

## CAPÍTULO VI

---

### *Estado actual y perspectivas*

El objetivo de este capítulo es mostrar los resultados obtenidos con la realización de la tesis de aplicación, el estado actual del sistema PIE, su validación y las mejoras que se pueden hacer al mismo, así como las conclusiones que se obtuvieron de la aplicación de los procesos de la inteligencia empresarial a las PyMEs. El capítulo VI se encuentra organizado de la siguiente forma: en la sección 6.1 se describen los resultados obtenidos de las actividades realizadas a lo largo de la tesis, en la sección 6.2 la validación del sistema y las métricas para implementar la metodología a otro tipo de PyMEs, en la sección 6.3 se describen los pasos a seguir para hacer extensible el sistema, en la sección 6.4 se describen las recomendaciones y mejoras contempladas para el sistema y por último en la sección 6.5 se describen las conclusiones y análisis finales de la tesis.

### **6.1 Resultados obtenidos**

El conjunto de actividades realizadas a lo largo de esta tesis permitió desarrollar conocimiento teórico-práctico para el área computacional, del cual obtuvimos los siguientes resultados, que corresponden a los objetivos específicos propuestos en el capítulo I:

- Se definió una arquitectura, que involucra las técnicas de la inteligencia empresarial, para PyMEs, dividida en 3 capas: integración, análisis y visualización.

- Se diseñó un modelo de datos multidimensional para PyMEs comercializadoras que apoya la toma de decisiones en el área de CRM ayudados de la técnica MBA, el cual también puede ser utilizado para otras PyMEs comercializadoras, debido a que se encuentra basado en un diagrama entidad-relación que es utilizado por otras empresas, y también puede ser utilizado para apoyar la toma de decisiones de otras áreas como ERP.
- Se definieron las transformaciones y trabajos necesarios para automatizar los procesos de extracción, limpieza y carga de la información que alimentó al *datawarehouse*.
- Se implementó un *datawarehouse*, siguiendo el modelo de datos multidimensional, descrito con anterioridad, el cual hizo posible los análisis de OLAP y *data mining* para PyMEs.
- Se definió un hipercubo de OLAP con 5 dimensiones, que permitió el análisis de la información almacenada para apoyar al área de CRM.
- Se aplicó el del algoritmo *Apriori* de *data mining* a los datos almacenados en el *datawarehouse*, con el fin de encontrar reglas que representen la asociación de productos que se compran juntos en una o varias transacciones por medio de la técnica MBA. Estas reglas permitieron establecer un modelo de comportamiento del cliente, identificando los productos de su preferencia.
- Se desarrolló un prototipo del sistema integrado que implementa las 3 técnicas de la inteligencia empresarial: *datawarehousing*, OLAP y *data mining*.
- El procedimiento que seguimos a lo largo de la tesis, ayuda a definir un marco metodológico para la aplicación del proceso de inteligencia empresarial a otro tipo de

PyMEs comercializadoras, esto pudo comprobarse porque se aplicó el modelo con los datos de otra empresa: *Abarrotes Tlacotepec*.

## 6.2 Validación

La capa de análisis, pasó por una etapa de validación, para corroborar que el sistema estuviera entregando resultados viables y confiables. Esta etapa fue muy importante, ya que los resultados entregados serán usados por el administrador de la empresa como justificación para las decisiones que sean tomadas.

Para los reportes de OLAP, se tuvo una entrevista con el contador de la PyME para verificar que los totales (medidas) tanto en pesos como en cantidades estuvieran correctos. Esta validación permitió comprobar que los resultados obtenidos de los reportes eran correctos.

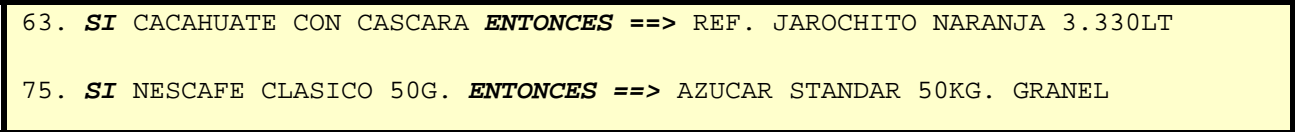
De esta manera el usuario tiene la seguridad de tomar decisiones basadas en resultados confiables, así mismo él puede realizar nuevos cálculos con los resultados obtenidos y de igual manera se estarán manejando resultados fiables.

