Data clear

Hayaat HEBIRET M1 AIGLE
10 Février 2021

(1) Managing data through the lens of an ontology

Ontology based data management
knowledge representation =/= automated reasoning

Why:
Proliferation of data sources VS cooperating environment (volume vs velocity)

Problem -> complex database and heterogeneous data sources

- **Accessing and querying data** (maintenance, adapt,formulation)
- **Data quality** (hight value information services : consistency, accuracy, completeness, confidentiality, integrity)
- **Open data** (transparency, structured description,where is it from)
- **Process and service specification** (what does it do/means, how does it affect the data)

Ontology based data management (OBDM) is extremely difficult.
"Applying suitable techniques from area of knowledge of representation and reasoning in AI for a new way to achieve data governance"

OBDM/OBDA

3 Layers <O,S,M>:

- **Data Layer** (existing data sources that are relevant for the organisation)
- **Ontology** (representation of the domain of interest for the organisation specified by means of a formal and hight level description of the statics and dynamic aspects)
- **Mapping** (declarative assertions specifying how the available sources in the data layer and computational resources used relate to the ontology)

It's a sophisticated form of information integration where the usual global schema is replaced by the conceptual model of an application domain formulated as an ontology.

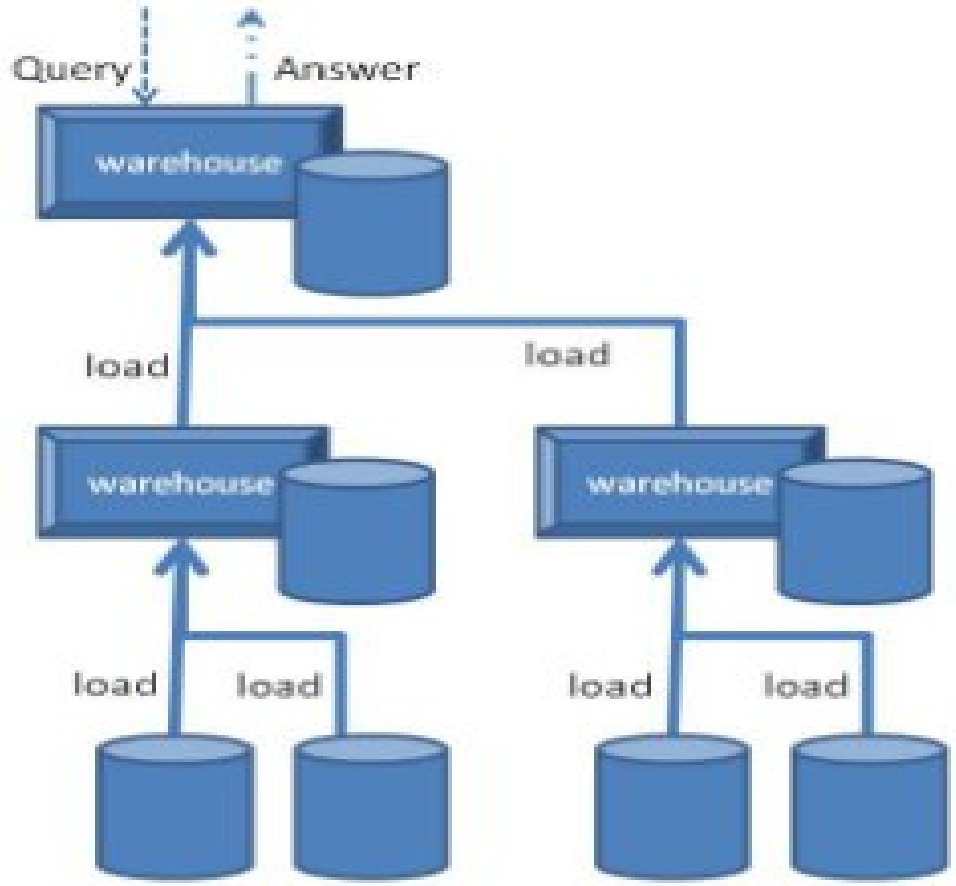
Query answering :

