Meta-modelización de Expresiones de Cálculo de Vistas

Trabajo de Pasantía
Laboratorio PRISM - Universidad de Versailles
Noviembre 2001 - Marzo 2002

Verónika Peralta
Abril 2002

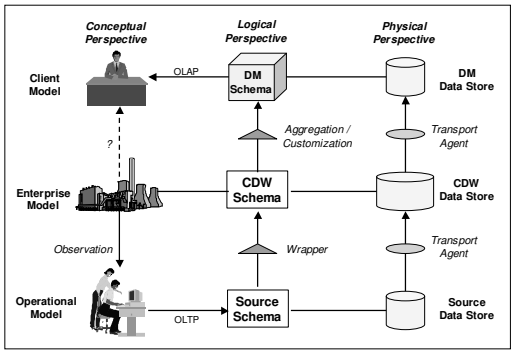
1

Plan

- **Contexto.**
- **Meta-modelo**
  - Expresiones de Cálculo de Vistas.
  - Reglas de Cálculo de Attributos.
- **Plataforma para MSIS.**
  - Prototipo actual.
  - Aplicaciones y herramientas.
- **Conclusions.**

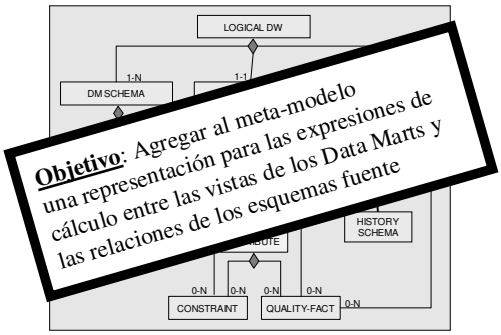
2

Contexto



3

Motivación

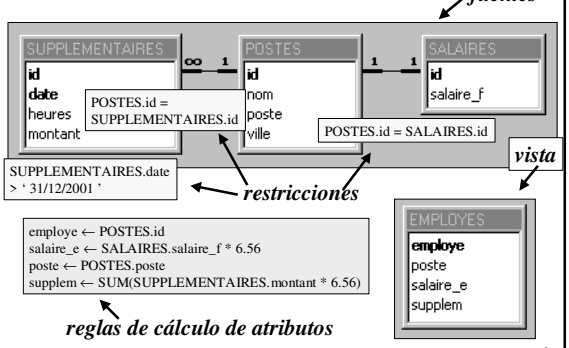


Objetivo: Agregar al meta-modelo una representación para las expresiones de cálculo entre las vistas de los Data Marts y las relaciones de los esquemas fuente

4

Problema a estudiar

relaciones fuentes



5

Problema a estudiar

- **Enriquecer el meta-modelo inicial:**
  - Representar las expresiones de cálculo de vistas (ECV)
    - Extensión del álgebra, SQL, nueva definición, ...
    - Una expresión compleja o varias expresiones simples.
  - Representar las reglas de cálculo de atributos (RCA)
    - Gramática, expresiones SQL, funciones de usuario, …
- **Explotar el meta-modelo:**
  - Componer las ECV a partir de las RCA.
  - Verificar la coherencia entre las ECV y las RCA.

6

Expresiones de Cálculo de Vistas

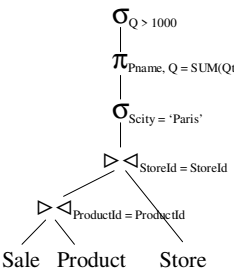
- **Para representar una vista simple: [Aki99]**
  - R= {R1 ,…Rn } : Relaciones
  - E = S ∪ A : Expresiones formadas con atributos de las relaciones:
    - S: Expresiones simples (sin agregados) para agrupar.
    - A: Expresiones con agregados.
  - C = CS ∪ CA : Restricciones (constraints) sobre los atributos de las relaciones:
    - CS: Restricciones sobre las tuplas de las relaciones.
    - CA: Restricciones sobre los agregados.
    - CR = {CR1,… CRn-1} : Restricciones de join entre las relaciones.

$$\sigma_{CA}\,(\pi_{S,A}\,(\sigma_{CS}\,(R_1 \rhd\lhd_{CR1} \ldots \rhd\lhd_{CRn-1} R_n)))$$ ***Normal Form [Gup95]***

- **Para representar vistas anidadas (nested views) [Gol00]**
  - Ri puede ser una relación o una vista.

