MASHING UP DATA ON THE CLOUD: e-CLOUDSS

Genoveva VARGAS SOLAR
CNRS, LIG & LAFMIA-UDLAP

José Luis Zechinelli Martini, UDLAP-LAFMIA (UMI 3175)
Regina Motz, Alberto Parro, U. de la República de Uruguay
Martin A. Musicante, Federal University of Rio Grande do Nord

Genoveva.Vargas@imag.fr
http://vargas-solar.imag.fr
Provide me with continuous information about the location of providers of fresh groceries with a description of their daily production that are located no farther than 8 Km of my current position and that are open in the afternoon.
Provide me with weakly information about the location and the street view’s images of the bio supermarkets that provide fresh groceries in sale and that are located in my neighbourhood and whose international providers belong to a responsible economy.
**QUERY**

Give me the location of the transport vehicles and the conditions of the groceries being transported that require to be delivered in the next hour and that are not farther than 3 km of the next warehouse or cargo station with available and required equipment for transporting them.
**Groceries Global Provision**

**Query**
Provide me with weakly information about the location and the street view’s images of the bio supermarkets that provide fresh groceries in sale and that are located in my neighbourhood and whose international providers belong to a responsible economy.

**Query**
Give me the location of the transport vehicles and the conditions of the groceries that require to be delivered in the next hour and that are not farther than 3 km of the next warehouse or cargo station with available and required equipment for transporting them.

**Query**
Provide me with continuous information about the location of the providers of fresh groceries with a description of their daily production that are located no farther than 8 Km of my current position and that are open in the afternoon.
Efficiently manage and exploit data sets according to given specific storage, memory, and computation resources
MASHING DATA **WITHOUT** RESOURCES CONSTRAINTS

Costly manage and exploit data sets according to unlimited storage, memory, and computation resources
Services coordination for evaluating queries

Combine service coordination, and (partial) results presentation with query evaluation
Optimize queries according to quality of service criteria
Propose a testbed for validating query evaluation based on service coordination within « real » application contexts (e-government, climatological, citizen’s assistance apps)

No off-the-shelf DBMS for evaluating different types of queries
**Research Objectives**

**Mashup specification declarative language**
- Service and data processing specification
- Spatio-temporal expressions on data presentation

**QoS contract language**
- Constraints on the execution context
- Recovery actions

**Services look up**
- Semantic descriptions
- Recommendation

**Efficiently querying data from services**

```plaintext
Mashup specification declarative language
- Service and data processing specification
- Spatio-temporal expressions on data presentation

QoS contract language
- Constraints on the execution context
- Recovery actions

Services look up
- Semantic descriptions
- Recommendation

Efficiently querying data from services
```

---

**Exception**

**Contract**

```plaintext
define Mashup as http://ficeda.com.mx/

(Mashup compound m2 http://smn.cna.gob.mx
  with Mashup.size_x, Mashup.size_y, m2 north beginning Mashup)

compound (m3 http://www.trackyourtruck.com/
  with m3.size_x, m3.size_y, m3 right Mashup)

C1

Exception
Contract
C1
```

```plaintext
do show emptydoc
```

---
Mashup specification, declarative language
• Service and data processing specification
• Spatio-temporal expressions on data presentation

QoS contract language, services look-up
• Constraints on the execution context
• Recoverable action on services

Efficiently querying data from services
• Semantic descriptions
• Recommendation
MASHING UP DATA IN DYNAMIC ENVIRONMENTS

EFFICIENT QUERYING DATA FROM SERVICES

CURRENT WORK & PERSPECTIVES

Hybrid query

Optimisation

Deployment
**HYBRID QUERY**

Give me the location of the transport vehicles and the conditions of the groceries being transported that require to be delivered in the next hour and are not farther than 3 km of the warehouse or cargo station with available and required equipment for transporting them.

- Is spatio-temporal or not
- Consumes on demand data or streams from static or nomad data services
- Is evaluated continuously and in batch
HYBRID QUERY
MASHING UP SERVICES FOR QUERYING DATA

MQList
- Declarative query language
- Combines service and ST
- Results presentation

Hypatia²

Query Model
- Query workflow: workflow-based service coordination model (ASM formalism)

Services Interfaces Representation

Data Model
- Primitive data types, tuples, sets, functions
- Operators: recursive relational, windows, combination, set, nesting/unnesting

² Evaluating Hybrid Queries through Service Coordination in HYPATIA, Victor Cuevas-Vicenttin; Genoveva Vargas-Solar; Christine Collet,. In Proceedings of the 15th International Conference on Extending Database Technology (EDBT), Berlin, Germany. 2012.
**QUERY WORKFLOW**

