Base de datos

|  |  |  |
| --- | --- | --- |
| **Grupo temático** | **Descripción** | **tópico** |
| Gestión de base de datos | Que es.  Es un conocimiento de la implementación de una base de datos, así como la extracción de información | * Implementación y consulta * Datamining |

Base de datos

Base de datos

Definir el término base de datos

* Definición de Bases de [Datos](http://www.monografias.com/trabajos11/basda/basda.shtml).- Un conjunto de [información](http://www.monografias.com/trabajos7/sisinf/sisinf.shtml) almacenada en [memoria](http://www.monografias.com/trabajos13/memor/memor.shtml) auxiliar que permite acceso directo y un conjunto de [programas](http://www.monografias.com/Computacion/Programacion/) que manipulan esos datos

Base de Datos es un conjunto exhaustivo no redundante de datos estructurados organizados independientemente de su utilización y su implementación en máquina accesibles en [tiempo](http://www.monografias.com/trabajos901/evolucion-historica-concepciones-tiempo/evolucion-historica-concepciones-tiempo.shtml) real y compatibles con usuarios concurrentes con necesidad de información diferente y no predicable en tiempo.

Ventajas de las bases de datos.-

1. Independencia de datos y tratamiento.

* Cambio en datos no implica [cambio](http://www.monografias.com/trabajos2/mercambiario/mercambiario.shtml) en programas y viceversa (Menor coste de [mantenimiento](http://www.monografias.com/trabajos15/mantenimiento-industrial/mantenimiento-industrial.shtml)).

1. Coherencia de resultados.

* Reduce redundancia :
  + Acciones lógicamente únicas.
  + Se evita inconsistencia.

1. Mejora en la disponibilidad de datos

* No hay dueño de datos (No igual a ser públicos).
* Ni aplicaciones ni usuarios.
* Guardamos [descripción](http://monografias.com/trabajos10/anali/anali.shtml) (Idea de catálogos).

1. Cumplimiento de ciertas [normas](http://www.monografias.com/trabajos4/leyes/leyes.shtml).

* Restricciones de [seguridad](http://www.monografias.com/trabajos/seguinfo/seguinfo.shtml).
  + Accesos (Usuarios a datos).
  + Operaciones ([Operaciones](http://www.monografias.com/trabajos6/diop/diop.shtml) sobre datos).

1. Otras ventajas:

* Más eficiente [gestión](http://www.monografias.com/trabajos15/sistemas-control/sistemas-control.shtml) de [almacenamiento](http://www.monografias.com/trabajos12/dispalm/dispalm.shtml).

Efecto sinérgico.

Colocar dibujo claroooooo

Describir La Arquitectura De Una Base De Datos  
Distintos Niveles de un SBD

Los SBD pueden ser estudiados desde 3 niveles distintos:  
1.- Nivel Físico.  
Es el nivel real de los datos almacenados. Es decir cómo se almacenan los datos, ya sea en [registros](http://www.monografias.com/trabajos7/regi/regi.shtml), o como sea..  
2.- Nivel Conceptual.  
Es el correspondiente a una visión de la base de datos desde el punto de visto del mundo real.

3.- Nivel Visión.  
Son partes del esquema conceptual. El nivel conceptual presenta toda la base de datos, mientras que los usuarios por lo general sólo tienen acceso a pequeñas parcelas de ésta. El nivel visión es el encargado de dividir estas parcelas. Un ejemplo sería el caso del empleado que no tiene por qué tener acceso al sueldo de sus compañeros o de sus superiores.

