

Enfoque MDA para el diseño de un data warehouse difuso

A model driven approach for fuzzy data warehouse design

Carolina Zambrano Matamala¹ Marcela Varas Contreras² Angélica Urrutia Sepúlveda³

Recibido 4 de agosto de 2011, aceptado 9 de abril de 2012

Received: August 4, 2011 Accepted: April 9, 2012

RESUMEN

Un data warehouse (DW) es un repositorio de datos que provienen de distintas fuentes. Es un sistema para el análisis de datos usado para el apoyo a la toma de decisiones. En el ámbito de análisis de datos es usual encontrar relaciones entre los datos de naturaleza difusa. Por ejemplo, en un contexto académico la respuesta a la consulta “qué alumnos obtuvieron buena nota final” no puede ser obtenida desde un DW tradicional pues no maneja información cualitativa. Debido a las limitaciones de los DW tradicionales es que se pueden extender usando lógica difusa a un DW Difusos (DWD). Por otro lado, en el ciclo de desarrollo de DW se han considerado diversas metodologías y modelos que tienen como desventaja que no integran los distintos niveles de abstracción ni el proceso de diseño de DW. Sin embargo, el enfoque MDA es una alternativa de integración que otorga un marco de desarrollo integrado basado en la arquitectura de modelos y transformaciones entre distintos niveles de abstracción, que la hace adecuada para este tema. En este artículo se presenta un enfoque MDA para el diseño de un DWD que considera medidas difusas, relaciones difusas y niveles difusos, para lo cual se proponen un metamodelo CWM OLAP Difuso, reglas QVT y un ejemplo. También se presenta la secuencia metodológica para la aplicación de las reglas QVT. Finalmente, se proporciona un caso de estudio de análisis en el ámbito educacional.

Palabras clave: Data warehouse, data warehouse difuso, arquitectura dirigida por modelos, lógica difusa.

ABSTRACT

A Data Warehouse (DW) is a repository of data from different sources. Is a system for data analysis used to support decision making. In the area of data analysis is common to find relationships between data from fuzzy nature. In this context, the questions give me the marks quantity of good, average and bad cannot be obtained from the traditional DW because do not handle as qualitative information. Due to the limitations of traditional DW is that it can be extended to Fuzzy DW (DWD) using fuzzy logic. On the other hand in the DW development life cycle are considered different methodologies and models that have the disadvantage that do not integrate the different levels of abstraction or the DW design process. However, MDA is an alternative approach to integration for the development of DW as it gives a comprehensive development framework based on the architecture of models and transformations between different levels of abstraction. This article presents a Fuzzy MDA approach to design fuzzy measures, fuzzy relations and fuzzy levels.

For which we propose a Fuzzy OLAP CWM metamodel, QVT rules and an example of rule. Is the sequence methodology for applying the rules QVT. Finally, it provides a case study analysis in the field of education.

Keywords: Data warehouse, fuzzy data warehouse, model driven architecture, fuzzy logic.

¹ Departamento de Ingeniería Informática y Ciencias de la Computación. Universidad de Atacama. Avenida Copayapu 485. Copiapó, Chile. E-mail: carolinazambrano@gmail.com

² Departamento de Ingeniería Informática y Ciencias de la Computación. Universidad de Concepción. Edmundo Larenas 215. Concepción, Chile. E-mail: mvaras@udec.cl

³ Departamento de Computación e Informática. Universidad Católica del Maule. Avenida San Miguel 3605. Talca, Chile. E-mail: aurrutia@spock.ucm.cl

INTRODUCCIÓN

Hoy en día las organizaciones cuentan con más información. Para analizar esta gran cantidad de información una alternativa es el uso de DW [13]. Un DW es un repositorio de datos que provienen de distintas fuentes. Los DW procesan esta información, que se puede utilizar para realizar análisis de datos a nivel estratégico y apoyar la toma de decisiones en una organización [15]. Los elementos principales de un DW son dimensiones, jerarquías, hechos y medidas. Una dimensión es un concepto abstracto que modela un contexto para el análisis. Sobre las dimensiones se pueden definir jerarquías, que permiten acceder a los datos a diferentes niveles o categorías de detalle [4]. Los hechos representan una tupla de claves primarias, foráneas y medidas. Las medidas son numéricas, por lo cual el tipo de análisis que se puede realizar es cuantitativo. Sin embargo, en el proceso de toma de decisiones es valorado el análisis cualitativo. Por ello es que el uso de la lógica difusa [34] nace como una interesante propuesta que permite extender un DW. Por ejemplo, la consulta entregue la cantidad de notas malas, regulares y buenas no puede ser obtenida desde un DW tradicional o sería un trabajo adicional para los usuarios realizar el proceso manualmente.

Al extender un DW a un DWD se permite dar un significado lingüístico a la información numérica. Por otro lado, en las jerarquías es bastante común que un elemento (instancia) no sea categorizable bajo una sola categoría sino que pertenezca a varias. En el ejemplo de la Figura 1 se observa una relación difusa entre Alumno y Nivel Avance. Luego en la Figura 2 se observa que una instancia de Alumno pertenece a más de un Nivel de Avance (relación difusa). Esta última relación es heterogénea [12] y no estricta [18].

Por todo lo anterior se llega a un problema interesante y motivante. Ya que por un lado se trata la relación difusa entre niveles y por otro lado la etiquetación permite clasificar de acuerdo a parámetros de una función de pertenencia como se muestra en el ejemplo de la Figura 3. Allí se ha etiquetado el atributo Nota de la tabla de hechos que se muestra en la Tabla 1, usando etiqueta buena, mala y regular. Por otra parte, el caso de las Bases de Datos Difusas (FDB) ha sido ampliamente estudiado, por ejemplo, en [9-10, 31-33].

En el ámbito de DW tradicionales existen variados enfoques de diseño y modelos tales como [2, 7, 11, 29, 30, 36]. También se ha aplicado el enfoque MDA a DW que puede ser revisado en [17]. Operaciones de DW se pueden revisar en [18]. Un enfoque lógico para el diseño multidimensional se puede encontrar en [1].

En el área de DWD se ha propuesto un modelo formal multidimensional difuso [5, 20]. Desde la perspectiva de diseño de DWD se encuentran los siguientes trabajos [3, 6, 28], pero ninguno de ellos presenta avances en el uso de Model Driven Architecture (MDA) para el diseño de un DWD. En el caso de transformación de modelos el trabajo [3] presenta algunos avances en diseño conceptual a diseño lógico de DWD usando MDA pero sólo para el caso de medidas difusas. Por lo cual la propuesta que se presenta en este artículo tiene la ventaja de considerar el enfoque MDA como la arquitectura que permite integrar todas las etapas de desarrollo de un DWD, por esto se propone una arquitectura MDA Difusa.

En esta investigación se define DWD como: “Un DW que permite almacenar y operar medidas difusas tipo 1 y tipo 2, relaciones difusas entre niveles y niveles difusos tipo 2”. Esto se explica en la sección Preliminares.

El artículo tiene la siguiente estructura: primero se presenta una sección de preliminares para explicar los conceptos de atributos difusos y la difusidad en los distintos aspectos del DWD que se proponen. Luego se presenta una comparación entre DWD versus DW Tradicional. Posteriormente se define el enfoque MDA para el Diseño de un DWD, seguido de la metodología de aplicación de las transformaciones con su secuencia lógica de aplicación. Posteriormente se presenta el caso de estudio. Finalmente se exponen la Conclusión y Trabajos Futuros que incluyen comentarios sobre los resultados y posibles extensiones futuras al trabajo presentado.

PRELIMINARES

Para comenzar es necesario comprender qué es un atributo difuso. En este contexto una clasificación de información difusa es presentada en [31] donde se definen los atributos difusos como atributos tipo 1, tipo 2 y tipo 3.

