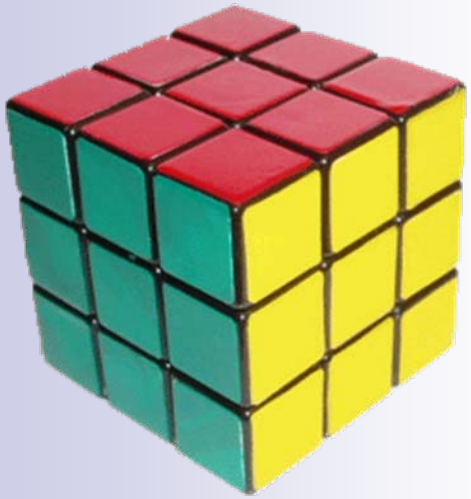
Towards a mobile Spatial OLAP (SOLAP) infrastructure to better support decision in mobility

OGRS 2009 – Nantes, France



Dr. Thierry Badard
Etienne Dubé
Belko Diallo
Jean Mathieu
Mamadou Ouattara

GeoSOA research group
Centre for Research in Geomatics
Laval University, Quebec City, Canada
http://geosoa.scg.ulaval.ca

Outline

- What is Spatial OLAP (SOLAP)?
- An open source software stack for Geospatial BI
  - GeoKettle
  - GeoMondrian
  - Spatialytics
- Towards a mobile SOLAP infrastructure
  - Need for mobile SOLAP?
  - Current state of the infrastructure
- Ongoing works and research challenges

Introduction to BI and OLAP

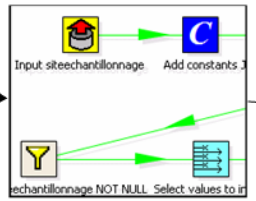
- Business Intelligence applications are usually used to better understand historical, current and future aspects of business operations.
- The applications typically offer ways to mine database- and spreadsheet-centric data, and produce graphical, table-based and other types of analytics regarding business operations.
- Rely on an architecture with complex components and applications:
  - Data warehousing
  - On-line Analytical Processing (OLAP) servers and clients
  - Reporting tools
  - Dashboards
  - Data mining

Classical BI infrastructure …

Data sources
(OLTP systems)

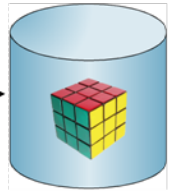
- Transactional databases
- Web ressources
- XML, flat files, proprietary file formats (Excel spreadsheets, …)
- LDAP
- …

