



Marchese Alicia  
Picco Alicia  
Pluss, Jorge J.  
Díaz Daniel  
Gaibazzi Ma. Florencia  
Repetto Luciano  
Ingrassia Ramiro

*Instituto de Investigaciones y Asistencia Tecnológica en Administración*

## **NUEVAS TECNOLOGÍAS PARA LA ADMINISTRACIÓN DE EMPRESAS**

### **1. Introducción - Tecnologías actuales y sus posibilidades**

Entre las nuevas pautas de producción de bienes, tangibles e intangibles, aparece últimamente una tendencia que se generaliza: el aprovechamiento de elementos que no hacen al producto principal, pero que son consecuencia de dicho proceso. Los subproductos derivados que antes se desechaban hoy son utilizados y reciclados para aportar al desarrollo de otros bienes.

Al igual que en la industria manufacturera, en la producción de información quedan elementos que hoy pueden ser reutilizados. La información relativa a transacciones que cumple con el objetivo de favorecer la toma de decisiones cotidianas, deja un enorme capital remanente cuya utilización racional va a modificar el concepto de toma de decisiones tal y como se la conoce hasta ahora.

Los datos se convierten en información mediante procesos sucesivos de clasificación, ordenamiento, agrupamiento, etc. La consecuencia de dichos procesos ya ha sido almacenada, ha llegado la etapa de reciclarla y convertirla en conocimiento útil para la estrategia del negocio. Las herramientas para su utilización existen hoy, y adoptan diferentes denominaciones comerciales. Es nuestro trabajo la recopilación de las mismas, su diferenciación y la búsqueda de los conceptos subyacentes en cada una.

Es sabido que en el ámbito de las tecnologías informáticas cada empresa lanza productos al mercado que cumplen objetivos muchas veces similares y su éxito (la captación de clientes/usuarios) se produce más por la difusión de los mismos y sus denominaciones comerciales que por el conocimiento que se tenga acerca de su utilidad final. No obstante, la aparición de determinados productos y su correlación en la bibliografía científica no deja dudas acerca de la etapa tecnológica que estamos transitando: La necesidad básica que satisfacen los productos que analizamos es la de brindar información con un alto grado de abstracción. Esto es, se trata de información dirigida a la toma de decisiones poco o muy poco estructuradas, con un alto índice de riesgo.

Las tecnologías de información han logrado integrar las distintas actividades de una empresa en un almacén de datos, lo que trajo como consecuencia la interoperatividad entre sectores de la misma. El viejo organigrama se nutre hoy con datos de los diferentes subsistemas que permiten la estructuración de funciones por proyectos, verdaderos desafíos administrativos que involucran áreas antes dispersas, como producción y marketing.

Pensemos en una gran empresa en red, que posee subsidiarias en distintos lugares del planeta, cada una de las cuales funciona como una unidad, que genera los datos necesarios para la toma de decisiones en su ámbito específico. Existe a su vez un estamento directivo dentro de la empresa que requiere de información compactada para la toma de decisiones estratégicas.

En una primer etapa la información que llegaba desde los niveles operativos a los niveles de decisión estratégicos era información "por excepción"; aquella que reflejaba los desvíos entre lo esperado y lo real. En la actualidad el concepto es que, con los datos integrados, se logre una visión totalizadora de los mismos, reuniéndolos en repositorios de información recolectada desde múltiples orígenes y guardados en un esquema homogéneo. Esto dinamiza enormemente la toma de decisiones estratégicas y permite redefinir objetivos y pautas de gestión en forma continua, sin afectar el normal desenvolvimiento de las operaciones cotidianas.



Una estructura de este tipo no viene dada, no se compra. El software provee el soporte para su diseño, pero es el administrador de negocios el que debe determinar qué datos se tomarán y con qué configuración y nivel de agrupamiento. Es tarea del gerente, involucrado en la toma de decisiones estratégicas, la definición del conjunto de datos y sus relaciones porque es él quien conoce sus necesidades y requerimientos administrativos.

Un mal diseño de datos torna inútil la utilización de estas herramientas. Debemos asumir el compromiso de conocer y reconocer las necesidades de la empresa y aprender a plasmarlos en un modelo que ayude al crecimiento cualitativo de nuestros gerentes.

En el presente trabajo profundizamos tres aspectos de vinculados con la utilización inteligente de la información de la empresa. Las redes neuronales, que intentan emular los patrones de procesamiento del cerebro biológico<sup>1</sup>, Data Warehousing, técnica que intenta unificar, tratar y suministrar la información necesaria para toma de decisiones racionales y Data Mining que utiliza las técnicas de inteligencia artificial para extraer de los datos tendencias o información que pasa desapercibida para el usuario final.

Distintas etapas en la administración han sido influenciadas por la evolución de tecnologías de la información disponibles. En la década del 80 el surgimiento de las PC y las interfaces gráficas marcaron un rumbo basado en la disponibilidad de herramientas cada vez más veloces, intuitivas y de menor costo.

La década del 90 fue caracterizada por un concepto que marcó un nuevo rumbo en la gestión: la "conectividad". El auge explosivo de Internet provocó que en la actualidad no se pueda concebir una empresa de mediano tamaño que no tenga una presencia en la red. Desde simples páginas con propaganda institucional o de productos, hasta portales de gestión dinámica con soluciones de e-commerce, se encuentran en constante crecimiento y renovación. La tendencia que marcan las nuevas tecnologías, en este siglo, es la de una profundización de ese concepto de conectividad.

Actualmente se piensa en términos de programación "totalmente distribuida", vale decir, programas que se buscan, se comunican e intercambian procesos a través de la red (web services). También se están comenzando a modelar sistemas pensados para suministrar información "totalmente personalizada", que permiten al usuario seleccionar la información que desea consultar, accediendo a la misma desde diversos medios tales como teléfonos celulares, dispositivos Palm, o su PC.

Para lograr todo esto los distintos operadores han tomado conciencia de la necesidad de utilizar estándares de codificación y comunicación de datos. A esto responde el auge que está adoptando en la comunidad de Internet la utilización de XML (extensible markup language) y sus lenguajes de codificación derivados.

La necesidad de compartir información y procesos, que establece la competitividad y globalización actual crea una imperiosa necesidad de actualización de la política del manejo de la información que genera o recibe la empresa. La utilización de técnicas inteligentes y la consolidación de la información son temas prioritarios para las Pymes actuales que intenten participar y protagonizar el cambio.

Sabemos que estos productos son hoy costosos en el mercado, inaccesibles para una PyME local, pero también proponemos formas alternativas para su utilización, en la creencia de que (como tantas veces ocurrió), **estas herramientas se generalizarán en el futuro y harán su aporte a la evolución administrativa, y los empresarios que sobrevivan serán aquellos que logren expandir sus potencialidades utilizándolas.**

El empresario PyME no puede hoy tener en su empresa todos los elementos necesarios para optimizar la gestión, pero puede aliarse con otros que tengan similares necesidades y compartir recursos estratégicos que le permitan plantearse objetivos de largo plazo individuales y compartidos para afrontar la nueva realidad global en la que tarde o temprano deberá insertarse.

## 2. Inteligencia Artificial: Redes neuronales

---

<sup>1</sup> LAUDON, Kenneth, LAUDON, Jane, "Administración de los Sistemas de Información: organización y tecnologías", Prentice Hall Hispanoamericana editora, México, 1996, pág.670



Hacia finales de la década de los 80 se produjo un resurgimiento de la Inteligencia Artificial, dejada de lado luego de los primeros intentos de los años 50, especialmente para la corriente Conexionista. Esta nueva etapa que comenzaba iba a modificar más aún el ámbito de las Tecnologías en Informática. El suceso fue el trabajo de David Rumelhart, científico de la Universidad de Stanford, sobre la "Solución del problema del Perceptrón Multicapa".