Por otro lado, para las reglas de asociación obtenidas con el algoritmo *Apriori* de *data mining*, la validación consistió, en el ajuste de las medidas de confianza y soporte y la verificación con el administrador.

Para los 4 archivos generados, se consideró que sólo serían válidas las reglas que tuvieran una confianza mayor a 0.25 o 25%, como se mencionó el capítulo 4, se seleccionó esta confianza para encontrar más reglas además de las que resultan obvias para el usuario, que mostraran información relevante, que pudiera serle de interés.

El soporte se definió en particular para cada archivo, esto difiere principalmente por lo que se quiere encontrar en cada uno. No se escogió el soporte máximo de “1” porque como se mencionó con anterioridad ya que, no existe cliente que haya comprado todos los productos de la tienda y buscar esta medida de soporte sólo ocasiona demasiado procesamiento y uso de memoria que perjudica el desempeño del sistema. El rango de soporte de los archivos se encuentra entre 0.005 y 0.5.

Las reglas obtenidas del algoritmo *Apriori*, se encuentran con mayor detalle en el apéndice C. Aquellas que resultaron más obvias fueron validadas por el administrador del supermercado, y las demás reglas que existen aunque que no resultan obvias, tienen lógica, como se muestra en la figura 6.1, lo cual indica que el resto de las reglas también son válidas.



```
63. SI CACAHUATE CON CASCARA ENTONCES ==> REF. JAROCHITO NARANJA 3.330LT
75. SI NESCAFE CLASICO 50G. ENTONCES ==> AZUCAR STANDAR 50KG. GRANEL
```

**Figura 6.1** Ejemplos de reglas de asociación obtenidas del algoritmo *Apriori*.

### **6.2.1 Métricas recomendables para PyMEs**

Como mencionamos en la sección 6.1, las actividades realizadas a lo largo de la tesis permitieron desarrollar una metodología para la aplicación del procedimiento de inteligencia empresarial, que también es aplicable para otras PyMEs comercializadoras. A continuación se describirán los pasos necesarios que se deben ejecutar para aplicar el modelo a otras empresas:

- ***Partir de una nueva implementación desde cero:***

1. Definir e implementar un modelo multidimensional, con al menos una estrella.

2. Realización de las transformaciones y trabajos pertinentes para obtener la información deseada que alimente al modelo multidimensional del *datawarehouse*.
3. Implementación del *datawarehouse*, ejecutando los trabajos definidos anteriormente.
4. Definición del cubo multidimensional representado por el documento XML, con el que se realizarán las consultas de OLAP.
5. Definición de las consultas básicas con las que se desea obtener la información relevante de la empresa.
6. Aplicación del algoritmo *Apriori* a los datos almacenados en el *datawarehouse*.

- ***Utilizando el modelo propuesto y el sistema PIE:***

1. Implementar el modelo multidimensional en una base de datos relacional.
2. Alimentar el *datawarehouse* con el modelo propuesto a partir de los datos de la PyME. Si se quisiera cambiar de manejador de base de datos se pueden utilizar scripts de SQL, que contienen la estructura de las tablas empleadas para la implementación del *datawarehouse*, estos scripts se ejecutarían en el nuevo manejador de base de datos.
3. Reutilizar las consultas ya predefinidas o diseñar nuevas consultas que sean necesarias, dependiendo del área a la que se quiera dar apoyo.
4. Aplicar el flujo de trabajo, utilizando los programas *ArffGenerator*, *ReportFormatter* y *Apriori*, para lo concerniente a *data mining*.

5. Reutilizar la interfaz del sistema PIE, para visualizar tanto los resultados de OLAP como los de *data mining*.

En la tabla 6.1 mostramos los tiempos que tardan en ejecutarse cada una de las etapas que conforman el proceso.

Actividad	Tiempo	Total tiempo
ETL (transformaciones)	4 min. Aprox. Cada una	32 min.
ETL (trabajo)	--	30 min.
Consultas OLAP	3 seg. por consulta	--
Algoritmo Apriori	13.5 min. por archivo	54 min.

