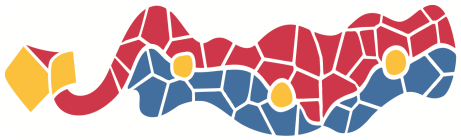
Moving from Oracle/ArcGIS to PostGresql/PostGIS

Yves Choquette

Natural Resources Canada – Earth Sciences Sector
Centre for Topographic Information



FOSS4G 2010
Barcelona

The Organization

- The Centre for Topographic Information (CTI) is part of the Mapping Information Branch of the Earth Sciences Sector, within Natural Resources Canada;
- The Centre's mandate is to create, maintain and disseminate Digital Topographic Data for Canada;
- To achieve its mandate, the CTI works in partnership with the provinces and territories.

The Geospatial Database

- All partners are free to work with the systems and technologies they prefer to produce the territory representation;
- However, everyone must deliver their data as 1:50K National Topographic System (NTS) tiles into the main repository, the Geospatial Database (GDB);
- The GDB is a protected database linked to a transaction services system that allows extraction and delivery of data in a secure way.

Technology Watch

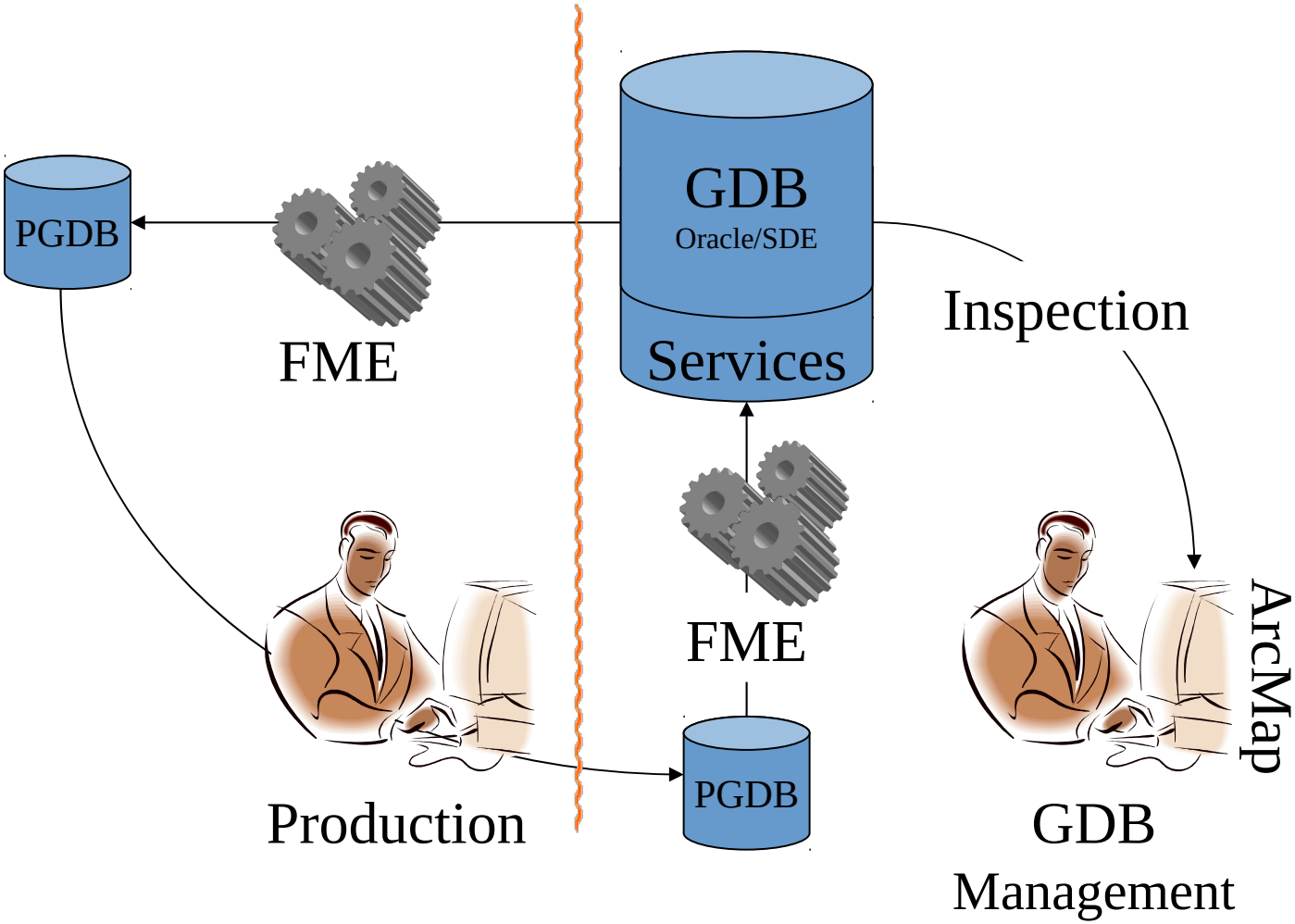
- The GDB management system prototyping project is part of the CTI technology watch activities;
- Every employee has the opportunity to invest 10% of one's time in a project with the aim of studying new technologies that could be used in daily operations;
- It is no secret that "Free and Open Source Software" is a subject full of promises for the Geomatics world.