Data extraction



* ETL systems

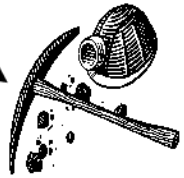
Data loading



Data Warehouse

Reporting tools

OLAP



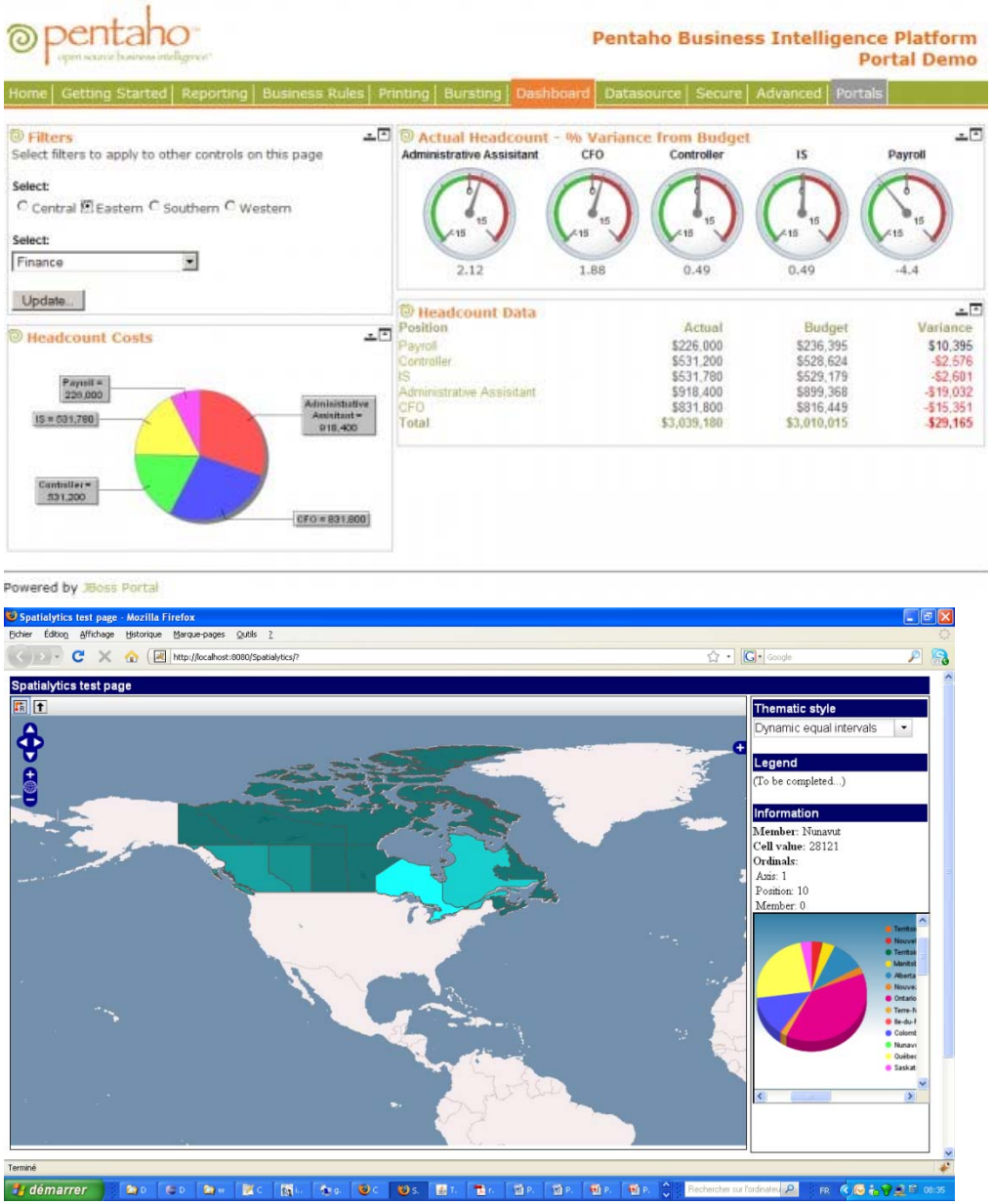
Data mining

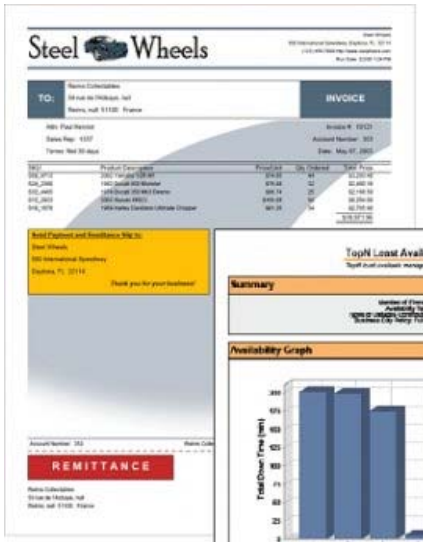
* ETL stands for Extract, Transform (integration, data cleansing, data structure, “updating”, …) and Load

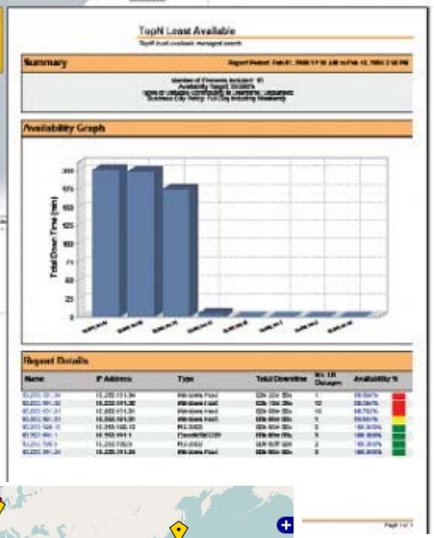
What is Spatial OLAP (SOLAP) ?

- SOLAP extends OLAP (On-Line Analytical Processing)
  - Dedicated to fast and complex analysis of large amount of data vs. transactional systems that are devoted to the rapid storage and updating of data without redundancy (transactions)
  - Analytical vs. operational needs
- SOLAP is more a concept than a precise software product!
- SOLAP adds the consistent handling of geospatial features, map displays and spatial analysis capabilities
- Why ? “About eighty percent of all data stored in corporate databases has a spatial component” – Franklin, 1992
- SOLAP “allows a **rapid and easy** navigation within **spatial datawarehouses** and offers many **levels** of information granularity, many **themes**, many **epochs** and many display modes of information that are synchronized or not: **maps**, tables and diagrams” – adapted from Rivest *et al.*, 2005

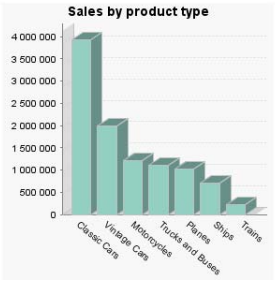
Different ways to present analytical information

