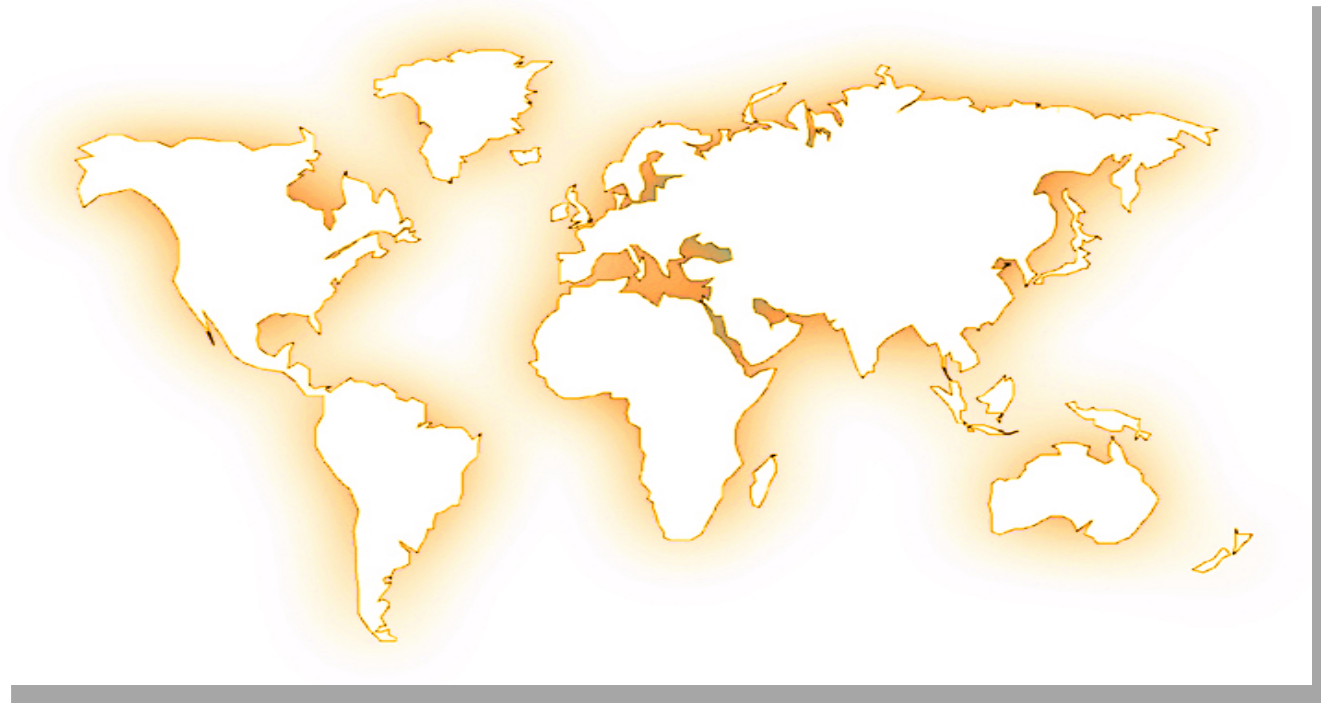




De la solubilité de Postgresql dans les SIG





**MAKINA
CORPUS**

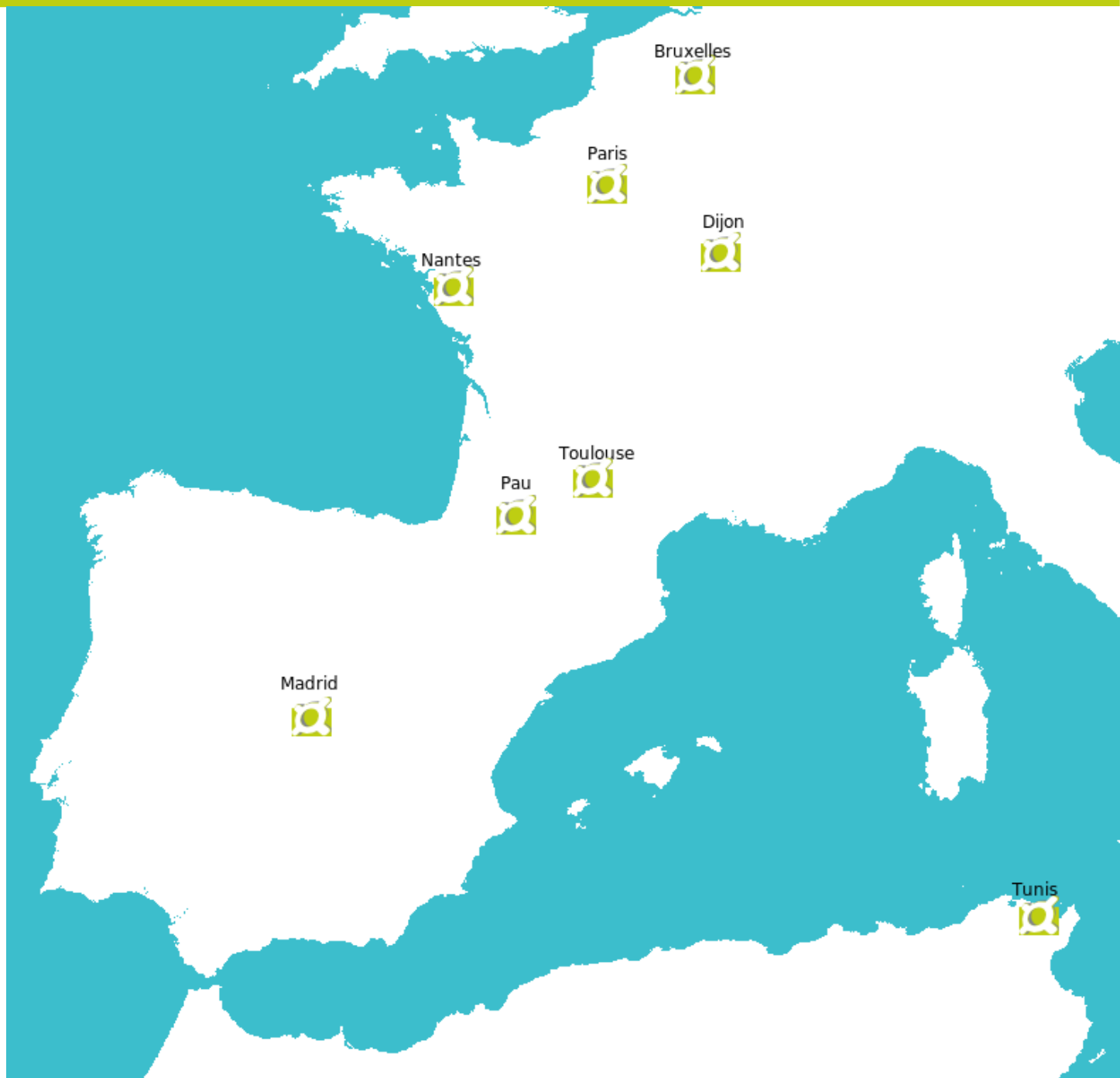
Vincent Picavet / Makina Corpus
vincent.picavet@makina-corpus.com

**Ingénieur Systèmes
d'Information
Géographique**

**SSLL Française
~ 50 personnes**

**Siège à Paris
5 agences en France
Bruxelles, Tunis,
Madrid.**

**SIG
Outils collaboratifs
Gestion de contenu
Système & réseau**



- **2 ans de développement SIG
chez Makina Corpus**
- **Mission actuelle chez le client X,
au service recherche &
développement**



PostgreSQL dans les architectures de systèmes d'information géographique orientés web



- But
 - Présenter de l'information géographique sur le web
- Qui ?
 - Collectivités, entreprises...
- Exemples
 - Géoportail,
 - Consultation du cadastre, PLU
 - Localisation de point d'intérêts

Plan Local d'Urbanisme

révision simplifiée n°1 (à Lyon 8e) du 1er août 2008

GRAND LYON
communauté urbaine

Aide © Communauté urbaine de Lyon 2008

Commune
Lyon 8ème

Localisation
Recherche sur voie/place/lieudit
du n° au n°
→ Localiser
Recherche sur parcelle
section n°
→ Localiser

