

OLAP en las PyMEs

Un modelo multidimensional común para empresas del sector yerbatero del nordeste argentino

Eduardo Zamudio, Mario R. Viale, Horacio D. Kuna

Facultad de Ciencias Exactas, Químicas, y Naturales. Universidad Nacional de Misiones.
Felix de Azara 1552. Posadas, Misiones, Argentina.
eduardo.zamudio@campus.unam.edu.ar, vicedecano@fceqyn.unam.edu.ar, hdkuna@unam.edu.ar

Resumen. El presente trabajo propone un conjunto de datamarts diseñados a partir de procesos de negocio claves para el sector de la Pequeña y Mediana Empresa (PyME) de un mercado vertical, como lo es la producción de yerba mate en la República Argentina. El desarrollo de un modelo multidimensional común de datos formado por los datamarts, se desarrolló en base a requerimientos comunes de estas organizaciones, y pretende ser de utilidad para el desarrollo de soluciones particulares en el sector productivo mencionado. Con este objeto, se ha llevado a cabo una prueba experimental sobre una organización de este tipo, y se ha documentado el proceso de adaptación del modelo propuesto al caso particular.

Keywords: modelo multidimensional, datawarehouse, datamart, OLAP, PyME

1 Introducción

El objetivo de este trabajo consiste en desarrollar un conjunto de datamarts que satisfagan requisitos comunes de pequeñas y medianas organizaciones dedicadas a la producción de yerba mate canchada¹, para el futuro desarrollo de sistemas de soporte a la toma de decisiones en el sector. Un modelo común de datos permitirá al desarrollador conocer de antemano cuál es el núcleo del problema al momento de desarrollar una solución particular.

De esta manera, el tomador de decisión tendrá un primer contacto con un entorno familiar, reconociendo los procesos comunes, y el equipo informático encargado del desarrollo comprenderá mejor las necesidades específicas de cada caso.

El modelo multidimensional de datos tiene origen en el álgebra de matrices, dando origen al concepto de cubo, el cual generaliza las hojas de cálculo a cualquier número de dimensiones [1]. El conjunto de cubos define la estructura del datawarehouse, que

¹ canchada: Nombre que se le da a al producto de la yerba mate una vez que ha pasado por el secadero.

integran los datamarts.

El modelo multidimensional está definido por un conjunto de hechos, dimensiones, y relaciones que representan los procesos de negocio que se pretenden modelar, para dar soporte a las necesidades del analista.

El diseño de una base de datos inicia con un análisis de requerimientos. Luego comienza la etapa de diseño del modelo multidimensional, el cual puede clasificarse en fases de acuerdo al nivel de detalle que representan en su diseño. De esta manera, un nivel conceptual (con distintos exponentes en [2],[3],[4],[5],[6],[7],[8],[9],[10], y [11]) pretende reflejar las ideas, independientemente de la forma de implementación. El nivel lógico tiene en cuenta el tipo de Sistema Gestor de Base de Datos (SGBD o DBMS) sobre el que se implementará el modelo. Finalmente, en el nivel físico, el modelo describe el SGBD en el cual se almacenarán los datos. Con el objetivo de compilar los trabajos en estas áreas y definir una terminología común, [12] realiza una pequeña descripción de cada uno de los métodos propuestos para el modelado multidimensional en los distintos niveles de detalle. Además agrega un nivel formal, el cual compila trabajos en el álgebra de los modelos.

Un esfuerzo por el diseño y uso de patrones para el modelado multidimensional se encuentra en [13], donde los autores proponen un conjunto de patrones de diseño dimensional (basado en un paradigma de objetos) que deben ser utilizados en conjunto con datos transaccionales y requerimientos de usuarios.

El enfoque de modelado lógico con mayor difusión utilizado en la actualidad corresponde al modelado dimensional [14]. A continuación se describen las características de dicho enfoque.

Para representar el modelado dimensional se utilizan Diagramas Entidad Relación (ERD). De esta manera, el área de presentación puede representarse a través de una base de datos relacional, recibiendo el nombre de esquema estrella, o bien, puede representarse a través de una base de datos multidimensional, y recibe el nombre de cubo.

El modelado dimensional se puede aplicar tanto a bases de datos relacionales como multidimensionales.

Una segunda cuestión a tener en cuenta es que los datos deben ser atómicos de modo que puedan servir de base a los datos sumarizados. Finalmente, los datamarts deben ser contruidos utilizando dimensiones y hechos ajustados.