**QUERY**
Give me the **location** of the **transport vehicles** and the conditions of the **groceries** being transported that require to be delivered in the **next hour** and that are **not farther than 3 km** of the next warehouse or cargo station with available and required equipment for transporting them.
ROADMAP

EFFECTIVELY QUERYING DATA FROM SERVICES

Hybrid query

Optimisation

Deployment
Find the query workflow(s) that implements the hybrid query and that best conveys with the SLA contract according to the available services.
MEASURES AND SLA

Low cost (SLA)
- Battery
- ASAP
- Economic

Throughput
- Latency
- Availability
- Execution time
- Price

Exported [51,59]
Calculated [1]
The cost function is defined as:

$$\text{Opt (SLA, R)} = \min \sum_j (f(A_j, W_j) - \gamma(\text{SLA, } R_j))$$

- **Opt (SLA, R)**: Find the best set of resources $R$ which satisfy the set of SLA requirements $\text{SLA}$.
- **$\gamma(\text{SLA, } R_j)$**: Gamma is the correlation function between SLA and the resources set $R_j$. E.g. is $Q$ available in $R_j$? How much?
- **$f(A_j, W_j)$**: Taking into account available ($A$) and used ($W$) involved resources.
Computing a solution space

Rewriting: determine the activities and a control flow of a workflow that

Parallelizes and factorizes as much as possible operations
Minimizes a cost function: available services and resources
CHOOSING AN "OPTIMUM" PLANS SET
[EUCLIDEAN DISTANCE]

HYBRID QUERY

SERVICE LEVEL AGREEMENT

Cost function

\[ QWF^1(p_1^1, p_2^1, \ldots, p_m^1) \]

\[ QWF^2(p_1^2, p_2^2, \ldots, p_m^2) \]

\[ \vdots \]

\[ QWF^n(p_1^n, p_2^n, \ldots, p_m^n) \]

\[ d(p_i^j, o) = \sqrt{\sum_{j=1}^{m} (p_j^i - o_j)^2} \]

SOLUTION SPACE
EFFICIENTLY QUERYING DATA FROM SERVICES

ROADMAP

Hybrid query

Optimisation

Deployment
Limitations

- Types of query that can be evaluated
- Types of services that can be utilized
- Fixed architecture, difficult to extend

- Do not use services at their full potential
- Query evaluation by the use of services
- Service composition to add new capabilities

<table>
<thead>
<tr>
<th>Spec. functionality</th>
<th>Amazon S3, Active XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interoperability</td>
<td>SoCQ, OGSA DQP, Active XML, Amazon S3</td>
</tr>
<tr>
<td>Data processing</td>
<td>OGSA DQP, Active XML, Panta Rhei (SeCo)</td>
</tr>
<tr>
<td>Data producers (services) with exposed schema</td>
<td>SoCQ, OGSA DQP, Active XML, Panta Rhei (SeCo)</td>
</tr>
</tbody>
</table>
Enable querying data services in dynamic environments

Autonomous, dynamic data services

No off-the-shelf DBMS

Users have a wide range of query requirements
ROADMAP

MASHING UP DATA IN DYNAMIC ENVIRONMENTS

EFFICIENTLY QUERYING DATA FROM SERVICES

CURRENT WORK & PERSPECTIVES
CURRENT COMPLEMENTARY WORK

Regina Motz  
U. DE LA R, URUGUAY  
http://www.researchgate.net/profile/Regina_Motz/  
Semantic services description, service recommendation with QoS

Martin Musicante  
UFRN, Brasil  
http://www.fing.edu.uy/~pardo/  
Services composition language with non functional properties

Alberto Pardo  
U. DE LA R, URUGUAY  
http://www.fing.edu.uy/~pardo/  
Formal expression of mashup language and its non functional properties

José Luis Zechinelli Martini  
UDLAP, Mexico  
joseluis.zechinelli@udlap.mx  
http://www.udlap.mx/ofertaacademica/  
Spatio temporal query languages services coordination, data management on the cloud
RESULTS AND IMPACT

Consolidate and enhance research on information integration through reliable data services coordination

Key aspects
- Services data integration by coordinating services
- Spatio-temporal presentation of mashed up data
- Reliability (e.g., security, data freshness)

Complement concerning data access by coordinating services

Simple solutions for effective e-Government procedures in LATAM
gracias

Contact: Genoveva.Vargas@imag.fr
http://vargas-solar.imag.fr