**Tabla 6.1 Tiempos de ejecución**

Como vemos el rango de tiempo de ejecución en el que nos encontramos es bastante satisfactorio, para el volumen de datos que estamos manejando, recordemos que tenemos alrededor de 10,000 ventas por mes.

Aunque el tiempo que se observa para la aplicación del algoritmo *Apriori* puede verse un poco elevado. Es importante recordar que esta actividad solo se realizará al momento en el que se actualicen los datos del *datawarehouse*, por lo que dicha actividad solo se realizará probablemente una vez al mes. Esta fue una de las principales razones por las que se decidió no ejecutar el algoritmo cada vez que el usuario quisiera consultar los resultados.

### **6.2.2 Infraestructura, costos y recursos humanos**

La infraestructura mínima necesaria para que el prototipo se desempeñe de manera aceptable, se define a continuación:

- Cualquier sistema operativo
- JVM versión 1.5 o superior

- Procesador Intel Pentium IV a 3.2Ghz
- 512MB en memoria RAM
- 100mb por cada 10,000 registros de espacio en disco (aproximadamente).

La infraestructura descrita anteriormente esta considerada para efectuar únicamente el proceso de inteligencia empresarial, si la PyME ya cuenta con una máquina con éstas características y puede compartirse para realizar las operaciones de análisis, entonces puede reutilizarse, de lo contrario será preferible designar una máquina especial para efectuar el procedimiento.

Todo desarrollo de sistemas computacionales implica costos, ya sean económicos de tiempo o de tecnología. En la tabla 6.2 se muestra un estimado de los costos que implicó desarrollar el sistema PIE.

Resulta complicado comparar el costo de adquisición con los sistemas de Oracle, Cognos y Microsoft, porque se trata de un proyecto individual que no iguala la funcionalidad de estos, sin embargo recordemos que las PyMEs no requieren de toda la funcionalidad que brindan los sistemas comerciales.

Costo	Herramientas	Diseño	Implementación	Pruebas	Correcciones	Total
<b>Horas/persona</b>	---	256 hrs.	192 hrs.	96 hrs.	96 hrs.	480 hrs.
<b>Pesos (\$93.75/hr)</b>	\$0.00	\$24,000.00	\$18,000.00	\$9000.00	\$9000.00	\$60,000.00

**Tabla 6.2 Costos de adquisición del sistema**

Quedarían pendientes los costos de operación del sistema, en lo que respecta a la PyME, el administrador fungiría como analista por lo que no se requiere personal adicional, como ocurre

por ejemplo con el grupo de expertos necesario para Oracle o Cognos mencionados en la sección 1.3.

Como se observa en la tabla 6.2, el costo de adquisición de este producto es mucho menor que adquirir un sistema comercial como los mencionados en el capítulo I, y la ventaja es que podemos reutilizarlo para otras PyMEs.

En la siguiente sección veremos las posibilidades de escalabilidad que tiene tanto el modelo multidimensional, como la metodología propuesta para las PyMEs, en cada una de las capas involucradas.

### **6.3 ¿Cómo escalar las funcionalidades del sistema PIE?**

Una de las principales ventajas que tiene el sistema PIE es que puede ser extensible en varios aspectos los cuales mencionaremos en seguida:

- ***El modelo multidimensional***

- *Nuevo hecho:* el modelo actual es una estrella que analiza el proceso de las ventas, en caso de que se deseara analizar otro proceso del negocio, se realizaría un modelo multidimensional semejante al presentado en la sección 4.3 que contenga el proceso a estudiar con sus respectivas dimensiones y atributos. Puede ser que las dimensiones del hecho existente sean las mismas que las del nuevo hecho, en este caso se pueden compartir dichas dimensiones y entonces tendríamos una constelación, como se muestra en la figura 6.2 donde se agrega el hecho *compra\_fact*, que se relaciona con la



dimensión *proveedor* (también agregada) y las ya existentes *región* y *tiempo*. Para almacenar estos datos en *datawarehouse*, se deben crear las transformaciones necesarias que extraigan los datos de las tablas de Paradox o donde se encuentren los datos de origen y crear nuevas tablas en el *datawarehouse* en MySQL o el manejador de destino. Con esto podemos concluir que haber utilizado una sola estrella facilita la representación del problema y se puede apreciar la escalabilidad del modelo ya que, a partir de varias estrellas podemos crear una constelación que de soporte a tantas áreas del negocio como deseemos, sin complicar la representación de los datos.