El trabajo está organizado de la siguiente manera: En la sección 2 se plantea el problema a resolver, basado en el inconveniente de obtener datos de calidad en forma instantánea, para dar soporte a la toma de decisiones en organizaciones del sector productivo de la yerba mate dedicado al secado del producto. En la sección 3 se describe la solución propuesta, a partir del desarrollo de un conjunto de datamarts, los cuales integren un modelo multidimensional de datos que represente requerimientos comunes de las organizaciones mencionadas. En la sección 4 se describe un caso de prueba donde se aplicó el modelo propuesto sobre una organización. Finalmente, en la sección 5 se presentan las conclusiones del trabajo y las posibles líneas de

investigación.

2 Situación problemática

Aquí se describe la problemática de las organizaciones vinculadas a la producción de yerba mate en relación a los sistemas de información para dar soporte a la toma de decisiones. Comenzando por una descripción de las organizaciones, su estructura organizacional, las actividades que realizan, los procesos de registración de datos, como así también los roles ejercidos dentro de ellas. Finalmente, se define claramente el problema que pretende atacar este trabajo, junto con sus objetivos.

2.1 Relevamiento organizacional

El sector elegido para el estudio pertenece a los secaderos de yerba mate pertenecientes a cooperativas de la provincia de Misiones, ya que el mismo presenta condiciones que favorecen la viabilidad del trabajo y acompañan los objetivos del mismo, como ser: cantidad de establecimientos, concentración geográfica, y volumen de producción.

El relevamiento se hizo sobre una población de 5 organizaciones cooperativas, de un total de 24, ubicadas en las zonas Este, Centro, y Sur, de la provincia de Misiones, según datos extraídos de [15].

En todos los casos, el relevamiento se realizó a través de entrevistas personales a directivos y administrativos de los establecimientos en forma individual.

2.2 Actividades

Junto con el proceso de transformación primaria, las cooperativas presentan un conjunto de actividades vinculadas, la cuales se definen a través de los siguientes procesos de negocio: Ingreso de Hoja Verde, Salida de Yerba Mate Canchada, Compra, Venta, Secado, Servicio de secanza, y Servicio de cuadrilla.

2.3 El problema

Las pequeñas y medianas organizaciones dedicadas a la producción de yerba mate canchada se encuentran con graves inconvenientes al momento de obtener información que le permita realizar la toma de decisiones en base a datos de calidad. Esto se debe a que las actividades de procesamiento de datos sumariados se realiza con herramientas genéricas (como ser planillas de cálculo), que no han sido diseñadas con este propósito exclusivo. De esta manera, recae en el administrativo de las organizaciones la responsabilidad, no solamente de la gestión transaccional, sino también de un complejo proceso de búsqueda y recolección de datos (no siempre en

soportes digitales) para la generación de reportes con datos sumariados. Los sistemas de soporte a la toma de decisión representan un nuevo modo de realizar el seguimiento de la gestión para los tomadores de decisión en el área de las PyMEs dedicadas a la producción de yerba mate canchada. Las alternativas de análisis, como ser Análisis en Línea (OLAP) a partir de cubos de datos, o el uso de tableros de control (Dashboards), son asistentes indispensables en procesos modernos de soporte a la toma de decisiones.

3 Solución Propuesta

El desarrollo del modelo de datos se encuentra dividido en las fases de modelado conceptual, lógico, y físico (Fig 1). Los distintos niveles de abstracción tienen como objetivo ampliar la visión de los procesos de negocio relevados, y flexibilizar cualquier posible implementación. Esto significa que éste es un modelo propuesto que pretende representar los requerimientos comunes de las organizaciones relevadas, de modo que cualquier desarrollo particular de una solución exigirá una revisión de dicho modelo para adaptarlo a las necesidades particulares de cada caso.

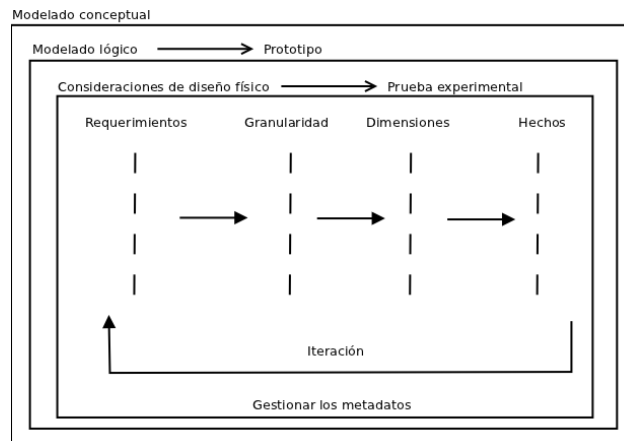


Fig 1. Fases de modelado y ciclo de vida

3.1 Requerimientos

El modelado conceptual implica la identificación de procesos de negocio, los cuales representan un aspecto funcional y cuantificable de la organización (por ej.: Ingreso de Hoja Verde). A partir de éstos, se identifican las entidades de alto nivel, que determinan las características de visualización del modelo; el conjunto de orígenes de