Plans
Zonage et autres prescriptions 1/5000
Photo aérienne ☐
→ Légende (pdf)
→ Imprimer le plan

Documents
→ Documents de la commune

écrire | aide | faq | plan du site | nous connaître | presse

géoportail le portail des territoires & des citoyens
Liberté • Égalité • Fraternité
République Française

ACCUEIL VOIR RECHERCHER CATALOGUER S'INFORMER GÉOPORTAIL SERVICES

Mode avancé
Mes lieux favoris | Lieux remarquables
Sélectionnez...

2D 3D BETA

Catalogue : 23 couches

- SURFACES BATIES
 - Bâtiments
- SERVICES PUBLICS
- RÉSEAUX DE TRANSPORT
- HYDROGRAPHIE
 - Réseau Hydrographique
- ZONES À RISQUES
- LITTORAL
- ALTITUDE
- SITES PROTÉGÉS
- SITES GÉODÉSQUES
- COMMUNES

Ma sélection : 4 couches

- Parcelles Cadastreales Opacité 100%
- Bâtiments Opacité 100%
- Limites Administratives Opacité 100%
- Photographies Aériennes Opacité 100%

Sélections disponibles
Profil personnalisé

Réseau Géodésique Français 1993 - coordonnées géographiques
Longitude : 01°26'19" E Latitude : 43°35'53" N

© 2007 - 2008 IGN BRGM Mentions légales Crédits

Ville de Rouen
Plan local d'Urbanisme

Photo aérienne +

Rechercher

Localisation | Légende | Liste | Infos | Rue

Quartiers
Centre ville rive gauche
Localiser

Recherche par numéro de rue
~ A B C D E F G H I J K L
M N O P Q R S T U V W X Y Z
--Nom de voirie-- --N°--
Localiser

Recherche de parcelles par sections et numéros
--Section-- --Numéro de parcel--
Localiser

Secteur : UBb
Nom : Maisons discontinues

local d'Urbanisme - juillet 2007 - informations délivrées à titre indicatif.
pour plus d'informations, consulter la Direction de l'Aménagement Urbain et de l'Habitat de la Ville de Rouen