Project Goals

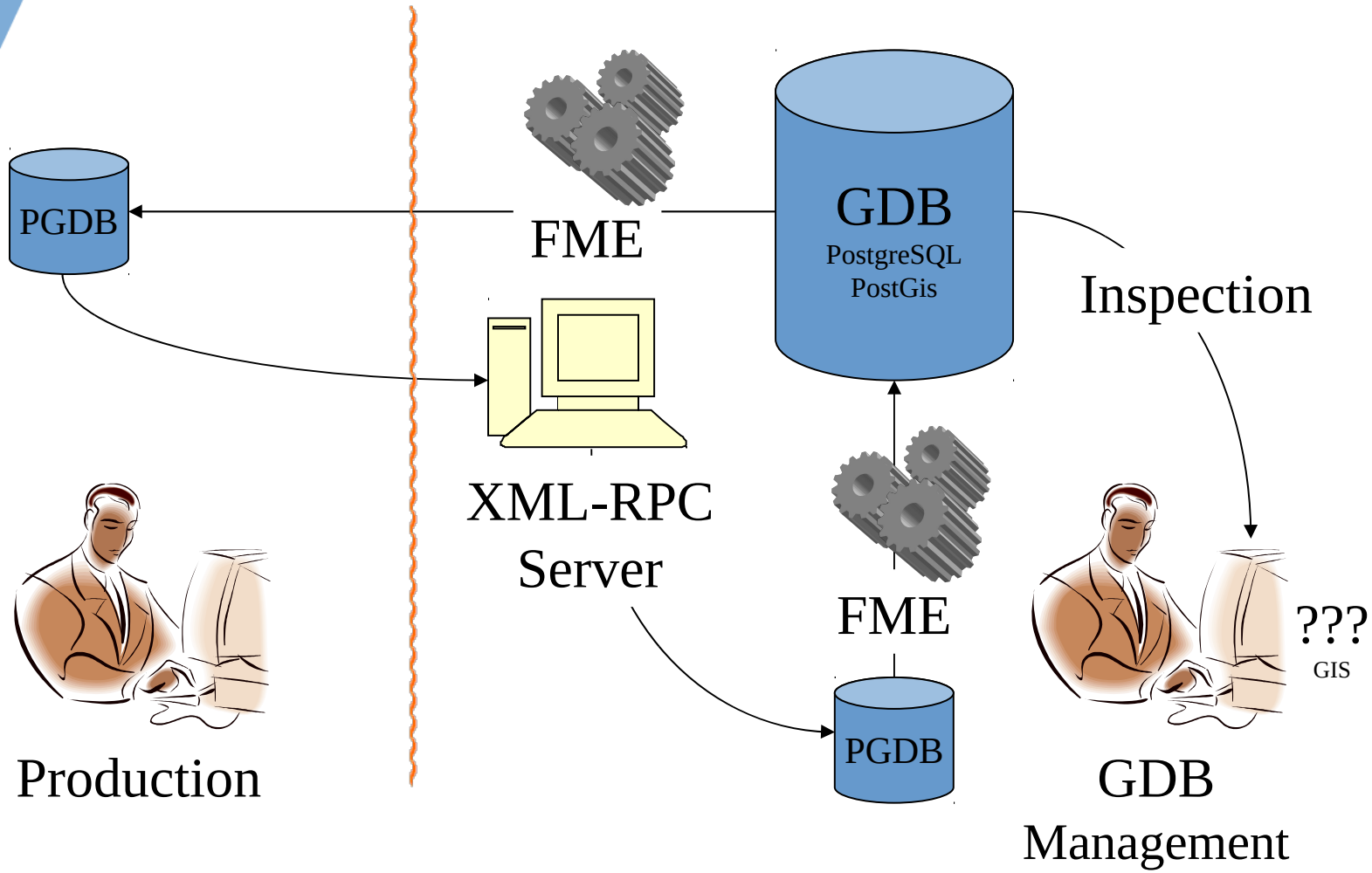
To ...