Modelos de Datos

Estos modelos son:

* Modelo Relacional de Datos
* Modelo de Red
* Modelo Jerárquico

1.- Modelo Relacional de Datos.  
Representa al mundo real mediante tablas relacionadas entre sí por columnas comunes. Ex.:

|  |  |  |
| --- | --- | --- |
| Num\_empleado | Nombre | Sección |
| 33 | Pepe | 25 |
| 34 | Juan | 25 |
| 35 | Kevin | 27 |

|  |  |
| --- | --- |
| Num\_sección | Nombre |
| 25 | Textil |
| 26 | Pintura |
| 27 | mantenimiento |

2.- Modelo de [Red](http://www.monografias.com/Computacion/Redes/).  
Representamos al mundo real como registros lógicos que representan a una entidad y que se relacionan entre sí por medio de flechas. Ex.:  
33 Pepe 25 Textil  
34 Juan  
3.- Modelo Jerárquico.  
Tiene forma de árbol invertido. Un padre puede tener varios hijos pero cada hijo sólo puede tener un padre. Ex.:  
Empresa  
25 Textil 26 [Pintura](http://www.monografias.com/trabajos13/histarte/histarte.shtml) 27 Construcción  
33 Pepe 34 Juan 36 Perico

Ex.: Tenemos dos entidades ([cliente](http://www.monografias.com/trabajos11/sercli/sercli.shtml) y cuenta), queremos que un cliente pueda poseer varias [cuentas](http://www.monografias.com/trabajos5/cuentas/cuentas.shtml), y que una cuenta pueda tener como titulares a varios [clientes](http://www.monografias.com/trabajos11/sercli/sercli.shtml). Usando el modelo jerárquico tendríamos que recurrir a una duplicación de los datos en el esquema.

Cliente Cuenta  
Cuenta Cliente  
Como ya se ha dicho esta duplicación sólo ocurriría a nivel esquemático, pero a nivel físico existiría un único [registro](http://www.monografias.com/trabajos7/regi/regi.shtml) por cliente, y por cuenta que relacionaríamos mediante varios punteros con todo lo necesario.

Cliente Cuenta Sucursal  
Con el modelo relacional podríamos tener ambas entidades definidas de la siguiente forma:  
Cliente = (Nº Cliente: Acceso Principal; Nombre, [Dirección](http://www.monografias.com/trabajos15/direccion/direccion.shtml), Nº Cuenta: Acceso Ajeno)  
Cuenta = (Nº Cuenta: Acceso Principal; Saldo)

Se podría hablar de otro modelo que está un nivel más arriba y que se denomina Modelo Entidad-Relación (E/R). Este modelo sólo realiza el [diseño](http://www.monografias.com/trabajos13/diseprod/diseprod.shtml), no realiza la implementación, por lo tanto una vez hecho el diseño se puede llevar al modelo relacional, de red o jerárquico. Define dos entidades y una forma de relación entre ellas.

Ex.:  
Nombre Dicción Num\_Cli Num\_cuenta Saldo  
CLIENTE Cliente - CUENTA  
Cuenta  
Entidad Relación Entidad

Clasificar Las Bases De Datos

Por Su Uso

TIPOS DE MODELOS DE DATOS Existen fundamentalmente tres alternativas disponibles para diseñar las bases de datos: el modelo jerárquico, el modelo de red y el modelo relacional.

MODELO JERARQUICO. Puede representar dos tipos de relaciones entre los datos: relaciones de uno a uno y relaciones de uno a muchos.  
MODELO DE RED. Este modelo permite la representación de muchos a muchos, de tal forma que cualquier registro dentro de la base de datos puede tener varias ocurrencias superiores a él. El modelo de red evita redundancia en la información, a través de la incorporación de un tipo de registro denominado el conector.  
MODELO RELACIONAL. Este modelo se está empleando con más frecuencia en la práctica, debido a las ventajas que ofrece sobre los dos modelos anteriores, entre ellas, el rápido entendimiento por parte de usuarios que no tienen conocimientos profundos sobre [Sistemas](http://www.monografias.com/trabajos11/teosis/teosis.shtml) de Bases de Datos.