datos, que proporcionan los datos transaccionales; y un resumen de análisis de requerimientos, identificando las jerarquías de visualización para cada entidad, así como un conjunto de valores numéricos que constituyen las medidas del negocio.

3.2 Procesos de negocio

Los datamarts desarrollados para la solución se corresponden con los procesos de negocio: Ingreso de Hoja Verde (HV), Secado, Personal, Servicio de Cuadrilla, Servicio de Secanza, Compras, Salida de Yerba Mate Canchada (YMC), y Ventas.

3.3 Granularidad

La Tabla 1 resume la granularidad de los procesos, las dimensiones intervinientes, y una primera definición del fact table, especificando la forma en que se cargan los datos en la misma, y los hechos.

Tabla 1. Relevamiento inicial de procesos de negocio, granularidad, dimensiones, y fact tables

Proceso	Granularidad	Tipo
Ingreso de HV	talon de ingreso de HV	Transaccional
Secado	período trabajado	Acumulado
Personal	período trabajado	Transaccional
Cuadrilla	periodo trabajado por el empleado	Transaccional
Servicio de secanza	planilla de convenio	Acumulado
Compras	producto comprado	Transaccional
Salida de YMC	registro de salida de YMC en planilla	Transaccional
Venta	comprobante original	Transaccional

3.4 Dimensiones

EL conjunto de dimensiones preliminares se puede observar en la Tabla 2. En la misma se ve ampliado el conjunto de dimensiones dotando al modelo de un menor nivel de abstracción.

Tabla 2. Arquitectura de bus

Procesos / Dimensiones	Fecha	Hora	Cliente	Campaña	Descuento motivo	Empleado	Empleado mini	Forma de pago	Molino	Lote	Transportista	Origen de HV	Asistencia	Productor	Proveedor	Producto	Planta / Secadero	Sector de planta
Ingreso de HV	X			X	X	X	X		X	X	X	X		X		X	X	
Secado	X	X		X		X	X										X	
Personal	X	X		X		X	X						X				X	X
Cuadrilla	X	X		X		X	X							X				
Secanza	X			X					X									
Compras	X			X				X							X	X		
Salida de YMC	X		X	X	X	X	X		X	X	X			X		X		
Venta	X		X	X				X								X		

3.5 Medidas

Para poder identificar claramente los hechos de un datamart, los mismos deben definirse con precisión. La Tabla 3 realiza una descripción de cada uno de estos elementos junto con una clasificación respecto del tipo de aditividad que utilizan. En el caso de aquellos hechos definidos por fórmulas, las mismas se definen en una columna.

Tabla 3. Medidas (para los procesos Ingreso de Hoja Verde, y Compras)

Proceso	Hecho	Descripción	Tipo	Formula
Ingreso de HV	bruto	Peso en kg del camión con la carga	no aditivo	
	tara	Peso en kg del camión sin la carga, o vacío	no aditivo	
	neto	Diferencia entre el bruto y la suma de la tara y el descuento	aditivo	(bruto) – (tara + descuento)
	descuento	Peso en kg del material rechazado	aditivo	
	precio	precio en pesos argentinos por kg de HV	no aditivo	
	importe total	Producto del neto y el precio	aditivo	(neto) * (precio)
Compras	cantidad	Cantidad del producto comprado	aditivo	
	flete	importe por el flete	semiaditivo	
	importe	Importe por el producto comprado	aditivo	

4 Prototipo

El desarrollo de un prototipo exige un nivel de abstracción menor. Con este fin, nos aproximamos a un enfoque de modelado lógico, en el cual debemos definir la arquitectura sobre la que se implementará el datawarehouse, sin que esto suponga la necesidad de describir aspectos físicos del mismo.

La arquitectura del datawarehouse para el desarrollo de este prototipo utiliza el paradigma relacional. Esta decisión está fundada en que varias de las herramientas actuales más populares, tanto comerciales como open source, de uso en el desarrollo de soluciones de business intelligence utilizan este paradigma.

