Physical Modeling of Data Warehouses using UML

Sergio Luján-Mora
Juan Trujillo

DOLAP 2004

Contents

• Motivation
• UML extension mechanisms
• DW design framework
• DW physical design
• Conclusions and future work
Motivation

• Data warehouses are complex information systems
• Support:
  – OLAP
  – Data mining
  – Decision Support Systems
  – …
• Building a DW: time consuming, expensive and prone to fail

Motivation

• Partial approaches:
  – ETL processes
  – Logical and conceptual design of the DW based on the MD paradigm
  – Derive DW schema from ER schemas of the data sources
  – …
• Most of the research efforts focused on MD data models
Motivation

• Implementation decisions:
  – Storage in different disks
  – Replication
  – Vertical and horizontal partitioning
  – Influence performance and maintenance
  – …

• Solution:
  – Tackle **physical design** from early stages
    • Allows the designer to anticipate physical design decisions
    • Reduce development time and cost

• Previous work: **Data Warehouse Engineering Process**
  – Modeling language that assists an entire DW project
  – Based on standards (UML, UP, XML)
  – Represent the models at different levels of granularity (from high-level to low-level)
  – Used at different stages of the DW project
  – Used by different personal (business users, administrators, etc.)
Motivation

• This work: Physical Design of DW
  – Component and deployment diagram from UML
  – Integrated in our DWEP: maps elements from the logical level into the physical level
  – Aimed to be used by DW designers (how to build) and administrators (how to implement and maintain)
Contents

• Motivation
• **UML extension mechanisms**
  • DW design framework
  • DW physical design
  • Applying modeling schemas
  • Conclusions and future work

---

**UML extension mechanisms**

• UML is a *general purpose* visual modeling language for systems
• Extension mechanisms allow the user to tailor it to specific domains
• Mechanisms:
  – Stereotypes → New building elements
  – Tagged values → New properties
  – Constraints → New semantics
Physical Modeling of Data Warehouses using UML

**UML extension mechanisms**

<table>
<thead>
<tr>
<th>Icon</th>
<th>Decoration</th>
<th>Label</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact 1</td>
<td>Fact 2</td>
<td>Fact 3</td>
<td>Fact 4</td>
</tr>
</tbody>
</table>

**Package stereotypes**

- StarPackage (Level 1)
- FactPackage (Level 2)
- DimensionPackage (Level 2)

**Class stereotypes**

- Fact (Level 3)
- Dimension (Level 3)
- Base (Level 3)
Contents

• Motivation
• UML extension mechanisms
• **DW design framework**
• DW physical design
• Applying modeling schemas
• Conclusions and future work

**DW diagrams**

• Development of DW can be structured into an integrated framework:
  
  – Five stages
  – Three levels

  Fifteen diagrams

• Diagrams spread throughout the five stages and the three levels
• Each diagram uses different formalisms (class diagram, component diagram, etc.) → Several UML profiles have been proposed:
  – Multidimensional profile
  – ETL Profile
  – Data Mapping Profile
  – **Database Deployment Profile**
**Physical Modeling of Data Warehouses using UML**

**Source (S)** (OLTP, external data, ...)

**Integration**

**Data Warehouse (DW)**

**Customization**

**Client (C)** (OLAP, data mining, ...)

**Conceptual**

<table>
<thead>
<tr>
<th>SCS</th>
<th>DM</th>
<th>SWDS</th>
<th>DMCS</th>
<th>SCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class diagram</td>
<td>ETL Diagram</td>
<td>Data Mapping Profile</td>
<td>Standard UML</td>
<td>Class diagram</td>
</tr>
<tr>
<td>Standard UML</td>
<td>Data Mapping Profile</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Logical**

<table>
<thead>
<tr>
<th>SL5</th>
<th>ETL Process</th>
<th>DWLS</th>
<th>Expanding Process</th>
<th>CL9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class diagram</td>
<td>ETL Process</td>
<td>Different data modeling profiles</td>
<td>Class diagram</td>
<td>Different data modeling profiles</td>
</tr>
<tr>
<td>Different data modeling profiles</td>
<td>ETL Process</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Physical**

<table>
<thead>
<tr>
<th>SPS</th>
<th>Transportation Diagram</th>
<th>EMPS</th>
<th>Transportation Diagram</th>
<th>CPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp. &amp; deploy. diagrams</td>
<td>Deployment diagram</td>
<td>Comp. &amp; deploy. diagrams</td>
<td>Deployment diagram</td>
<td>Comp. &amp; deploy. diagrams</td>
</tr>
<tr>
<td>Database Deployment Profile</td>
<td>Database Deployment Profile</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