- Show that the GDB management system can run on PostgreSQL/PostGis;
- Use as much as possible free and open source software to perform the prototype tasks;
- Assess the system performance;
- Explore the editing and visualization tools that can be used on the data stored in PostGIS.

Current System Description



Prototype Description



Proprietary Software Used

- **Feature Manipulation Engine (FME):**
  - Is an integrated collection of Spatial ETL (Extract, Transform and Load) tools for spatial data transformation and data translation
  - Produced by ***Safe Software***
  - Considered to be a GIS utility to help users convert data between formats as well as process data geometry and attributes.

- **Load Sharing Facility (LSF):**
  - Is a commercial computer software job scheduler
  - Sold by ***Platform Computing***
  - Used to execute batch jobs on networked Unix and Windows systems

Open Source Software Used

- **Development Environment:**
  - Eclipse 3.4.1
    - Pydev + Subversive - Pylint
  - Python 2.4
    - psycopg2 + typecheck - lxml
- **Database:**
  - PostgreSQL 8.4 + PostGis 1.3.6
- **GIS Tested:**
  - uDig 1.1.1 and 1.2 RC3
  - Quantum Gis 1.4.0
  - gvSIG 1.9

Implementation Phases

- Software Installation
- Database Configuration
- Data Loading
- Delivery Service Development
- Delivery Test
- GIS Evaluation

Software Installation

- Generally, open source software installation methods are reliable, whatever the platform used.

- In case of problem, communities present on the WEB are helpful and competent.

Database Configuration

The section 5.7 of the PostgreSql documentation on schemas explains how an Oracle Database can be converted to PostgreSQL.

http://www.postgresql.org/docs/8.4/interactive/ddl-schemas.html

Oracle		**PostgreSQL**	
Instance:	BDG_GEST_DEV Server + Port	Cluster: Database:	BDG840D Server + Port BDG_GEST_DEV
Account:	BDG_DBA Schema + Login	Schema: Login role:	BDG_DBA BDG_DBA

Database Configuration

- During the creation of the database:
  - Use the template ***template_postgis*** to integrate the spatial functionalities;
  - Create a **schema** and a **login role** with the same name;
  - Make sure the **search_path** points to the **schema** just created and then, to the public schema:

