DATABASES:
TEMI AVANZATI

Roberto Basili
Corso di Basi di Dati e Conoscenza
a.a. 2013-14

Outline

- Motivations for the second emisemester

- Main topics
  - From RDMS to deductve DBs
  - Deduction, Inference and Programming
  - Knowledge Representation
  - Multimedia DBs
  - Management of Unstructured Data

- Projects of the Course

Motivations

- Database = a piece of software to handle data:
  - store,
  - maintain, and
  - query

- Most ideal system situation-dependent
  - data type: simple / semi-structured / complex / …
  - types of queries: simple lookup / analytical / …
  - type of usage: multi-user / single-user / distributed / …
  - …

Motivations

- Relational databases are tuned towards:
  - simple data, highy structured
  - simple, ad-hoc queries
  - multiple users

- Other models are more suitable for other types of data
  - Object-Oriented,
  - Deductive,
  - Semi-Structured Databases,
  - Data warehouses

Motivations

- Study different data models
  - Advantages, disadvantages
  - What is important at the conceptual level?
  - What's underneath?
    - Technology
    - Social needs

Motivations

- First emisemester:
  - Relational database model
- Motivations:
  - why they should be extended
  - Which further conceptual notions are required
  - How they interact with the existing core notions (e.g. SQL interpretation)
- Student should be then able to:
  - Frame DB technology in a wider perspective
  - Join groups dealing with more specific technologies (e.g. Semantic Web)

Organization

Two lesons per week …

- Wed 11:30 → 13:30 Aula 9 NED
  - lessons
  - opportunity to practice, ask questions
- Gio 9:00 → 11:30 Aula 4 PP
  - New topics
  - Solve exercises

Practical Organization

- Important information

  http://ai-nlp.info.uniroma2.it/basili/didattica/DKBs1/

- basili@info.uniroma2.it

Practical Organization

- Course material: Books
  - Brachman e Levesque: "Knowledge representation and reasoning", Morgan Kaufmann, Cap. 2,3,4,10. ISBN 1-55860-932-6.
  - Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (Eds.). 2003. The Description Logic Handbook: Theory, Implementation, and Applications. Cap. 1, 2, 10. Cambridge University Press, New York, NY, USA.
  - V. S. Subrahmanian. 1998. Principles of Multimedia Database Systems. Part I, II. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Tech Reports, Papers e Slides distribuited through the Web site

Course Topics

- Limitations of the relational model
- Knowledge Representation and Reasoning
- Deductive databases
- Further Adanced Topics
  - Object-Oriented Databases
  - Data Warehousing & OLAP
  - Semi-Structured data
- Multimedia DBs

Project

- Pick one of the topics:
  - deductive databases / rule-based systems
  - object-oriented databases
  - Multimedia and semi-structured databases
- Formulate your own project
  - illustrating the different course concepts
  - agree upon it with the responsible
  - show you mastered the technology

Limitations of the relational model

- Not every query can be expressed
  - Transitive closure cannot be expressed in Relational Algebra
    - Give all cities reachable from Antwerp by plane
    - Give all smallest components of a part
    - Give all decendants of person X
  - Not even if you're very smart …
    - proof
  - Extension to other relational query languages

Deductive Databases

- Motivation is two-fold:
  - add deductive capabilities to databases; the database contains:
    - facts (intensional relations)
    - rules to generate derived facts (extensional relations)

    Database is *knowledge base*
  - Extend the querying
    - datalog allows for recursion

Deductive Databases

- Datalog as engine of deductive databases
  - similarities with Prolog
  - has *facts* and *rules*
  - rules define -possibly recursive- views
- Semantics not always clear
  - safety
  - negation
  - recursion

Deductive Databases

```
g(a,b). g(b,c). g(a,d).
reach(X,X) :- g(X,Y).
reach(X,Y) :- g(X,Y).
reach(X,Z) :- reach(X,Y), reach(Y,Z).
node(X) :- g(X,Y).
node(Y) :- g(X,Y).
unreach(X,Y)    :-    node(X), node(Y),
                      not reach(X,Y).
```

Deductive Databases

- Relevant topics :
  - How are relational (i.e. extensional) DBs used to support recursive computation
  - How to handle negation and recursion in the same program
  - How to efficiently evaluate Datalog queries

OO Databases

- Many applications require the storage and manipulation of complex data
  - design databases
  - geometric databases
  - …
- Object-Oriented programming languages manipulate complex objects
  - classes, methods, inheritance, polymorphism

OO Databases

- Very simple example:
  - Class book
    - set of authors
    - title
    - set of keywords

Extremely simple to model in OO language

Hard in relational database!

