

**Microsoft**



Microsoft® Research

# Faculty Summit 2012

Riviera Maya, Mexico | May 23-25 | In partnership with CONACYT



# MASHING UP DATA ON THE CLOUD: E-CLOUDSS

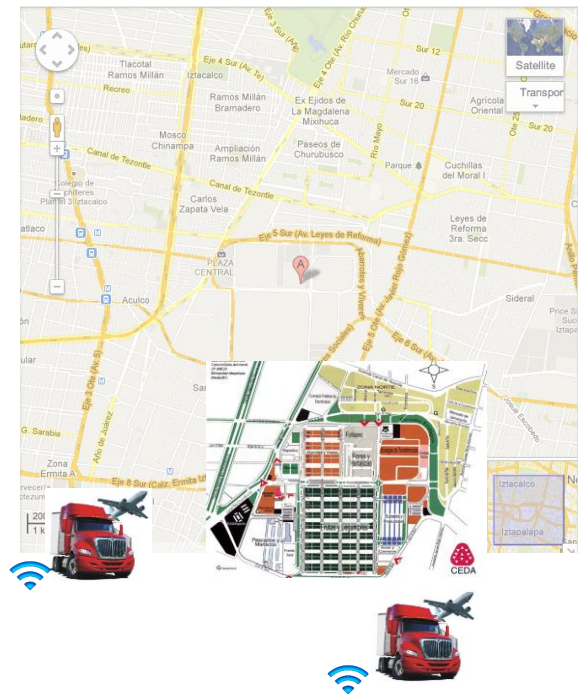
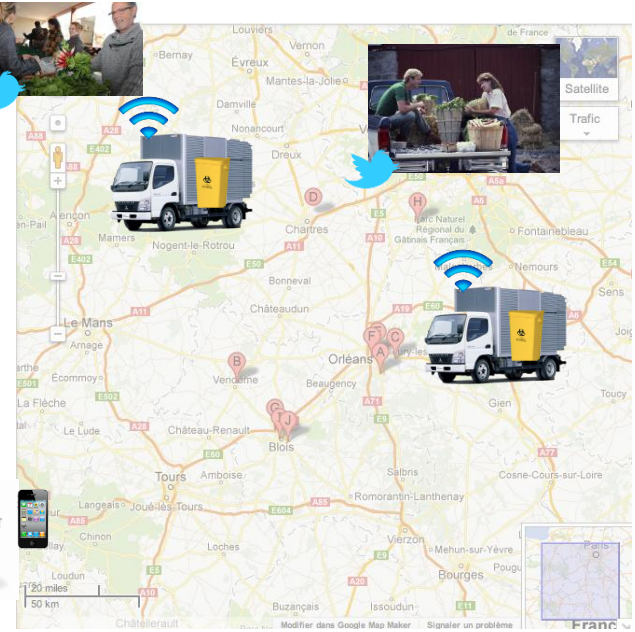
Genoveva VARGAS SOLAR  
CNRS, LIG & LAFMIA-UDLAP

José Luis ZECHINELLI MARTINI, UDLAP-LAFMIA (UMI 3175)  
Regina MOTZ, Alberto PARDO, U. de la República de Uruguay  
Martin A. MUSICANTE, Federal University of Rio Grande do Nord

[Genoveva.Vargas@imag.fr](mailto:Genoveva.Vargas@imag.fr)

<http://vargas-solar.imag.fr>

# GROCERIES GLOBAL PROVISION

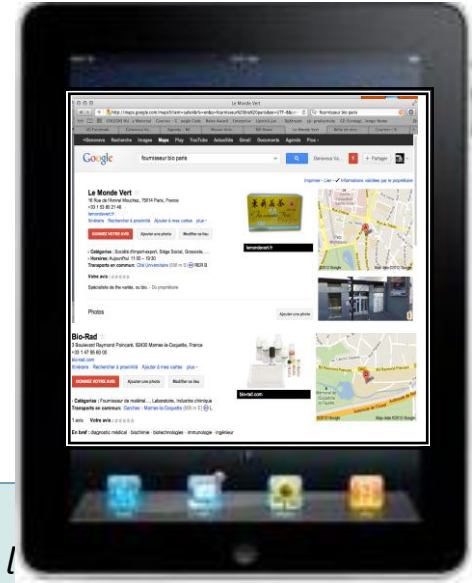


# GROCERIES GLOBAL PROVISION



## QUERY

*Provide me with continuous information about the location 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*