Ejemplo de ECV

Find the average selling price and the total quantity sold, for each product whose total quantity sold is greater than 1000, at Paris.

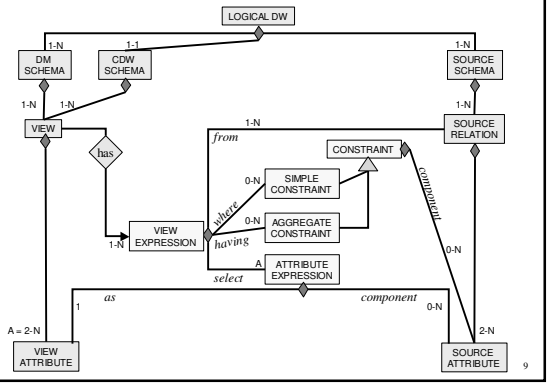


STORE (Storeid, Sname, Scity)
PRODUCT (Productid, pname, ptype, pcategory)
SALE (Timeid, Storeid, Productid, Qty, Prc)

```
SELECT   pname, sum(qty) as Q, avg(prc) as P
FROM     Sale, Product, Store
WHERE    Sale.storeid=Store.storeid
AND      Sale.productid= Product.productid
AND      Scity = ' Paris '
GROUP BY pname
HAVING   sum(qty) > 1000
```

Otros ejemplos: [Gol00]

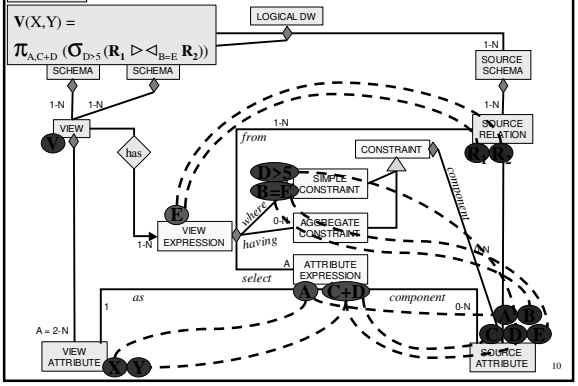
Meta-modelo simple



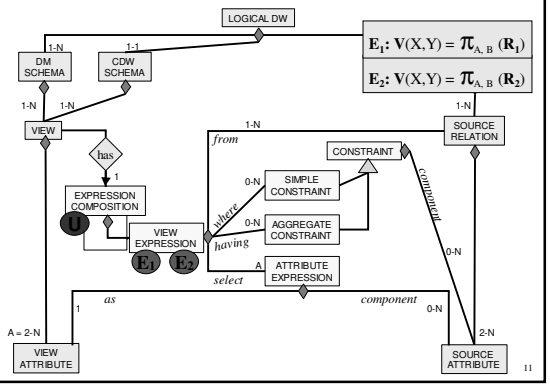
Ejemplo del Meta-modelo

$\mathbf{R_1}(A,B)$
$\mathbf{R_2}(C,D,E)$

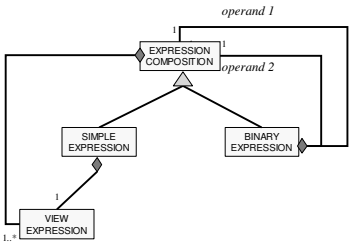
$\mathbf{V}(X,Y) = \pi_{A,C+D}\,(\sigma_{D>5}\,(\mathbf{R_1} \rhd\lhd_{B=E} \mathbf{R_2}))$



