Opérateurs OLAP dans les bases de données multidimensionnelles multifonctions

Ali Hassan
Franck Ravat
Olivier Teste
Ronan Tournier
Gilles Zurfluh

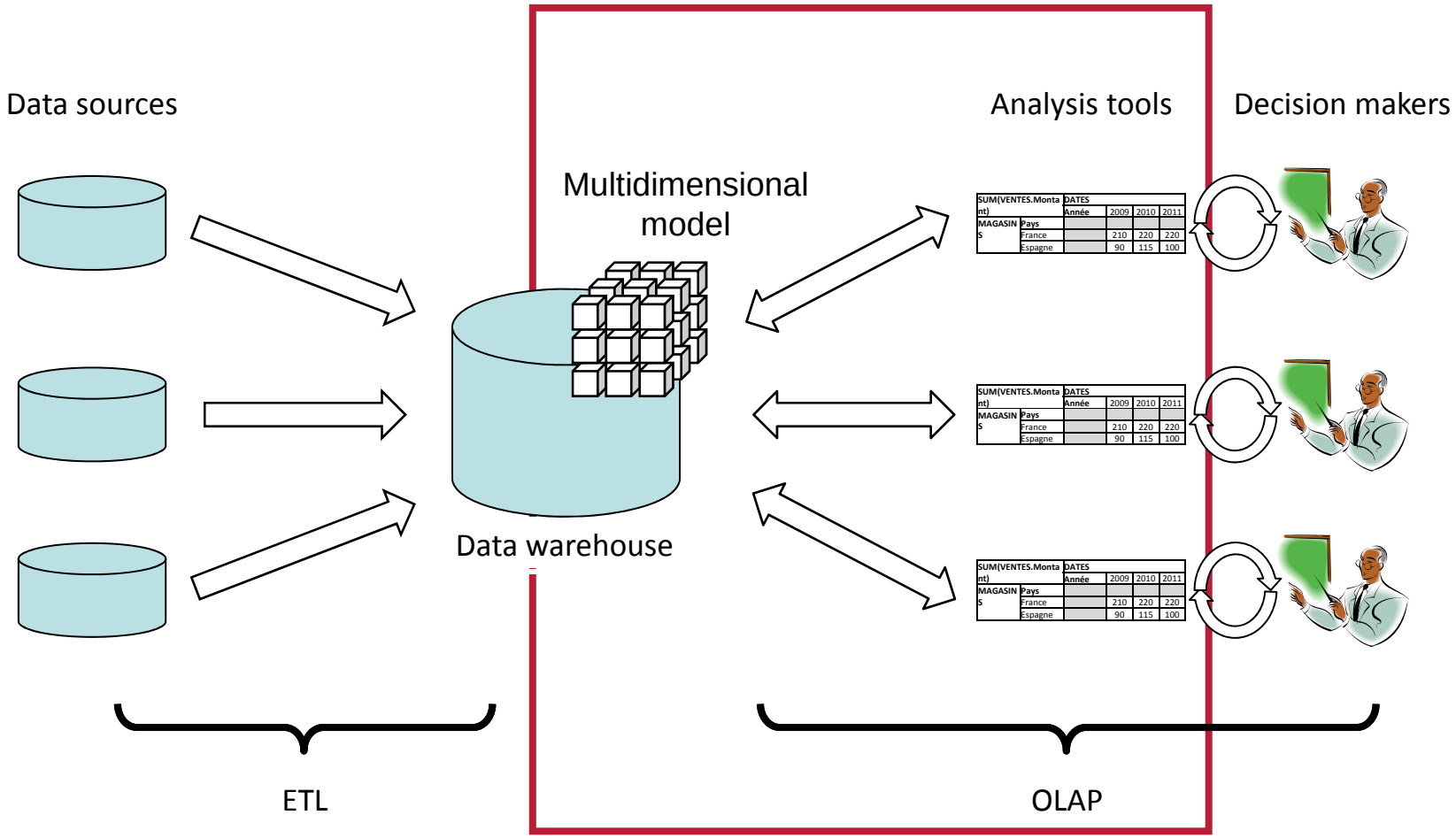
Layout

- Introduction
- Multifunction Conceptual Data Model
- Extended Multidimensional Table
- Extended OLAP Operators
- Conclusion and Future Works