# GROCERIES GLOBAL PROVISION



## QUERY

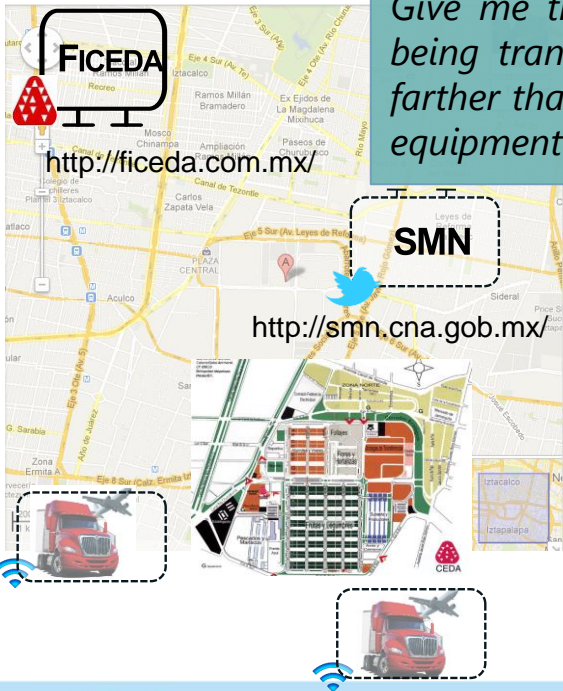
*Provide me with weakly information about the location and the street view's images of the bio supermarkets that provide fresh groceries in sale and that are located in my neighbourhood and whose international providers belong to a responsible economy*



# GROCERIES GLOBAL PROVISION

## 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 international providers belong to a responsible economy*

## LOW COST (SLA)

- Battery
- ASAP
- Economic

*Give me information of the transport vehicles and the conditions of the groceries being delivered that require to be delivered in the next hour and that are not farther than 8 km of the next warehouse or cargo station with available and required equipment for transporting them*