Esto se dio, gracias dos hechos que se sucedieron en la misma década. El primero, es la entrada en la cuarta generación del hardware de computadora, cuyos avances más significativos fueron :

- ✓ Ampliación del tamaño de la memoria,
- ✓ Aumento en la velocidad de procesamiento (se llega a 200 millones de instrucciones por segundo –MIPS-),y
- ✓ Disminución del costo de las computadoras junto con su tamaño.

El segundo hecho ocurrido más precisamente en el año 1982, fue la exposición del trabajo sobre "Sistemas Artificiales" de John Hopfield, investigador del departamento de Biología Molecular de la Universidad de Princeton. Estos hechos marcan el final de una etapa muy teórica de esta corriente, de la cual ya no se esperaban grandes resultados y su paso a la visión científica actual, con aplicaciones prácticas antes impensadas.

### **2.1. La Corriente Conexionista**

La corriente conexionista, junto al simbolismo , son las ramas de la Investigación en el campo de la Inteligencia Artificial (IA) Laudon y Laudon la definen como: "el esfuerzo por desarrollar sistemas computacionales que puedan comportarse como los seres humanos, con la capacidad de aprender idiomas, llevar a cabo tareas físicas, usar aparatos de percepción y emular el conocimiento humano y la toma de decisiones"<sup>2</sup>

El conexionismo o enfoque Ascendente (Bottom Up), se propone simular artificialmente la estructura física del cerebro humano, para llegar a la Inteligencia. La idea central es que si el cerebro es el responsable de nuestro pensamiento, acciones y nuestro sentido común, al imitar su estructura, se tendría que llegar al mismo resultado: la Inteligencia.

Sin entrar en discusiones filosóficas, debemos tener en cuenta que "*no existen ordenadores capaces de soportar diez billones de unidades de procesamiento en un mismo programa, ni seguramente alguien que se proponga interconectarlo...*"<sup>3</sup>

La inquietud de reproducir el funcionamiento del cerebro humano puede remontarse a David Hume quien escribió un tratado sobre la naturaleza humana, pasando por Luigi Galvani que expone sobre los efectos de la electricidad en la actividad muscular y afirmándose con Ramón y Cajal, quien presenta a la neurona como la unidad mínima cerebral.

En verdad, en 1943 se realizó un modelo lógico de cómo trabaja el cerebro. Fue propuesto por Mc Culloch y Pitts, y se trataba de un modelo matemático computacional del cerebro constituido por pequeños procesadores denominados neuronas artificiales. Luego fue modificado y mejorado, en 1958 por Frank Rosenblatt, profesor de Física y Ciencia Macromolecular del MIT. Aquí nace el llamado Perceptrón. Gracias a importantes investigaciones realizadas anteriormente, como la dirigida por John Mc Cartthy y Marvin Minsky entre otros en la Dartmouth College, donde se acuñó el término Inteligencia Artificial, acompañado por los avances de la Fisiología en el estudio del sistema nervioso superior y la tecnología que se potenciaba para más complejos emprendimientos.

Pero este avance se frena a fines de los sesenta, cuando Marvin Minsky y Seymour Papert, critican el modelo lógico del Perceptrón, un concepto que no se podía implementar con el hardware existente en la época, imprescindible para continuar con las investigaciones (por ejemplo, el tamaño de la memoria). Se produce un silencio en las investigaciones conexionistas, aunque algunos siguieron con sus búsquedas.

<sup>2</sup> LAUDON, Kenneth, LAUDON, Jane, "Administración de los Sistemas de Información: organización y tecnologías", Ed. Prentice Hall Hispanoamerica editora, 3er. Edición, 1.996, México

<sup>3</sup> Iovaldi, R, "Inteligencia Artificial y Aprendizaje: investigación y metodología"



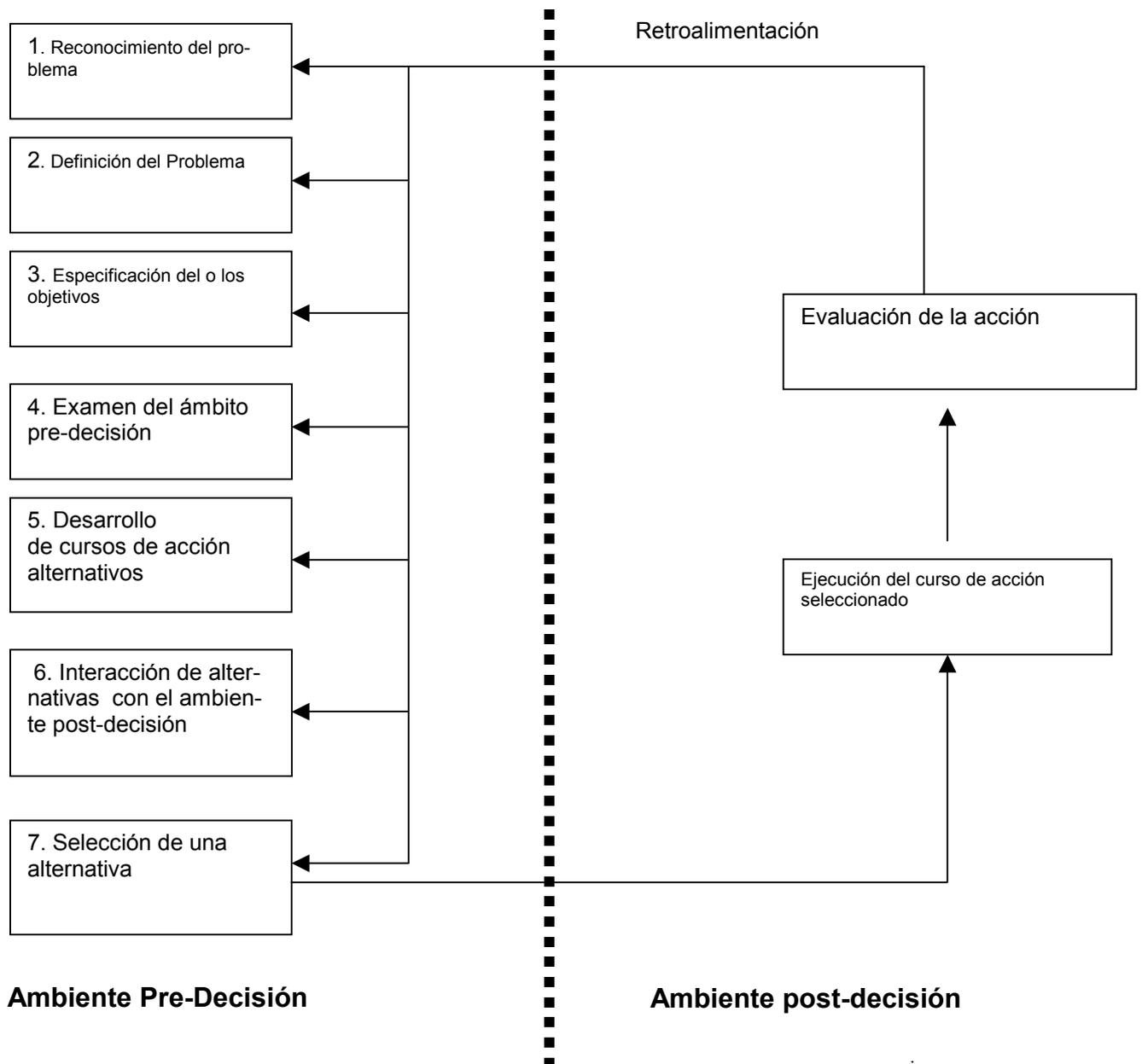
## 2.2 Conexionismo y Sistemas Expertos

"La tarea del ejecutivo implica no sólo tomar decisiones él mismo, sino procurar que la organización o sector que dirige lo haga de una manera eficaz"

Simon, Herbert

Mientras tanto, otros estudios abordan la toma de decisiones analizando su proceso, características (estructuradas o no estructuradas) y componentes. Pero la Inteligencia Artificial (IA) aporta una herramienta concreta en esta etapa: los sistemas expertos, que pueden seleccionar una alternativa gracias al recorrido de un camino de razonamiento que tienen incorporado (reglas).