Y Aplicación

BASES DE DATOS DISTRIBUIDAS. Las bases de datos distribuidas se están utilizando cada vez más en la misma medida en que se usan las arquitecturas de cliente-[servidor](http://www.monografias.com/trabajos12/rete/rete.shtml) y groupware. Los principales [problemas](http://www.monografias.com/trabajos15/calidad-serv/calidad-serv.shtml#PLANT) que se generan por el uso de la [tecnología](http://www.monografias.com/Tecnologia/index.shtml) de bases de datos distribuidas son en lo referente a duplicidad de datos y a su integridad al momento de realizar actualizaciones a los mismos.

TENDENCIAS FUTURAS. En el futuro la mayoría de las [organizaciones](http://www.monografias.com/trabajos6/napro/napro.shtml) cambiarán la forma convencional de manejo de la información a la arquitectura de base de datos a las ventajas [derivadas](http://www.monografias.com/trabajos6/esfu/esfu.shtml#tabla) de su uso. El uso de las bases de datos distribuidas se incrementará de manera considerable en la medida en que la tecnología de [comunicación](http://www.monografias.com/trabajos12/fundteo/fundteo.shtml) de datos brinde más facilidades para ello. El uso de bases de datos facilitará y soportará en gran medida a los Sistemas de Información para la Toma de Decisiones.

2. DBMS

Definir Que Es Un "DBMS"  
El DBMS: es un conjunto de programas que se encargan de manejar la creación y todos los accesos a las bases de datos, está compuesto por:  
DDL: [Lenguaje](http://www.monografias.com/trabajos35/concepto-de-lenguaje/concepto-de-lenguaje.shtml) de Definición de Datos  
DML: Lenguaje de Manipulación de Datos  
[SQL](http://www.monografias.com/trabajos14/sqlserver/sqlserver.shtml): Lenguaje de Consulta.  
SQL (Structured Query Lenguaje).  
Los DBMS más comunes son [Oracle](http://www.monografias.com/trabajos14/base-datos/base-datos.shtml), SqlServer, Informix, Sysbase,

1. Oracle

2. SQL Server

3 Sysbase

4MySql

Describir las [funciones](http://www.monografias.com/trabajos7/mafu/mafu.shtml) de cada componente de un sistema de base de datos  
1. Los datos.  
Es lo que se conoce como base de datos propiamente dicha. Para manejar estos datos utilizamos una serie de programas.  
2.- Los Programas.  
Son los encargados de manejar los datos, son conocidos como DBMS (Data Base Management System) o también SGBD (Sistema Gestor de Base de Datos). Los DBMS tienen dos funciones principales que son:  
- La definición de las estructuras para almacenar los datos.  
- La manipulación de los datos.  
1.- Describir los sistemas de procesamiento de archivos  
La información necesita ser guardada y manipulada para que sea útil y este es uno de los [objetivos](http://www.monografias.com/trabajos16/objetivos-educacion/objetivos-educacion.shtml) principales que han tenido la [computadoras](http://www.monografias.com/trabajos15/computadoras/computadoras.shtml) dentro de la [empresa](http://www.monografias.com/trabajos11/empre/empre.shtml)

. 2.- Explicar su [evolución](http://www.monografias.com/trabajos16/teoria-sintetica-darwin/teoria-sintetica-darwin.shtml) histórica  
numerosas distribuciones es el PostgreSQL.

Sistema [UNIX](http://www.monografias.com/trabajos36/sistema-unix/sistema-unix.shtml), fue portado a Linux y modificado para utilizar el lenguaje de consultas SQL, donde se llamó Posgres95.  
Mejorado en el 96 tomo su nuevo PostgeSQL y adopto el nuevo carácter del SQL - 92.  
desde internet a través de una página [Web](http://www.monografias.com/trabajos5/laweb/laweb.shtml).  
PostgreSQL requiere de cierto [conocimiento](http://www.monografias.com/trabajos/epistemologia2/epistemologia2.shtml) del lenguaje SQL por lo que es