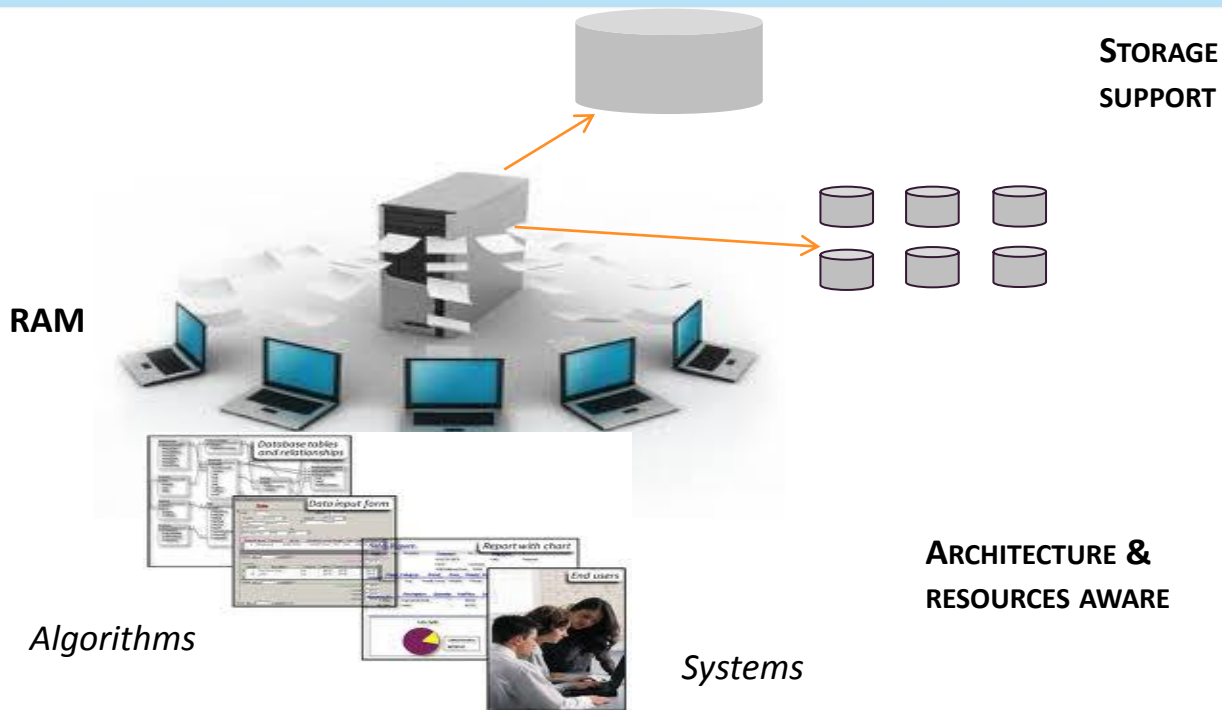
## QUERY

*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*

SMN

<http://smn.cna.gob.mx/>

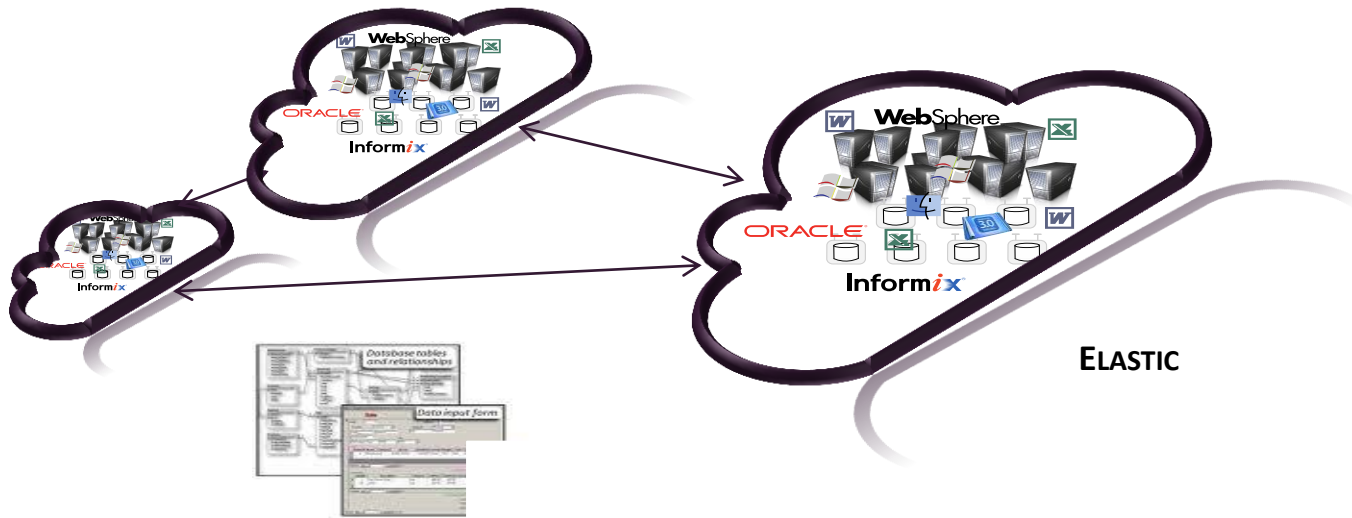
# MASHING UP DATA WITH RESOURCES CONSTRAINTS



Efficiently manage and exploit data sets according to given specific storage, memory, and computation resources



# MASHING DATA **WITHOUT** RESOURCES CONSTRAINTS



*Algorithms*

*Systems*

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

Exception Contract

*C1*

```
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 ) )
```

= labmember

do

show

emptydoc

# RESEARCH OBJECTIVES

## Mashup specification declarative language

- Service and data processing specification
- Spatio-temporal expressions on data presentation

