. ¿Considera que la información que obtiene por medio de estas herramientas es suficiente para la toma de decisiones, y/o conocer de forma detallada a su empresa?*
7. *Conoce el término Minería de Datos. Si es así, de una definición de Minería de Datos*
8. *¿Conoce alguna persona que se dedique a la investigación y promoción de la Minería de Datos en nuestro país?*



## CUADRO SINÓPTICO

### Minería de Datos

Por su área de aplicación	<ul style="list-style-type: none"> <li>En empresas</li> <li>En consultorías</li> <li>En universidades</li> </ul>	<ul style="list-style-type: none"> <li>Sector público</li> <li>Sector privado</li> <li>De servicios               <ul style="list-style-type: none"> <li>Nacionales</li> <li>Extranjeras</li> </ul> </li> <li>Públicas</li> <li>Privadas</li> </ul>
Por su finalidad	<ul style="list-style-type: none"> <li>Desarrollo de nuevos sistemas</li> <li>Para conocimiento de un área en específico de la empresa</li> </ul>	
Por quien lo realiza	<ul style="list-style-type: none"> <li>Por empleados de la empresa (Información confidencial)</li> <li>Por empresas especializadas</li> </ul>	<ul style="list-style-type: none"> <li>Consultorías</li> <li>Fabricantes de software especializado</li> </ul>
Por las herramientas que se utilizan	<ul style="list-style-type: none"> <li>Conocimientos técnicos, personas dedicadas a esta área.</li> <li>Software especializado</li> </ul>	<ul style="list-style-type: none"> <li>Análisis estadístico</li> <li>Descubrimiento de conocimientos (IA)</li> <li>Propietario</li> </ul>
Por quién lo utiliza	<ul style="list-style-type: none"> <li>Analistas empresariales</li> <li>Peritos en estadísticas</li> <li>Profesionales en tecnologías de Información</li> </ul>	
Por quienes obtienen sus beneficios	<ul style="list-style-type: none"> <li>Directivos</li> <li>Gerentes empresariales</li> <li>Ejecutivos del negocio</li> </ul>	
Por las técnicas que se utilizan	<ul style="list-style-type: none"> <li>Redes Neuronales</li> <li>Descubrimiento de asociaciones</li> <li>Clasificación</li> <li>Clustering</li> <li>Descubrimiento secuencial</li> </ul>	

## 1.5 OPINIONES PROFESIONALES

### Definición de Personal.

#### ❖ Profesionales

Dr. Adolfo Guzmán Arenas

M. en C. Gilberto Lorenzo Martínez Luna

Lic. Roberto García

Ing. Agustín Melchor

#### ❖ Teóricas

Lic. Oscar Ramírez Contreras

Lic. Edwin Erik Rodríguez

Lic. Manuel Nicolás Ramírez Contreras

Lic. Daniel Pérez Hernández

## 1.6 HIPÓTESIS PRELIMINAR

### Correlaciones Causa - Efecto

Variables Independientes	Variables Dependientes
Grandes volúmenes de Información	Decisiones no tan certeras Decisiones lentas Desaprovechamiento de información
Uso de herramientas automatizadas	Mejor aprovechamiento de la información
Reportes o informes detallados	Toma de decisiones más certeras y rápidas.

### Ejemplos de correlaciones que alteran información:

- ❖ El manejo y almacenamiento de grandes volúmenes de información, puede conducir a decisiones no muy certeras.
- ❖ El manejo y almacenamiento de grandes volúmenes de información, puede llevarnos a tomar decisiones lentas.
- ❖ El manejo y almacenamiento de grandes volúmenes de información, puede conducir al desaprovechamiento de la información.

- 
- ❖ El uso de herramientas automatizadas nos lleva al mejor aprovechamiento de la información
  - ❖ La elaboración de reportes o informes detallados, nos ayuda a la toma de decisiones más certeras y rápidas.

### *Hipótesis Preliminar.*

"El conocimiento a detalle de una organización y la elevación de la productividad se aumentará por el uso de una herramienta informática denominada "*minería de datos*". La cual nos generará reportes o informes que coadyuvarán a una mejor toma de decisiones dentro de las empresas."

### Definición Temática

La minería de datos la considero como una combinación de técnicas semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos y que pueda resultar más provechosa que los propios datos de partida.

---

## 1.7 OBJETIVOS PERSONALES Y GENERALES

### *PERSONAL*

- ❖ Graduarme en la Licenciatura en Informática de acuerdo al Reglamento General de Exámenes Profesionales de la Universidad Nacional Autónoma de México conforme a sus artículos 18 al 31 del mismo reglamento. También de acuerdo al Reglamento de Exámenes Profesionales de la Facultad de Contaduría y Administración de acuerdo a su apartado E, artículos 44 a 56 de dicho reglamento.
- ❖ Por otra parte también deseo volverme una persona experta en el tema de Minería de Datos, debido a que es un tema de actualidad y sobre todo de mucho provecho para la comunidad informática de nuestro país.

### *GENERAL*

- ❖ Aportar una forma de conceptualizar, estudiar y aplicar la Minería de Datos entendible para cualquier persona que este involucrada en la informática.
- ❖ Aportar un conocimiento útil de utilizar la minería de datos y que pueda ser aplicable a cualquier tipo de empresa, para que esto coadyuve al aumento de la productividad de la misma.



# Marco Teórico

## 2. MARCO TEÓRICO

### 2.1 Acopio Bibliográfico

Para la mejor comprensión del tema "minería de datos", se hizo uso de diversas fuentes documentales, las cuales se clasificarán en :

- 2.1.1 *Lectura total o libros de estudio: se realiza de todo el libro*
- 2.1.2 *Lectura ligera: se consulta de uno o varios capítulos*
- 2.1.3 *Lectura Rápida: referente a un tema en específico*
- 2.1.4 *Lectura Superficial: de uno o varios conceptos.*

### LIBROS

#### 2.1.1 *Lectura Total o libros de estudio*

**LIBRO :**

Data Mining

Pieter Adriaans and Dolf Zantinge

Addison Wesley; 1ª. Edición , 1996

ISBN 0-201-40380-3

**DESCRIPCIÓN :**

**CAPÍTULO 1.**

Nos da una breve introducción al área de conocimiento y descubrimiento en la bases de datos (Knowledge Discovery in Databases KDD); también da algunas definiciones importantes para las organizaciones modernas; así como también explica el ambiente y las condiciones de las organizaciones para aplicar la minería de datos.

**CAPÍTULO 2.**

En este capítulo se explican los conceptos físicos y abstractos que se deben entender en la metodología de la investigación de la minería de datos; por otra parte nos muestra algunos algoritmos para la minería de datos; así como algunos aspectos a considerar para determinar el algoritmo de aprendizaje que sea el más adecuado para lo que se desea descubrir.

**CAPÍTULO 3.**

En este capítulo se muestran las relaciones entre la minería de datos y el data warehouse. Por otra parte explica que los datos son muy necesarios, si no es que indispensables para los procesos de la minería de datos, y el data warehouse es considerado como la mejor estructura para la minería de datos

en un ambiente KDD; así como también puede ser integrado en un sistema de soporte de decisiones y el diseño de bases de datos.

#### **CAPÍTULO 4.**

Describe el proceso completo de KDD, y muestra un ejemplo muy amplio para poder comprenderlo correctamente, también muestran varios usos de los procesos de la minería de datos, para la especificación de la información requerida en un select.

#### **CAPÍTULO 5.**

Este capítulo nos ayuda a comprender ¿Qué es lo que tenemos que considerar cuando iniciamos un proyecto de minería de datos? ¿Qué pasos debemos llevar a cabo?; y al final del capítulo nos muestra algunas reglas para un KDD.

#### **CAPÍTULO 6.**

Nos describe algunas aplicaciones de la vida real, y de la experiencia diaria; como por ejemplo de un banco, tomando información llamada "piloto" y experiencias anteriores, y muchas otras técnicas de reingeniería de bases de datos. Estos casos ilustran algunas posibilidades, para aplicar las técnicas de minería de datos.

#### **CAPÍTULO 7.**

Este capítulo describe algunos aspectos formales para los algoritmos de conocimiento y la teoría relacional de la minería de datos: dichos conceptos se refieren a la teoría completa y las bases de datos primitivas para la minería de datos.

#### **CONCLUSIÓN :**

Este libro nos permite iniciar a conceptualizar lo que es la minería de datos y cuales son sus técnicas y herramientas que se podrían utilizar al aplicar a minería de datos en una organización.

**LIBRO :**

Data Mining Techniques: For Marketing, Sales and Customer Support

Michael J.A. Berry and Gordon Linoff

Wiley Computer Publishing: 1<sup>st</sup> edition, 1997, ISBN 0-471-17980-9

**DESCRIPCIÓN :**

**CAPÍTULOS 1, 3 - 6.**

Los primeros seis capítulos de este libro nos introducen a la minería de datos y discute su contexto en los negocios. Nos muestra que el valor de los datos es solo el principio; también explica como dentro de una red neuronal se puede predecir como van a responder los datos; pero esta información no puede ser evaluada cuando esta es incorporada en el proceso del negocio.

**CAPÍTULO 2.**

Este capítulo explica el ciclo de la minería de datos, sus técnicas y aplicaciones en los negocios.

**CAPÍTULO 7.**

Este capítulo nos introduce a algunas metodologías para llevar a cabo la minería de datos.

**CAPÍTULOS 8 - 14.**

Estos capítulos nos muestran siete técnicas, como son : detección de cluster, memoria basada en el conocimiento, análisis de mercado, algoritmos genéticos, ligas de análisis, árboles de decisiones y redes neuronales. Estos capítulos se ilustran con ejemplos.

**CAPÍTULOS 15 - 18.**

Estos últimos capítulos nos enseñan las tareas de la minería de datos. Una de las principales herramientas de la minería de datos son las de Proceso Analítico en Línea (OLAP - On-Line analytic processing).



## 2.1.2 Lectura Ligera

Libro	Descripción
<p>Data Warehousing: La integración de la información para la mejor toma de decisiones. Harjinder S. Gill y Prakash C. Rao, Prentice Hall: 1ª. Edición, 1996 ISBN 968-880-792-3</p>	<p>Se consulto el capítulo 1 "Introducción al data warehousing" y el capítulo 11 de "Minería de Datos", dentro del cual se analizaron los puntos ingredientes de minería de datos análisis estadístico, técnicas y herramientas.</p>
<p>Data Preparation for Data Mining: Dorian Pyle Morgan Kaufmann Publishers, Inc.: 1ª edition San Francisco, California</p>	<p>Dentro de este libro se consulto el capítulo de "Data Exploration as a Process"</p>
<p>Data Minig with Neural Networks - Solving Bussiness Problems from Application Development to Design Support Joseph D. Bigus Mc Graw - Hill</p>	<p>Dentro de este libro se consulto la relación que existe entre la minería de datos y las redes neuronales dentro de la primera parte del libro "The Data Mining Process Using Neural Networks"</p>
<p>Advances in Knowledge Discovery and Data Minig Usama M. Fayyad, Gregory Piatetsky - Shapiro, Padharaic Smyth and Romasamy Uthurusamy AAAI Press / The MIT Press: 1996.</p>	<p>La lectura de este libro ha sido principalmente para tomar los ejemplos que ahí se tienen.</p>
<p>The Second international Conference on Knowledge Discovery &amp; Data Minig: AAAI Press Evangelous Simoudis, Jiawei Hon &amp; Usama Foyyad Portland, Oregon.</p>	<p>Se utilizo el primer tema de este libro el cual habla a cerca de la Minería de datos como una herramienta para el descubrimiento del conocimiento.</p>
<p>Teoría de encuestas por muestreo con aplicaciones: Pandurang V. Sukhatme, Traducción de Ana María Flores y José Nieto de Pascual, Fondo de Cultura Económica 2ª. Edición, 1962.</p>	<p>Se utilizaron diversos capítulos de este libro para conocer a cerca de las diversas formas de elaborar un análisis estadístico.</p>

### 2.1.3 Lectura Rápida

Libro	Descripción
Rough Sets and Data Mining - Analysis of Imprecise Data T.Y. Lin and N. Cercone Kluwer Academic Publishers.	Este libro se consulto para conocer que se tiene a cerca de la minería de datos y el análisis de los datos.
Principles of Data Mining and Knowledge Discovery Jan Komorowski, Jan Zythow First European Symposium, PKDD; 1997.	Este libro habla a cerca de los principios de la minería de datos y el descubrimiento de l conocimiento.
Estadística Básico en Administración, Conceptos y Aplicaciones: Mark L. Berenson, David M. Levine, Traducción Julio Coro Pando, Prentice-Hall Hispanoamericana, S.A. 4ª. Edición, México, 1992, ISBN 58259.	Este libro se consulto para conocer los conceptos de la estadística moderna.

### 2.1.4 Lectura Superficial

Libro	Descripción
Research and Development in Knowledge Discovery and Data Mining Xindong Wu Ramamohanarao Kotagiri, Kevin B. Karlo Second Pacific - Asia Conference, PAKDD - 98, Melbourne Australia, Abril, 1998.	Este libro se consulto para analizar los casos prácticos a cerca de minería de datos.

## 2.2 TESIS

Tesis	Descripción
Seminario de Investigación Informática: La Minería de Datos como el Proceso para el Descubrimiento y Generación de Conocimiento en una Base de Datos María Isabel Ángeles Larrieta, Angélica María Santillan Gómez: UNAM - FCA, 1998.	Esta tesis se consulto en su totalidad para analizar que es lo que se tiene hasta este momento en minería de datos dentro de la comunidad informática de nuestro País.

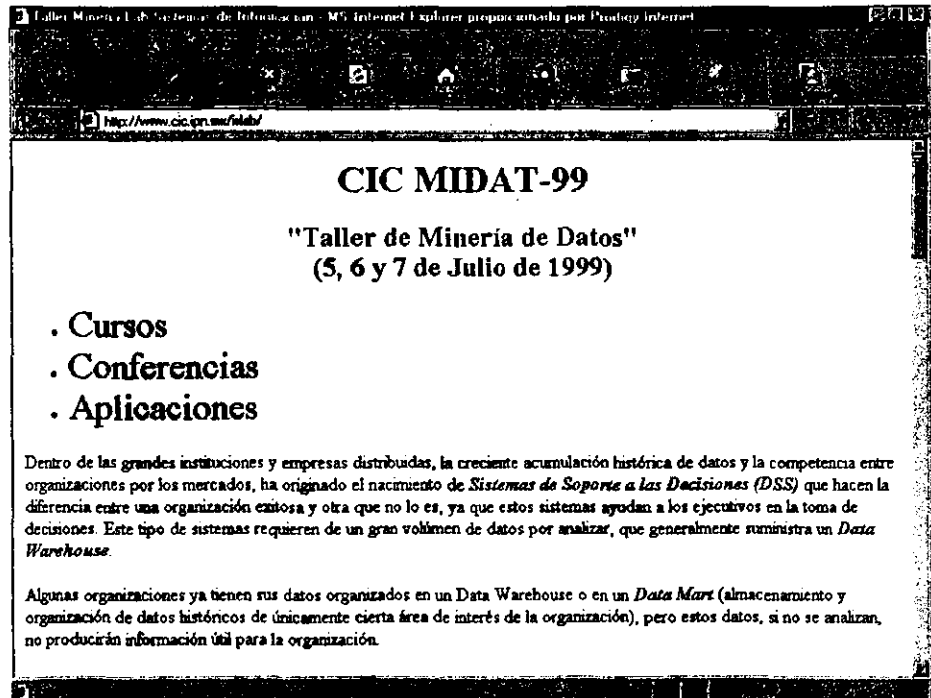
### 2.3 FOLLETOS

Folleto	Descripción
Minería de Datos con Búsqueda de Patrones de conocimiento; Centro de Investigación en Computación; subdirección de Vinculación; IPN, No. 16 Serie Roja.	La lectura de este folleto es total debido a que es lo que se ha investigado a cerca de la minería de datos dentro de nuestro país.
Estado del Arte y de la Práctica en Minería de Datos, Análisis y Crítica; Centro de Investigación en Computación; subdirección de Vinculación; IPN No. 3 Serie Verde.	La lectura de este folleto es total debido a que es lo que se ha investigado a cerca de la minería de datos dentro de nuestro país.

## 2.4 DIRECCIONES DE INTERNET

Dirección : [Http://www.cic.ipn.mx/islab/](http://www.cic.ipn.mx/islab/)

Fecha Octubre de 1999.



Esta dirección nos muestra lo que fue presentado en un taller de minería de datos, presentado en el *CIC* (Centro de Informática y Computación del IPN).

Dirección : <http://www.ictnet.es/esp/comunid/atec-inf/mining/1/default.htm>

Fecha : Octubre de 1999.



## Data Mining

[Home](#) [Comunidad](#) [Servicios](#) [Tiempo Libre](#) [Miembros](#) [Buscar](#)



**¿Perdistes todo en el  
Casino?**

Conoce nuestra Comunidad..  
¡Suscríbete a la lista de la Comunidad!

### Estudios, trabajos y artículos



Esta página es una comunidad de mineros de datos, en la cual se pueden encontrar trabajos o artículos referentes a ese tema.

Dirección: <http://w3.mor.itesm.mx/~portillo/Thesis/kdd.html>

Fecha : Enero 2000

KDD M: Internet Explorer proporcionado por Frudap Internet

<http://w3.mor.itesm.mx/~portillo/Thesis/kdd.html>

***Arquitectura Inteligente para  
Descubrimiento de Conocimiento  
en Bases de Datos***



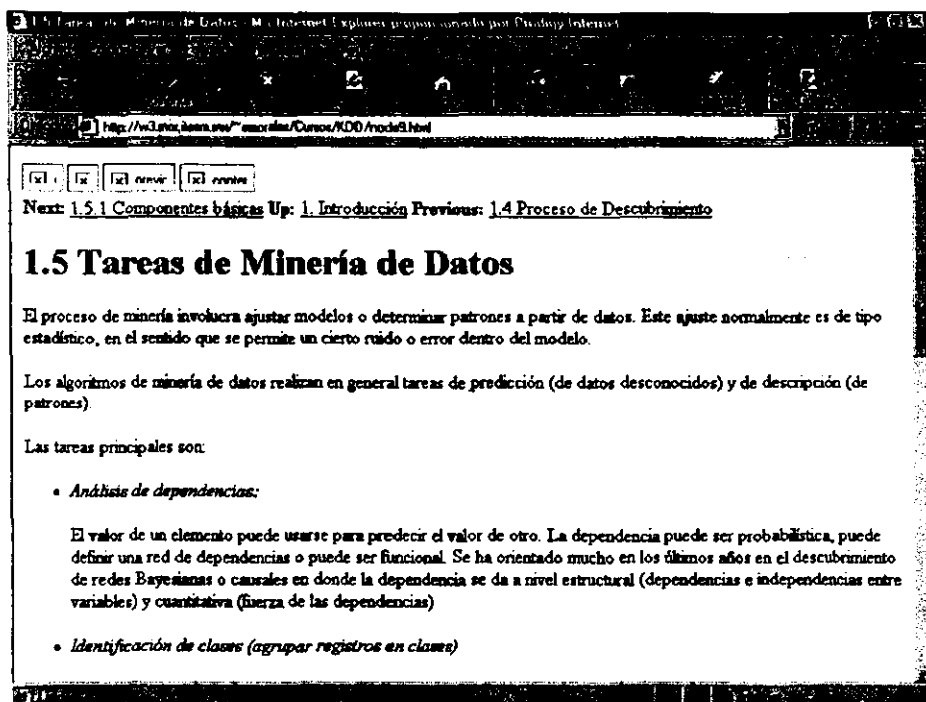
**Jesús Portillo Carbajal**  
[portillo@compuc.mor.itesm.mx](mailto:portillo@compuc.mor.itesm.mx)

**Eduardo Morales Manzanares**  
[emorales@compuc.mor.itesm.mx](mailto:emorales@compuc.mor.itesm.mx)

Investigación realizada por maestros del ITESM.

Dirección: <http://w3.mor.itesm.mx/~emoraes/Cursos/KDD/node9.html>

Fecha: noviembre de 1999.



The screenshot shows a web browser window with the address bar containing the URL <http://w3.mor.itesm.mx/~emoraes/Cursos/KDD/node9.html>. The page content includes a navigation menu with buttons for 'Inicio', 'Inicio', 'Inicio', and 'Inicio'. Below the menu, there are links for 'Next: 1.5.1 Componentes básicas', 'Up: 1. Introducción', and 'Previous: 1.4 Proceso de Descubrimiento'. The main heading is '1.5 Tareas de Minería de Datos'. The text describes the process of mining as adjusting models or determining patterns from data, often allowing for some error. It lists two main tasks: prediction and description. The prediction task is further divided into 'Análisis de dependencias' and 'Identificación de clases (agrupar registros en clases)'. The 'Análisis de dependencias' section explains that the value of one element can be used to predict the value of another, and that dependencies can be probabilistic or functional. It also mentions the use of Bayesian or causal networks to analyze dependencies at a structural level and quantify their strength.

