Beyond GIS: Spatial On-Line Analytical Processing and Big Data

Yvan Bedard, PhD, P.Eng.
Independant consultant
Strategic and Scientific Advisor, Intelli3
Researcher: Centre for Research in Geomatics, Laval University, Quebec

UNIVERSITÉ LAVAL

University of Maine
Orono, USA
September 18th, 2014

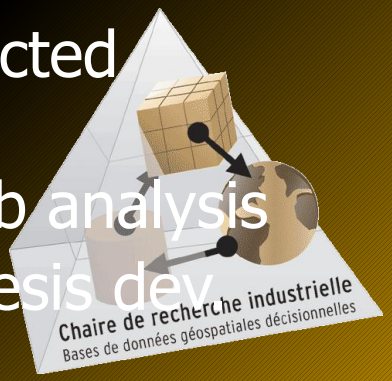
Presentation

- Origin of SOLAP
- Nature
- Evolution
- Examples of applications
- State-of-the-art for today's technology
  - Challenges that remain
- SOLAP and Big Data

ORIGINS OF SOLAP

Origins

- Organisations worldwide invest hundreds of millions of dollars annually to acquire large amounts of data about the land, its resources and uses
- These data however prove difficult to use by managers who need:
  - aggregated information - trends analysis
  - spatial comparisons - space-time correlations
  - fast synthesis over time queries - unexpected
  - interactive exploration - crosstab analysis
  - geogr. knowledge discovery - hypothesis dev.
  - etc.



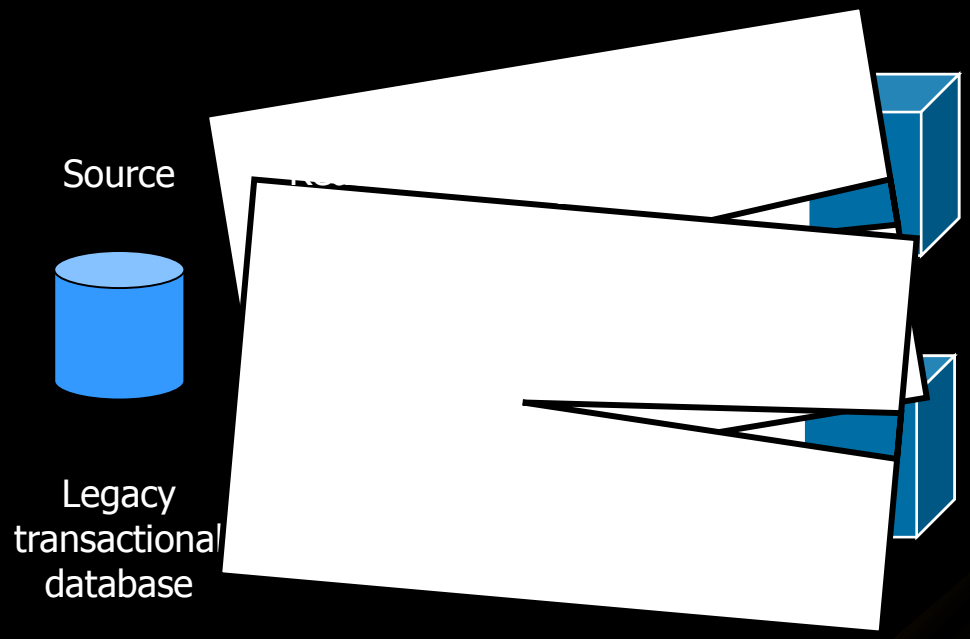
Barriers to make analysis with transactional systems

- GIS and DBMS design are transactional by nature
  - Oriented towards data acquisition, storing, updating, integrity checking, simple querying
- Transactional databases are usually normalized so duplication of data is kept to a minimum :
  - To preserve data integrity and simplify data update
- A strong normalization makes the analysis of data more complex :
  - High number of tables, therefore high number of joins between tables (less efficient).
  - Long processing time
  - Development of complex queries

Analytical approach vs transactional approach

No unique data structure is good for BOTH managing transactions and supporting complex queries. Therefore, two categories of databases must co-exist: transactional and analytical (E.F. Codd).

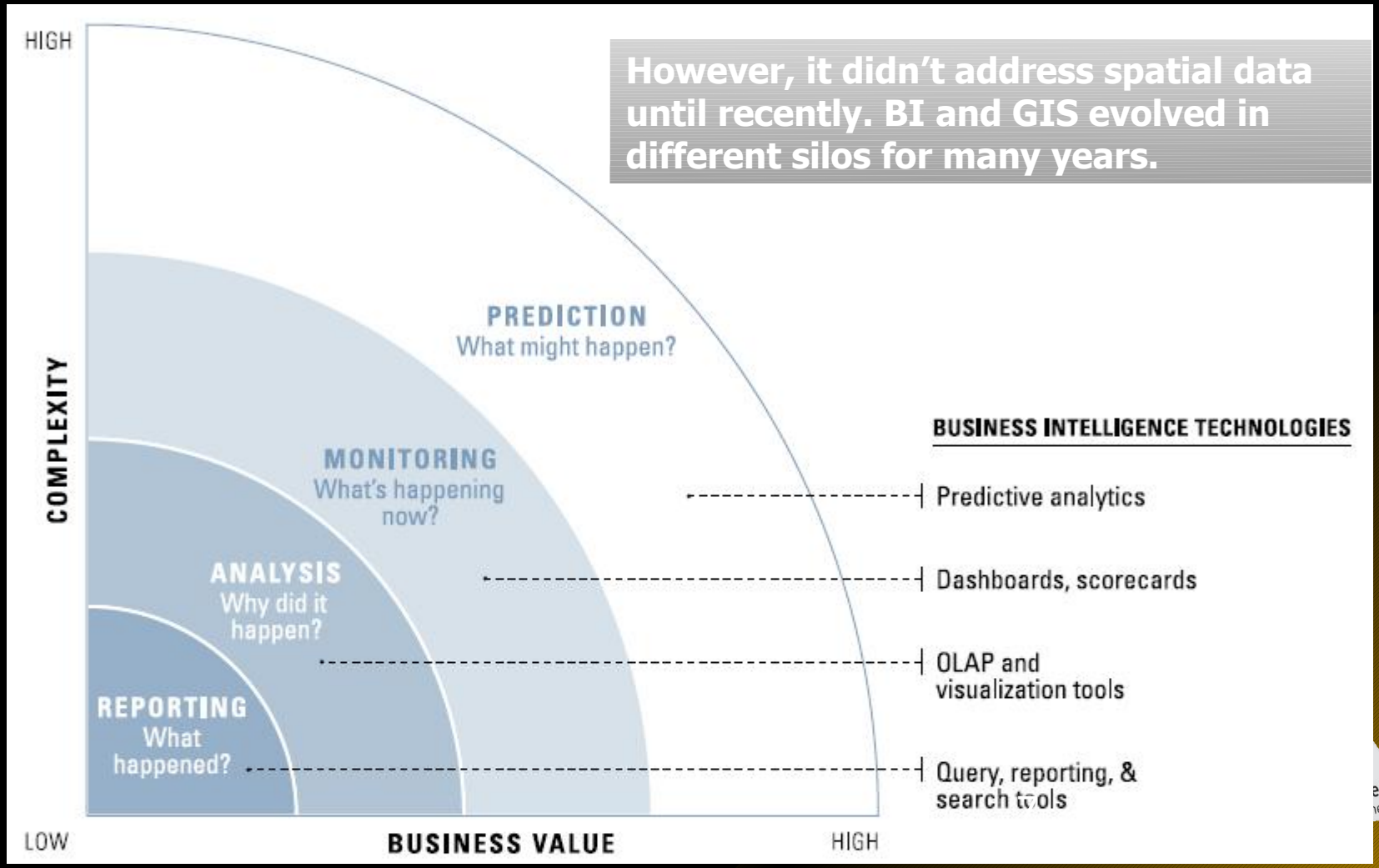
Example of co-existence: one source -> several datacubes



Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

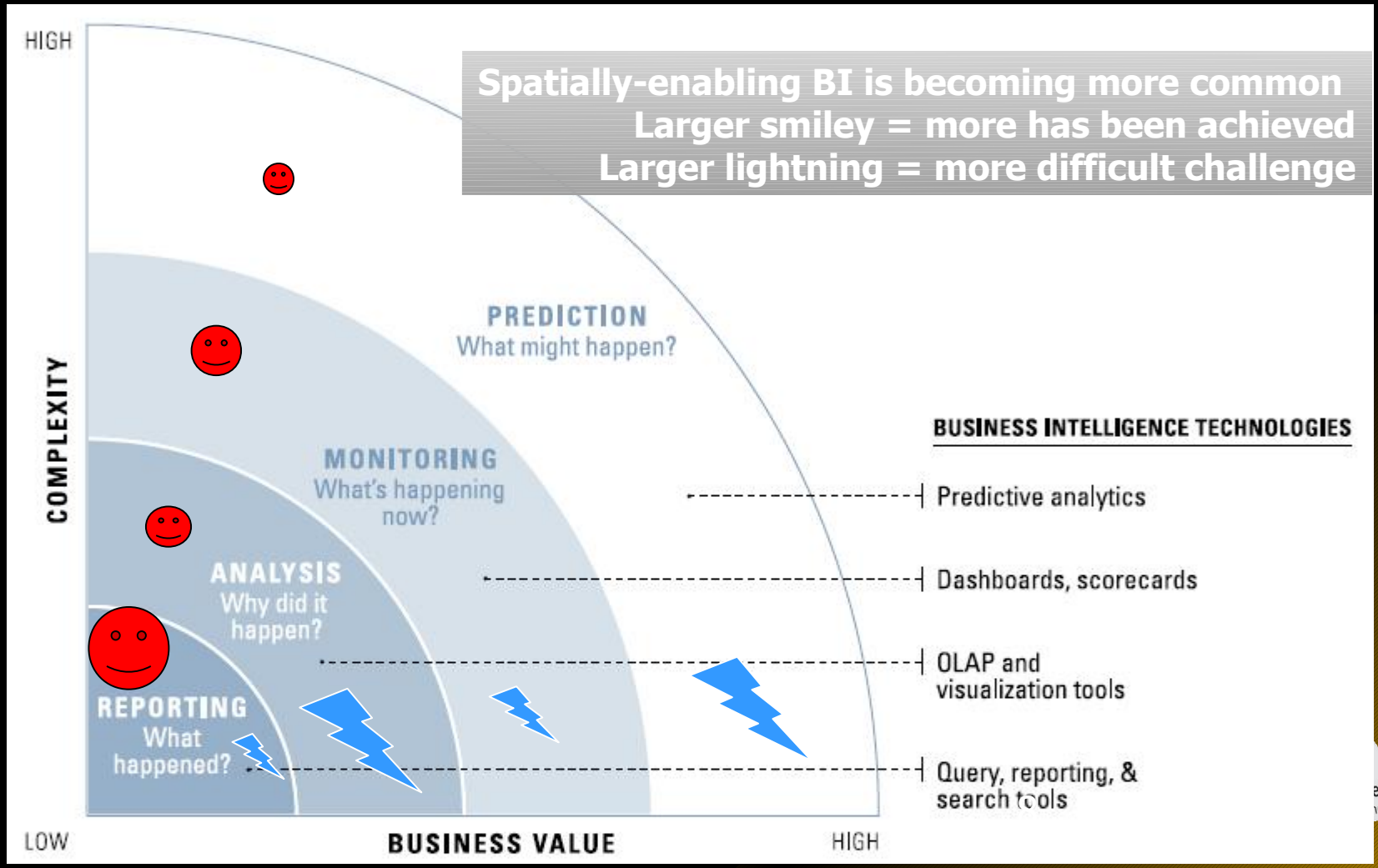
BI Market

- Business Intelligence exists since the early 1990s and its market is larger that the GIS market.