## QoS contract language

- Constraints on transaction execution context
- Recommendation

## Users look

- Semantic description
- Recommendation



presentation  
expressions on data

# ROADMAP

**MASHING UP DATA  
IN DYNAMIC  
ENVORINMENTS**

**ENVORINMENTS  
IN DYNAMIC**

Hybrid query

**EFFICIENT  
QUERYING I  
FROM SERV**

Optimisation

**FROM SERVICES  
ORIENTED DATA**

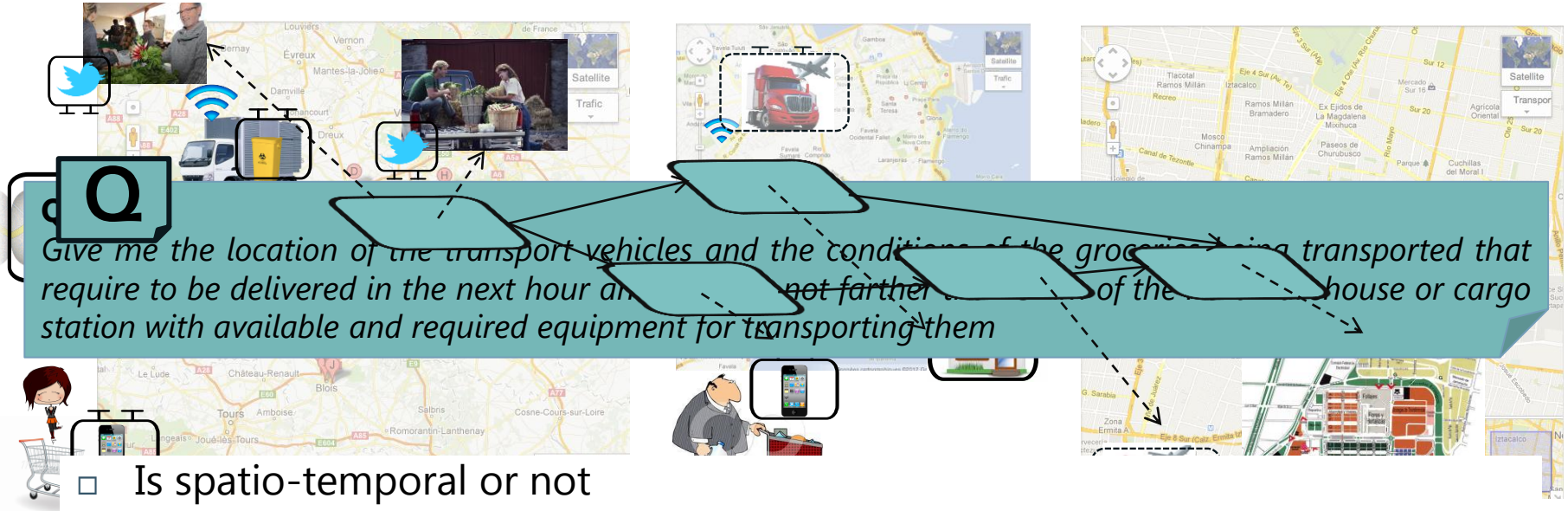
**CURRENT WORK  
& PERSPECTIVES**

Deployment

**ORIENTED  
DATA**

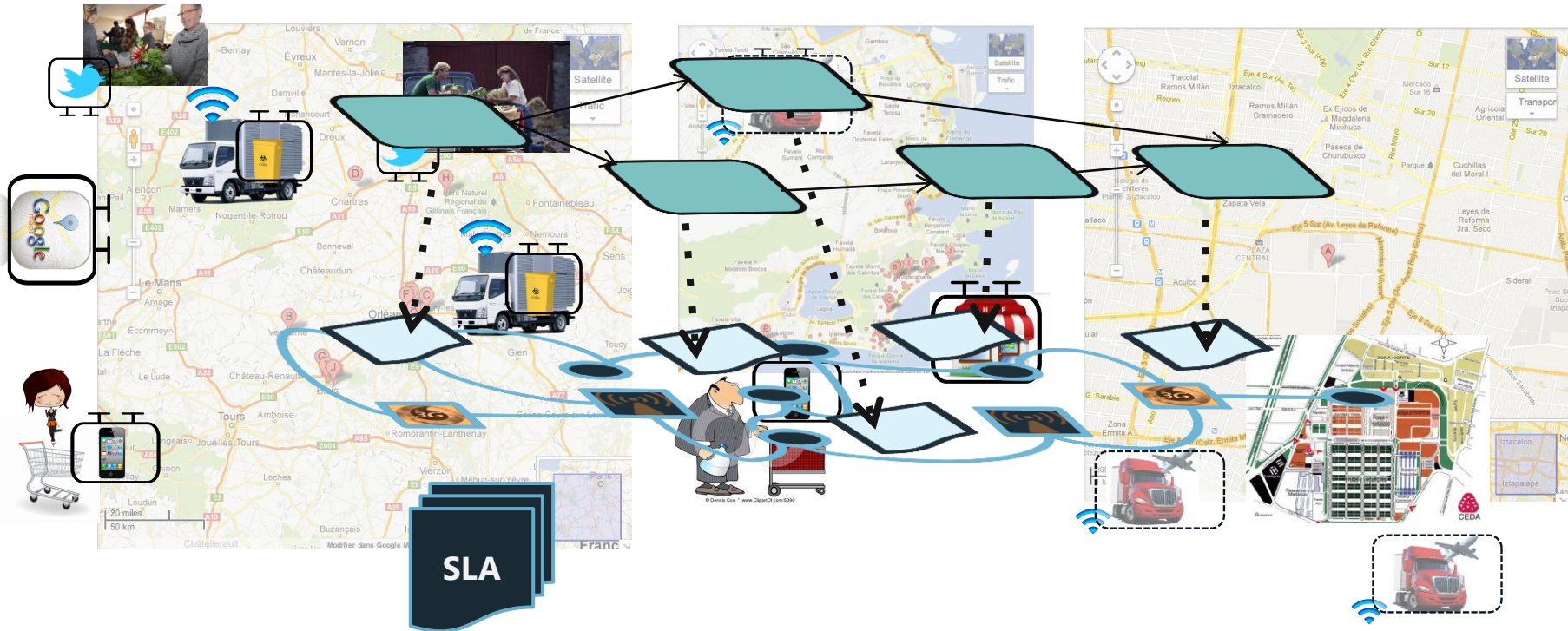


# HYBRID QUERY



- 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

HYPATIA<sup>2</sup>

## QUERY MODEL

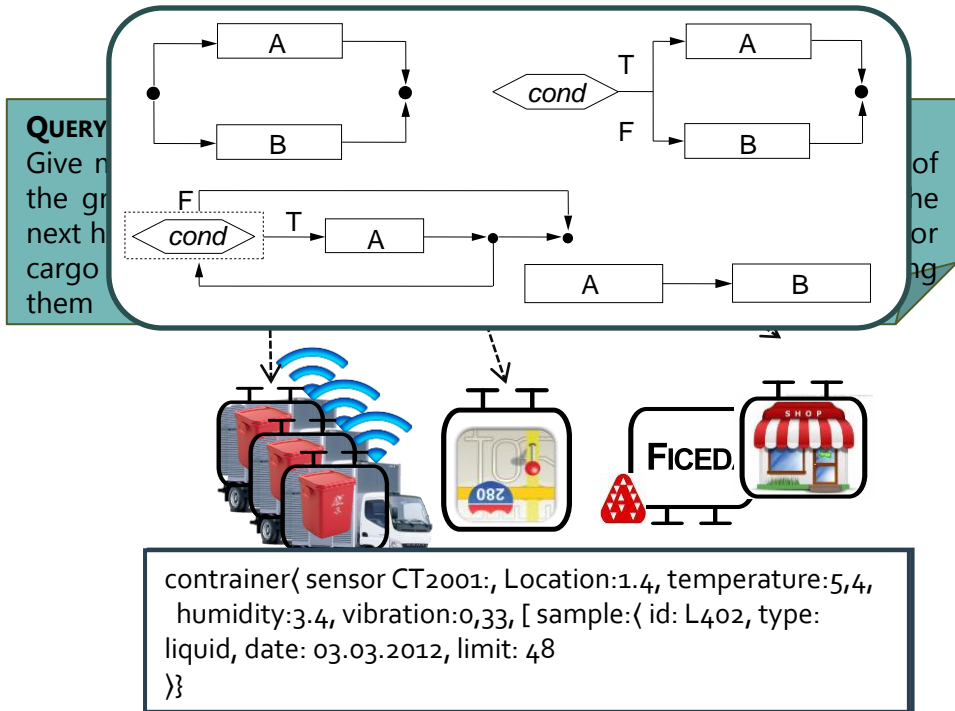
- **Query workflow:** workflow-based service coordination model (ASM formalism)