4.1 Proceso de adaptación del modelo al caso particular

La organización elegida para el desarrollo de la prueba experimental es una cooperativa cuya actividad principal es la transformación primaria dentro del ciclo de producción de yerba mate, esto es, el secado de hoja verde de yerba mate.

El proceso de adaptación (Fig 2) comienza con el relevamiento de los requerimientos del caso particular. El conjunto de requerimientos comunes predefinidos ayudará a identificar aquellos que ya han sido relevados, y a enfocar el proyecto a la priorización de los mismos, así como la identificación de los orígenes de datos requeridos, y la ampliación del modelo en base a los requerimientos particulares obtenidos del relevamiento.

En cualquier caso, el desarrollo es un proceso iterativo que permitirá refinar la solución hasta cumplir con los objetivos y así obtener la aceptación del usuario.

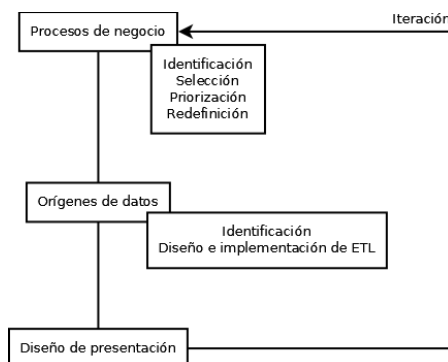


Fig 2. Proceso de adaptación al caso particular

4.1.1 Procesos de negocio

Con objeto de probar la adaptabilidad del modelo propuesto en un escenario real, se optó por utilizar el grupo predefinido de procesos, y determinar la selección y priorización para el desarrollo de los mismos a partir de la definición del conjunto de factores de evaluación [16].

4.1.2 Selección y priorización

En la Tabla 4, se refleja el cruzamiento de los factores de evaluación con los procesos de negocio, cuyas ponderaciones fueron determinadas junto con los entrevistados durante el relevamiento.

Los procesos con mayor ponderación acumulada serán los seleccionados para continuar el proceso de implementación. Por otra parte, **la digitalización de los datos** representa una restricción **imprescindible** para el cumplimiento de los objetivos de este trabajo. Por este motivo, se ha definido un corte, estableciendo como candidatos aquellos procesos cuyos orígenes de datos cumplan dicha condición (indicados con un * en la Tabla 4).

Tabla 4. Priorización de procesos.

Proceso	Disponibilidad de datos	Calidad de datos	Complejidad de datos	Complejidad	Significancia estratégica	Puntos
(*)Ingreso de HV	Alta (3)	Media (2)	Bajo (3)	Media (2)	Alto (6)	16
(*)Compras	Alta (3)	Media (2)	Bajo (3)	Baja (3)	Media (4)	15

4.1.3 Orígenes de datos

Las datos transaccionales de la organización son gestionados principalmente por un sistema DOS, con archivos planos, con registros de longitud fija, como soporte de almacenamiento. Además se utilizan un conjunto de planillas de cálculo, y es amplio el uso del papel como soporte de registración, sin presentar éstos un formato digitalizado.

4.1.4 Granularidad

La granularidad de los procesos seleccionados se corresponden con la descripción de granularidad de los procesos del modelo propuesto tanto en el caso de los procesos, como así también para las dimensiones.

4.1.5 Dimensiones

No se disponen de datos digitalizados para la totalidad de las dimensiones. Por este motivo se han definido las dimensiones (Tabla 5) que fueron efectivamente cargadas con datos reales de la organización.

Tabla 5. Dimensiones

Procesos / Dimensiones	Fecha	Campaña	Descuento motivo	Empleado	Empleado mini	Formade pago	Molino	Transportista	Origen de HV	Productor	Proveedor	Producto	Planta / Secadero
Ingreso de HV	X	X	X	X	X		X	X	X	X		X	X
Compras	X	X				X					X	X	

5 Conclusiones

El presente trabajo ha conseguido establecer, a partir de un modelo de datos multidimensional común, las bases en la investigación de sistemas orientados a la toma de decisiones de un mercado vertical, compuesto por organizaciones PyMEs, como son las cooperativas dedicadas a la transformación primaria de la yerba mate en la provincia de Misiones. Esto fue posible gracias a la identificación de requisitos comunes obtenidos de dichas organizaciones, permitiendo el diseño y desarrollo experimental de un prototipo para una organización local.

El proceso de desarrollo de dicho modelo ha permitido realizar una primera identificación de las necesidades recurrentes en materia de información resumida, como así también el gran déficit en relación a la informatización de los procesos transaccionales.