Los atributos difusos **tipo 1** [31] son atributos clásicos que admiten el tratamiento impreciso, donde las etiquetas lingüísticas definidas sólo se usarán en las condiciones difusas de las consultas. Los atributos difusos **tipo 2** [31] son atributos que admiten tanto datos clásicos (*crisp*) como difusos (imprecisos), en forma de distribuciones de posibilidad sobre un dominio subyacente ordenado. El tipo 2 permite también la representación de información incompleta en forma de datos de tipo Unknown, Undefined y Null. Los atributos difusos **tipo 3** [31] son atributos sobre datos de dominio discreto sobre dominio subyacente no ordenado con analogía. En estos atributos se pueden definir conjuntos de etiquetas con una relación de similitud o proximidad definida sobre ellas. Además, en el tipo 3 se pueden admitir distribuciones de posibilidad sobre el dominio. A continuación se explica la difusidad en el DWD propuesto.

Difusidad entre niveles (Relación Difusa): Cuando una relación entre niveles es difusa, cada instancia de nivel inferior estará asociada a más de una instancia del nivel superior, generando una relación muchos a muchos, entre el nivel padre y el nivel hijo. Esta jerarquía es no estricta. La Figura 1 muestra un ejemplo de esquema de dimensión difusa donde se ha definido la relación entre Alumno y Nivel Avance como difusa representada por una línea punteada. Que la relación sea difusa implica que hay asociado un grado de pertenencia entre las instancias de dimensión que participan en dicha relación. Por ejemplo, un Alumno puede ser asociado a más de un Nivel Avance como se muestra en la Figura 2.

Difusidad en las medidas: Cuando una medida es difusa en esta propuesta se considera que puede ser difusa tipo 1 o difusa tipo 2. Si es medida difusa tipo 1 sólo se le agrega una etiqueta a la medida. Si es medida difusa tipo 2 significa que se ha asociado un concepto difuso que tiene asociadas etiquetas lingüísticas que se representan por funciones de pertenencia. Por ejemplo, en la Figura 1 la medida NotasT2 será difusa tipo 2 y su concepto difuso será calidad de las notas donde se definen tres etiquetas para clasificar las Notas en mala, regular o buena como se muestra en la Figura 3. El referencial ordenado que se ocupa es Nota como se muestra en la tabla de hechos de la Tabla 1.

Difusidad de un nivel: Para el caso de nivel difuso tipo 2 se presenta en la Figura 1 el nivel Región_Distancia que ha sido marcado como difuso tipo 2 mediante la representación de un círculo punteado. En el caso del nivel difuso tipo 2 también se agrega un concepto difuso que tiene funciones de pertenencia que contendrán las etiquetas. Por ejemplo en el caso de Región_Distancia el concepto difuso puede ser la localización donde las etiquetas serian cerca y lejos (cada una con un grado de pertenencia dependiendo de la distancia a algún lugar).

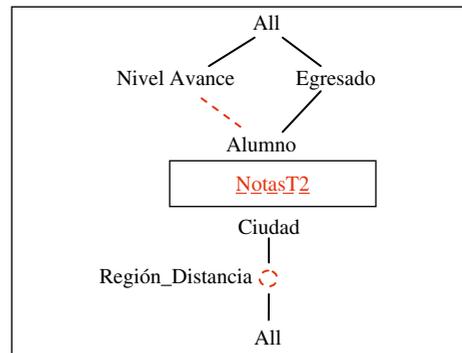


Figura 1. Esquema de DWD con relación difusa, nivel difuso tipo 2 y medida difusa tipo 2.

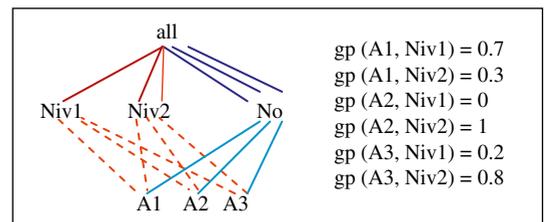


Figura 2. Extracto de instancia de dimensión para la relación difusa.

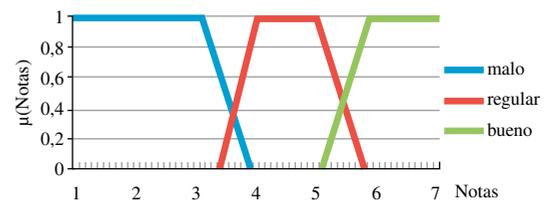


Figura 3. Funciones de pertenencia trapezoidales para el concepto calidad de las notas y las etiquetas lingüística mala, regular y buena definidos sobre el atributo notas.

Tabla 1. Tabla de hechos para el esquema de la Figura 1 con dimensiones base alumno y ciudad.

Alumno	Ciudad	Prom. Nota
A1	C1	6.1
A2	C2	5.2
A4	C1	4.8
A5	C3	4.2
A6	C4	4.9
A7	C5	5.3

Por otro lado, es necesario comprender el enfoque MDA. MDA [8, 19, 21] es una propuesta de la Object Management Group (OMG) [24], que es un consorcio a nivel internacional sin fines de lucro, cuyo objetivo principal es potenciar el desarrollo de aplicaciones orientadas a objetos, reduciendo el costo y la complejidad de los desarrollos. Para cumplir su objetivo define estándares abiertos. En este contexto el enfoque MDA permite especificar sistemas con modelos independientes de la plataforma, separando la especificación de la implementación y de la funcionalidad en una plataforma específica, para lo cual define tres capas, CIM, PIM y PSM, donde cada una refleja un nivel de abstracción que se explica a continuación:

CIM (computation independent model): Modelo que tiene relación con el negocio, donde se representa una vista del sistema. Tiene que ver con los requerimientos [8, 19, 21].

PIM (platform independent model): Modelo de un subsistema que no contiene información específica de la plataforma o tecnología específica que se utiliza, posee un alto nivel de abstracción. Representa la estructura, funcionalidad y restricciones del sistema independientemente de la plataforma tecnológica sobre la que se va a implementar [8, 19, 21].

PSM (platform specific model): Modelo de un subsistema que incluye información sobre la tecnología específica que se utiliza; este modelo complementa al PIM especificando cómo toma forma el sistema al ser implementado en una plataforma determinada [8, 19, 21].

El enfoque MDA permite realizar transformación de modelos a distintos niveles de abstracción que son transformaciones verticales, como CIM a PIM, PIM a PSM, y también al mismo nivel de abstracción que son transformaciones horizontales, como por ejemplo CIM a CIM.

En esta propuesta se realizará una transformación desde PIM DIFUSO a PSM DIFUSO como se explica en la sección MDA para el diseño de un DWD. A continuación el siguiente apartado muestra una tabla comparativa entre DW Tradicional y DWD a modo de enfatizar las ventajas del DWD.

DW TRADICIONAL VERSUS DWD

Como se indicó anteriormente, un DWD se define para efectos del presente trabajo como: “Un DW que permite almacenar y operar medidas difusas *tipo 1* y *tipo 2*, relaciones difusas *tipo 3* entre niveles y niveles difusos *tipo 2*”.

Un DWD permite entregar información cualitativa y cuantitativa en el proceso de análisis, esto es muy importante, dado que en el proceso de toma de decisiones es bien valorado poder obtener información clasificada como lo permite el DWD. La Tabla 2 muestra un resumen comparativo.

Tabla 2. Comparación de un data warehouse tradicional (DW) versus el enfoque de data warehouse difuso (DWD) propuesto.

Característica	Data Warehouse Tradicional (DW)	Data Warehouse Difuso Propuesto (DWD)
Instancias	Una instancia no puede pertenecer a más de una categoría.	Permite que una instancia pertenezca a más de una categoría debido a que se extiende la función de agregación a agregación difusa.
Tipo de información a interpretar	Sólo permite interpretación cuantitativa de los hechos.	Además de la interpretación cuantitativa, permite interpretar información cualitativa, permitiendo clasificar la información.
Conceptos lingüísticos	Un DW no permite incluir conceptos lingüísticos para la interpretación de los datos.	Permite el uso de etiquetas lingüísticas como se mostró en el ejemplo de la Figura 3 donde se etiquetó el atributo total ventas.
Tipo de datos	Sólo permite datos <i>crisp</i> para el proceso de análisis y toma de decisiones.	Permite datos <i>crisp</i> y difusos <i>tipo 1</i> , <i>tipo 2</i> y <i>tipo 3</i> .
Esquemas	El esquema de un DW consta de dimensiones, niveles, hechos, medidas y jerarquías.	Se pueden modelar las características difusas.