## SERVICES INTERFACES REPRESENTATION

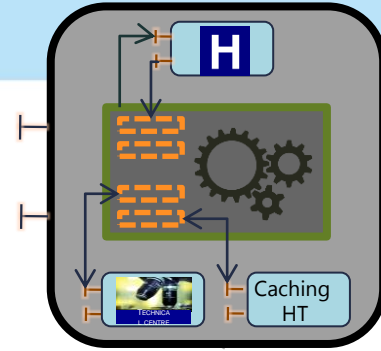
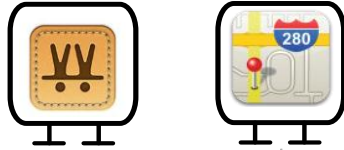
## DATA MODEL

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

<sup>2</sup> 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

**EFFICIENTLY  
QUERYING DATA  
FROM SERVICES**

**FROM SERVICES  
OPERATING DATA**

Hybrid query

Hybrid query

Optimisation

Optimisation

Deployment

Deployment

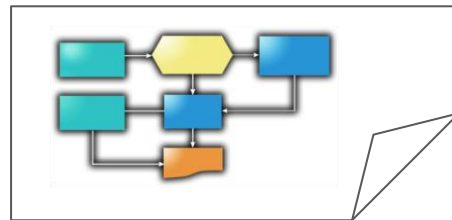


# HYBRID QUERY OPTIMIZATION

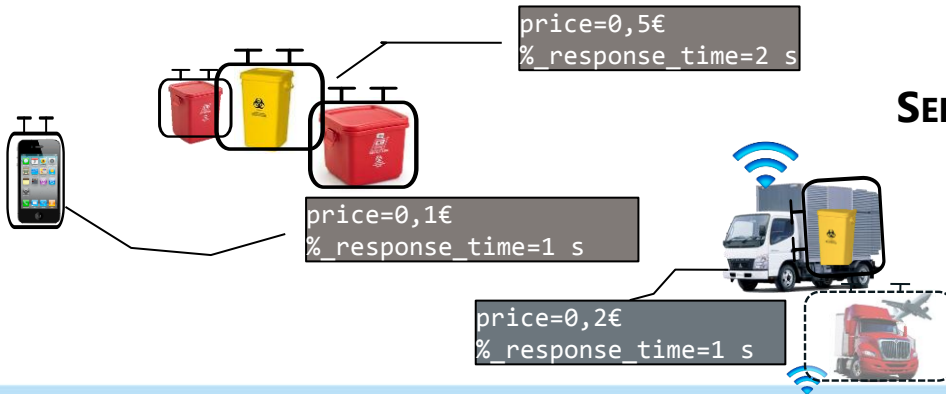
## HYBRID QUERY

## SERVICE LEVEL AGREEMENT

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



price=0,2€  
%\_response\_time=8 s



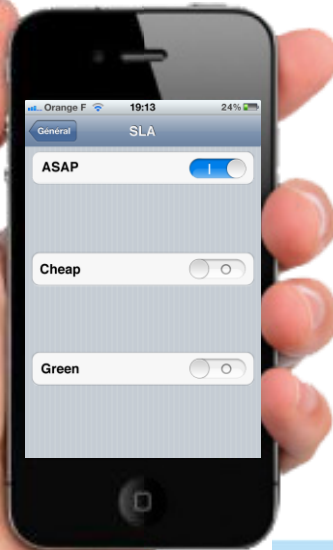
## SERVICES MEASUREMENT



# MEASURES AND SLA

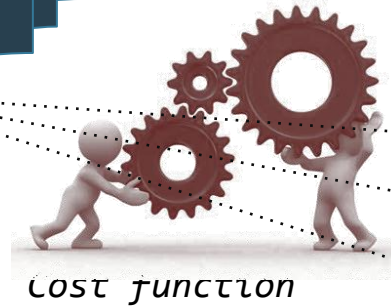


END-TO-END FASHION



## Low cost (SLA)

- Battery
- ASAP
- Economic



- Throughput
- Latency
- Availability
- Execution time
- Price

Exported [51,59]  
Calculated [1]

