Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP (DOLAP 2004), p. 48-57, Washington D.C. (USA), November 12-13 2004.

Department of Software and Computing Systems

Universitat d'Alacant
Universidad de Alicante

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

Physical Modeling of Data Warehouses using UML

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

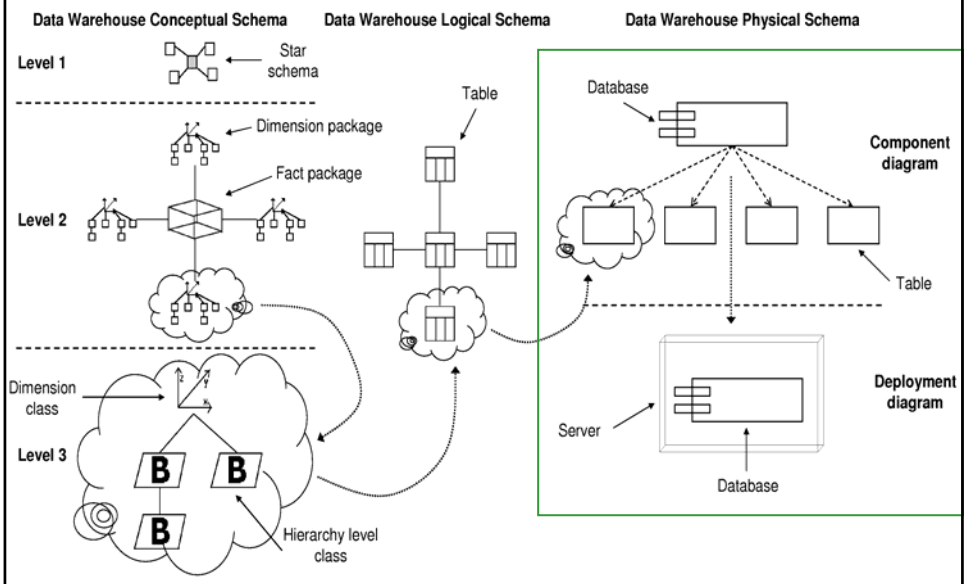
- Implementation decisions:
  - Storage in different disks
  - Replication
  - Vertical and horizontal partitioning
  - Influence performance and maintenance
  - …
- Solution:
  - Tackle physical design from early stages
    - Allows the designer to anticipate physical design decisions
    - Reduce development time and cost

Motivation

- 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)



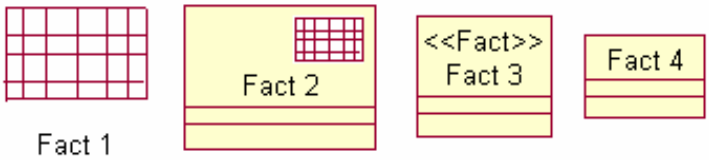
Contents

- 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

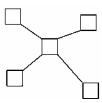
UML extension mechanisms



Icon Decoration Label None

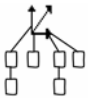
UML extension mechanisms

Package stereotypes



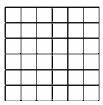
StarPackage
(Level 1)

FactPackage
(Level 2)

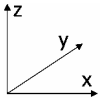


DimensionPackage
(Level 2)

Class stereotypes



Fact
(Level 3)



Dimension
(Level 3)

Base
(Level 3)

