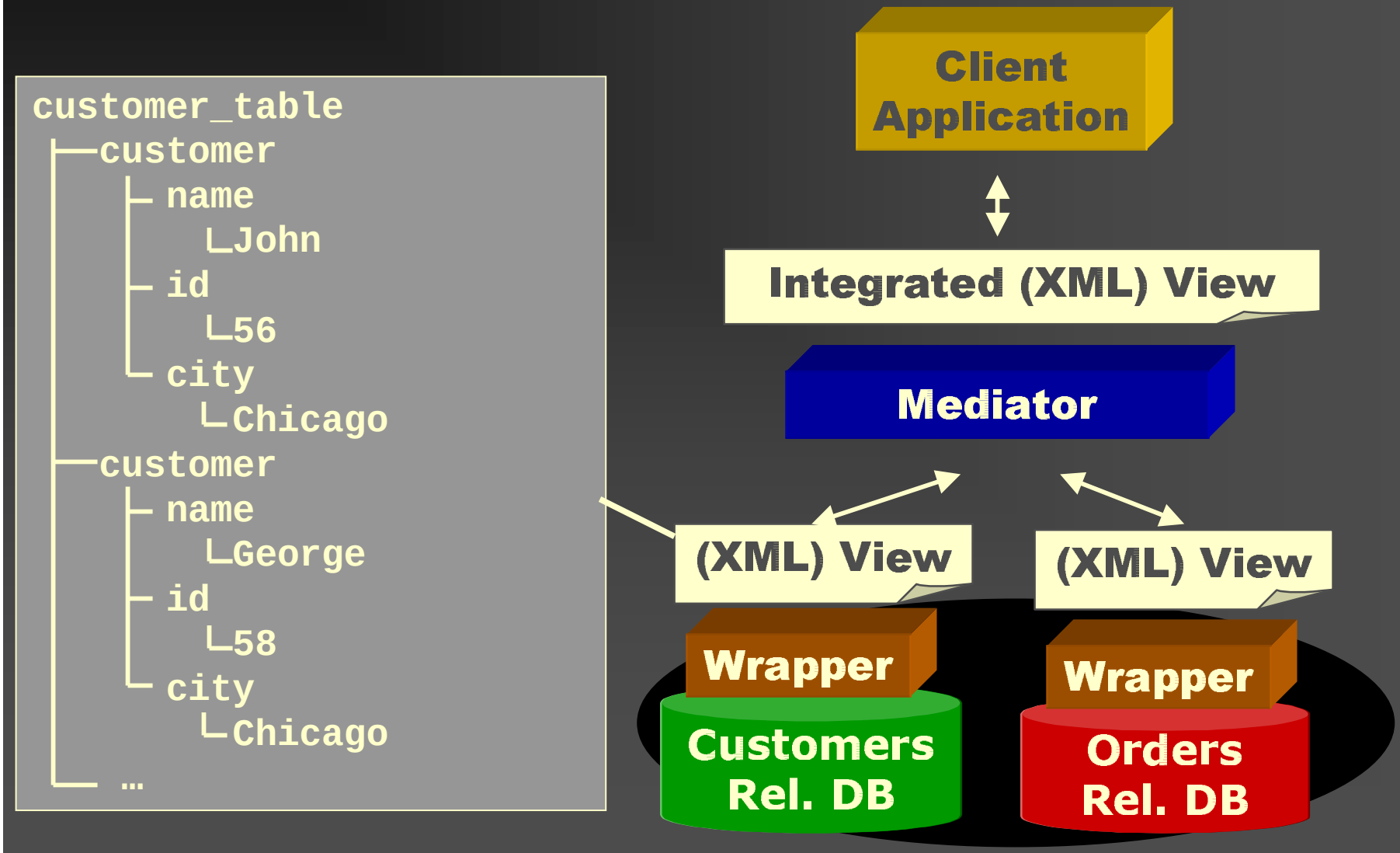
XML View-Based Mediation: TSIMMIS & Beyond

Yannis Papakonstantinou
UCSD

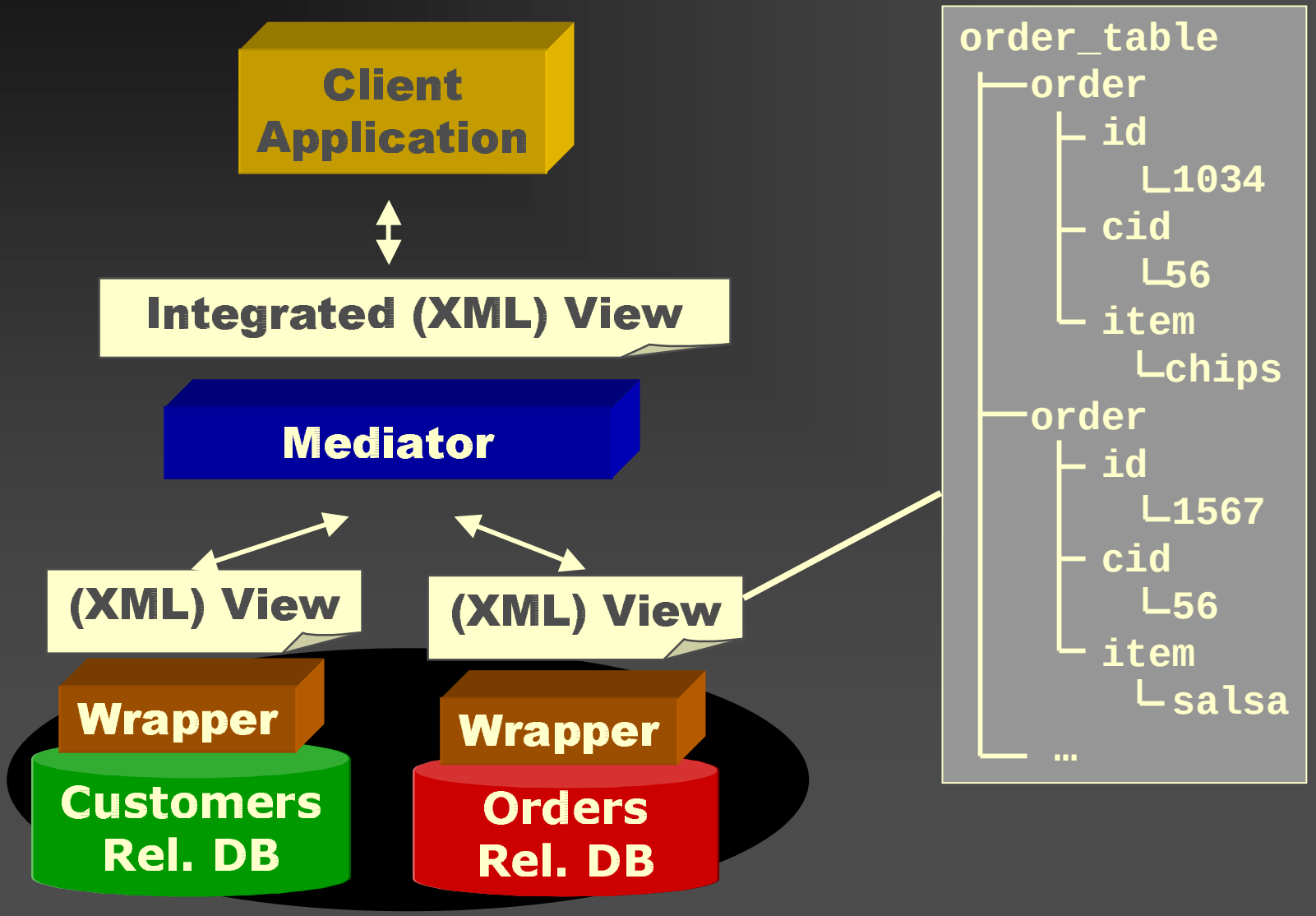
Overview

- Virtual XML (Semistructured) Views in Data Integration
  - Global-As-View (TSIMMIS, MIX, Enosys Software)
- Query Processing Issues in Mediators
- Retrospective

