

Data Integration

* Explain Ontology-based mediators.

=> Ontology based mediator are used in data integration system to connect multiple data source.

They rely on an ontology to define the relationships and constraints between different data sources.

It is used to combining information from different, data sources into a global schema.

It enables system to share and interpret data based on a common understanding of terms.

Using ontologies, we can represent relationship and constraints across different datasets.

-> Working :

1 Mapping : The mediator uses mapping between the global schema and individual source schema.

2 Rewriting Queries: When a user queries the global schema, the mediator rewrites the query

3 Reasoning: The mediator may perform logical reasoning over the ontology to get new information

4 Combining Results: Result from different data sources are combined and presented in a single integrated system

-> There are main Two types of the Ontology-based mediator

(a) Global-as-View

(b) Local-as-View

(a) Global-as-View:

The Global schema is defined as a view over the data sources schema.

Mappings specify how each global schema element is computed from data sources.

In Global-as-View mediator, Global schema is less Flexible and required simpler query rewriting.

Simple implementation and efficient for simple queries.

-> Working:

(i) Schema Definition: The global schema is defined based on the underlying data sources.

(ii) Query Translation: When a user queries the global schema than the mediator translate the query into sub-query.

(iii) Data Retrieval: The mediator executes the generated sub-queries against data source.

(iv) Result Integration: Retrieved results from a various sources and combined

(v) Static Mapping: The mappings are predefined which makes the system less Flexible.

Ex. Source 1:

student (ID, Name, DOB)

Source 2:

Course (CourseID, CName, ID)

Global Schema:

StudentCourse (ID, Name, CName)

(b) Local-as-View:

The data sources are described as views over the global schema.

Mapping specify how data sources can be derived from global schema.

The Local-as-View mediator is more flexible and required more complex query.

LAV is suitable for environments where data sources are dynamic.

-> Working:

(i) Source Descriptions: Each data source is described in term of global schema.

cii) Query Rewriting: It is rewritten into queries over the individual data sources.

ciii) Subquery Generation: The mediator generates one or more subqueries that target the relevant data sources.

civ) Data Retrieval: The mediator executes the generated queries on the corresponding data sources.

cv) Result Aggregation: The result from the different data sources are aggregated.

* Explain Peer-to-Peer Data Management System.

=> Peer-to-Peer Data Management Systems provides a decentralized approach to the data integration.

P2PMS allowing multiple data source to collaborate without the need of central authority.

In PDMS, there is no central mediator instead each data source which called as peer.

Each peer can manage its own data while serving to other peer.

Each peer can act as both a data server and a mediator.

The PDMS architecture allows for dynamic participation which means peer can join or leave the system at any time.

PDMS allows peers to maintain their own schemas.

Each peer defines mappings that relates its schema to those of other peers.

Mappings can represent various relationships such as a one-to-one, one to many or many-to-many.

-> Advantages:

- 1 Scalability: The decentralized system allows an easy addition of new data source without complete reconfiguration of the system.
- 2 Flexibility: Peer can modify their schemas or mappings independently.
- 3 Robustness: The absence of the central point make system more powerfull.
- 4 Dynamic Participation: New peers join or existing peer leaving in the system without major disruptions.

-> Disadvantages:

- 1 Complexity
- 2 Access Control and Security
- 3 Performance Overhead
- 4 Consistency Management