Today's Level of Integration

- Integrating GIS and BI is a recent field with a lot of potential

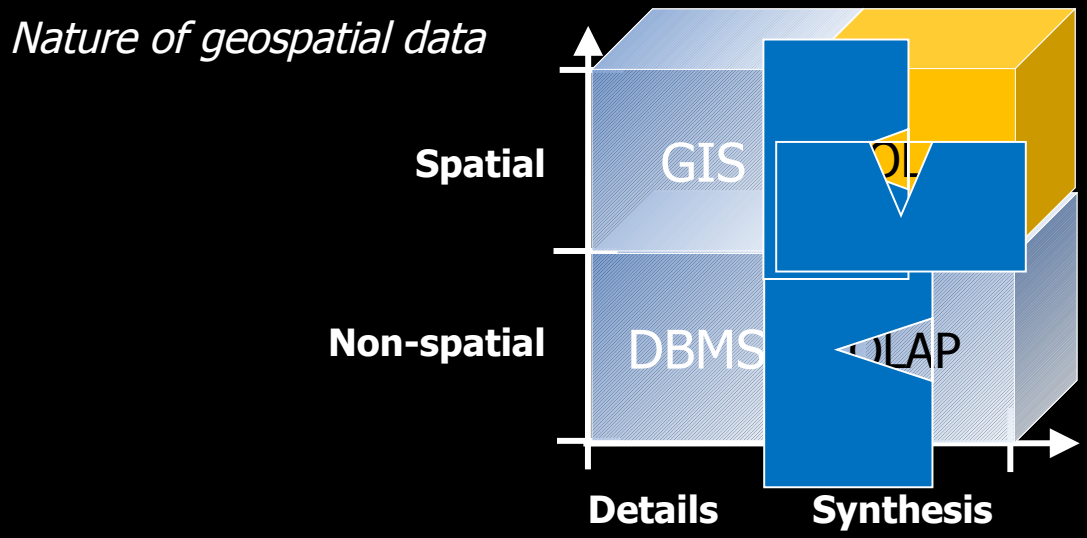


SOLAP Epochs

- 1996-2000: pionneering
  - early prototypes in universities
    - Laval U. - Simon Fraser U. - U. Minnesota
- 2001-2004: early adopters
  - advanced prototypes in universities
  - first applications in industry
- 2005-... : maturing
  - larger number of ad hoc applications
  - First commercial SOLAP technologies
- 2010-...: wide adoption
  - About 40 commercial products

NATURE OF SOLAP

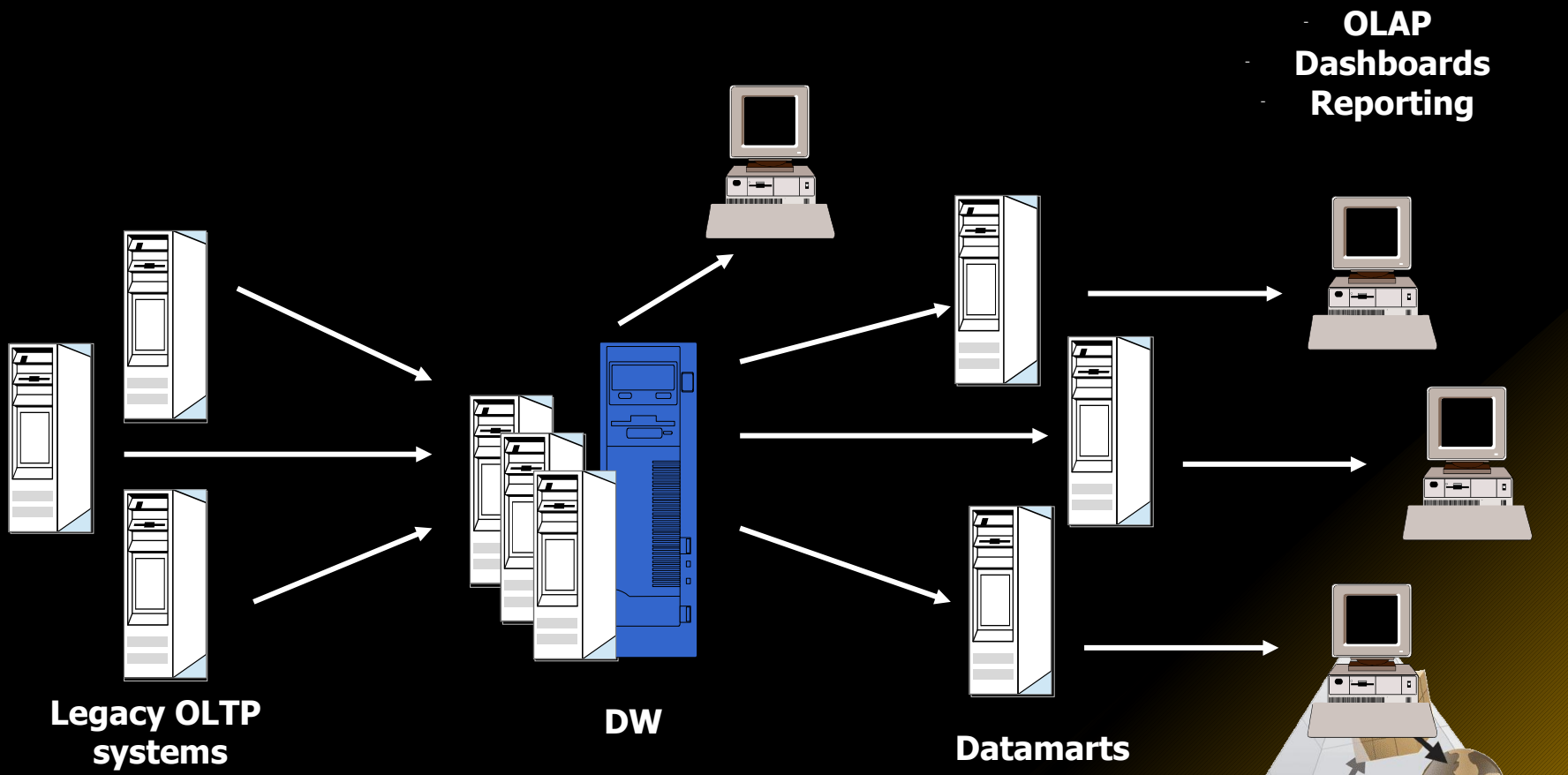
A Natural Evolution



- Add capabilities to existing systems, don't aim at replacing them
- Add value to existing data, no attempt to manage these data

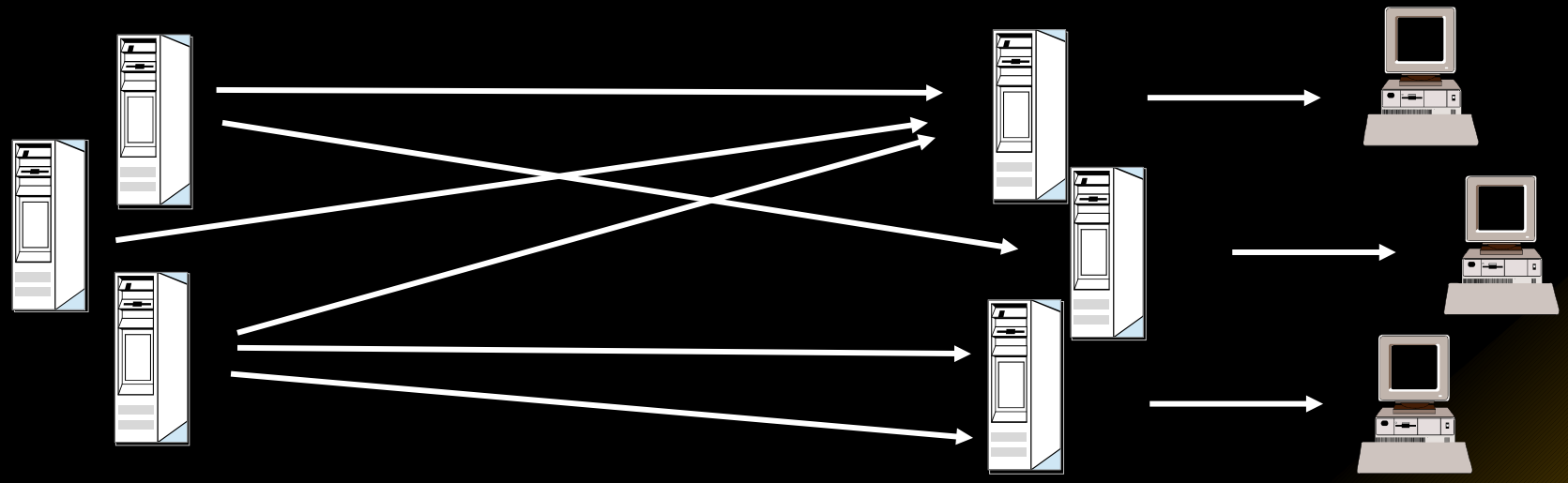
Analytical System Architectures

(ex. standard data warehouse)



Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

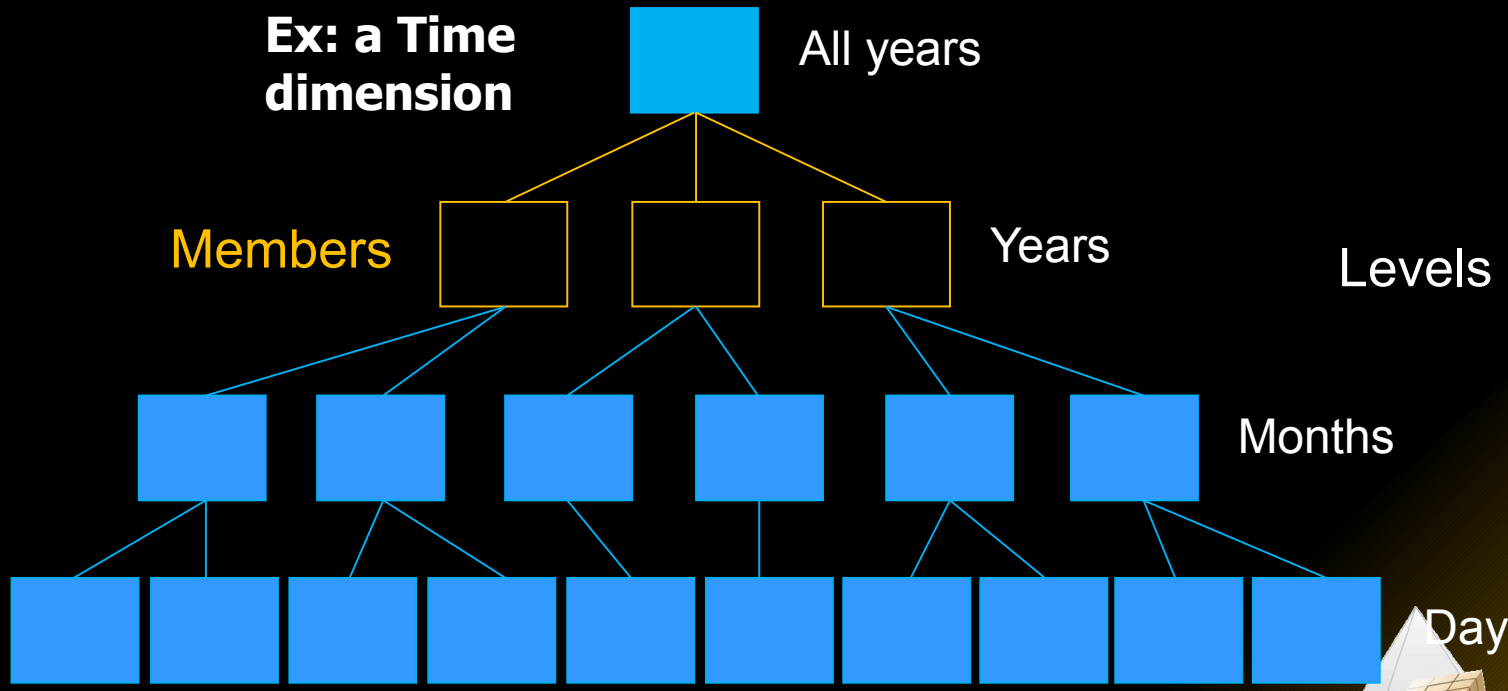
Analytical System Architectures
(ex. direct, without data warehouse)