Introduction

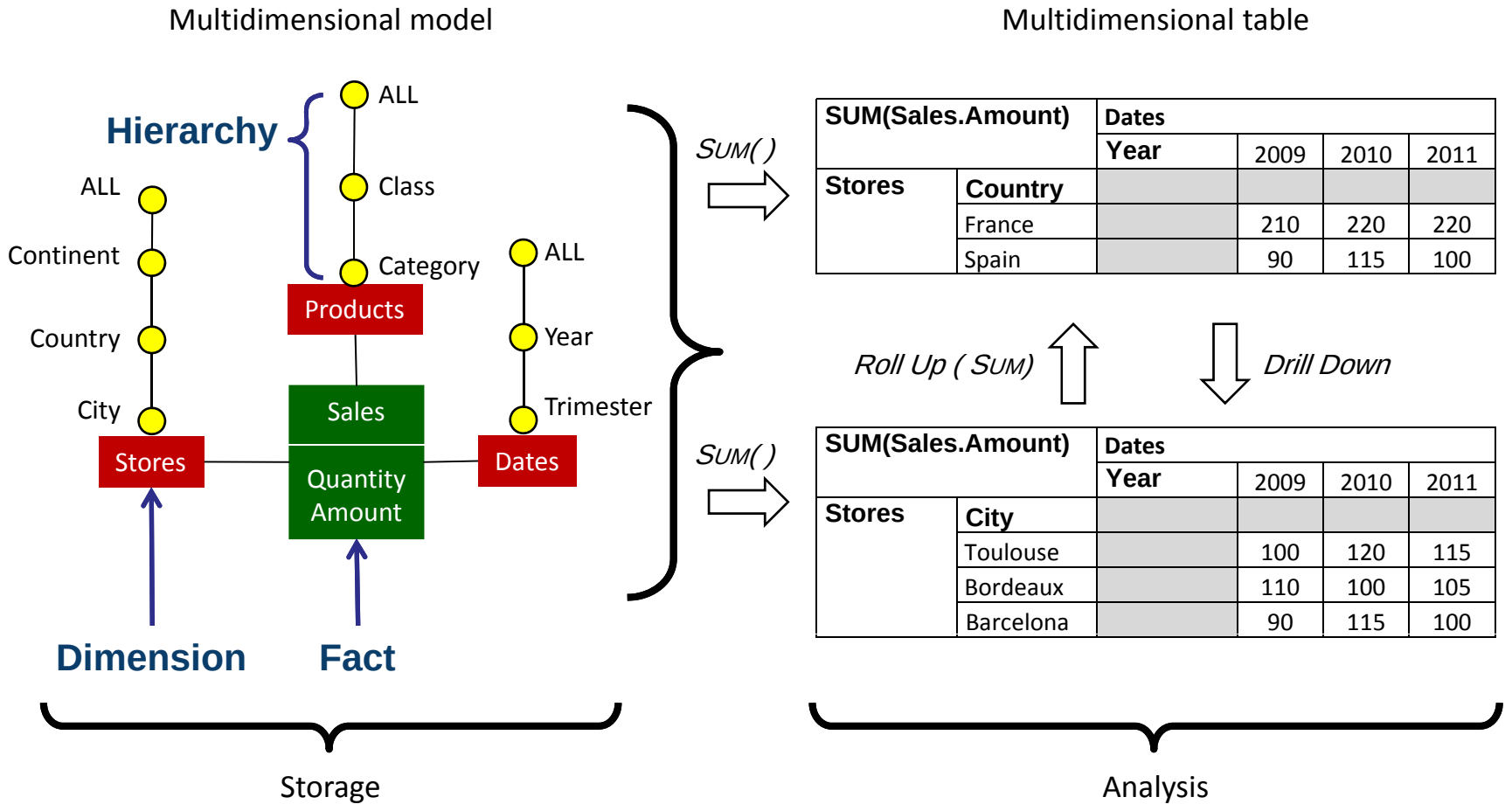
- Context
- Problem statement

Context (1/2)