3.- Describir los sistemas de procesamiento en línea y por lotes  
  
4.- Explicar los lenguajes de 3era y 4ª generación  
Los lenguajes de tercera generación (3GL; third-generation lenguajes) son los lenguajes propiamente como los conocemos (poseen instrucciones, funciones, sintaxis, [semántica](http://www.monografias.com/trabajos29/semantica-conectores-aplicaciones-obras-literarias/semantica-conectores-aplicaciones-obras-literarias.shtml)); ya trascienden el uso de los términos nemotécnicos. Una instrucción puede indicar 1 o más tareas para el [computador](http://www.monografias.com/trabajos15/computadoras/computadoras.shtml). Pueden dividirse como sigue:  
Orientados a problemas  
A procedimientos  
A objetos  
4GL (Generadores de aplicaciones que no dependen de una [metodología](http://www.monografias.com/trabajos11/metods/metods.shtml)).

* Son de alto nivel y amigables; los programadores profesionales que utilizan 4GL sostienen que experimentan incrementos en la [productividad](http://www.monografias.com/trabajos6/prod/prod.shtml) de 200 a 1,000% en comparación con los lenguajes orientados al [procedimiento](http://www.monografias.com/trabajos13/mapro/mapro.shtml) de la tercera generación ([COBOL](http://www.monografias.com/trabajos/marcohistocomp/marcohistocomp.shtml), FORTRAN, BASIC y demás ). Hay dos tipos de 4GL:

Orientados a la producción  
Al usuario  
5.- Ejemplificar las aplicaciones de los 3GL y los 4GL

**Datamining**

Datamining

**Descubriendo** [**Información**](http://www.monografias.com/trabajos7/sisinf/sisinf.shtml) **Oculta**

[Data Mining](http://www.monografias.com/trabajos54/data-mining/data-mining.shtml), **la extracción de información oculta y predecible de grandes bases de datos**,

Es una poderosa [tecnología](http://www.monografias.com/Tecnologia/index.shtml) nueva con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información ([Data Warehouse](http://www.monografias.com/trabajos6/dawa/dawa.shtml)).

Las [herramientas](http://www.monografias.com/trabajos11/contrest/contrest.shtml) de Data Mining

Estas herramientas exploran las [bases de datos](http://www.monografias.com/trabajos11/basda/basda.shtml) en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas.

Las [técnicas](http://www.monografias.com/trabajos6/juti/juti.shtml) de Data Mining pueden ser implementadas rápidamente en plataformas ya existentes de [software](http://www.monografias.com/Computacion/Software/) y [hardware](http://www.monografias.com/Computacion/Hardware/) para acrecentar el [valor](http://www.monografias.com/trabajos14/nuevmicro/nuevmicro.shtml) de las [fuentes](http://www.monografias.com/trabajos10/formulac/formulac.shtml#FUNC) de información existentes y pueden ser integradas con nuevos [productos](http://www.monografias.com/trabajos12/elproduc/elproduc.shtml) y sistemas pues son traídas en línea (on-line).

pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como, "¿Cuáles [clientes](http://www.monografias.com/trabajos11/sercli/sercli.shtml) tienen más [probabilidad](http://www.monografias.com/trabajos54/resumen-estadistica/resumen-estadistica.shtml) de responder al próximo mailing promocional, y por qué? y presentar los resultados en formas de tablas, con [gráficos](http://www.monografias.com/trabajos11/estadi/estadi.shtml#METODOS), reportes, [texto](http://www.monografias.com/trabajos13/libapren/libapren.shtml), hipertexto, etc.

**Los Fundamentos del Data Mining**

Las técnicas de Data Mining son el resultado de un largo [proceso](http://www.monografias.com/trabajos14/administ-procesos/administ-procesos.shtml#PROCE) de [investigación](http://www.monografias.com/trabajos11/norma/norma.shtml) y [desarrollo](http://www.monografias.com/trabajos12/desorgan/desorgan.shtml) de productos.

Esta [evolución](http://www.monografias.com/trabajos16/teoria-sintetica-darwin/teoria-sintetica-darwin.shtml) comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para

Permitir a los usuarios navegar a través de los datos en tiempo real.