TSIMMIS View-Based Approach: Wrappers Export Source Views



Wrappers Export Source Views



Mediators Export Inte s, Tailored to Applicatio

MSL (oriented to views)
- id's as skolems
- silly angles <foo … >

XQuery defines integrated view as f(source views)

lient
ication

Integrated (XML

Mediator

(XML) View

(XM

Wrapper

Customers Rel. DB

Wrapper

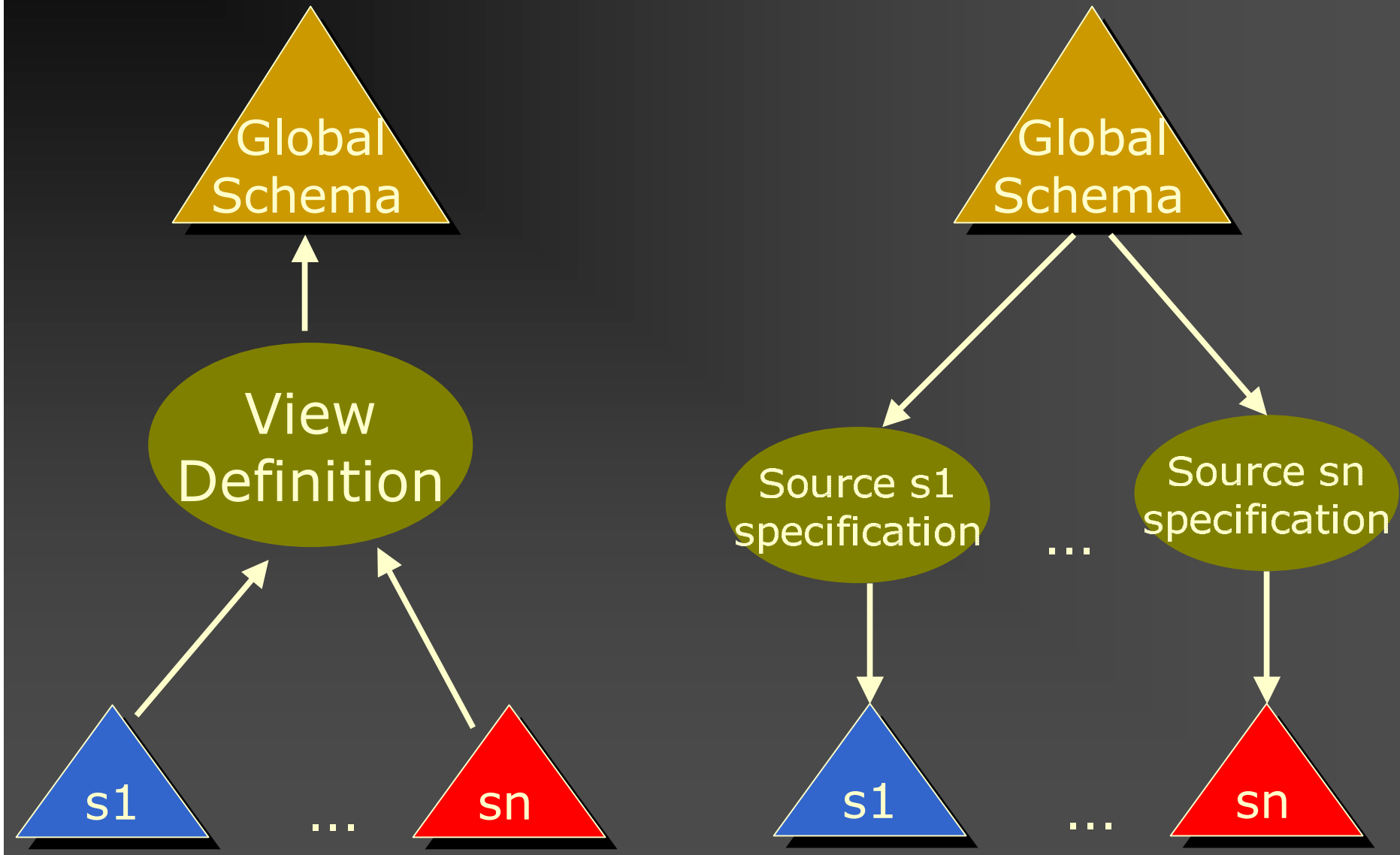
Orders Rel. DB

```
customers
├─ customer
│   ├─ name
│   │   └─ John
│   ├─ id
│   │   └─ 56
│   └─ city
│       └─ Chicago
└─ orders
    ├─ order
    │   ├─ id
    │   │   └─ 1034
    │   └─ item
    │       └─ chips
    ├─ order
    │   …
    └─ customer
        …
```

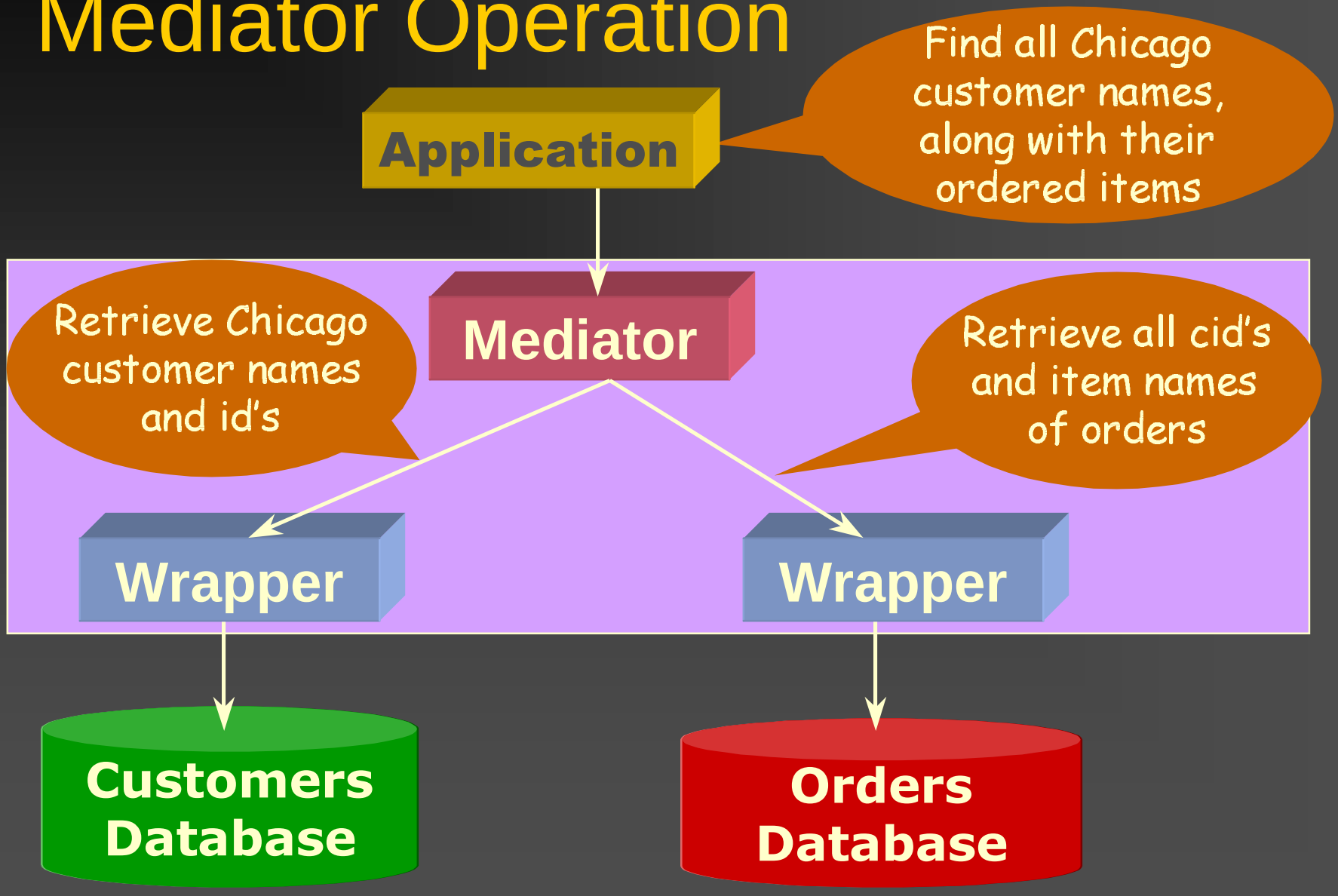
```
customer_table
├─ customer
│   ├─ name
│   │   └─ John
│   ├─ id
│   │   └─ 56
│   └─ city
│       └─ Chicago
```

```
├─ order
│   ├─ id
│   │   └─ 1034
│   ├─ cid
│   │   └─ 56
│   └─ item
│       └─ chips
```