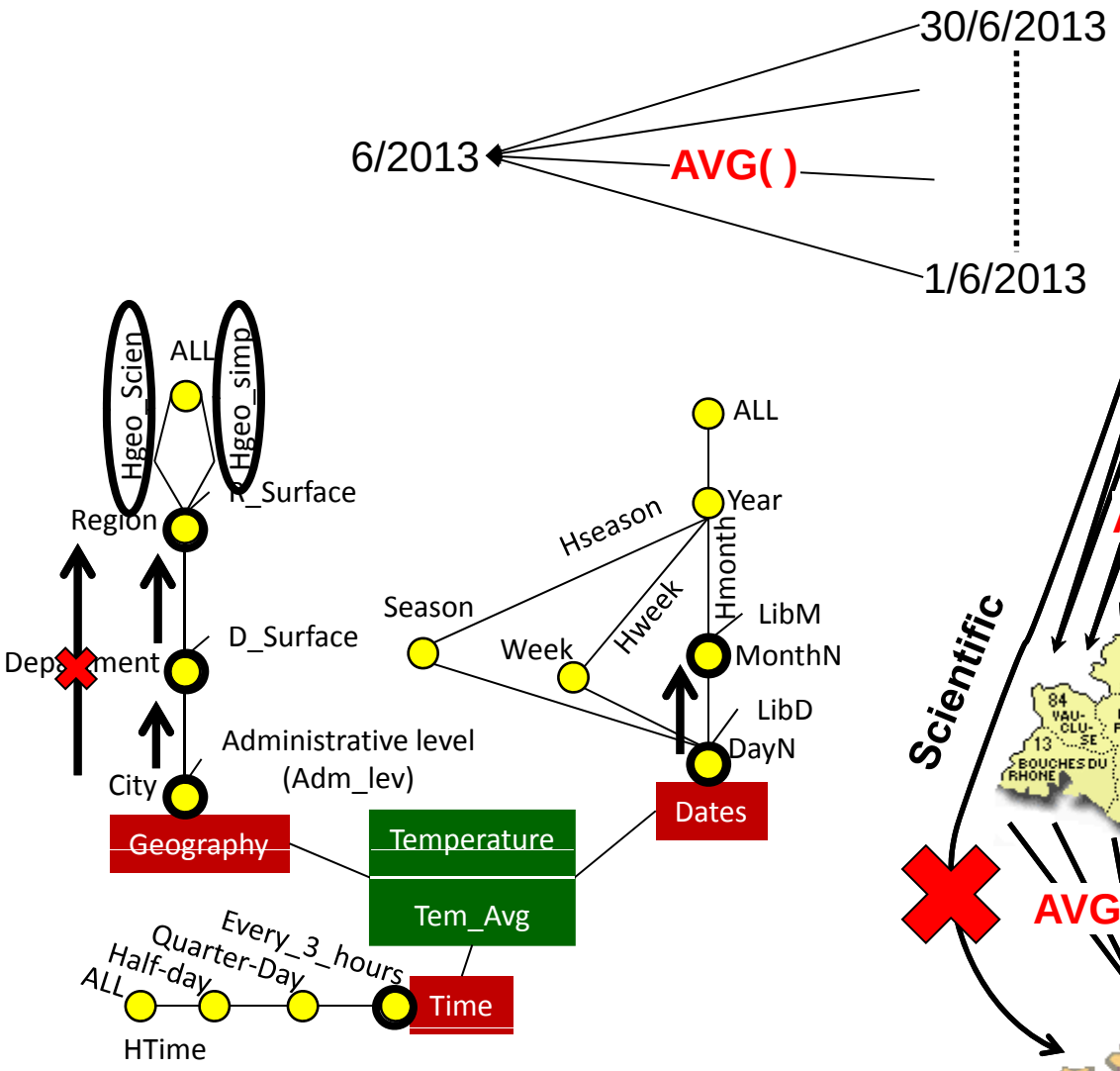


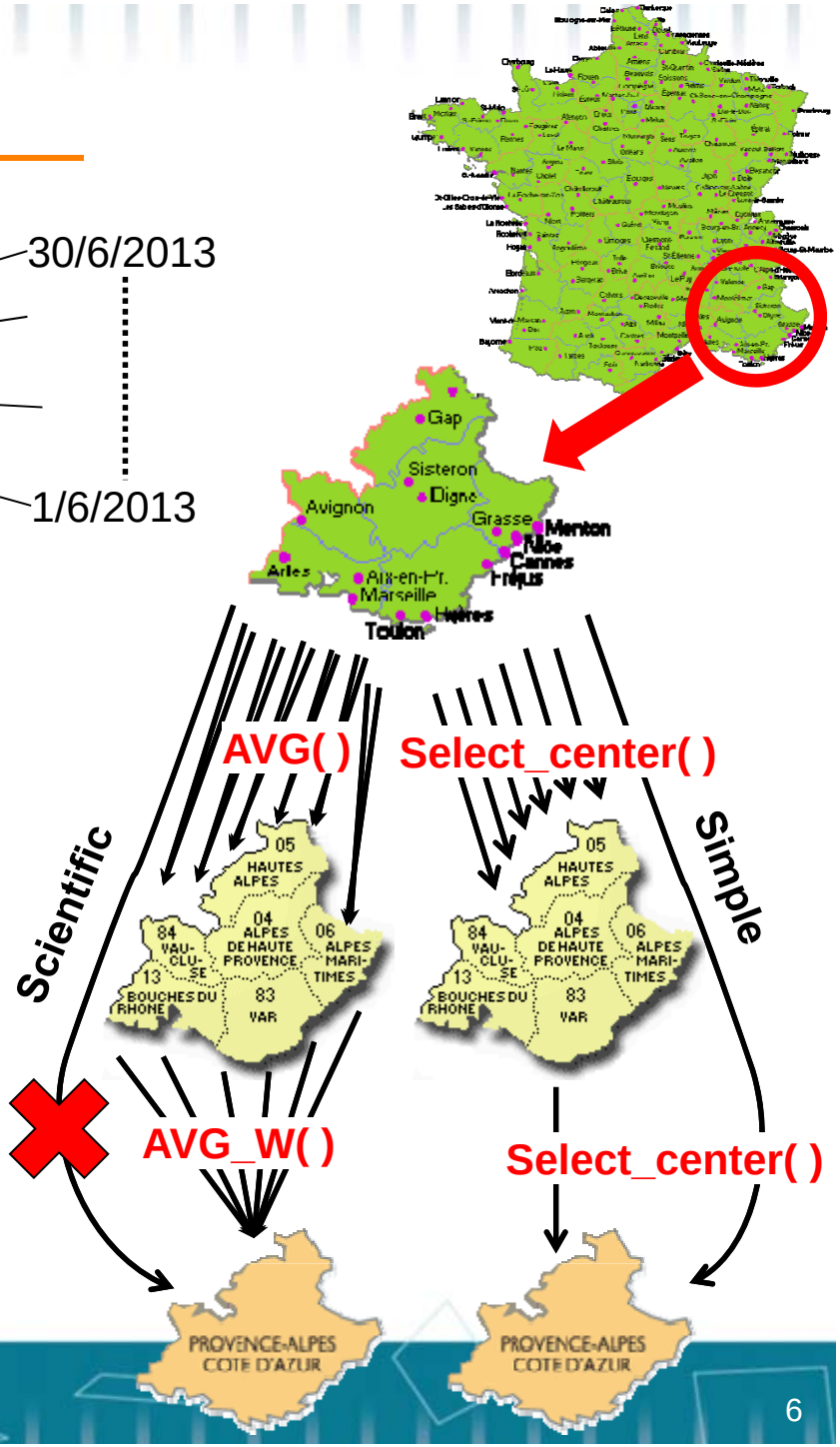
Contribution

Context (2/2)


Multidimensional model
Multidimensional table
Hierarchy
ALL
Class
Category
ALL
Continent
Country
City
Products
Year
Trimester
Sales
Quantity
Amount
Stores
Dates
Dimension
Fact
SUM( )
SUM( )
Roll Up (SUM)
Drill Down
SUM(Sales.Amount)
Dates
Year
2009
2010
2011
Stores
Country
France
210
220
220
Spain
90
115
100
SUM(Sales.Amount)
Dates
Year
2009
2010
2011
Stores
City
Toulouse
100
120
115
Bordeaux
110
100
105
Barcelona
90
115
100
Storage
Analysis

Problem Statement (1/2)





Problem Statement (2/2)