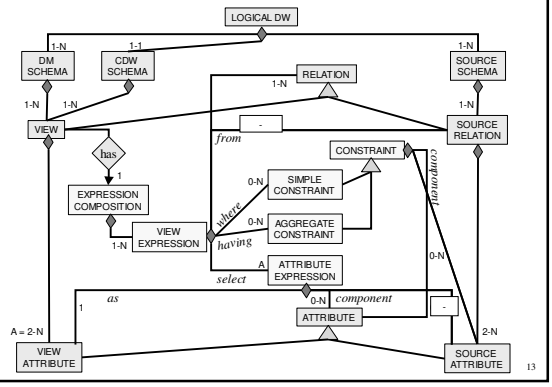
Composición de ECV



Composición de ECV



Generalización del Meta-modelo



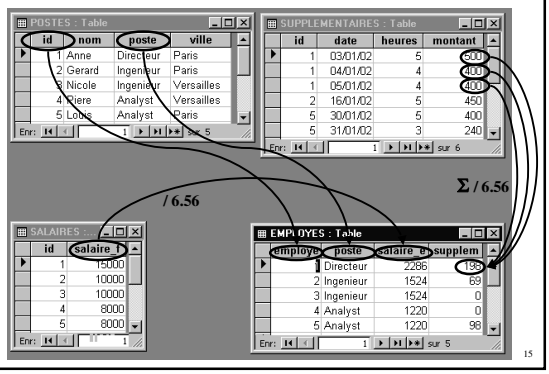
13

Reglas de Cálculo de Atributos

- **Un atributo de una tabla fuente.** — *SExpr*
  - POSTES.nom
- **Un cálculo con atributos de tablas fuentes.**
  - Majuscule (POSTES.nom) — *SExpr*
  - Concaténation (POSTES.id, POSTES.nom) — *SExpr*
  - Premier-mot (POSTES.nom) — *SExpr*
  - SALAIRES.salaire_f * 6.56 — *SExpr*
- **Un agregado con atributos de tablas fuentes.**
  - Sum (SALAIRES.salaire_f) — *AExpr*
  - Sum (SALAIRES.salaire_f * 6.56) — *AExpr*
  - Sum (SALAIRES.salaire_f) / Count (SALAIRES.salaire_f) — *AExpr*
  - Sum (SALAIRES.salaire_f) / VILLES.viatique — *CExpr*
- **Un valor externo a las fuentes.**
  - Aujourd'hui () — *VExpr*
  - 6.56 — *VExpr*
  - Auto-numérique () — *VExpr*

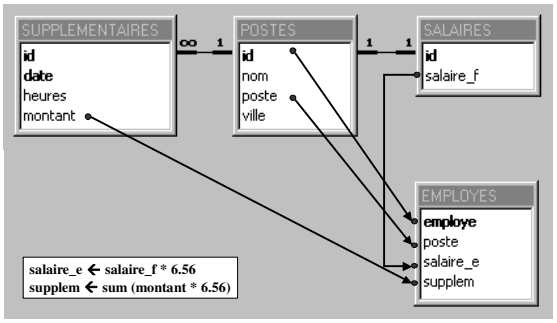
14

RCA: Un ejemplo



15

Repr. Gráfica de las RCA



16

Gramática para las RCA

Let BasicTypes be the set of basic types, such as *boolean, integer, float, string* and *date*.
Let Values be a set of constant values of basic types.
Let Attributes be a set of attributes. For each A ∈ Attributes . Dom(A) ∈ BasicTypes.