### Esquema de un Proceso de Decisión (fig.1)



Se señalan distintas limitaciones de los sistemas expertos, que sólo se deben a la falta de sentido común ó heurística y la cantidad de datos de ingreso ó alternativas que puede soportar. Es decir, las decisiones tienen un sentido fáctico y un sentido ético, según Herbert Simon, siendo este último "una cualidad imperativa: seleccionan un estado futuro de cosas con preferencias a otro y dirigen el comportamiento hacia la alternativa elegida"<sup>4</sup>.

Es por ello, que en un proceso de decisión, tal como se lo ve en la fig.1, ocurren distintos hechos cuya modelización es sumamente dificultosa. Quienes deseen plasmar en un sistema experto el proceso decisorio deberán en un primer momento enumerar las alternativas posibles y luego definir cuántas de ellas podremos ingresar en una forma "si  $a_1, a_2, a_3, a_4, \dots, a_n$  entonces  $b_1, b_2, b_3, b_4, \dots, b_n$ " y dentro de cada regla, establecer la ponderación de cada alternativa, que al ser contrastada con la realidad, cambiará constantemente porque la realidad es un cambio continuo.

Volvemos a lo mismo, la experiencia de los Gerentes, su capacitación constante, son las que adecuan sus decisiones a los distintos cambios de la realidad, simplemente porque a medida que cambia la realidad, cambia nuestro sentido ético o ponderación de alternativas, es decir, nuestra heurística. Si bien los sistemas expertos elucidan el aspecto ético de la decisión, el desafío de la Inteligencia Artificial, consiste hoy en apostar al nuevo conexionismo, a través de su producto mejorado: Las Redes Neuronales.

A pesar de todo, la situación actual es muy distinta a la de los finales de la década de los sesenta, por ello no se puede pensar que los sistemas expertos tendrán la misma suerte que las redes neuronales.

### 2.3 Las redes Neuronales

Las redes neuronales son un conjunto de neuronas artificiales, las cuales son pequeñas unidades de procesamiento de información. En conjunto, emulan la corteza cerebral en su estructura y funcionamiento.

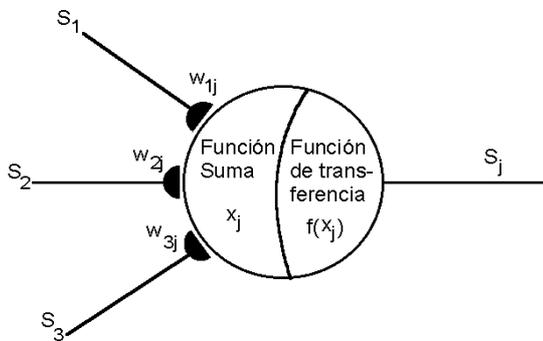
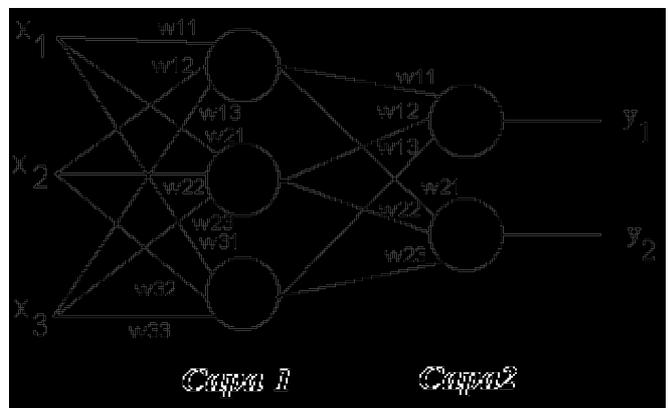


Fig.2: Neurona Artificial

5



– Fig.3: Red Neuronal Típica

<sup>4</sup> SIMON, Herbert, "El comportamiento Administrativo", Ed. Aguilar, Buenos Aires, 1984, pág.45

<sup>5</sup> Figuras extraídas del trabajo de Redes Neuronales, una introducción; de Xavier Padern, <http://www.eupmt.es/cra/inform/info11.htm>



Las partes físicas de una neurona artificial (fig.2) son:

1. Datos de entrada y pesos relativos de los mismos
2. Una función de base: la suma de los productos de los datos con su "peso" o ponderación,
3. Una función de activación: función matemática que se compara con la función de base, y
4. una salida

La organización de las redes, toma en general la forma de una "matriz", que en lugar de números, cuenta con neuronas que pueden combinarse, según:

- ✓ El número de neuronas que asignemos,
- ✓ La cantidad de columnas: hay tres tipos de columnas; la columna de entrada, son las neuronas que reciben la información; las columnas ocultas, son la intermedias y son las que pueden variar en número; y la columna de salida, son las neuronas que se corresponden con la información de entrada
- ✓ La cantidad de entradas que tendrá
- ✓ El número de respuestas que fijemos
- ✓ Cuántas conexiones tendrán cada una de las neuronas y la ponderación que reciben
- ✓ Qué funciones de activación tendrá cada neurona, (no es necesario que tengan todas las mismas funciones)

Imitando a las neuronas biológicas, las redes neuronales (fig.3), funcionan de acuerdo al siguiente esquema: Cuando una neurona artificial recibe los datos (el equivalente a las sustancias neurotransmisoras biológicas), se construye la función de base con ellos, es decir, se incorporan a la ecuación que contiene la suma de los productos de los datos por el peso relativo (la sustancia es captada por las dendritas)

A continuación se los compara con la función de activación (equivalente a la carga eléctrica que se aloja en el núcleo de la neurona), y si la función de base fue mayor que la de activación habrá una salida (nueva carga eléctrica) que se retransmitirá a la siguiente neurona (sinapsis).

Hay tres formas de aprendizaje

➤ En el **Aprendizaje Supervisado** se presentan a la neurona unas entradas con las correspondientes salidas deseadas. La red primeramente calcula una salida para esas entradas y luego, conociendo la diferencia entre la salida deseada y la encontrada, ajusta sus pesos proporcionalmente al error que ha cometido (si la diferencia entre salida calculada y salida deseada es nula, no se varían los pesos).

➤ En el **Aprendizaje No Supervisado**, solo se presentan a la neurona las entradas y dará una salida, la cual por camino inverso, tendrá que llegar a la misma entrada.

➤ En el **Aprendizaje Por Refuerzo** se combinan los dos anteriores, y de cuando en cuando se presenta a la red una valoración global de cómo lo está haciendo.

Según el mecanismo que utilizan para aprender, actualmente existen tres tipos de redes neuronales:

- ◆ Perceptrón Multicapa con Retropropagación:  
Desarrollado por Rumelhart, Hinton y Willoes del grupo PDP (Parallel Distributed Processing) de Cambridge Massachusset,
- ◆ Memoria Asociativa de Hopfield: (fig.4)  
Expuesto por John Hopfield, investigador del Instituto de Tecnología de California
- ◆ Red Kohonen: (fig.5)  
Desarrollada por Teuvo Kohonen, del Centro de Investigación en Redes Neuronales de Helsinki, Universidad de Finlandia.