Most projects we do have such an architecture: simpler, faster, less costly
Requires highly open SOLAP technology to connect to a variety of legacy systems (DBMS, BI, GIS, CAD, Big Data engines, etc.)

Datacube Concepts

Dimension = axis of analysis organized hierarchically

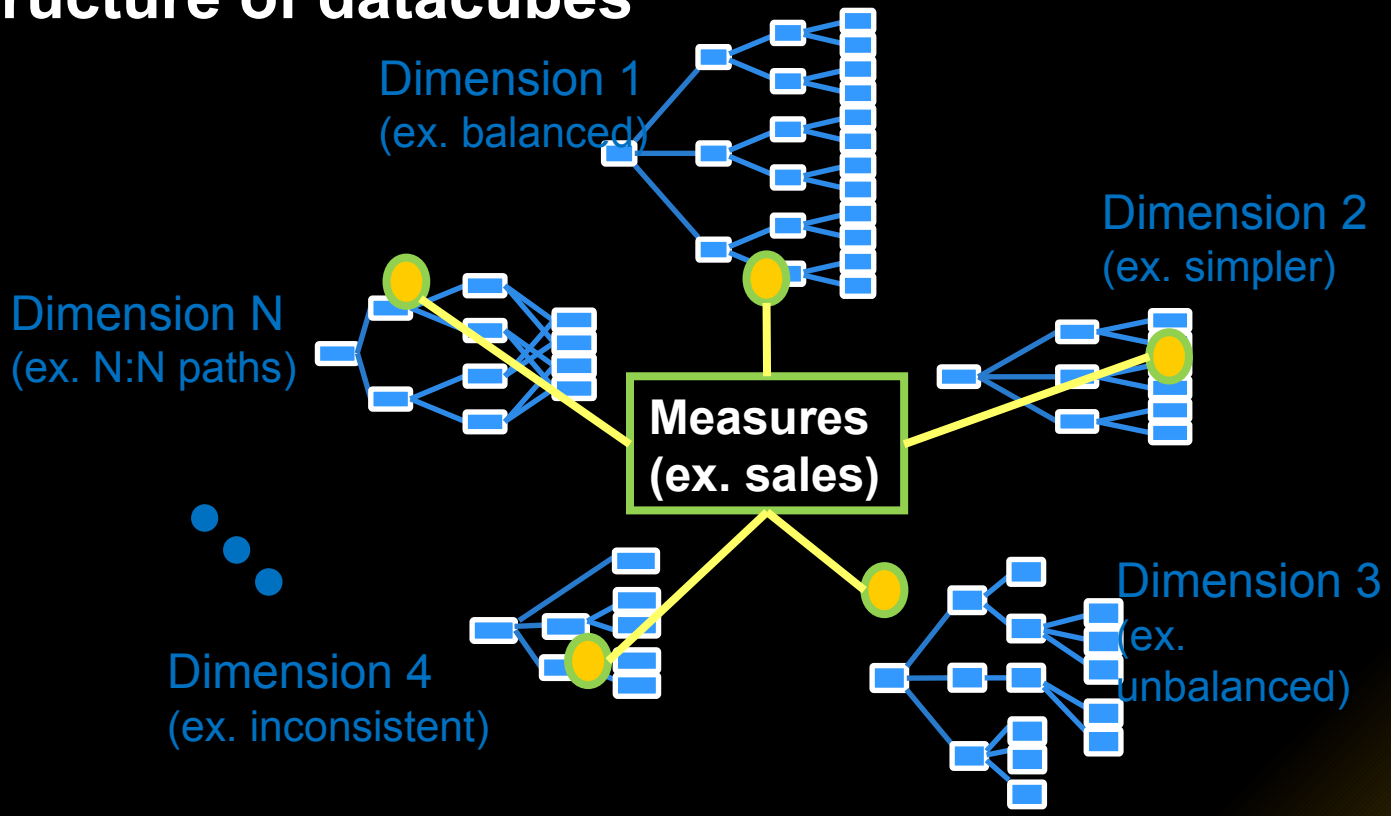


Hypercube = N dimensions

***Datacube* = casual name = hypercube**

Datacube Concepts

Structure of datacubes



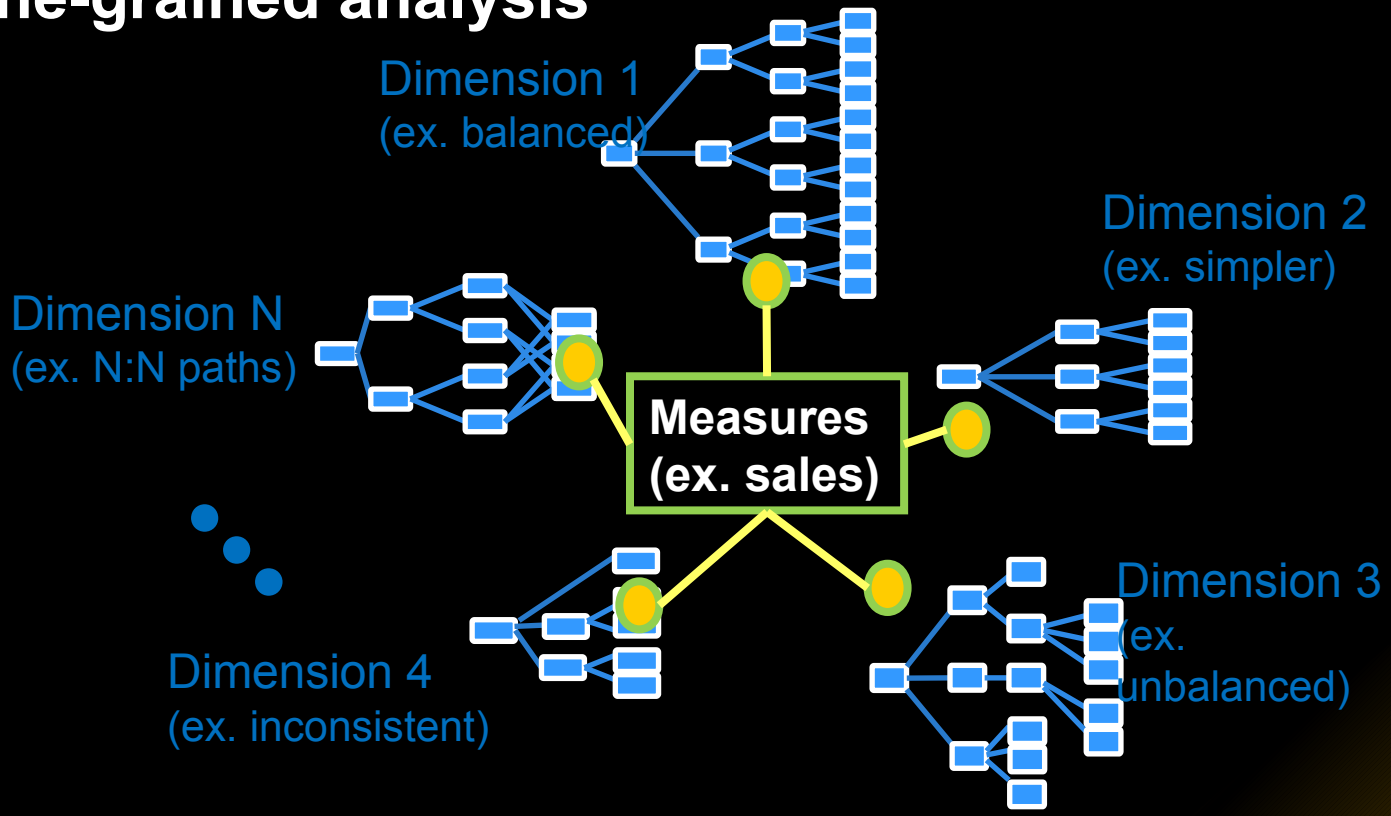
Members = filters (similar to independant variables)

Measures = result (similar to dependant variables)

Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

Datacube Concepts

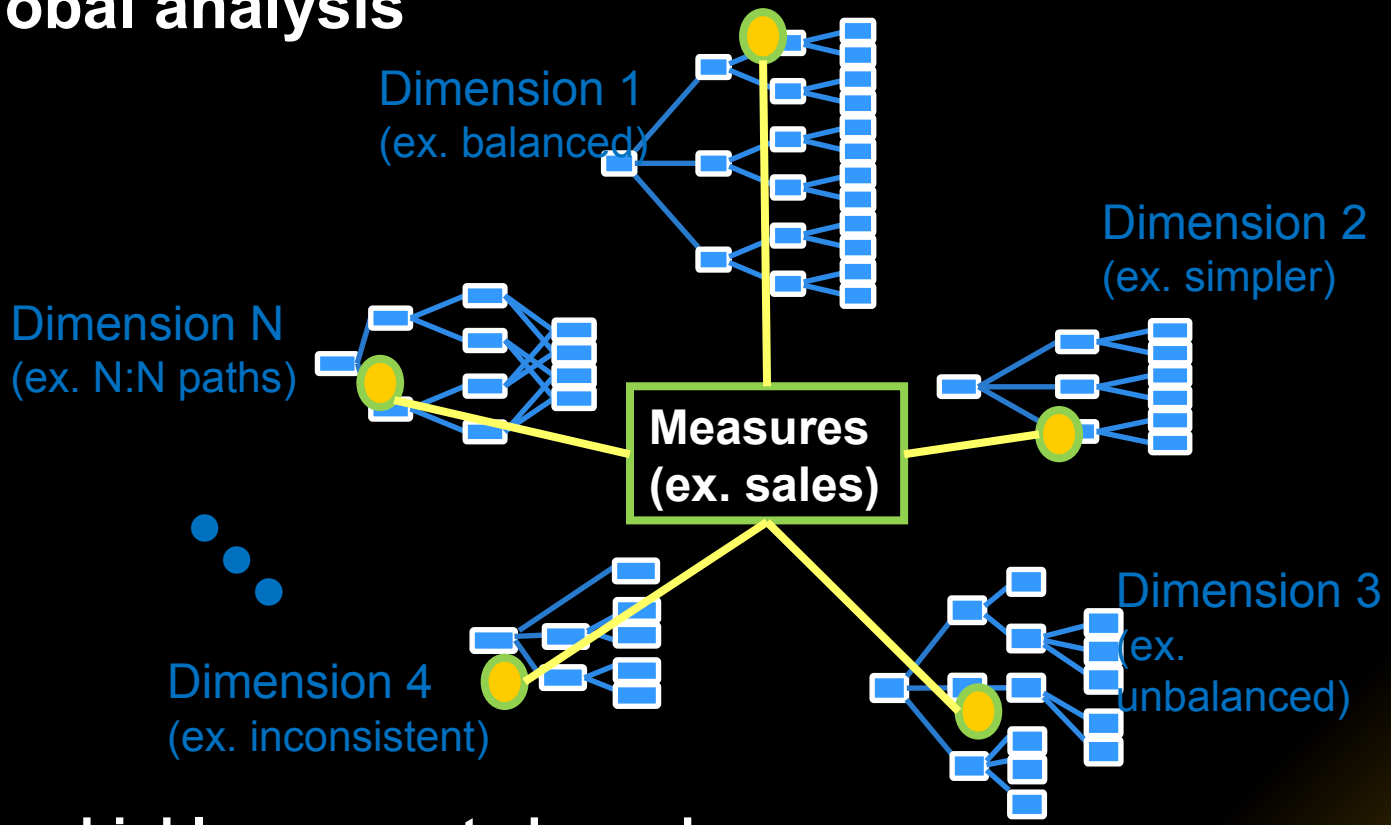
Fine-grained analysis



Uses detailed members of dimensions hierarchies

Datacube Concepts

Global analysis



Uses highly-aggregated members

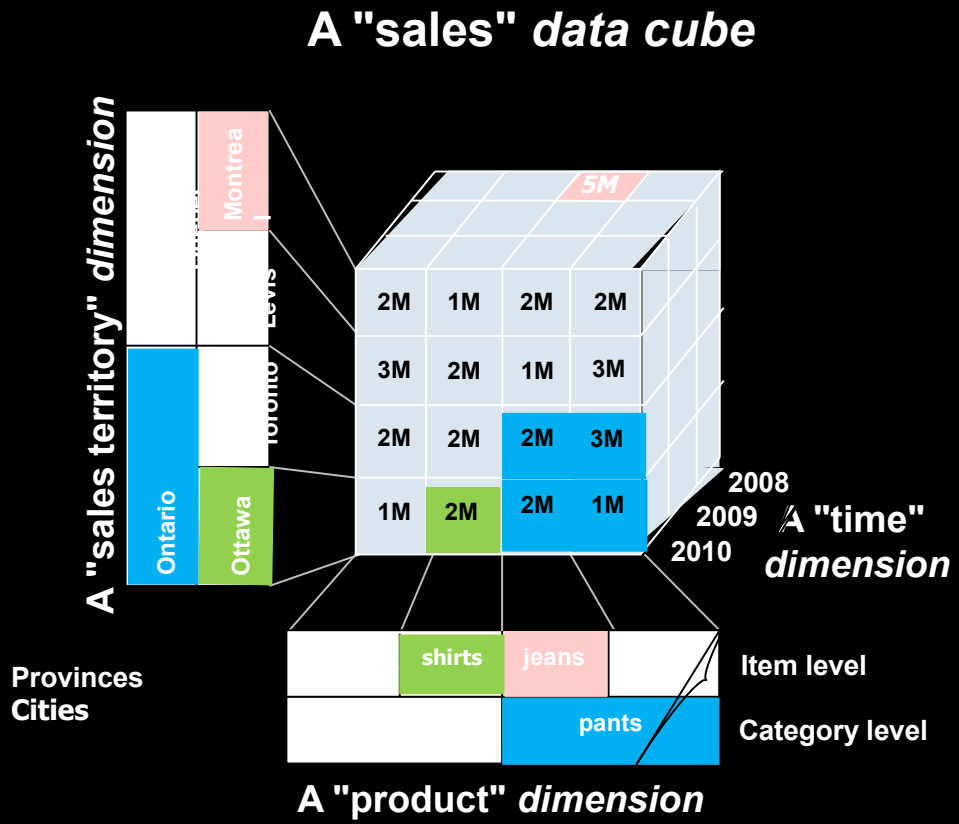
As fast as fine-grained analysis (always <10 sec.)

Requires only a few mouse clics (no query language)

Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

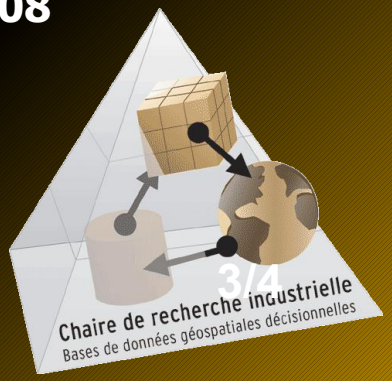
Datacube Concepts

Cube (hypercube) = all facts



Fact: **each unique combination of fine-grained or aggregated members and of their resulting measures**

Ex.: sold for 2M$ of shirts in Ottawa in 2010
Ex. : sold for 8M$ of pants in Ontario in 2010
Ex. : sold for 5M$ of jeans in Montreal in 2008



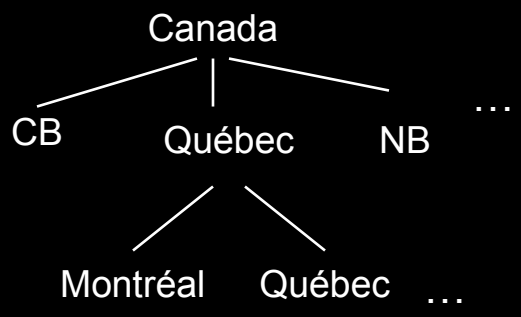
Datacube Concepts

- Data structures (MOLAP, ROLAP, HOLAP):
  - Multidimensional (proprietary)
  - Relational implementation of datacubes
    - Client and server provides the multidimensional view
      - Star schemas, snowflake schemas, constellation schemas
  - Hybrid solutions
- Query languages:
  - SQL = standard for transactional database
    - Used in ROLAP
  - MDX = standard for datacubes
    - Used in MOLAP

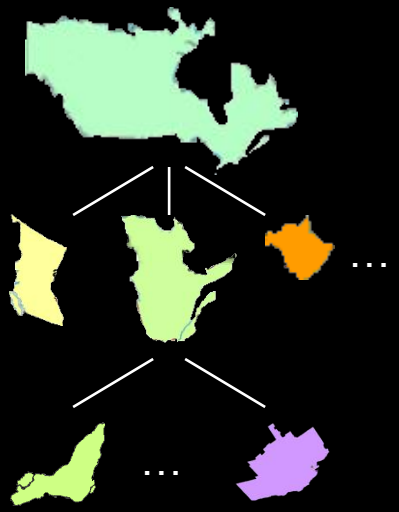
Spatial Datacube Concepts

Spatial dimensions

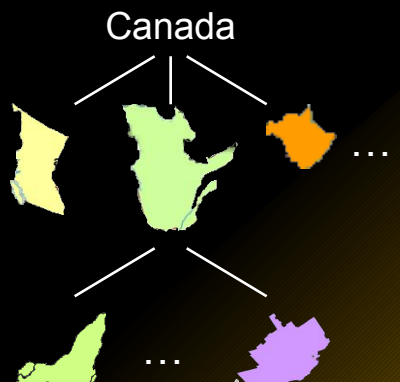
Non-geometric spatial dimension



Geometric spatial dimension



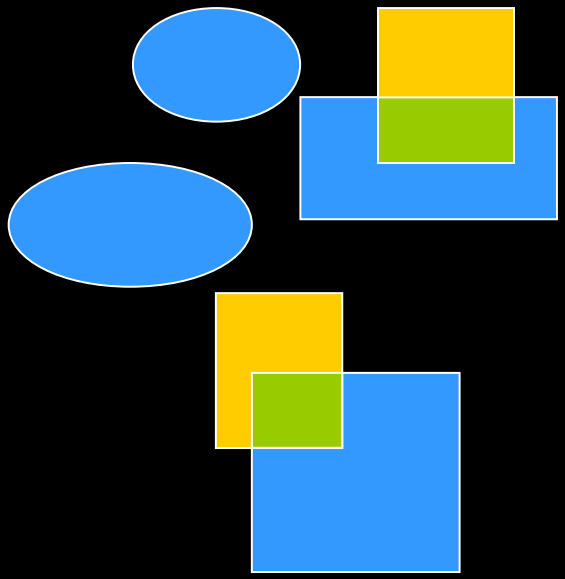
Mixed spatial dimension



N.B. more concepts exist

Spatial Datacube Concepts

Spatial measures



Spatial dimension 1

Spatial dimension 2

Metric operators

- Distance
- Area
- Perimeter
- …

Topological operators

- Adjacent
- Within
- Intersect
- …

N.B. more concepts exist

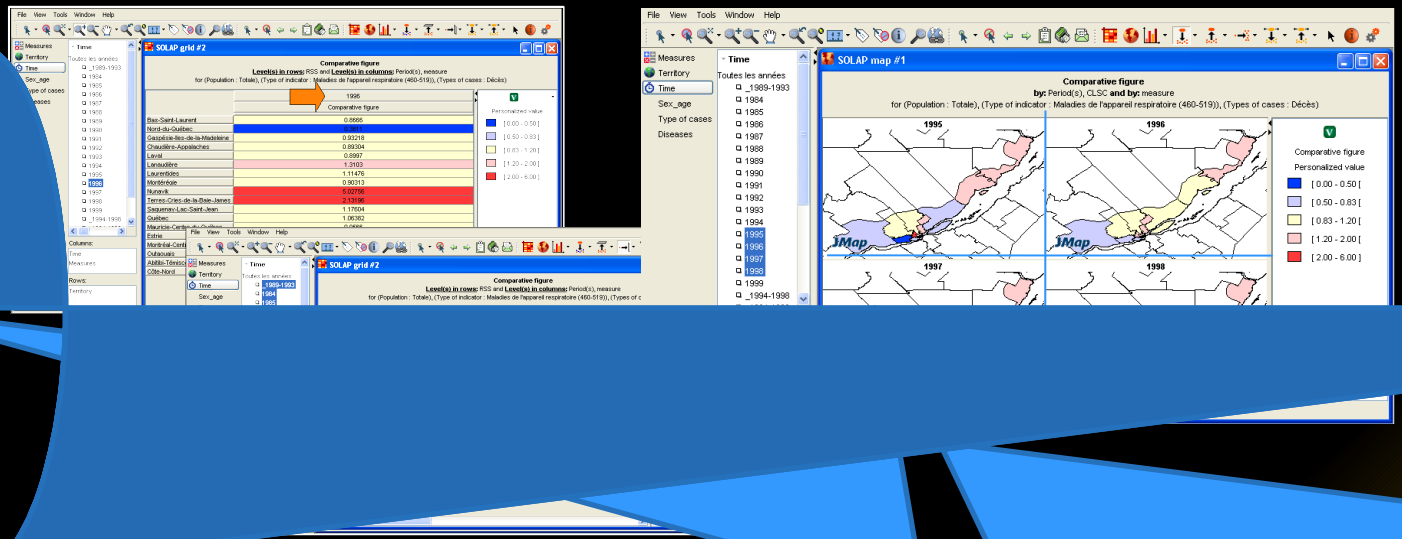
Spatial Datacube and SOLAP

- Spatial OLAP (On-Line Analytical Processing)
- SOLAP is the most widely used tool to harness the power of spatial datacubes
  - It provides operators that don't exist in GIS
- SOLAP = generic software supporting rapid and easy navigation within spatial datacubes for the interactive exploration of spatio-temporal data having many levels of information granularity, themes, epochs and display modes which are synchronized or not: maps, tables and diagrams