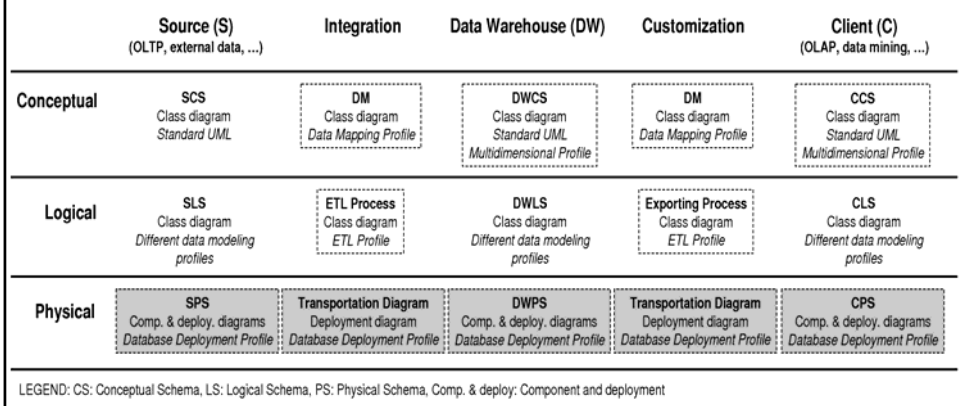
Contents

- 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

DW diagrams

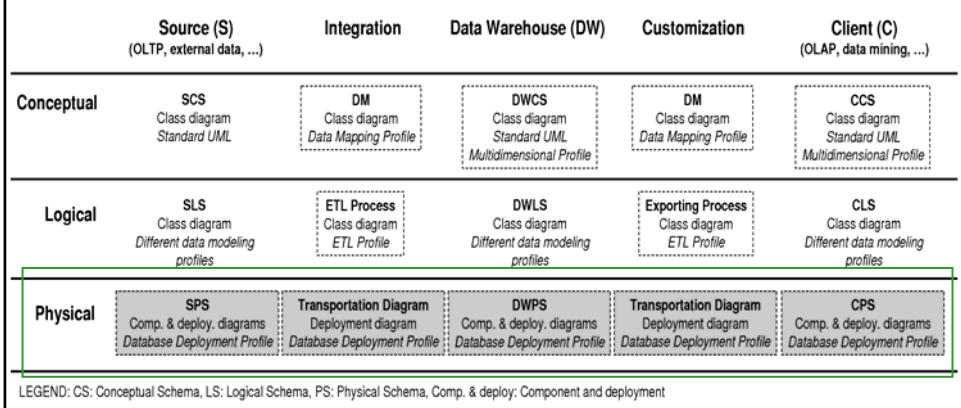
- Stages:
  - Source: data sources (OLTP, external data sources, etc.)
  - Integration: mapping between source and data warehouse
  - Data Warehouse: structure of the DW
  - Customization: mapping between data warehouse and clients' structures
  - Client: structures used by the clients to access the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

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

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

  } Component and deployment diagram

  - Integration Transportation Diagram
  - Customization Transportation Diagram

  } Deployment diagram

DW physical design

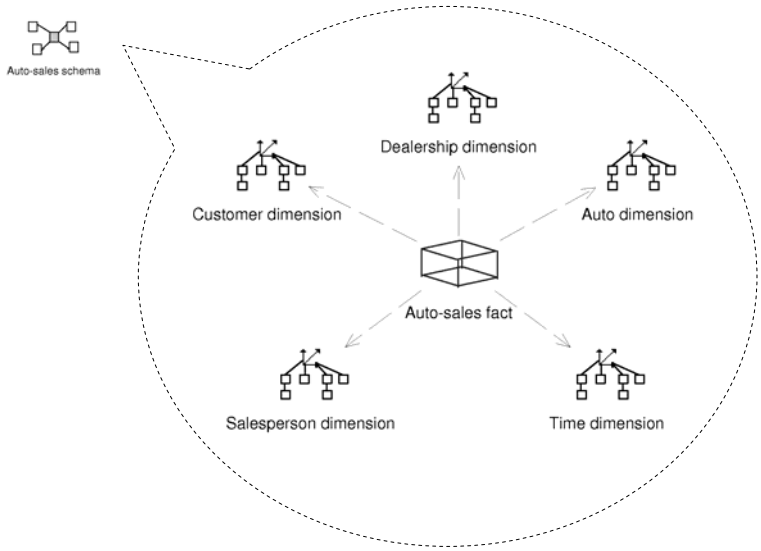
- Example:
  - DW with daily sales of a company that sales automobiles (cars and trucks)
  - Dimensions of analysis: automobile, customer, dealership, salesman, time
  - Two data sources:
    - Sales server: transactions and sales
    - CRM server: customers
  - Different final users' requirements:
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