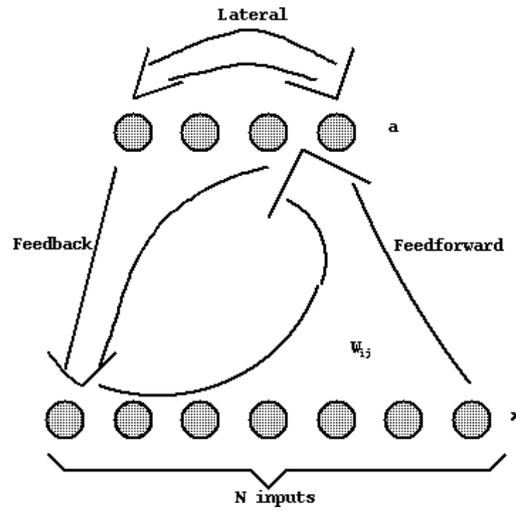


Fig.4: Red de memoria Asociativa con pesos fijos: John Hopfield<sup>6</sup>

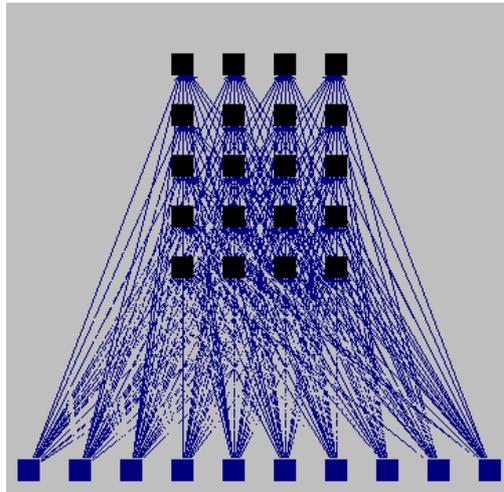
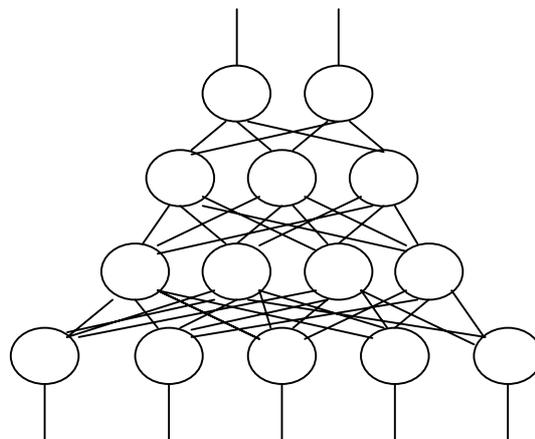


Fig.5: Red Kohonen: Teuvo Kohonen<sup>7</sup>



<sup>6</sup> Figura extraída del Trabajo Redes Neuronales, en la pagina web:  
<http://www.gc.ssr.upm.es/inves/neural/ann2/anntutor.htm>

<sup>7</sup> Figura extraída del Trabajo Redes Neuronales, en la pagina web:  
<http://www.gc.ssr.upm.es/inves/neural/ann2/anntutor.htm>



Fig.6: Red Perceptrón Multicapa con retropropagación: Rumelhart, Hinton y Willoes

## **2.4 Sentido Etico/Heurístico**

Luego de la breve descripción anterior de los elementos básicos de las redes neuronales, tratando de no caer en el riguroso modelo matemático de la red neuronal, retomemos el ámbito de las decisiones, para no perder el objetivo de nuestro trabajo, el cual es conocer estas herramientas para promover su uso y aplicación en el ámbito empresarial.

El aspecto lógico de las redes neuronales, es tratar de alcanzar (de imitar) el sentido común. Cada alternativa que se le presenta a la red neuronal, será ponderada de acuerdo a la visión subjetiva de la realidad. ¿Cómo sabremos si esa ponderación es exacta? Debemos enseñarle a la red la meta deseada. Así, en el inicio podemos ponderar arbitrariamente y en el proceso de aprendizaje, irá recalculando las nuevas ponderaciones, hasta que encuentre las que son correctas. Esto significa que aprendió y está en condiciones de tomar una decisión.

Recordemos que las ponderaciones, además, deben pasar por un criterio de relación (la función de activación), y según este, se dará una respuesta. No sólo se pondera, se establece un cierto "hilo de conducción" para no caer en la "insensatez". Por supuesto que el concepto de decisión "sensata" implica una disquisición que no podemos abordar en este trabajo. Basta decir que no existen las computadoras inteligentes o sensatas ni las redes neuronales pueden realizar hoy un aprendizaje ilimitado. Por ahora lo cierto es que incorporan mucho mejor que los sistemas expertos el sentido heurístico que tiene toda decisión.

Cómo superar las limitaciones actuales, se encuentra ligado al futuro de los sistemas expertos que, lejos de desaparecer, se potenciarán con el desarrollo de las Redes, ya que el razonamiento humano utiliza ambos procedimientos: conexionistas (bottom up) y simbólico (top down). Si en el futuro se lograra emular esta capacidad será como lo expresa Marvin Minsky:

*“¿Cuál aproximación (corriente) es la mejor para continuar? Esta es simplemente una mala pregunta. Cada una tiene virtudes, deficiencias y necesitamos sistemas integrados que puedan explotar las ventajas de ambas...Pero ahora estamos estancados en el medio; Ni puramente simbólico ni puramente conexionista pueden ser capaces de soportar los tipos de rendimiento intelectual que tomamos por supuesto,..., yo argumentaré que la solución recae en algún lugar entre los dos sistemas y nuestro problema es construir un puente compatible entre ambos.”*

En nuestro Grupo de Investigación, de acuerdo con el concepto anterior, pensamos que las Redes Neuronales podrán en un futuro recrear las reglas con que cuenta el SEDAF (Sistema Experto de Diagnóstico y Asesoramiento Financiero). De esta forma, lograremos una mayor flexibilidad y, en definitiva, uniremos los aportes de ambas herramientas en los sentidos fáctico y heurístico que se necesitan para poder llegar a decidir.

## **3. Gestión experta de Datos: Data Warehousing**

La técnica del Data Warehousing ha ido evolucionando con las necesidades de decisión que se han generado en el ámbito empresarial. Desde los años 80 ha habido una marcada caída del costo de almacenamiento de datos por Megabyte, como así también un vertiginoso crecimiento de la velocidad de procesadores y recursos con los que cuentan las computadoras. Esto ha provocado que las corporaciones comenzaran a almacenar inmensa cantidad de datos, sin tener, en ese momento, un objetivo definido.

La consecuencia lógica de este proceso, fue converger en una paradójica riqueza de datos contrapuesta con una pobreza de información útil. Un hecho que agravó este fenómeno, fue el que en una misma organización se fuesen utilizando diversas fuentes de datos, originadas en los distintos sistemas y módulos de sistemas, que a través de la historia de la empresa se fueron desarrollando, muchas veces basados en dispares tecnologías de bases de datos.



También se debe considerar que el auge que ha tenido la utilización de aplicaciones de ofimática (planillas de cálculo, procesadores de texto, etc...) ha derivado también, en que las empresas cuenten con una gran cantidad de fuentes de datos en los formatos que manejan estas aplicaciones.

Surge de todo lo expuesto la necesidad de tener en las organizaciones una clara política tendiente a utilizar racionalmente técnicas que capturen, reúnan y asocien datos de diversas fuentes, con el fin de presentar la información en forma útil para la toma de decisiones.