TRABAJOS RELACIONADOS

Actualmente en el área de DWD se han presentado trabajos orientados a la formalización, el diseño y la implementación.

En el aspecto de formalización se encuentra el trabajo de Molina [5, 20] que propone un modelo formal multidimensional difuso. El modelo permite información imprecisa en hechos y jerarquías de una dimensión, donde se permite el uso de etiquetas lingüísticas. Cuando una dimensión es difusa se permite que los nombres o etiquetas de los niveles representen conjuntos difusos, de esta forma se asocia un grado de pertenencia a cada uno de los conjuntos que representan la etiqueta. Para el caso de hecho difuso se incluye un valor que representa el grado de certeza de dicho hecho. Este enfoque es el mismo utilizado por Laurent en [16]. Para implementar el modelo formal multidimensional difuso definen una herramienta propia basada en arquitectura cliente servidor. Esta es una desventaja del enfoque pues al ser una arquitectura independiente de un sistema de DW tradicional sólo funcionará si se tiene acceso a la plataforma diseñada por los autores de [20]. Además este trabajo no presenta ningún aspecto de diseño de DWD.

El trabajo de Sapir [28] presenta una metodología para diseñar un DWD, donde se considera como metodología base para la extensión al caso difuso la metodología de Kimball [15]. La desventaja de este trabajo es que cada etapa de diseño no es integrada, por lo cual el uso de MDA como proponemos en esta investigación es una ventaja.

En [6] Fasel introduce un modelo para DWD basado en estructuras de metatablas para el caso de dimensiones con atributos difusos y medidas difusas, además de un caso de estudio para representar el uso de su propuesta. Este trabajo sólo se enfoca en la parte de diseño e implementación del DWD y por cada atributo difuso que se desee implementar genera una estructura de metatabla al igual que el caso de la propuesta de Sapir [28]. Sin embargo, el trabajo de Fasel no presenta ningún aspecto de diseño conceptual.

El trabajo de Carrera [3] tiene como objetivo transformar un esquema multidimensional difuso desde el nivel conceptual al nivel lógico, para el que

se define un conjunto de reglas basadas en QVT. El enfoque de la propuesta está basado en MDA, para lo cual transforma un PIM en un PSM. Sin embargo, esta propuesta sólo considera atributos difusos en las medidas. Nuestra propuesta considera medidas, relaciones y niveles difusos, con los cuales se genera un metamodelo completo para DWD, se propone la arquitectura MDA para diseño y desarrollo de DWD, aporta conceptualmente y además se prueba con un caso de estudio real analizado en el ámbito estudiantil. Por lo tanto, nuestra propuesta es completa en relación a la propuesta de [3] que sólo trabajó con medidas difusas. En comparación a las propuestas de [6, 28] nuestra propuesta presenta la ventaja de integración dada por el enfoque MDA.

MDA PARA EL DISEÑO DE DWD

En este apartado se describe la propuesta con enfoque MDA para el diseño de un DWD, para lo cual primero se explica la arquitectura MDA Difusa, luego la extensión de las metaclases de CWM con nuevos estereotipos con el fin de dar soporte al DWD. Luego el PSM Difuso y las reglas QVT.

La Figura 4 muestra la arquitectura MDA Difusa. En esta se muestran los tres niveles de abstracción, CIM DIFUSO, PIM DIFUSO y PSM DIFUSO. Una transformación de modelos T (círculo con una T) es el proceso de convertir un modelo en otro. Una transformación está conformada por un conjunto de reglas que describen cómo el modelo en el lenguaje fuente puede ser transformado en un modelo en el lenguaje de destino. A continuación se explican los niveles de abstracción relacionados con el ciclo de desarrollo de los DWD:

CIM DIFUSO: En esta capa se generará un modelo que permita analizar qué requerimientos de negocios pueden ser favorecidos con un análisis difuso y cualitativo.

PIM DIFUSO: En esta capa se realiza la extensión del metamodelo CWM OLAP [25] a CWM OLAP Difuso, definiendo un conjunto de metaclases de CWM con nuevos estereotipos con el fin de dar soporte al DWD.

PSM DIFUSO: Para el enfoque propuesto se utiliza el Metamodelo CWM Relacional [25] con metatablas para soportar los elementos difusos. Sólo se ha desarrollado un PSM Difuso. Como trabajo futuro se estudiarán más PSM.

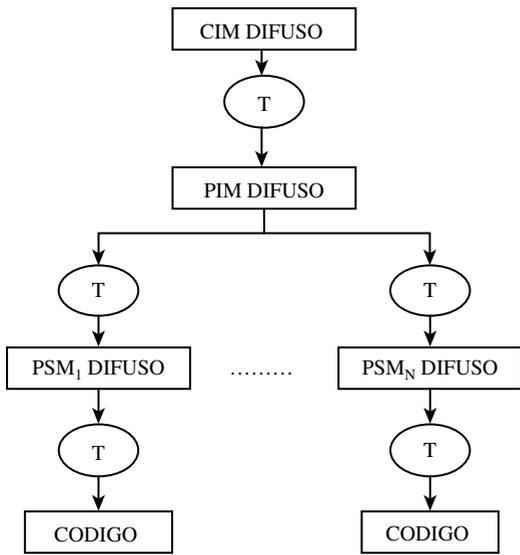


Figura 4. Arquitectura MDA difusa.

Las etapas del enfoque MDA Difuso desarrolladas en el artículo son PIM DIFUSO a PSM DIFUSO.

Extensión de las metaclasses de CWM con nuevos estereotipos para dar soporte al DWD

Los estereotipos definidos para las medidas etiquetadas difusas se desprenden de la clase *Measure* del Metamodelo CWM OLAP (ver Figura 5). Los estereotipos definidos para los niveles etiquetados difusos *tipo 2* se desprenden de la clase *Level* del Metamodelo CWM OLAP (ver Figura 6). Por último, los estereotipos definidos para relaciones difusas *tipo 3* entre niveles se desprenden de la clase principal del Metamodelo CWM OLAP (ver Figura 7). Son:

Estereotipo «*FuzzyMeasureT1*»: Los atributos de este estereotipo representan si una medida es difusa *tipo 1*. Se extiende a partir de la Clase Base *Measure* y el icono que lo representa es *FMT1*.

Estereotipo «*Label*»: Los atributos de este estereotipo representan la metadata para representar un atributo difuso *tipo 1* en una medida difusa *tipo 1*. Se extiende a partir de la Clase Base *FuzzyMeasureT1* y el icono que lo representa es *LT1*.

Estereotipo «*FuzzyMeasureT2*»: Los atributos de este estereotipo representan si una medida ha sido etiquetada como difusa *tipo 2*. Se extiende a partir de la Clase Base *Measure* y el icono que lo representa es *FMT2*.

Estereotipo «*ConceptT2*»: Los atributos de este estereotipo representan la metadata para representar un concepto difuso *tipo 2* en una medida. Se extiende a partir de la Clase *FuzzyMeasureT2* y el icono que lo representa es *CT2*.

Estereotipo «*MemberFunction*»: Los atributos de este estereotipo representan la metadata para representar una función de membresía cuando se ha etiquetado una medida como difusa *tipo 2*. Se extiende a partir de la Clase *FuzzyMeasureT2* y el icono que lo representa es *MF*.

Estereotipo «*MetaDegree*»: Los atributos de este estereotipo representan la metadata para representar el grado de pertenencia de un nivel difuso *tipo 2* y/o medida difusa *tipo 2* a las distintas funciones de membresía de un concepto difuso.

Estereotipo «*FuzzyHierarchyLevelAssociationT3*»: Los atributos de este estereotipo representan si existe una relación difusa *tipo 3* entre niveles de una jerarquía. Se extiende a partir de la Clase Base principal (from Core) y el icono que lo representa es *FHLA*.

Estereotipo «*ConceptT3*»: Los atributos de este estereotipo representan la metadata para representar una relación difusa *tipo 3* entre niveles de una jerarquía. Se extiende a partir de la Clase Base *FuzzyHierarchyLevelAssociation* y el icono que lo representa es *CT3*. Esta metadata representa la tabla puente que se genera cuando existe una relación difusa entre niveles que también sirve para relaciones no estrictas.