Apuntes de la materia de Minería de Datos del Instituto Tecnológico de Monterrey (ITESM)

Dirección: <http://www.map.es/csi/silice/DW2253.html>

Fecha: noviembre de 1999.

Data Mining - Minería de Datos - MS Internet Explorer proporcionado por Prodigy Internet

<http://www.map.es/csi/silice/DW2253.html>

**2.2.5.3.- DATA MINING O MINERÍA DE DATOS**

2.2.5.3.1.- Introducción

2.2.5.3.2.- Técnicas de Data Mining

2.2.5.3.3.- Metodología de aplicación

**2.2.5.3.1.- Introducción**

El Data Mining es un proceso que, a través del descubrimiento y cuantificación de relaciones predictivas en los datos, permite transformar la información disponible en conocimiento útil de negocio. Esto es debido a que no es suficiente "navegar" por los datos para resolver los problemas de negocio, sino que se hace necesario seguir una metodología ordenada que permita obtener rendimientos tangibles de este conjunto de herramientas y técnicas de las que dispone el usuario.

Constituye por tanto una de las vías clave de explotación del Data Warehouse, dado que es este su entorno natural de trabajo.

Se trata de un concepto de explotación de naturaleza radicalmente distinta a la de los sistemas de información de gestión, dado que no se basa en coeficientes de gestión o en información altamente agregada, sino en la información de detalle contenida en el almacén. Adicionalmente, el usuario no se conforma con la mera visualización de datos, sino que trata de obtener una relación entre los mismos que tenga repercusiones en su negocio.

**2.2.5.3.2. Técnicas de Data Mining**

Apuntes sobre minería de datos (introducción, técnicas y metodología de la aplicación)



Dirección : [http://www.fcee.ulpgc.es/Acede98/acede/mesa11/11\\_05c.htm](http://www.fcee.ulpgc.es/Acede98/acede/mesa11/11_05c.htm)

Fecha : noviembre de 1999.

The screenshot shows a web browser window with the address bar containing the URL [http://www.fcee.ulpgc.es/Acede98/acede/mesa11/11\\_05c.htm](http://www.fcee.ulpgc.es/Acede98/acede/mesa11/11_05c.htm). The page content includes a header with the text "VIII CONGRESO NACIONAL DE ACCTE" and "COMISIONE VICEPRESIDENCIA INSTITUCIONAL". Below this, there is a section titled "Apuntes de DATA WAREHOUSING" and a list of authors: Enrique López, Paula Rodríguez, Francisco José Martínez López, José Luis Rodríguez López, Rafael Rodríguez Barrera, and UNIVERSIDAD DE SEVILLA. The word "ABSTRACT" is visible below the authors' names. The main body of the page contains the beginning of an abstract, starting with "Como es sabido, en el mundo de las organizaciones, las T.I. han supuesto, en primera instancia, la automatización de los procesos de carácter técnico o administrativo para, posteriormente, dar apoyo en áreas más relacionadas con la aplicación del conocimiento, experiencia y propia coordinación juegan".

*Apuntes de DATA WAREHOUSING, como un paso más hacia la gestión del conocimiento en las empresas*

## 2.5 CONGRESOS Y SEMINARIOS

Dirección : <http://www.ati.es/NOTICIAS/doc/19980414a.html>

Fecha : Octubre de 1999.

ATI - Conferencia-Coloquio: Minería de datos (Data Mining): Fundamentos y perspectivas - MS Internet Explorer (http://www.ati.es/NOTICIAS/doc/19980414a.html)

Inicio

▲ Simposio de

▼ Anuncio

### Noticias ATI

## Conferencia-Coloquio: "Minería de datos (Data Mining): Fundamentos y perspectivas"

### ATI - CAPITULO TERRITORIAL DE MADRID CONFERENCIA-COLOQUIO

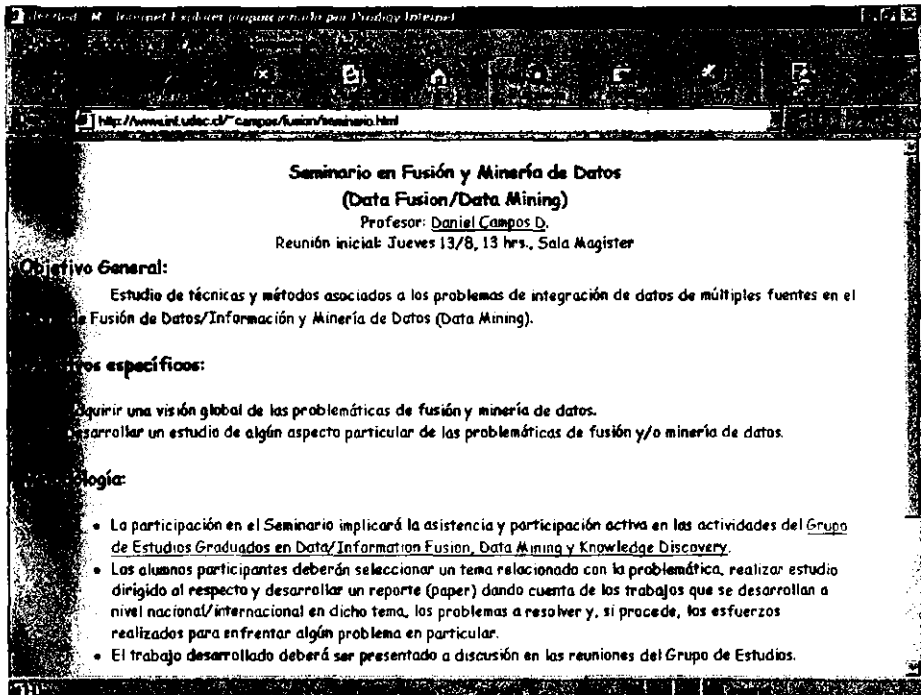
El Capítulo Territorial de Madrid de la Asociación de Técnicos de Informática, en colaboración con la Escuela de Informática de la Universidad Pontificia Comillas, tiene el placer de invitarle a la conferencia sobre el tema:

### "Minería de datos (Data Mining): Fundamentos y perspectivas"

Publicaciones de Minería de Datos, de un Seminario realizado en la ciudad de Madrid, España.

Dirección: <http://www.inf.udec.cl/~campos/fusion/seminario.html>

Fecha: Noviembre de 1999.



The image is a screenshot of a web browser window. The title bar reads "Internet Explorer (page opened via Proxy Internet)". The address bar shows the URL "http://www.inf.udec.cl/~campos/fusion/seminario.html". The main content area of the browser displays the following text:

**Seminario en Fusión y Minería de Datos  
(Data Fusion/Data Mining)**  
Profesor: Daniel Campos D.  
Reunión inicial: Jueves 13/8, 13 hrs., Sala Magister

**Objetivo General:**  
Estudio de técnicas y métodos asociados a los problemas de integración de datos de múltiples fuentes en la Fusión de Datos/Información y Minería de Datos (Data Mining).

**Objetivos específicos:**  
Adquirir una visión global de las problemáticas de fusión y minería de datos.  
Desarrollar un estudio de algún aspecto particular de las problemáticas de fusión y/o minería de datos.

**Logía:**

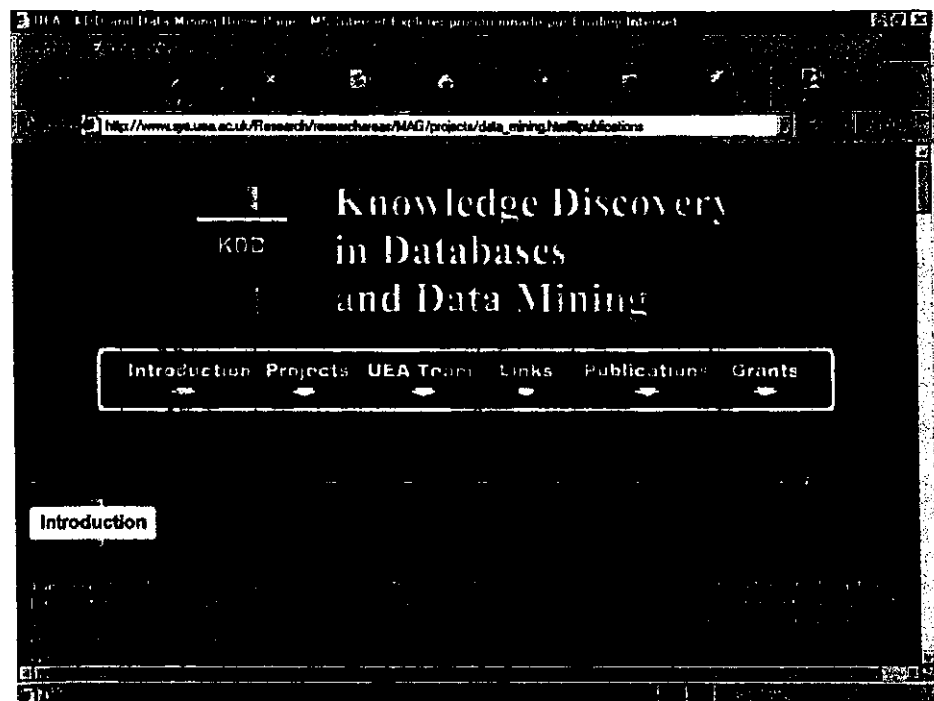
- La participación en el Seminario implicará la asistencia y participación activa en las actividades del Grupo de Estudios Graduados en Data/Information Fusion, Data Mining y Knowledge Discovery.
- Los alumnos participantes deberán seleccionar un tema relacionado con la problemática, realizar estudio dirigido al respecto y desarrollar un reporte (paper) dando cuenta de los trabajos que se desarrollan a nivel nacional/internacional en dicho tema, los problemas a resolver y, si procede, los esfuerzos realizados para enfrentar algún problema en particular.
- El trabajo desarrollado deberá ser presentado a discusión en las reuniones del Grupo de Estudios.

**Información a cerca de un seminario de la "Fusión de la Información y de la Minería de Datos."**

## 2.6 DESARROLLO DE SOFTWARE ESPECIALIZADO

**Dirección:** [http://www.sys.uea.ac.uk/Research/researchareas/MAG/projects/data\\_mining.html](http://www.sys.uea.ac.uk/Research/researchareas/MAG/projects/data_mining.html)

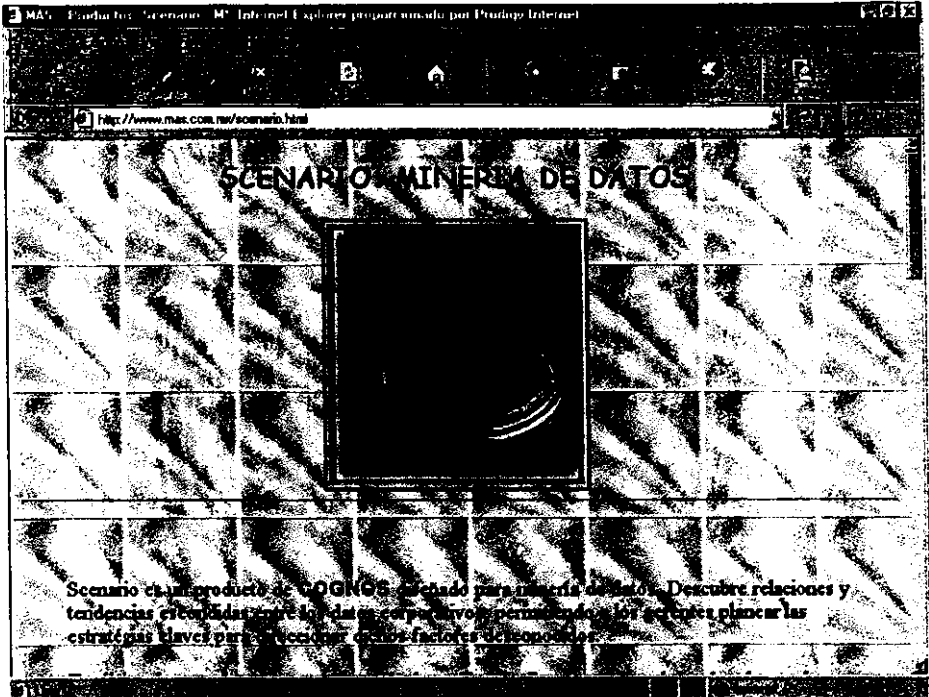
**Fecha:** noviembre de 1999.



Documentales e informes de investigaciones y software de minería de datos,

Dirección: <http://www.mas.com.mx/scenario.html>

Fecha: noviembre de 1999.



Software de minería de datos de datos realizado por una empresa de consultoría mexicana.

Dirección: <http://www.isid.com/home.htm>

Fecha: Noviembre de 1999.

2 | PÁGINA DE INICIO DEL SISTEMA DE INFORMACIÓN Y DOCUMENTACIÓN - MS Internet Explorer - página creada por Eudwig L. | 1/1/99

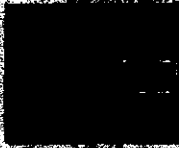


http://www.isid.com/home.htm



ANCHURA OPTIMA DEL NAVEGADOR

### VISUAL DATA MINING



#### Spotfire Pro

\* La problemática de las empresas de todo el mundo es encontrar una forma de almacenamiento y explotación de cantidades ingentes de datos que permitan día a día



**Trabajamos para Usted y su Empresa**  
En ISID somos conscientes del alto grado de competitividad que existe hoy día en el entorno empresarial tanto dentro, como fuera de nuestras fronteras. Por ello, creemos que los Sistemas de Información y tecnologías relacionadas, juegan un papel estratégico en los procesos diarios de las empresas. La correcta planificación, y su posterior gestión, son las claves para el éxito de un Sistema de Información.

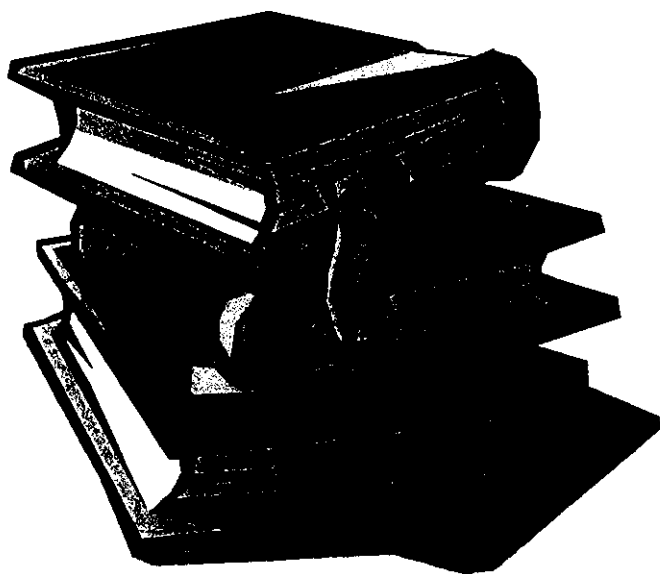
*"Confrontos su problemática... Podemos ayudarle con la solución tecnológica más adecuada a sus requerimientos y necesidades."*



#### Cómo Podemos Ayudarle: Actividades, Soluciones y Servicios

Ni mucho menos queremos compararnos con un genio como Leonardo da Vinci, al que corresponde esta imagen, pero sí queremos transmitir nuestra dedicación, trabajo diario, creatividad e inquietud, con el ánimo de posicionar a nuestra empresa en un

Software de minería de datos con ejemplos de aplicación



# Marco Conceptual

---

### 3. MARCO CONCEPTUAL

#### 3.1 ANTECEDENTES<sup>1</sup>

Desde hace muchos años el ser humano se ha preocupado por el manejo, almacenamiento y distribución de la información; debido a esto ha ideado formas que se adaptan a las necesidades que cada persona; por ejemplo antiguamente si una persona contaba con dinero posiblemente lo almacenaba debajo del "colchón" y en el tallo de un árbol registraba la cantidad de dinero que guardaba; pero esto podría no ser un método muy confiable, por lo que comenzó a idear otros métodos más eficientes para contabilizar el dinero que guardaba, porque ahora ya hemos llegado a utilizar software especializado.

En general se puede decir que el origen del interés de los datos, se da desde el surgimiento de las primeras ciudades, los principios de la producción en masa, la especialización de la mano de obra, el empleo del dinero, y la posibilidad de alquilar servicios y productos para las necesidades de vida; todo esto requería la conservación de datos en registros confiables. Conforme ha pasado el tiempo se han conservado diversos tipos de datos y registros diferentes, los cuales incluyen calendarios, datos de censos, investigaciones especializadas, registros de propiedad de la tierra, de nacimiento, o de matrimonio, árboles genealógicos, entre otros.

En el siglo XVII las personas se empezaron a interesar por dispositivos en los que pudieran procesar sus datos automáticamente, aunque fuera de forma rudimentaria. El desarrollo de dichos dispositivos de cómputo y tabulación prácticos requirió dos cosas:

1. Una capacidad de almacenamiento y procesamiento de datos a gran escala.
2. Un cierto nivel básico de capacidades de manufactura y dispositivos mecánicos y eléctricos.

Hacia fines del siglo XIX, las capacidades de maquinado habían llegado a un nivel de avance suficiente y ocurrió también una necesidad que comprendía una cantidad masiva de datos.

---

<sup>1</sup> Larry Long: "Introducción a las computadoras y al Procesamiento de Información", 2ª. Edición, Prentice Hall Hispanoamericana, México, 1988.



---

De forma paralela al crecimiento del equipo para procesar datos se dio el desarrollo de nuevos medios para almacenarlos. La primera forma moderna de almacenamiento de datos fueron las perforaciones en algún medio adecuado de papel.

En la década de 1870 y 1880 la cinta de papel perforada se presentó junto con los primeros equipos de teletipo. Herman Hollerith hacia 1890 y James Powers a principios de este siglo utilizaron tarjetas perforadas como medio de almacenamiento.

Hacia 1936 Eastman Kodak como Bush, con su Rapid Selector, registraban datos digitales en forma fotográfica y aunque estos podían leerse de nuevo con tecnología de celdas fotoeléctricas, éste era un medio no borrable.

Dentro de la segunda mitad de la década de 1930 se inició la era de los medios de almacenamiento magnéticos con los experimentos Bell Laboratories con cinta magnética para almacenamiento de sonido; en 1937 se realizaron trabajos, en Alemania, con óxidos para utilizarlos en el almacenamiento magnético.

En 1947 Eckert y Mauchly desarrollaron una unidad de cinta magnética.

En 1952, tanto la UNIVAC como Raytheon ofrecieron en el mercado en forma comercial unidades de cinta magnética.

En 1968, Ascher Olper, que en ese momento pertenecía la personal de IBM, advirtió que era necesario unificar el conocimiento sobre la programación de manera que estuviese al alcance de todos los programadores de sistemas. En respuesta a la petición hecha por de Olper IBM, decide patrocinar la serie sobre programación de sistemas, un ambicioso proyecto cuya finalidad era reunir, organizar y publicar los principios y las técnicas que tuvieran utilidad permanente en la industria en general.

El hecho de asociar una estructura de datos con una relación fue identificado por investigadores y documentadores en el año de 1967, por R.E. Levein y M.E. Maron. Posteriormente D.L. Childs, en 1968, efectuó una siguiente publicación al respecto. Pero es hasta después del artículo de junio

---

de 1970 en el volumen 13, número 6 de *Communications of the ACM* donde aparece el artículo "A Relational Model of Data for Large Shared Data Banks" de Eddgar F. Codd que da por iniciado el modelo de datos relacional.

Este artículo fue el resultado de un trabajo que se inició en 1968 en el IBM Research Laboratory en San José, California, con un modelo abstracto de información. El objetivo era encontrar un fundamento teórico de los diferentes aspectos de un Data Base Management System completamente ajeno de los aspectos de un proceso físico de una máquina o CPU en particular. Este modelo es el que actualmente se conoce como modelo relacional o estructuras de datos relacionales.

En 1969, el Dr. E.F. Codd publicó el primer documento que define un modelo de bases de datos basado sobre el concepto matemático de los conjuntos relacionales. El Modelo de Bases de Datos Relacional (RDBM) ha sido refinado constantemente desde entonces, y mucho más a partir de la publicación que hizo Codd de las "12 reglas" para las bases relacionales en 1985, y en 1990 su libro que define la versión 2 (RV/2) del modelo relacional mediante 333 reglas que son un subconjunto y expansiones de las 12 originales.