### **3.1. Conceptos De Data Warehousing**

Existen muchas y variadas definiciones sobre Data Warehousing cada una de las cuales perfilan diversos aspectos de su implementación. Por ejemplo:

<b>Barry Devlin</b>	<i>"Es un simple, completo y consistente almacenamiento de datos obtenidos de una variedad de fuentes y hecho accesible para usuarios finales en una manera que puedan entender y usar en el contexto de sus negocios"</i>
<b>Bill Inmon</b>	<i>"Conjunto de datos integrados orientados en una materia, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración"</i>
<b>Ken Orr</b>	<i>"Una facilidad de proveer rápido acceso a datos empresariales, calificados e integrados por usuarios finales profesionales y no profesionales"</i>

De estas definiciones podemos inferir que Data Warehousing es el conjunto de datos históricos de una empresa que poseen algunas o todas estas características:

- Independencia de aplicación
- Son referidas a un momento en el tiempo, representativo de un ciclo del negocio
- Almacenadas en una manera fácilmente accesible a personas sin conocimientos técnicos
- Generan un modelo
- Generan "metadatos"

Es conveniente, en este momento, definir algunos conceptos con los que se relaciona estrechamente la tecnología del Data Warehouse, tales como, "metadatos", "data mart", y "OLAP"

**Metadatos:** Una definición sencilla es: "Un metadato es información que describe un dato"<sup>8</sup>, vale decir que son datos acerca de datos.

Un metadato puede contener:

- Información de cómo ubicar diferente información contenida en el Data Warehouse.
- Información sobre los mecanismos por los cuales se extraen datos de las bases de datos operacionales.
- Reglas que determinan los controles de accesos a los datos y condiciones de seguridad de los mismos.
- Reglas y procedimientos contenidos por el usuario final.
- Información sobre la historia del dato.
- Algoritmos de acumulación / agregación de datos.
- Identificación del responsable del almacenamiento del dato.

Existen dos pasos que se deben llevar a cabo antes del desarrollo del Data Warehouse, los cuales hacen un uso estrecho de los "metadatos" de las diversas fuentes de datos a las que se accede:

---

<sup>8</sup> LARDENT, Alberto R. "Sistemas de Información para la gestión empresarial (Tomo I) – Planeamiento, Tecnología y Calidad", Ed. Pretince Hall, Buenos Aires, 2001, Capítulo 18



<p><b>Refinamiento (Cleaning)</b></p>	<p>Las diversas fuentes de datos que utiliza el Data Warehouse, realizan diferentes tratamientos de los datos que contienen. De esta forma determinados tipos de datos que son almacenados en un sistema, no reciben el mismo tratamiento al ser almacenados en otros sistemas. Por ejemplo: un campo donde se almacene un dato puede tener un impedimento del sistema de almacenar registros en blanco, nulos, o utilizar un grado de precisión predeterminada, que no se corresponde con un almacenamiento del mismo tipo de dato en otro sistema.</p> <p>En este paso, en definitiva, se intenta lograr que los datos que va a utilizar el DW lleguen a tener un grado de conformidad aceptable por la empresa.</p>
<p><b>Reingeniería (Transformation)</b></p>	<p>El proceso del reingeniería de datos pretende producir un cambio en los datos a fin de que sean legibles para el usuario. La base de la necesidad de transformar los datos se encuentra en decodificar, a fin de hacer comprensibles, datos previamente codificados por razones de optimización o de espacio de almacenamiento. Por ejemplo: se puede haber guardado en una tabla una codificación de acciones representadas por números (1: actualizar mensualmente, 2: Semanalmente, 3: Diariamente)</p>

**Data Mart:** Algunas de las definiciones que se han enunciado de Data Mart son:

<p><b>Aaron Zornes (Meta Group)</b></p>	<p><i>"Un ámbito o departamento orientado del Data Warehouse. Puede incluir duplicación de datos de un Data Warehouse corporativo y/o datos locales"</i></p>
<p><b>Gartner Group</b></p>	<p><i>"Un conjunto descentralizado de datos de un almacenamiento diseñado para soportar los requerimientos de un mercado específico de operación o análisis"</i></p>
<p><b>Bill Inmon</b></p>	<p><i>" el Data Warehouse departamental es utilizado para mantener la información departamental que es extraída del Data Warehouse de la empresa. Este es a veces descripto como livianos y altamente resumidos datos, o como data marts"</i></p>

Estas definiciones nos perfilan al Data Mart como *"una implementación del Data Warehouse pero referida a un ámbito de datos y funciones más limitados"*<sup>9</sup>

La figura 7 muestra la relación existente entre las fuentes de datos operacionales, el Data Warehouse y un Data Mart de una empresa.

<sup>9</sup> Data Warehousing with DB2 for OS/300 – Diciembre 1997 – First Edition – IBM Corporation, International Support Organization. Cap. 1 y 2

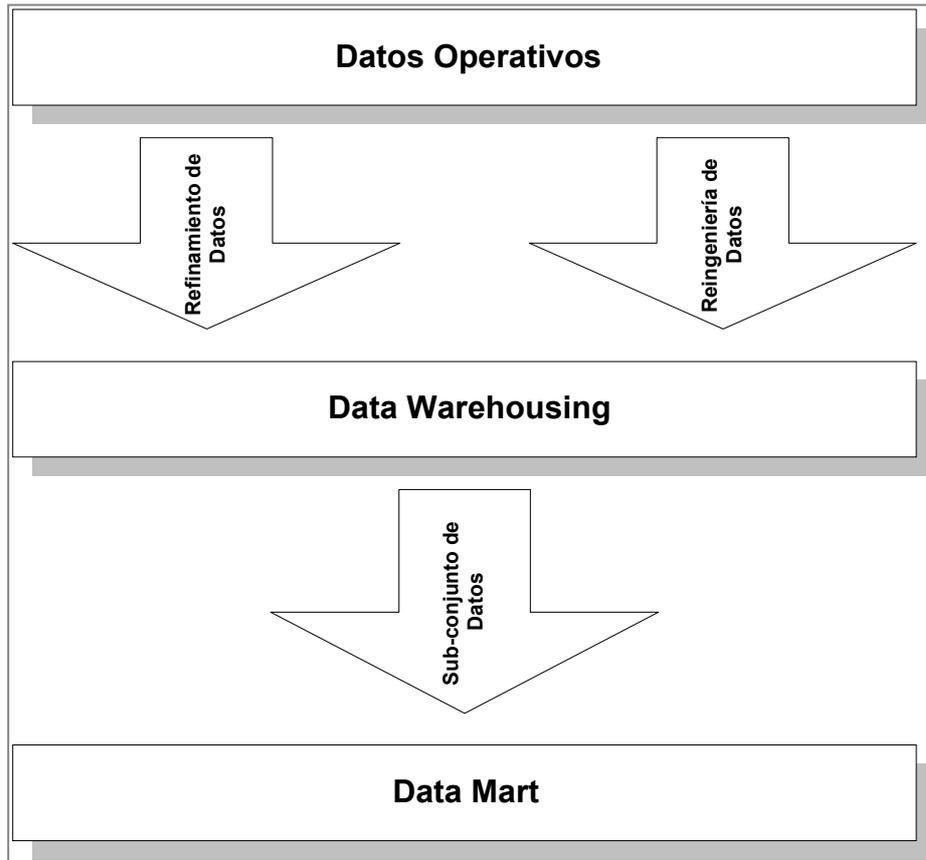


Figura 7 : Relación entre Fuentes de Datos Operativas, Data Warehouse y Data Mart

### **OLAP (Procesamiento analítico en línea)**

Otra de las técnicas que se debe emplear para implementar un sistema de Data Warehousing es desarrollar un sistema de Procesamiento analítico en línea.

Para obtener el mismo se debe tener en consideración:

- Proveer fácil acceso a los usuarios finales.
- Se debe proveer una interfaz desde la cual los usuarios puedan determinar los parámetros de la información que están buscando.
- Debe proveer una rápida respuesta.

Una comparación utilizada muy frecuentemente, y en algunos casos adoptada para denominar a los sistemas OLAP<sup>10</sup> es la de utilizar la imagen de un "cubo". Esta imagen, transmite al usuario en forma intuitiva la idea de tres dimensiones. De esta manera se puede asimilar la visualización de datos de acuerdo a un ordenamiento dado (por ejemplo, temporal, espacial, etc. ...).