El verdadero éxito del trabajo será puesto a prueba cuando los conocimientos aquí descriptos sean tomados como marco referencial al momento de desarrollar una solución particular en alguna de las organizaciones previamente mencionadas. Permitiendo, de esta manera, modificar cualquiera de las estructuras propuestas para adaptarlas a las necesidades del caso particular.

5.1 Futuras líneas de investigación

El modelo multidimensional de datos basado en requerimientos comunes ha sido el inicio de una línea no explorada de investigación en relación a sistemas de información en el mercado yerbatero. Por este motivo, trabajos futuros podrían enfocarse en el refinamiento y evolución de dicho modelo.

La falta de informatización de los procesos en organizaciones del sector, es un hecho que no favorece el desarrollo de sistemas orientados a dar soporte a la toma de decisiones. Por este motivo, y con objeto de reducir los mayores esfuerzos de desarrollo, como son los procesos de Extracción, Transformación, y Carga (ETL), podría desarrollarse un sistema de gestión open source que tenga en cuenta los requerimientos aquí obtenidos.

Por otra parte, el mismo enfoque podría aplicarse a mercados verticales de características similares, particularmente hablando de PyMEs, mercados concentrados geográficamente, procesos simples y repetitivos, y grandes volúmenes de producción.

Bibliografía

1. PEDERSEN, T.B., JENSEN, C.S.: Multidimensional database technology. Computer. Vol 34 Issue12. Dic 2001. IEEE (2002)
2. CABIBBO, L., TORLONE, R.: A Logical Approach to Multidimensional Databases. Advances in Database Technology — EDBT'98. LNCS Vol 1377, 183. Springer (1998)
3. GOLFARELLI, M., MAIO, D., RIZZI, S.: Conceptual design of data warehouses from E/R schemes. System Sciences, 1998., Proceedings of the

- Thirty-First Hawaii International Conference on. Vol 7. IEEE (1998)
4. HÜSEMANN, B., LECHTENBÖRGER J., VOSSEN, G.: Conceptual Data Warehouse Design. In Proc. of the International Workshop on Design and Management of Data Warehouses (DMDW 2000). CiteSeer (2000)
 5. LEHNER, W.: Modeling large scale OLAP scenarios. Advances in Database Technology—EDBT'98. LNCS Vol 1377. 153. Springer (1998)
 6. MEDINA, E., TRUJILLO, J.: A standard for representing multidimensional properties: The common warehouse metamodel (CWM). Advances in Databases and Information Systems. LNCS Volume 2435/2002. Springer (2002)
 7. RIZZI, S., et al.: Research in data warehouse modeling and design: dead or alive?. DOLAP '06 Proceedings of the 9th ACM international workshop on Data warehousing and OLAP. ACM (2006)
 8. SAPIA, C. et al.: Extending the E/R model for the multidimensional paradigm. ER '98 Proceedings of the Workshops on Data Warehousing and Data Mining: Advances in Database Technologies. ACM (1998)
 9. TSOIS, A., KARAYANNIDIS, N., SELLIS, T.: MAC: Conceptual data modeling for OLAP. 3rd International Workshop on Design and Management of Data Warehouses (DMDW 2001). Citeseer (2001)
 10. TRUJILLO, J., PALOMAR, M.: An object oriented approach to multidimensional database conceptual modeling (OOMD). DOLAP '98 Proceedings of the 1st ACM international workshop on Data warehousing and OLAP. ACM (1998)
 11. TRYFONA, N., BUSBORG, F., BORCH CHRISTIANSEN J.G.: starER: a conceptual model for data warehouse design. DOLAP '99 Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP. ACM (1999)
 12. ABELLO, A., SAMOS, J., SALTOR, F.: A Data Warehouse Multidimensional Data Models Classification. Technical Report LSI-2000-6. Dept. Lenguajes y Sistemas Informáticos. Universidad de Granada (2000)
 13. JONES, M.E., SONG, I.: Dimensional modeling: Identification, classification, and evaluation of patterns. Decision Support Systems. Volume 45, Issue 1. Elsevier (2008)
 14. KIMBALL, Ralph y ROSS, Margy: The data warehouse toolkit: the complete guide to dimensional modelling. 2ed. pp. 421. Wiley Computer Publishing. ISBN: 0-471-20024-7 (2002)
 15. Instituto Nacional de la Yerba Mate, <http://www.inym.org.ar/>
 16. BALLARD, C., et al.: Dimensional Modeling: In a Business Intelligence Environment. pp. 664. IBM RedBooks. ISBN: 0738496448 (2006)