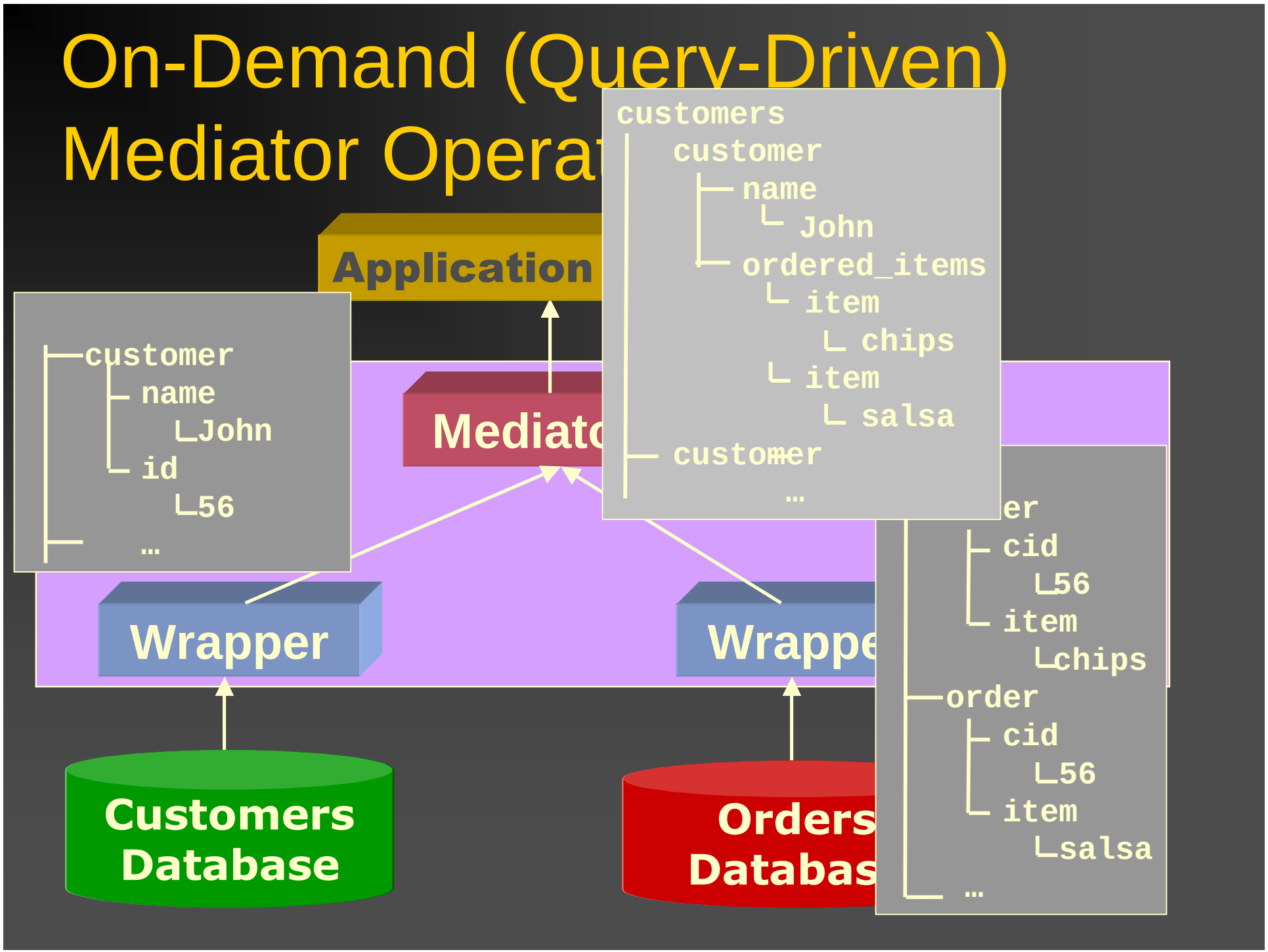
Global-As-View Vs Local-as-View



Virtual Views:Query-Driven Mediator Operation



On-Demand (Query-Driven) Mediator Operat



Multiple Plans are Possible

- Retrieve customers
- For each customer find matching orders

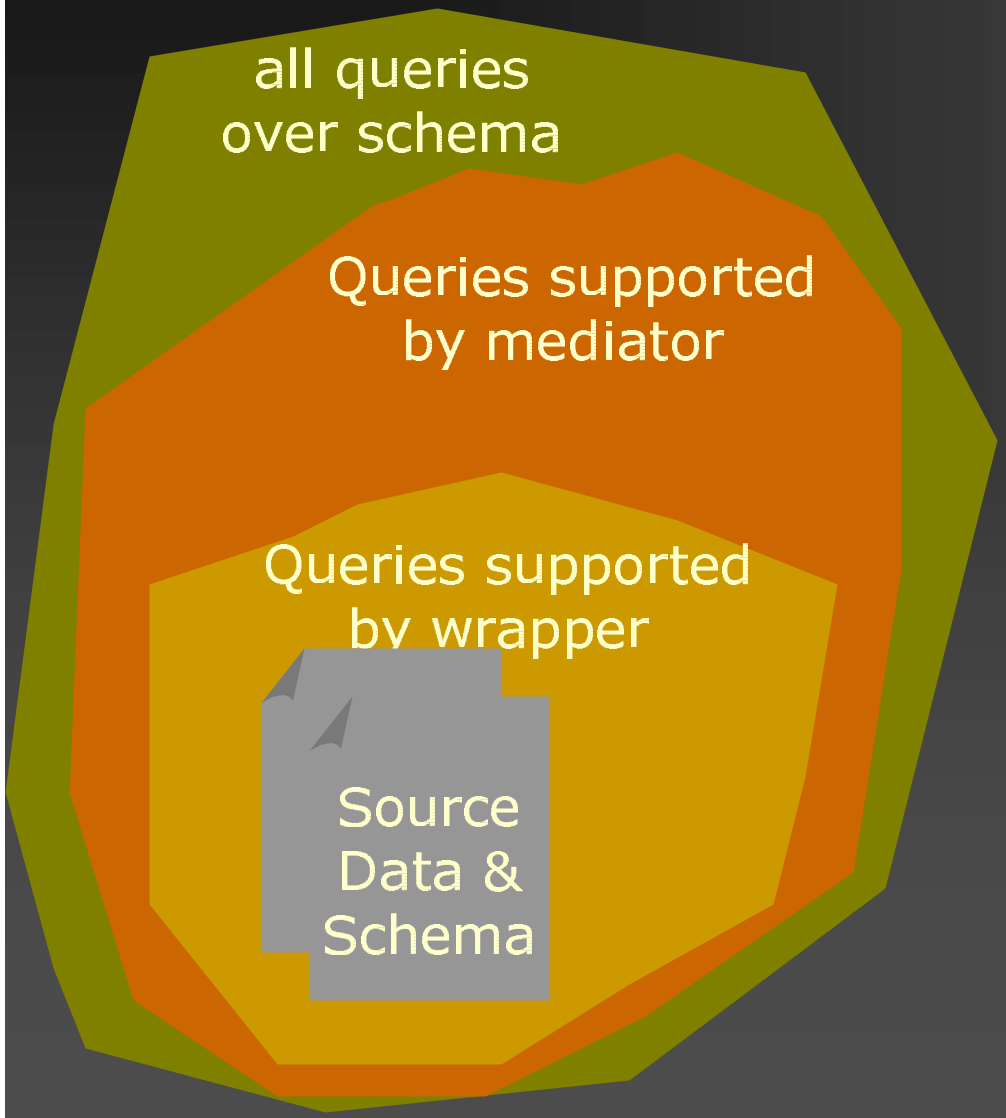
A New Kind of Query Processing Problem

- Build and Run “Optimal” Plan
  - Consisting of operators that
    - Collect source info using supported queries and commands
    - Combine info into XML result