Por ejemplo, una entidad podría requerir que su sistema OLAP ordene las ventas de un determinado grupo de productos (golosinas), en un determinado periodo de tiempo (1º trimestre de 2002) y en una determinada área de cobertura (centro del país).

Las bases de datos que implementan estos sistemas son llamadas comúnmente "multidimensionales". Cabe destacar que la comparación del sistema con un cubo, no implica que las dimensiones del mismo se restrinjan a tres, sino que el administrador de datos de la empresa puede crear la cantidad de dimensiones que necesite visualizar de sus datos a fin de brindar información útil para las decisiones.

<sup>10</sup> El sistema gestor de base de datos SQL SERVER de Microsoft, trabaja con objetos creados por el administrador de la base llamados "cubos".

En la figura 8 se puede ver una representación gráfica de un ordenamiento de datos de acuerdo a tres dimensiones: línea de productos, temporalidad y espacio. Una vez construido, el cubo resulta

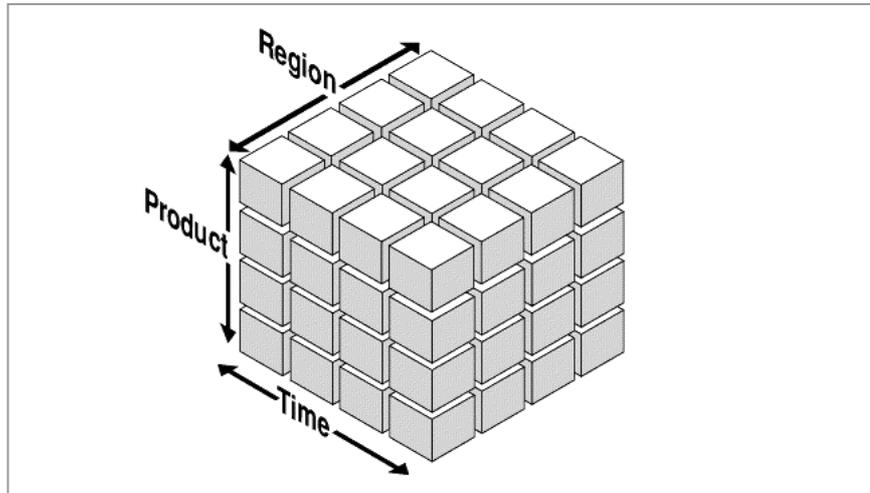


Figura 8: La imagen de un cubo permite graficar la ordenación de datos según distintas dimensiones (Imagen extraída de "ORACLE 8i – Data Warehousing Guide")

sencillo de modificar y analizar. Puede ser particionado, para consulta de uno u otro dato que resulte relevante y permite la expansión independiente de los datos (dimensiones elegidas). También puede unirse lógicamente con otros cubos, conformando un denominado "cubo virtual", que no está constituido por datos sino por relaciones entre los mismos.

"Un cubo es un operador que genera un conjunto multidimensional que constituye una expansión de datos fácticos, datos que explican eventos individuales. La expansión se basa en columnas que el usuario desea analizar. Esas columnas se denominan dimensiones. El cubo es el resultado que contiene en una tabla las relaciones cruzadas de todas las posibles combinaciones de las dimensiones."<sup>11</sup> En otras palabras, el cubo reúne datos de diferentes orígenes y crea una vista particular que los relaciona, brindando al usuario una visión integrada que supera a la que brinda una tabla de doble entrada.

### **Como construir un Data Warehouse**

El diseño de datos dimensionales es toda una especialidad que requiere no sólo del manejo de técnicas de modelado sino también conocimiento de los procesos de negocio con que vayamos a tratar.

Habiendo visto las técnicas y herramientas que podemos utilizar, podemos ver en detalle los pasos metodológicos para construir un sistema de Data Warehousing.

<b>1</b>	<b>PLANEAMIENTO</b>	<ul style="list-style-type: none"> <li>• Se crea la "agenda" de tareas, con consideración de los plazos propuesto para lograr el desarrollo.</li> <li>• Se debe realizar el "estudio de factibilidad económica" del proyecto.</li> <li>• Determinación de los destinatarios de la información</li> </ul>
----------	---------------------	--

<sup>11</sup> BEJELETICH, Sharon, MABLE, Greg, et.al, "Microsoft SQL Server 7", Prentice Hall, Madrid, 1999, págs.779-790



2	<b>FUNCIONES</b>	Se ha de considerar en esta etapa las expectativas de los distintos sujetos asociados al Data Warehouse. Se debe tomar en cuenta: <ul style="list-style-type: none"><li>• La perspectiva del empresario</li><li>• La perspectiva de la tecnología utilizada</li><li>• La perspectiva del usuario final</li></ul>
3	<b>ANALISIS</b>	Se define los requerimientos de funciones de la etapa anterior
4	<b>DISEÑO</b>	Se transforma el modelo "lógico" desarrollado en un modelo "físico"
5	<b>CONSTRUCCION</b>	Se ensamblan los componentes desarrollados en la etapa anterior
6	<b>PRUEBA</b>	Es sugerible el desarrollo de un "piloto" para realizar los distintos controles de calidad que debe soportar el sistema
7	<b>INSTALACION Y UTILIZACION</b>	En esta etapa se debe prestar atención a la sincronización de los distintos componentes así como verificación de accesos
8	<b>OPTIMIZACION</b>	De la interacción con el sistema, seguramente surgirán nuevos requerimientos que deben ser atendidos

Se estima un tiempo prudente de desarrollo de un sistema de Data Warehousing en aproximadamente 9 - 12 meses.

### **DATA WAREHOUSE – DATA MART. Hacia su aplicación en el ámbito de las Pymes**

De todo lo expuesto podemos valorar el aporte sustancial que ha dado a la Tecnología Informática el Data Warehouse. Ahora bien, cabe preguntarse, ¿cómo podríamos aplicarlo a empresas locales?.

En la introducción de este trabajo vimos que el Data Warehouse tuvo su origen a partir del crecimiento anárquico de fuentes de datos, por falta de planificación de recursos informáticos, por parte de las empresas. En un escenario ideal, las pequeñas empresas al expandirse, estableciendo sucursales, centros de negocios o administraciones descentralizadas, deberían haber tenido que migrar la totalidad de sus datos a medida que cambiaban sus sistemas.

En la práctica, las empresas forzadas por la inestabilidad económica, cambios de reglas y falta de previsibilidad, han ido mostrando lo contrario. El hecho es que en el ámbito de las Pymes argentinas, nos encontramos con una urgente necesidad de replanificar los recursos de información, muchas veces dispersos y desestructurados, con que cuentan las empresas, a fin de disponer contenidos útiles para la toma de decisiones racionales.

Una de las tendencias que muestra actualmente la tecnología informática, es la de ir avanzando constantemente en amplios consensos de aceptación de estándares de codificación de datos, programas y recursos utilizables. Internet y sus lenguajes universalmente aceptados como HTML o XML son una clara muestra de esto.

De esto podemos prever un aceleramiento en la circulación e intercambio de información motivado por la facilidad de transferencia de datos estandarizados, lo cual derivará en la necesidad de optimizar la información decisional en tiempo y calidad, para poder mantener una empresa competitiva

El hecho de que una pequeña o mediana empresa, cuente con una clara estrategia de utilización de sus datos, sin duda, la posicionará en mejor lugar para competir y expandir sus posibilidades. Los desarrollo de un Data Warehouse implica un esfuerzo que muchas PyMEs no pueden encarar, para lo cual han surgido las aplicaciones de Data Mart, productos que utilizan una tecnología similar, pero permiten una definición de datos más fácil y rápida, disminuyendo notablemente el costo del proyecto.