Characteristics of SOLAP

- Provides a high level of interactivity
  - response times < 10 seconds independently of
    - the level of data aggregation
    - today's vs historic or future data
    - measured vs simulated data
- Ease-of-use and intuitiveness
  - requires no SQL-type query language
  - no need to know the underlying data structure
- Supports intuitive, interactive and synchronized exploration of spatio-temporal data for different levels of granularity in maps, tables and charts that are synchronized at will

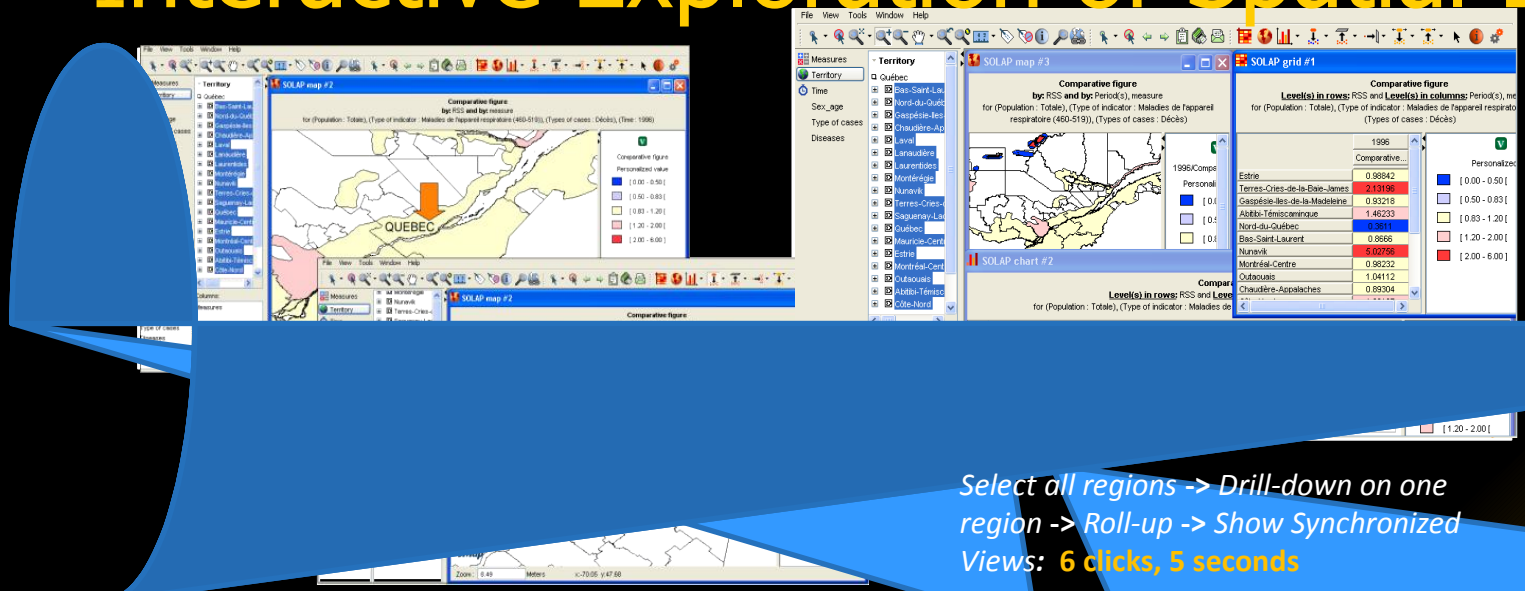
The Power of SOLAP Lies on its Capability to Support Fast and Easy Interactive Exploration of Spatial Data



Select 1 year -> Select all years ->
Select 4 years -> Multimap View:
7 clicks, 5 seconds

Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

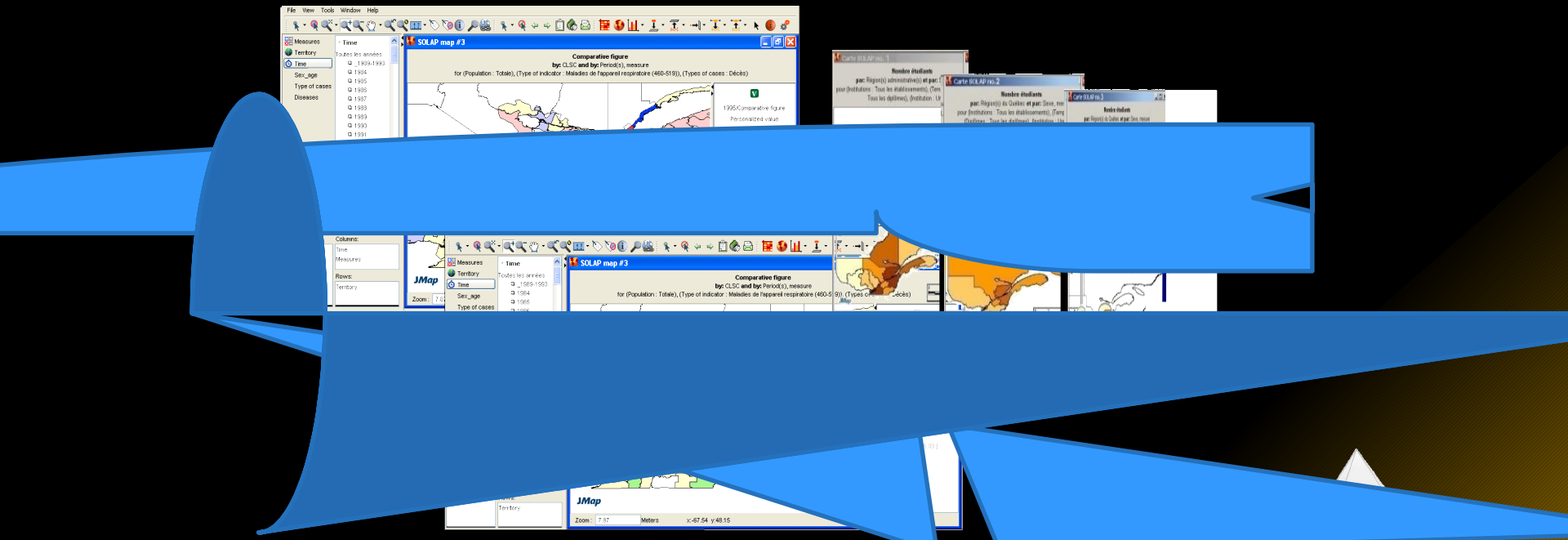
The Power of SOLAP Lies on its Capability to Support Fast and Easy Interactive Exploration of Spatial Data



Select all regions -> Drill-down on one region -> Roll-up -> Show Synchronized Views: **6 clicks, 5 seconds**

Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

The Power of SOLAP Lies on its Capability to Support Fast and Easy Interactive Exploration of Spatial Data

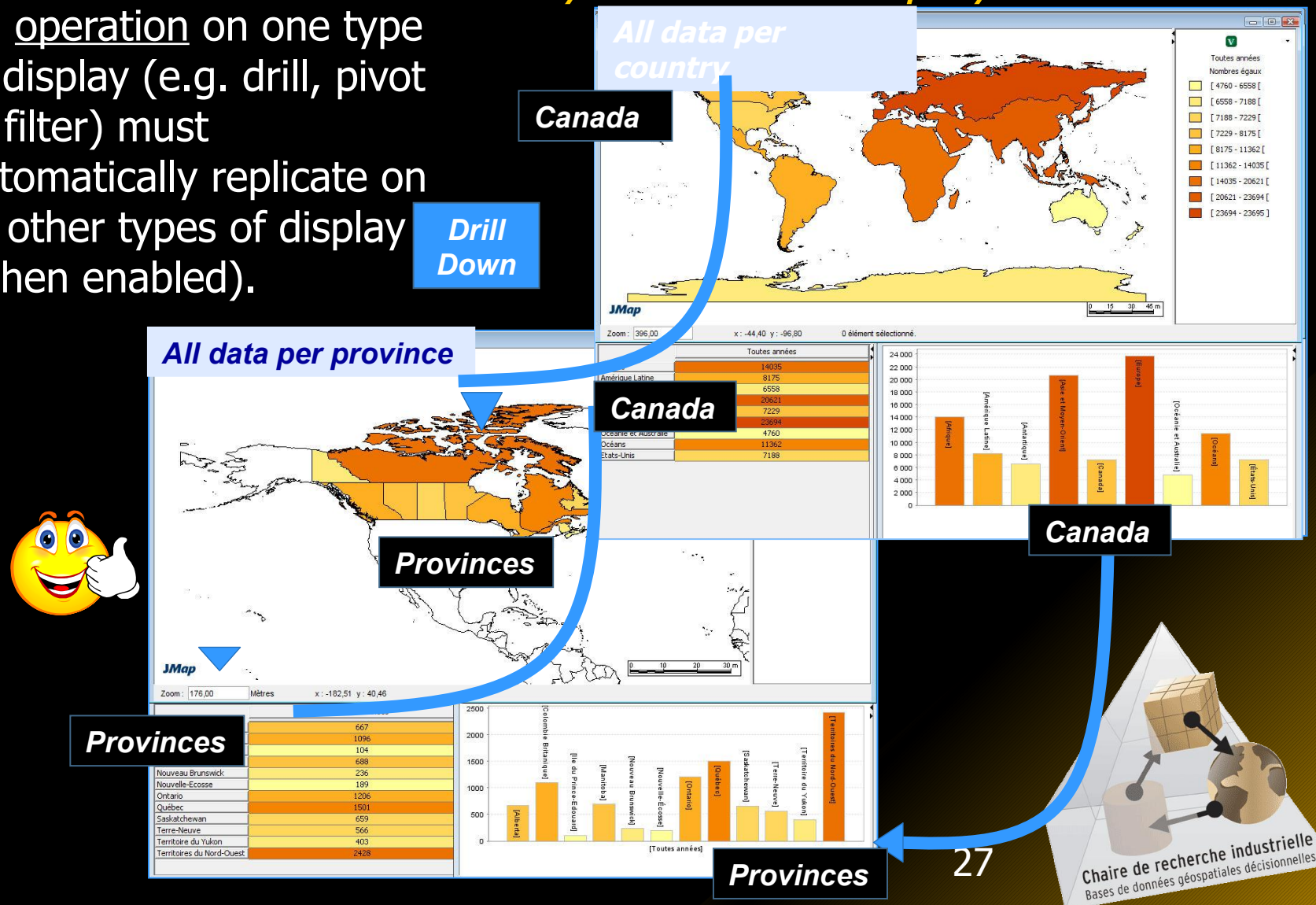


Change data -> Roll-up -> Roll-up -> Pivot ... : 6 click, 5 seconds

Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

Functionalities: *Exploration-oriented Visualization and synchronized displays*

- An operation on one type of display (e.g. drill, pivot or filter) must automatically replicate on all other types of display (when enabled).



Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

Functionalities: *Exploration-oriented Visualization and intelligent automatic mapping*

- ✓ Intelligent automatic mapping:
  - ✓ Supports user's knowledge
  - ✓ Generates coherent maps by using predefined display rules in accordance to the user's selection
  - ✓ Instantaneous display
  - ✓ No SQL involved

map thematic

classification

display type

- ✓ Manual processing:
  - ✓ Involve specific knowledge by the user (database, semiology, mapping)
  - ✓ Is time-consuming

What color, symbol , pattern ?

Which advanced map ?

EVOLUTION OF SOLAP
and
TODAY'S STATE OF THE ART

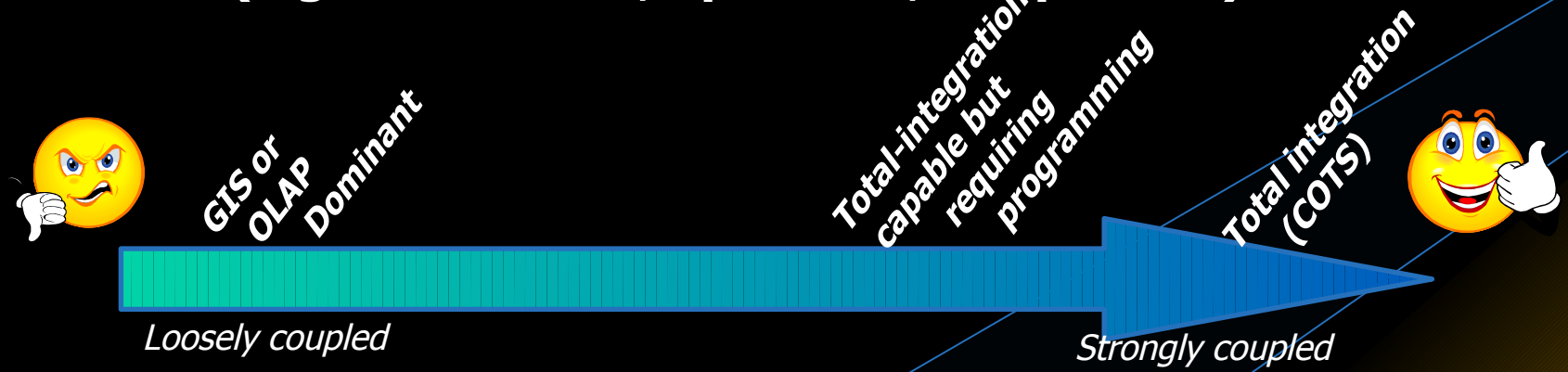
Approaches to Develop SOLAP Applications

- *Ad hoc*, proprietary programming specific to one application
- Combining GIS + OLAP capabilities
  - GIS-centric
  - OLAP-centric
  - Integrated SOLAP
    - Ad hoc programming (ex. using diverse open-source softwares)
    - SOLAP technology (the most efficient)

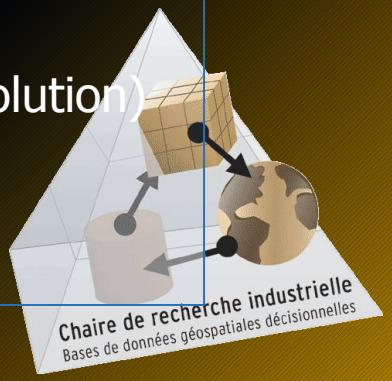
-The dominant tool offers its full capabilities but gets minimal capabilities from the other tool
-GUI provided by the dominant tool

Off-the-Shelf Integrated SOLAP

Facilitates the deployment of a SOLAP application by offering built-in elements (e.g. Framework, operators, unique GUI)

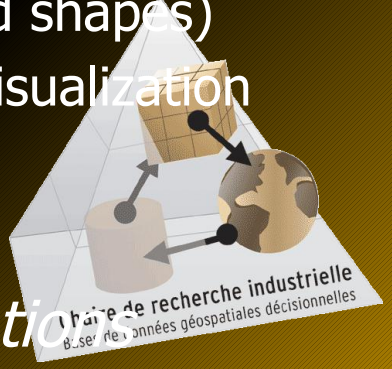


- ✓ 2 GUI *vs* common and unique GUI
- ✓ Built-in integration framework (no need to program the solution)
- ✓ Offers built-in functionalities to visualize and explore data
- ✓ No dominant component



Commercial Offerings

- About 40 SOLAP-like products exist
  - Most :
    - Run with only one GIS or DBMS or BI software
    - Are OLAP-centric or GIS-centric
    - Are limited to one type of datacube (ROLAP or MOLAP)
    - Have limited cartographic capabilities
      - Geometry:
        - number of spatial dimensions
        - types of spatial dimensions (ex. alternate)
        - Types of geometry (ex. lines, aggregated shapes)
      - "Intelligent" mapping rules for efficient geovisualization
      - Often ignore ISO or OGC standards
- *The technology that came out of our lab, Map4Decision, doesn't suffer from these limitations*

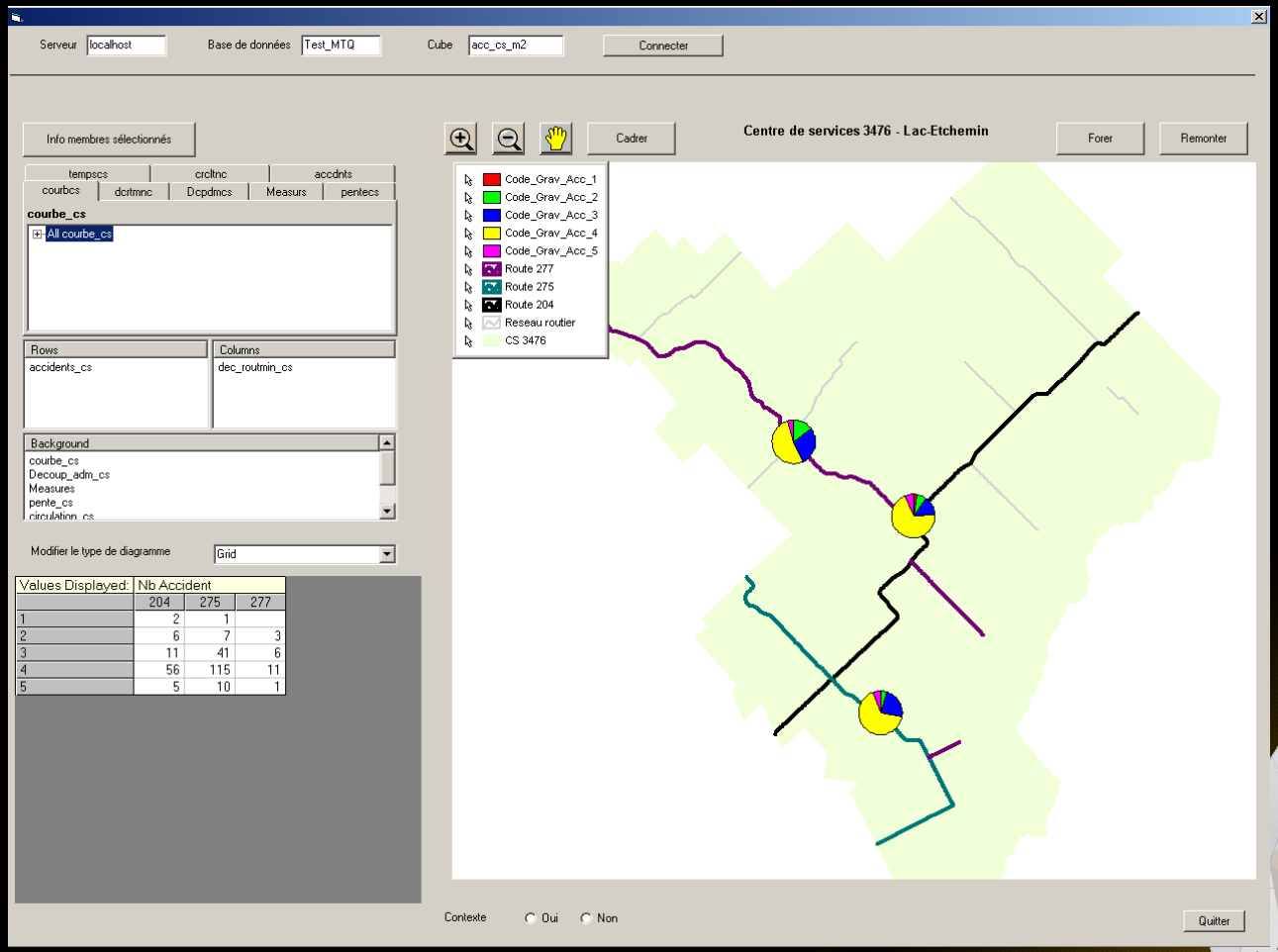


EXAMPLES OF SOLAP APPLICATIONS

Experiences since 1996

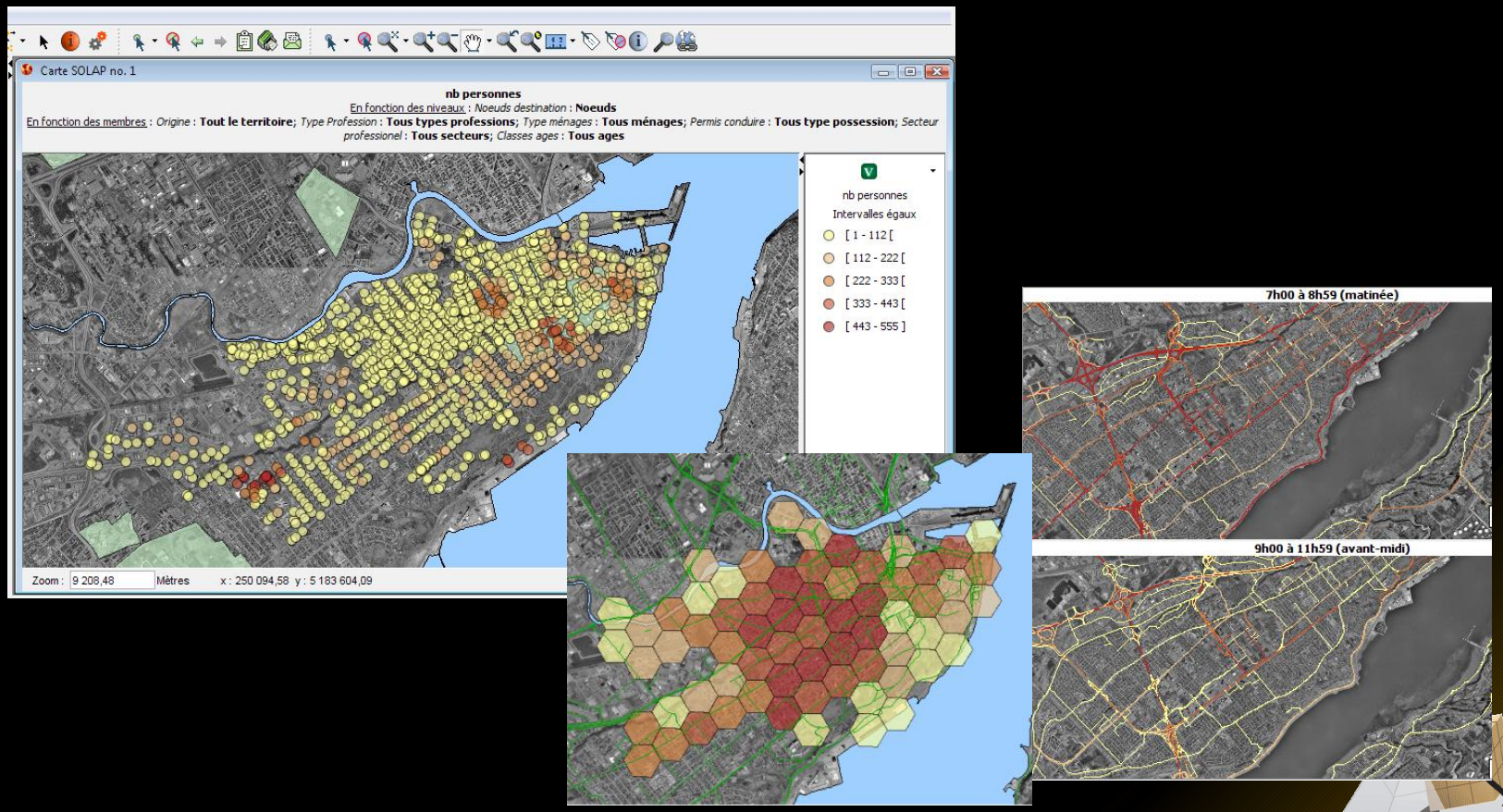
- Besides developing theoretical concepts, we have experimented with several technologies to build SOLAP applications and test concepts
- Experimentations in:
  - forestry - agriculture - public health
  - **transport** - search & rescue - sports
  - recruitment - archeology - infrastructures
  - climatology - erosion - etc.
- Experimentations with:
  - MapX - ArcGIS - Geomedia - SoftMap
  - Oracle - Access - SQL-Server - MySQL
  - Proclarity - Cognos - etc.