1. Variability of the aggregation function
   a) with granularity levels
   b) with hierarchies
   c) with analysis axes

2. Aggregation constraints: when the measure cannot be aggregated from the base level

Contributions

1. Extend the conceptual model

2. Extend the OLAP operators

3. Extend the multidimensional table

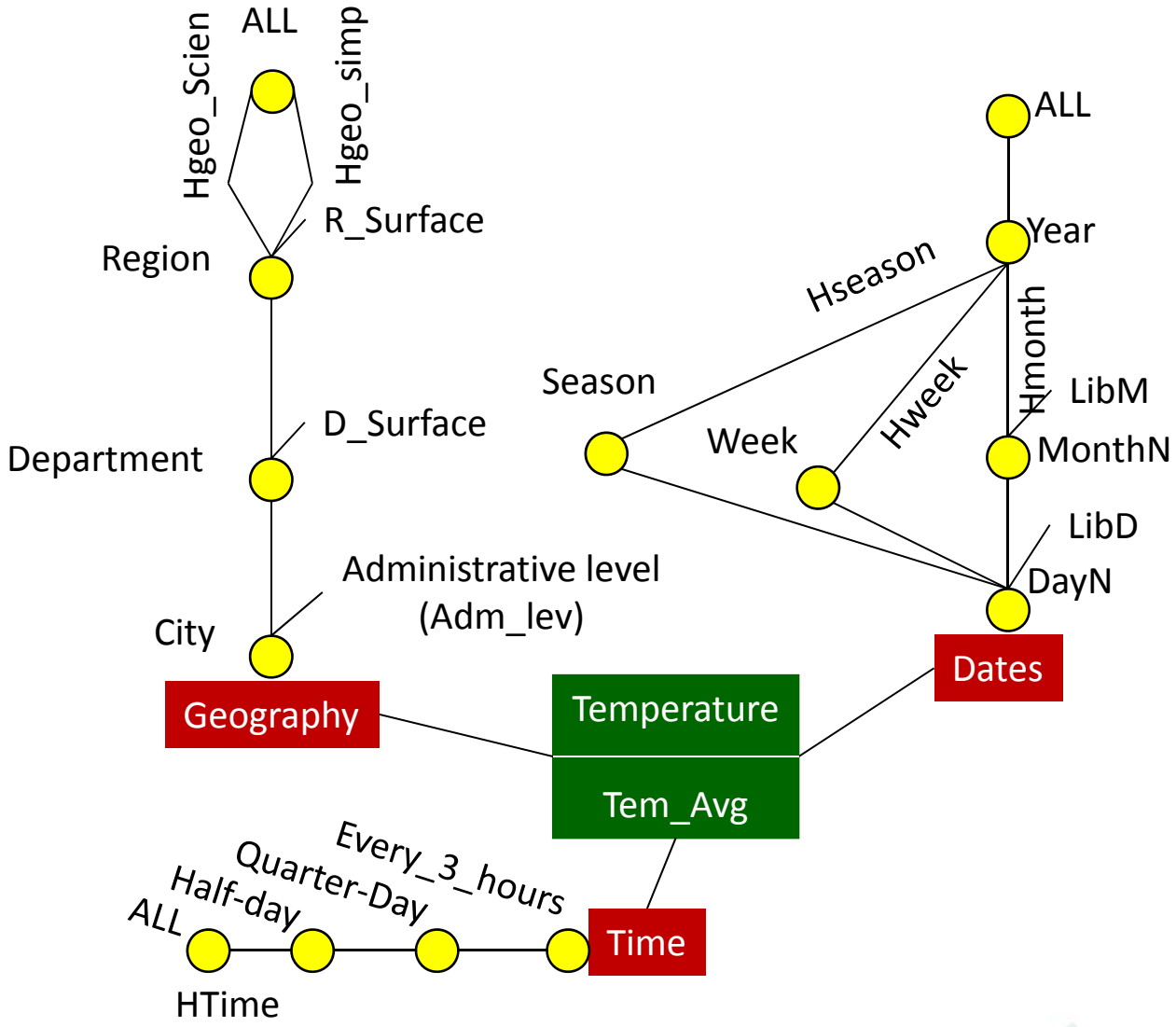
Multifunction Conceptual Data Model

- Extensions for differentiated multiple aggregations
- Graphical formalisms
  - Structural schema
  - Aggregation schemas

Multifunction Conceptual Data Model

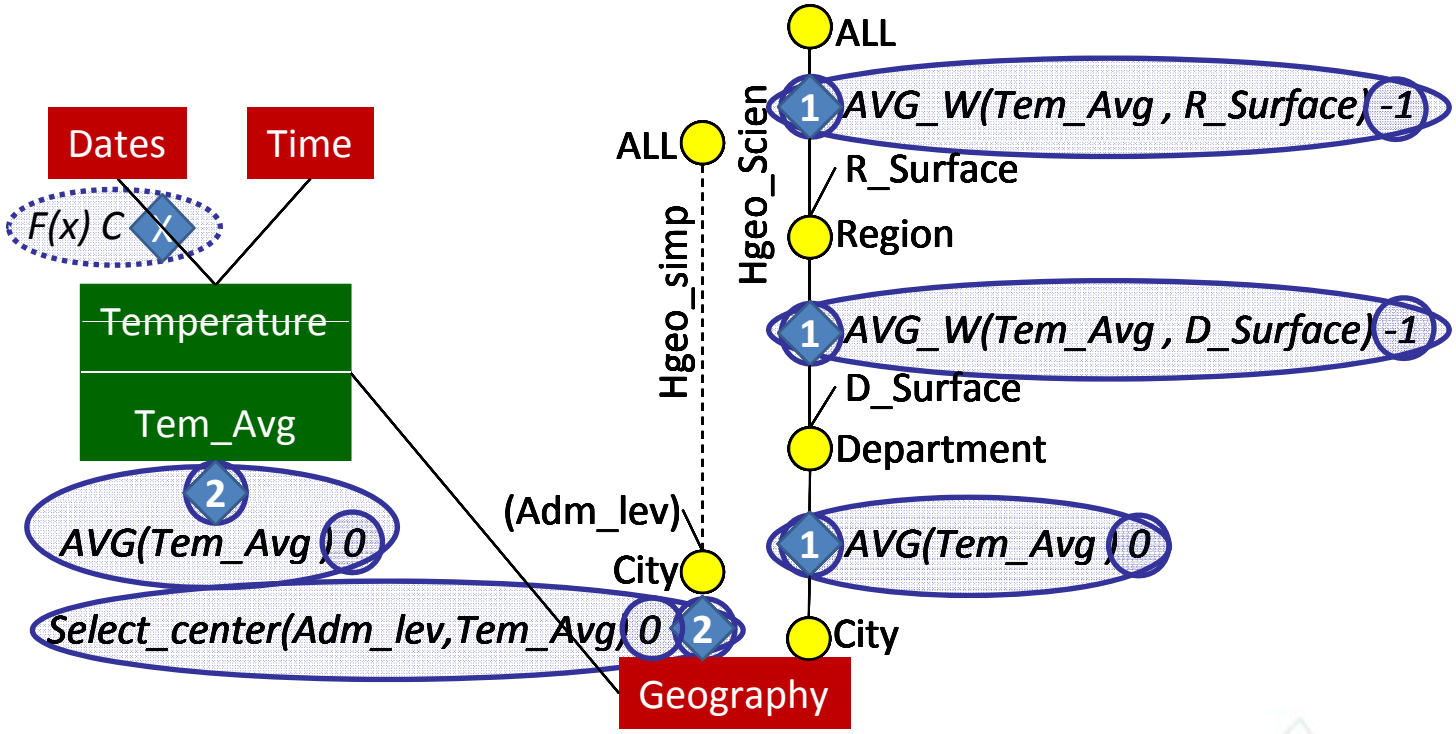
- Multidimensional schema S =
  - $F = \{F_1, \ldots, F_n\}$
  - $D = \{D_1, \ldots, D_m\}$
  - $Star : F \rightarrow 2^D$
  - $Aggregate : M \rightarrow 2^{\mathbb{N}^* \times \mathcal{F} \times 2^D \times 2^H \times 2^P \times \mathbb{N}^-}$
    - **General aggregation** : $(x_1, f, \{\}, \{\}, \{\}, x_2)$
    - **Multiple dimensional aggregation** : $(x_1, f, \{D_j\}, \{\}, \{\}, x_2)$
    - **Multiple hierarchical aggregation** : $(x_1, f, \{D_j\}, \{H_s\}, \{\}, x_2)$
    - **Differentiated aggregation** : $(x_1, f, \{D_j\}, \{H_s\}, \{P_q\}, x_2)$