A continuación se mencionarán las 12 reglas de Codd.<sup>2</sup>

- ⊙ *Regla 0*: "Cualquier DBMS que proclame ser relacional, deberá manejar, completamente, las bases de datos por medio de sus capacidades relacionales"
- ⊙ *Regla 1 (The Information Rule)*: "Toda la información dentro de una Base de datos relacional se representa de manera explícita a nivel lógico y exactamente de una sola manera, como valores en una tabla".
- ⊙ *Regla 2 (Guaranteed Access Rule)*: "Se garantiza que todos y cada uno de los datos (valor atómico) en una base de datos relacional pueden ser leídos recurriendo a una combinación de nombre de la tabla, valor de la llave primaria y nombre de la columna".
- ⊙ *Regla 3 (Systematic Treatment of Null Values)*: "En un DBMS totalmente relacional se soportan los valores nulos (que son distintos de una cadena de caracteres vacía o de una cadena con caracteres en blanco o de cero o cualquier otro número), para representar información faltante o no aplicable de una forma consistente, independiente del tipo de dato".

---

<sup>2</sup> Rubén Adad, Alfredo Careaga, Miguel A. Medina: "Fundamentos de las Estructuras de Datos Relacionales", Noriega Editores, México, 1992.

- ☺ **Regla 4 (Dynamic on-line Catalog Based on the Relational Model):** "La descripción de la Base de datos se representa en el nivel lógico de la misma forma que los datos ordinarios, de tal suerte que los usuarios autorizados puedan aplicar el mismo lenguaje relacional para consultarla, que aquel que emplean con sus datos habituales".
- ☺ **Regla 5 (Comprehensive Data Sublanguage Rule):** Que cuente con un sublenguaje de datos completo. a)Definición de datos, b)Definición de vistas, c)Manipulación de datos, d)Restricciones de integridad, e)Autorización, f)Inicio y fin de una transacción.
- ☺ **Regla 6 (View Updating Rule):** "Todas las vistas que teóricamente sean actualizables, deberán ser actualizadas por medio del sistema"
- ☺ **Regla 7 (High-Level Insert, Update, and Delete):** "La posibilidad de manejar una relación base o una relación derivada como un solo operador se aplica a la lectura, inserción, modificación y eliminación de datos".
- ☺ **Regla 8 (Physical Data Independence):** "Los programas aplicativos y la actividad en terminales no deberán ser afectados por cambios en el almacenamiento físico de los datos o en los métodos de acceso".
- ☺ **Regla 9 (Logical Data Independence):** "Los programas aplicativos, y la actividad en terminales no deberán ser afectados por cambios de cualquier tipo, que preserven la información y que teóricamente permitan la no afectación, en las tablas base".
- ☺ **Regla 10 (Integrity Independence):** "Las restricciones de integridad de una Base de datos deberán poder definirse con el mismo sublenguaje de datos relacional y deberá almacenarse en el catálogo, no en los programas aplicativos".
- ☺ **Regla 11 (Distribution Independence):** "Un DBMS relacional debe tener independencia de distribución".

A mediados de los años 70's IBM desarrollo como lenguaje estándar el Lenguaje de Consulta Estructurado (SQL), fue diseñado como un lenguaje para acceder explícitamente a DBMS's basados en el modelo relacional que corría sobre mainframes. La versión inicial del lenguaje apareció como SEQUEL, posteriormente se nombre fue reducido a SQL.

En la actualidad nos podemos caracterizar por generar, recopilar y procesar información. La cantidad de actividades de este tipo que realizamos hoy en día no tiene precedentes; debido a que el volumen de estas aumenta a un ritmo vertiginoso, y así continuará en el futuro; por lo que también los avances

---

tecnológicos han ido aumentando conforme a las necesidades de almacenamiento.

Lo que se ha buscado desde los inicios del manejo de los datos es:

1. Que deben representarse y almacenarse en cierta forma para accederlos posteriormente.
2. Deben organizarse de manera adecuada para accederlos selectiva y eficientemente.
3. Deben procesarse y presentarse de tal manera que puedan apoyar eficientemente al usuario.
4. Deben protegerse y manejarse para que no pierdan su valor.

Actualmente no solo se almacena información; esta tecnología en desarrollo maneja: fotografías, grabaciones de voz, archivos de vídeo y otros tipos de información.

# MONOGRAFÍA MINERÍA DE DATOS

## Historia

Surge de la necesidad de explorar y explotar al 100% los recursos de una base de datos, y sobre todo por curiosidad de una persona por conocer qué es lo que podría ocultar una base de datos y qué puede llegar a ser información de gran utilidad para una empresa.

## Antecedentes

La evolución de los Manejadores de Base de Datos se da de la siguiente forma:

- ❖ **SQL:** Lenguaje para el acceso a la información almacenada en una base de datos relacional.
- ❖ **Data Ware Housing:** Conjunto de datos integrados orientados a una materia, que varían con el tiempo y que son transitorios, los cuales soportan el proceso de toma de decisiones de una administración.
- ❖ **Data Mining:** Búsqueda de relaciones insospechadas, de información histórica almacenada en una base de datos, por medio de técnicas semiautomáticas.

## Utilidad

La minería de datos auxilia a los usuarios empresariales en el procesamiento de vastas reservas de datos para descubrir "relaciones insospechadas", por ejemplo, entre productos y clientes o patrones de compra de los clientes. La meta de la minería de datos es descubrir "revelaciones estratégicas competitivas" para controlar la participación en el mercado y las utilidades. Los seres humanos tenemos la agudeza para percibir

excepciones y anomalías, pero no tenemos la potencial capacidad para inferir relaciones en volúmenes de datos.

## Concepto

En Estados Unidos se conoce con el término minería de datos. En nuestro País se puede llegar a conceptualizar como: *"una combinación de técnicas semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos y que pueda resultar más provechosa que los propios datos de partida."*

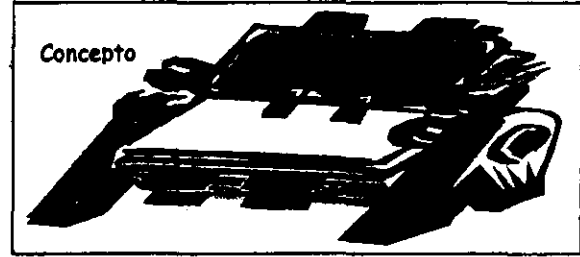
## Usuarios potenciales

Los usuarios clave son los analistas empresariales, los peritos en estadísticas y los profesionales en tecnología de la información que auxilian a los usuarios empresariales. Quienes obtienen los beneficios de los resultados de minería de datos son los gerentes empresariales y los ejecutivos, que desean entender los factores de éxito del negocio con base en datos completos del cliente.

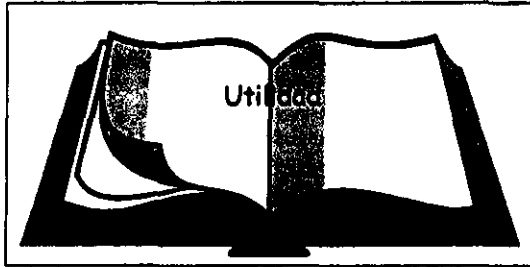
## Tecnologías y herramientas

Se pueden clasificar en:

- ✓ **Análisis estadístico de datos:** Detectar patrones no usuales de datos.
- ✓ **Descubrimiento de conocimientos:** Extraer de los datos información implícita, no trivial, que no se conocía y potencialmente útil.
- ✓ **Otros:** sistemas de visualización, sistemas de información geográfica, análisis fractal y herramientas propietario.



Usuarios potenciales



Técnicas y Herramientas



**CUADRO SINÓPTICO**

Minería de Datos

Por su área de aplicación	<ul style="list-style-type: none"> <li>En empresas</li> <li>En consultorías</li> <li>En universidades</li> </ul>	<ul style="list-style-type: none"> <li>Sector público</li> <li>Sector privado</li> <li>De servicios</li> <li>Nacionales</li> <li>Extranjeras</li> <li>Públicas</li> <li>Privadas</li> </ul>
Por su finalidad	<ul style="list-style-type: none"> <li>Desarrollo de nuevos sistemas</li> <li>Para conocimiento de un área en específico de la empresa</li> </ul>	
Por quien lo realiza	<ul style="list-style-type: none"> <li>Por empleados de la empresa (información confidencial)</li> <li>Por empresas especializadas</li> </ul>	<ul style="list-style-type: none"> <li>Consultorías</li> <li>Fabricantes de software especializado</li> </ul>
Por las herramientas que se utilizan	<ul style="list-style-type: none"> <li>Conocimientos técnicos, personas dedicadas a esta área.</li> <li>Software especializado</li> </ul>	<ul style="list-style-type: none"> <li>Análisis estadístico</li> <li>Descubrimiento de conocimientos (IA)</li> <li>Propietario</li> </ul>
Por quién lo utiliza	<ul style="list-style-type: none"> <li>Analistas empresariales</li> <li>Peritos en estadísticas</li> <li>Profesionales en tecnologías de información</li> </ul>	
Por quienes obtienen sus beneficios	<ul style="list-style-type: none"> <li>Directivos</li> <li>Gerentes empresariales</li> <li>Ejecutivos del negocio</li> </ul>	
Por las técnicas que se utilizan	<ul style="list-style-type: none"> <li>Redes Neuronales</li> <li>Descubrimiento de asociaciones</li> <li>Clasificación</li> <li>Clustering</li> <li>Descubrimiento secuencial</li> </ul>	

## 3.2 DEFINICIONES

### Etimológicas

Término	Definición
Del griego Myo	Lo oculto
Mina Mena	Sustantivo. Del latín "minari"
Mena	De latín "minari", llevar, conducir
Mina	Sustantivo. Excavación subterránea o a cielo abierto para extraer un mineral.
Minería	Arte de laborear las minas. Conjunto de los individuos que se dedican a ello.
Minero	Adjetivo. Persona que trabaja en las minas

*Derivados:* Minador, minal, minar, minera, mineral, minería, minero

Término	Definición
De	Proviene del latín. Preposición Elemento de relación que introduce tanto complementos del verbo como del nombre: Denota: propiedad, posesión y pertenencia: <i>casa de mi padre; los árboles del jardín.</i> <sup>3</sup>

Término	Definición
Dato	Del latín "datum". Sustantivo. Lo que se da. Antecedente que se da para llegar a un conocimiento. <sup>4</sup>

<sup>3</sup> *Enciclopedia Microsoft Encarta 99. VOX - Diccionario General de la Lengua Española, 1997*  
Bibliograf, S.A., Barcelona.






<sup>4</sup> *Gran Diccionario Enciclopédico Ilustrado de Selecciones del Reader's Digest, México, 1972.*



 Sinónimos

Palabra	Sinónimo <sup>4</sup>
Mina	Almacén Excavación Filón Galería Túnel Yacimiento
Minar	Excavar Horadar Socavar
Dato	Antecedente Detalle Documento Nota Noticia

 De diccionarios<sup>5</sup>

Término	Definición
Minería	<p> Excavación que hace por pozos, galerías y socavaciones a cielo abierto para extraer mineral.</p> <p> Aquello que abunda en cosas dignas de aprecio o de que puede sacarse algún provecho o utilidad.</p> <p> Oficio, empleo o negocio de que con poco trabajo se saca mucho interés y ganancia.</p>
Datos	<p> Antecedente necesario para llegar al conocimiento exacto de una cosa.</p> <p> Documento (que aclara).</p>

<sup>4</sup> Larousse de la Conjugación, Sinónimos y Antónimos

<sup>5</sup> Gran Diccionario Enciclopédico Ilustrado de Selecciones del Reader's Digest, México, 1972.

Autor	Definición
Dr. Adolfo Guzmán Arenas	<p>"Un sistema de minería de datos está formado por varios programas de cómputo que realizan la búsqueda en una base de datos, de manera automática, de tendencias, desviaciones, anomalías, patrones y situaciones "interesantes". Estas desviaciones o tendencias a menudo son reportadas inmediatamente, o más a menudo en un archivo, para su posterior visualización y decisión final. Existen varios algoritmos generales para "minería de datos". El sistema a menudo posee un configurador que permite particularizarlos a casos y situaciones específicas; una agenda de trabajo guía a cada minero en sus búsquedas. La prioridad de búsqueda (orden en la agenda), los criterios de "interés" y "tolerancia" y las definiciones de "situación anómala" son modificables por el usuario, de suerte que le minero originalmente posee un criterio general pero quizá fuera de foco de lo que es "interesante" para el usuario, y termina como una colección de límites de valores y criterios de éxito muy específicos, que busca en determinada área de la base de datos y bajo condiciones o predicados igualmente ajustados a la medida".<sup>6</sup></p>
Joseph P. Bigus	<p>"Minería de datos es el descubrimiento eficiente de información valiosa, no - obvia de una gran colección de datos; cuyo objetivo es ayudar a buscar situaciones interesantes con los criterios correctos, complementar enormemente una labor que hasta ahora se ha considerado "intelectual" y de alto nivel, privativa de los gerentes planificadores y administradores. Además, de realizar la búsqueda fuera de horas pico, usando tiempos de máquina excedentes".<sup>7</sup></p>
William J. Frawley, Gregory Piatetski - Shapiro y Cristoper J. Matheus	<p>"Minería de datos o recuperación del conocimiento en bancos de datos como también se le conoce, es la extracción implícita no trivial previamente conocida y potencialmente útil de los datos. Esto comprende un</p>

<sup>6</sup> Estado del Arte y de la Práctica en Minería de Datos, Análisis y Crítica; Dr. Adolfo Guzmán Arenas

<sup>7</sup> Minería de datos con Búsqueda de Patrones de Comportamiento; Dr. Adolfo Guzmán Arenas

	número diferente de aproximaciones técnicas diferentes como índices, resumen de la información, reglas de clasificación del aprendizaje, dependencia de redes, análisis de cambios y detección de anomalías" <sup>8</sup>
Harjinder S. Gill y Prakash C. Rao	"La minería de datos apoya la modalidad de descubrimiento del soporte de decisiones. Las herramientas de minería de datos recorren los datos detallados de transacciones para desenterrar patrones y asociaciones ocultos. Por lo regular los resultados generan extensos reportes o de les analiza con herramientas de visualización de datos descubiertos". <sup>9</sup>
Macel Holshemire y Arno Siebes	"Minería de datos es la búsqueda de relaciones y patrones globales que existen en grandes bases de datos pero ocultos entre gran cantidad de datos, tal como una relación entre datos del paciente y el diagnóstico de su médico. Estas relaciones representan valiosos conocimiento acerca del banco de datos y los objetos en el banco de datos". <sup>4</sup>

### Propia

*"La minería de datos la considero como una combinación de técnicas semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos y que pueda resultar más provechosa que los propios datos de partida".*

<sup>8</sup> Advances in Knowledge Discovery and Data Mining

<sup>9</sup> Data Warehousing: La integración de la información para la mejor toma de decisiones

### 3.3 EVOLUCIÓN

La minería de datos es el resultado de un largo proceso de búsqueda. Esta evolución inició con el almacenamiento de datos en computadoras, continuó con el mejoramiento en el acceso de los datos almacenados, y más recientemente con la generación de tecnologías que permita a los usuarios navegar a través de sus datos en tiempo real.

En la evolución de la tecnología de información cada nuevo paso fue construido a partir de uno previo. Lo cual lo podemos observar en el acceso dinámico a los datos que fue un antecedente necesario para la navegación de aplicaciones y en la habilidad para almacenar grandes bases de datos que fue el antecedente de la minería de datos.

El siguiente cuadro muestra los pasos evolutivos de la minería de datos ✓

Pasos evolutivos	Cuestionamientos de la empresa	Características	Tecnología disponible
Almacenamiento de los datos (1960's)	¿Cuál fue mi ganancia total en los últimos cinco años?	Retrospectiva, datos estáticos distribuidos	Computadoras, cintas y discos
Acceso a los datos (1980's)	¿Cuáles fueron mis ventas totales en la sucursal Centro el mes de marzo?	Retrospectiva, datos dinámicos distribuidos a nivel registro	Bases de datos relacionales (RDBMS), Lenguajes de consulta estructurada (SQL).
Data warehousing y soporte de decisiones (1990's)	¿Cuál fue la diferencia de ventas totales en la sucursal Centro con relación a la sucursal Sur en el mes de Marzo?	Retrospectiva, datos distribuidos dinámicos a múltiples niveles.	Procesamiento analítico en línea (OLAP), Bases de datos multidimensionales, Data warehouse.
Minería de datos	¿Qué es lo más	Prospectiva,	Algoritmos

✓ [http://www.pilotsw.com/r\\_and\\_t/whpaper/datamine/dmfrnd.htm](http://www.pilotsw.com/r_and_t/whpaper/datamine/dmfrnd.htm)

(Desarrollándose hoy en día)	probable que pase en las ventas de la sucursal Centro en el próximo mes de Junio? ¿Porqué?	información proactiva distribuida.	avanzados, computadoras multiproceso, bases de datos masivas.
------------------------------	--	------------------------------------	---

A continuación se hace una descripción de cada uno de los pasos evolutivos de la minería de datos.

#### a. Almacenamiento de los datos (1960's)

En un principio las empresas tenían como principal preocupación el almacenamiento de sus datos, debido a que generaban grandes cantidades de los mismos era necesario diseñar alguna forma en la que pudieran ser guardados; por lo que se creó el concepto de banco de datos, el cuál es descrito a continuación:

- ⊕ *Banco de Datos:* son un conjunto de registros listados uno después de otro, almacenados en archivos. Al hablar de bancos de datos se ha considerado que estos existen totalmente independientes entre sí. En la realidad, el manejo de combinaciones de tales archivos resulta difícil porque los datos se almacenan en diferentes formatos y en distintos archivos, una de las características de los archivos es que los datos no se pueden compartir entre los diferentes programas, a menudo no son recuperables ni están seguros.

Hay cuatro acciones que pueden registrarse con un archivo: 1) simplemente consultarlo sin modificarlo, 2) modificar un registro, 3) insertar un nuevo registro y 4) borrar un registro existente.

Existen varias formas de organizar y acceder archivos para su recuperación subsecuente.

#### Organización de archivos:

- ⊕ *Archivos secuenciales:* La forma más simple de almacenar un conjunto de registros es en una lista larga. Si los registros están en secuencia, en términos de uno o más campos, se dice que el arreglo es un archivo secuencial. En un archivo secuencial la única forma de recuperar los datos es empezar al inicio de archivo y leer un registro después de

otro hasta llegar al final.

- Ⓢ *Archivos no secuenciales*: Tienen la ventaja de que no es necesario dedicar tiempo o esfuerzo para mantener el archivo en secuencia física al unir, clasificar o realizar cualquier otra manipulación, ya sea al crear el archivo por primera vez, o cuando más adelante se agregan nuevos registros al final del archivo.
- Ⓢ *Archivos indizados*: Se espera tener acceso a esos registros en forma directa, se espera recuperar aquellos registros en secuencia de acuerdo con algún campo, ya sea que el archivo esté o no almacenado en secuencia física, de acuerdo con ese u otro campo.

Desdichadamente la sencillez de los bancos de datos (independientes), deja mucho que desear en cuanto a eficiencia a algunas situaciones comunes del procesamiento y almacenamiento de datos.<sup>10</sup>

#### b. Acceso a los datos (1980's)

El concepto de bases de datos surgió a partir del manejo de combinaciones de los bancos de datos, llevando a un concepto más poderoso de almacenamiento de datos.

- Ⓢ *Bases de datos*: una base de datos es una colección de datos, que guardan una relación entre sí. Un sistema de bases de datos, es un sistema de mantenimiento de registros basado en computadoras, es decir, un sistema cuyo propósito general es registrar y mantener información. Un sistema de bases de datos incluye cuatro componentes principales: datos, hardware, software y usuarios.

Los puntos clave que se tomaron en cuenta para la creación de bases de datos fueron: ciertas características del medio ambiente de procesamiento de datos, los problemas asociados para almacenar datos redundantes, la capacidad para almacenar datos que tiene que ver con relaciones múltiples y el concepto de independencia de datos.

Conforme creció el procesamiento de datos empezaron a cambiar cierto número de reglas básicas. El hardware se volvió más barato, el desarrollo del software tomó una forma más estandarizada y estructurada, y se

<sup>10</sup> Fundamentos de la Estructuras de datos relacionales, Adad Rubén, Careaga Alfredo y Miguel Ángel Medina, MEGABYTE, México, 1983.

acumularon muchas aplicaciones nuevas que debían implantarse.