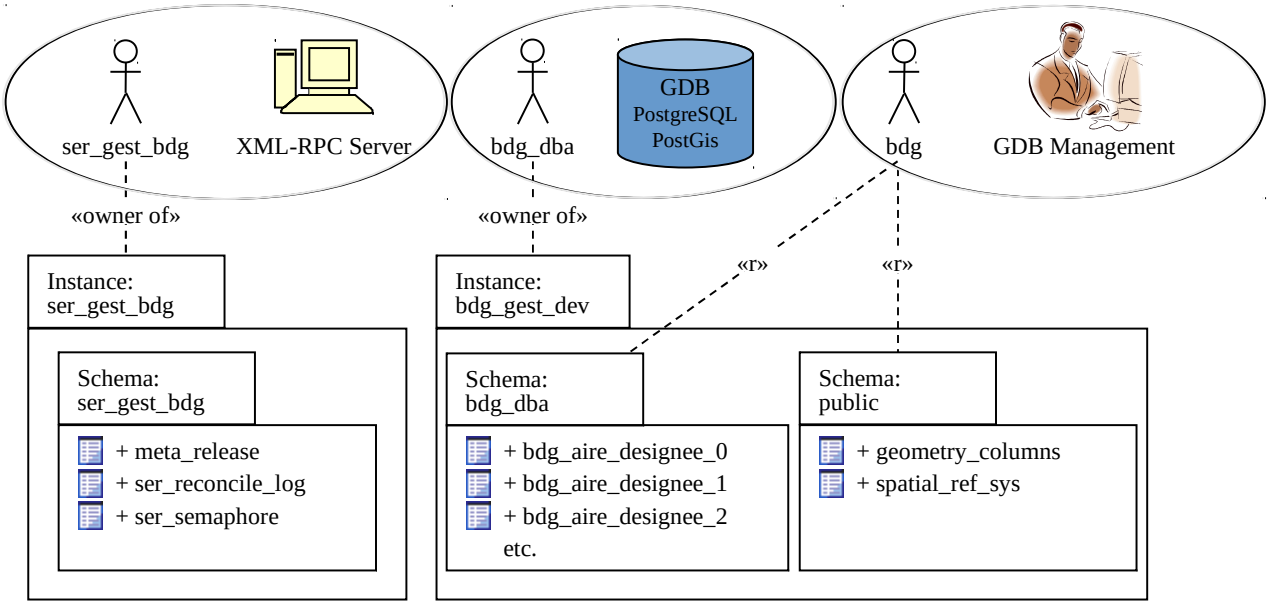
set search_path=bdg_dba,public

Oracle		**PostgreSQL**	
Instance:	BDG_GEST_DEV Server + Port	Cluster: Database:	BDG840D Server + Port BDG_GEST_DEV
Account:	BDG_DBA Schema + Login	Schema: Login role:	BDG_DBA BDG_DBA

Database Configuration

Access account to the GDB Management System:

- The XML-RPC server uses an account *(ser_gest_bdg)* to store the state of the active processes;
- The *bdg_dba* account gives read/write access to the main repository data;
- The *bdg* account gives read-only access to the *bdg_dba* account data via the use of a group role named *bdg_gest_pro_read*



Data Loading

- The data has been loaded per NTS tile, using FME batches.
- To speed up the process, linear and spatial indexes have been created after the loading.
- This type of loading does not involve any spatial processing;
- The loading time is comparable to Oracle/SDE loading time.

Delivery Service Development

- The development took 16 weeks and was made by Stéphane Janvier, a COOP student from Sherbrooke University;
- All PL/SQL services from the old system were rewritten in Python, resulting in a better system security;
- The Eclipse development environment is ideal to manage the system:
  - Subversion repository integration;
  - Ability to trace the code execution on client and server sides using the same session.