En principio resolvían el problema a nivel departamental de la gran empresa, y actualmente están siendo aplicados a la resolución integral del problema en empresas medianas. Mejora el índice precio/rendimiento y permite que estas empresas logren implementar sus propios Sistemas de Información para Ejecutivos (EIS).

Para ello se requiere de la creación de estructuras de datos multidimensionales (cubos), y la determinación de las relaciones entre las mismas.

Una vez establecidas las variables (atributos) a analizar en el nuevo entorno multidimensional, estamos en condiciones de asociarlas y de conformar reglas del tipo: "SI  $A_1$  y  $A_2$  ... $A_i$  ocurren, ENTONCES es usual el caso en el cual  $B_1$  y  $B_2$ ...  $B_n$  también ocurran en la misma transacción." Este esquema surge como consecuencia de la observación de un gran número de transacciones, y de su clasificación y agrupamiento mediante las técnicas apropiadas. Se basa en dos conceptos básicos: El soporte, es decir la probabilidad de que una misma transacción contenga elementos de A y de B y la confianza, o probabilidad de que una transacción que contiene A contenga también B.

Las reglas de asociación pueden abarcar múltiples niveles de abstracción si los datos a ser utilizados son jerárquicamente organizados. Un nivel de abstracción se refiere a un nivel conceptual o jerárquico de un cubo de datos. Es función del usuario determinar el nivel de cada dimensión utilizada en el proceso. El menor nivel de jerarquía es aquél que tiene una sola dimensión (ej: un dato crudo, como el ticket de supermercado, que integra la facturación de una caja, de una sucursal, de una ciudad, de un momento dado).

#### 4. La inteligencia artificial aplicada a los datos: Data Mining

La minería de datos es un conjunto de tecnologías o herramientas capaces de analizar la información almacenada en un Data Warehouse con el objeto de ayudar a descubrir relaciones insospechadas, tendencias, modelos de comportamiento o correlaciones entre los datos contenidos en el mismo. A través del procesamiento de información, utilizando algoritmos de búsqueda estadística, DataMining provee las bases para transformar la información en conocimiento. Basándose en su capacidad de organizarlos en diferentes dimensiones de análisis potencial.

Las etapas en la definición de un DataMining son simples y para su mejor análisis puede tomarse nuestro trabajo anterior sobre el tema<sup>12</sup>. Sólo remarcaremos los elementos constitutivos más importantes, a ser desarrollados en forma secuencial<sup>13</sup>:

1. **Definir los orígenes de datos.** Mediante el establecimiento de la ubicación de los mismos.
3. **Construir los cubos.** De acuerdo a lo establecido previamente.
2. **Definir las dimensiones a compartir.** Permite adaptar la representación obtenida a las necesidades de reportes standard que cada empresa necesita. La clave consiste en encontrar las dimensiones críticas para guiar las decisiones particulares de cada negocio.
4. **Desarrollo de modelos de minería** Siguiendo con el ejemplo anterior, si agregamos en función de tiempo (período de ventas), producto y región de venta, perderemos la dimensión color del producto o la dimensión de sexo de los compradores. El modelo siempre implica una elección.
3. **Definir los roles de las bases de datos.** Esto es depurar los distintos procedimientos SQL para corroborar sus resultados con lo deseado por los usuarios finales.

La existencia de la gran cantidad de datos, y la posibilidad de que los mismos sean consultados en forma interactiva por parte de los distintos agentes a través de lenguajes como el SQL, genera una continua intromisión en la BdeD que puede llegar a perjudicar la operatoria normal del sistema. Surge la necesidad de contar con archivos intermedios que resuman las actividades históricas en forma de totales.

<sup>12</sup> Pluss, Jorge et.al "El conocimiento en los Sistemas de Información", 6tas. Jornadas de Investigación en la Facultad de Ciencias Económicas, 2001

<sup>13</sup> DE VILLE, Barry, "Modelling Data", Digital Press, 2001

La demanda de este tipo de datos se ha incrementado, dado que los gestores de empresas comienzan a entender y apreciar los diferentes niveles de inteligencia de negocio que los sistemas de BdeD permiten conseguir. En la realidad de nuestras PyMEs este ámbito no ha sido aún descubierto, ya que tampoco existe una experiencia generalizada de cooperación entre las mismas.

Es necesario realizar una considerable labor de análisis y recopilación de requisitos del negocio para asegurar que el diseño de un almacén será flexible y ampliable y satisfará las necesidades de los usuarios. Una vez diseñado, comienza la tarea de transformación y migración de datos, procedentes de diversas fuentes, lenguajes y estructuras.

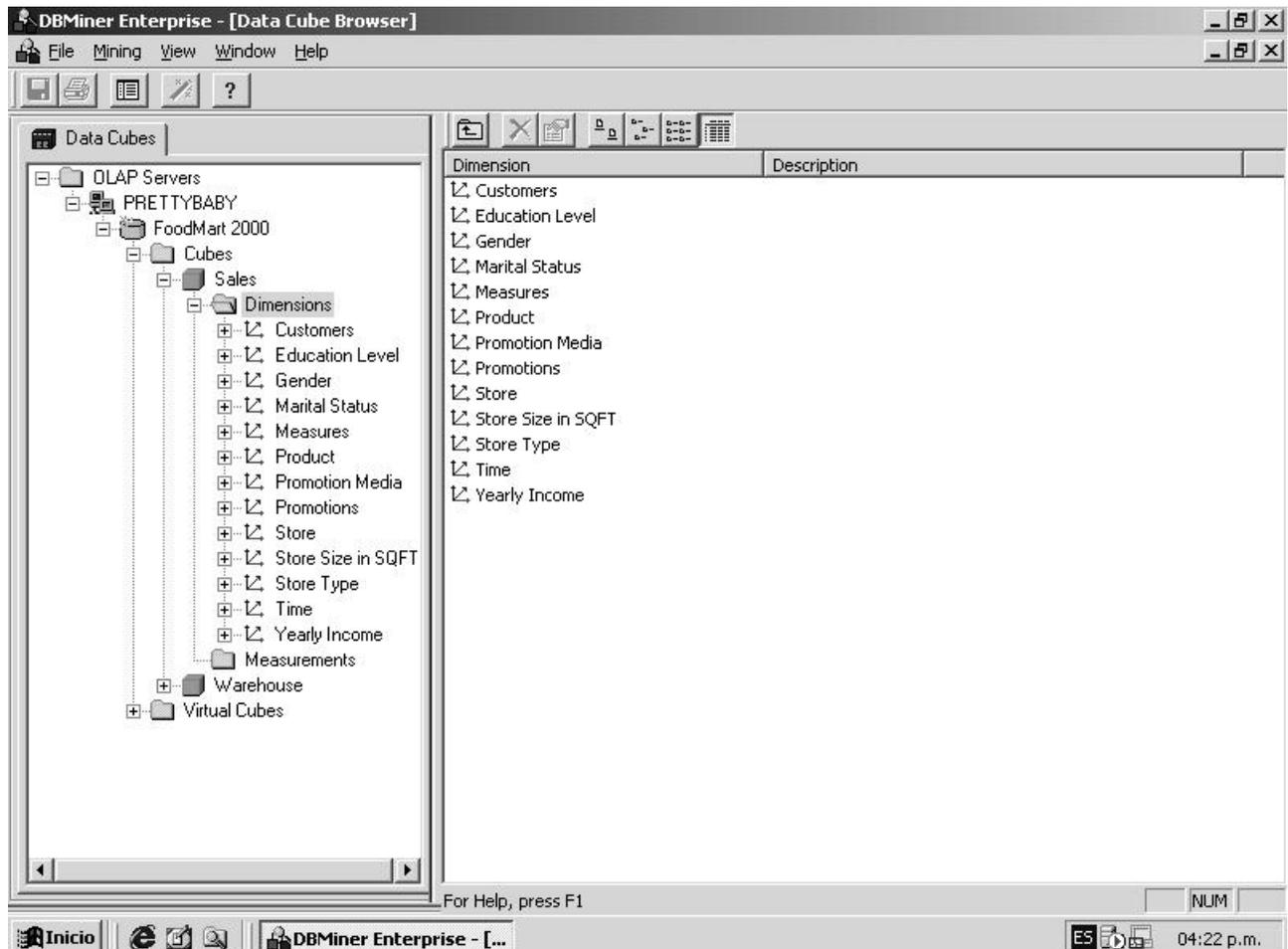


Fig.9: Pantalla del Software DBMINER Enterprise de Micorsoft

En la pantalla anterior puede visualizarse desde la empresa ejemplo, cómo se desarrollan los cubos y dentro del cubo de ventas "Sales" las diferentes dimensiones del análisis de datos.