Con el medio ambiente de las bases de datos se logró compartir archivos de datos entre distintas aplicaciones y emplear descripción de datos y proposiciones de acceso estandarizada de todo el proceso de desarrollo de la aplicación.<sup>11</sup>

Con esto se puede decir que las bases de datos proporcionan el marco para tratar los datos como un recurso estandarizado, administrable y compatible, preocupándose por la seguridad, respaldo y recuperación, concurrencia y capacidad de auditoría de los datos.

⊕ **DBMS (Data Base Management System):** un sistema manejador de bases de datos (DBMS) es probablemente mejor definido como una pieza de software, la cual soporta la creación, manipulación y administración de sistemas de bases de datos. Un sistema de bases de datos constituye en sí mismo un sistema completo de información.

Los elementos de un DBMS son:

1. Lenguaje de definición de datos (DDL Data Definition Language)
2. Diccionario de datos (DD Data Dictionary)
3. Lenguaje de manipulación de datos (DML Data Manipulation Language)

Un sistema de bases de datos consta de dos partes: el "Sistema Manejador de Bases de Datos (DBMS)", el cual es el programa que organiza y mantiene estas listas de información, y la "Aplicación de Bases de Datos", la cual es un programa que nos permite recuperar, consultar y modificar la información almacenada por el DBMS.

Para llevar a cabo el almacenamiento de datos, el DBMS cuenta con algún tipo de servicio de definición de datos para definir registros y campos en la base de datos. Además se necesita un mecanismo interno que mantenga os datos en el disco y conozca en dónde reside cada elemento. Un DBMS provee los siguientes servicios:

<sup>11</sup> [http://www.pilotsw.com/r\\_and\\_t/whtpaper/datamine/dmfnd.htm](http://www.pilotsw.com/r_and_t/whtpaper/datamine/dmfnd.htm)

<sup>12</sup> A. Luis Joyanes Aguilar: Glosario de términos y siglas Diccionario Inglés.

<sup>13</sup> <http://rcr.csun.edu/ChrisJ/computers/sqk/sqlch3.htm>

<sup>14</sup> Mark L. Gillenson, Introducción a las bases de datos, IBM System Research Institute.

<sup>15</sup> [http://users.demag.rwth-aachen.de/donald/Diversen/W3Encyc/\\_quer003.htm](http://users.demag.rwth-aachen.de/donald/Diversen/W3Encyc/_quer003.htm)

1. Definición de datos: Provee un método de definición y almacenamiento de una cierta cantidad de datos.
2. Mantenimiento de datos: Mantiene los datos utilizando un registro para cada elemento, los campos contienen información particular que describe cada elemento.
3. Manipulación de datos: provee servicios que permiten al usuario insertar, modificar, borrar y ordenar los datos de la base de datos.
4. Integridad de datos: Provee uno o más métodos para asegurarse que los datos son correctos.
5. Despliegue de datos: Provee algún método para presentarle los datos al usuario.

⊕ *Modelos de bases de datos:* Existen cuatro tipos principales de modelos de bases de datos para definir registros o entidades y las relaciones que guardan, esto es, las estructuras de las bases de datos para visualizar y manipular los datos a nivel lógico:

#### 1. Sistema Manejador de Archivos (FMS File Management System)

Es el modelo de base de datos más sencillo de comprender y es el único que describe como los datos están almacenados en el disco. En este modelo, cada campo o dato es almacenado secuencialmente sobre el disco en un archivo muy largo. El sistema manejador de archivos fue el primer método utilizado para almacenar datos en una base de datos computarizada. Las desventajas de este modelo son claras; primero no hay indicación de la relación entre los elementos más que la secuencia de almacenamiento. El programador y a veces el usuario tiene que conocer exactamente como los datos están almacenados en el archivo y el orden en el que pueden ser accedados. La segunda desventaja del FMS es que crea problemas con la integridad de los datos; los valores de todos los archivos tienen que ser verificados por el programa de aplicación antes de ser almacenados en el disco, además que no hay forma de buscar un registro específico rápidamente, cada búsqueda debe comenzar desde el principio del archivo examinando registro por registro. La desventaja más grande del FMS es que no permite cambiar fácilmente la estructura de la base de datos.



## 2. Sistema de Bases de Datos Jerárquico (Hierarchical Database System)

En este modelo los datos son organizados en una estructura de árbol que se origina desde una raíz. Cada clase de datos es localizada en diferentes niveles de una rama particular que proviene de la raíz. La estructura de datos en cada nivel de clase es llamado nodo; si no nacen de él nuevas ramas el último nodo en la serie es considerado una hoja.

En términos de bases de datos, la estructura de árbol define las relaciones padre - hijo y hermano entre los distintos elementos en la base de datos y esto muestra una ventaja sobre los sistemas manejadores de archivos (FMS), ya que permiten definir las relaciones uno a muchos. Además la estructura jerárquica hace fácil y rápida la búsqueda de datos.

En una estructura jerárquica siempre existe un nodo raíz único el cual es generalmente propietario por el sistema o el DBMS. Los nodos de nivel 1 representan una clase particular de datos. Cada nodo de nivel 1 puede tener uno o más hijos de nivel 2.

La estructura física de los datos en el disco no tiene importancia en el modelo jerárquico: el DBMS puede almacenar los datos como una lista ligada de campos, con apuntadores que van del padre al hijo y de rama a rama, finalizando en un valor nulo o apuntador terminal en la última hoja.

Este diseño facilita la adición de nuevos nodos en cualquier nivel, ya que el DBMS únicamente tiene que modificar el apuntador terminal al siguiente nodo de la rama en la lista. Por conveniencia, podemos definir un registro como un padre y todos sus hijos.

El primer problema radica en la estructura inicial de la base de datos, la cual es arbitraria y debe ser definida por el programador cuando la base de datos es creada. La relación padre - hijo no puede ser modificada sin rediseñar la estructura. Otro problema creado por la rigidez de la estructura jerárquica, es la dificultad para modificar la definición de los niveles de clases, ya que se tiene que redefinir la estructura.

La desventaja más significativa de este modelo es que no provee un método de definición sencillo para el uso de relaciones muchos a muchos;

una solución a este problema es el almacenamiento de múltiples copias del mismo dato en múltiples niveles. Otro enfoque de solución al problema de la relación muchos a muchos, es ir aumentando relaciones padre - hijo secundarias y apunadores en la estructura jerárquica.

### 3. Sistema de Base de Datos en Red (Network Database System)

Las primeras especificaciones fueron escritas en 1971 por CODASYL. Hay que notar, que el término "red" no tiene relación con el medio físico en el que actualmente corren las bases de datos, el modelo de red define conceptualmente las bases de datos en las cuales existen las relaciones muchos a muchos. Las relaciones entre los diferentes datos son referidas comúnmente como conjuntos que los distinguen estrictamente de las relaciones padre-hijo definidas en el modelo jerárquico.

Un modelo de base de datos de red se identifica por líneas o apunadores cíclicos para mapear las relaciones entre los diferentes elementos de datos. El modelo de red puede ser utilizado para describir relaciones cada vez más complejas. La flexibilidad del modelo de red en las relaciones muchos a muchos es fuerte.

Las interrelaciones entre los diferentes conjuntos pueden convertirse en un modelo cada vez más complejo y difícil de mapear. Tal como las bases de datos jerárquicas, las bases de datos de red pueden ser muy rápidas, especialmente mediante el uso de índices de apunadores que permiten la ubicación directa en el primer elemento de un conjunto en una búsqueda.

Sin embargo, el modelo de red sufre del mismo problema estructural mencionado en la descripción del modelo de sistemas jerárquico. El diseño inicial de la base de datos es arbitrario, una vez que este es instalado cualquier cambio requiere crear una nueva estructura. El modelo de red permite adicionar nuevos datos o modificaciones a los ya existentes de manera simple, ya que sólo se tiene que definir un nuevo conjunto de relaciones propias con el resto del conjunto de datos.

#### 4. Modelo de Base de Datos Relacional (RDBM Relational Database Model)

El modelo relacional abandona el concepto de relaciones padre - hijo entre diferentes elementos de datos. Además, el dato es organizado en conjuntos lógicos matemáticos dentro de una estructura tabular. En un RDBM, cada campo se convierte en una columna dentro de una tabla, y cada registro se convierte en un renglón. Diferentes relaciones entre varias tablas, son definidas a través del uso de funciones matemáticas, tales como el JOIN y UNION.

Cada tabla tiene una o más columnas con el mismo nombre que se encuentra en otra tabla; son estos nombres de columnas comunes los que son utilizados para relacionar diferentes tablas. Sin embargo, los nombres de columnas no tienen que ser idénticos en el modelo relacional.

El modelo relacional tiene distintas ventajas sobre el modelo jerárquico y de red, la más importante de las cuales es su completa flexibilidad en la descripción de las relaciones entre los diferentes elementos de datos. El programador define la base de datos creando las tablas y decidiendo cuáles columnas serán relacionadas. Desde este punto, los usuarios pueden consultar la base de datos sobre alguna de las columnas en una tabla o sobre las relaciones entre diferentes tablas. Modificar la estructura de la base de datos es tan simple como aumentar o borrar columnas de una tabla, lo cual no afecta a otra tabla de ningún otro modo. Pueden ser creadas nuevas tablas como proyecciones (subconjuntos) de tablas existentes, y otras tablas pueden ser removidas. No se tiene que reconstruir la estructura de la base de datos completa para hacer cambios, esto representa un incremento en la preservación de la integridad de datos.

#### 5. Un nuevo tipo de DBMS es el llamado Orientado a Objetos (Object Oriented Databases), el cual ha sido muy difundido en los últimos dos años.

##### ④ *Arquitecturas de DBMS*

El tipo de sistemas de computadoras en que las bases de datos pueden ejecutarse; pueden ser clasificados en cuatro categorías o plataformas principales: centralizada, PC, Cliente/Servidor y distribuidas. La arquitectura del DBMS mismo no determina necesariamente el tipo de

sistema computacional en que la base de datos tiene que ejecutarse.

### *1. Plataforma Centralizada:*

En un sistema centralizado, todos los programas se ejecutan sobre una computadora principal, incluyendo el DBMS, las aplicaciones que accesan a la base de datos y las facilidades de comunicación que envían y reciben datos de las terminales de usuarios.

Los usuarios accesan la base de datos a través de una conexión local o terminales remotas. Las terminales son generalmente mudas, tienen o no poder de procesamiento propio, y constan únicamente de una pantalla, teclado y hardware para comunicarse con el host. La ventaja de los microprocesadores ha llevado al desarrollo de terminales más inteligentes en los últimos años, donde la terminal comparte responsabilidad para manejar la salida en pantalla y la entrada de datos por parte del usuario.

Mientras los sistemas en mainframe y microcomputadoras son la plataforma principal para sistemas de bases de datos corporativas, los sistemas basados en PC's pueden comunicarse además con sistemas centralizados a través de comunicaciones de hardware/software que emulan (imitan) los tipos de terminal utilizados con un host particular.

Todo el procesamiento de datos en un sistema centralizado toma lugar en el host, y el DBMS debe estar corriendo antes de que las aplicaciones puedan accesar la base de datos.

Las principales ventajas de un sistema centralizado es su seguridad y la habilidad de manejar enormes montos de datos en dispositivos de almacenamiento. Además, soportan un gran número de usuarios simultáneos. Las desventajas se refieren generalmente a los costos de instalación y mantenimiento. Grandes sistemas de mainframes y de minicomputadoras requieren soporte especializado.

### *2. Sistemas de Bases de Datos Cliente/Servidor*

De manera sencilla, una base de datos Cliente/Servidor divide el procesamiento de la base de datos entre dos sistemas: el cliente (el cual ejecuta la aplicación de la base de datos) y el servidor (el cual ejecuta todo o parte del DBMS actual). El servidor de archivos LAN provee

recursos compartidos, tales como espacio en disco para las aplicaciones, e impresoras. El servidor de bases de datos puede encontrarse corriendo en la misma PC como el servidor de archivos.

La aplicación del cliente, se identifica como front-end, maneja todas las pantallas y el procesamiento de entrada y salida del usuario. El back-end del servidor de bases de datos maneja el procesamiento de datos y el acceso a disco.

Por ejemplo, un usuario en el front-end crea un requerimiento de datos (consulta) para el servidor, y la aplicación del front-end envía el requerimiento a través de la red al servidor. El servidor ejecuta la búsqueda y envía de regreso únicamente los datos que corresponden a la solicitud del usuario.

La ventaja de un sistema cliente/servidor resulta obvia, la división del procesamiento entre dos sistemas reduce la cantidad de tráfico de datos en la red. Mientras los sistemas cliente corren generalmente sobre PC's, el servidor de bases de datos puede correr sobre otra PC o mainframe.

La desventaja de los sistemas de bases de datos descritos anteriormente es que requieren que los datos se encuentren almacenados en un sistema único. Esto puede ser un problema cuando se tiene la necesidad de atender a usuarios dispersos en un área geográfica, o bien, que necesitan compartir porciones de sus bases de datos departamentales con otros departamentos o un host central.

## 2.1 La tecnología Cliente/Servidor

Un sistema cliente/servidor es aquel que divide el procesamiento de datos entre dos componentes distintos. Por esta definición, los sistemas cliente/servidor no se limitan a aplicaciones de bases de datos; una aplicación que tiene una interfaz de usuario (front - end) y que corre localmente sobre un cliente y una parte del procesamiento corre sobre el servidor (back-end) es una forma de cómputo cliente/servidor.

Actualmente el enfoque principal es sobre aplicaciones de bases de datos. La razón de ello es que una gran cantidad de empresas están reduciendo sus costos de cómputo mediante "downsizing" de sus bases de datos. La idea del "downsizing" es simple, mover las bases de datos corporativas de sistemas grandes y centralizados, a sistemas pequeños y menos costosos que no requieren de soporte y mantenimiento extensivos. La división en el poder de procesamiento, que es la base de la arquitectura cliente/servidor, hace posible que los sistemas pequeños manipulen los datos.

Las principales ventajas de un sistema cliente/servidor surgen de la división del procesamiento entre el sistema cliente y el servidor de bases de datos. De que el procesamiento de la base de datos se lleva a cabo en el back-end, la velocidad del DBMS no se encuentra ligada con la velocidad de la estación de trabajo. Como resultado, la estación de trabajo necesita estar habilitada para correr el software front-end.

Otro beneficio que se obtiene, es la independencia entre estaciones de trabajo; los usuarios no se encuentran limitados a un solo tipo de sistema o plataforma. En un sistema cliente/servidor, las estaciones de trabajo pueden ser PC's compatibles con IBM, Macintosh, estaciones de trabajo UNIX, o una combinación de las anteriores, y pueden correr múltiples sistemas operativos, tales como MS/PC-DOS, MS Windows, IBM OS/2, o System 7 de Apple.

Otra ventaja es la preservación de la integridad de datos. Actualmente, varios servidores de bases de datos corren en un DBMS con base en el modelo relacional, de modo que los usuarios no pueden acceder a los datos fuera del DBMS. Además, el DBMS puede proveer servicios de protección de datos, tales como almacenamiento de archivos encriptados; respaldos en cinta en tiempo real, lo cual ocurre mientras la base de datos está siendo accesada; generación de discos espejo, en el cual los datos son escritos automáticamente en una base de datos duplicada sobre un disco duro distinto.

Por otra parte, el DBMS puede proveer procesamiento de transacciones, lo cual permite tener un registro de los cambios a la

base de datos y ayuda a corregir los errores que se presenten en la base de datos en caso de que el servidor tenga algún problema.

El procesamiento de transacciones es un método por el cual el DBMS mantiene un log (registro) de todas las modificaciones hechas a la base de datos en un periodo determinado. Se utiliza principalmente para las bases de datos que están siendo modificadas constantemente, tal como un sistema de procesamiento de órdenes, para asegurarse que las modificaciones a los datos están siendo almacenadas adecuadamente en la base de datos.

El log es utilizado para restaurar la base de datos (tanto como sea posible) debido a un estado de error por el cual el sistema haya fallado durante alguna modificación. El DBMS es responsable de manejar la seguridad necesaria para prevenir múltiples cambios al mismo registro o campo; los conflictos y deadlocks entre usuarios que modifican el mismo registro son reducidos significativamente cuando son manejados por un DBMS central.

## 2.2 Sistemas de Procesamiento Distribuido

El procesamiento distribuido ha existido en su forma simple desde hace varios años; de una manera muy limitada, los datos son compartidos entre varios hosts mediante el envío a través de conexiones directas (sobre la misma red) o a través de conexiones remotas vía telefónica o líneas de datos dedicadas. Una aplicación que es ejecutada sobre uno o más hosts extrae la porción de datos que ha sido modificada durante un periodo definido por el programador, y después se transmiten los datos a un host centralizado u otros hosts que se encuentran distribuidos. Posteriormente, las otras bases de datos son actualizadas, de manera que todos los sistemas están en sincronía unos con otros. Este tipo de procesamiento distribuido ocurre generalmente entre computadoras departamentales o redes LAN y sistemas propietarios.

## 3. *Lenguajes de Programación de Aplicaciones de Bases de Datos*

Un DBMS sofisticado pierde importancia si los usuarios no tienen una manera de acceder a los datos.

Una aplicación de bases de datos es un programa que permite a los usuarios ingresar, modificar, borrar y obtener reportes de los datos que se encuentran en la base de datos. Las aplicaciones generalmente son escritas por los programadores, en uno o más lenguajes de programación especializados. Sin embargo, recientemente ha existido una tendencia al uso de herramientas de acceso a bases de datos orientadas al usuario que simplifican el proceso de uso de un DBMS y eliminan la necesidad de la programación a la medida.

Los lenguajes utilizados para crear aplicaciones de bases de datos pueden ser agrupados en tres categorías: lenguajes procedurales, lenguajes SQL y todos los demás lenguajes.

### 3.1 Lenguajes Procedurales

La gran mayoría de los lenguajes de programación pueden ser descritos como procedurales. Cuando un programador crea una aplicación de bases de datos en uno de estos lenguajes, ha escrito el código de la aplicación como una serie de procedimientos. Cada procedimiento hace el trabajo de una porción de la aplicación, tal como una consulta a la base de datos o un procedimiento para modificar los datos contenidos en la base de datos. Los diferentes procedimientos son ligados unos con otros mediante otros procedimientos que forman la interfaz con el usuario (por ejemplo, un menú) y son ejecutados en un punto determinado de la aplicación.

Los lenguajes de programación estándar, tales como Pascal, COBOL, BASIC y C, son lenguajes procedurales. Estos lenguajes pueden ser utilizados para crear aplicaciones de bases de datos mediante el uso de una Interfaz de Programas de Aplicación (Application Programming Interface -API-) la cual consiste de un conjunto de funciones estándar (o llamadas) que apoyan al lenguaje para darle acceso a los datos de la base de datos. Las funciones del API generalmente se encuentran contenidas en "bibliotecas" que se encuentran incluidas en la aplicación cuando ésta es compilada. Varios vendedores de DBMS's tienen estas bibliotecas disponibles como parte del paquete del DBMS o bien, como una opción adicional que implica un costo. Todos estos lenguajes de alto nivel, que además, pueden ser utilizados para crear aplicaciones sin bases de datos, son referidas



generalmente como "lenguajes de tercera generación (3GLs)".

Algunos lenguajes de programación procedurales son específicos para un DBMS particular. Estos lenguajes son referidos comúnmente como "lenguajes de cuarta generación (4GLs)". Los ejemplos más comunes de lenguajes procedurales para bases de datos específicas son: DBASE, PAL (Paradox Application Language) y el R/BASIC Language (utilizado por Advanced Revelation).<sup>12</sup>

### 3.2 Lenguajes de Consulta Estructurados (SQL Structured Query Language)

El Lenguaje de Consulta Estructurado (SQL) fue diseñado como un lenguaje para acceder explícitamente a DBMS's basados en el modelo relacional. La versión inicial del lenguaje apareció como SÉQUEL a mediados de los años 70's y fue desarrollado por IBM como un lenguaje estándar que corría sobre mainframes IBM. Posteriormente, su nombre fue reducido a SQL.