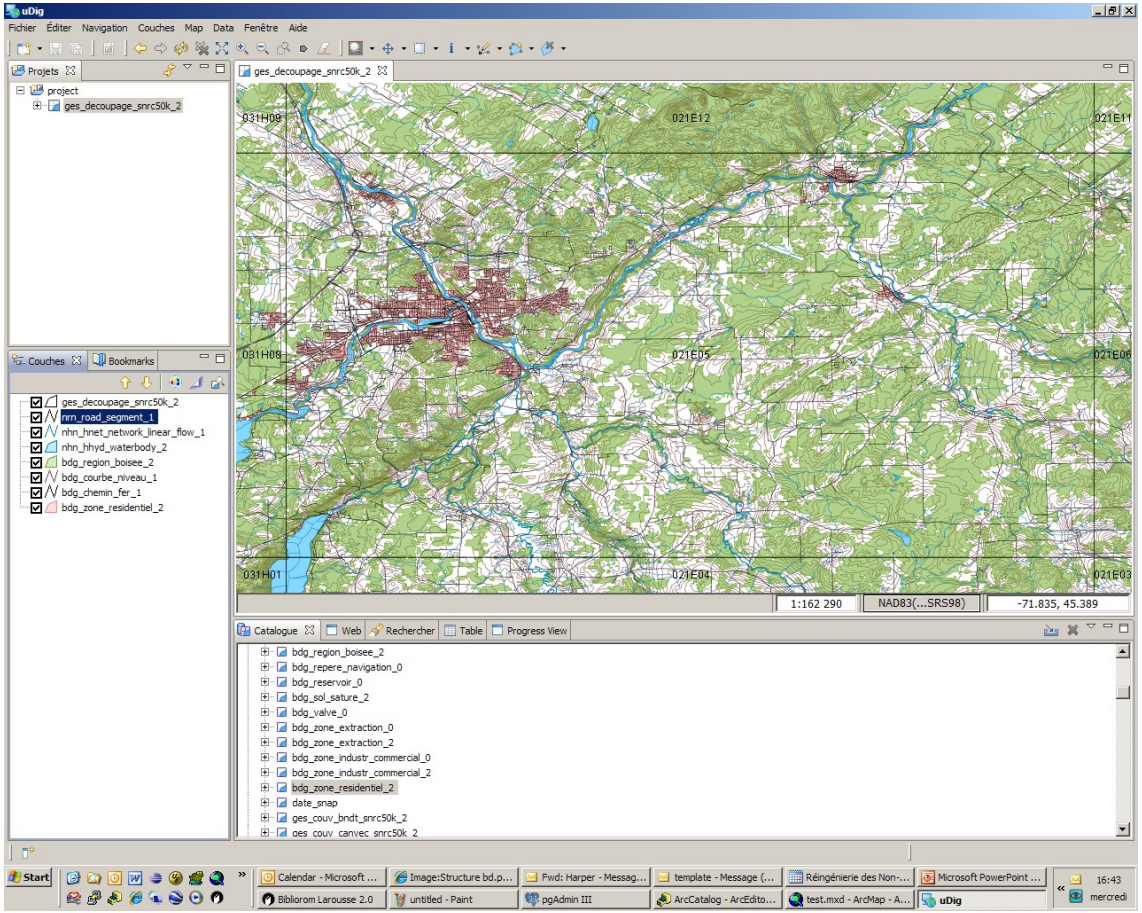
Delivery Tests

- The execution time is comparable to Oracle/SDE execution time on the old system:
  - The loading involves no spatial operations;
  - The processing done resembles the massive loading described previously, thus the similarity in performance;
  - It is difficult to make an accurate comparison between the production server and the prototype server because of the configuration difference.

GIS Evaluation

- Sought-after GIS Characteristics
  - The GIS takes into account the display window coordinates when extracting the information to be displayed;
  - It **uses the spatial index** to get the data from the database;
  - The interface is intuitive.
- Work Method
  - Display of the NTS grid that covers the Canada;
  - Zoom on the interest zone;
  - Activation of the layers to work with.

Simulation


uDig
Fichier Éditer Navigation Couches Map Data Fenêtre Aide
Projets
project
ges_decoupage_snrc50k_2
Couches
Bookmarks
ges_decoupage_snrc50k_2
nrn_road_segment_1
nhn_hnet_network_linear_flow_1
nhn_hhyd_waterbody_2
bdg_region_boisee_2
bdg_courbe_niveau_1
bdg_chemin_fer_1
bdg_zone_residentiel_2
ges_decoupage_snrc50k_2
031H09
021E12
021E11
031H08
021E05
021E06
031H01
021E04
021E03
1:162 290
NAD83(...SRS98)
-71.835, 45.389
Catalogue
Web
Rechercher
Table
Progress View
bdg_region_boisee_2
bdg_repere_navigation_0
bdg_reservoir_0
bdg_sol_sature_2
bdg_valve_0
bdg_zone_extraction_0
bdg_zone_extraction_2
bdg_zone_industr_commercial_0
bdg_zone_industr_commercial_2
bdg_zone_residentiel_2
date_snap
ges_couv_bndt_snrc50k_2
Start
Calendar - Microsoft ...
Image:Structure bd.p...
Fwd: Harper - Messag...
template - Message (...
Réingénierie des Non-...
Microsoft PowerPoint ...
Bibliorom Larousse 2.0
untitled - Paint
pgAdmin III
ArcCatalog - ArcEdito...
test.mxd - ArcMap - A...
uDig
16:43
mercredi

GIS Evaluation

- gvSIG
  - The interface is not intuitive;
  - Although the installation went without problem, first attempts to display the NTS grid were laborious and unsuccessful;
  - A training course is needed to figure out how it works and what the potential is.

GIS Evaluation

- Quantum GIS
  - The interface is more intuitive;
  - Since the display zone is not taken into account, it is preferable to specify a filter (“where” clause);
  - The application is slow: it may take a few minutes to get a result after a button clic.

GIS Evaluation

- uDig 1.1.1
  - This version satisfies the Sought-after GIS characteristics and the performance is nearly as good as with ArcMap
    - Some database tuning could improve the performance
  - The application does not like impatient people: it is preferable to let commands end before sending a new one.
- uDig 1.2 RC3
  - Shows a display problem, both on Windows and on Mac OSX.

About Costs

- Some may say it is not free:
  - Need to pay for training and consultants
  - But, is it really different from proprietary software?
- There are no license fees to pay when Internet is involved:
  - This means a lot of savings for the WEB distribution applications.
- Have you seen “The surprising truth about what motivates us”

http://www.youtube.com/watch?v=u6XAPnuFjJc&feature=player_embedded

Conclusion

- Although interesting, Free GIS does not seem to have reached its full potential:
  - This is probably due to the tasks complexity to be managed
    (multi-thread, spatial functions, graphic display, DB connection, etc.)
- All the GIS tests show stability problems:
  - These may be caused by a bad system configuration, particularly with Java and its set of libraries.

Conclusion

- PostgreSQL is a mature and reliable DBMS
  - Its existence goes back to 1985, when the Ingres software was completely rewritten to bring Postgres;
  - The SQL language was integrated in 1995.
- PostGis integrates well into PostgreSQL
  - uDig tests have shown that the spatial index system works well;
  - Its model is simple in comparison with SDE

    Based on the “Keep It Simple” paradigm (KIS) as opposed to “Keep It Complex” (KIC)

Demo

Questions

?