conception Sig@rouen - mars 2006
réalisation Géosignal

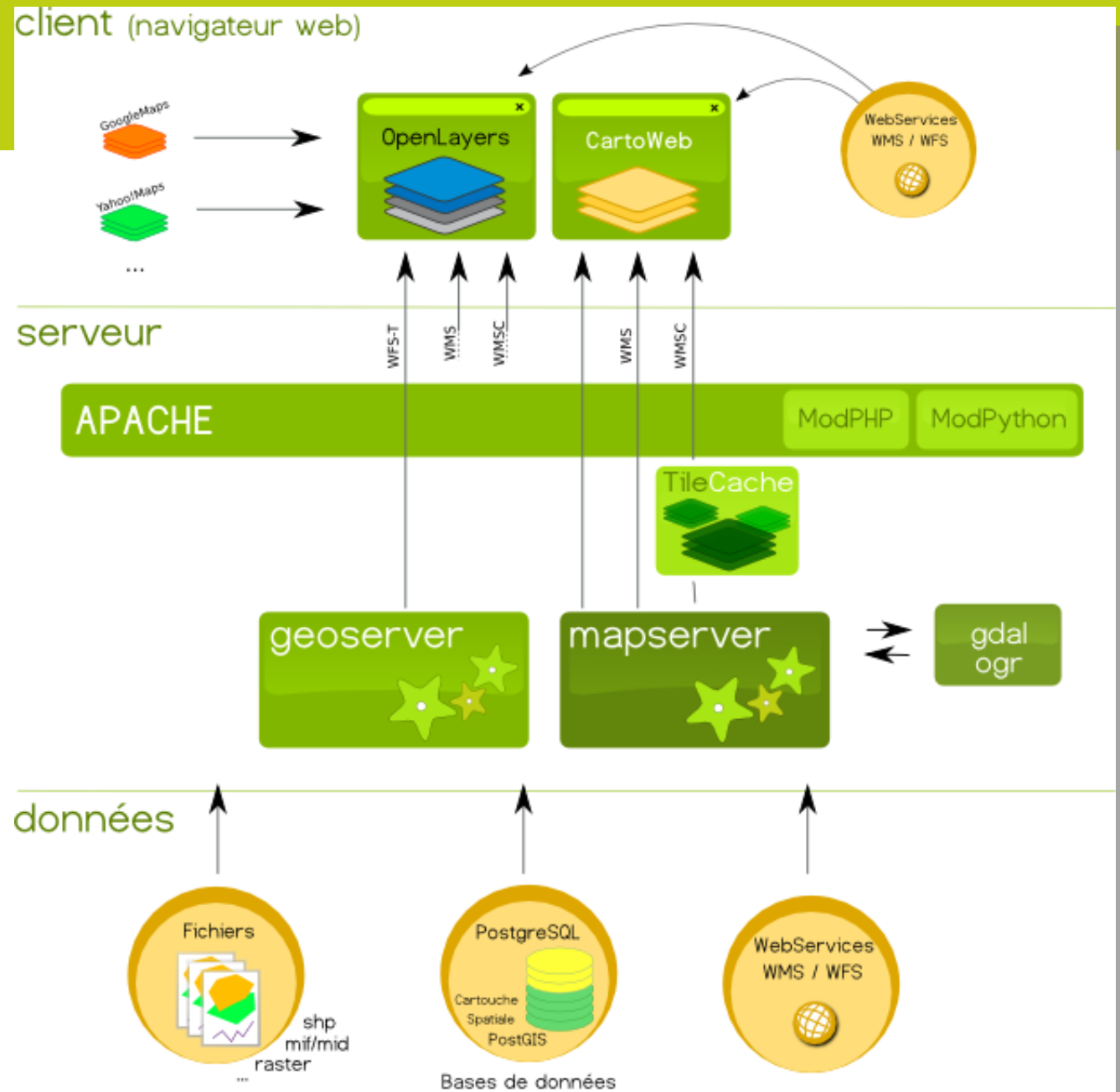


- Architecture
SIG web

client (navigateur web)

serveur

données



WMS: Web Map Service, WMSC: WMS Cache,
WFS: Web Feature Service, WFS-T: WFS-Transactional