Level 1: Model definition

Auto-sales schema

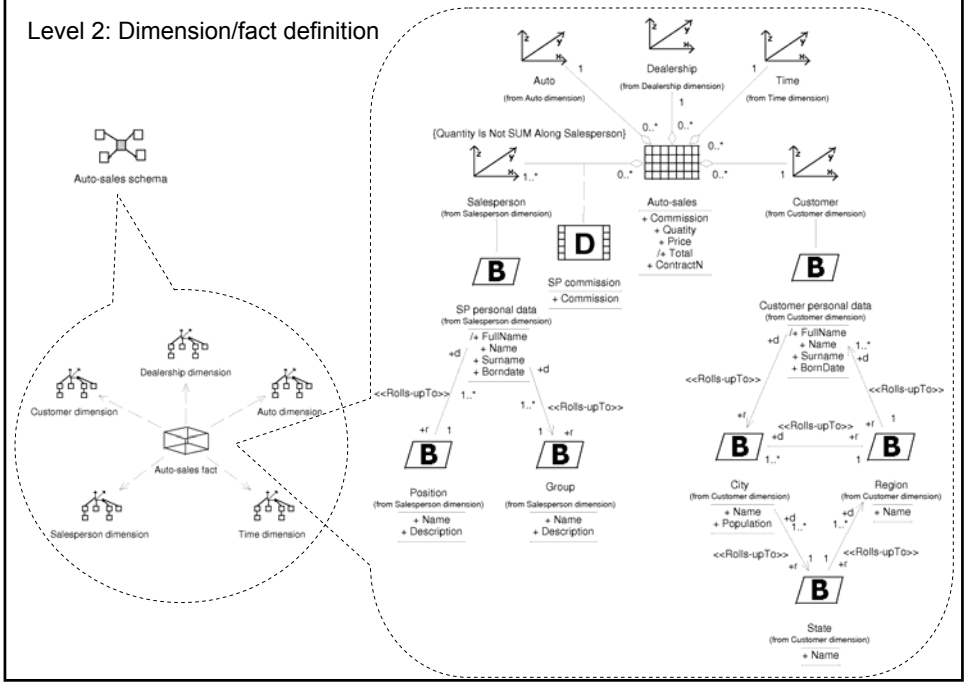
Physical Modeling of Data Warehouses using UML