Graphical Formalisms (1/2) : Structural Schema



Graphical Formalisms (2/2) : Aggregation Schema

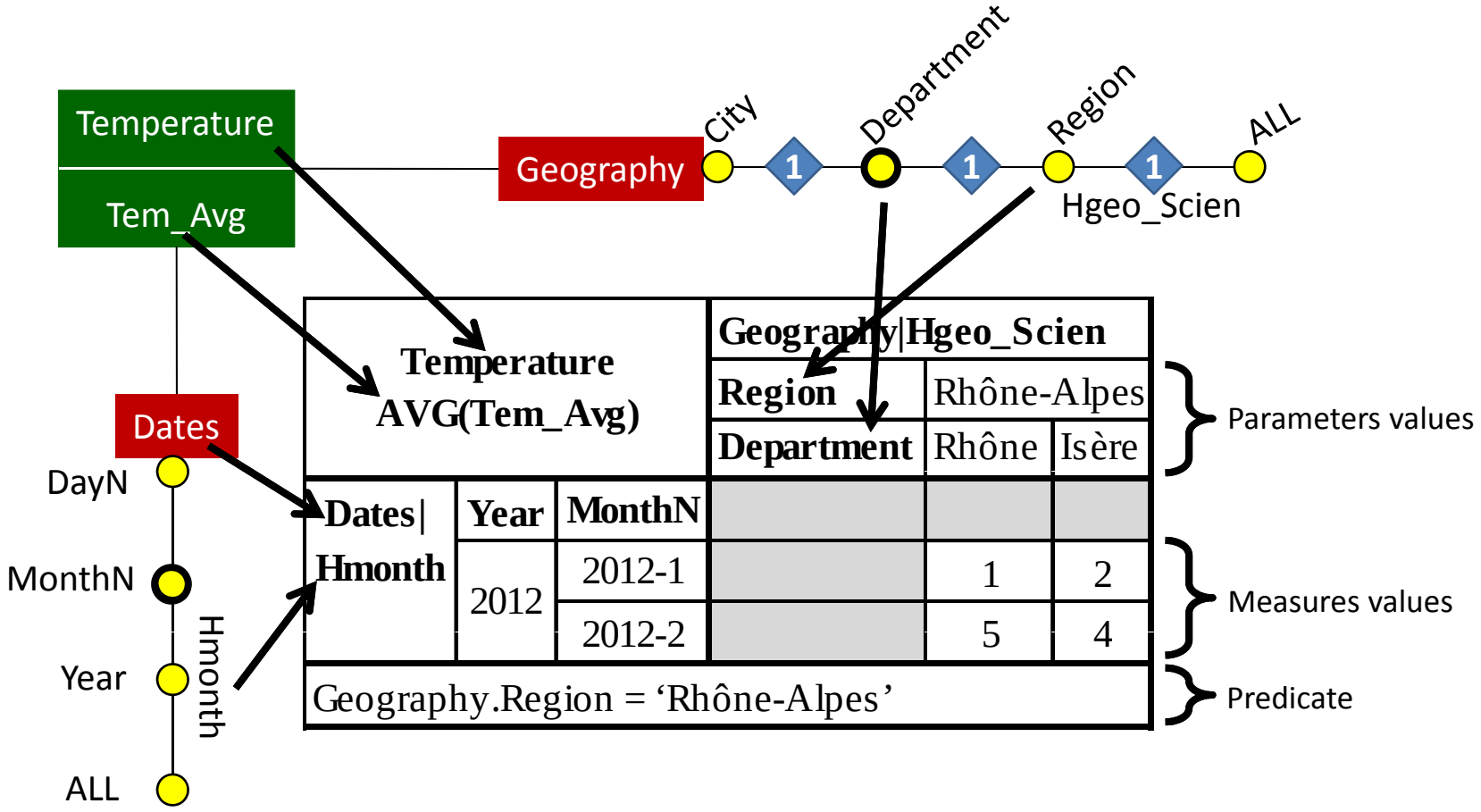
- General aggregation
- Multiple dimensional aggregation
- Multiple hierarchical aggregation
- Differentiated aggregation
- Execution order
- Aggregation constraints