SYSTEM AND NETWORK MEASURES

# COST FUNCTION

## INITIAL STRATEGY

Choice of the set of QoS metrics to be taken into account

Define single-objective optimization function

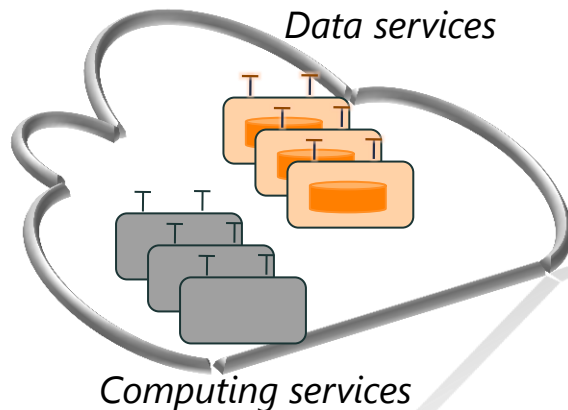
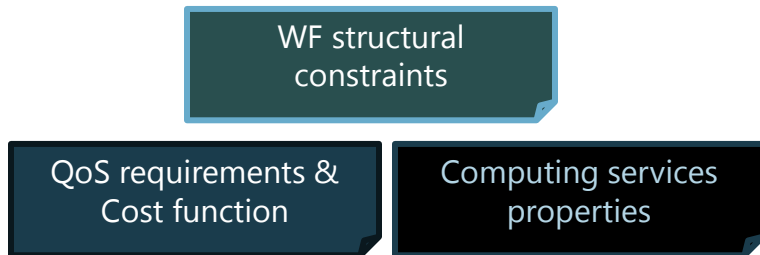
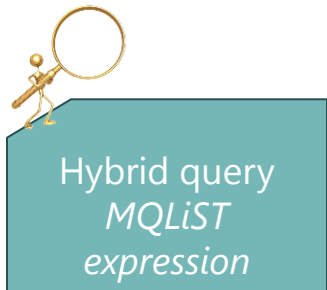
$$\text{Opt (SLA, R)} = \min \sum_j ( f(A_j, W_j) - \gamma(\text{SLA}, R_j) )$$

find the best set of resources R  
which satisfy the set of SLA  
requirements SLA

taking into account  
available (A) and used  
resources (W) involved

gamma is the correlation  
function between SLA and the  
resources set  $R_j$ .  
E.g. is Q available in  $R_j$ ? How  
much?

# COMPUTING A SOLUTION SPACE



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

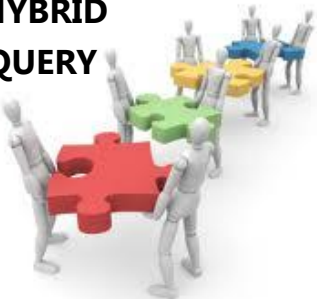
Parallelizes and factorizes as much as possible operations

Minimizes a cost function: available services and resources

# CHOOSING AN "OPTIMUM" PLANS SET

[EUCLIDEAN DISTANCE]