Data Mining está listo para su aplicación en la [comunidad](http://www.monografias.com/trabajos910/comunidades-de-hombres/comunidades-de-hombres.shtml) de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras:

* Recolección masiva de datos
* Potentes computadoras con multiprocesadores
* Algoritmos de Data Mining

[Algoritmos](http://www.monografias.com/trabajos15/algoritmos/algoritmos.shtml) de Data Mining utilizan técnicas que han existido por lo menos desde hace 10 años, pero que sólo han sido implementadas recientemente como herramientas

**El Alcance de Data Mining**

El nombre de Data Mining deriva de las similitudes entre buscar valiosa información de negocios en grandes bases de datos - por ej.: encontrar información de la [venta](http://www.monografias.com/trabajos12/curclin/curclin.shtml) de un producto entre grandes montos de Gigabytes almacenados

Las bases de datos pueden ser grandes tanto en profundidad como en ancho:

* **Más columnas**. Los analistas muchas veces deben limitar el número de [variables](http://www.monografias.com/trabajos12/guiainf/guiainf.shtml#HIPOTES) a examinar cuando realizan análisis [manuales](http://www.monografias.com/trabajos6/maca/maca.shtml) debido a limitaciones de tiempo. Sin embargo, variables que son descartadas porque parecen sin importancia pueden proveer información acerca de modelos desconocidos. Un Data Mining de alto rendimiento permite a los usuarios explorar toda la base de datos, sin preseleccionar un subconjunto de variables.
* **Más filas.** Muestras mayores producen menos errores de estimación y desvíos, y permite a los usuarios hacer inferencias acerca de pequeños pero importantes segmentos de población.

Las técnicas más comúnmente usadas en Data Mining son:

* **Redes neuronales artificiales:** modelos predecible no-lineales que aprenden a través del [entrenamiento](http://www.monografias.com/trabajos14/mocom/mocom.shtml) y semejan la [estructura](http://www.monografias.com/trabajos15/todorov/todorov.shtml#INTRO) de [una red](http://www.monografias.com/Computacion/Redes/) neuronal biológica.
* **Arboles de decisión:** [estructuras](http://www.monografias.com/trabajos15/todorov/todorov.shtml#INTRO) de forma de árbol que representan [conjuntos](http://www.monografias.com/trabajos10/historix/historix.shtml) de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de [árboles](http://www.monografias.com/trabajos15/composicion-follaje/composicion-follaje.shtml) de decisión incluyen [Arboles](http://www.monografias.com/trabajos15/composicion-follaje/composicion-follaje.shtml) de Clasificación y Regresión (CART: Classification And Regression Tree) y Detección de [Interacción](http://www.monografias.com/trabajos901/interaccion-comunicacion-exploracion-teorica-conceptual/interaccion-comunicacion-exploracion-teorica-conceptual.shtml) Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection)
* **Algoritmos genéticos:** técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y [selección](http://www.monografias.com/trabajos5/selpe/selpe.shtml) natural en un [diseño](http://www.monografias.com/trabajos13/diseprod/diseprod.shtml) basado en los conceptos de evolución.
* **Método del vecino más cercano:** una técnica que clasifica cada [registro](http://www.monografias.com/trabajos7/regi/regi.shtml) en un conjunto de datos basado en una combinación de las clases del/de los k registro (s) más similar/es a él en un conjunto de datos históricos (donde k ≥ 1). Algunas veces se llama la técnica del vecino k-más cercano.
* **Regla de** [**inducción**](http://www.monografias.com/trabajos6/elme/elme.shtml#induccion)**:** la extracción de reglas if-then de datos basados en significado estadístico.

**¿Cómo Trabaja el Data Mining?**

En Data Mining se llama Modelado.

Modelado es simplemente el acto de construir un [modelo](http://www.monografias.com/trabajos/adolmodin/adolmodin.shtml) en una situación donde usted conoce la respuesta y luego la aplica en otra situación de la cual desconoce la respuesta.

Por ejemplo, si busca un galeón [español](http://www.monografias.com/trabajos5/oriespa/oriespa.shtml) hundido en los mares lo primero que podría hacer es investigar otros tesoros españoles que ya fueron encontrados en el pasado.