Challenges in Query Processing & Optimization

- Operate within the Limited and Different Capabilities of the Sources
  - Describe sets of supported queries
  - Use most efficient supported queries
- Optimize plans/queries sent to sources
  - Estimate Costs of Plans
  - Beyond Conjunctive Queries
  - Compose Queries/Views Efficiently
  - Adapt Plans Along the Way
- Schema inference & optimization
- Iterator models
  - Combine navigation & querying

From Limited Wrappers to Efficient Plans for Extended Query Sets

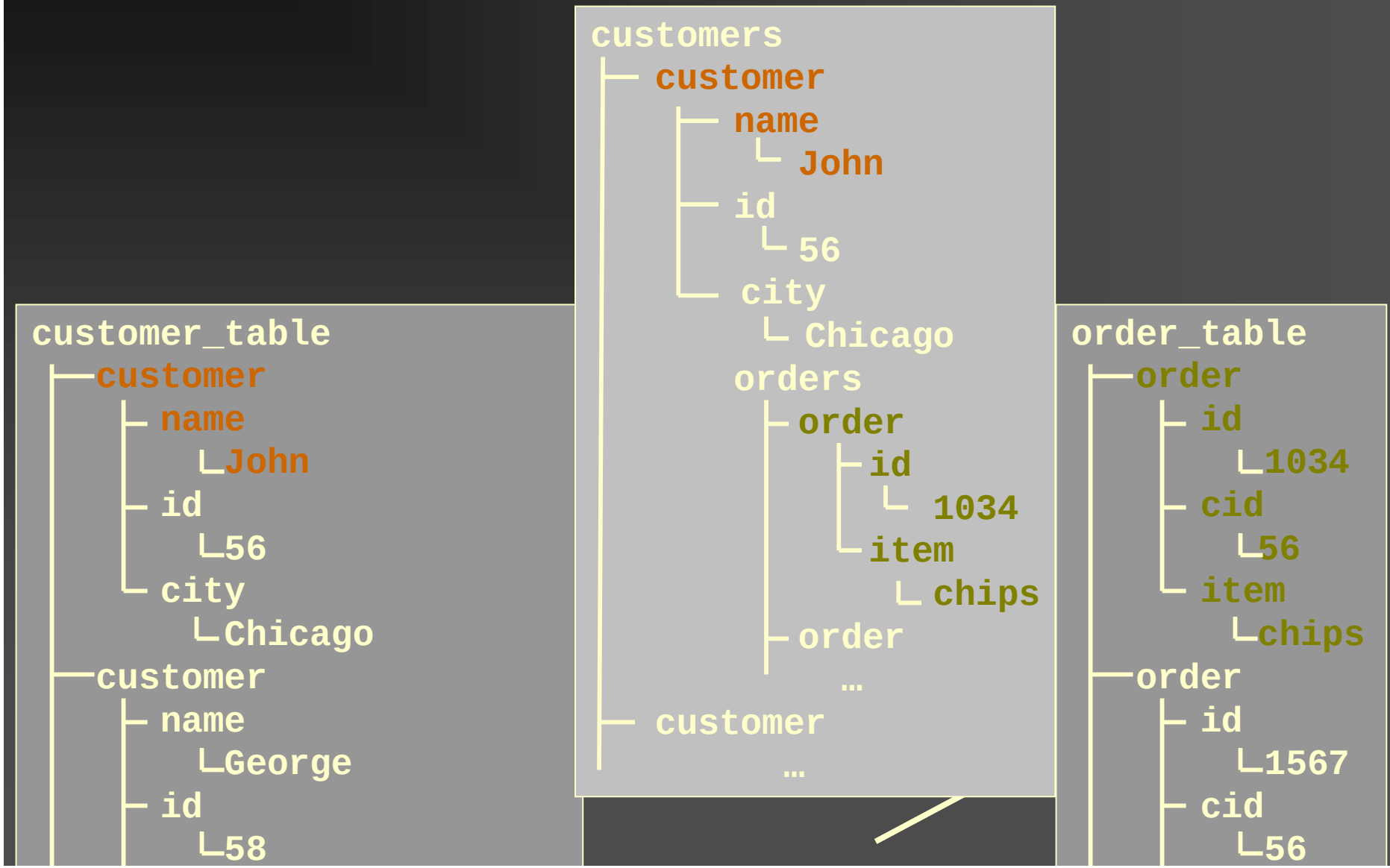


- Answering Queries Using Views
- But with Infinite Sets of Views
- And not only conjunctive
- Increasing Relevance due to Web Services

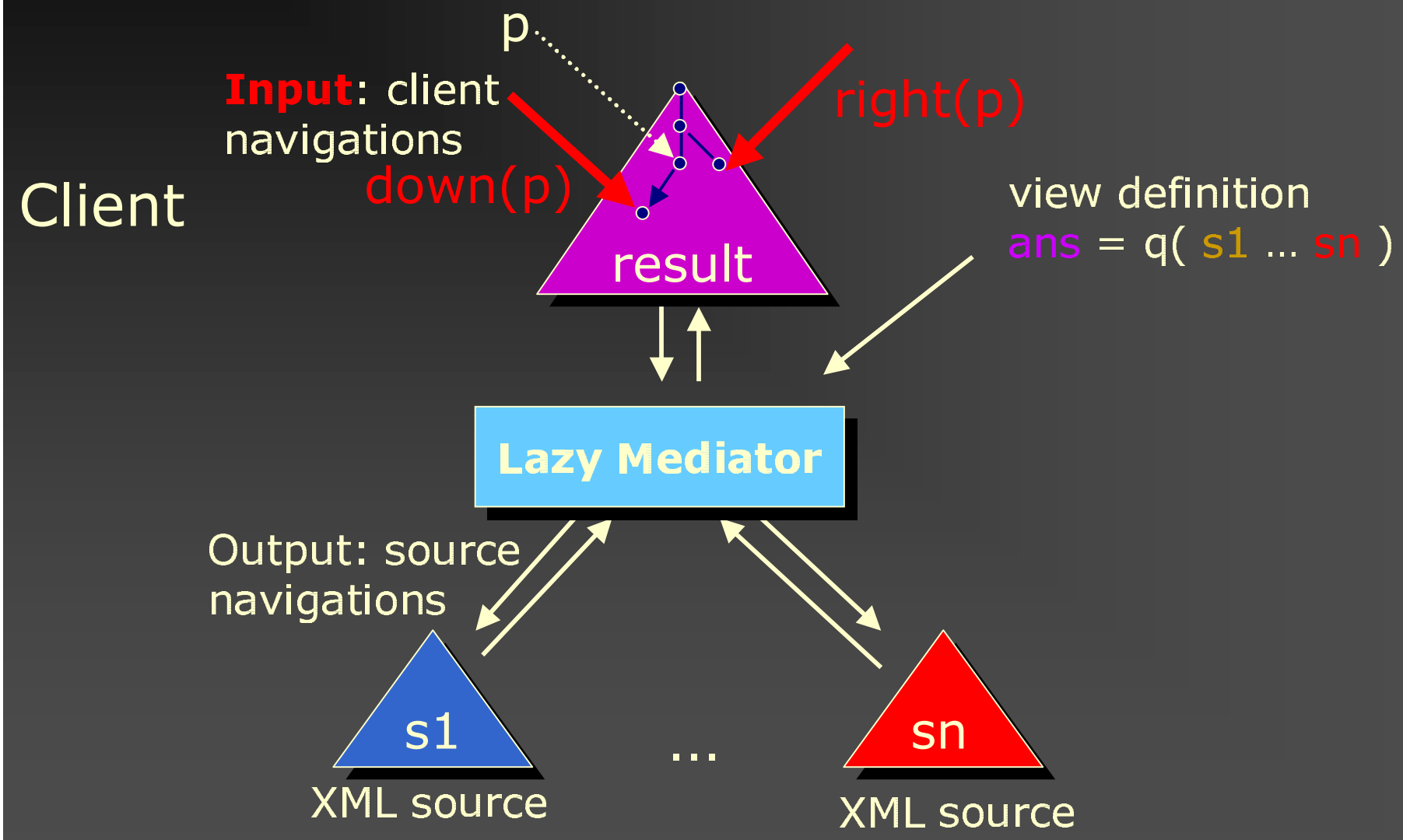
Challenges in Query Processing for Distributed XML Queries