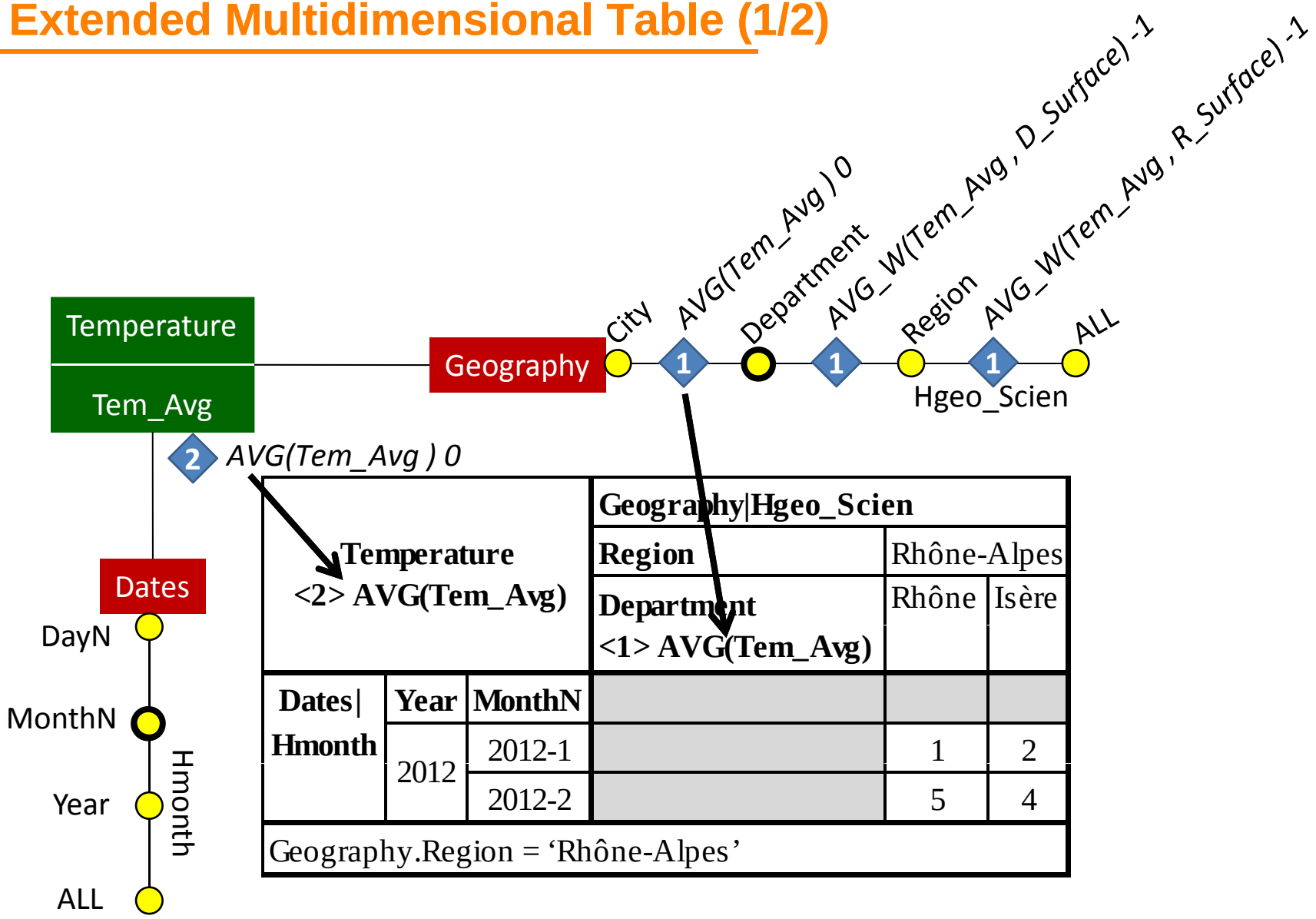
Multidimensional Table

- Classical multidimensional table
- Extended multidimensional table

Classical Multidimensional Table

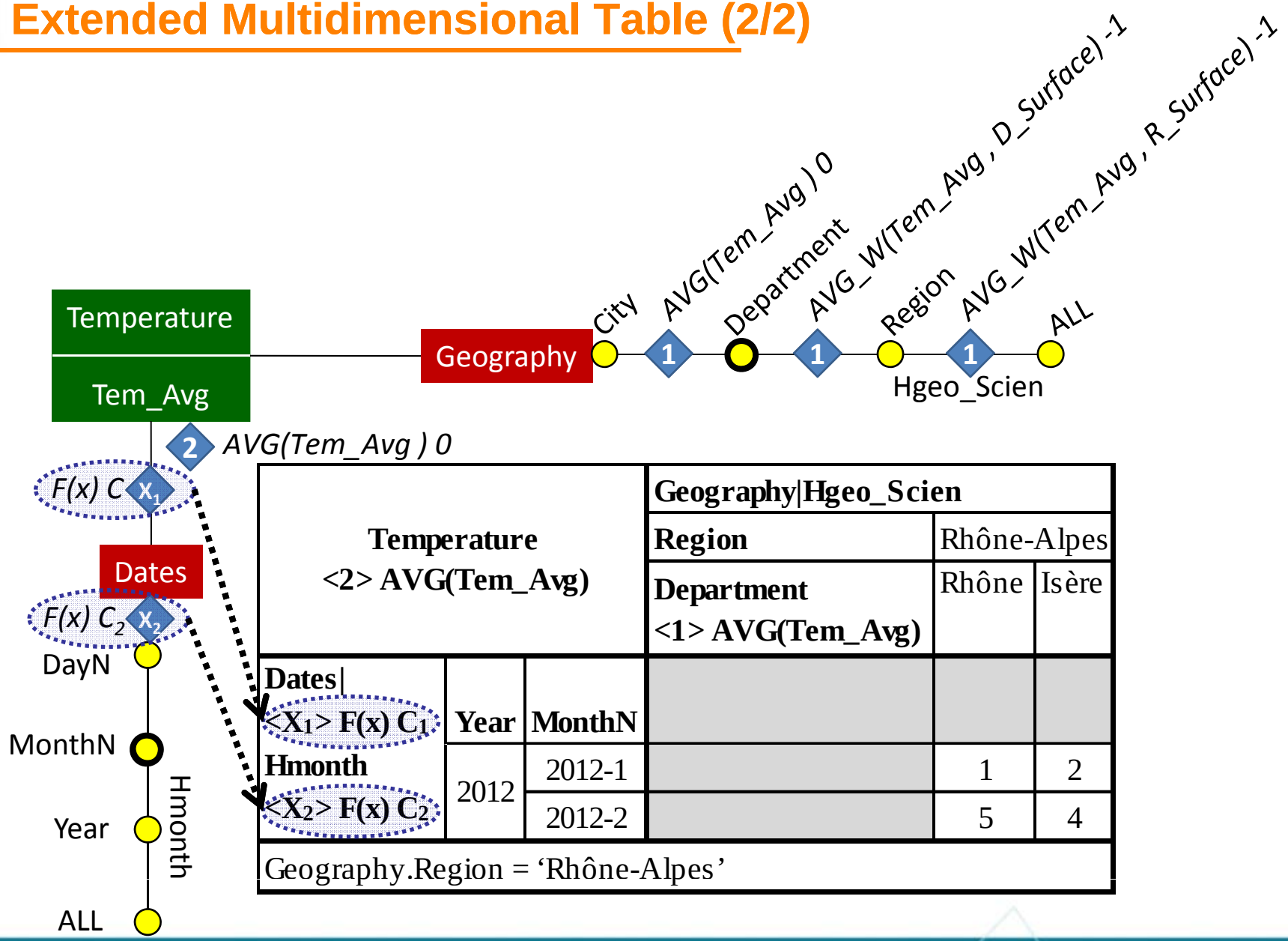


Extended Multidimensional Table (1/2)



Temperature <2> AVG(Tem_Avg)			Geography|Hgeo_Scien		
			Region	Rhône-Alpes	
			Department <1> AVG(Tem_Avg)	Rhône	Isère
Dates| Hmonth	Year	MonthN			
	2012	2012-1		1	2
		2012-2		5	4
Geography.Region = 'Rhône-Alpes'					

Extended Multidimensional Table (2/2)