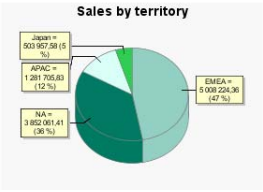






Product Line and Territory analysis

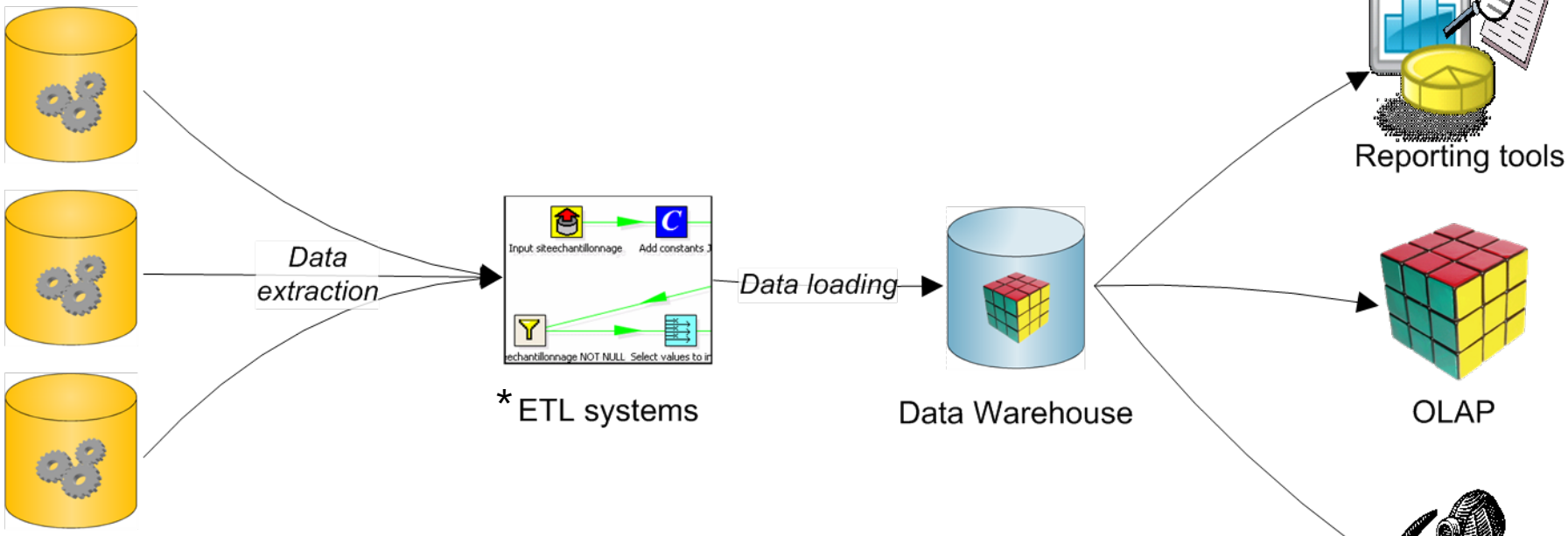




OK, but …

- How to efficiently implement SOLAP concepts?
- What kind of tools or software infrastructure does it require?
- …
- Many R&D questions the GeoSOA research group at Laval University has started to investigate and to implement …

From a classical BI infrastructure …



* ETL systems

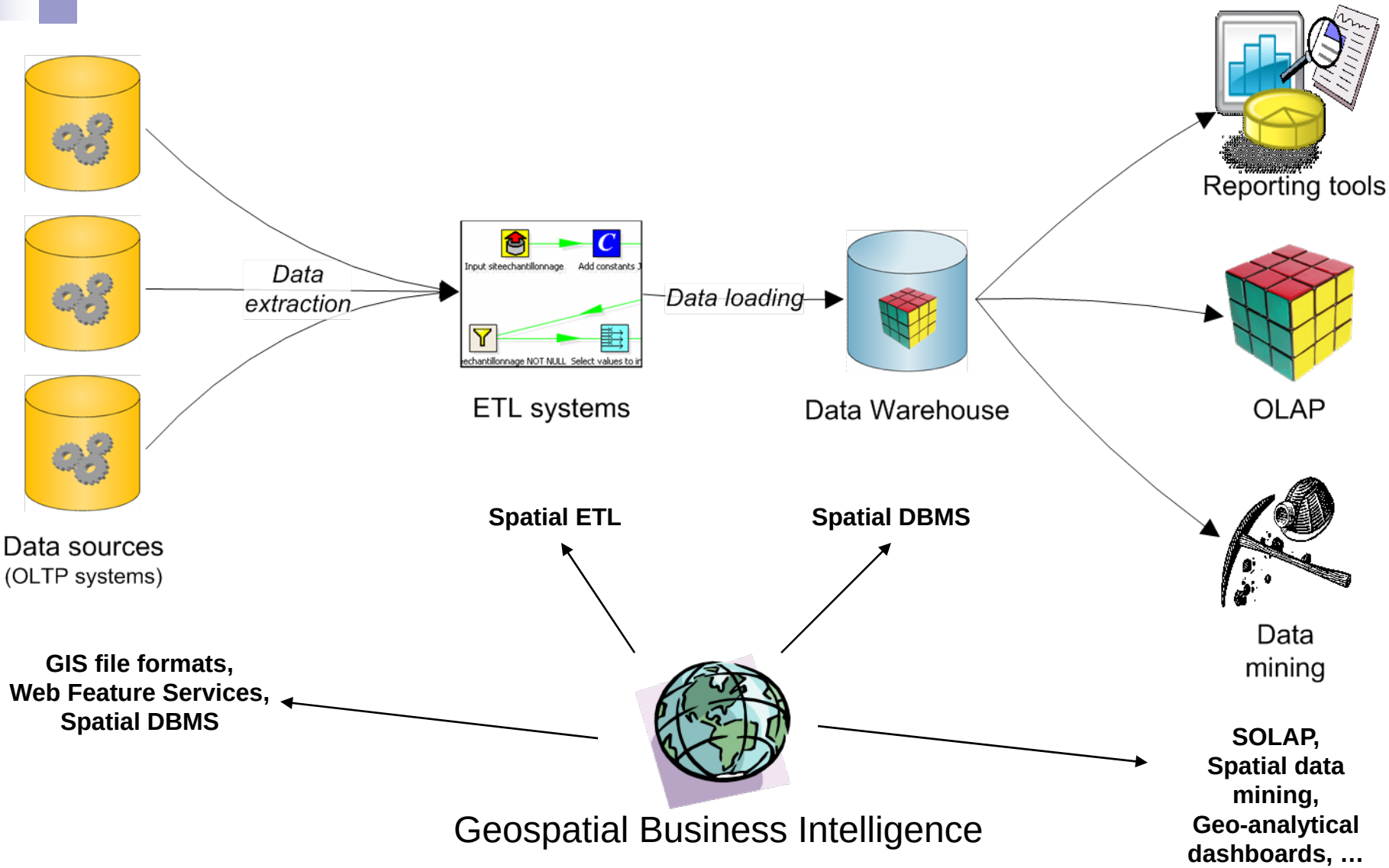
Data Warehouse

Data sources
(OLTP systems)

- Transactional databases
- Web ressources
- XML, flat files, proprietary file formats (Excel spreadsheets, …)
- LDAP
- …

* ETL stands for Extract, Transform (integration, data cleansing, data structure, “updating”, …) and Load

… to a Spatial BI infrastructure!



