HyperQueries: Dynamic Distributed Query Processing on the Internet

Alfons Kemper  Christian Wiesner

Universität Passau
Outline

- Motivation
- Hyperlinks & HyperQueries
- Execution of HyperQueries
- Optimization Issues
- Related Work
- Conclusions
Motivation: E-Market Places (1)

Traditional techniques
- Replicate data in a central data warehouse
- Severe problems
  - Security and privacy violations
  - Coherence problems
  - Schema integration problems
  - Fixed query operators
**Motivation: E-Market Places (2)**

- **HyperQuery framework**
  - Market place is intermediary
  - Distribute query processing along allocation schema of data
  - Objects & queries flow through the WWW
  - Scalability
Hierarchical Flow of Objects

Client

select p.ProductDescription, c.Supplier, c.Price
from NeededProducts p, Catalog@MarketPlace c
where p.ProductDescription = c.ProductDescription
order by p.ProductDescription
expires Friday, Sep 14, 2001 11:30:00 AM CET
Broadcast-like Flow of Objects

Client

select p.ProductDescription, c.Supplier, c.Price
from NeededProducts p, Catalog@MarketPlace c
where p.ProductDescription = c.ProductDescription
order by p.ProductDescription
expires Friday, Sep 14, 2001 11:30:00 AM CET

Market Place

Supplier 1

Supplier 2

Supplier 3

Supplier 4
Hyperlinks

- Embed hyperlinks as virtual attributes in databases
- Hyperlinks refer to HyperQueries
## Hyperlinks - Example

<table>
<thead>
<tr>
<th>ProductDescription</th>
<th>Supplier</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battery, 12V 32A</td>
<td>Supplier 1</td>
<td><code>hq://supplier1.com/Electrical/Price?ProdID=CB1232</code></td>
</tr>
<tr>
<td>Battery 12V 55A</td>
<td>Supplier 1</td>
<td><code>hq://supplier1.com/Electrical/Price?ProdID=CB1255</code></td>
</tr>
<tr>
<td>Tires 175/65TR14</td>
<td>Supplier 2</td>
<td><code>hq://supplier2.com?ProdKey=175_65TR14</code></td>
</tr>
<tr>
<td>Spark Plug VX</td>
<td>Supplier 3</td>
<td><code>hq://supplier3.com/PriceForUSA!Currency=USD?ID=1234</code></td>
</tr>
</tbody>
</table>

**Virtual Attribute:**
- `!Currency=USD` - Global Parameter
- `?ProdID=CB1232` - Object-Specific Parameter

**HyperQuery Protocol**
- `hq` - HyperQuery Protocol
- `supplier1.com` - DNS
- `Electrical/Price` - HyperQuery ID
- `?ProdID=CB1232` - Object-Specific Parameter
- `!Currency=USD` - Global Parameter
HyperQueries

- Sub-plans at remote hosts
- Virtual table `HyperQueryInputStream`
- SQL Dialect

Electrical/Price@Supplier1.com:

```
select h.*, p.Price as Price
from HyperQueryInputStream h, Products p
where h.ProdID = p.ProdID
```
Templates for Sub-Plans

Nesting

Sequencing

Inner

Dispatch

recv

send
Example

Query

```
select p.ProductDescription, c.Supplier, c.Price
from NeededProducts p, Catalog@MarketPlace c
where p.ProductDescription=c.ProductDescription
order by p.ProductDescription
expires Friday, Sep 14, 2001 11:30:00 AM CET
```

NeededProducts@Client:

<table>
<thead>
<tr>
<th>ProductDescription</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battery, 12V 32A</td>
<td>500</td>
</tr>
<tr>
<td>Battery, 12V 55A</td>
<td>750</td>
</tr>
<tr>
<td>Tires 175/65TR14</td>
<td>1000</td>
</tr>
<tr>
<td>Spark Plug VX</td>
<td>8000</td>
</tr>
</tbody>
</table>
HyperQuery Execution
HyperQuery Execution

["Battery, 12V 32A", "Supplier 1", "hq://supplier1.com/Electrical/Price?ProdID=CB1232", 500]
HyperQuery Execution
HyperQuery Execution

Supplier 1
HyperQuery Execution
HyperQuery Execution

Supplier 1
HyperQuery Execution

["Battery, 12V 55A", "Supplier 1", "hq://supplier1.com/Electrical/Price?ProdID=CB1255", 750]
HyperQuery Execution
HyperQuery Execution

Supplier 1
HyperQuery Execution
HyperQuery Execution

Supplier 1

Supplier 2

['Tires 175/65TR14', 'Supplier 2', 'hq://supplier2.com/Price?ProdKey=175_65TR14', 1000]
HyperQuery Execution

Supplier 1

Supplier 2
HyperQuery Execution
HyperQuery Execution

["Spark Plug VX","Supplier 3","hq://supplier3.com/PriceForUSA!Currency=USD?ID=1234", 8000]
HyperQuery Execution

Supplier 1

Supplier 2

Supplier 3
HyperQuery Execution

Supplier 1
Supplier 2
Supplier 3
HyperQuery Execution
Real HyperQuery Execution
Real HyperQuery Execution

Supplier 1
Real HyperQuery Execution
Real HyperQuery Execution
Real HyperQuery Execution
Real HyperQuery Execution
Real HyperQuery Execution
Real HyperQuery Execution

Supplier 1
Supplier 2
Supplier 3

Currency = USD
Real HyperQuery Execution

Supplier 1

Supplier 2

Supplier 3

Currency=USD
Real HyperQuery Execution

Supplier 1
Supplier 2
Supplier 3
Origin of Data

- SQL database queries
- Relational representation
- Extensibility
- Applications
- Legacy systems
  - ERP systems (e.g., SAP R/3)
- Human interaction
Optimization: Bulk Bypassing

- Strip off bulky attributes
- Bypass Sub-plans
- Re-merge
- Reduce network traffic
Optimization: Multiple Virtual Attributes

- Parallelize requests

Availability & Price
Optimization: Caching
Optimization: Caching
Optimization: Caching
Optimization: Caching

Supplier 1

Supplier 2
Optimization: Caching

Supplier 1

Supplier 2
Optimization: Caching

Supplier 1  Supplier 2
Optimization: Caching
Optimization: Caching
Optimization: Predicate Migration
Related Work

- Stonebraker et.al.: Quel as a Datatype, SIGMOD 1984
- Braumandl et.al.: ObjectGlobe: Ubiquitous query processing on the Internet, VLDBJ 2001
- Roth and Schwarz: Don’t Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources, VLDB 1997
- Stonebraker, Hellerstein: Content Integration for E-Business, SIGMOD 2001 (Cohera)
Conclusions

- Sub-plans "sitting behind" hyperlinks
- Market place is intermediary
- Embed hyperlinks at the market place
- Provide HyperQueries at remote sites
- Data integration at remote sites
- Optimization techniques