SQL se describe más propiamente como un sub lenguaje, ya que no contiene facilidades de manejo de pantalla o de entrada/salida. Su objetivo principal es proveer un método estándar para acceder a las bases de datos, independientemente del lenguaje en que se encuentre escrita el resto de la aplicación. Está diseñado para consultas interactivas de una base de datos o como parte de una aplicación escrita en algún lenguaje procedural.<sup>13</sup>

Desde que fue creado, ha sido revisado en distintas ocasiones; a principios de los 80's la organización ANSI (American National Standards Institut) intentó estandarizar el lenguaje SQL, lo que condujo a la liberación de las especificaciones ANSI-86 SQL y posteriormente las especificaciones ANSI-89 SQL y ANSI-92 SQL. IBM se ha enfocado a la expansión del lenguaje SQL, trabajando especialmente en ampliar las capacidades de su base de datos relacional para mainframes DB2. Cada vendedor de DBMS incluye sus propias extensiones al estándar SQL, y estas extensiones pueden hacer a las distintas versiones de SQL incompatibles una con otra. SQL fue inicialmente introducido comercialmente como un sistema de bases de datos en 1979 por Oracle.

En 1991, ANSI actualizó el estándar. El nuevo estándar es conocido como SAG SQL.<sup>14</sup>

Se han derivado una serie de lenguajes de consulta, todos basados en el SQL estándar, a continuación se presentan el orden alfabético algunos de ellos:

**ESQL/C:** Creado por Informix. Hace una combinación de sentencias SQL y sentencias del lenguaje C.

**Inquery:** Es un lenguaje de consultas para bases de datos. Soporta dos tipos de consultas: Lenguaje natural y estructurado. Evalúa información de acuerdo a modelos probabilísticos, y no a modelos booleanos. Es decir, encuentra documentos que concuerden con todas las palabras en la consulta y decide que documentos son los más adecuados. Inquery's es otro método, de consultas estructuradas, que permite obtener información más exacta acerca de las relaciones de términos en la consulta.

**MiniSQL:** Desarrollado por David Hughes Technologies.

**MSQL:** Es un lenguaje de consultas para uso ideal con páginas html generadas a partir de consultas de bases.

**PL/SQL:** es una extensión de SQL.

**Query-By-Example:** No es estrictamente un lenguaje; es una interfaz que presenta al usuario con una o más tablas en blanco que corresponden a las tablas en la base de datos. El usuario elige las columnas que serán incluidas en la consulta a través de una combinación de teclas, y define las condiciones de la consulta llenando las condiciones dentro de las columnas apropiadas. El DBMS traslada el QBE en las acciones necesarias para cumplir con el requerimiento del usuario.

**SQL\*Plus:** Es un intérprete de comandos SQL para diversos sistemas operativos.

**SQL92:** Lenguaje nativo de Oracle, es el estándar más reciente de ISO

**SQLDB:** (Structured Query Language Database). Es un lenguaje creado para definir y manipular bases de datos relacionales como la especifica la ANSI en el estándar X3.135-986 y el estándar X3.115-1989 de la FIPS (Federal Information Processing Standard).

**Transact-SQL:** Lenguaje nativo de Sybase, es una extensión de

**c. Data warehousing y soporte de decisiones (1990's)**

Un Data warehouse o almacén de datos o bodega es usualmente definido como una colección de datos integral variable en el tiempo, no volátil y orientada a temas importantes para el soporte a la toma de decisiones en la administración de la organización.

**Características:**

1. Debe ser integral; no importando de donde provienen las fuentes de datos que lo integran deberán formalizarse como un solo tipo descrito de datos.
2. Debe estar organizado en relación al objetivo principal de la aplicación y las variables más importantes de la organización. Esta información es el soporte real para la toma de decisiones por lo que su objetivo particular debe ser servir a la misión central de la organización.
3. Debe ser incrementable en el tiempo. Debe almacenar información de varios años.
4. No debe ser volátil. No debe permitir agregaciones, modificaciones o eliminaciones de los datos que contiene cuando estas provengan del usuario.

Data warehouse es un concepto relativamente nuevo tanto en México como en otros países, que viene a resolver los problemas del manejo y uso adecuado de grandes fuentes de datos y de muy diversos tipos, para apoyar la toma de decisiones oportunas y fundamentadas.

Un sistema operacional o de procesamiento en línea (OLTP On-Line Transaction Processing), es un sistema tal como el de administración de recursos humanos, de asignación de créditos bancarios, de recuperación y control de cartera o de control de seguros, y su función principal es dar el soporte a las necesidades diarias de la empresa; son sistemas normalmente optimizados para el manejo de un conjunto predefinido de transacciones.

Los sistemas operacionales de los cuales se transferirá la información seleccionada, pueden haber sido construidos utilizando manejadores de datos relacionales (RDBMS, del inglés), manejadores de archivos jerárquicos, de archivos planos y otro tipo de manejadores.

Data warehouse no es ni un producto de software ni una máquina o tecnología de bases de datos en particular, sino una serie de componentes y procesos que en conjunto forman la arquitectura data warehouse. Sus partes importantes son:

- ④ *Análisis, selección y extracción de datos.* Procesos requeridos para seleccionar datos de sistemas operacionales, extraerlos y convertirlos a un formato o formatos que permitan manejarlos en común, de acuerdo al modelo de datos de la empresa, y de acuerdo a la información para toma de decisiones que se desee contar. Los datos deberán ser actualizados (extraídos de nuevo) como un proceso cíclico y periódico.
- ④ *Almacenamiento de los datos extraídos.* Elementos necesarios para almacenar y manejar los datos. A este nivel se cuenta con los detalles de los datos, así como con los metadatos o información de alto nivel que los describe, obtenidos en el análisis del paso anterior. Para el almacenamiento y manejo de los datos extraídos se puede utilizar un RDBMS, el cual, por su manejo de grandes volúmenes de información, del orden de terabytes incluso, por su implementación destinada a explotar las facilidades de hardware tales como memorias de 4 gigabytes o mayores y procesamiento simétrico o SMP, los hace eficientes para consultas, selección y procesamiento de datos.
- ④ *Análisis de los datos.* Análisis simple. Explotación de los datos realizando consultas simples basándose en herramientas de productividad de oficina tales como hojas de cálculo o paquetes de análisis estadísticos.
- ④ *Análisis complejo.* Explotación de los datos basándose en consultas y análisis multidimensional, utilizando herramientas de software para procesamiento analítico u OLAP (On-Line Analytical Processing)

Los objetivos de los sistemas OLTP y data warehouse para toma de decisiones son muy diferentes: tratar de diseñar un data warehouse pensando en un sistema operacional es garantía de fracaso. El éxito en la implantación de una arquitectura de data warehouse en las empresas radica en parte en el éxito en el diseño de los sistemas operacionales, ya que éstos son los proveedores de datos y los que se deben adaptar rápida y flexiblemente a los cambios del negocio. Se deben tener ideas muy claras de los siguiente ¿Qué datos se deben utilizar?; ¿cómo se deben transformar?, ¿cómo se deben transferir, almacenar y organizar?, y

finalmente cómo se deben acceder y analizar?

Por lo anterior podemos definir al Data warehousing, como: el proceso de extracción y transformación de datos obtenidos en fuentes operacionales (OLTP) llevándolos a una base de datos centralizada reconocida como un Data warehouse. Por lo que una vez en el repositorio es explotada usando herramientas para la toma de decisiones.<sup>16</sup>

#### **d. Minería de datos (hoy en día)**

A lo largo de la vida productiva de una empresa se van acumulando grandes cantidades de datos que son almacenados en algún lugar para posteriormente hacer uso de ellos.

Gracias a los sistemas de cómputo, las empresas tienen la capacidad de almacenar y acceder, en archivos o bases de datos, grandes cantidades de datos históricos sobre las operaciones diarias de su negocio; información que en su momento fueron usados para satisfacer las necesidades propias de la empresa y para el soporte de decisiones.

Todos esos archivos deben contener gran cantidad de información que sería de gran utilidad si fuera posible aprovecharla.

La mayoría de las organizaciones no sufre por falta de información, sino más bien por exceso de información redundante a la que resulta cada vez más complicado acceder para buscar datos específicos y significativos que permitan obtener una visión más completa de la situación operacional de la empresa y así lograr una mejor toma de decisiones.

Las áreas de sistemas han venido trabajando para crear extractos de información de las bases de datos operacionales y almacenar estos datos en archivos, tratando de responder a las peticiones de los usuarios para obtener información que les ayude a tomar mejores decisiones. Las necesidades de información han hecho que se diseñen sistemas de información ejecutiva y de apoyo a la toma de decisiones -- tienen como objetivo primordial proveer de toda la información necesaria a los ejecutivos de alto nivel para apoyarlos en la toma de decisiones, además de que les permite tener acceso rápido y efectivo a la información compartida y crítica del negocio --; sin embargo, las demandas

<sup>16</sup> Vizcaíno Carlos, "Soluciones Avanzadas, Tecnologías de Información y Estrategias de Negocios"

---

---

de las empresas, en relación a la información, van más allá de simples consultas, cruces de información o reportes consolidados; lo que ha hecho que se creen nuevas formas de análisis de la información.

Existen algunos tipos de empresas que su principal negocio es manejar la información; por lo que siempre tratan de convertirla por todos los medios posibles en conocimiento.

La minería de datos esta lista para aplicaciones de las empresas actuales ya que esta soportada por tres tecnologías que actualmente están lo suficientemente maduras:

- ✓ Colección masiva de datos
- ✓ Poderosas computadoras de multiproceso
- ✓ Algoritmos de minería de datos

---

### 3.4 CLASIFICACIÓN

La minería de datos esta clasificada, en tres grandes categorías:

- ❖ *Análisis estadístico o de datos:* se usan para detectar patrones no usuales de datos, estos patrones de datos se explican después mediante modelos estadísticos y matemáticos. Algunas técnicas de modelado son el análisis lineal y no lineal, análisis de regresión continua y logística, análisis de variación y multivariación y el análisis de series históricas. Las herramientas de análisis estadístico se utilizan en varias aplicaciones empresariales: incrementar la participación en el mercado y las utilidades determinando las mejores oportunidades; mejorar la satisfacción del usuario mejorando la calidad en productos y servicios.
- ❖ *Descubrimiento de conocimiento:* consiste en un conglomerado de componentes que identifican y extraen patrones y relaciones interesantes y útiles. Las principales entradas al sistema de descubrimiento de conocimiento son: los datos del Data Warehouse, la guía del analista empresarial y los conocimientos en la materia que almacena la base de conocimiento del sistema.
- ❖ *Otros tales como la visualización y sistemas de visualización geográfica.*

---

## 3.5 TÉCNICAS

Las técnicas más comunes de la minería de datos son:

### ☐ Redes Neuronales <sup>17</sup>

La minería basada en redes neuronales esta específicamente indicada para identificar patrones y prever tendencias basadas en comportamientos previamente identificados. Una tendencia identifica un movimiento en el hábito basado en el comportamiento anterior. El mercado de valores es un campo perfecto para analizar las tendencias. Cuando se analiza el precio de un producto se cotiza en bolsa los últimos seis meses se puede predecir una cuota de mercado para las próximas semanas.

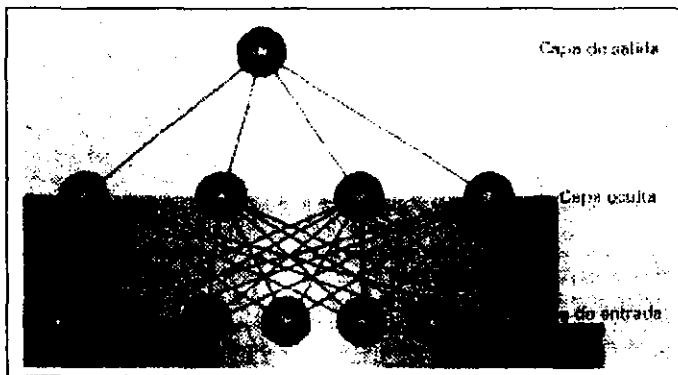
La base de este tipo de procesamiento es lo que se aprendió trabajando con el sistema nervioso central. El conocimiento se puede aprender de una serie de datos ampliamente dispares, complejos o imprecisos. Hay tres capas de la red: la capa inferior recibe los datos de entrada, la capa oculta (intermedia) realiza el trabajo, y la exterior presenta las salidas al analista. Por ejemplo, en una empresa de marketing, las entradas pueden ser los datos históricos correspondientes a los hábitos de consumo de los clientes en momentos cercanos en el tiempo a los lanzamientos de nuevas campañas de marketing por parte de la compañía. La capa oculta, o intermedia, procesa los datos de entrada y entrega los resultados en forma de patrones o tendencias a la capa exterior. La capa de entrada, la de salida, y la oculta están compuestas por nodos. Estos nodos son otra forma de llamar a los elementos de procesamiento, que se asemejan a las neuronas del cerebro; de ahí la terminología "red neuronal".

Cuando esta red se entrena mediante información en la capa de entrada, asume un misterioso componente de humanidad a medida que se hace experta tomando datos de elementos de datos sin aparente relación y devolviendo resultados a la capa de salida. La siguiente figura muestra la estructura de una red neuronal y cómo cada nodo en cada capa está conectado con todos los nodos de las capas adyacentes.

---

<sup>17</sup> Michael J:A: Berry, Gordon Linoff: Data Mining Techniques, For Marketing, Sales an Customer Support  
60





*Estructura de una red neuronal*

□ Descubrimiento de asociaciones

Esta técnica concierne al estudio de conjuntos de datos e intenta descubrir asociaciones entre las ocurrencias de los atributos dentro de los datos. El descubrimiento de asociaciones intenta descubrir ocurrencias similares de los valores de los datos de los registros, y produce una salida que podría ser expresada como una regla: "El 86 por 100 de los pacientes examinados en la clínica A para un remedio de hernia necesitaron también, dentro de los seis mese siguientes, una intervención relacionada con una dolencia estomacal".

Se puede descubrir un factor de confianza haciendo minería de datos, usando descubrimiento de asociaciones o reglas de asociaciones. El factor de confianza está expresado en un porcentaje que mide el poder de una predicción.

Dentro de esta técnica cuando se habla de reglas de asociaciones, se habla de componentes de la parte izquierda (LHS) y de componentes de la parte derecha (RHS) de la regla de asociación. Para determinar un factor de confianza, debemos buscar en qué porcentaje el suceso (en este caso, adquirir una mesa de café) de la parte derecha ocurrirá a la vez que los dos sucesos de la parte izquierda (en este ejemplo, comprar un sofá y, al menos, otro mueble de asiento más para el salón). En la parte izquierda pueden existir uno o más sucesos, mientras que en la derecha tiende a ser uno solo. Una vez que se han

---

procesado los datos y encontrado las asociaciones, se hayan examinado y sopesado, la empresa podrá determinar qué decisiones de marketing son adecuadas.

### Clasificación

La clasificación consiste en la agrupación de los datos basándose en una serie de características predefinidas por el analista antes de que comience el ejercicio. Esta técnica examina los datos ya clasificados y agrupados mediante la aplicación de una regla de pertenencia. Esta regla de pertenencia puede tener un componente temporal, por ejemplo, un año natural o fiscal o un mes del año; un componente geográfico, por ejemplo, al este o al oeste del río Mississippi; o un componente cuantitativo, por ejemplo, los clientes con un gasto anual superior o inferior a una determinada cantidad.

Una de las aplicaciones más frecuentes de esta técnica es en el área de conservación de clientes.

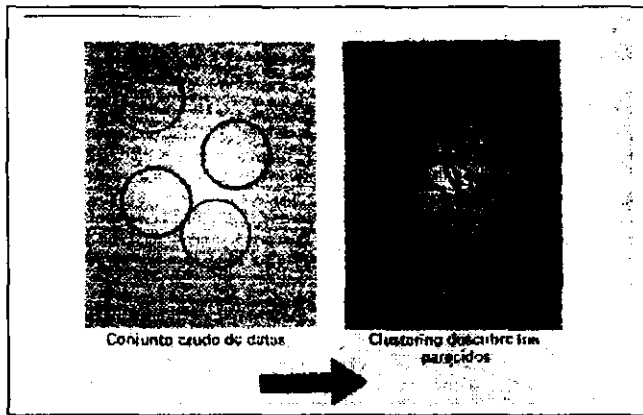
### Clustering <sup>18</sup>

El clustering (agrupación) consiste en agrupar información de datos similares partiendo de un conjunto más grande de datos. Contrariamente a la clasificación, la técnica de clustering descubre las agrupaciones a medida que trabaja con los datos de entrada. Se identifican las similitudes que lleven a las segmentaciones en grandes grupos de datos que se parezcan unos a otros. Una vez que las agrupaciones y sus miembros son identificados, se pueden descubrir generalizaciones, patrones y tendencias basándose en las características de los miembros de cada agrupación.

La base de esta técnica es el descubrimiento de que lo que parecían valores de datos y de atributos sin ninguna relación dentro de una serie de registros, de hecho, contiene información acerca de un conjunto claro de similitudes. Cada usuario se encuentra con resultados agrupados en la vida diaria. La siguiente figura ilustra cómo se solapan datos previamente sin ninguna relación a través de la aplicación de una técnica de clustering.

---

<sup>18</sup> Michael J. Corey, Michael Abbey, Oracle Data Warehousing, Oracle Press, 1997.



### *Técnica de Clustering*

#### Descubrimiento secuencial

El descubrimiento secuencial intenta encontrar patrones entre sucesos que ocurren en una progresión a lo largo de un rango de tiempo. Para hacer este proceso significativo en la estructura de proveedor/comprador, por ejemplo, en las ventas al por menor, el componente temporal de los datos que están siendo analizados se decide previamente. Una vez que se ha tomado esta decisión, el minero de datos puede manejar los datos que se van a inspeccionar y el proceso comienza.

Esta técnica de minería de datos se puede usar, por ejemplo, para buscar patrones de compras de objetos de compras, el esfuerzo de minería de datos se concentra en buscar más ocurrencias de los mismos hábitos. La minería de datos analiza conjuntos de registros referentes a compras, buscando patrones que aparezcan frecuentemente durante el periodo de tiempo elegido.

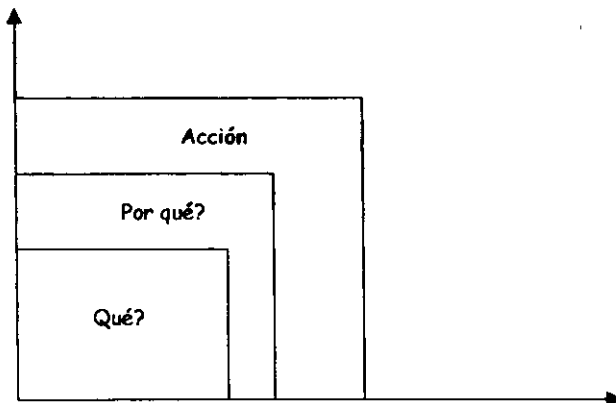
**NOTA :** Todas estas técnicas que pueden ser aplicadas a la minería de datos, han sido estudiadas y definidas previamente por medio de algoritmos bien definidos, de acuerdo a una ardua labor tanto analítica como de programación para ser aplicadas en distintas herramientas (software) que ya están a la venta en el mercado.

### 3.6 INGREDIENTES DE LA MINERÍA DE DATOS

La minería de datos auxilia a los usuarios empresariales en el procesamiento de vastas reservas de datos para describir "relaciones insospechadas", por ejemplo entre productos y clientes o patrones de compra de los clientes. La meta es descubrir "relaciones estratégicas competitivas" para controlar la participación en el mercado y las utilidades.

Los analistas empresariales tienen un rango de necesidades. La primera necesidad es comprender qué está sucediendo en el negocio. La siguiente es por qué está sucediendo, ¿cuál es el comportamiento de clientes y mercados?. La última necesidad es ¿qué puede hacerse?, ¿cuáles acciones se pueden tomar?. El valor de un análisis para los gerentes es más alto cuando genera una recomendación factible, y lo que puede hacerse, son retos para las técnicas tradicionales de análisis. Las consultas, reportes y análisis multidimensionales tradicionales se encuentran en lo que está sucediendo y en menor medida, en el por qué. "La minería de datos" se concentra en llenar la necesidad de descubrir el por qué, para luego predecir y pronosticar las posibles acciones con cierto factor de confianza para cada predicción.

#### *Metas de soporte de decisiones*



Las herramientas de minería de datos son un componente importante del sub-bloque de análisis y reportes del bloque de acceso y uso de la arquitectura de referencia descrita a continuación. Las herramientas de minería de datos utilizan el sub-bloque de análisis y recuperación para tener una interfaz con el data warehouse y con el mercado de datos. Muchas de las herramientas de la minería de datos también emplea el componente de depósito local del bloque de acceso y recuperación, a fin de almacenar los datos en estructuras de datos para análisis subsecuentes y presentaciones de los resultados.