## HYBRID QUERY



## SERVICE LEVEL AGREEMENT



Cost function

$$QWF^1(p_1^1, p_2^1, \frac{1}{4}, p_m^1)$$

$$QWF^2(p_1^2, p_2^2, \frac{1}{4}, p_m^2)$$

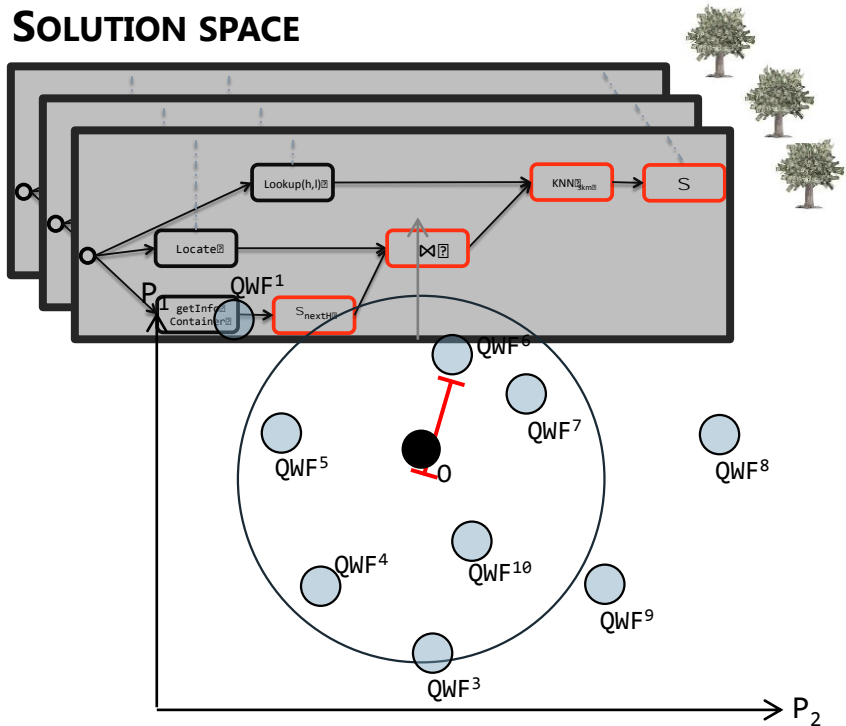
⋮

$$QWF^n(p_1^n, p_2^n, \frac{1}{4}, p_m^n)$$

"  $i:1 \in i \in n$

$$d(p^i, o) = \sqrt{\sum_{j=1}^m \dot{a} (p_j^i - o_j)^2}$$

## SOLUTION SPACE





# ROADMAP

EFFICIENTLY  
QUERYING DATA  
FROM SERVICES

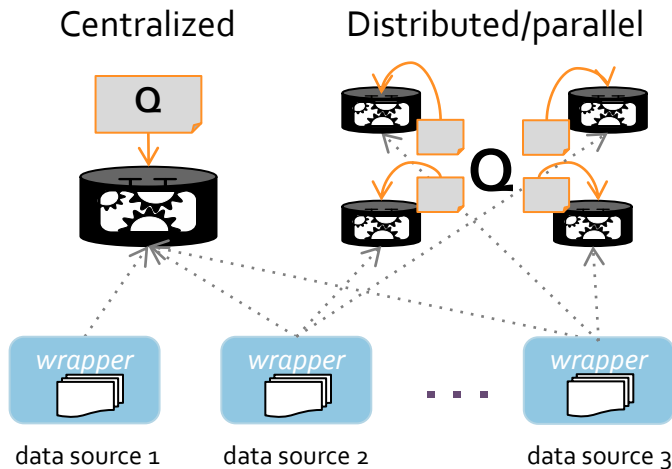
KNOW SERVICES  
OPERATING DATA

Hybrid query

Optimisation

Deployment

# QUERY PROCESSING ARCHITECTURES



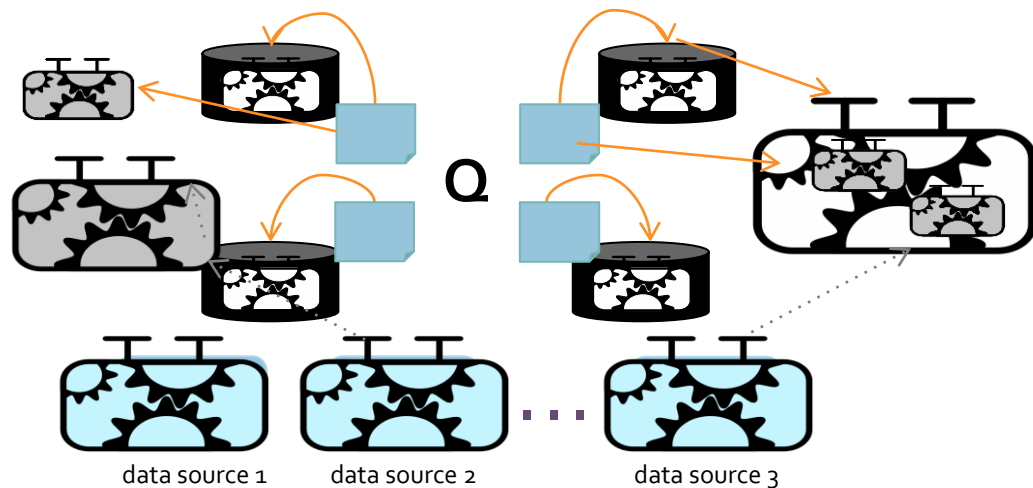
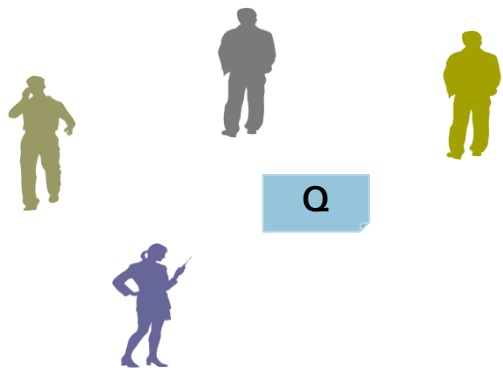
## ■ Limitations

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

Spec. functionality	Amazon S3, Active XML
Interoperability	SoCO, OGSA DQP, Active XML, Amazon S3
Data processing	OGSA DQP, Active XML, Panta Rhei (SeCo)
Data producers (services) with exposed schema	SoCO, OGSA DQP, Active XML, Panta Rhei (SeCo)