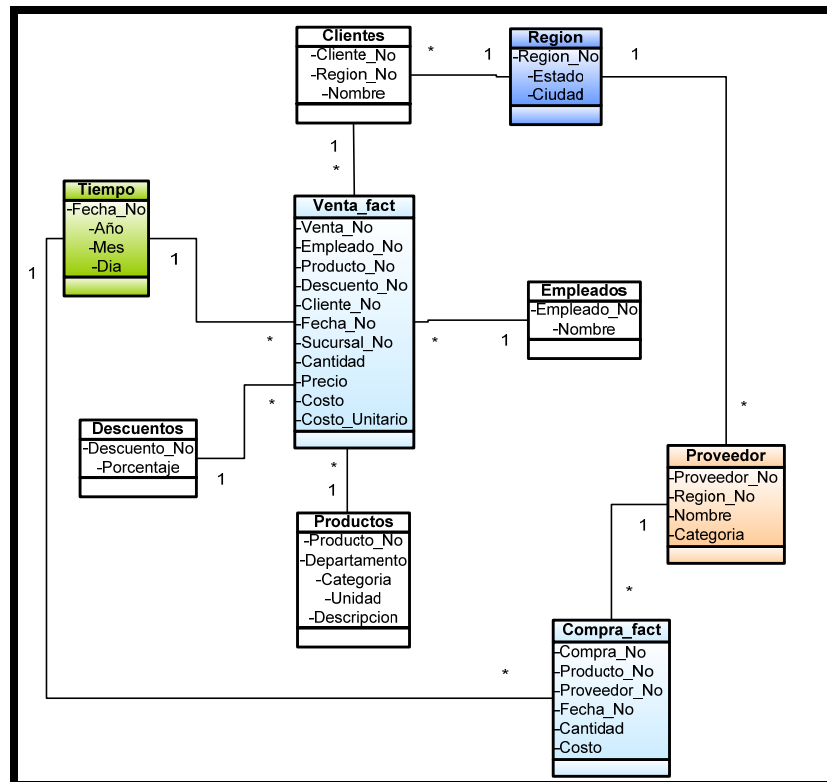


Figura 6.2 Agregación de un nuevo hecho al modelo multidimensional con sus dimensiones

- Nueva dimensión: para añadir una nueva dimensión al modelo existente, simplemente se agrega la dimensión con sus respectivos atributos y a la tabla hecho se le agrega la llave primaria de la nueva dimensión, como vemos en la figura 6.3, con la nueva dimensión *sucursal*.

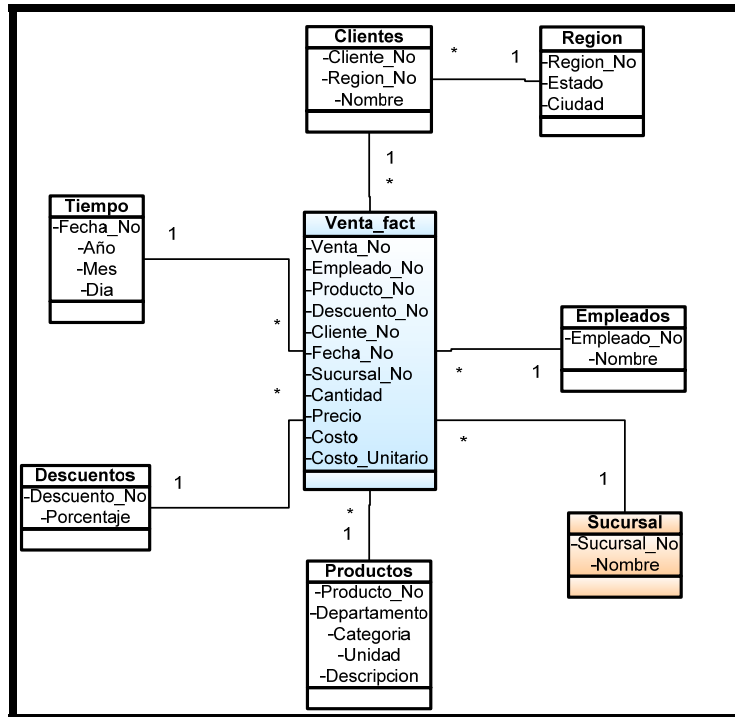


Figura 6.3 Agregación de una nueva dimensión al modelo multidimensional.