**Una** [**arquitectura**](http://www.monografias.com/trabajos6/arma/arma.shtml) **para Data Mining**

Varias herramientas de Data Mining actualmente operan fuera del warehouse, requiriendo pasos extra para extraer, importar y analizar los datos. Además, cuando. El Data warehouse analítico resultante puede ser aplicado para mejorar procesos de negocios en toda la [organización](http://www.monografias.com/trabajos6/napro/napro.shtml), en áreas tales como manejo de campañas promocionales, detección de fraudes, lanzamiento de nuevos productos, etc.

**Glosario de Términos de Data Mining**

* **Algoritmos genéticos:** Técnicas de optimización que usan procesos tales como combinación [genética](http://www.monografias.com/trabajos/genetica/genetica.shtml), mutación y selección natural en un diseño basado en los conceptos de evolución natural.
* **Análisis de** [**series de tiempo**](http://www.monografias.com/trabajos31/series-tiempo-internet/series-tiempo-internet.shtml) **(time-series):** Análisis de una secuencia de medidas hechas a intervalos específicos. El tiempo es usualmente la dimensión dominanate de los datos.
* **Análisis prospectivo de datos:** Análisis de datos que predice futuras tendencias, comportamientos o eventos basado en datos históticos.
* **Análisis exploratorio de datos:** Uso de técnicas estadísticas tanto [gráficas](http://www.monografias.com/trabajos11/estadi/estadi.shtml#METODOS) como descriptivas para aprender acerca de la estructura de un conjunto de datos.
* **Análisis retrospectivo de datos:** Análisis de datos que provee una visión de las tendencias , comportamientos o eventos basado en datos históricos.
* **Árbol de decisión:** Estructura en forma de árbol que representa un conjunto de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Ver CART y CHAID.
* **Base de datos multidimensional:** Base de datos diseñada para procesamiento analítico on-line (OLAP). Estructurada como un hipercubo con un eje por dimensión.
* **CART Árboles de clasificación y regresión**: Una técnica de árbol de decisión usada para la clasificación de un conjunto da datos. Provee un conjunto de reglas que se pueden aplicar a un nuevo (sin clasificar) conjunto de datos para predecir cuáles [registros](http://www.monografias.com/trabajos7/regi/regi.shtml) darán un cierto resultado. Segmenta un conjunto de datos creando 2 divisiones. Requiere menos preparación de datos que CHAID .
* **CHAID Detección de interacción automática de Chi cuadrado**: Una técnica de árbol de decisión usada para la clasificación de un conjunto da datos. Provee un conjunto de reglas que se pueden aplicar a un nuevo (sin clasificar) conjunto de datos para predecir cuáles registros darán un cierto resultado. Segmenta un conjunto de datos utilizando tests de chi cuadrado para crear múltiples divisiones. Antecede, y requiere más preparación de datos, que CART.
* **Clasificación**: Proceso de dividir un conjunto de datos en [grupos](http://www.monografias.com/trabajos11/grupo/grupo.shtml) mutuamente excluyentes de tal manera que cada miembro de un [grupo](http://www.monografias.com/trabajos14/dinamica-grupos/dinamica-grupos.shtml) esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a variable(s) específica(s) las cuales se están tratando de predecir. Por ejemplo, un problema típico de clasificación es el de dividir una base de datos de compañías en grupos que son lo más homogéneos posibles con respecto a variables como "posibilidades de [crédito](http://www.monografias.com/trabajos15/financiamiento/financiamiento.shtml)" con [valores](http://www.monografias.com/trabajos14/nuevmicro/nuevmicro.shtml) tales como "Bueno" y "Malo".
* **Clustering (agrupamiento)**: Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a todas las variables disponibles.
* **Computadoras con multiprocesadores**: Una [computadora](http://www.monografias.com/trabajos15/computadoras/computadoras.shtml) que incluye múltiples [procesadores](http://www.monografias.com/trabajos5/sisope/sisope.shtml) conectados por una [red](http://www.monografias.com/Computacion/Redes/). Ver procesamiento paralelo.