- Operate within the Limited and Different Capabilities of the Sources
  - Describe sets of supported queries
  - Use most efficient supported queries
- Optimize plans/queries sent to sources
  - Estimate Costs of Plans
  - Adapt Plans Along the Way
  - Beyond Conjunctive Queries
  - XQuery processing
- Schema inference & optimization
- Iterators for streaming and navigation pull
  - Combine navigation & querying

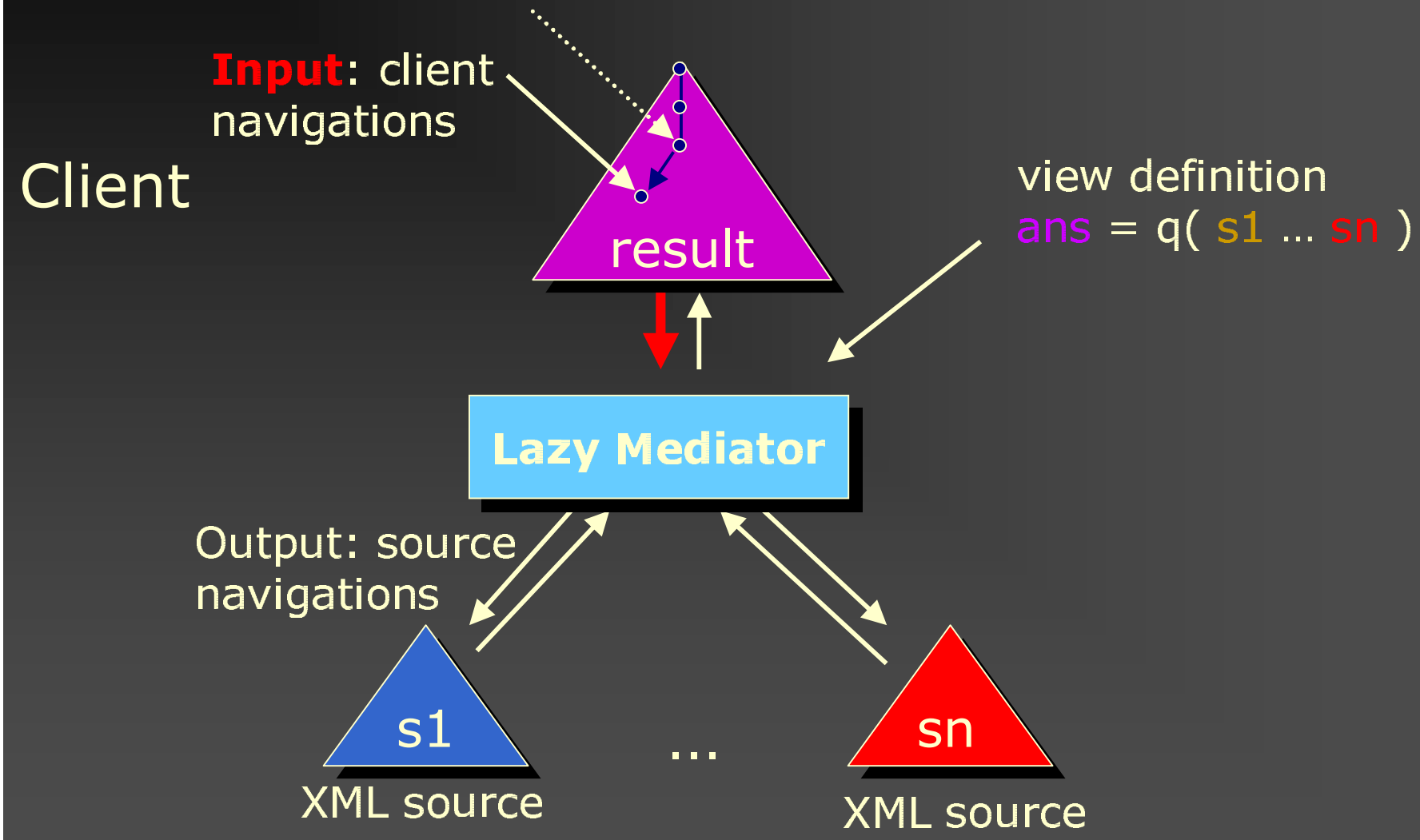
Navigation-Driven Evaluation of Query Result



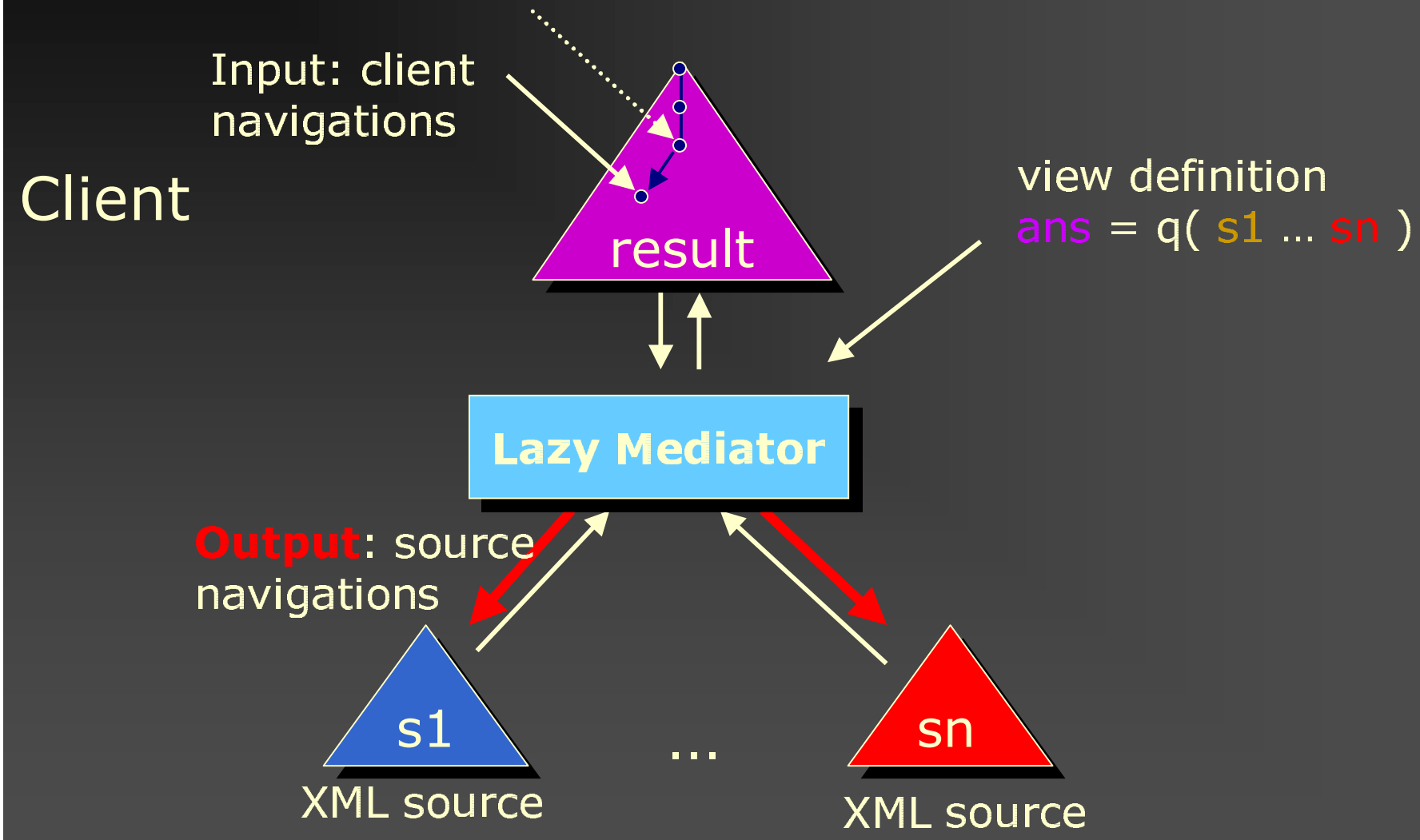
Navigation-Driven Evaluation
(LPV00)