Example: Road Safety Analysis (Transport Quebec)



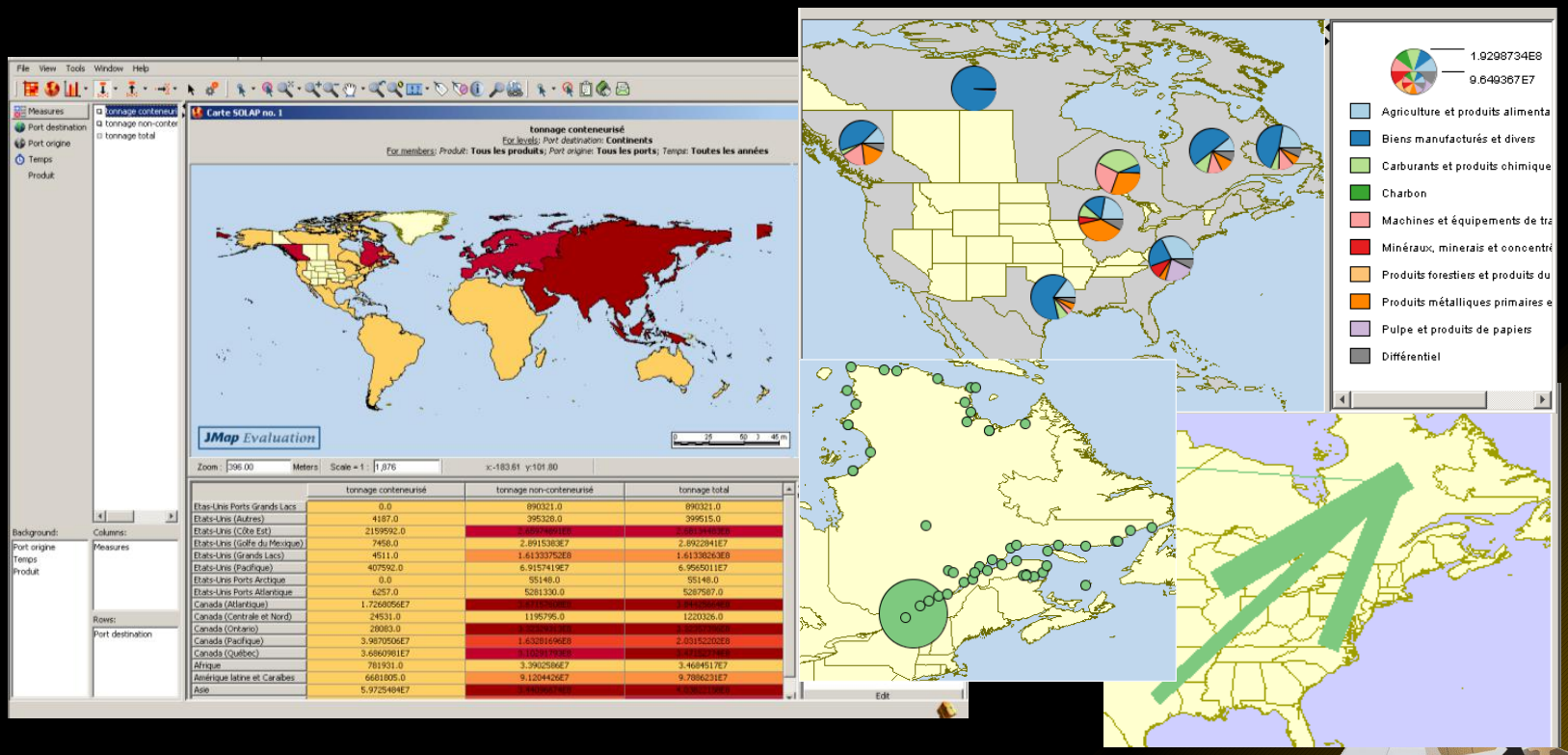
Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

Example: Origin-Destination Analysis (Cities + Transport Quebec)



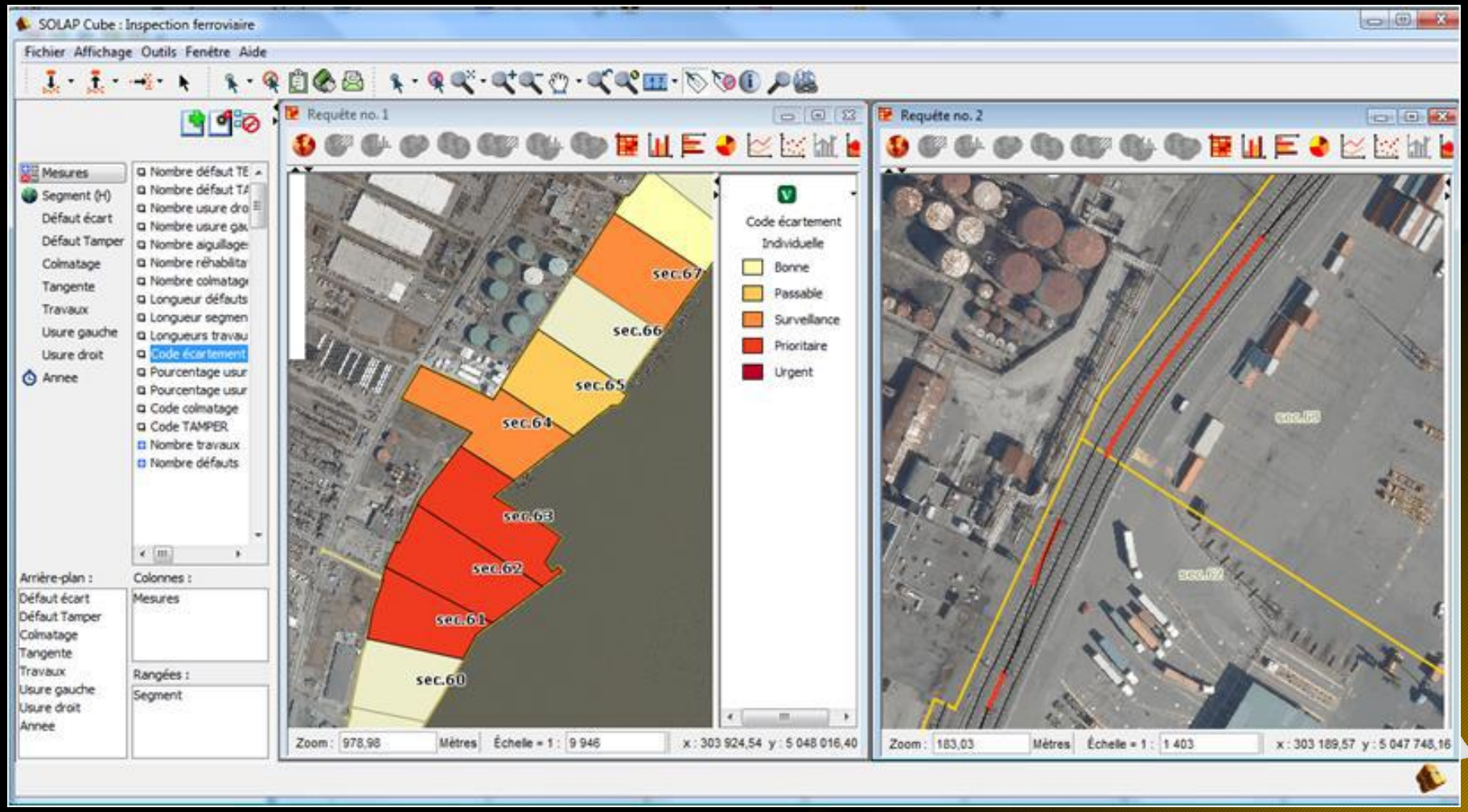
Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

Example: Marine Transportation (Transport Quebec)