Require to consistently integrate the geospatial component in all parts of the architecture!

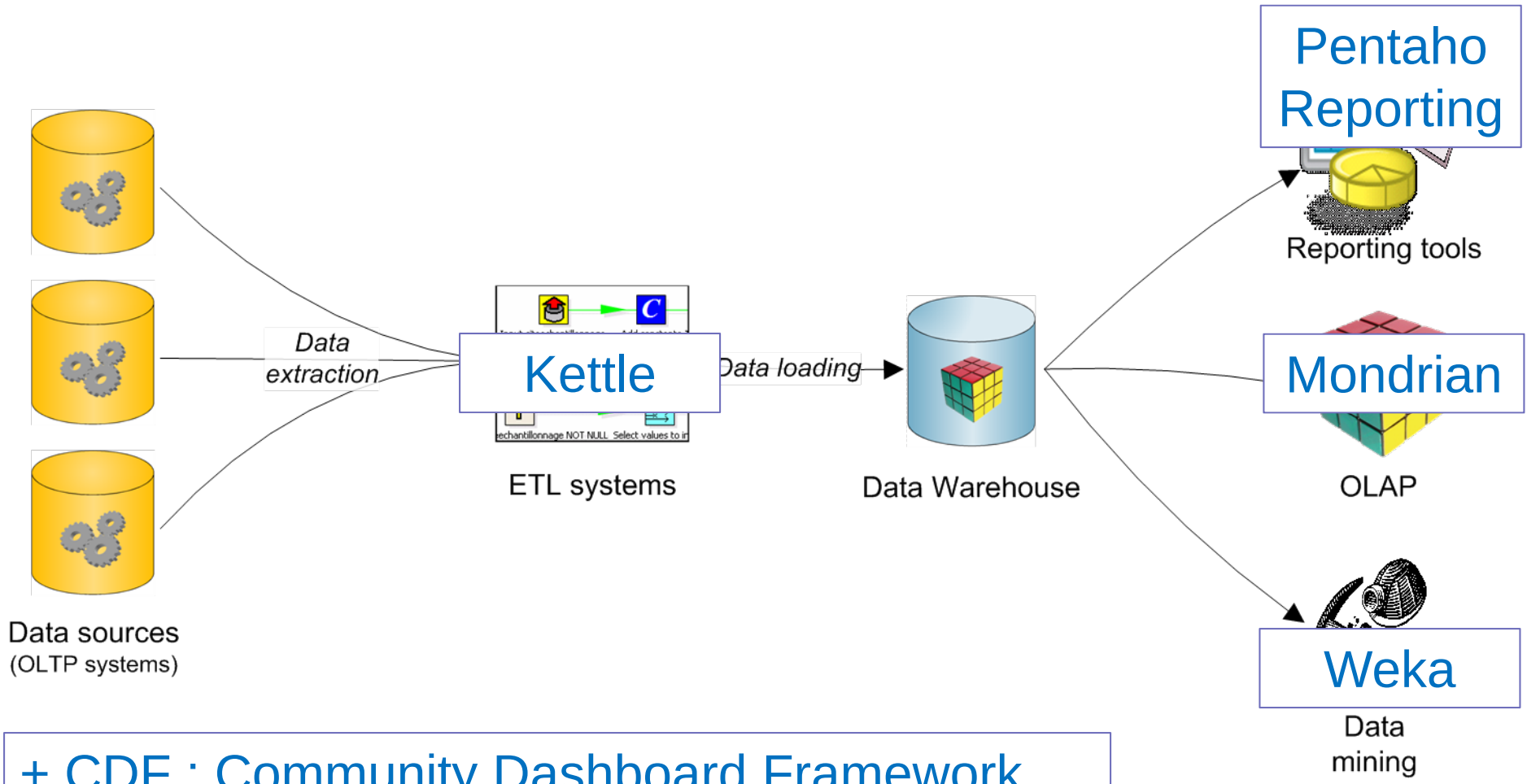
An open source software stack for Geospatial BI

To avoid reinventing the wheel …

- We have used components from the Pentaho open source BI software stack.
- http://www.pentaho.org