Estereotipo «*FuzzyLevelT2*»: Los atributos de este estereotipo representan si un nivel ha sido etiquetado como difuso *tipo 2*. Se extiende a partir de la Clase Base *Level* y el icono que lo representa es *FuzzyLevelT2*.

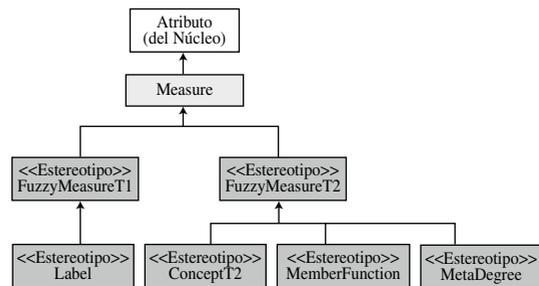


Figura 5. Extracto extensión metamodelo CWM OLAP, caso medida difusa.

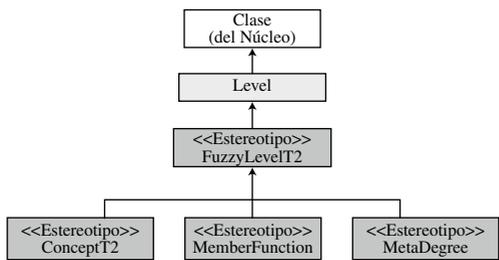


Figura 6. Extracto extensión metamodelo CWM OLAP caso nivel difuso tipo 2 de una jerarquía.

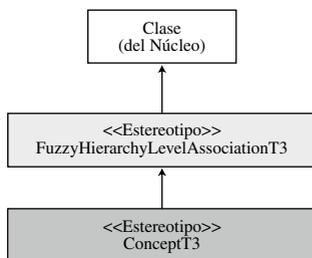


Figura 7. Extracto extensión metamodelo CWM OLAP, caso relación difusa entre niveles de una jerarquía.

Definición del Metamodelo CWM OLAP DIFUSO (PIM DIFUSO)

CWM [25] se compone de muchos metamodelos; en este artículo se trabajará con CWM OLAP, cuya extensión a CWM OLAP Difuso (PIM DIFUSO) define un conjunto de metaclasses de CWM con nuevos estereotipos con el fin de dar soporte al DWD, valores etiquetados y restricciones. En la Figura 8 se pueden apreciar el Metamodelo CWM OLAP Difuso propuesto, las clases propuestas son *FuzzyMeasureT1*, *Label*, *FuzzyMeasureT2*, *MetaDegree*, *ConceptT2*, *ConceptT3*, *MemberFunction*, *FuzzyLevelT2*, *FuzzyHierarchyLevelAssociation*.

Metamodelo de salida a las transformaciones (PSM DIFUSO)

Como se indicó, CWM [25] se compone de muchos metamodelos. Como PSM DIFUSO para el enfoque propuesto se utiliza el Metamodelo CWM Relacional con metatablas para soportar los elementos difusos. Estas son:

- *Meta_ConceptT2*: Esta metatabla permite almacenar los conceptos cuando una medida o un nivel ha sido etiquetado como difuso tipo 2. Se permite que un concepto pueda pertenecer a más de una función de membresía.

- *Meta_MemberFunction*: Esta metatabla permite almacenar los parámetros de la función de membresía, su etiqueta y el concepto que representa dicha función de membresía de un nivel o medida etiquetada como difusa tipo 2.
- *Meta_Degree*: Esta metatabla permite almacenar el grado de pertenencia a las distintas funciones de membresía de una medida difusa tipo 2 o un nivel difuso tipo 2. Además guarda el nombre de la medida y/o nivel que ha sido marcado como difuso tipo 2.
- *Meta_ConceptT3*: Esta metatabla permite almacenar el grado de pertenencia entre las instancias que participan en la relación difusa entre los niveles. Se puede decir que esta tabla es una tabla puente entre los niveles que tienen una relación difusa, por lo cual se sigue la misma solución que para el caso de una jerarquía no estricta.
- *Meta_Label*: Esta metatabla permite almacenar el nombre de la etiqueta que se ha asignado a una medida tipo 1. Sólo necesita el nombre de la etiqueta y la clave foránea de la tabla de hechos desde la cual proviene.

Definición de reglas de transformación QVT PIM DIFUSO a PSM DIFUSO

El PIM DIFUSO está representado por el Metamodelo CWM OLAP Difuso propuesto (ver Figura 8) y el PSM DIFUSO por el Metamodelo CWM Relacional con las metatablas antes descritas. Es importante indicar que las reglas que generan las estructuras para soportar los elementos difusos tienen funciones OCL [22] para restringir por ejemplo que su creación se realice una sola vez. A continuación se muestra una breve descripción de cada regla de transformación:

- *RMain*: Todo esquema del Metamodelo CWM OLAP Difuso se transforma en esquema en el metamodelo CWM Relacional.
- *RCube_TO_FactTable*: Todo cubo del Metamodelo CWM OLAP Difuso se transforma en una tabla en el Metamodelo CWM Relacional.
- *RMeasure_TO_Column*: Toda medida del metamodelo CWM OLAP Difuso se transforma en un atributo en el Metamodelo CWM Relacional.
- *RLabel_TO_Meta_Label*: La primera medida etiquetada *tipo 1* del Metamodelo CWM OLAP Difuso permite transformar un *Label* del Metamodelo CWM OLAP Difuso en tabla del Metamodelo CWM Relacional.

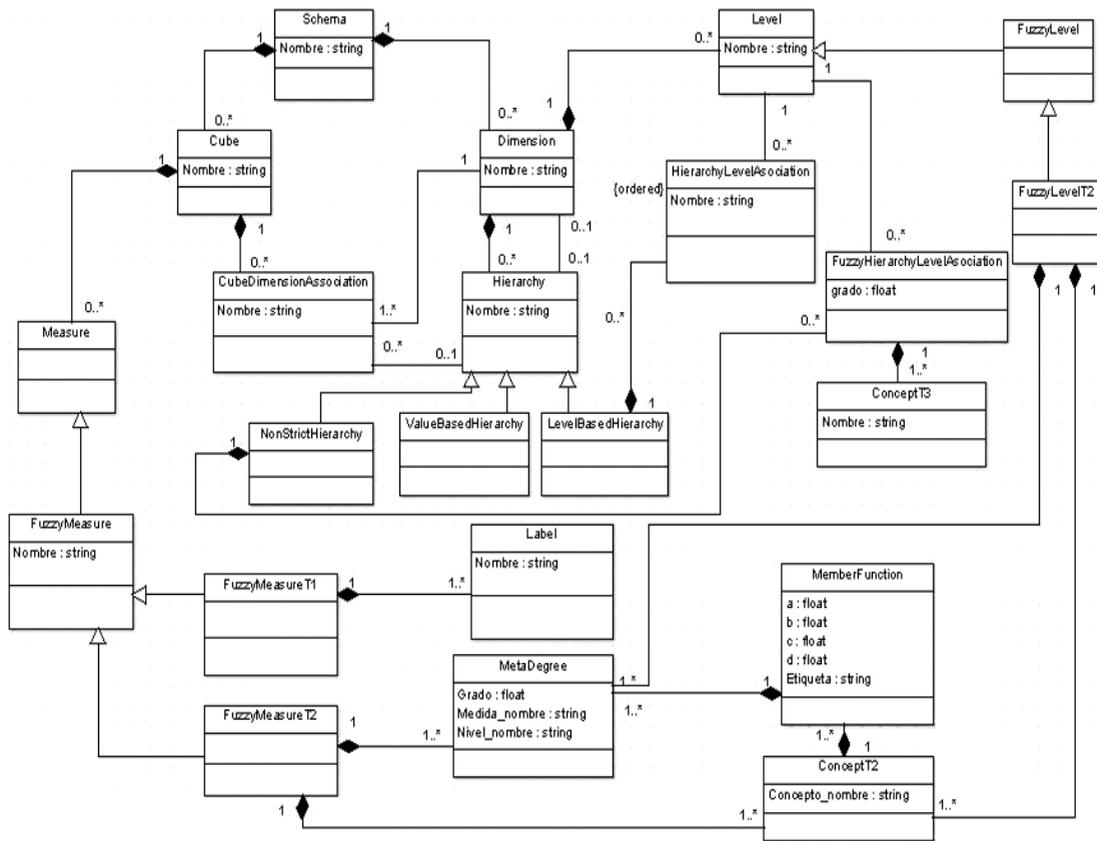


Figura 8. Extracto Metamodelo CWM OLAP Difuso con medidas difusas, relaciones difusas de una jerarquía y niveles difusos de una jerarquía.