*Arquitectura de la minería de datos\**

	<b>Acceso y Recuperación</b>	<b>Análisis y Reportes</b>	
	Acceso directo al data warehouse	Herramientas de reporte	
	Acceso al mercado de datos	Herramientas de análisis Herramientas SSD	
	Reingeniería	Herramientas de modelos empresariales	
	Transformación a estructuras multidimensionales	Herramientas de minería de datos	
	Creación de depósito local	Aplicaciones de producción nueva	OLAP
Manejo de metadatos de data warehouse	Examen y navegación de meta datos	Administración de metadatos y reportes	

El componente de Acceso y Recuperación proporciona un acceso directo al data warehouse evitando el mercado de datos. Este componente también es responsable de transformar los datos recuperados en vistas multidimensionales o de almacenarlos en una base de datos multidimensional para un análisis posterior. Los datos del data warehouse y del mercado de datos pueden ser "copiados" en un depósito local, con o sin traducciones y transformaciones, para el análisis local, minería de datos y elaboración de reportes. Una consideración importante es el rango de consultas: predefinidas, de acuerdo, repetitivas, para profundizar; y estas consultas tienen acceso a

\* Data Warehousing: La integración de la información para la mejor toma de decisiones

---

cantidades de datos pequeñas o grandes. Tales aspectos se solventan en este componente. El acceso y el examen de los metadatos también se incluyen en este componente.

El componente de Análisis y Reporte es responsable de la familia de herramientas y aplicaciones necesarias para aprovechar el data warehouse y el mercado de datos. Estas se clasifican en herramientas de reOporte, de análisis y soporte de decisiones, de modelado empresarial y procesamiento analítico, y de minería de datos.

La mayoría de las herramientas de minería de datos puede con facilidad saltarse el data warehouse o el mercado de datos y acceder de manera directa a la fuente de datos. Tradicionalmente, las herramientas de minería de datos acceden a los datos fuente. Sin embargo, los datos del data warehouse o del mercado de datos están refinados, integrados y estandarizados. La estandarización elimina aspectos como las convenciones y nombre múltiples, las estructuras ocultas de codificación y los campos faltantes. Los datos operacionales en la fuente son por lo general inconsistentes y están dispersos en muchas aplicaciones. Además, se requieren datos históricos para descubrir patrones temporales de interés.

La minería de datos define en varias formas el procesamiento informático y analítico. \*

	<i>El procesamiento Informático/analítico en contra de la minería de datos</i>	<i>Minería de datos</i>
Enfoque	Datos de resumen	Datos de transacción o detalle
Dimensiones	Limitadas	Muchas
Cantidad de atributos	Total de decenas	Cientos por cada dimensión
Tamaño del dataset	De reducido a mediano para cada dimensión	Millones para cada dimensión
Enfoque del análisis	¿qué está sucediendo en el negocio?	¿por qué está sucediendo?. Acciones de predicción y pronóstico
Técnicas de análisis	Recabar y picar	Descubrir automáticamente
Proceso de análisis	Análisis empresarial iniciado y controlado	Datos y sistema iniciado. Orientación mínima al analista empresarial
Factor de confianza	Derivado por el analista empresarial	Derivado de los datos
Estado de tecnología	Desarrollada	Desarrollada en análisis estadístico. Incipiente en descubrimiento de conocimientos

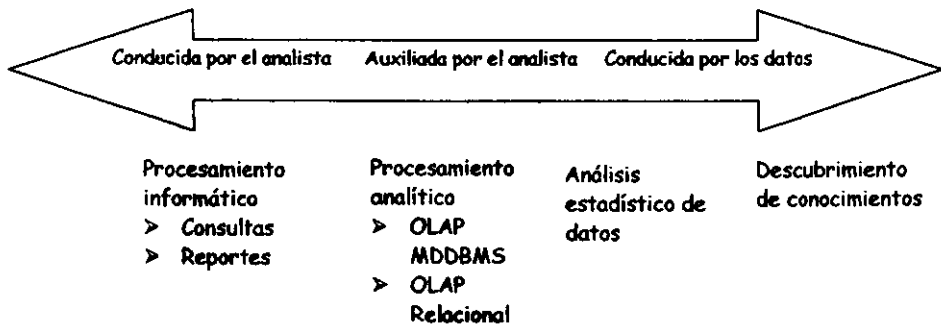
Los datos en el data warehouse deben estar al nivel de detalle correcto. Debido a la naturaleza incipiente de la tecnología de minería de datos, es necesaria - en especial al principio - una estrecha cooperación entre los analistas empresariales y los profesionales en tecnología de la información.

Para formar la mezcla correcta de actividades de minería de datos, son cruciales tres ingredientes: usuarios, aplicaciones empresariales y tecnología y herramientas.

\* Data Warehousing: La integración de la información para la mejor toma de decisiones

- ④ Usuarios de la minería de datos: son los analistas empresariales, los peritos en estadística y los profesionales en tecnología de la información que auxilian a los usuarios empresariales y los ejecutivos, que desean entender los factores de éxito del negocio con datos completos del cliente, y utilizar luego este conocimiento para afinar las estrategias de producción, precios y comercialización, mejorar el nivel de éxito de las estrategias e impulsar el balance.

*Función del analista empresarial y de los datos*



- ④ Surgimiento de aplicaciones de minería de datos: En las aplicaciones empresariales, a la fecha, la tecnología de minería de datos se ha utilizado principalmente en aplicaciones de comercialización, ventas y análisis de crédito; y se ha aplicado con éxito en áreas empresariales con el más alto potencial, tales como la segmentación de clientes y del mercado y el análisis de comportamiento del cliente, en particular en los sectores de menudeo, bancario y financiero. Hasta aquí, la tecnología por lo general era costosa de aplicar y desplegar, pero esta situación está cambiando con rapidez. Hoy en día, una confluencia de tres fuerzas importantes conduce al crecimiento de la minería de datos:

- > La tecnología del data warehouse para proporcionar un gran banco de datos bien organizados e históricos.
- > Hardware en paralelo, productos de base de datos y herramientas a precios razonables.
- > Tecnología y herramientas para minería de datos cada vez más desarrolladas.



---

④ Tecnología y herramientas de la minería de datos: Existe una amplia variedad de tecnología para la minería de datos, todavía va apareciendo más en el mercado. Estas herramientas y tecnologías de minería de datos se clasifican en tres grandes categorías:

- Análisis estadístico o de datos
- Descubrimiento de conocimientos
- Otros, como sistemas de visualización, sistemas de información geográfica, análisis fractal (modelos matemáticos de simulación) y herramientas propietario.



# Marco Metodológico

## 4. MARCO METODOLÓGICO

Debido al trabajo realizado en el marco problemático y el marco conceptual de la presente investigación, llegue a la conclusión de que las variables y la hipótesis subsisten de la siguiente manera.

### 4.1 VARIABLES

<b>Variables Independientes</b>	<b>Variables Dependientes</b>
Grandes volúmenes de Información	Decisiones no tan certeras Decisiones lentas Desaprovechamiento de información
Uso de herramientas automatizadas	Mejor aprovechamiento de la información
Reportes o informes detallados	Toma de decisiones más certeras y rápidas.

### 4.3 HIPÓTESIS DEFINITIVA

*"El aprovechamiento en mayor grado de los grandes volúmenes de información contenida en una base de datos, se dará por el uso de una herramienta informática que es la minería de datos. La cuál nos ayudará a generar reportes o informes que coadyuvarán a una eficaz y eficiente toma de decisiones dentro de las empresas, así como a un conocimiento más a detalle del negocio."*

### 4.4 DEFINICIÓN DEL UNIVERSO

La práctica de la minería de datos se dará en empresas que cuenten con una base de datos relacional y sobre todo que hagan uso de herramientas para la explotación de la información contenida en sus bases de datos.

---

## 4.5 DEFINICIÓN DE LA MUESTRA

A lo largo de toda la historia, la curiosidad ha conducido a la experimentación y la investigación, con el objeto de auxiliar al proceso de toma de decisiones. Este proceso es un denominador en todos los campos de la actividad humana.

La definición de las metas propias conduce a la consideración de cursos alternativos de acción. Un tomador de decisiones sensato busca evaluar información con el objeto de elegir el curso de acción que rinda el máximo beneficio para los objetivos. Para el investigador la información necesaria son los datos. A fin de que el análisis estadístico resulte útil en el proceso de toma de decisiones, los datos de entrada iniciales deben ser apropiados, si los datos están sesgados, son ambiguos o tienen otro y tipo de error, es probable que aun las pruebas mejor diseñadas no compensen tales deficiencias.

Dentro de la estadística moderna, se pueden realizar análisis que incluyan recopilación, presentación y caracterización de la información a fin de que auxilie tanto en el análisis de los datos como en el proceso de toma de decisiones.

Existen básicamente dos tipos de muestra: la muestra no probabilística y la muestra probabilística. Para la mayor parte de los estudios analíticos sólo existe disponible una muestra no probabilística, como la muestra de juicio. En este caso, la opinión de un experto en la materia de estudio es crucial para avalar la utilización de los resultados obtenidos con el fin de hacer cambios en algún proceso.<sup>1</sup>

- Para la presente investigación he decidido realizar una Muestra de Juicio: Son aquellas muestras que resultan de una opinión, tomando en cuenta la concurrencia de efectos; es decir, tomando en cuenta todo aquello que pueda afectar el objeto de estudio.

Debido a que se desconoce el número exacto de empresas que utilizan herramientas automatizadas para el manejo de la información, he decidido

---

<sup>1</sup> Mark L. Berenson, David M. Levine: Estadística Básica en Administración, Conceptos y Aplicaciones, Prentice - Hall, Hispanoamericana, S.A. 4ª. Edición, México, 1992, ISBN 58259

---

definir un número de empresas que si utilizan herramientas automatizadas para la toma de decisiones. Dichas empresas son las que se enlistan a continuación:

- ⊕ Impulsora Mexicana de Telecomunicaciones S.A. de C.V.
- ⊕ Grupo Financiero Inbursa S.A. de C.V.
- ⊕ Canal 11 (Instituto Politécnico Nacional)
- ⊕ Consultoría IISI S.A. de C.V. (Integración Inteligente de Sistemas de Información)
- ⊕ Consultoría Saul Cantú y Asociados
- ⊕ Seguros Monterrey Atenia S.A. de C.V.
- ⊕ Consultoría INFOINTER S.A. (Informática Interdisciplinaria)
- ⊕ Consultoría KCS S.A. (Knowledge and Consulting Services)
- ⊕ General Electric Capital Services

#### **4.6 DEFINICIÓN DEL MÉTODO DE LA INVESTIGACIÓN**

El método que se utilizará para esta investigación será por medio de encuestas, debido a que presentan las siguientes ventajas:

- ✓ Es una técnica por medio de la cual se puede tener una conversación personal
- ✓ Se puede llegar a conocer vivencias de otras personas y/o conocer a expertos
- ✓ Abundar temáticamente
- ✓ Aclarar con más profundidad algunos temas tratados
- ✓ Abundar en los cuestionamientos que se realicen, etc.

---

## 4.7 COSTO DE LA INVESTIGACIÓN

### Desarrollo de Tesis Profesional "La minería de datos una herramienta para toma de decisiones"

#### Cédula de costos incurridos

Concepto	Costo Mensual	Costo del Trabajo
Sueldo y Salarios (investigador)	\$ 8,000.00	\$ 80,000.00
Renta	\$ 500.00	\$ 5,000.00
Mobiliario y equipo (computadora, impresora, escáner)	(depreciación mensual) \$250.00	\$ 2,500.00
Servicios		
Teléfono	\$ 100.00	\$ 1,000.00
Luz	\$65.00	\$ 650.00
Agua	\$45.00	\$ 450.00
Papelería	\$ <u>200.00</u>	\$ 2,000.00
Costo TOTAL	\$12,160.00	\$ <b>89,600.00</b>

## 4.8 CONSTRUCCIÓN DEL CUESTIONARIO

### CUESTIONARIO

Pregunta	Razón de ser	Respuesta Esperada
1. <i>¿En su ámbito laboral cuentan con alguna metodología para el análisis de información?</i> SI ( ) NO ( )	Saber si en distintas empresas del país tienen alguna metodología definida para el análisis de la información contenida en una base de datos.	Sí.
2. Si contestó afirmativamente la pregunta anterior, señale que modelos o metodologías utiliza: ( ) Modelo relacional ( ) Modelo de red ( ) Modelo jerárquico ( ) Modelo distribuido ( ) Otra. ¿Cuál?	Saber si se cuenta con una metodología propia de la empresa o alguna de las conocidas, como la del modelo jerárquico, de red, etc.	Sí, utilizamos una combinación del modelo jerárquico y el modelo distribuido para analizar la información.
3. Las bases de datos que se diseñan en su ámbito laboral, se utilizan para explotarse: ( ) Inmediatamente ( ) A futuro ( ) Ambas	Conocer que grado de aprovechamiento se tiene de la información	Para su explotación inmediata y futura.
4. <i>¿Utiliza algún modelo o método para el análisis de la información histórica?</i> SI ( ) NO ( )	Saber si se utilizan modelos o metodologías para el análisis de la información histórica	No
5. Si contestó afirmativamente, diga cuales utiliza:	Conocer los modelos que utilizan las empresas para el análisis de información	Seleccionar algunos modelos utilizados, o determinar cuál es la

<ul style="list-style-type: none"> <li><input type="checkbox"/> Clasificación</li> <li><input type="checkbox"/> Asociación</li> <li><input type="checkbox"/> Árboles de decisiones</li> <li><input type="checkbox"/> Redes neuronales</li> <li><input type="checkbox"/> Algoritmos paralelos</li> <li><input type="checkbox"/> Híbridos</li> <li><input type="checkbox"/> Reglas de inducción</li> <li><input type="checkbox"/> Estadísticas</li> <li><input type="checkbox"/> Análisis Fractal</li> <li><input type="checkbox"/> Clustering</li> <li><input type="checkbox"/> Secuencias</li> <li><input type="checkbox"/> Algoritmos genéticos</li> <li><input type="checkbox"/> Memoria basada en el razonamiento</li> <li><input type="checkbox"/> Reglas de clasificación</li> <li><input type="checkbox"/> Patrones generales y búsqueda de excepciones</li> <li><input type="checkbox"/> Sistemas de visualización</li> <li><input type="checkbox"/> Técnicas difusas</li> <li><input type="checkbox"/> Sistemas de información geográfica</li> <li><input type="checkbox"/> Otros ¿Cuál?</li> </ul>	<p>histórica.</p>	<p>metodología que utiliza; y si es una metodología propia describirla brevemente.</p>
<p>6. ¿Considera que los modelos y/o métodos utilizados son eficientes para el apoyo a la toma de decisiones y el conocimiento del negocio?</p> <p>SI ( ) NO ( )</p>	<p>Conocer en nivel de detalle de los modelos y/o metodologías utilizadas.</p>	<p>No, porque en ocasiones falta detallar otro tipo de información que no se obtiene tan fácilmente.</p>
<p>7. ¿Utiliza alguna herramienta automatizada para la extracción y manejo de la información?</p> <p>SI ( ) ¿Cuál?</p> <p>NO ( ) ¿Por qué?</p>	<p>Para identificar si se apoyan de herramientas automáticas de análisis de información para la toma de decisiones.</p>	<p>Sí, Informix.</p>
<p>8. Considera que la utilización de herramientas que le ayuden a generar información adicional a la que se obtiene comúnmente</p>	<p>Demostrar la importancia que tienen las herramientas de análisis de información histórica.</p>	<p>Sí, porque nos ayudan a analizar la información existente y poder tomar decisiones más certeras.</p>



<p>de sus sistemas de información, podrían ayudar a conocer más a detalle su negocio e incrementar la productividad del mismo</p> <p>I ( ) NO ( )</p> <p>Por qué?</p>		
<p>1. Conoce el término "minería de datos"? Si es así, de una definición.</p>	<p>Conocer la definición personal de Minería de Datos</p>	<p>Es una combinación de técnicas, por medio de las cuales podemos encontrar información muy valiosa</p>
<p>2. ¿Conoce alguna persona que se dedique a la investigación y promoción de la "minería de datos" en nuestro País?</p>	<p>Saber si tienen el conocimiento de alguna persona que se dedique a esta área del conocimiento</p>	<p>Sí, el Dr. Adolfo Guzmán Arenas.</p>



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CONTADURÍA Y  
ADMINISTRACIÓN



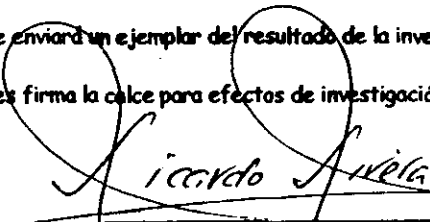
A QUIEN CORRESPONDA:

Por medio de la presente, presento a la Srita. Esmeralda Martínez Montes, con número de cuenta 9221692-9 como pasante de la carrera de Licenciatura en Información, y que actualmente se encuentra realizando su tesis profesional denominada "La minería de datos una herramienta para la toma de decisiones".


Ella realiza una investigación, y para tal efecto se encarece se le brinde la facilidad de aplicar un cuestionario, cuyo contenido será tratado de forma absolutamente confidencial.

Al final de su trabajo ella le enviará un ejemplar del resultado de la investigación.

La señorita Martínez Montes firma la calce para efectos de investigación.



Dr. Ricardo Rivera Soler  
Tels. 56 76 62 86 particular  
54 82 00 13 oficina



Esmeralda Martínez Montes  
Tels. 56 62 72 36 oficina  
56 17 57 66 particular

# ESTA TESIS NO SALE DE LA BIBLIOTECA

## 4.10 CUESTIONARIO PILOTO

### Cuestionario

*¿En su ámbito laboral cuentan con alguna metodología para el análisis de información?  
SI ( ) NO ( )*

*Si contestó afirmativamente la pregunta anterior, señale que modelos o metodologías  
liza :*

- ) Modelo relacional*
- ) Modelo de red*
- ) Modelo jerárquico*
- ) Modelo distribuido*
- ) Otra. ¿Cuál?*

*Las bases de datos que se diseñan en su ámbito laboral, se utilizan para explotarse:*

- ) Inmediatamente*
- ) A futuro*
- ) Ambas*

*¿Utiliza algún modelo o método para el análisis de la información histórica?  
SI ( ) NO ( )*

*Si contestó afirmativamente, diga cuales utiliza:*

- ) Clasificación*
- ) Asociación*
- ) Árboles de decisiones*
- ) Redes neuronales*
- ) Algoritmos paralelos*
- ) Híbridos*
- ) Reglas de inducción*
- ) Estadísticas*
- ) Análisis Fractal*
- ) Clustering*
- ) Secuencias*
- ) Algoritmos genéticos*
- ) Memoria basada en el razonamiento*
- ) Reglas de clasificación*

- ) Patrones generales y búsqueda de excepciones
- ) Sistemas de visualización
- ) Técnicas difusas
- ) Sistemas de información geográfica
- ) Otros ¿Cuál?

2. ¿Considera que los modelos y/o métodos utilizados son eficientes para el apoyo a la toma de decisiones y el conocimiento del negocio?

SI ( ) NO ( )

3. ¿Utiliza alguna herramienta automatizada para la extracción y manejo de la información?

SI ( ) ¿Cuál?

NO ( ) ¿Por qué?

3. Considera que la utilización de herramientas que le ayuden a generar información adicional a la que se obtiene comúnmente de sus sistemas de información, le podrían ayudar a conocer más a detalle su negocio e incrementar la productividad del mismo

SI ( ) NO ( )

¿Por qué?

9. Conoce el término "minería de datos"? Si es así, de una definición.

10. ¿Conoce alguna persona que se dedique a la investigación y promoción de la "minería de Datos" en nuestro País?