Pentaho open source BI software stack

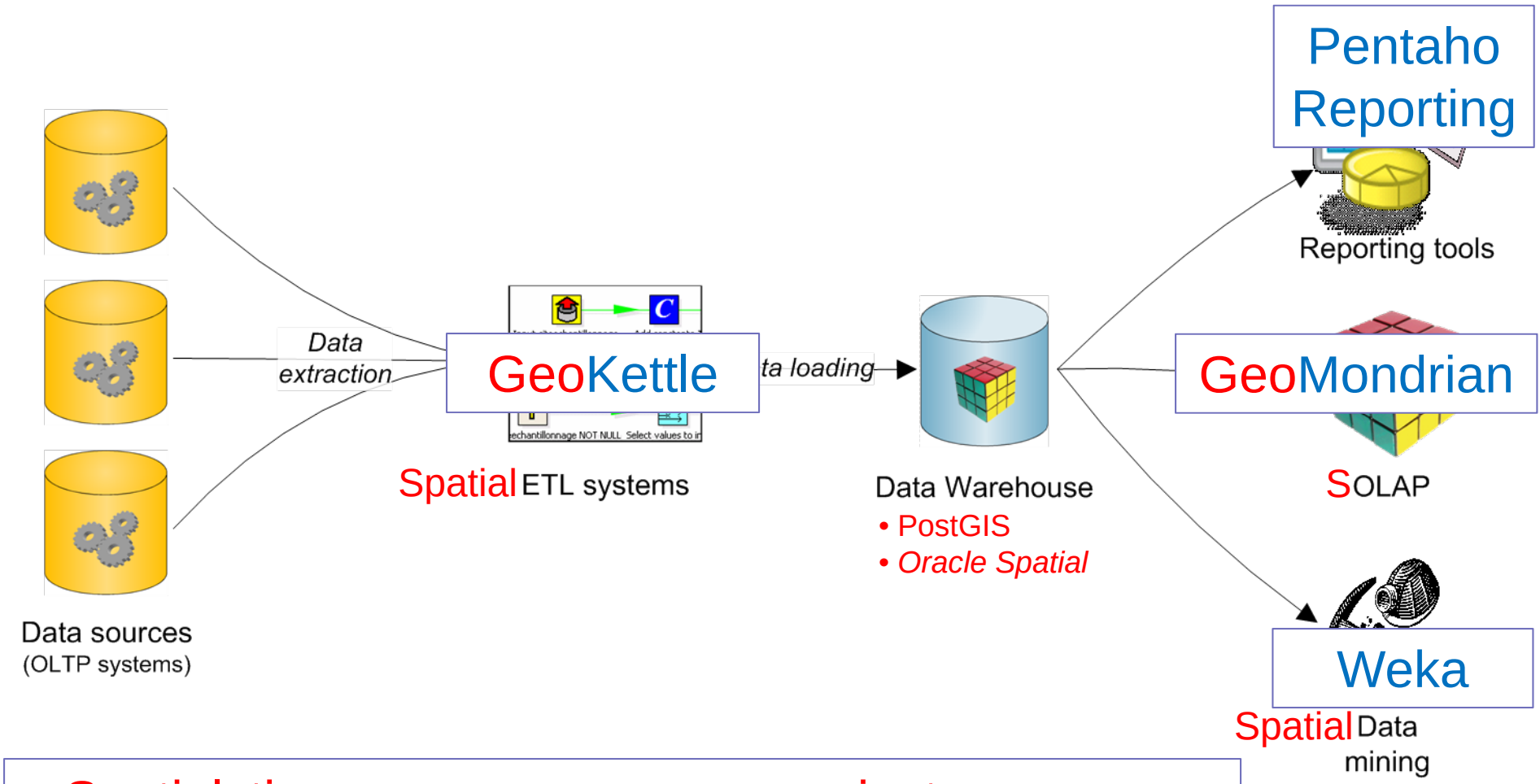
- Pentaho (http://www.pentaho.org)



+ CDF : Community Dashboard Framework
+ Other projects : olap4j, JPivot, Halogen, …

Open source geospatial BI software stack

- GeoSOA group (http://geosoa.scg.ulaval.ca)



+ Spatialytics : new open source project
and ongoing experiments with CDF & JasperServer

Open source geospatial BI software stack

- GeoKettle:
  - Spatial ETL tool
  - http://www.geokettle.org
- GeoMondrian:
  - First implementation of a real SOLAP Server
  - http://www.geo-mondrian.org
- Spatialytics:
  - A drillable cartographic client which enables the navigation in SOLAP datacubes
  - Aims to be integrated in geo-analytical dashboards
  - http://www.spatialytics.org

MDX query language

- MDX stands for MultiDimensional eXpressions
- Multidimensional query language
- *De facto* standard from Microsoft for *SQL Server OLAP Services* (now *Analysis Services*)
- Also implemented by other OLAP servers (Essbase, Mondrian) and clients (Proclarity, Excel PivotTables, Cognos, JPivot, …)
- MDX is for OLAP data cubes what SQL is for relational databases
- Looks like a SQL query but relies on a different model (close to the one used in spreadsheets)
-

```
SELECT
  { [Measures].[Store Sales] } ON COLUMNS,
  { [Date].[2002], [Date].[2003] } ON ROWS
FROM Sales
WHERE ( [Store].[USA].[CA] )
```

} query axes
} cube name
} filter axis ("slicer")

SOLAP requires spatially enabled MDX …

- … to complete the consistent integration of the spatial component.
- This enables spatial analysis in multidimensional queries (i.e. perform geo-analytical queries).
- We have thus brought to Mondrian and MDX what SQL spatial extensions do for relational DBMS (i.e. *Simple Features for SQL and implementations such as PostGIS).*
- Example query: filter spatial dimension members based on distance from a feature

```
SELECT
  {[Measures].[Population]} on columns,
  Filter(
    {[Unite geographique].[Region economique].members},
      ST_Distance([Unite geographique].CurrentMember.Properties("geom"),
      [Unite geographique].[Province].[Ontario].Properties("geom")) < 2.0
  ) on rows
FROM [Recensements]
WHERE [Temps].[Rencensement 2001 (2001-2003)].[2001]
```