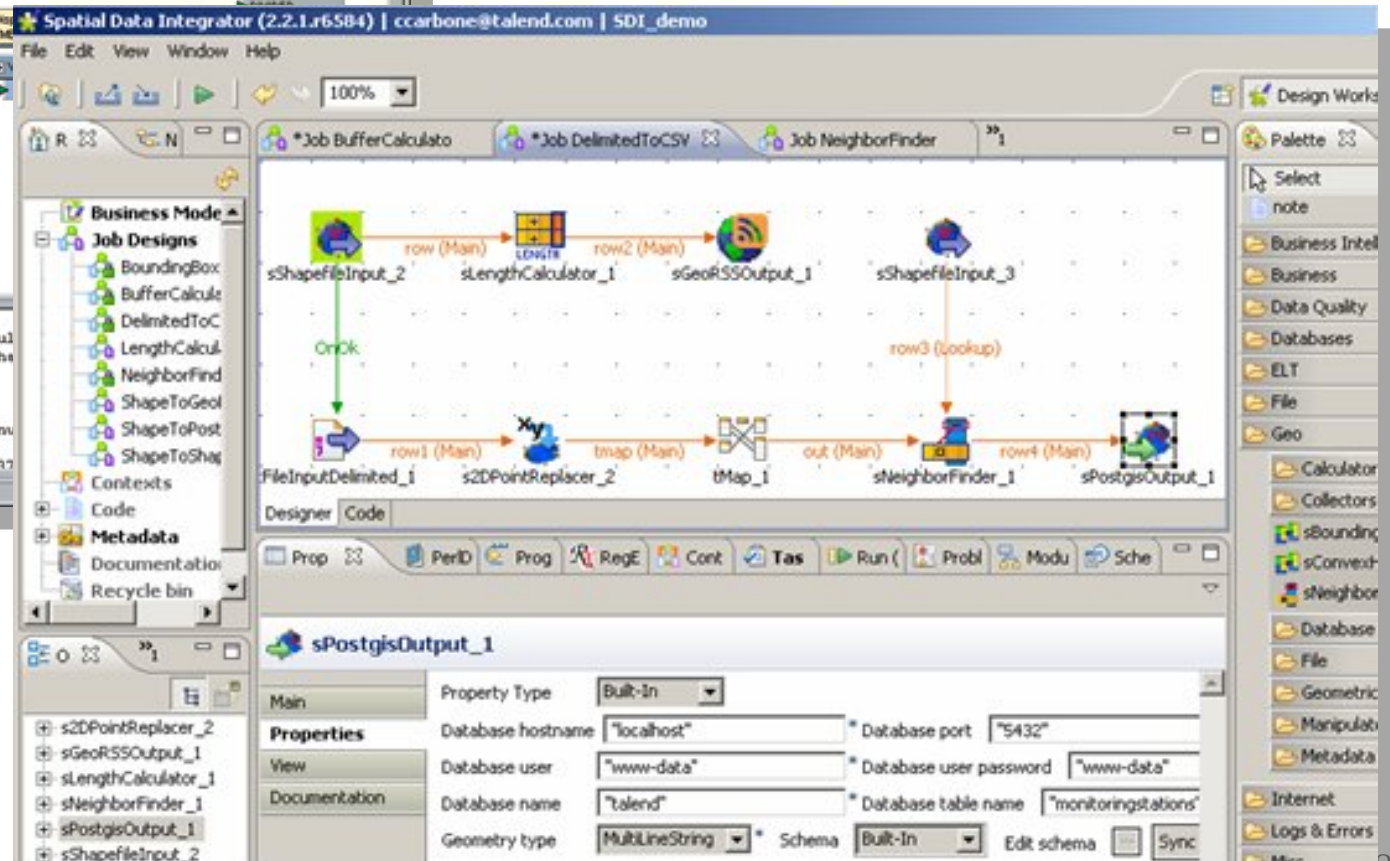
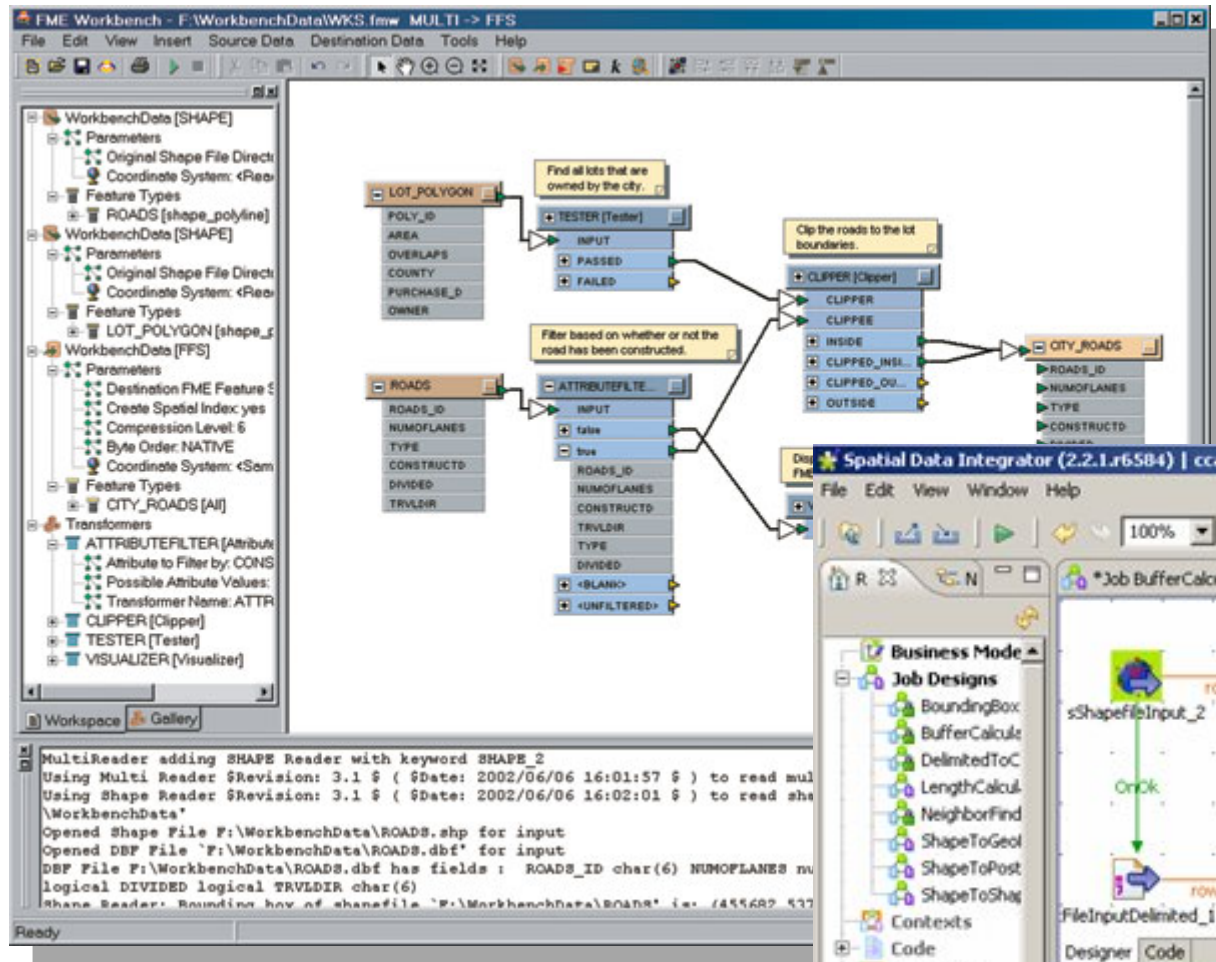
- Types de données particuliers (géométries)
- Gros volumes
- Accès concurrentiels forts, rapidité cruciale
- Principalement de la lecture, avec serveurs cartos
- Gestion des droits au niveau applicatif
- Lecture / écriture avec des clients SIG desktop
- Peu de mise à jour en mode web