Se crean las transformaciones pertinentes para obtener los datos y almacenarlos en el *datawarehouse*. Posteriormente se le agrega al documento XML, las líneas que representan la nueva dimensión y sus jerarquías, como se ve en la figura 6.4.

```
<Dimension name="nombre_dimension" foreignKey="Dimension_No">
  <Hierarchy name="nombre_dimension" hasAll="true" allMemberName="All nombre_dimension"
  primaryKey="Dimension_No">
    <Table name="dimension"/>
    <Level name="dimension.jerarquia1" table="dimension" column=jerarquia1" uniqueMembers="false"/>
    <Level name="dimension.jerarquia2" table="dimension" column="jerarquia2" uniqueMembers="false"/>
  </Hierarchy>
</Dimension>
```

```
<Level name="dimension.jerarquiaN" table="dimension" column="jerarquiaN" uniqueMembers="false"/>
</Hierarchy>
</Dimension>
```

**Figura 6.4** Líneas para definir una nueva dimensión en el documento XML

- **Consultas de OLAP:** para agregar más consultas predeterminadas de OLAP se modifican los archivos de XML como se describió en la sección 5.2.1.
- **Algoritmos de data mining:** utilizar los programas desarrollados en esta tesis (*ArffGenerator*, *ReportFormatter*, *Apriori*) para generar reportes nuevos cuyos resultados puedan ser incorporados a la interfaz de PIE. Para generar archivos para otro tipo de algoritmos, es necesario identificar los requerimientos de dichos algoritmos y modificar los programas pertinentes para obtener los archivos con los formatos correctos. Es importante recordar que se deben mantener los parámetros especificados en el capítulo IV, porque de lo contrario los resultados que se obtengan de las reglas de asociación pueden no ser tan óptimos.

En la siguiente sección, mostraremos las mejoras que pueden hacerse al sistema en estudios y desarrollos futuros.

## 6.4 Recomendaciones al sistema

Algunas de las recomendaciones y mejoras observadas para el sistema son las siguientes:

- **Formalizar el modelo multidimensional y la metodología:** ya que en esta tesis sólo se trabajó con un caso de estudio práctico de prueba, que sirve como paso inicial para llegar a una generalización más estricta y definida.

- ***Automatización del proceso de actualizado y mantenimiento de los datos en el datawarehouse.*** Se podría implementar un nuevo menú en la interfaz del sistema PIE, en donde el usuario pudiera especificar la periodicidad de actualización de los datos, y la hora en la que desea llevar a cabo dicha actualización. También podrían considerarse algunos otros aspectos configurables para facilitar esta actividad, como la incorporación de funcionalidades de *Spoon* y *Pan* que permitan generar transformaciones y trabajos de manera más automática y transparente para el usuario.
- ***Agregación de dimensiones y hechos de manera dinámica:*** en esta parte el usuario podría especificar, a través de una interfaz de tipo *wizard*, los datos de su nuevo hecho o dimensión, y estos se agregarían automáticamente al *datawarehouse* y al documento XML, para que pudieran ser utilizados inmediatamente.
- ***Integración del algoritmo de clustering de data mining:*** como vimos a lo largo de la tesis, también se describió el algoritmo de clustering, que podría aplicarse a los clientes, para formar grupos de clientes comunes. Así como se agregaron los submenús con los archivos a los que se les aplicó el algoritmo de *data mining*, se podría considerar un apartado de archivos a los que se les aplicara el algoritmo de clustering y mostrar los resultados entregados por Weka en la interfaz del sistema PIE. También podríamos considerar la incorporación de otros algoritmos, según la información que quisiéramos obtener y la naturaleza de los datos que manejemos. Este algoritmo no se consideró para el prototipo actual porque presenta otras complejidades técnicas ya que se necesita identificar que tipos de grupos analizar, por ejemplo los grupos de clientes, ya que los grupos de productos, los obtenemos con las reglas de asociación. Un buen inicio para aplicar este algoritmo sería el archivo

*clientes vs. ventas* porque de ahí se pueden identificar grupos de clientes y sugerir productos para determinados grupos.