- *RFactTable_Meta_Label_TO_FK*: Una tabla de hecho de un cubo etiquetado del metamodelo CWM OLAP Difuso es asociado a la metatabla de etiquetas en el metamodelo CWM Relacional mediante la creación de claves foráneas.
- *RConceptT2_TO_Meta_Concept*: La primera medida etiquetada *tipo 2* del metamodelo CWM OLAP Difuso permite transformar un ConceptT2 en la tabla Meta_Concept del metamodelo CWM Relacional. Dicha tabla se crea sólo una vez, cuando se aplica la regla, al encontrar la primera medida etiquetada tipo 2 o nivel etiquetado *tipo 2*. Esto se da porque las estructuras que soportan *tipo 2* son las mismas tanto para medidas *tipo 2* como para niveles tipo 2.
- *RMemberFunction_TO_Meta_MemberFunction*: La primera medida etiquetada *tipo 2* del metamodelo CWM OLAP Difuso permite transformar un MemberFunction del metamodelo CWM OLAP Difuso en la tabla Meta_MemberFunction del metamodelo CWM Relacional.
- *RMetaDegree_TO_MetaDegree*: La primera medida etiquetada *tipo 2* del metamodelo CWM OLAP Difuso permite transformar un MetaDegree en una tabla MetaDegree del metamodelo CWM Relacional.
- *RMetaConcept_MemberFunction_TO_FK*: La tabla MetaConcept es asociada a la tabla de funciones de membresía en el metamodelo CWM Relacional mediante la creación de claves foráneas.
- *RMeta_MemberFunction_Meta_Degree_TO_FK*: La tabla Meta_MemberFunction es asociada a la tabla Meta_Degree en el metamodelo CWM Relacional mediante la creación de claves foráneas.
- *RFactTable_Meta_Degree_TO_FK*: La tabla de hecho es asociada a la tabla Meta_Degree

en el metamodelo CWM Relacional mediante la creación de claves foráneas.

- *RLevel_TO_Table*: Todo nivel del metamodelo CWM OLAP Difuso se transforma en una tabla en el metamodelo CWM Relacional.
- *RConceptT3_TO_Meta_ConceptT3*: Toda relación difusa tipo 3 entre una dimensión y un nivel o entre nivel padre y nivel hijo del metamodelo CWM OLAP Difuso se transforma en una tabla *Meta_ConceptT3* en el metamodelo CWM Relacional.
- *RDimPadre_MetaConceptT3_TO_FK*: Cada relación difusa entre dimensión y nivel es transformada a sus correspondientes claves foráneas de las tablas del metamodelo CWM Relacional.
- *RLevHijo_MetaConceptT3_TO_FK*: Cada relación difusa entre dimensión y nivel es transformada a sus correspondientes claves foráneas de las tablas del metamodelo CWM Relacional.
- *RAttributeLevel_TO_Column*: Todo atributo de un nivel del metamodelo CWM OLAP Difuso se transforma en una columna en el metamodelo CWM Relacional.
- *RDimension_TO_Table*: Toda dimensión del metamodelo CWM OLAP Difuso se transforma en una tabla en el metamodelo CWM Relacional.
- *RAttributeDimension_TO_Column*: Todo atributo de una dimensión del metamodelo CWM OLAP Difuso se transforma en una columna en el metamodelo CWM Relacional.
- *R_Dh_TO_FK*: Toda clase de asociación dimensión/jerarquía del metamodelo CWM OLAP Difuso se transforma en una clave foránea en el metamodelo CWM Relacional.
- *RHla_TO_FK*: Toda clase de asociación de nivel/jerarquía del metamodelo CWM OLAP Difuso se transforma en una clave foránea en el metamodelo CWM Relacional.
- *RCda_TO_FK*: Toda clase de asociación del metamodelo CWM OLAP Difuso se transforma en una clave foránea en el metamodelo CWM Relacional.

Especificación de transformaciones

Una regla de transformación de modelos debe definir, evitando cualquier ambigüedad, la relación implícita que existe entre los elementos de los modelos fuentes de transformación. Por otro lado, un metamodelo es un mecanismo que permite

construir formalmente lenguajes de modelado. El metamodelo de un lenguaje es una descripción de todos los conceptos que pueden usarse en el mismo, por ejemplo, los conceptos de package, clase, atributo y operaciones en UML [26].

MDA propuso el estándar QVT [23] que permite crear consultas, vistas y transformaciones de modelos. Las transformaciones, en el contexto de QVT, se clasifican en relación (Relation) y función (Operational Mapping), donde Relation es un lenguaje de naturaleza declarativa. Las transformaciones del tipo Relations especifican transformaciones multidireccionales y chequean la consistencia entre dos o más modelos relacionados. Las funciones, en cambio, implementan la transformación, es decir, transforman elementos de un dominio en elementos de otro. El mapping se expresa mediante el lenguaje llamado Core, el cual es utilizado internamente por las herramientas de transformación y posee una notación imperativa.

En este artículo las reglas de transformación han sido expresadas por medio del lenguaje Relations, de QVT, el cual puede ser expresado mediante una notación gráfica y/o textual. Estas transformaciones se realizan a nivel de metamodelos, siendo el Metamodelo CWM OLAP Difuso el PIM DIFUSO fuente y el Metamodelo CWM Relacional Difuso el PSM DIFUSO destino. Esto permite instanciar un modelo conceptual multidimensional difuso a nivel lógico. A continuación se muestra a modo de ejemplo la formalización de la regla de transformación *ConceptT2_TO_Meta_Concept* mediante notación gráfica del lenguaje Relations de QVT donde la cláusula Where define la condición que debe ser satisfecha por los elementos de los modelos que participan en la relación, mientras que en la cláusula When se establece la condición que debe cumplirse antes de la transformación. Para más información sobre MDA y QVT se pueden revisar [8, 19, 21, 23, 27].

Regla *ConceptT2_TO_Meta_ConceptT2*

El objetivo de esta regla de transformación es convertir un concepto difuso tipo 2 *c* del Metamodelo CWM OLAP Difuso en una tabla *tmc* del Metamodelo CWM Relacional (ver Figura 9). Esta regla de transformación se compone de la regla *Measure_TO_Column* (*m, co*) y de la función OCL *ifNotExistsMetaConcept* (*t*) en la cláusula

When (precondiciones). En la poscondición Where, se compone de la regla *Meta_Concept_Meta_MemberFunction_TO_FK(fk)*, que permite asociar por medio de una clave foránea el concepto con una función de membresía y de la regla *MemberFunction_TO_Meta_MemberFunction*.

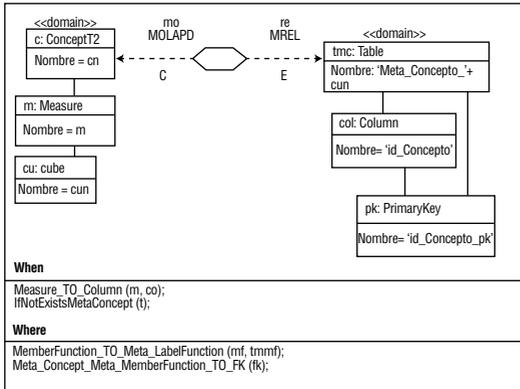


Figura 9. Notación gráfica de regla ConceptT2_TO_Meta_Concept

METODOLOGÍA DE APLICACIÓN DE LAS TRANSFORMACIONES

Para la transformación de un esquema conceptual de DWD a esquema lógico se debe como primer paso

generar el esquema conceptual multidimensional difuso por medio de una instancia del metamodelo CWM OLAP Difuso. Luego se aplican las transformaciones siguiendo el orden que se muestra en la Figura 10 que representa la secuencia de aplicación metodológica para la aplicación de las reglas de transformación. Finalmente el último paso es la generación del esquema lógico mediante una instancia del Metamodelo CWM Relacional Difuso.