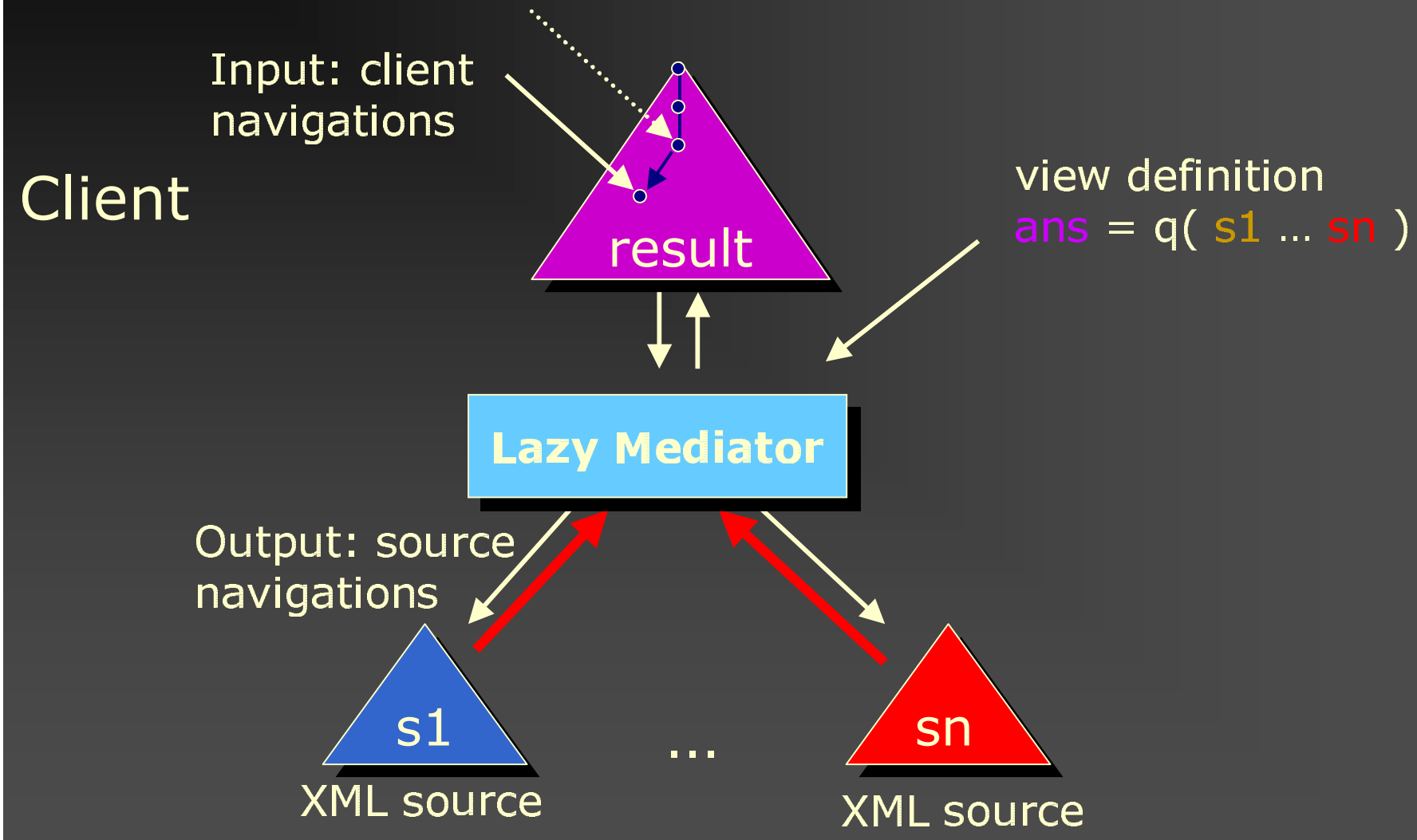
Navigation-Driven Evaluation



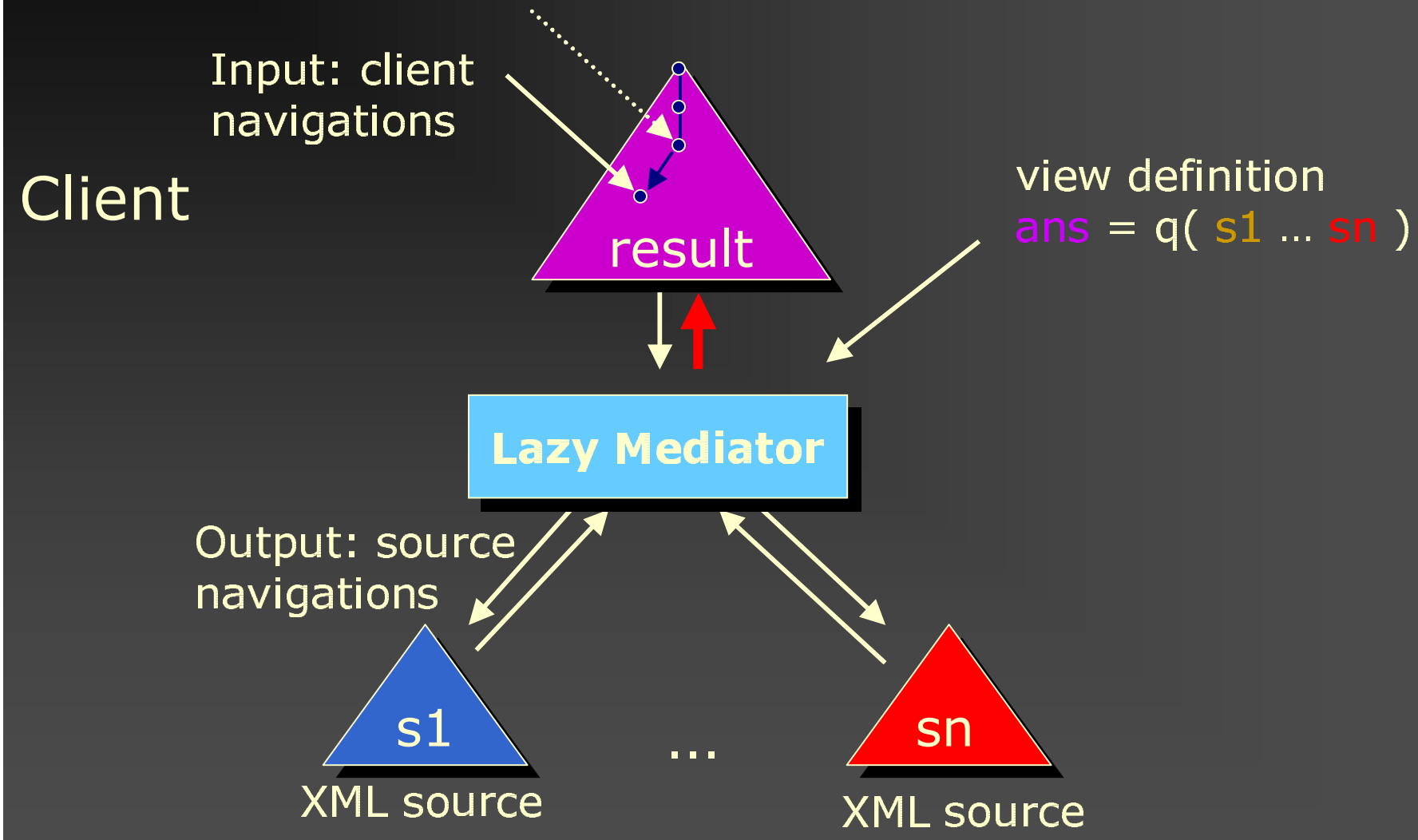
Navigation-Driven Evaluation