ValueExpressions

- If E ∈ Values
  → E ∈ VExpr

SimpleExpressions

- If A ∈ Attributes
  → A ∈ SExpr
- If $E_1 \ldots E_n \in$ SExpr, n ≥ 1, f ∈ Dom(E_1) x … x Dom(E_n) → T, T ∈ BasicTypes
  → f($E_1, \ldots E_n$) ∈ SExpr

AggregateExpressions

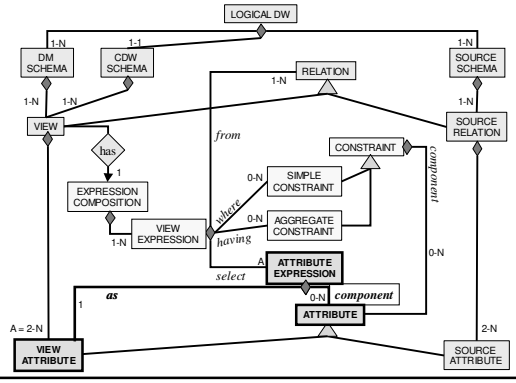
- If E ∈ SExpr, g ∈ Set(Dom(E)) → T, T ∈ BasicTypes → g(E) ∈ AExpr
  → g(E) ∈ SExpr
- If $E_1 \ldots E_n \in$ AExpr, n ≥ 1, f ∈ Dom(E_1) x … x Dom(E_n) → T, T ∈ BasicTypes
  → f($E_1, \ldots E_n$) ∈ AExpr

ComposedExpressions

- If $E_1 \in$ SExpr, $E_2 \in$ AExpr, $E_3 \ldots E_n \in$ Sexpr ∪ Aexpr, n ≥ 2, f ∈ Dom(E_1)x…x Dom(E_n) → T, T ∈ BasicTy
  → f($E_1, \ldots E_n$) ∈ CExpr