Como se aprecia en la Figura 10 el procedimiento para aplicar las reglas de transformación siempre comienza con la regla *RMain*. Posteriormente se verifica si se aplicarán reglas para cubos, dimensiones o niveles. En caso de que se apliquen reglas para el cubo la primera en gatillarse es la regla *Cube_TO_FactTable* que permite generar la tabla de hechos. Luego la regla *Measure_TO_Column* convierte las medidas en columnas de la tabla de hechos. Es en este caso que se pueden aplicar las reglas para medidas tipo 1 y/o medidas tipo 2. Es importante indicar que basta con que sólo una medida haya sido modelada con los estereotipos «FMT1» y/o «FMT2» para que se utilicen las reglas que generan las estructuras para medidas tipo 1 y tipo 2. Estas estructuras se generan sólo una vez, independientemente de la cantidad de medidas tipo 1 o tipo 2 que existan. También se observa que si hay medidas etiquetadas

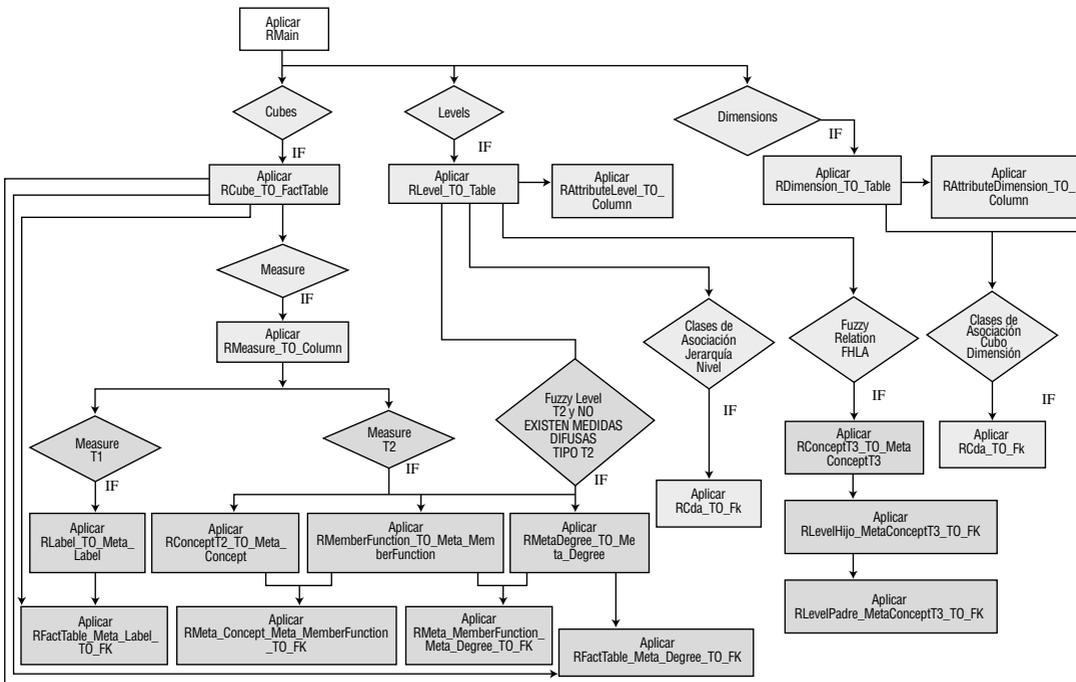


Figura 10. Diagrama con secuencia a seguir en la aplicación de las transformaciones.

tipo 1 se aplica la regla *Label_TO_MetaLabel* y la regla *FactTable_MetaLabel_TO_FK*. Por otro lado, si existen medidas etiquetadas como tipo 2 se aplican las reglas *ConceptT2_TO_MetaConcept*, *MeasureT2_TO_MetaDegree*, *MemberFunction_TO_Meta_MemberFunction* y las reglas *Meta_Concept_Meta_MemberFunction_TO_FK*, *Meta_MemberFunction_Meta_Degree_TO_FK* y *FactTable_Meta_Degree_TO_FK*. Estas tres últimas son las que asocian las distintas metatablas usando las claves foráneas.

Para el caso de las dimensiones, se aprecian las reglas *Dimension_TO_Table*, *AttributeDimension_TO_Column* y *Cda_TO_Fk*.

La secuencia de reglas para el caso de niveles con relaciones difusas se aplica cuando se ha modelado en el esquema conceptual una relación como difusa con el estereotipo «FHLA». En este caso la secuencia indica que se debe aplicar la regla *ConceptT3_TO_MetaConceptT3* que permite crear una tabla puente entre los niveles que son difusos (jerarquía no estricta). Luego las reglas *RLevelHijo_MetaConceptT3_TO_FK* y *RLevelPadre_MetaConceptT3_TO_FK* permiten asociar la clave primaria del nivel hijo como clave foránea de la tabla puente *MetaConceptT3* y asociar la clave primaria del nivel padre como clave foránea de la tabla puente *MetaConceptT3*.

Por otro lado se puede apreciar que las reglas para niveles difusos tipo 2 son las mismas que para las medidas etiquetadas como difusas tipo 2, por lo cual estas reglas se aplican sólo si no existen medidas etiquetadas como difusas tipo 2 y si existe un nivel con estereotipo

«FuzzyLevelT2» definido en el esquema conceptual. Esto se debe a que las estructuras que soportan el tipo 2 son las mismas tanto para medidas como para niveles.

La Figura 12 muestra las tablas *Meta_ConceptT3*, *Meta_ConceptT2*, *Meta_De-gree*, *Meta_MemberFunction* y con sus respectivos atributos, en un esquema lógico final después de aplicar la secuencia de reglas propuestas en esta investigación. A continuación el siguiente apartado muestra el caso de estudio.

CASO DE ESTUDIO

En el caso de estudio que se presenta se aplican las reglas de transformación del enfoque MDA Difuso propuesto. Para esto se presenta un Cubo de Indicadores de Estudiantes [35]. Éste se modela conceptualmente usando una instancia del metamodelo CWM OLAP Difuso propuesto.

Cubo Indicadores de Estudiantes

El Cubo Indicadores de Estudiantes está compuesto por seis dimensiones de análisis que son Cohorte, Carrera, Tiempo, Estudiante, Localización y Curso. Las dimensiones Cohorte, Carrera y Tiempo no tienen niveles (ver Figura 11). Las dimensiones Estudiante, Localización y Curso sí tienen niveles.

Los datos que se ocuparon para el análisis son los datos de pregrado de los estudiantes de la Facultad de Ingeniería de la Universidad de Atacama. El propósito de implementación del indicador es proporcionar información de las notas promedio, haciendo un análisis por cohorte, carrera, estudiante, localización y curso. En este contexto, se definió como medida difusa «FMT2» las Notas Promedio. También se aplica un nivel difuso «FuzzyLevelT2» que es Región y una relación difusa «FHLA» entre los niveles Estudiante y Nivel Estudiante.