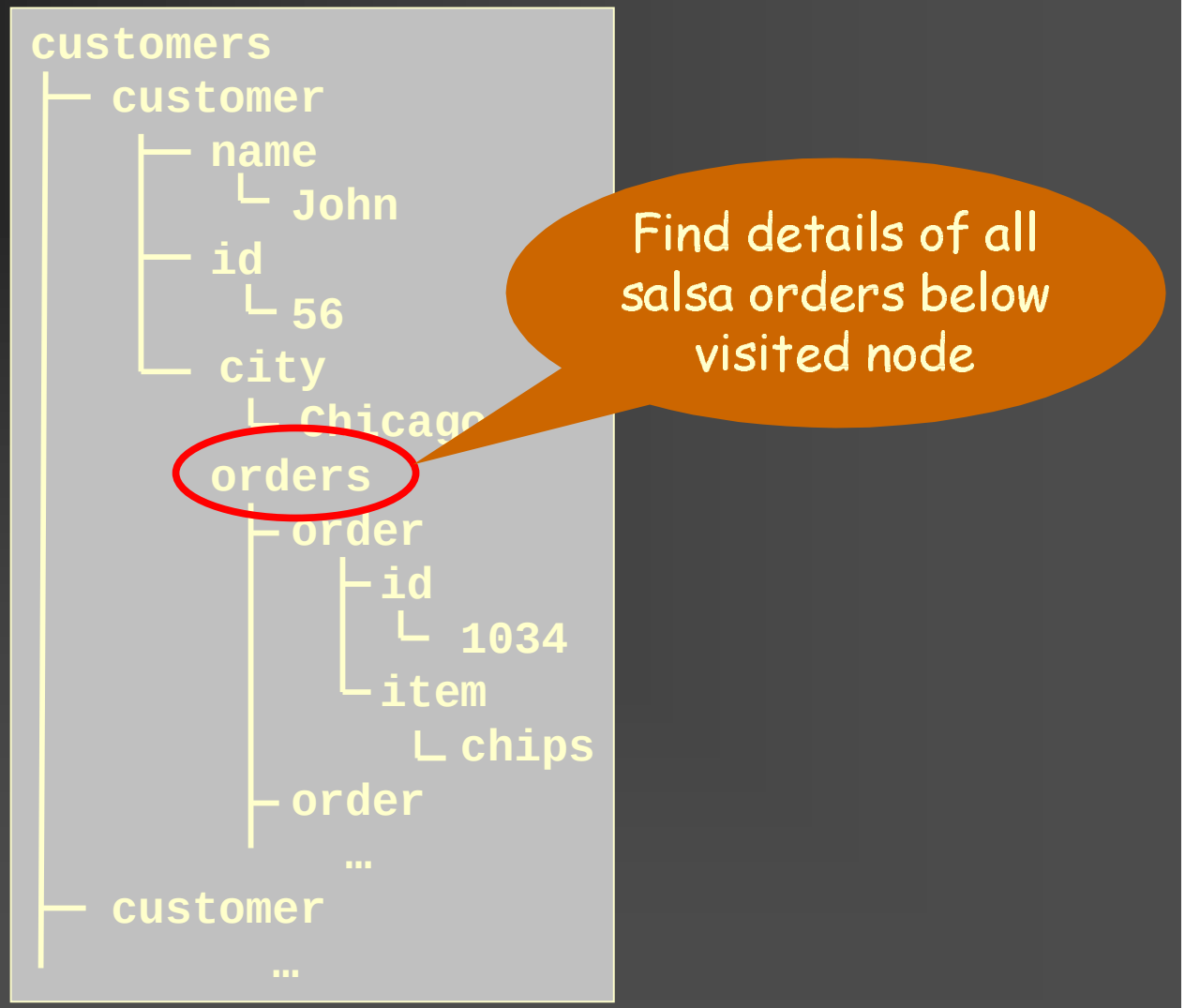
Navigation-Driven Evaluation



Navigation-Driven Evaluation



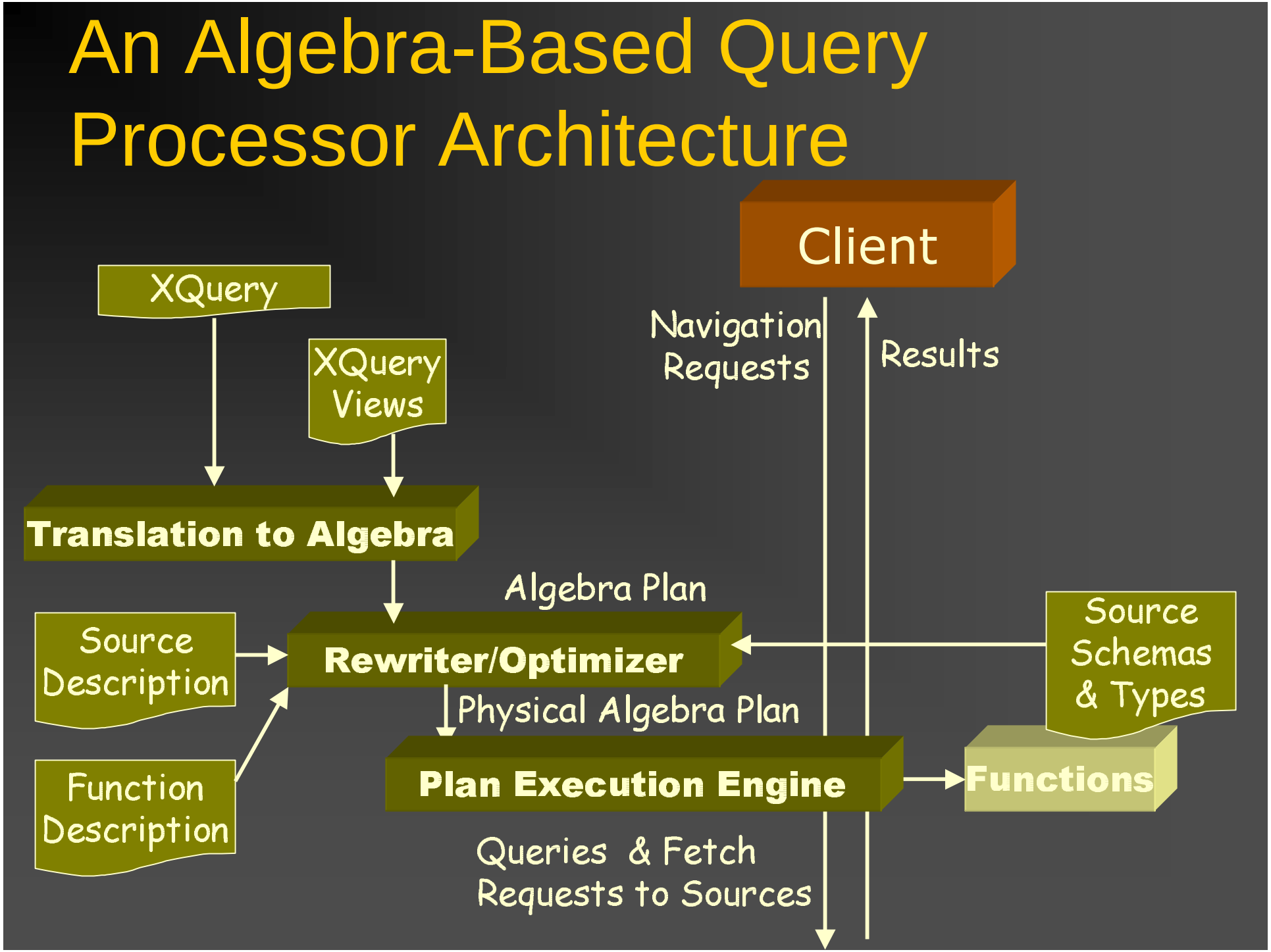
Mixing Querying & Navigation
(MP02)