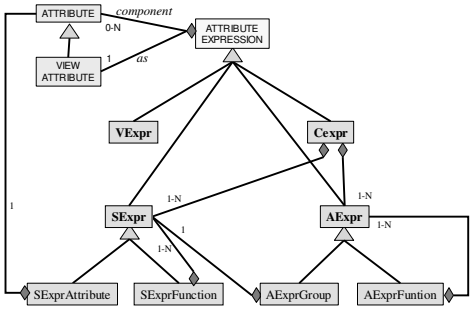
17

Meta-modelo lógico del DW



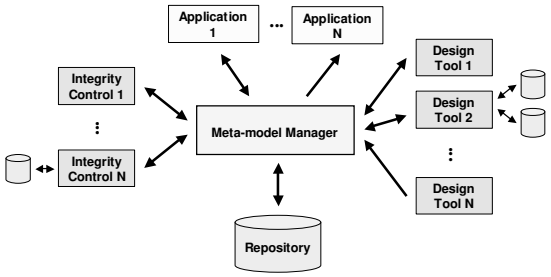
18

Meta-modelo de las RCA



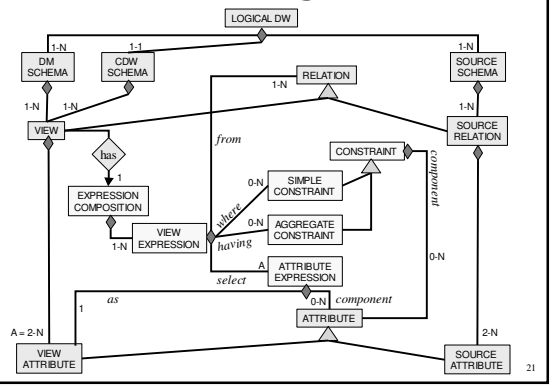
19

Plataforma para MSIS



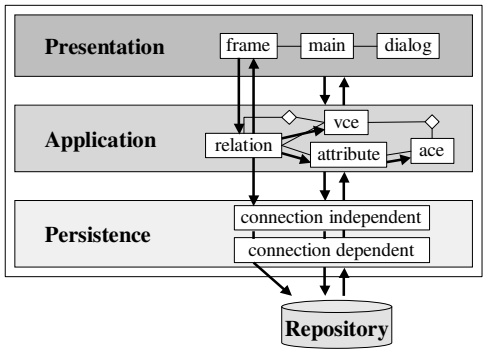
20

Meta-modelo lógico del DW

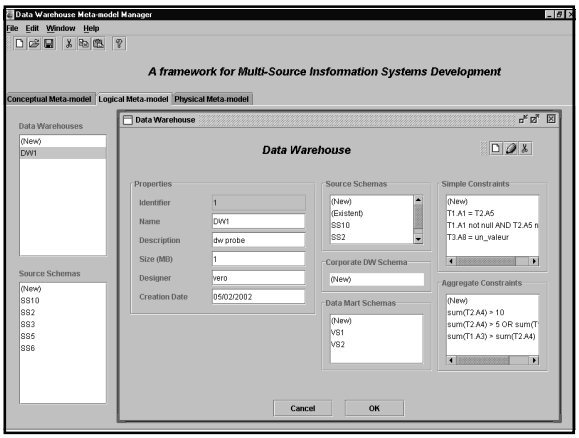


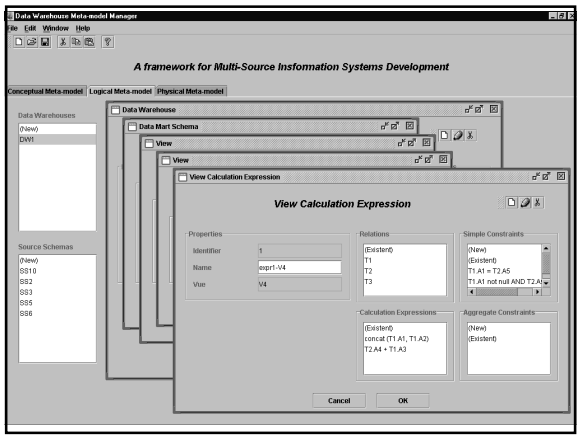
21

Arquitectura del Prototipo



22





Aplicaciones y herramientas

- **Discovering view calculation expressions:**
  - Cómo calcular las ECV accediendo a fuentes existentes. [Ked99]
  - Input: views, source relations.
  - Output: ECV.
- **Automatic query rewriting using user profiles:**
  - Los requerimientos de usuarios pueden ser re-escritos integrando perfiles de usuarios (en forma de vistas) [Lah02]
  - Input: requerimientos (externos a la plataforma).
  - Output: views, ECV.

25

Aplicaciones y herramientas

- **DW cleanning processes:**
  - [Sou02]
  - Input: source schema, CDW schema.
  - Output: ECV, loading processes information.
- **CDW logical design and evolution**
  - [Mar00], [Mar01], [Per01]:
  - Input: source schemas.
  - Output: CDW schema.

26

Conclusiones

- **Se realizó:**
  - Estudio de ECV y RCA.
  - Meta-modelo para ECV y RCA.
  - Prototipo de la plataforma.
- **Trabajo futuro:**
  - Verificación de la coherencia entre ECV y RCA.
  - Composición de ECV a partir de RCA.
  - Enriquecer meta-modelo conceptual y físico.
  - Derivación de ECV y RCA a partir del meta-modelo conceptual.
  - Desarrollo de aplicaciones y herramientas.
  - Meta-modelizar los procesos.

27

Bibliografía

- [Aki99] Akinde, M. Böhlen, M.
  "*Constructing GPSJ View Graphs*". DMDW'99.
- [Gol00] Golfarelli, M. Rizzi, S.
  "*View Materialization for Nested GPSJ Queries*". DMDW'00.
- [Gup95] Gupta, A. Harinarayan, V. Quass, D.
  "*Aggregated-query processing in Data Warehousing environments*". VLDB'95.
- [Jar97] Jarke, M. Vassiliou, Y.
  "*Data Warehouse Quality Design: A Review of the DWQ Project*". Information Quality'97.
- [Jar99] Jarke, M. Lenzerini, M. Vassiliou, Y. Vassiliadis, P.
  "*Fundamentals of Data Warehouses*". Springer-Verlag, 1999.
- [Ked99] Kedad, Z. Bouzeghoub, M.
  "*Discovering View Expressions from a Multi-Source Information System*". CoopIS'99.

28