Esta última relación es difusa por esencia, dado que regularmente los alumnos no se encuentran al día en el avance de su malla curricular, sino que, por el contrario, tienen asignaturas de distintos niveles (semestres), lo cual nos indica que su grado de pertenencia a un nivel de su carrera es difuso. En el caso de la dimensión localización, se observa que tiene dos niveles que son ciudad y región, donde el nivel región ha sido marcado como difuso, esto indica que se puede etiquetar como difuso tipo 2. Resumiendo, en la Figura 11 se aprecia el Esquema Conceptual para el Cubo Indicadores de alumnos que se representa mediante una instancia del Metamodelo CWM OLAP Difuso. En esta figura se puede observar el estereotipo «FMT2» que representa las medidas difusas tipo 2, el estereotipo «FuzzyLevelT2» que representa al nivel difuso tipo 2 y el estereotipo «FHLA» que representa a las relaciones difusas tipo 3 en una jerarquía. Luego se aplica el paso que corresponde a la identificación de los cubos, medidas, dimensiones, atributos de dimensiones, niveles, atributos de niveles, clases de asociación

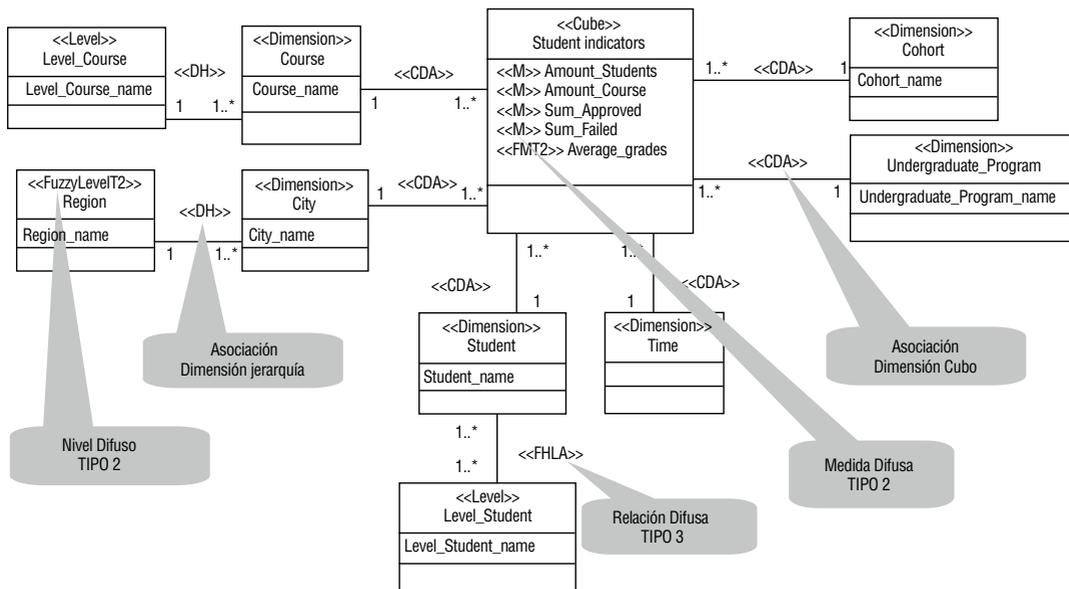


Figura 11. Esquema conceptual para el indicador notas promedio, el estereotipo <<FMT2>> representa las medidas difusas tipo 2, el estereotipo <<FuzzyLevelT2>> representa al nivel difuso tipo 2 y el estereotipo <<FHLA>> representa a las relaciones difusas tipo 3 en una jerarquía.

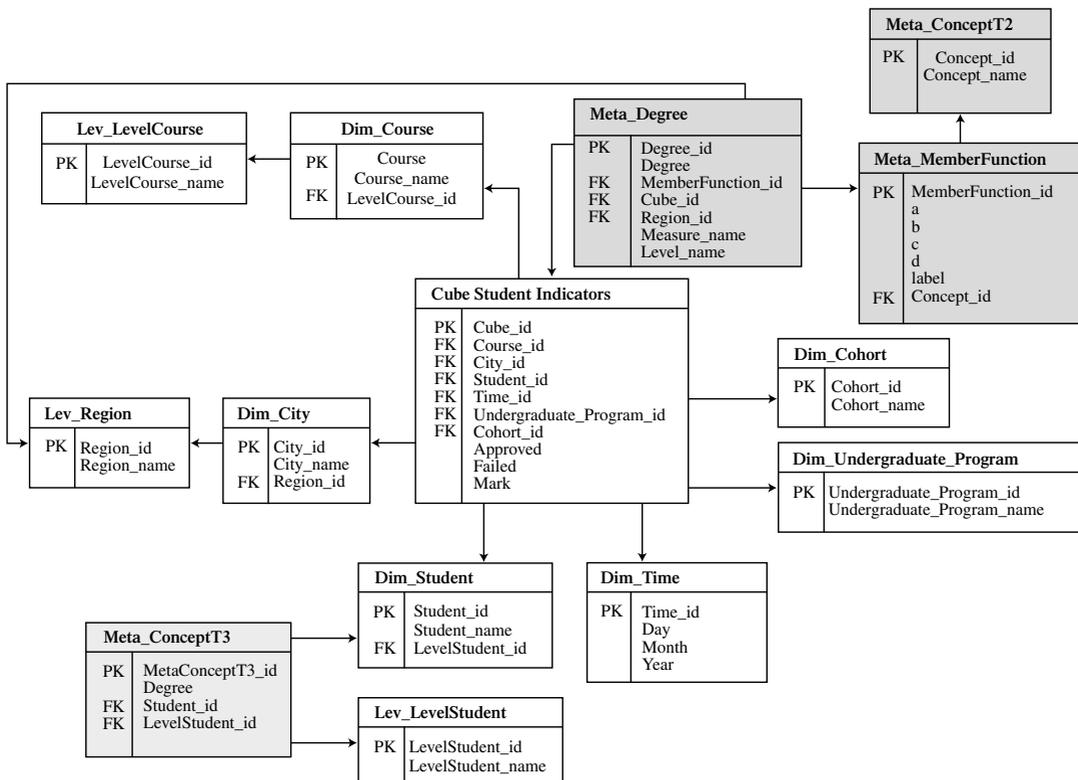


Figura 12. Esquema lógico para el indicador de alumnos notas promedio, con estructuras para soportar medidas difusas tipo 2, niveles difusos tipo 2 y relaciones difusas tipo 3 en una jerarquía.

y las medidas difusas tipo 1, tipo 2, niveles difusos tipo 2 y relaciones difusas tipo 3 de la instancia del Metamodelo CWM OLAP Difuso.

Después se aplican las reglas según las secuencias mostradas en la Figura 10 generando así el Modelo Lógico Multidimensional Difuso.

En la Figura 12 se aprecia el Esquema Lógico para el indicador de alumnos Notas Promedio, con las estructuras para soportar los elementos difusos que se muestran en color plomo.

CONCLUSIONES Y TRABAJO FUTURO

En este artículo se ha introducido el enfoque MDA para el diseño de un DWD presentando una arquitectura MDA Difusa. Se ha trabajado en la transformación de un PIM DIFUSO a un PSM DIFUSO, para lo cual se extendió el Metamodelo CWM OLAP para soportar medidas etiquetadas como difusas *tipo 1* y *tipo 2*, niveles etiquetados *tipo 2* y relaciones difusas *tipo 3*, dando origen al Metamodelo CWM OLAP Difuso (PIM DIFUSO) que sirvió como metamodelo de entrada para definir una serie de transformaciones que se ejemplificaron y se definieron metodológicamente.

Es importante resaltar que el estereotipo definido para la relación difusa también sirve para relaciones muchos a muchos entre niveles de una jerarquía que no son difusos. Esto debido a que el enfoque de solución, para este tipo de jerarquías llamadas no estrictas, es incorporar una tabla puente entre la relación muchos a muchos que en nuestro caso es también difusa. Por lo tanto, la extensión del Metamodelo CWM OLAP Difuso aporta extendiendo las jerarquías no estrictas, las que no eran soportadas por el Metamodelo CWM OLAP. Desde el punto de vista de implementación lo anterior no es problema, puesto que el Sistema de Gestión de Base de Datos SQL Server provee un mecanismo para su manipulación.

El DWD propuesto permite relaciones difusas entre niveles, lo que otorga una ventaja a nivel de instancias del DW debido a que al permitir que una instancia pertenezca a más de una clase (categoría o nivel) se otorga un nivel de análisis mucho más natural que se encuentra en muchos casos. Lo anterior obliga a que se necesite extender la función de agregación

tradicional a agregación difusa, tarea que estamos desarrollando.