Level 2: Star schema definition

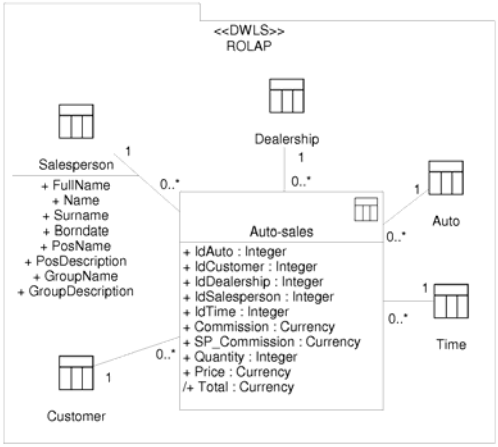


Physical Modeling of Data Warehouses using UML

Level 2: Dimension/fact definition



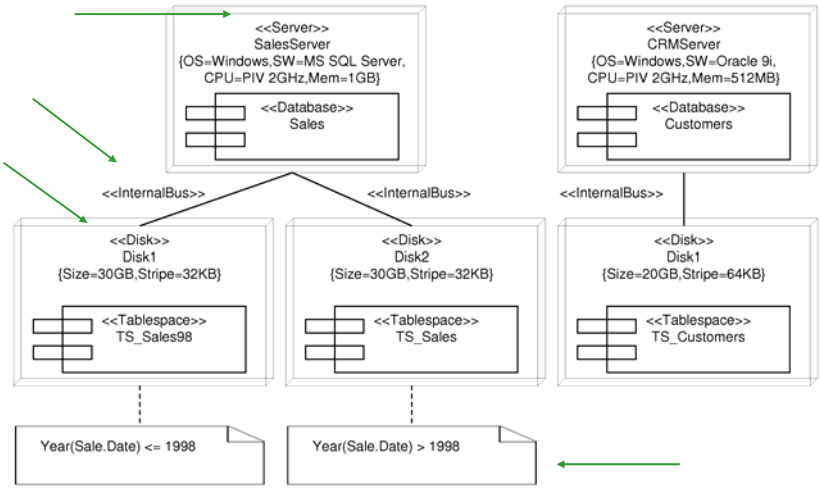
DW physical design



Data Warehouse Logical Schema: ROLAP

Physical Modeling of Data Warehouses using UML

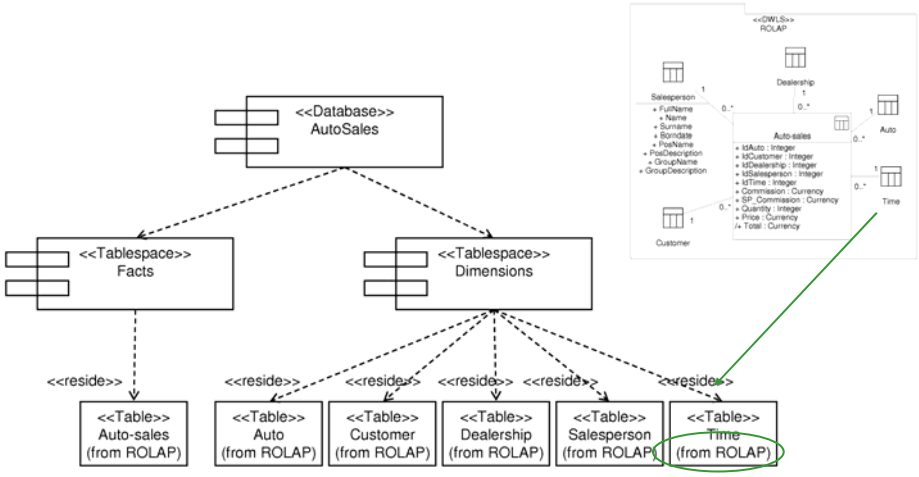
DW physical design



Source Physical Schema: deployment diagram

DW physical design

DWLS

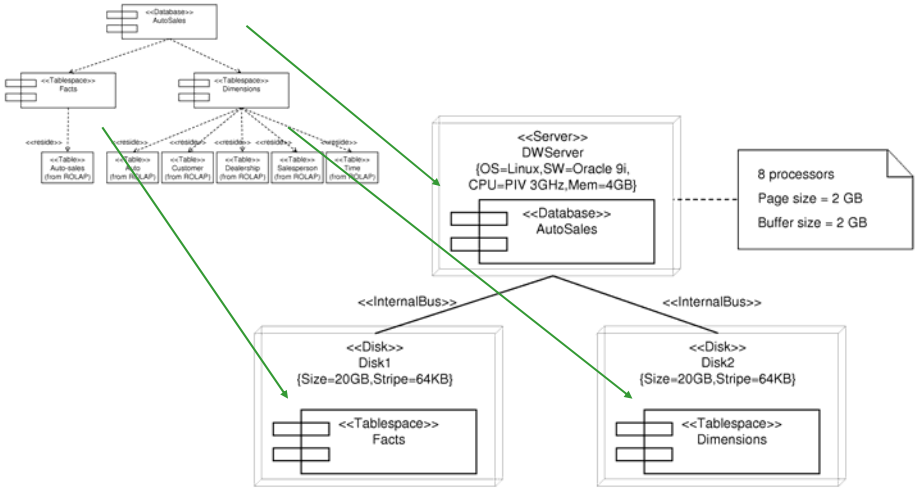


Data Warehouse Physical Schema: component diagram

Physical Modeling of Data Warehouses using UML

DWPS

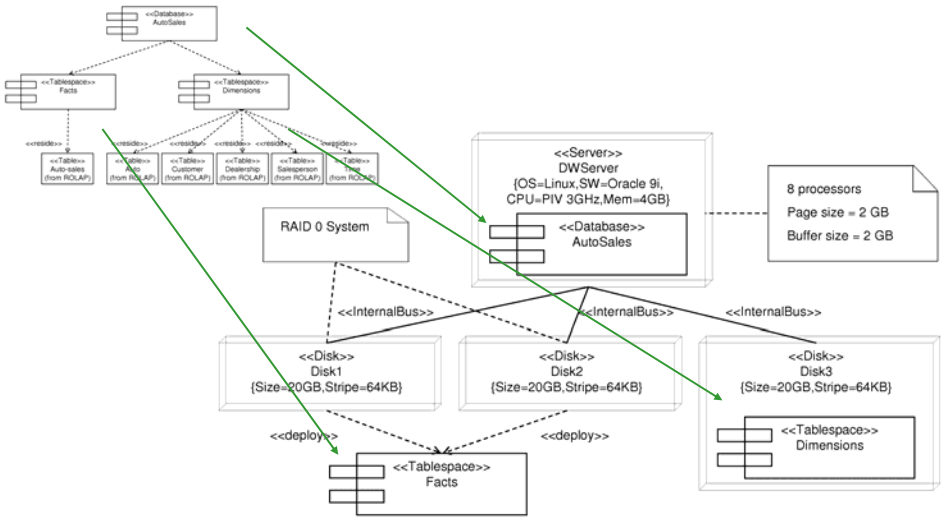
DW physical design



Data Warehouse Physical Schema: deployment diagram

DWPS

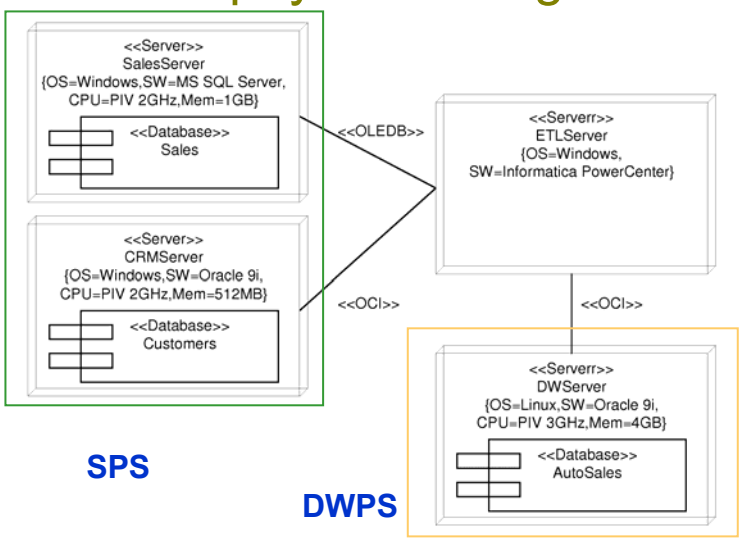