- query over ontology (different from traditional model of logical theory)
- Various model of the whole system (incomplete info)
- automated deduction technique are relevant

(2) Web data management

Data integration : providing a uniform access to a set of autonomous/heterogeneous data sources in a particular application domain

Warehousing : data extracted from data sources ahead of a query, transformed, loaded in the warehouse.
(used for data mining,complex query)



Virtual data integration : access to "fresh" informations via mediator approach

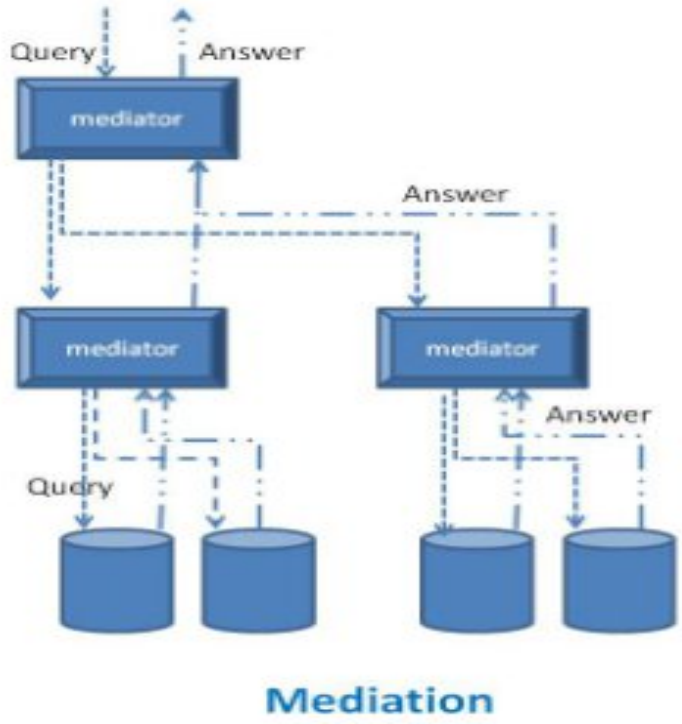
Mediator : designing a mediated/global schema as a unique entry point on which global queries are posed by users.

Main issue: semantic mapping between schemas of data sources and global schemas.

Wrapper: small programs that translate local relational query to an appropriate request understood by specific data sources ans transform the result into a relation.
'allow the mediator to see each data as relational no matter which actual format it uses)

Steps:

1. User query posed over global schema (independently of data sources)
2. Transformation to local queries
3. Global query combine the data provided by the sources
4. Queries are optimised ans transformed to query plans
5. The local query plans are executed
6. Their result is combined by the global query plan



Mediation

Approach 1: Global As View (GAV)

The global relations are defined as views over the local relations (query processing is simple)

Pro:

- Query rewriting by unfolding each atom in the query (use of disjunctions/conjunctions) is simple

Cons:

- Adding/removing data sources imply to revise all the GAV mapping defining the global schema
- When a new data source arrives we have to consider how it may be combined with all the existing data sources to produce tuples of any global relation

Approach 2: Local As View (LAV)

The local relations are defined as views over the global relations (flexibility and robustness)

Pro:

- Adding/removing data sources does not affect the mediated schema (we can define without knowing the sources)
- When a new data source arrives/leaves it has no impact on the rest (We can define the mapping without knowing the sources)

Cons:

- Building the rewriting require more work than the simple unfolding of the GAV approach

Bibliography

1. Maurizio Lenzerini : Managing data through the lens of an ontology. Dipartimento di Ingegneria Informatica, Automatica e Gestionale Antonio Ruberti Universita di Roma La Sapienza,
   https://core.ac.uk/download/pdf/188826491.pdf

2. Serge Abiteboul, Ioana Manolescu, Marie-Christine Rousset, Pierre Senellart : Data Integration
   http://webdam.inria.fr/Jorge/files/wdm-data-integration.pdf