Nombre:

Puesto:

Empresa:

No. de Cuestionario	Pregunta 1		Pregunta 2						Pregunta 3			Pregunta 4	
	¿En su ámbito laboral cuenta con alguna metodología para el análisis de la información?		Si contestó afirmativamente la pregunta anterior, señale que modelos o metodologías utiliza:						Las bases de datos que se diseñan en su ámbito laboral, se utilizan para explotarse:			¿Utiliza algún modelo o método para el análisis de la información histórica?	
	SI	NO	Modelo Relacional	Modelo de Red	Modelo Jerárquico	Modelo Distribuido	Otra	¿Cúal?	Inmediatamente	A futuro	Ambas	SI	NO
1		x					x	trabajo en la que se toman en cuenta los costos, tiempos y recursos	x			x	
2	x		x				x	Yourdon, en el es una combinación de ambas, dependiendo de la empresa a la que se le este dando la consultoría		si aunque en algunas ocasiones es a futuro			x
3	x		x		x				x			x	
4	x						x	Oracle Methodo una vez se hace por cada analista de la consultoría			x	x	
5	x						x				x	x	
6	x						x	Propia	x				x
7		x						El trabajo se realiza directamente	x		en ocasiones		x
8	x		x				x	Propia	x				x
9	x						x	Combinación de lo de Yourdon			x	x	
10	x		x								x		x
11	x				x				x				x
12	x						x	Propia			x	x	
Totales	10	2	4	0	2	0	8		7	0	6	6	6

Pregunta 3

Si contestó afirmativamente, diga cuales utilizó:

Clasificación	Asociación	Arboles de Decisiones	Redes Neuronales	Algoritmos paralelos	Híbridos	Reglas de Inducción	Estadísticas	Fractal	Clustering	Secuencias	Algoritmos genéticos	Mem. Bor.	Reglas de Clasificación	Patrones generales...	Sist. Visualización	Técnicas Difusas	Sist de Inf. Geográfica	Otros	
x	x																		
x	x						x												
																		x Oracle Methods	
		x																	
x																			
																		x	
																		x	
									x									x	
3	2	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	3	1

Pregunta 6 ¿Cuántos métodos utilizados son eficientes para el apoyo a la toma de decisiones y el conocimiento del negocio?		Pregunta 7 ¿Utiliza alguna herramienta automatizada para la extracción y manejo de la información?			Pregunta 8 ¿Cree que el uso de herramientas que le ayuden a generar información adicional o que se obtiene comúnmente de sus sistemas de información, le podrán ayudar a conocer más a detalle su negocio e incrementar la productividad del mismo.			Pregunta 9 ¿Conoce el término "minería de datos"? Si es así, de una definición		Pregunta 10 ¿Conoce alguna persona que se dedique a la investigación y promoción de la "minería de datos" en nuestro País?	
SI	NO	SI	¿Cuál?	NO	¿Por qué?	SI	NO	¿Por qué?			
x		x	Sybase, y se explota con Visual Basic				x	Los sistemas que hay tienen la funcionalidad que necesitan, o si es necesario algo se hacen las modificaciones e los sistemas.	se refiere a mucha información, que se encuentra en muchos lados. Relación entre distintos sistemas, información que se utilice comúnmente.		NO
	x	x	En algunas empresas es Oracle en otras es Sybase			x		Es muy importante conocer a detalle mi negocio, para saber que fortalezas y debilidades tengo.	No, pero me imagino que ayuda a extraer información oculta.		NO
x		x	realizadas en la empresa, se usará Oracle Data Mart			x		La toma de decisiones es más sencilla cuando el cúmulo de información es menor y nos refleja con poca la realidad de un todo.			NO
x		x	Oracle Applications, OFA, Oracle Discoverer			x		puede estar involucrada en un negocio, el éxito consiste en utilizar las herramientas adecuadas para observar la información requerida en el momento que se requiere.	El proceso de encontrar patrones y regularidad en conjuntos de datos. El análisis de datos en busca de relaciones que no hayan sido encontradas.		NO
	x	x	Oracle 7.0			x		reconocer la información almacenada, pero no se aprovecha como se pudiera.	NO		NO
	x				Es dependiendo del sistema que es está desarrollando	x		Deben ser herramientas que traigan beneficios a mi organización en todos los aspectos, económico, administrativo, productivo entre otros.	NO		NO
	x	x	Informix 7.0			x		Muchos sistemas desarrollados en otras empresas son recomendables que en toda organización se tuviera una herramienta de este tipo.	NO		Si, trabajo en la comisión Nacional de Valores
x		x	Oracle			x		Aunque no contamos con alguna herramienta de este tipo.	NO		NO
	x	x	Sybase y Java			x		En ocasiones las utilizamos, pero no siempre. Cuando se ha experimentado herramientas de este tipo para aquí no contamos con alguna.	me imagino que se refiere a la búsqueda de información para el conocimiento del negocio.		NO
	x	x	Progress			x			NO		NO
	x	x	Informix 7.0			x		En alguna ocasión utilizamos una herramienta así y nos ayudo mucho a la toma de decisiones. Porque nos ayudo a conocer más a fondo nuestro negocio.	es la indagación de información en una base de datos que nos ayudará a tomar decisiones.		NO
x						x			Es buscar información oculta para que nos ayude a conocer nuestro negocio.		NO
5	7	10				11	1				

Observaciones
Empresa Representada
ETSE
KCS
Conel II
General Electric Capital Services
Soll Comó
INPOINTER
IAI S.A.
IAI S.A.
INPOTEC
BORLAND
AFIANZADORA INSURGENTES
BANCOMER



---

***Personas a entrevistar:***

Lic. Manuel Nicolás Ramírez Contreras

Lic. Oscar Carmona Cruz

Ing. Agustín Melchor Rangel

Lic. Daniel Pérez Hernández

Lic. Maricela Montes Márquez

Mat. Antonio González

Lic. Oscar Ramírez

#### **4.11 CONCLUSION Y APROBACIÓN DE LA HIPÓTESIS**

Tomando en cuenta los resultados obtenidos por medio de las encuestas he llegado a la conclusión de que en muchas organizaciones se hacen uso de herramientas automatizadas para el apoyo en la toma de decisiones, como lo son Sybase, Oracle, entre otros, pero también se nota que no se cuentan con métodos o procedimientos adecuados para explotar en un mayor grado la información histórica, y en su gran mayoría solo se explota la información de manera inmediata y no se toma muy en cuenta la información histórica de la organización.

En la mayoría de las encuestas se puede observar que aunque si se cuenta con metodologías para el análisis de la información, también consideran que es muy útil contar con otras herramientas que coadyuven a la toma de decisiones y al conocimiento de la empresa para que con esto se llegue a la elevación integral de la productividad de la empresa.

Por otra parte las personas entrevistadas dicen que si es bueno contar con herramientas adicionales para el apoyo de la toma de decisiones y sobre todo que les ayuden a conocer de una forma más profunda su empresa, para que esto coadyuve a una toma de decisiones más certero, pero lamentablemente no en muchas organizaciones se conocen o utilizan este tipo de Herramientas, como podría ser la minería de datos ya que también se denota que la gran mayoría de las personas entrevistadas tienen una idea y muy pocas conocen y aplican este término.



# Marco Instrumental

---

## 5. MARCO INSTRUMENTAL

Como resultado de la investigación realizada doy, a continuación, algunas de las actividades realizadas para dar a conocer los resultados obtenidos de la misma

5.1 Debido a que considero que el tema de minería de datos es de actualidad, doy una propuesta de un breve temario para la carrera de licenciatura en informática de la Facultad de Contaduría y Administración el cual podrá ser integrado al temario general de la materia de Bases de Datos.

Dicha propuesta ha sido presentada a la jefa de informática avanzada de la carrera de la Licenciatura de la Facultad, la Licenciada Reyna Mier V. (ver Anexo 1)

5.2 Elabore un breve artículo para ser publicado en alguna revista.  
(ver Anexo 2)

Dicha propuesta ha sido presentada a el Licenciado Gustavo Almaguer Pérez, Jefe de la Gaceta de la Facultad de Contaduría y Administración.

5.3 Cree una página web con la información necesaria para conocer las generalidades de la "minería de datos". Dicha página se encuentra localizada en el site de internet que se menciona a continuación :

<http://orbita.starmedia.com/~siteesme/>

Un ejemplo de las páginas mostradas puede ser visto en el anexo 3 de esta tesis.

5.4 He preparado el material necesario para realizar pláticas o conferencias en diversas organizaciones y/o instituciones educativas para que conozcan los beneficios de la "minería de datos".

Un ejemplo de dicho material podría ser observado en el anexo 4 de esta tesis.



# Conclusiones

---

## 6. CONCLUSIONES

1. El antecedente de la "minería de datos" son las bases de datos, con lo que conforma su recolección y utilización de la información almacenada en dichas bases de datos.

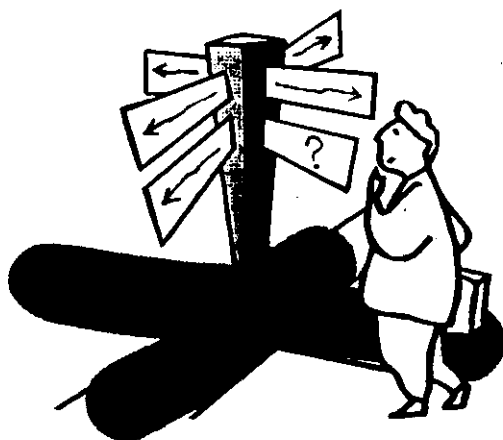
2. La "minería de datos", se puede considerar como el resultado de una largo proceso de captación y acumulación de datos a través del tiempo, el almacenamiento y acceso en diferentes medios y / o formas hasta llegar al análisis y de explotación de los datos para convertirlos en información que coadyuve a la toma de decisiones.

3. En la evolución de la información nos encontrarnos a nuestro paso la creación de bancos de datos, bases de datos, data warehouse hasta lo que hoy en día se conoce como minería de datos. Todos los avances en la información tienen como objetivo primordial proporcionar información que apoye al ser humano en la difícil tarea de tomar decisiones certeras que lo orienten hacia el logro de sus objetivos.

4. La "minería de datos" es el proceso cuyo objetivo es descubrir, extraer y almacenar información relevante y previamente ignorada de amplias bases de datos a través de programas de búsqueda de criterios definidos por el analista, en grandes cantidades de datos.

5. La "minería de datos" se ha convertido en una fuente importante de creatividad, y apoyo; sobre todo para aquellos que ofrecen herramientas auxiliares para los tomadores de decisiones, lo cual lo puede ser observado en la gran cantidad de software minero existente en el mercado que se va acrecentando cada día más.

6. Al tomar a la "minería de datos" como una herramienta para toma de decisiones, pretende orientar a los administradores y ejecutivos a que con esta herramienta pueden tomar mejores decisiones y llegar en cierto momento a elevar la productividad de la empresa, aunque las metodologías utilizadas o sistemas sean distintos en diversas organizaciones, la "minería de datos" bien aplicada siempre coadyuvará a la toma de decisiones.



# Anexos

# **ANEXO 1**

## **PROPUESTA DE TEMA**

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CONTADURÍA Y ADMINISTRACIÓN




Lic. Reyna Mier Valdéz  
Jefa de la División de Informática Avanzada  
PRESENTE

Por medio de la presente, y debido a al incremento del conocimiento tecnológico que se va dando día con día, me permito hacer una propuesta para agregar al temario de la materia de bases de datos un tema de actualidad e interés para los futuros profesionales de la materia, dicha propuesta es dada como resultado de la tesis profesional denominada "La minería de datos una herramienta para la toma de decisiones", la cual su servidora ha sido responsable, conjuntamente con el Dr. Ricardo Rivera Soler (asesor de la misma).

Anexo a la presente carta el formato del temario propuesto.

Agradeciendo de antemano su atención y esperando que dicha propuesta sea tomada en cuenta, estoy a sus ordenes para cualquier duda y si lo desea para impartir alguna platica referente a este tema y en su caso para impartir este tema dentro de la materia de bases de datos.

Recibi  
Reyna Mier U.  
Agosto 21, 2000

  
Esmeralda Martínez Montes  
Tels. 56 62 72 36 oficina  
56 17 57 66 particular



**INFORMÁTICA IV  
(BASES DE DATOS)**

CLAVE:

PLAN:

LICENCIATURA(SEMESTRE): INFORMÁTICA (5°.)

DEPTO. ACADÉMICO: INFORMÁTICA AVANZADA

ÁREA: BASES DE DATOS

REQUISITOS: INFORMÁTICA III

CRÉDITOS: 8

HORAS POR CLASE: 2

CLASES POR SEMANA: 2

HORAS POR SEMESTRE: 3

**OBJETIVO GENERAL: DAR A CONOCER UNO DE LOS CONCEPTOS DE ACTUALIDAD EN EL ÁMBITO DE LAS BASES DE DATOS, ASÍ COMO PROPORCIONAR UNA CLARA VISIÓN DE LA UTILIDAD QUE LA APLICACIÓN DE ESTE CONCEPTO PUEDA LLEGAR A TENER EN UNA ORGANIZACIÓN PARA LA TOMA DE DECISIONES.**

HORAS	TEMÁTICA	OBJETIVOS EDUCACIONALES	SUGERENCIAS DIDÁCTICAS	REFERENCIAS BIBLIOGRÁFICAS
30 min.	1. CONCEPTO a. Origen etimológico b. ¿Qué es la minería de datos?	1. PARTICULAR DE LA UNIDAD Al término de esta unidad el alumno conocerá el origen etimológico y conceptual de la minería de datos. 2. ESPECÍFICOS El alumno será capaz de: Explicar y entender el término minería de datos.	Exposición oral y/o audiovisual por parte del profesor.	1
30 min.	2. EVOLUCIÓN a. Origen b. Evolución c. Beneficios	1. PARTICULAR DE LA UNIDAD Al término de esta unidad el alumno conocerá el origen evolución y utilidad de la minería de datos. 2. ESPECÍFICOS El alumno será capaz de: Explicar y entender el origen de la minería de datos, así como su utilidad.	Exposición oral y/o audiovisual por parte del profesor.	2
90 min.	3. TÉCNICAS a. Redes neuronales b. Descubrimiento de asociaciones c. Clasificación d. Clustering e. Descubrimiento secuencial	1. PARTICULAR DE LA UNIDAD Al término de esta unidad el alumno conocerá algunas de las principales técnicas utilizadas en la minería de datos. 2. ESPECÍFICOS El alumno será capaz de : Explicar y aplicar las técnicas utilizadas en	Exposición oral y/o audiovisual por parte del profesor.	3

		la minería de datos.		
30 min.	<p>4. SOFTWARE EXISTENTE</p> <p>a. Características del software de minería de datos</p> <p>b. Algunos software en el mercado</p> <p>c. Futuro de la minería de datos</p>	<p>1. PARTICULAR DE LA UNIDAD</p> <p>AL término de la presente unidad el alumno conocerá las características generales de los software de minería de datos.</p> <p>2. ESPECÍFICOS</p> <p>El alumno será capaz de:</p> <p>Conocer algunos software de minería de datos.</p> <p>Analizar el futuro de la minería de datos</p>	Exposición oral y/o audiovisual por parte del profesor.	3

**BIBLIOGRAFÍA:**

1. Pieter Adrians and Dolf Zantinge, Data Mining, Addison Wesley; 1ª. Edición , 1996, ISBN 0-201-40380-3
2. Tesis, Esmeralda Martínez Montes, Minería de datos una herramienta para la toma de decisiones, 2000.
- 3 Michael J.A. Berry and Gordon Linoff, Data Mining Techniques: For Marketing, Sales and Customer Support, Wiley Computer Publishing; 1ª edition, 1997, ISBN 0-471-17980-9

**PERFIL PROFESIOGRÁFICO DEL DOCENTE:**

Los requisitos que debe reunir el docente para impartir la signatura de **INFORNÁTICA IV** son los siguientes:

**ACADÉMICOS:**

Se Licenciado en Informática

**PROFESIONALES**

Tener experiencia y/o conocimiento de la minería de datos.

**DOCENTES**

Acreditar dos cursos de Didáctica de 30 horas en la Coordinación de Calidad Académica de la Facultad de Contaduría y Administración.

Conocer las asignaturas de:

Informática I a Informática III

**ANEXO 2**  
**ARTÍCULO PARA REVISTA**

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE CONTADURÍA Y ADMINISTRACIÓN




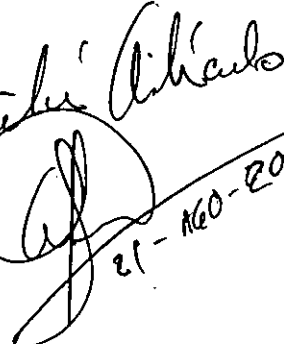
L.A. Gustavo Almaguer Pérez  
PRESENTE

Por medio de la presente, me permito, proporcionarle un pequeño artículo de revista, que ha sido uno de los resultados obtenidos de la tesis profesional denominada "La minería de datos una herramienta para la toma de decisiones", la cual su servidora ha sido responsable, conjuntamente con el Dr. Ricardo Rivera Soler (asesor de la misma); para que sea tomado en cuenta para su posible publicación en la gaceta de la facultad.

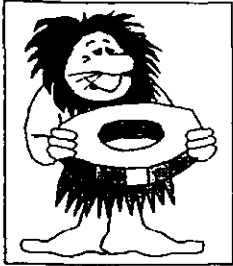
Anexo a la presente carta el artículo, para ser tomado a su consideración.

Agradeciendo de antemano su atención y esperando que dicha propuesta sea tomada en cuenta, estoy a sus ordenes para cualquier duda o aclaración.

  
Esmeralda Martínez Montes  
Tels. 56 62 72 36 oficina  
56 17 57 66 particular

  
21-1160-2000

## "La Minería de las Computadoras"



Seguramente en muchas ocasiones te has preguntado "¿cómo puedo hacer para conocer más a detalle mi empresa?, ¿esto me puede ayudar para incrementar la productividad? ¿qué necesito hacer?"; para dar respuesta a estas y otras preguntas te voy a explicar un nuevo concepto denominado "minería de datos".

Este término surge principalmente de la necesidad de explorar y explotar en un 100% los recursos contenidos en una o unas bases de datos; y sobre todo por la curiosidad de algunas personas por conocer que es lo que podría ocultar; "ya vez como en ocasiones la curiosidad no es mala"; y que esta información es de gran utilidad para las empresas.

Dentro de una organización los principales beneficiarios de la "minería de datos" son los gerentes empresariales y los ejecutivos, ya que los resultados que arroja esta herramienta los ayuda a la toma de decisiones.

Pero te seguirás preguntando ¿qué es la minería de datos", pues yo la concibo como una combinación de técnicas semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos, y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos y que pueda resultar más provechosa que los datos iniciales.

La minería de datos al ser aplicada correctamente, auxilia los usuarios empresariales en el procesamiento de grandes cantidades de datos, para descubrir "relaciones insospechadas", por ejemplo, entre productos y clientes o patrones de compra de los clientes. La meta de la "minería de datos" es descubrir relaciones estratégicas competitivas para controlar la participación en el mercado y las utilidades; esto se da debido a que los seres humanos contamos con la agudeza de percibir, excepciones y anomalías, pero no tenemos la potencial capacidad para inferir relaciones en volúmenes de información.



Existen diversas técnicas utilizadas para la minería de datos entre las cuales se encuentran:

- ⊗ **Redes Nueronales:** esta técnica se utiliza para identificar patrones y prever tendencias basadas en comportamientos previamente identificados. El mercado de valores es un campo perfecto para analizar las tendencias, cuando se analiza el precio de un producto que se cotiza en bolsa los últimos seis meses se puede predecir una cuota de mercado para las próximas semanas. La base de este tipo de procesamiento es lo que se asimila al sistema nervioso central. El conocimiento se puede aprender de una serie de datos ampliamente dispares, complejos e imprecisos. Existen 3 capas, la capa inferior recibe los datos de entrada, la capa oculta (intermedia) realiza el trabajo, y la exterior presenta las salidas al analista; esta forma de procesar la información asemejan las neuronas del cerebro, de ahí la terminología "red neuronal".
- ⊗ **Descubrimiento de asociaciones:** esta técnica intenta descubrir ocurrencias similares de los valores de los datos de los registros, y produce una salida que podría ser expresada como una regla; a las cuales se les conoce como reglas de asociaciones.
- ⊗ **Clasificación:** consiste en la agrupación de datos basándose en una serie de características predefinidas por el analista con anterioridad. Esta técnica examina los datos ya clasificados y agrupados mediante la aplicación de una regla de pertenencia; dicha regla puede tener un componente temporal, un componente geográfico o un componente cuantitativo.
- ⊗ **Clustering:** también conocido como agrupación; consiste en agrupar información de datos similares partiendo de un conjunto más grande de datos; esta técnica descubre agrupaciones a medida de que trabaja los datos de entrada.
- ⊗ **Descubrimiento secuencial:** esta técnica intenta encontrar patrones entre sucesos que ocurren en una progresión a lo largo de un tiempo.

En la actualidad existen muchos fabricantes de software, entre los que tenemos:

- ⊗ NEOVISTA dicha compañía tiene en el mercado distintos softwares de minería de datos entre los cuales se encuentran DecisionNet, DecisionAccess, DecisionSeries.
- ⊗ Otro software es el llamado DataMind que es uno de los más utilizados en Estados Unidos.

Por último te proporciono unas direcciones de internet en donde podrás encontrar mas información sobre "minería de datos".