- This work is ongoing …

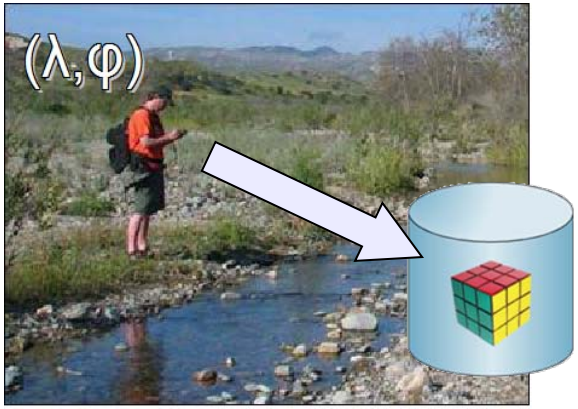
Towards a mobile SOLAP infrastructure

Need for mobile SOLAP …

- Development of wireless, radio and sensors networks
  - With an increasing bandwidth and data volume
- Widespread adoption of web services architectures
- Development of Location Based Services (LBS)
- Democratisation of mobile devices such as PDA and smartphones
- Need for informed decisions to be taken:
  - Anywhere
  - Anyplace
  - Anytime
  - And based on information “synchronised” with the ground reality
- Have prompted the need for research and development in the mobile geospatial BI domain

Use case: Mobile SOLAP

- Mobile geospatial BI:



1. Enhancing data warehouses with terrain data: **utilization of location information from users**

Mobile SOLAP

2. On the field decision-making: **delivering geo-decisional data to mobile users**

Constraints in mobile computing

1 – User interface and Input methods: reduced screen size, pen input

2 – Network links: wireless, limited bandwidth, sporadic coverage

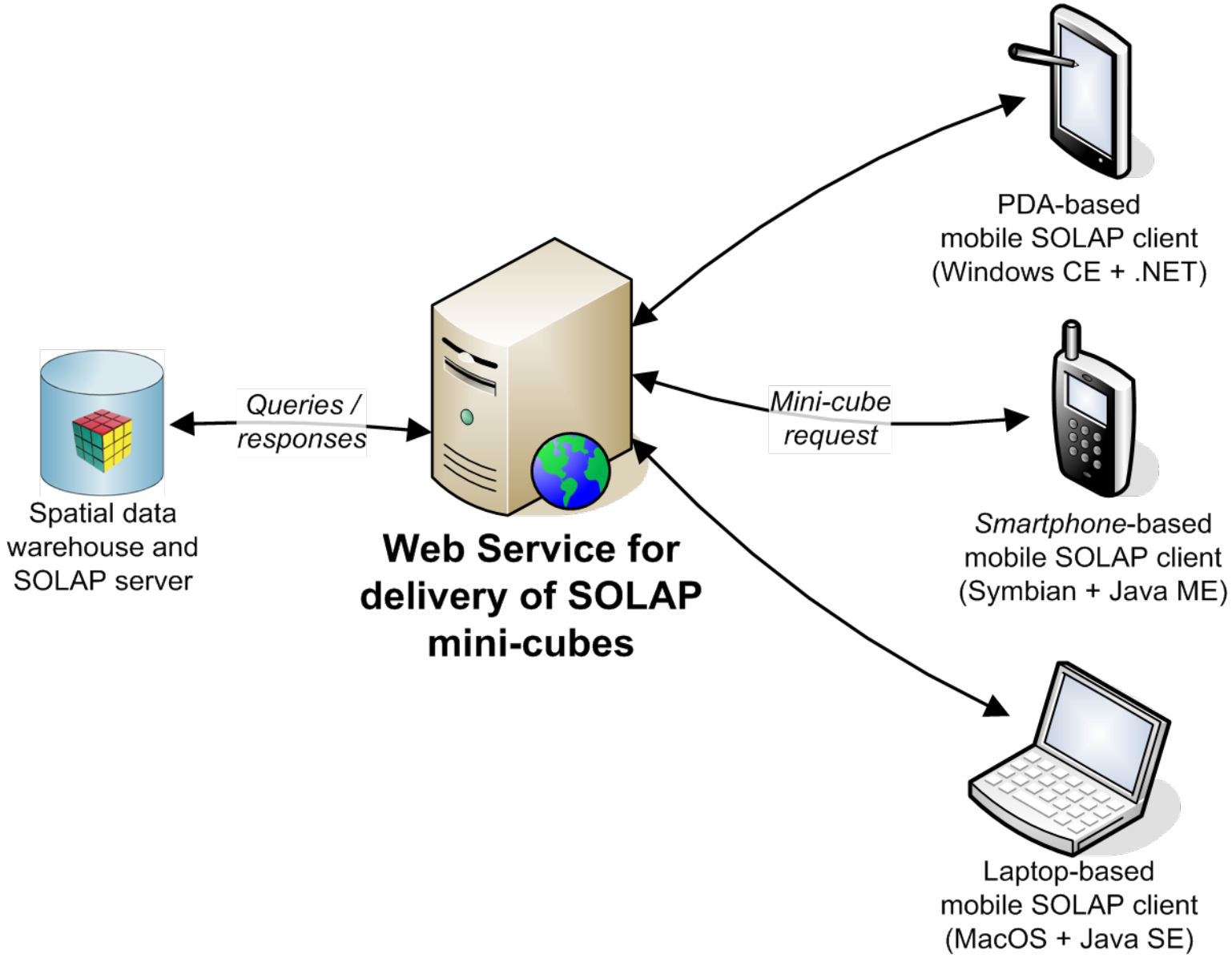
3 – Interoperability: heterogeneous hardware and software

4 – Data size and structure: limited storage and computing capacity

SOLAP in SOA environments

- Current implementations of SOLAP are client/server solutions.
- Binary, closed protocols: no interoperability between different platforms.
- We need a way to encode SOLAP data objects in XML.
  - Previous proposals (XMLA, XCube) do not fully address the requirements for spatial cubes.
  - This prompted the development of a new encoding.
- This enables the development of Web Services for SOLAP, leveraging this XML encoding for the exchange of cube data. Example applications:
  - Web-based geo-decisional dashboards using various on-line data sources.
  - Delivery of SOLAP “mini-cubes” to mobile devices.
  - …