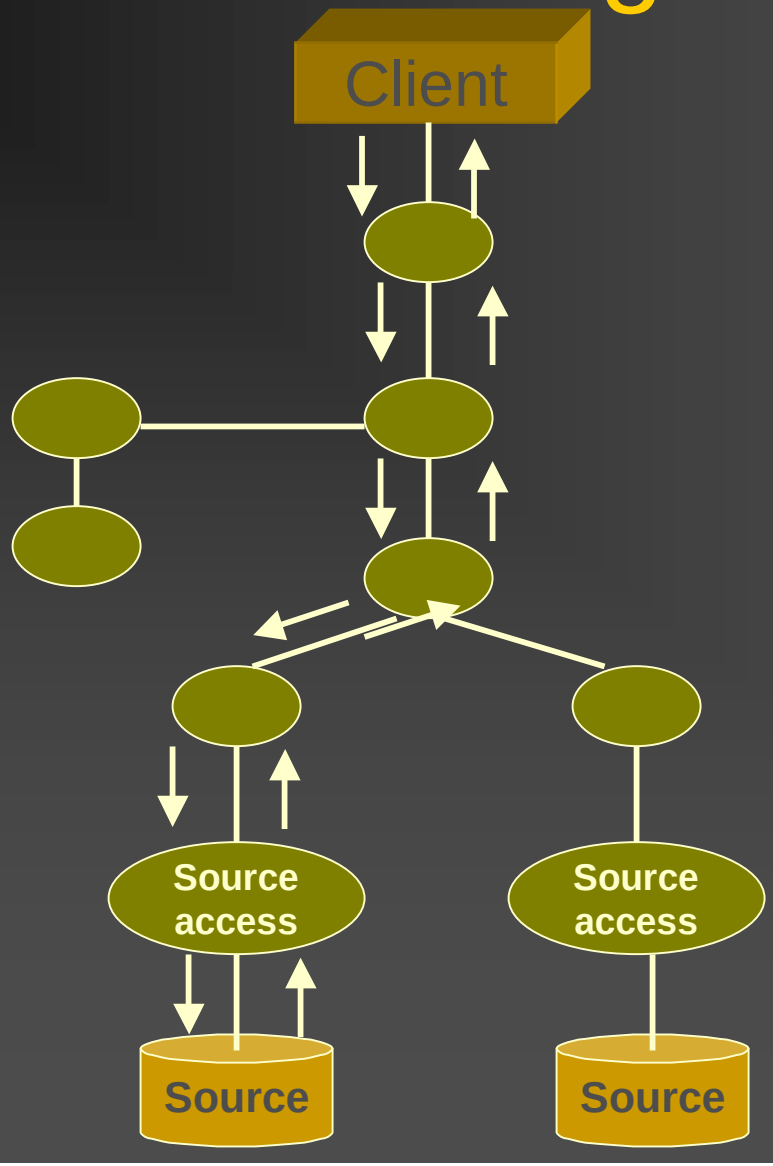
Challenges in Mixing Querying & Navigation

- Two-dimensional navigation
- Reminds of cursors but there are multiple continuation points
- Need semantic id’s (skolems) !
- Contextualizing queries by navigation

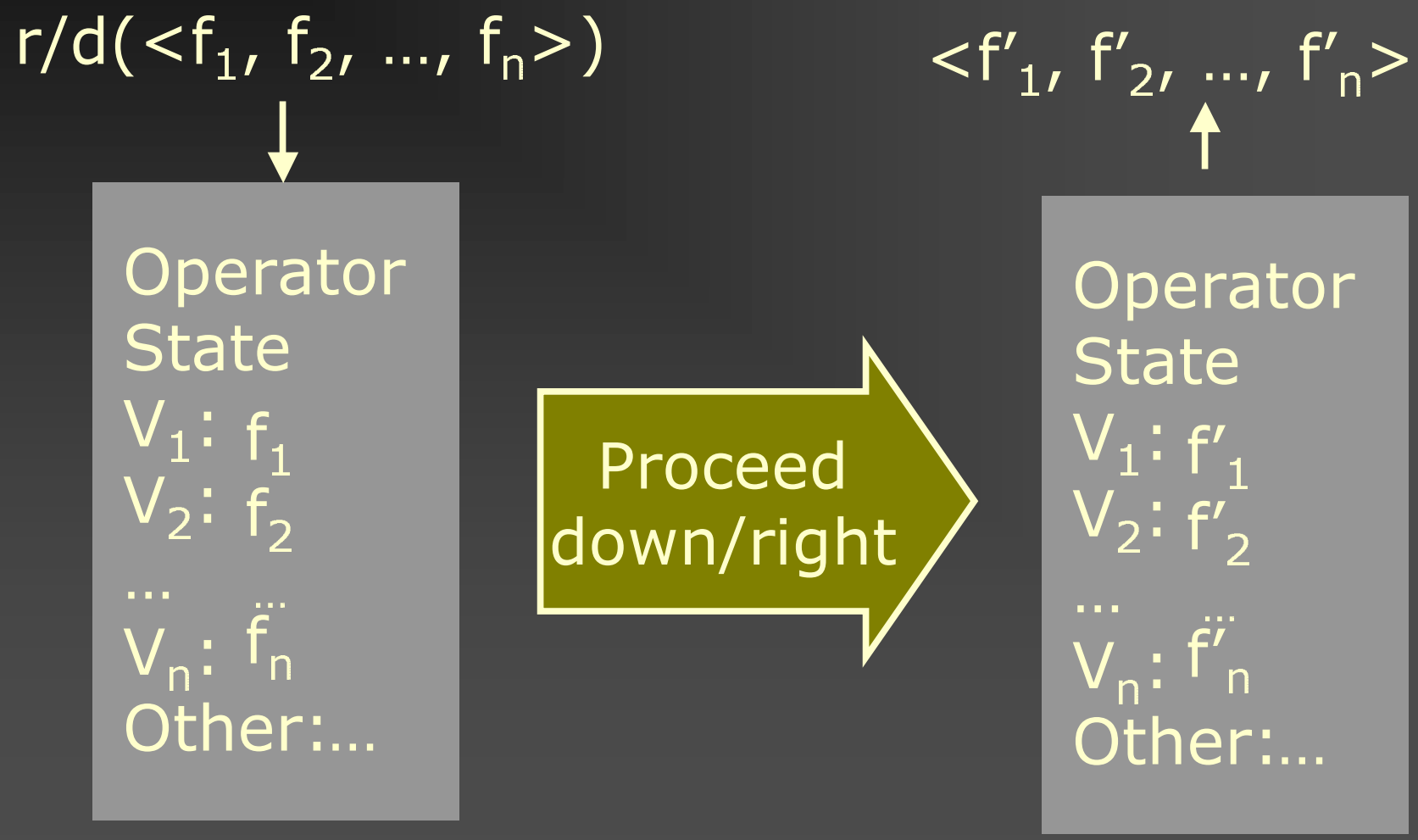
An Algebra-Based Query Processor Architecture



Building Navigation-Driven Evaluation on the Algebra



Use of Semantic Id's in Navigation-Driven Evaluation



TSIMMIS Retrospective & Status

- Semistructured Data Model => XML
- Query and View Definition Languages
- Capabilities-Based Rewriting
- Database Architecture for Semistructured Data Mediator
- Role of Schemas
- Non-Conjunctive Views
- Adaptive Optimization / Learning
- Supporting “stuff”: metadata repositories, query builders, form-and-report tools
- Download one from www.enosyssoftware.com