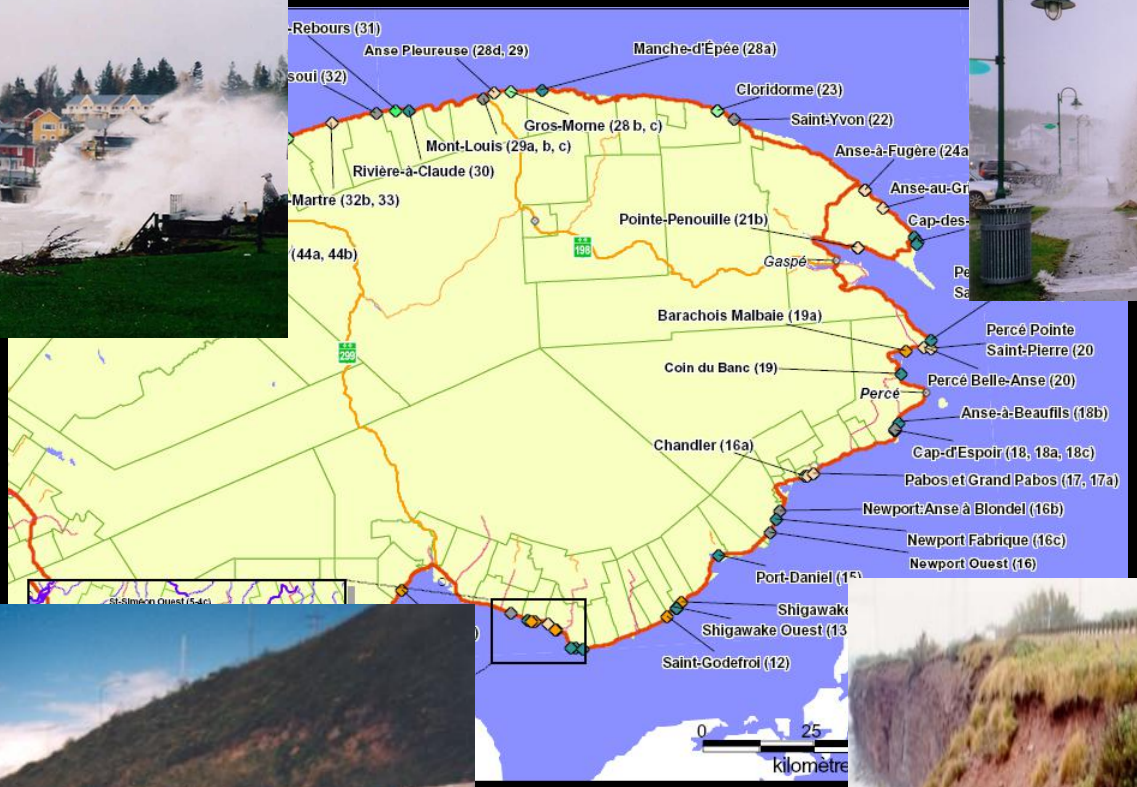


Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

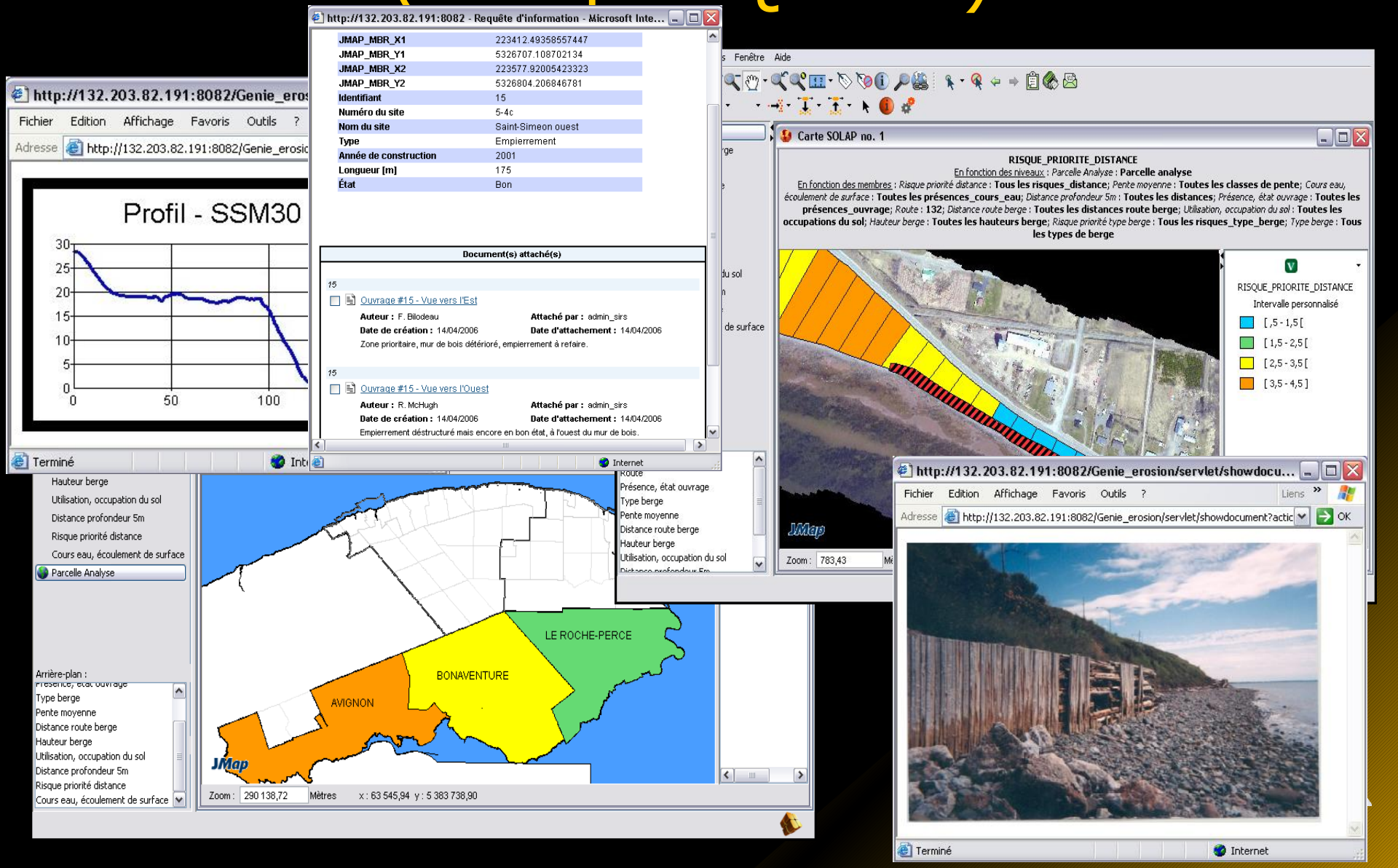
Example: Managing Infrastructures (Port of Montreal)



Example: Coastal erosion management (Transport Quebec)

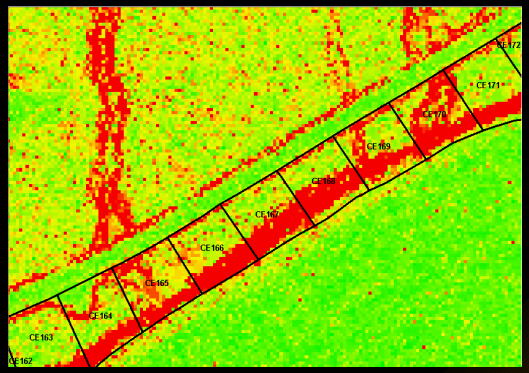


Example: Coastal Erosion Management (Transport Quebec)



Example: Coastal Erosion Management (Transport Quebec)

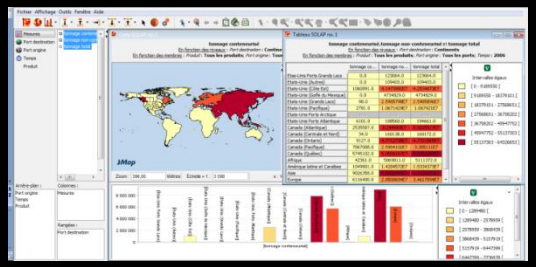
- Criteria to assess the risk of erosion and landslide
  1. Distance between road and bank
  2. Type of bank
  3. Height of the bank
  4. Average slope of parcels
  5. Presence of watercourses, surface water spilling
  6. Presence and quality of protection infrastructures
  7. Distance between the bank and 5m waterline
  8. Land use and occupation
- Divided the coastal zone into parcels
- Each parcel has values for each criteria



Analyse de chaque parcelle pour chaque critère

Example of Measured Benefits in a Project for Transport Quebec

M.J.Proulx, Intelli3 (2009)



Annual Report :

150 maps and tables
Static data

Analysis & page editing (3 months-person)
Updating (1 month-person)

Ad hoc queries continuously
Delays to produce outputs

Depend upon an expert in cartography

Solution géodécisionnelle :

200 000 maps and tables
Dynamic applications

Data structuration (15 days-person)
Updating (5 dayx-person)

Application in intranet
Fast response

Easy user interface

Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

Benefits of SOLAP applications

- In our projects, positive results in many applications have been achieved, such as:
  - cutting by a factor of 10 the time required to produce maps and reports that summarize key information
  - allowing new users having never heard of GIS to produce hundred of thousands of synchronized maps, reports and tables on demand with only three hours of training
  - providing keyboardless access to geospatial data at different levels of detail with a facility never achieved before

SOLAP AND BIG DATA

Big Data

- Big Data characteristics (the Vs)
  - Volume
  - Velocity
  - Variability
  - *And 5 more Vs*
    - *Value, Validity, Veracity, Vulnerability and Visualization*
- These characteristics are happening at an unprecedented pace. Examples include:
  - Mining social networks (ex. Facebook)
  - Monitoring web surfing (ex. Google)
  - Tracing users interactions (ex. Amazon)
  - Exploring smartphone usages (ex. Apple apps)

Business Intelligence and SOLAP

- BI transforms large volumes of structured raw data into meaningful information for more effective decision-making
- BI provide historical, current, and predictive views
- Over the last 20 years, BI has developed a strong data analytics culture, powerful data visualization solutions and proven methods to integrate with organizations' structured database ecosystems
- *Business Analytics* (BA) has been used recently to highlight analytical capabilities
- OLAP is widely used for Business Analytics

BI -> Big Data

- SOLAP has its roots in BI
- An important part of the Big Data discourse is similar to the discourse of BI
- However, Big Data is not BI with bigger data
- The main differences are in velocity, variety and the underlying technology to tackle these two characteristics
- Another difference is that Big Data often comes from outside, typically from the cloud.

BI -> Big Data

- While some see Big Data as the new generation of BI, others see it as a different family of products
- The boundary between Big Data and BI is not clear as there exist two groups of technologies:
  - Big Data core technologies vs. Big-Data-enabled technologies
- History repeats itself:
  - Spatially-enabled DBMS vs. GIS
  - BI-enabled DBMS vs. genuine BI technology
  - Database-enabled CAD vs. GIS
  - 3D-enabled GIS vs. real 3D software
  - Spatially-aware Big Data vs Big Earth Data

Big GeoData

- Two categories:
  - Geolocalized Big Data
    - Location simply as one additional, accessory data
    - Sources: mostly points (smartphones GPS position, web surfing IP address position, Amazon's clients addresses, etc.)
  - Spatially-centered Big Data
    - Location, shape, size, orientation, spatial relationships are core data, a « raison d'être »
    - Sources: ITS, sensor networks, high-resolution imagery (drones, satellites) raw data, interpreted imagery polygonal and line data, terrestrial 3D laser scans, LIDAR, etc.

SOLAP and Big GeoData

- Today's SOLAP is Spatially-centered and Big Data-aware
  - More powerful than simple point location analysis
  - Integrates well in geospatial dataflow ecosystems
  - Fast analysis of large Volumes
  - Does Just-in-Time, very high Velocity expected soon
  - Excellent tool to analyse Veracity
  - The move to Variable, unstructured data hasn't been done yet but is possible (ex. text)

Conclusion

- GIS and BI have evolved in silos for many years
- R&D bridging both universes started mid-90s
- Market is reaching maturity
- A scientific community exists as well as products
- R&D will bring stronger bridges with Big GeoData
- We live in complex technological ecosystems where data (geodata) has the potential to deliver new powerful insights

Food for thought

"As the IT infrastructure inevitably changes over time, analysts and vendors (especially new entrants) become uncomfortable with what increasingly strikes them as a 'dated' term, and want to change it for a newer term that they think will differentiate their coverage/products... When people introduce a new term, they inevitably (and deliberately, cynically?) dismiss the old one as 'just technology driven' and 'backward looking', while the new term is 'business oriented' and 'actionable'" (Elliott, 2011).

Thank you !

More info at these web sites:

http://sirs.scg.ulaval.ca/yvanbedard/

Technology transfer =
Map4Decision (www.intelli3.com)