Use case: Mobile SOLAP



Web service for delivery of SOLAP mini-cubes

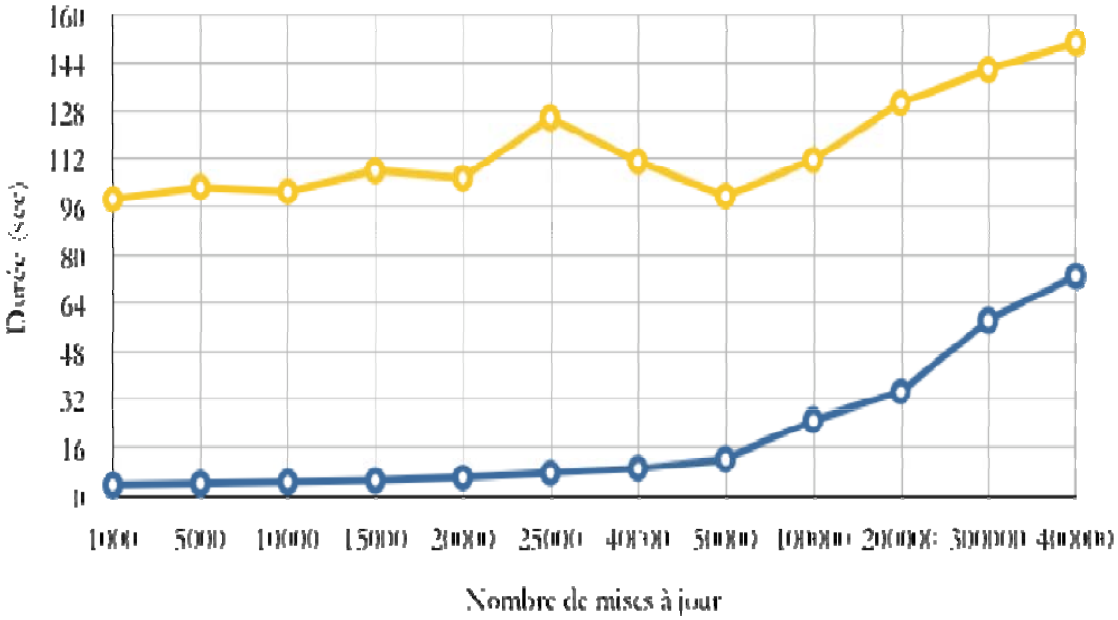
- A kind of WFS but for SOLAP data cubes
- Relies on the definition of a new XML/GML encoding for the interoperable exchange of geospatial datacubes
- Provides different OGC-like operations
  - GetCapabilities, DescribeSchema, GetMembers & GetCells
- A working implementation exists
  - Developed in Java and based on GeoMondrian for SOLAP data retrieval, PostgreSQL+PostGIS for the spatial data warehouse and Apache Axis2 for the Web Service framework.
- The WXS (W3C XML Schema) definition for the encoding is available at:
  - http://geosoa.scg.ulaval.ca/GeoCubeML .
- Specification expected to be extended and evolve, along with the development of new Web Services.

Real-time updating of spatial data warehouses

- To take informed decision, up-to-date information is required
- Updating of data warehouse usually relies on:
  - Building anew the entire data warehouse from sources
    - But could be excessively time consuming (e.g. may be done each night)
    - Does not enable real-time updating
  - Incremental methods
    - But none are dedicated to spatial data warehouses
- We have:
  - Adapted some incremental methods used in classical data warehouses to handle the spatial component
    - Consistent incremental updating of the spatial component in the data warehouse
  - Exposed these updating capabilities as a web service