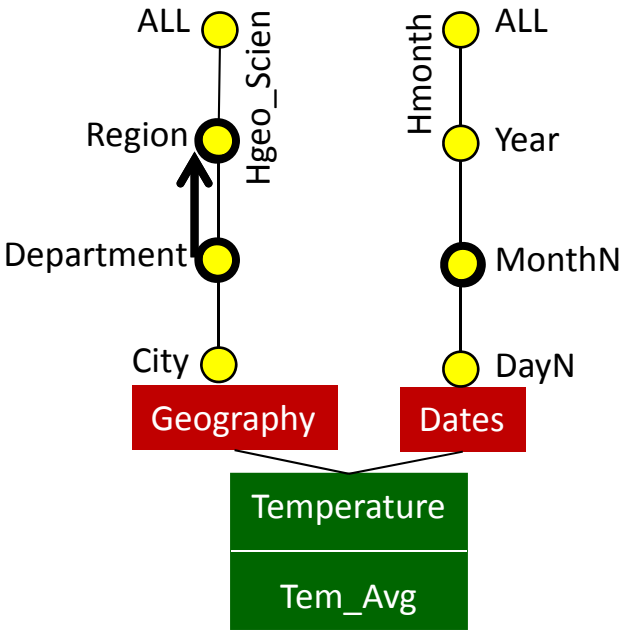


Temperature <2> AVG(Tem_Avg)			Geography|Hgeo_Scien		
			Region	Rhône-Alpes	
			Department <1> AVG(Tem_Avg)	Rhône	Isère
Dates| <X1> F(x) C1	Year	MonthN			
Hmonth <X2> F(x) C2	2012	2012-1		1	2
		2012-2		5	4
Geography.Region = 'Rhône-Alpes'					

OLAP Operators

- Classical OLAP operators (RollUp)
- Extended OLAP operators (RollUp)

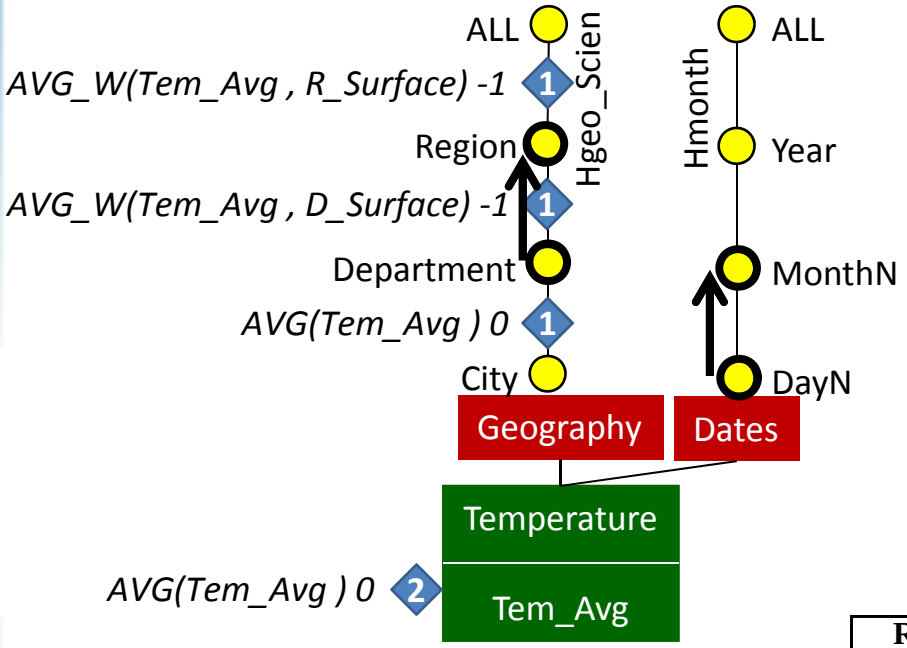
Classical OLAP Operators (RollUp)



Result1 =
SELECT G.REGION, G.DEPARTMENT, D.YEAR, D.MonthN, AVG(TT.TEM_AVG) **AS** TEM_AVG,
FROM DATES D, GEOGRAPHY G, TEMPERATURE TT
WHERE TT.ID_CITY = G.ID_CITY **AND** TT.ID_DATE = D.ID_DATE
GROUP BY G.REGION, G.DEPARTMENT, D.YEAR, D.MonthN

SELECT REGION, YEAR, MonthN, SUM(sum_Tem_Avg)/SUM(count_Tem_Avg) **AS** TEM_AVG
FROM Result1
GROUP BY REGION, YEAR, MonthN

Extended OLAP Operators (RollUp)(1/3): Motivating Example



Region	Month	Tem_Avg
Rhône-Alpes	2012-1	**0.67**

AVG(Tem_Avg)

Region	Date	Time	Tem_Avg
Rhône-Alpes	1/1/2012	00:00	-1
Rhône-Alpes	1/1/2012	12:00	1.7
Rhône-Alpes	2/1/2012	00:00	0
Rhône-Alpes	2/1/2012	12:00	2

AVG_W(Tem_Avg , D_Surface)

Region	Department	D_Surface	Date	Time	Tem_Avg
Rhône-Alpes	Rhône	3249	1/1/2012	00:00	-1
Rhône-Alpes	Rhône	3249	1/1/2012	12:00	1
Rhône-Alpes	Isère	7431	1/1/2012	12:00	2
Rhône-Alpes	Rhône	3249	2/1/2012	00:00	0
Rhône-Alpes	Rhône	3249	2/1/2012	12:00	2
Rhône-Alpes	Isère	7431	2/1/2012	12:00	2