---

**Physical Modeling of Data Warehouses using UML**

**DW diagrams**

- **Stages:**
  - Source: data sources (OLTP, external data sources, etc.)
  - Integration: mapping between source and data warehouse
  - Data Warehouse: structure of the DW
  - Customization: mapping between data warehouse and clients' structures
  - Client: structures used by the clients to access the DW (data marts, OLAP applications, etc.)
Physical Modeling of Data Warehouses using UML

**DW diagrams**

- For each stage, different levels:
  - Conceptual
  - Logical
  - Physical
- Remarks:
  - Every DW project does not need the fifteen diagrams
  - The different diagrams of the same DW are not independent but overlapping (UML importing mechanism)

**Contents**

- Motivation
- UML extension mechanisms
- DW design framework
- **DW physical design**
- Applying modeling schemas
- Conclusions and future work
Physical Modeling of Data Warehouses using UML

**Source (S)**
- Conceptual: SCS Class diagram, Standard UML
- Logical: SLS Class diagram, Different data modeling profiles
- Physical: SPS Comp. & deploy. diagrams, Database Deployment Profile

**Integration (I)**
- Conceptual: DI Class diagram, ETL Profile
- Logical: ETL Process Diagram, ETL Profile
- Physical: Transportation Diagrams, Comp. & deploy. diagrams, Database Deployment Profiles

**Data Warehouse (DW)**
- Conceptual: DWS Class diagram, Standard UML, Multidimensional Profile
- Logical: DWLS Class diagram, Different data modeling profiles
- Physical: Transportation Diagrams, Comp. & deploy. diagrams, Database Deployment Profiles

**Customization (C)**
- Conceptual: CNS Class diagram, Standard UML, Multidimensional Profile
- Logical: CLS Class diagram, Different data modeling profiles
- Physical: CPS Comp. & deploy. diagrams, Database Deployment Profile

**Client (C)**
- Conceptual: Client Physical Schema
- Logical: Source Physical Schema
- Physical: Data Warehouse Physical Schema

---

**Physical Modeling of Data Warehouses using UML**

**DW physical design**

- **UML component and deployment diagrams extended** → **Database Deployment Profile**: `<<Database>>`, `<<Tablespace>>`, `<<Table>>`, etc.

**Diagrams**
- Source Physical Schema
- Data Warehouse Physical Schema
- Client Physical Schema
- Integration Transportation Diagram
- Customization Transportation Diagram

**Component and deployment diagram**

**Deployment diagram**
Physical Modeling of Data Warehouses using UML

**DW physical design**

- Example:
  - DW with daily sales of a company that sales automobiles (cars and trucks)
  - Dimensions of analysis: automobile, customer, dealership, salesman, time
  - Two data sources:
    - Sales server: transactions and sales
    - CRM server: customers
  - Different final users' requirements:
    - MacOS and Windows
    - Web and desktop application
Physical Modeling of Data Warehouses using UML

Level 2: Star schema definition
Physical Modeling of Data Warehouses using UML

**DW physical design**

Data Warehouse Logical Schema: ROLAP

Source Physical Schema: deployment diagram
Physical Modeling of Data Warehouses using UML

DW physical design

Data Warehouse Physical Schema: component diagram

Physical Modeling of Data Warehouses using UML

DWPS

Data Warehouse Physical Schema: deployment diagram
Physical Modeling of Data Warehouses using UML

**Data Warehouse Physical Schema: deployment diagram**

**Integration Transportation Diagram: deployment diagram**
Physical Modeling of Data Warehouses using UML

**DW physical design**

**Contents**

- Motivation
- UML extension mechanisms
- DW diagrams
- DW engineering process
- Applying modeling schemas
- **Conclusions and future work**
Conclusions

- UML component and deployment diagrams for DW physical design
- Advantages:
  - Part of a DW Engineering Process based on the UML & UP
  - Traces a project from the conceptual to the physical level
  - Reduces development cost thanks to tackle implementation issues in early stages
  - Different levels of abstraction

Future work

- Index representation
- Formal definition with OCL
- Design guidelines
- CASE tool support with Rational Rose → Add-in
Physical Modeling of Data Warehouses using UML

Sergio Luján-Mora
Juan Trujillo

DOLAP 2004