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

# QUERYING IN DYNAMIC ENVIRONMENTS



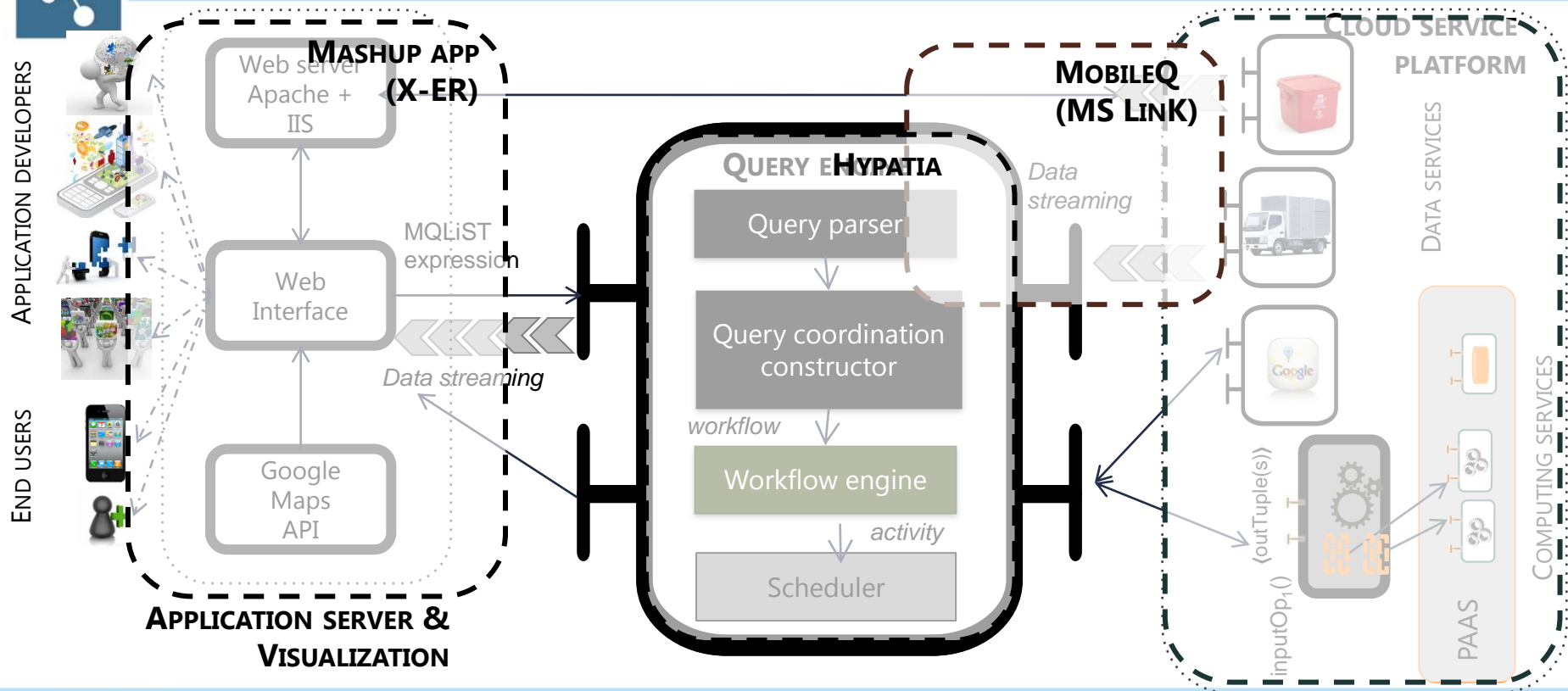
Enable querying data services in dynamic environments

Autonomous, dynamic data services

No off-the-shelf DBMS

Users have a wide range of query requirements

# IMPLEMENTATION ISSUES





# ROADMAP

**MASHING UP DATA  
IN DYNAMIC  
ENVIRONMENTS**

ENVIRONMENTS  
IN DYNAMIC

**EFFICIENTLY  
QUERYING DATA  
FROM SERVICES**

FROM SERVICES  
QUERYING DATA

**CURRENT WORK &  
PERSPECTIVES**

PERSPECTIVES

# CURRENT COMPLEMENTARY WORK



**Regina MOTZ**  
**U. DE LA R, URUGUAY**

[http://www.researchgate.net/profile/Regina\\_Motz//](http://www.researchgate.net/profile/Regina_Motz//)

Semantic services description,  
service recommendation with  
QoS



**Martin MUSICANTE**  
**UFRN, BRASIL**

<http://www.fing.edu.uy/~pardo/>

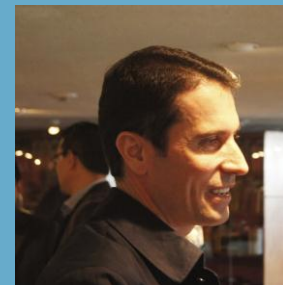
Services composition language  
with non functional properties



**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](mailto:joseluis.zechinelli@udlap.mx)  
[http://www.udlap.mx/ofertaaca](http://www.udlap.mx/ofertaacademica/)

**demica/**  
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



*obrigado*

*gracias*

*thank you*



Contact: [Genoveva.Vargas@imag.fr](mailto:Genoveva.Vargas@imag.fr)  
<http://vargas-solar.imag.fr>