DW physical design



Data Warehouse Physical Schema: deployment diagram

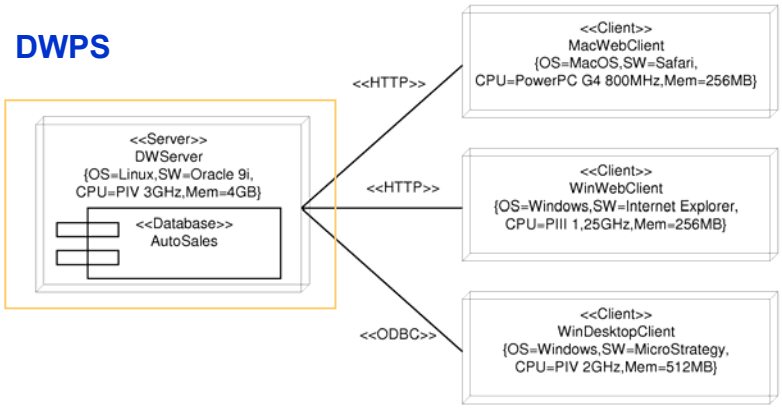
Physical Modeling of Data Warehouses using UML

DW physical design



Integration Transportation Diagram: deployment diagram

DW physical design



Customization Transportation Diagram: deployment diagram

Contents

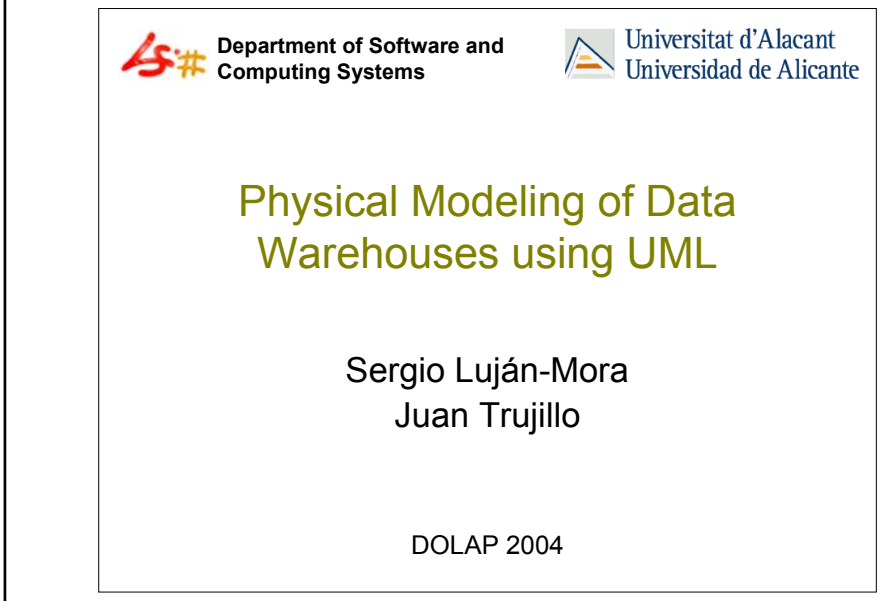
- Motivation
- UML extension mechanisms
- DW diagrams
- DW engineering process
- Applying modeling schemas
- **Conclusions and future work**

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


Department of Software and Computing Systems
Universitat d'Alacant
Universidad de Alicante
Physical Modeling of Data Warehouses using UML
Sergio Luján-Mora
Juan Trujillo
DOLAP 2004