* **Data cleansing:** Proceso de asegurar que todos los valores en un conjunto de datos sean consistentes y correctamente registrados.
* **Data Mining**: La extracción de información predecible escondida en grandes bases de datos.
* **Data Warehouse**: [Sistema](http://www.monografias.com/trabajos11/teosis/teosis.shtml) para el [almacenamiento](http://www.monografias.com/trabajos12/dispalm/dispalm.shtml) y [distribución](http://www.monografias.com/trabajos11/travent/travent.shtml) de cantidades masivas de datos
* **Datos anormales**: Datos que resultan de errores (por ej.: errores en el tipeado durante la carga) o que representan eventos inusuales.
* **Dimensión**: En una base de datos relacional o plana, cada campo en un registro representa una dimensión. En una base de datos multidimensional, una dimensión es un conjunto de entidades similares; por ej.: una base de datos multidimensional de ventas podría incluir las dimensiones Producto, Tiempo y Ciudad.
* **Modelo analítico**: Una estructura y proceso para analizar un conjunto de datos. Por ejemplo, un árbol de decisión es un modelo para la clasificación de un conjunto de datos
* **Modelo lineal:** Un modelo analítico que asume relaciones lineales entre una variable seleccionada (dependiente) y sus predictores (variables independientes).
* **Modelo no lineal**: Un modelo analítico que no asume una relación lineal en los coeficientes de las variables que son estudiadas.
* **Modelo predictivo:** Estructura y proceso para predecir valores de variables especificadas en un conjunto de datos.
* **Navegación de datos:** Proceso de visualizar diferentes dimensiones, "fetas" y niveles de una base de datos multidimensional. Ver OLAP.
* **OLAP Procesamiento analítico on-line (On Line Analitic prossesing):** Se refiere a aplicaciones de bases de datos orientadas a array que permite a los usuarios ver, navegar, manipular y analizar bases de datos multidimensionales.
* **Outlier:** Un item de datos cuyo valor cae fuera de los [límites](http://www.monografias.com/trabajos6/lide/lide.shtml) que encierran a la mayoría del resto de los valores correspondientes de la [muestra](http://www.monografias.com/trabajos11/tebas/tebas.shtml). Puede indicar datos anormales. Deberían ser examinados detenidamente; pueden dar importante información.
* **Procesamiento paralelo**: Uso coordinado de múltiples procesadores para realizar tareas computacionales. El procesamiento paralelo puede ocurrir en una computadora con múltiples procesadores o en una red de estaciones de [trabajo](http://www.monografias.com/trabajos34/el-trabajo/el-trabajo.shtml) o PCs.
* **RAID**: Formación redundante de discos baratos (Redundant Array of inexpensive disks). Tecnología para el almacenamiento paralelo eficiente de datos en sistemas de computadoras de alto rendimiento.
* **Regresión lineal:** Técnica [estadística](http://www.monografias.com/trabajos15/estadistica/estadistica.shtml) utilizada para encontrar la mejor relación lineal que encaja entre una variable seleccionada (dependiente) y sus predicados (variables independientes).
* **Regresión** [**logística**](http://www.monografias.com/trabajos15/logistica/logistica.shtml)**:** Una regresión lineal que predice las proporciones de una variable seleccionada categórica, tal como Tipo de [Consumidor](http://www.monografias.com/trabajos5/comco/comco.shtml#aspe), en una población.
* **Vecino más cercano:** Técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases del/de los k registro (s) más similar/es a él en un conjunto de datos históricos (donde k ≥ 1). Algunas veces se llama la técnica del vecino k-más cercano.
* **SMP Multiprocesador simétrico (Symmetric multiprocessor):** Tipo de computadora con multiprocesadores en la cual [la memoria](http://www.monografias.com/trabajos16/memorias/memorias.shtml) es compartida entre los procesadores