Advantages of UML for Multidimensional Modeling

Sergio Luján-Mora (slujan@dlsi.ua.es)
Juan Trujillo (jtrujillo@dlsi.ua.es)
Department of Software and Computing Systems
University of Alicante (Spain)

Panos Vassiliadis (pvassil@cs.uoi.gr)
Department of Computer Science
University of Ioannina (Greece)

Content

• Introduction
• MD Modeling based on the UML
• Advantages of UML for MD Modeling
• Conclusions
Advantages of UML for Multidimensional Modeling

Content

- Introduction
- MD Modeling based on the UML
- Advantages of UML for MD Modeling
- Conclusions

Introduction

Data warehouses (DW)
Multidimensional databases (MDB)
On-Line Analytical Processing (OLAP)

Different approaches → No standard
Introduction

• **Unified Modeling Language (UML)** →
  Standard object-oriented (OO) modeling language for describing and designing software systems
  – Widely accepted
  – General
  – Visual support
  – …

• Our previous work:
  – A UML extension for MD modeling
  – Based on a three-layered schema

• In this work:
  – How our approach resolves important problems of MD modeling: multistar models, shared dimensions, multiple and alternative hierarchy levels, etc.
Advantages of UML for Multidimensional Modeling

Content

• Introduction
• MD Modeling based on the UML
  • Advantages of UML for MD Modeling
  • Conclusions

Advantages of UML for Multidimensional Modeling

MD Modeling based on the UML

• 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
MD Modeling based on the UML

- Stereotype is the main mechanism:
  - Tagged values define stereotype properties
  - Constraints specify stereotype behavior

- Different representations of a stereotype:

```
<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>
```

Advantages of UML for Multidimensional Modeling

Model definition  | Star schema definition  | Dimension/fact definition
Level 1            | Level 2                | Level 3

```
StarPackage1

DimensionPackage1

DimensionPackage2

FactPackage

<<Roll-up Tenancy>>

<<Roll-up Tenancy>>

Base1           

Base2

Base3

B

B
```

Advantages of UML for Multidimensional Modeling

MD Modeling based on the UML

Package stereotypes

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

Class stereotypes

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

Content

- Introduction
- MD Modeling based on the UML
- Advantages of UML for MD Modeling
- Conclusions
Advantages of UML for MD Modeling

1. Multistar models
2. Support for different building perspectives
3. Shared dimensions
4. Shared hierarchy levels
5. Multiple and alternative classification hierarchies

6. Heterogeneous dimensions
7. Shared aggregation
8. Derivation rules
9. Computer-Aided Software Engineering (CASE) tool support
Advantages of UML for Multidimensional Modeling

Advantages of UML for MD Modeling

- Running example: a simplified version of the warehouse example from Kimball’s *The Data Warehouse Toolkit*:
  - **Inventory snapshot**: measures the inventory levels in a regular period of time
  - **Delivery status inventory**: tracks the disposition of all the items in a delivery
  - **Transaction inventory**: every change of status of delivered products is recorded throughout the delivery flow

1. Multistar models:
   - Fact constellation
   - A single MD model has multiple facts → Multiple star schemas
Advantages of UML for Multidimensional Modeling

Advantages of UML for MD Modeling

2. Support for different building perspectives:
   - Top-down: DW $\rightarrow$ data marts
   - Bottom-up: data marts $\rightarrow$ DW

3. Shared dimensions:
   - Two or more star schemas can share some dimensions
   - Benefits:
     - Reduces the development time
     - Drill-across
     - Avoids inconsistency
Advantages of UML for Multidimensional Modeling

Inventory Delivery Status Star → Inventory Snapshot Star → Inventory Transaction Star

Advantages of UML for Multidimensional Modeling

Warehouse Dimension

Inventory Snapshot Fact → Time Dimension

Product Dimension
Advantages of UML for Multidimensional Modeling

- Inventory Delivery Status Star
- Inventory Snapshot Star
- Inventory Transaction Star
- Warehouse Dimension
  - Product Dimension
  - Time Dimension
- Vendor Dimension
- Time Dimension (from Inventory Snapshot Star)
- Warehouse Dimension (from Inventory Snapshot Star)

Advantages of UML for Multidimensional Modeling

- Inventory Delivery Status Star
- Inventory Snapshot Star
- Inventory Transaction Star
- Warehouse Dimension
  - Product Dimension
  - Time Dimension
- Vendor Dimension
- Time Dimension (from Inventory Snapshot Star)
- Warehouse Dimension (from Inventory Snapshot Star)
Advantages of UML for MD Modeling

4. **Shared hierarchy levels:**
   - Two or more dimensions share some hierarchy levels
   - Similar benefits to shared dimensions: reduces development time, avoids inconsistency, etc.
   - Salient feature: two dimensions do not need to share the whole hierarchy → Higher level of flexibility
Advantages of UML for Multidimensional Modeling

5. Multiple and alternative classification hierarchies:
   - Classification hierarchies provide the basis for the subsequent data analysis
   - **Multiple**: when a dimension has two or more classification hierarchies
   - **Alternative**: where two or more classification hierarchies converge into the same hierarchy level
Advantages of UML for Multidimensional Modeling

6. Heterogeneous dimensions:
   - A dimension that describes a large number of heterogeneous items with different attributes
   - Different categorization levels thanks to generalization/specialization hierarchies
Advantages of UML for Multidimensional Modeling

7. Shared aggregation:
   - Allows us to represent many-to-many relationships between Fact and Dimension
   - Named relationships allows us to define more than one relationship between two classes
Advantages of UML for Multidimensional Modeling

8. Derivation rules:
   • Derived attribute $\rightarrow /$
   • Derivation rules are explicitly defined
Advantages of UML for MD Modeling

9. Computer-Aided Software Engineering (CASE) tool support:
   • Add-in for Rational Rose 2002
   • A well-known CASE tool available in the market
   • We do not have to develop our own CASE tool from scratch
Advantages of UML for Multidimensional Modeling

Content

• Introduction
• MD Modeling based on the UML
• Advantages of UML for MD Modeling
• Conclusions

Conclusions

• UML packages: clean MD design of huge and complex systems
• UML importation mechanism:
  – Simplifies the models
  – Avoids redundancy, inconsistency, and ambiguity
Advantages of UML for Multidimensional Modeling

Conclusions

• Future work:
  – Implementation of MD models \( \rightarrow \) ORDB and OODB
  – Different implementation strategies: snowflake, constellation, etc.
  – Incorporate UML use cases into MD conceptual modeling