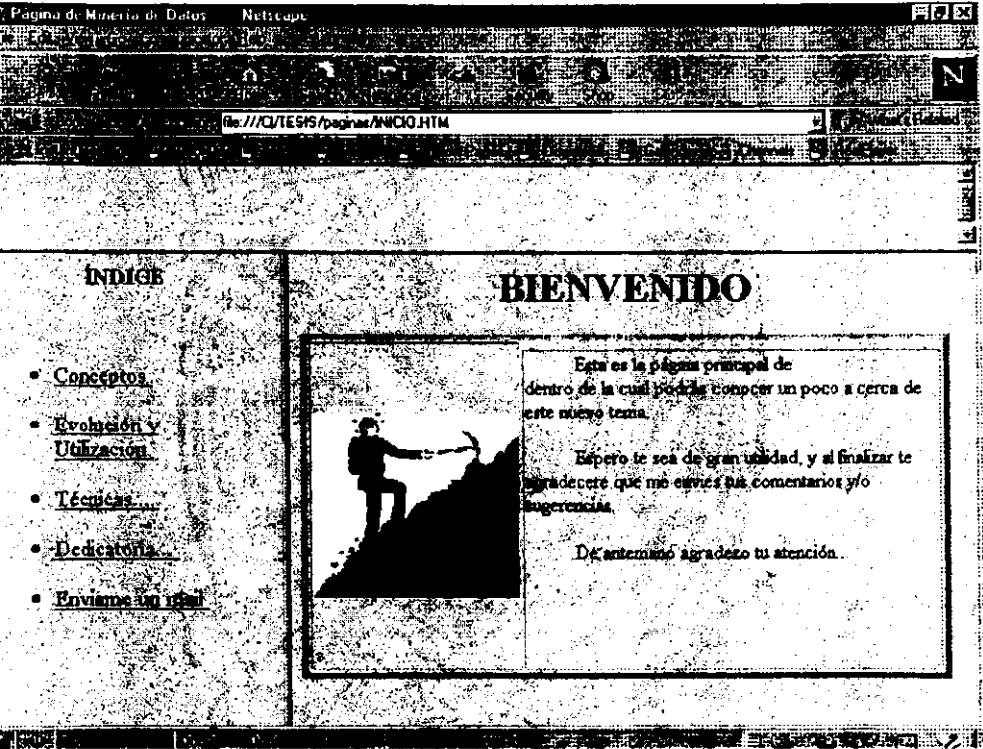
Fuente	Dirección
Ultragem Data Mining Co.	<a href="http://ultrgem.com">http://ultrgem.com</a>
IBM Corporation	<a href="http://www.software.ibm.com/sq/issues/uoL4/data.htm">http://www.software.ibm.com/sq/issues/uoL4/data.htm</a>
Silicon Graphics Inc.	<a href="http://www.sgi.com/Products/hardware/challenge/Datamining">http://www.sgi.com/Products/hardware/challenge/Datamining</a>

# **ANEXO 3**

# **PÁGINA WEB**



Estos son algunos ejemplos de lo que podrán encontrar en la página Web que desarrollé acerca de "Minería de Datos".



## INDICE

- [Conceptos](#)
- [Evolución y Utilización](#)
- [Técnicas...](#)
- [Dedicatoria...](#)
- [Enviar un mail](#)

## Definiciones de Minería de Datos

### Etimologías De algunos verbos Propia

### Etimologías

- Del griego *Myo* --> Lo oculto  
Mina --> Mina
- Derivados: *Minador, mineral, minar, minera, mineral, minería, minero*
- *De* --> Proviene del latín. Preposición Elemento de relación que introduce tanto complementos del verbo como del nombre. Denota propiedad, posesión y pertenencia: casa de mi padre; los árboles del jardín.
- *Dato* --> Del latín "datum". Sustantivo. Lo que se da. Antecedente que se da para llegar a un conocimiento.

# MINERIA DE DATOS

## INDICE

## Técnicas de Minería de Datos

- [Conceptos](#)
- [Evolución y Utilización](#)
- [Técnicas...](#)
- [Dedicatoria...](#)
- [Envíame un mail](#)

- [Redes Neuronales](#)
- [Descubrimiento de asociaciones](#)
- [Clasificación](#)
- [Clustering](#)
- [Descubrimiento secuencial](#)

### Redes Neuronales

La minería basada en redes neuronales está específicamente indicada para identificar patrones y prever tendencias basadas en comportamientos previamente identificados. Una tendencia identifica un movimiento en el hábito basado en el comportamiento anterior. El

# MINERIA DE DATOS

## INDICE

- [Conceptos](#)
- [Evolución y Utilización](#)
- [Técnicas...](#)
- [Dedicatoria...](#)
- [Envíame un mail](#)

## ENVIAME UN MAIL

Nombre:

Teléfono:

E-mail:

Sexo:  
 Masculino  
 Femenino

País:

Comentario y/o Sugerencia:

# Anexo 4

# Presentación

A continuación se mostrarán algunos ejemplos de las diapositivas que serían utilizadas para alguna plática o conferencias.

# Minería de Datos

## Historia

«Surge de la necesidad de explorar y explotar al 100% los recursos de una base de datos, y sobre todo por curiosidad de la persona por conocer que es lo que podrá ocultar una base de datos y que puede llegar a ser información de gran utilidad para una empresa.»

# Minería de Datos

## Antecedentes

La evolución de los **Manejadores de Base de Datos** se da de la siguiente forma:

- **SQL**: lenguaje para el acceso a la información almacenada en una base de datos relacional.
- **Data Ware Hoisting**: Conjunto de datos integrados orientados a una materia, que varían con el tiempo y que son transitorios, los cuales soportan el proceso de toma de decisiones de una administración.
- **Data Mining**: Búsqueda de relaciones inspeccionadas de información histórica almacenada en una base de datos, por medio de técnicas semi automáticas.

# Minería de Datos

## Utilidad

La minería de datos ayuda a los usuarios empresariales en el procesamiento de vastas reservas de datos para descubrir "relaciones insospechadas", por ejemplo, entre productos y clientes o patrones de compra de los clientes. La meta de la minería de datos es descubrir "revelaciones estratégicas competitivas" para controlar la participación en el mercado y las utilidades. Los seres humanos tenemos la agudeza para percibir excepciones y anomalías, pero no tenemos la potencial capacidad para inferir relaciones en volúmenes de datos.



# Minería de Datos

## Concepto

En Estados Unidos se conoce con el término Data Mining.

Yo lo conceptualizo como:

*"una combinación de técnicas semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos y que pueda resultar más provechosa que los propios datos de partida."*

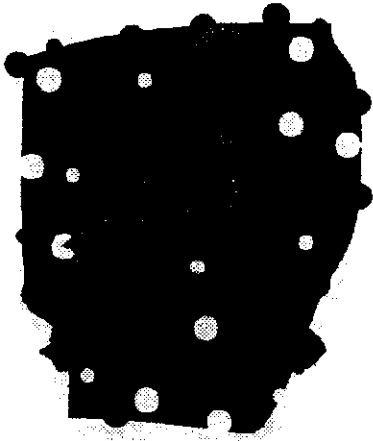
# Minería de Datos

---

## Técnicas :

- \* Redes Neuronales
- \* Descubrimiento de asociaciones
- \* Clasificación
- \* Clustering
- \* Descubrimiento secuencial





# Glosario

---

## 8. - GLOSARIO

<i>Término</i>	<i>Definición</i>
<i>A</i>	
<i>Actualización de Datos</i>	La actividad de actualizar continuamente el contenido de los datos de la base de datos a partir de las fuentes de datos después de terminar la carga inicial. Este es un proceso continuo donde las operaciones de datos de hoy se convierten en los datos históricos de mañana.
<i>Adición</i>	La actividad de combinar datos de diversas tablas para formar una unidad de información más compleja que se necesita con frecuencia para responder a consultas a la base de datos con mayor facilidad y rapidez. Como todas las bases de datos operacionales guardan datos en la unidad más sencilla en que es posible (normalización) para prevenir anomalías de actualización y eliminación.
<i>Agrupamiento, agrupar</i>	Agrupar tablas es el acto de requerir que las tablas físicas de una base de datos residan físicamente juntas en el medio de almacenamiento. Las búsquedas secuenciales previas producen espectaculares en el desempeño cuando se accesa una gran cantidad de filas o renglones.
<i>Ambiente administrado de consultas</i>	Capacidad de procesamiento informático donde las herramientas esconden la complejidad de las estructuras de datos con una "capa semántica" de términos y reglas empresariales.
<i>Análisis de datos estáticos</i>	Comparar un dato de valor estático con otro.

<i>Análisis dinámico de datos</i>	Comparar valores de datos de diferentes fuentes de datos y dimensiones distintas.
<i>Análisis multidimensional de datos</i>	de Análisis de simultáneo de múltiples dimensiones de datos.
ANSI	<p>(American National Standards Institute). Instituto Nacional de Normas Americanas. Organización de establecimiento de normas patrocinadas por la industria. Fue fundada en 1918 y fijaba las normas industriales de los Estados Unidos y su correspondencia con las establecidas por la OSI (International Organization for Standardization). ANSI determina las normas relativas al hardware, en puntos tales como protocolos de nivel de enlace, posiciones y significado de las pastillas en los chips, registro en disco y algunas normas para el software.</p>
API	<p>(Application Program Interface) Interfaz de programador de aplicaciones. Especificación de la comunicación entre un programa de aplicaciones y uno de utilidad.</p>
<i>Aplicaciones nuevas</i> <i>producción</i> <i>Asociaciones</i>	de Aplicaciones operacionales nuevas.
	<p>Técnica utilizada en la minería de datos, el objetivo de la asociación es encontrar tendencias a través de un gran número de transacciones que pueden ser utilizadas para comprender y explotar los patrones naturales de compra, ajustar inventarios, analizar clientes, identificar servicios financieros que la gente compra.</p>
<i>B</i>	
<i>Banco de datos</i>	<p>Sin un conjunto de registros listados uno después de otro, almacenados en archivos. Es una colección de datos que no tienen ningún tipo de relación entre ellos. (verificar)</p>

<i>Base de datos</i>	<i>Una base de datos es una colección de datos, que guardan una relación entre sí.</i>
<i>Base de datos activa de muestra</i>	Una base de datos que es un pequeño subconjunto de datos operacionales e históricos, pero que aporta una muestra representativa válida para propósitos de consulta, soporte de decisiones, pronóstico y simulación. La base de datos se actualiza periódicamente para asegurar que los cambios en los datos reales se transfieran a la base de datos muestra.
<i>Base de datos activa de muestra</i>	Una muestra estadísticamente correcta de un conjunto de datos.
<i>Base de datos multidimensional</i>	Una base de datos diseñada alrededor de un conjunto de dimensiones, se usa en el análisis multidimensional.
<i>Base de datos operacional</i>	Una base de datos que apoya sistemas de software que están soportando normalmente a las operaciones empresariales de una organización. También llamada base de datos OLTP.
<i>Bloque de fuentes de datos</i>	El componente de la arquitectura de referencia del data warehouse que representa las actividades que tratan con las bases de datos operacionales y las fuentes externas que aportan datos.
<b>G</b>	
<i>Capa de administración de datos.</i>	El componente de la arquitectura de referencia del data warehouse que maneja las opciones de administración de datos.
<i>Capa de infraestructura</i>	Componente de la base datos relacionado con las plataformas y el soporte computacional y el hardware y software de ambiente adquiridos en la industria que se requieren para apoyar a la empresa.
<b>CASE</b>	Siglas de Computer Aided Software Engineering (ingeniería de software apoyada por computadora). Una metodología

---

	de desarrollo de software soportada por herramientas basadas en computación para el apoyo a tareas de análisis, diseño y desarrollo.
<i>Ciclo de vida de desarrollo</i>	En flujo de actividades que representa un proyecto de implementación y despliegue de la información.
<i>Clasificación</i>	<i>Técnica de análisis de minería de datos, emplea un conjunto de transacciones preclasificados para desarrollar un modelo que pueda clasificar los registros de una gran base de datos. Este tipo de análisis es particularmente adecuado para aplicaciones de análisis de riesgo y detección de fraudes. Emplea frecuentemente algoritmos basados en arboles o redes neuronales.</i>
<i>Cluster</i>	<i>Segmentación de problemas similares, esta técnica de minería de datos es a menudo uno de los primeros pasos en el análisis de la minería de datos. Identifica grupos de relaciones entre los registros que pueden ser usadas como punto de partida para futuras exploraciones en las relaciones; además ayuda a determinar las características de los segmentos con respecto a los resultados deseados.</i>
<i>CODASYL</i>	<i>(CO)Nference on DAta SYstems Languages) Formada por representantes de las áreas de la defensa y gobierno de los Estados Unidos, así como representantes mundiales de las áreas de negocios, con el objetivo inicial de proponer un lenguaje de programación de alto nivel para usar en el desarrollo de programas aplicables a negocios. Este objetivo inicial fue alcanzado con la publicación del primer COBOL.</i>
<i>Consulta</i>	Una petición formal y claramente especificada de información planteada por

---

un usuario o una herramienta operada por un usuario.

## *D*

*Data warehouse*

*Almacén de datos, es una colección de datos integrales, variables en el tiempo, no volátiles y orientados a temas importantes para el soporte a la toma de decisiones en la administración de la organización.*

*Datawarehousing*

*Es el proceso de extracción y transformación de datos obtenidos en fuentes operacionales (LTP) llevándolos a una base de datos centralizada reconocida como un data warehouse. Por lo que una vez en el repositorio es explotada usando herramientas para la toma de decisiones.*

*Dato*

*Del latín "Datum", lo que se da. Antecedente necesario para llegar al reconocimiento exacto de una cosa o para deducir las consecuencias legítimas de un hecho. Se emplea el término dato para referirse a los valores registrados físicamente en la base de datos.*

*Datos extraños o irregulares*

*Son datos desconocidos o no comunes en la información que se este manejando.*

*DBMS*

*(Data Base Management System) Es una pieza sofisticada de software, la cual soporta la creación y administración de sistemas de bases de datos. Esto significa que una base de datos y un DBMS de propósito específico pueden ser usadas para reemplazar un DBMS de propósito general. Un sistema de bases de datos constituye en sí mismo un sistema completo de información, más comúnmente, es un componente de una gran sistema con otros componentes que incluyen programas que hacen uso de sus facilidades.*

*Depósito*

*Compendio de todas las definiciones y*



---

especificaciones tecnológicas que son importantes para las aplicaciones y las bases de datos

Un compendio de definiciones y especificaciones categóricas de datos y sus relaciones.

*Diccionario de datos*

*E*

*EAP*

Siglas de Enterprise Architecture Planning (Planeación de la arquitectura de la empresa), una metodología estructurada para la arquitectura y el diseño de sistemas de información. Los pasos importantes del diseño de la arquitectura con la formulación de una arquitectura de datos, una arquitectura de aplicaciones y una arquitectura tecnológica.

*Proceso por el que un remitente convierte un mensaje inteligible para receptores autorizados en ininteligible para otras personas no autorizadas.*

Son las especificaciones de cómo está organizada una base de datos.

Es una forma de ver algo complejo y hacerlo entendible usando un conjunto de analogías simplificadas para ayudar a convertir una solución compleja en componentes pequeños.

*Una especificación para hardware o software que esta extensamente usada y aceptada o que esta sancionada por una organización de las normas.*

Modalidad de descubrimiento del análisis de datos, o analizar datos de detalle para revelar relaciones, patrones y asociaciones insospechados o desconocidos.

*Encriptación*

*Esquema de bases de datos*

*Estructura*

*Extándar*

*Extracción de datos*

*F*

*FAQ*

Siglas de Frequently Accessed Query (consulta frecuentemente accesada)

FMS

*(File Management System) Sistema manejador de archivos. Es le único modelo que describe cómo son almacenados los datos en el disco. En este modelo, cada campo o dato es almacenado secuencialmente sobre el disco en un gran archivo. Fue el primer método usado para almacenar datos en una base de datos computarizada y la simplicidad es su una única ventaja. Los productos existentes actualmente sobre este modelo son de bajo nivel. Sus desventajas son claras. Primero, no hay indicación de la relación entre los elementos más que la secuencia de almacenamiento. El programador, y algunas veces el usuario, tiene que conocer exactamente como son almacenados los datos en el archivo para poder manipularlos.*

*G*

*Indizado*

Técnica de usos frecuente para optimizar el desempeño de la base de datos mejorando el método de acceso para buscar y recobrar registros de la base de datos.

*Información*

*Cualquier mensaje capaz de ser representado y manipulado. Resultado del tratamiento de los datos de un programa presentado en un formato determinado.*

*Información histórica*

Información que se reopilo del pasado y que por lo general se archivo y se retiro de las bases de datos operacionales.

*Integración de datos*

La actividad de combinar datos de diversas fuentes a fin de representar una colección única de datos para el usuario.

*Inteligencia Artificial (IA)*

La ciencia de programar una computadora para ejecutar funciones que imitan la inteligencia humana.

*Interface*

*La conexión e interacción entre hardware, software y los usuarios. Frontera común*

---

*entre dos sistemas, dispositivos o programas.*

*IS o IT*

Siglas de Information Systems (Sistemas de Información) o de Information Technology (Tecnología de la Información). Técnicos que usa el departamento funcional responsable del desarrollo y despliegue de sistemas de información de una organización.

*ISO*

*(International Standards Organization)*  
*Modelo de referencia ISO/ OSI.*  
*Arquitectura general propuesta por la Organización Internacional de Normas para los Sistemas de Comunicación que permite la conexión de sistema abiertos.*

*III*

*Mercado de datos (data mart)*

Una implementación de data warehouse con un ámbito de datos y funciones de data warehouse más pequeño y restringido, que sirva a un departamento único o una parte de la organización. Una organización por lo general tiene varios mercados de datos.

*Método de desarrollo de cascada.*

El ciclo de desarrollo clásico que se usa tradicionalmente en proyectos de desarrollo de software. El nombre viene del hecho de que el flujo del agua en una cascada siempre tiene una sola dirección. En este método. Le flujo es desde la recopilación de requerimientos y el análisis hasta el desarrollo y entrega del sistema

*Método de desarrollo en espiral*

Un método iterativo de desarrollo del software en el que la funcionalidad del software se produce en etapas y las mejoras se identifican en realidad despegando sucesivamente el software con funcionalidad creciente pero bajo control estricto.

*Middleware*

El término común que se aplica al software

---

que intercambia información en forma transparente entre aplicaciones y bases de datos.

*Modelado de datos*

Actividad de representar las categorías de datos y relaciones entre ellas como una abstracción en forma de diagrama.

*O*

*OLAP*

*(Online Analytical processing database) Base de datos diseñada para acceder a datos resumidos. Usando técnicas especiales de indexación, procesa consultas que pertenecen a grandes cantidades de datos más rápido que una base de datos transaccional.*

*OLTP*

Siglas de On - Line Transaction Processing (procesamiento de transacciones en línea). El término se utiliza para definir cualquier sistema de software que reúne datos usando las transacciones (en el momento en que ocurren) entre la fuente de datos y la base de datos.

*R*

*RDBMS*

Siglas de Relational Database Management System (sistema de administración de base de datos relacional). Un sistema para almacenar datos construido alrededor del modelo relacional con base en tablas, columnas y vistas.

*Redes Neuronales*

*Técnica de análisis de minería de datos que construye modelos atendiendo los patrones en los datos que se utilizan. Son modelos predictivos que basados en principios similares a aquellos que rigen el cerebro humano. En una red de nodos (neuronas), cada nodo recibe una entrada y envía una salida a los nodos subsecuentes basándose en lo que recibió como entrada. Una vez que*

---

*la red neuronal ha sido validada, puede ayudar a analizar y predecir eventos a partir de entradas de datos nuevos.*

*Refinado de datos*

Las actividades colectivas aplicadas a los datos de las fuentes a fin de prepararlos.

*Registro*

*Conjunto de campos relacionados que almacenan información a cerca de un elemento.*

*Reporte*

Documento que se produce en respuesta a una consulta.

*S*

*Sistema cliente - servidor*

Un sistema de aplicación de software donde el procesamiento de la aplicación lo proporcionan en conjunto dos componentes arquitectónicos claramente separados: el cliente y el servidor.

*Sistemas de visualización*

*Técnica de análisis de minería de datos que permite hacer descubrimientos analizando los datos de manera gráfica con muchas variables, y después ver patrones y relaciones que sería muy difícil determinar mediante algoritmos de máquina, sin importar las capacidades de cómputo del sistema.*

*SQL*

*(structured Query Language) lenguaje utilizado para realizar consultas a bases de datos relacionales. Desarrollado por IBM*

*U*

*UNIVAC*

*Primer computadora comercial exitosa, introducida en 1951 por Remington Rand.*

*V*

*Visualización de datos.*

Presentación de datos de una forma gráfica para facilitar el análisis de datos complejos y voluminosos.