En la siguiente sección discutiremos algunos de los puntos y conclusiones obtenidas a lo largo de la tesis y los beneficios computacionales obtenidos de este estudio.

## **6.5 Discusión final**

Con esta tesis se comprueba la factibilidad de la aplicación del proceso de inteligencia empresarial, mediante la metodología propuesta, en las PyMEs, lo que hasta ahora no se ha abordado con profundidad, por cuestiones de infraestructura y recursos limitados para el proceso como se explicó en capítulos anteriores.

Se comprobó que si es factible la implementación de las técnicas de *datawarehousing*, OLAP y que no se requieren de enormes volúmenes de datos, con los que generalmente cuentan las grandes empresas, como comúnmente se piensa, para la aplicación de los algoritmos de *data mining* para que devuelvan resultados interesantes.

Se demostró también que no se requiere de personal experto en el área de inteligencia empresarial para poder tener acceso a dichos procesos y entender la información que se entrega como resultados. Esto se corroboró gracias a la participación, de la empresa *Procesamiento Avanzado de Información para Negocios S.A de C.V*, quien nos proporcionó, los datos reales de la *abarrotera Coscomatepec*, uno de sus clientes, y que hizo posible la interacción con el administrador de la misma. El administrador utilizó el sistema PIE, y después de una breve explicación, comenzó a manipular su información, encontrando datos interesantes.

En primer lugar, se percató de que su hipótesis de los productos “gancho”, explicada en la sección 4.1.5, no era del todo cierta, ya las reglas de asociación obtenidas de la aplicación del algoritmo de *data mining*, muestran interacción entre los mismos productos “gancho”, por lo que se puede pensar que los clientes han identificado que su supermercado vende a buen precio esos productos, pero no otros. Esto sirvió para que el administrador pudiera replantear nuevas estrategias.

Otro de los beneficios de este tipo de análisis es que por medio de las reglas de asociación, se puede diseñar una nueva distribución de los productos dentro de la tienda, colocando los productos que se compran juntos en estantes cercanos. Así mismo también se puede diseñar una nueva distribución para la bodega, lo que facilitará el problema de la carga de la camioneta para la venta en ruta y disminuir el tiempo que se toma realizar esta actividad.

Así mismo, los análisis de OLAP y *data mining*, pueden ayudarles a definir una política de descuentos, algo que no habían podido implementar con anterioridad, porque les parecía un proceso complicado y difícil de controlar y mantener.

Al administrador del supermercado le pareció que el sistema era muy intuitivo y amigable, y que, análisis de este tipo sobre su propia información, si podría traer beneficios a su negocio. Sin embargo también nos hizo algunos comentarios, que fueron considerados en la sección de mejoras al sistema que se mencionaron anteriormente.

La metodología y prototipo propuestos en la tesis, muestran una versión viable que puede ser aplicable no sólo a PyMEs abarroteras, sino a PyMEs comercializadoras en general. Esto lo podemos afirmar porque el modelo entidad - relación utilizado como base para nuestro modelo

multidimensional, el cual proviene del sistema SAN, es un modelo general que da servicio a muchas otras PyMEs comercializadoras que no necesariamente son abarroteras, lo que nos hace pensar que el modelo multidimensional propuesto en esta tesis también puede aplicarse a otras PyMEs comercializadoras.

La metodología tampoco se encuentra limitada al soporte en el área de CRM, como vimos en la escalabilidad del modelo podemos agregar información que apoye a la toma de decisiones en múltiples áreas como las compras y el manejo de inventarios.

Después de haber realizado todas estas actividades, se espera que el proyecto de inteligencia empresarial para el apoyo a la toma de decisiones en las PyMEs tenga continuación, que existan más personas interesadas en apoyar e impulsar el crecimiento de sistemas computacionales que involucren las técnicas de inteligencia empresarial: *datawarehousing*, OLAP y *data mining*, que beneficien al mayor número de unidades productivas que existen en México: las pequeñas y medianas empresas.