OO Databases

- In many applications persistency of the data is nevertheless required
  - protection against system failure
  - consistency of the data

- Mapping: object in OO language → tuples of atomic values in relational database is often problematic

OO Databases

- Either we ignore the multivalued dependencies

Title	Author	Keyword
Database System Concepts	**Silberschatz**	**Database**
Database System Concepts	**Korth**	**Database**
Database System Concepts	**Sudarshan**	**Database**
Database System Concepts	**Silberschatz**	**Storage**
Database System Concepts	**Korth**	**Storage**
Database System Concepts	**Sudarshan**	**Storage**

- This table is in 3NF, BCNF

OO Databases

- Or we go to 4NF

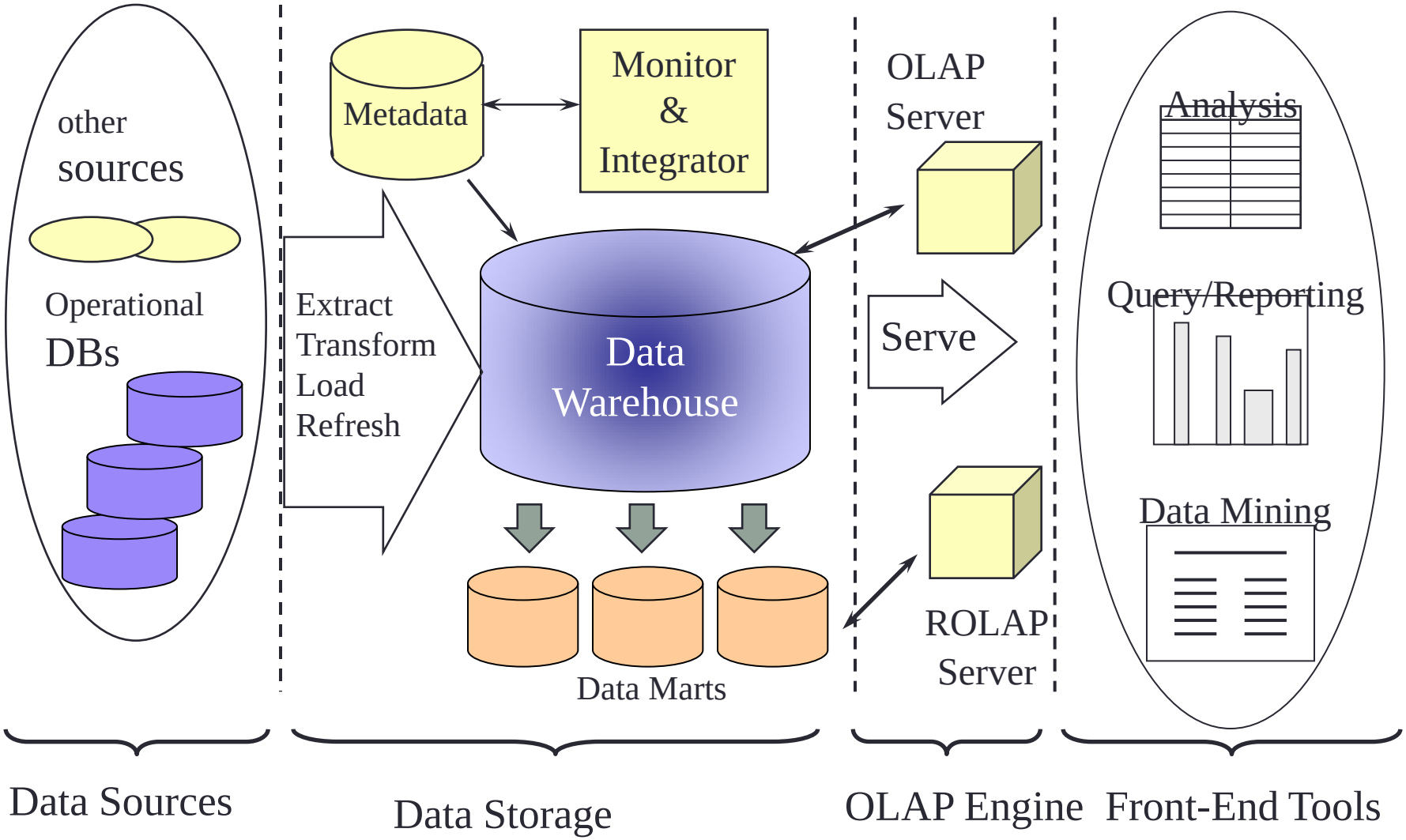
Title	Author
Database System Concepts	**Silberschatz**
Database System Concepts	**Korth**
Database System Concepts	**Sudarshan**

Title	Keyword
Database System Concepts	**Database**
Database System Concepts	**Storage**

OO Databases

- Basically OODB = persistent OO programming language
  - Very important concept
  - rather uninteresting scientifically

Data Warehousing & OLAP



Data Warehousing & OLAP

Transaction processing	Flight reservations
• Operational setting • Up-to-date = critical	• ticket sales • do not sell a seat twice
• Simple data	• reservation, date, name
• Simple queries; only « touch » a small part of the database	• Give flight details of X List flights to Y

Data Warehousing & OLAP

Decision support

- Off-line setting
- « Historical » data
- Summarized data

- Integrate different databases
- Statistical queries

Flight company

- Evaluate ROI flights
- Flights of last year
- # passengers per carrier for destination X
- Passengers, fuel costs, maintenance info
- Average % of seats sold/month/destination

Data Warehousing & OLAP

- Relevant topics :
  - Conceptual models for decision support
  - Database explosion problem
  - Efficient implementation strategies
    - indexing, view materialization

Summary

- Relational model has limitations
  - simple queries
  - simple data
- OODBs allow complex data types
- Deductive databases, datalog complex queries
- Somewhere in-between: datawarehouses and OLAP
  - special requirements, special datastructures