Real-time updating of spatial data warehouses

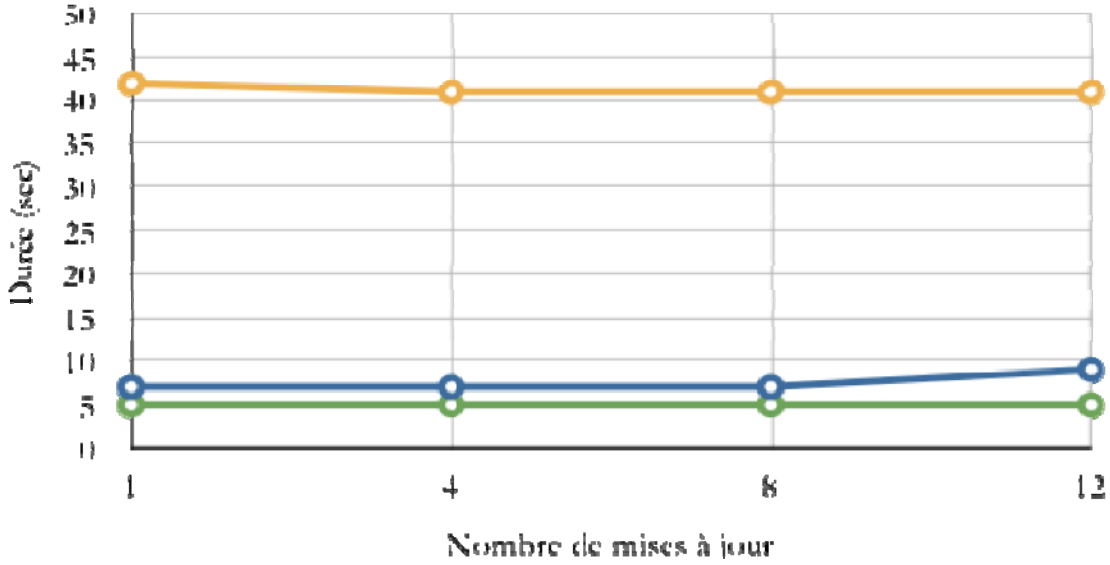
- First results are very good …



Non spatial data warehouse with 1 000 000 records in the fact table

Real-time updating of spatial data warehouses

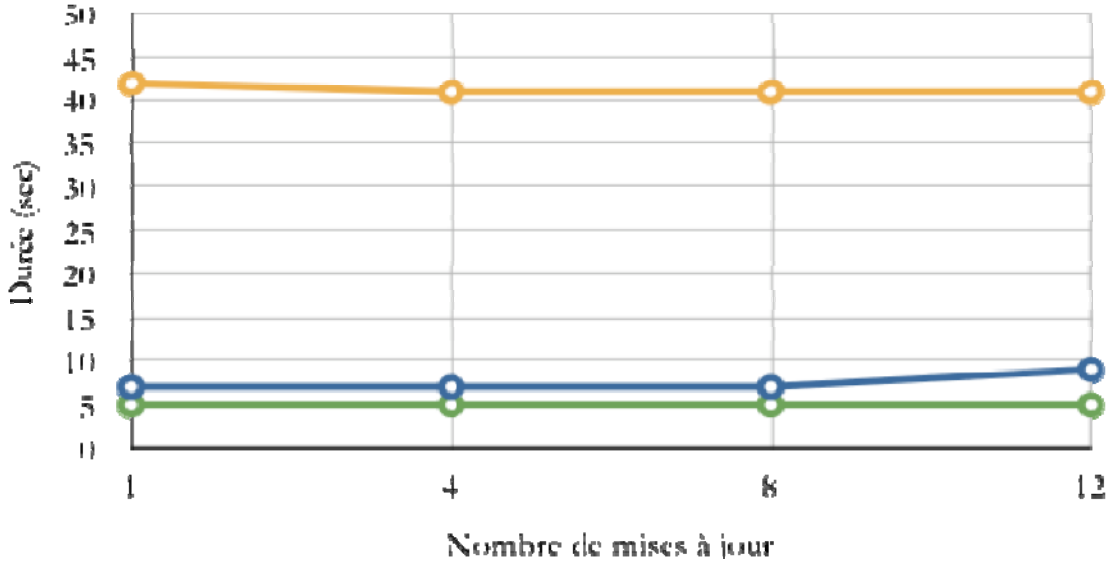
- First results are very good …



Spatial data warehouse with 50 000 records in the fact table

Real-time updating of spatial data warehouses

- First results are very good …



Spatial data warehouse with 50 000 records in the fact table

- It opens the door to the real-time updating of large spatial data warehouses ...

Ongoing works and research challenges

Ongoing works & research challenges

- They mainly deal with:
  - Real-time integration of data streams/sensor data
  - Spatial data mining algorithms and techniques
  - Definition of mobile SOLAP apps and clients
  - Information visualisation and intensive computing

- A kind of GeoSOA research agenda!

Real-time integration of streams/sensor data

- More and more data available from different types of sensors
- OGC has standardized a set of services to deal with geo-sensors (Sensor Web Enablement)
- Need for real-time integration of data flows stemming from distributed geo-sensors into data warehouses
- To support better decision making processes by including data captured on the fly such as the location of users or of tracked mobile objects
- Require to investigate:
  - Query languages on data streams
  - Add spatial analysis capabilities to these languages
  - Manage the integration of large data streams into the data warehouse : need for real-time or active spatial data warehouses?

Spatial data mining

- The definition of innovative data mining algorithms and techniques
- To automatically retrieve trends, patterns or phenomena in large geospatial data volume stored in the data warehouses
- Able to directly and consistently handle the geospatial component without a human intervention
  - Many research works in this domain relies on classical data mining algorithms which requires to pre-process the spatial data in order to fit the data model of the algorithm
  - Such a model does not natively support geospatial data types
- Must deal with complex spatiotemporal data such as multiple trajectories of mobile users or objects

Mobile SOLAP apps and clients

- The design of a mobile client which enables better decision support in a mobile context.
- Such a decision support mobile system should be able to gather data from different sources including the interoperable data cubes delivered by the web services infrastructure.
- It should also encompass the limitations related to mobile computing: disconnected operation when roaming between mobile networks, limited bandwidth and processing capabilities, poor rendering and graphical resolution, different input/output devices (touch screen, stylus, ...).
- It should finally account for the location of the user by issuing queries to the web services infrastructure which pertain to this information in order to deliver information which better fulfill the requirements of the users.
- This requires the definition of new geo-analytical web services to complement the infrastructure.

Information visualisation & intensive computing

- The definition of innovative cartographic representations adapted to Spatial OLAP analyses, especially dedicated to mobile devices
- The definition of a new “SLD-like” standard in order to allow automated portrayal of the results of a geo-analytical query
  - From multidimensional data to cartographic/thematic maps
  - Separation of contents and presentation
- Cloud computing and Grid infrastructure
  - To support very large data volume
  - To support numerous simultaneous users
  - To enable the real-time processing of data
  - Meet the enterprise IT evolutions

Questions?

- Thanks for your attention!

- Contact:

Dr. Thierry Badard
GeoSOA research group
Laval University, Quebec, Canada
Email: Thierry.Badard@scg.ulaval.ca
Web: http://geosoa.scg.ulaval.ca