Por otro lado, se ha mostrado el proceso de transformación de nuestro PIM DIFUSO al PSM DIFUSO usando un caso de estudio real basado en el análisis de indicadores de estudiantes. Por lo tanto, la propuesta aporta en el diseño de un DWD tanto a nivel de diseño conceptual como a nivel de diseño lógico.

Una de las ventajas de la propuesta de esta investigación es que puede ser fácilmente implementada bajo arquitectura ROLAP. Además el uso del enfoque MDA ha permitido integrar todo el proceso de diseño de un DWD, quedando algunas etapas pendientes que fácilmente se integrarán bajo este enfoque.

Como trabajos futuros se considera la automatización de las reglas presentadas, el trabajo de ingeniería de requerimientos para DWD de forma de representar el nivel de abstracción CIM DIFUSO en el enfoque MDA. También se trabaja en la implementación de reglas de transformación para las típicas funciones de agregación y para la agregación difusa necesaria en el caso de las relaciones difusas entre niveles. Todo lo anterior con el objetivo de implementar un Framework basado en enfoque MDA para el desarrollo de DWD. También se puede hacer un estudio de los PSM.

AGRADECIMIENTOS

Este trabajo fue parcialmente financiado por la Dirección de Investigación de la Universidad de Atacama, Chile, Proyecto 221219 "Data Warehouse para Análisis con Jerarquías Difusas".

REFERENCIAS

- [1] L. Cabibbo and R. Torlone. "A logical approach to multidimensional databases". Lecture Notes in Computer Science. Vol. 1377. 1998.
- [2] F. Carpani. "CMDM: Un modelo conceptual para la especificación de bases de datos multidimensionales". Tesis para optar al grado de Maestría. Universidad de la República. Uruguay. 2000.

- [3] S. Carrera, M. Varas y A. Urrutia, "Transformación de esquemas multidimensionales difusos desde el nivel conceptual al nivel lógico". *Ingeniare. Revista Chilena de Ingeniería*. Vol. 18 N° 2, pp. 165-175. Mayo-Agosto 2010.
- [4] S. Chaudhuri and U. Dayal. "An overview of data warehousing and OLAP technology". *SIGMOD Record*. Vol. 26, pp. 65-74. 1997. Pearson. 2004. ISBN 8420540250.
- [5] M. Delgado, C. Molina, D. Sánchez, A. Vila and L. Rodríguez-Ariza. "A fuzzy multidimensional model for supporting imprecision in OLAP". *Proceedings of IEEE International Conference on Fuzzy Systems*. ISBN 0-7803-8353-2. 2004.
- [6] D. Fasel and K. Shahzad. "A data warehouse model for integrating fuzzy concepts in meta table structures". In *Proc. ECBS 2010*, pp. 100-109. 2010.
- [7] E. Franconi and A. Kamblet. "A data warehouse conceptual data model". *Proceedings of the 16th International Conference on Scientific and Statistical Database Management*, pp. 435-436. June, 2004.
- [8] D.S. Frankel. "Model driven architecture. applying MDA to enterprise computing". Wiley. Indianapolis, Indiana. 2003.
- [9] J. Galindo, A. Urrutia and M. Piatinni. "Fuzzy databases: Modeling, design and implementation". Idea Group, Inc. 2009.
- [10] J. Galindo, A. Urrutia and M. Piatinni. "Handbook of research on fuzzy information processing in databases". Edited By José Galindo, Universidad de Málaga. Spain. 2008.
- [11] M. Golfarelli, D. Maio and S. Rizzi. "The dimensional fact model: A conceptual model for data warehouses". *International Journal of Cooperative Information Systems*. Vol. 7, Issue 3, pp. 215-247. 1998.
- [12] C. Hurtado, C. Gutierrez and A. Mendelzon. "Capturing summarizability with integrity constraints in OLAP". *ACM Transactions on Database Systems*. Vol. 30, Issue 3, pp. 854-886. 2005.
- [13] W. Inmon. "Building the data warehouse". John Wiley & Sons. 2002.
- [14] C. Jensen, K. Augustas, T. Bach Pedersen and I. Timko. "Multidimensional data modelling for location-based services". *VLDB J*. Vol. 13, Issue 1, pp. 1-21. 2004.
- [15] R. Kimball, M. Ross and R. Merz, "The data warehouse toolkit: The complete guide to dimensional modeling". John Wiley & Sons. 2002.
- [16] A. Laurent and F. Hoffmann. "Generating fuzzy summaries from fuzzy multidimensional databases". *IDA 2001, LNCS 2189*, pp. 24-33. Springer. 2001.
- [17] J.-N. Mazón and J. Trujillo. "An MDA approach for the development of data warehouses". *Decision Support Systems*. Vol. 45, pp. 41-58. 2008.
- [18] J.-N. Mazón, J. Lechtenböcker and J. Trujillo. "A survey on summarizability issues in multidimensional modeling". *Data Knowl. Eng.* Vol. 68, Issue 12, pp. 1452-1469. 2009.
- [19] S. Mellor, K. Scott, A. Uhl and D. Weise, "MDA distilled: Principles of model driven architecture". Addison Wesley. 2004.
- [20] C. Molina, L. Rodríguez-Ariza, D. Sánchez and M. Vila. "A new fuzzy multidimensional model". *IEEE Transactions On Fuzzy Systems*. Vol. 14, Issue 6. December, 2006.
- [21] OMG: "MDA guide". Version 1.0.1. URL: <http://www.omg.org/docs/omg/03-06-01.pdf>. June, 2003.
- [22] OMG: "Object constraint language (OCL) especification 2.0". URL: <http://www.omg.org/cgi-bin/doc?ptc/03-10-14>. October, 2003.
- [23] OMG: "MOF QVT final adopted specification". URL: <http://www.omg.org/cgi-bin/doc?ptc/2005-11-01>. November, 2005.
- [24] OMG: "Object management group". URL: <http://www.omg.org/>.
- [25] OMG: "Common warehouse metamodel CWM specification". Versión 1.1. URL <http://www.omg.org/cgi-bin/doc?formal/03-03-02>. March, 2003.
- [26] OMG: "Unified modeling language specification". Version 1.5. URL: <http://www.omg.org/cgi-bin/doc?formal/03-03-01>. March, 2003.
- [27] OMG: "MOF 2.0/XMI mapping". Version 2.1.1. URL: <http://www.omg.org/docs/formal/07-12-01.pdf>. December, 2007.
- [28] L. Sapir, A. Shmilovici and L. Rokach. "A methodology for the design of a fuzzy

- data warehouse". 4th IEEE Conference on Intelligent Systems. Bulgaria. 2008.
- [29] J. Trujillo, M. Palomar, J. Gómez and Il-Yeol Song. "Designing data warehouses with OO conceptual models". IEEE Computer. Vol. 34, pp. 66-75. 2001.
- [30] N. Tryfona, F. Busborg and J. Borch. "StarER: A conceptual model for data warehouse design". In Proceedings of the ACM International Workshop on Data Warehousing and OLAP. Kansas City, Kansas. 1999.
- [31] A. Urrutia. "Definición de un modelo conceptual para bases de datos difusas". Tesis para optar al grado de doctor. Universidad de Castilla-La Mancha. España. 2003.
- [32] A. Urrutia, M. Varas y J. Galindo. "Diseño de una base de datos difusa modelada con UML". 6° Workshop Iberoamericano de Ingeniería de Requisitos y Ambientes Software. IDEAS, pp. 145-155. Asunción, Paraguay. Mayo 2003.
- [33] M. Varas y A. Urrutia. "Bases de datos difusas modeladas con UML". VIII Congreso Argentino de Ciencias de la Computación. CACIC. Buenos Aires, Argentina. 2002.
- [34] L.A. Zadeh. "Fuzzy Sets". Information and Control. Vol. 8, pp. 338-353. 1965.
- [35] C. Zambrano. "Modelo Conceptual de Data Warehouse Difuso con manejo de jerarquías difusas". Tesis para optar al grado de Magíster en Ciencias de la Computación. Universidad de Concepción. Dirigida por: Marcela Varas Contreras.
- [36] L. Zepeda and M. Celma. "A model driven approach for data warehouse conceptual design". In Proceedings of the Seventh International Baltic Conference on Databases and Information Systems. 2006.