#### **4.1. Knowledge Discovery y Data mining:**

Existe una confusión acerca del verdadero significado de ambos términos, siendo que algunos autores los consideran sinónimos, pero en la primera conferencia internacional sobre KDD (KDD Knowledge Discovery in Databases) que se realizó en Montreal en 1995 se propuso el término KDD para describir todo el proceso de extracción de conocimiento de las datos. Conocimiento, significa relaciones y patrones entre los datos. Data mining debería ser considerado como una etapa descubrimiento dentro del procesos de KDD.

El descubrimiento de conocimiento en bases de datos está íntimamente relacionado con el desarrollo de otro concepto importante, el Data warehousing o almacenes de datos. La información en el Data Warehousing es orientada al sujeto, no volátil y de naturaleza histórica, por lo que los almacenes de



datos tienden a contener una extrema cantidad de datos. La combinación de los almacenes de datos, soporte de decisiones, y Data mining muestran un innovador y revolucionario enfoque de la administración de información.

### **5. Conclusión: La información como factor de producción:**

Muchas de las compañías internacionales producen más información en una semana de la que una persona podría leer en toda su vida. En el futuro la habilidad de una persona para leer e interpretar no va ser suficiente para sobrevivir en el ambiente profesional, tanto sea para organizaciones científicas como comerciales. La producción rutinaria y reproducción de datos nos fuerza a adaptar nuestras estrategias y desarrollar nuevos sistemas para filtrar, seleccionar e interpretar los datos.

Las operaciones de las bolsas de valores se han convertido, en un sentido, en un juego de computadoras contra computadoras donde los humanos solo intervienen a un nivel de "diseño" (meta-level).

El proceso de data warehousing debe orientarse a proveer la información correcta, a la persona indicada, en el formato adecuado y en el tiempo preciso. Su aplicación es ventajosa en cualquier área de negocios donde sea necesario mejorar el proceso de toma de decisiones, acceder a información clave, obtener valor agregado de la articulación de los diferentes sistemas operativos, soportar la toma de decisiones en los niveles tácticos y estratégicos, además del operativo. Entre los sectores que mayor esfuerzo invierten en este área se encuentran el financiero, el de telecomunicaciones y el de retail. Existe una fuerte tendencia a ver el data warehousing como una necesidad estratégica sin la cual es imposible competir.

Las herramientas de data mining permiten explotar el conocimiento oculto en los grandes volúmenes de datos. El notable incremento en la última década de la relación performance / costo de las plataformas de hardware y software permiten el uso de técnicas de inteligencia artificial aplicadas al campo de los negocios, a fin de detectar tendencias, descubrir relaciones, identificar nuevos patrones de comportamiento de nuestros clientes, segmentar el mercado sobre la base de nuevas dimensiones o calcular el impacto de numerosas variables en las diferentes estrategias y tácticas de negocio.

En el ámbito de mercadotecnia se pueden descubrir relaciones entre clientes y productos ( y así gestionar una promoción comercial más específica o dirigida), detectar hábitos de consumos ocultos dentro de esa información histórica y definir perfiles de consumidores o usuarios ( esto reduce los gastos de promoción e incrementa su efectividad), ayuda a corregir u orientar estrategias en materia de televentas, consecuencias de análisis de respuestas.

En el ámbito del seguro, las compañías pueden efectuar análisis de riesgo para localizar las características de clientes de alto riesgo. Analizando los hábitos de pago, una entidad crediticia puede prever el nivel de riesgo de conceder un crédito a un solicitante. Con relación a la medicina, ya se aplica en el análisis de respuestas a los diagnósticos y tratamientos.

Una diferencia entre las herramientas de ayuda a la decisión y el Data Mining, es en que en las primeras, la iniciativa sobre qué elementos se deben observar o analizar corresponde al usuario. En cambio, el Data Mining, toma la iniciativa por sí mismo y descubre correlaciones existentes pero ocultas entre los datos. De ahí que mediante esta herramienta sea posible anticipar una hipótesis o detectar un comportamiento inusual en un cliente.

Podemos ver al Data Warehousing como la base de un depósito de conocimiento que, usado eficazmente, permite identificar estrategias de reducción de costo, lograr servicios de valor agregado a un costo manejable y generar dentro de la organización una mayor eficacia de los datos. También es la base de la infraestructura de una organización de comercio electrónico, la de la gestión del conocimiento.

Toda organización que desee hacer más eficiente el proceso de toma de decisiones debe en primer término fijar sus objetivos, hacerlos explícitos y luego encontrar la metodología que mejor se adapte para la consecución de los mismos. Sabiendo que su implantación es un proceso que implica trabajo y compromiso de todos los sectores, deberá generar sus propios almacenes de datos y capacitar a su personal para una nueva etapa: la de decidir con información. Se trata de un paso cualitativo que incidirá en toda la estructura administrativa, aportando importantes resultados, más positivos cuando mejor esté diseñada la estructura de datos.



## BIBLIOGRAFIA:

BJELETICH, Sharon, MABLE, Greg, "Microsoft SQL Server 7 al Descubierta", Prentice Hall, Madrid, 1999

EDMOND, David, "Information Modeling", Prentice Hall, 1992

FLIPPO, Edwin, "Dirección de Empresas", Musinger

GROS, Jorge, "Del Query al Mata Mining", Windows NT Magazine, julio-agosto 1997

IBM Corporation, International Support Organization, "Data Warehousing with DB2 for OS/300", Diciembre 1997

IOVALDI, Raul, "Inteligencia Artificial y Aprendizaje: investigación y metodología", Universidad Nacional de Rosario editorial, Rosario, 2001

LARDENT, Alberto R. "Sistemas de Información para la gestión empresarial (Tomo I) – Planeamiento, Tecnología y Calidad", Prentice Hall, Buenos Aires, 2001

LAUDON, Kenneth, LAUDON, Jane, "Administración de los Sistemas de Información: organización y tecnologías", Prentice Hall Hispanoamerica editora, 3er. Edición, México, 1996

Microsoft Corp. "SQL Server - Books Online", 1998-2000

ORACLE CORP., "Oracle8i Data Warehousing Guide Release 2 (8.1.6) A76994-01",

<http://www.oracle.com>

PLUSS, Jorge, MARCHESE, Alicia, PICCO, Alicia et.al. "El conocimiento en los Sistemas de Información", 6tas. Jornadas de Investigación en la Facultad de Ciencias Económicas, 2001

PADERN, Xavier, "Redes Neuronales, una introducción",

<http://www.eupmt.es/cra/inform/info11.htm>

RUSSEL, Stuart, NORVIG, Peter, "Artificial Intelligence: a modern approach", Prentice Hall editor, serie Artificial Intelligence, New Jersey, 1995

SIMON, Herbet, "El comportamiento administrativo", Aguilar Argentina S.A., 2da. Edición, Buenos Aires, 1984

Tutorial Redes Neuronales, <http://www.gc.ssr.upm.es/inves/neural/ann2/anntutor.htm>