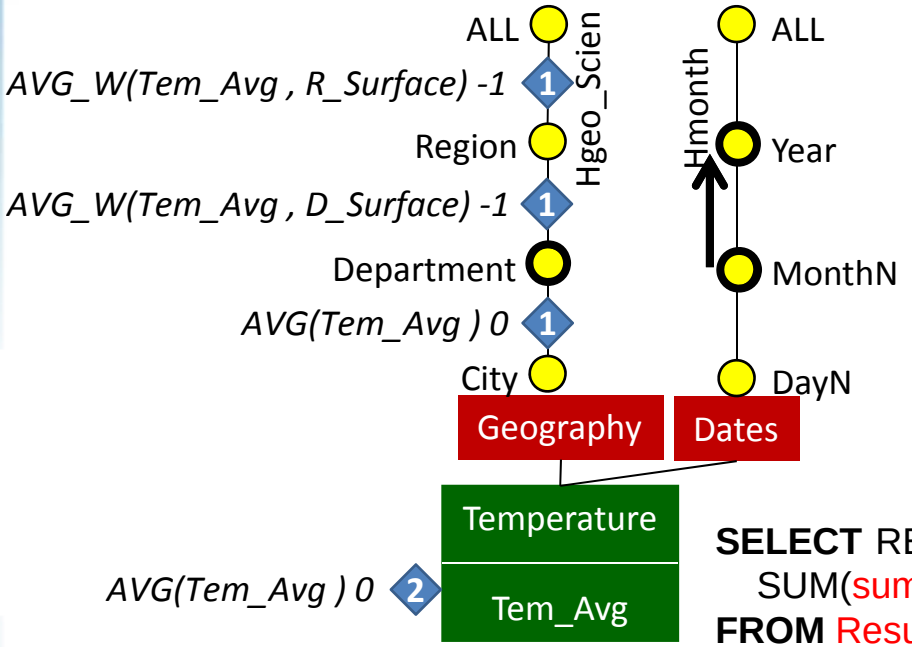
AVG(Tem_Avg)

Region	Department	D_Surface	Month	Tem_Avg
Rhône-Alpes	Rhône	3249	2012-1	0.5
Rhône-Alpes	Isère	7431	2012-1	2

AVG_W(Tem_Avg , D_Surface)

Region	Month	Tem_Avg
Rhône-Alpes	2012-1	**1.54**

Extended OLAP Operators (RollUp) (2/3)

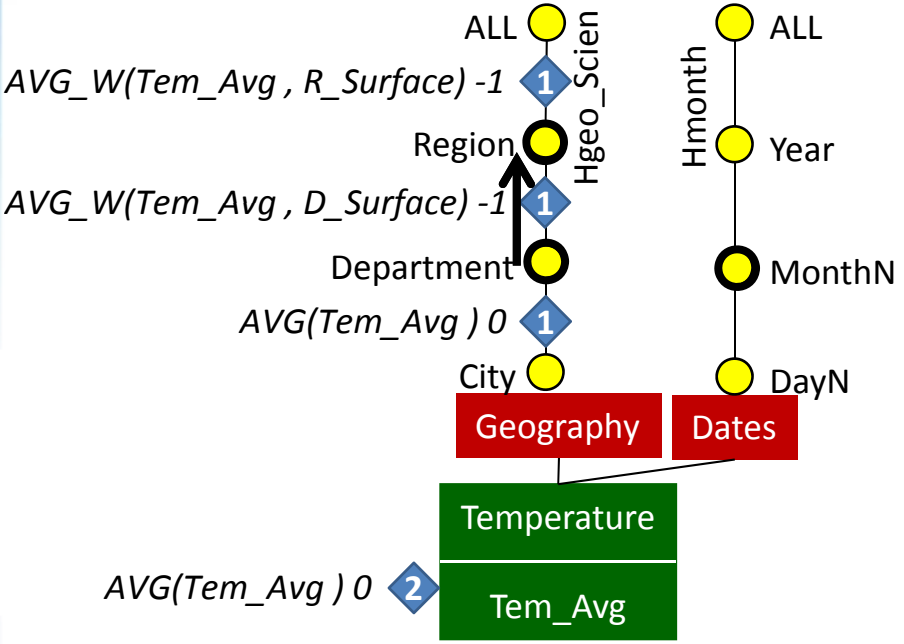



```
SELECT REGION, DEPARTEMENT, YEAR,
   SUM(sum_Tem_Avg)/SUM(count_Tem_Avg)AS TEM_AVG
FROM Result2
GROUP BY REGION, DEPARTEMENT, YEAR
```

Result2 =

```
SELECT REGION, DEPARTEMENT, YEAR, MonthN, AVG(TEM_AVG) AS TEM_AVG
FROM ( SELECT G.REGION, G.DEPARTEMENT, D.YEAR, D.MonthN, D.DayN, T.EVERY_3_HOURS,
                AVG(TT.TEM_AVG) AS TEM_AVG
       FROM DATES D, GEOGRAPHY G, TEMPERATURE TT, TIME T
       WHERE TT.ID_TIME = T.ID_TIME AND TT.ID_CITY = G.ID_CITY
          AND TT.ID_DATE = D.ID_DATE
       GROUP BY G.REGION, G.DEPARTEMENT, D.YEAR, D.MonthN, D.DayN,
                T.EVERY_3_HOURS)
GROUP BY REGION, DEPARTEMENT, YEAR, MonthN
```

Extended OLAP Operators (RollUp) (3/3)




```
SELECT REGION, YEAR, MonthN, AVG(TEM_AVG) AS TEM_AVG
FROM (SELECT REGION, YEAR, MonthN, DayN, EVERY_3_HOURS,
        AVG_W(DATA_WEIGHTED(TEM_AVG, D_SURFACE)) AS TEM_AVG
        FROM (SELECT G.REGION, G.DEPARTEMENT, D.YEAR, D.MonthN, D.DayN,
                T.EVERY_3_HOURS, G.D_SURFACE, AVG(TT.TEM_AVG) AS TEM_AVG
              FROM DATES D, GEOGRAPHY G, TEMPERATURE TT, TIME T
              WHERE TT.ID_TIME = T.ID_TIME AND TT.ID_CITY = G.ID_CITY
                AND TT.ID_DATE = D.ID_DATE
              GROUP BY G.REGION, G.DEPARTEMENT, D.YEAR, D.MonthN, D.DayN,
                T.EVERY_3_HOURS, G.D_SURFACE)
        GROUP BY REGION, YEAR, MonthN, DayN, EVERY_3_HOURS)
GROUP BY REGION, YEAR, MonthN
```

Conclusion et perspectives

Conclusion

- Extended conceptual model
  - general, multiple dimensional and hierarchical and differentiated functions
  - execution order
  - aggregation constraints
- Graphical formalisms
  - Structural schema
  - Aggregation schemas
- Extended and simplified multidimensional table
- Change in the internal mechanism of the OLAP operators

Future Work

- Optimize the algorithms of the extended OLAP operators
- Perform experiments based on a prototype
- Study the other impacts on the OLAP operators

Questions?