- PostgreSQL (8.3 sauf contrainte)
- PostGIS (1.3.3)
 - Types de données géométriques
 - Fonctions de traitement
 - Indexes spatiaux
- Serveurs Linux (Debian) sauf exception
 - Recompilation de PostgreSQL et Postgis, procédures et outils de déploiement maison
- Clients
 - Stations de développement sous Linux
 - Outils clients sous Linux ou Windows

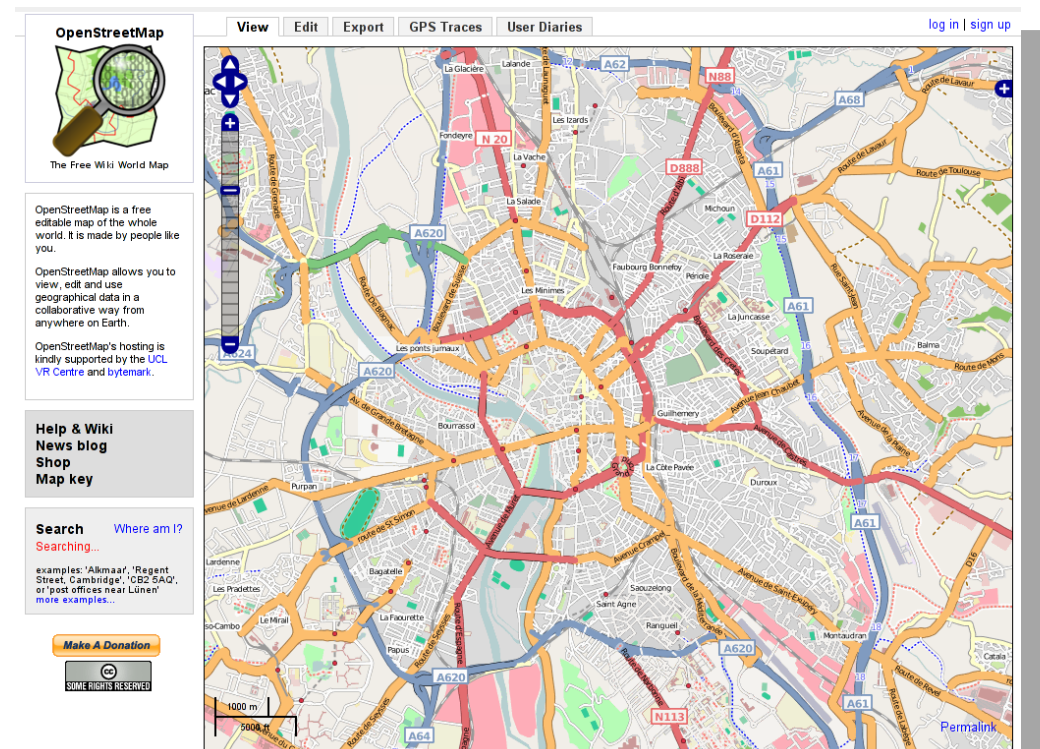
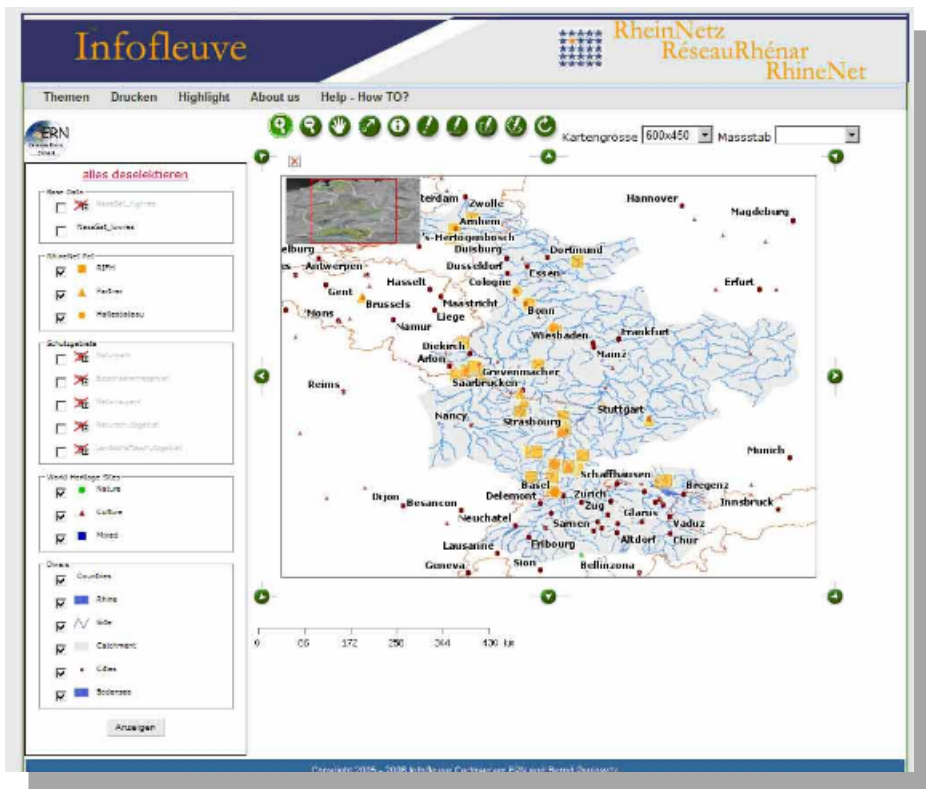
- Phase importante en SIG
 - Données hétéroclites, en type ou formats
 - SIG : Mapinfo, SHP, Oracle Spatial...
 - Néogéographie : GeoJson, KML, GeoRSS...
 - Rasters pas en base
 - Souvent fait une fois pour toute (ou à chaque relivraison)
- Outils
 - Scripts maison
 - Shp2pgsql
 - Ogr2ogr (GDAL/OGR)
 - De plus en plus des ETL : FME, Talend (+SDI)



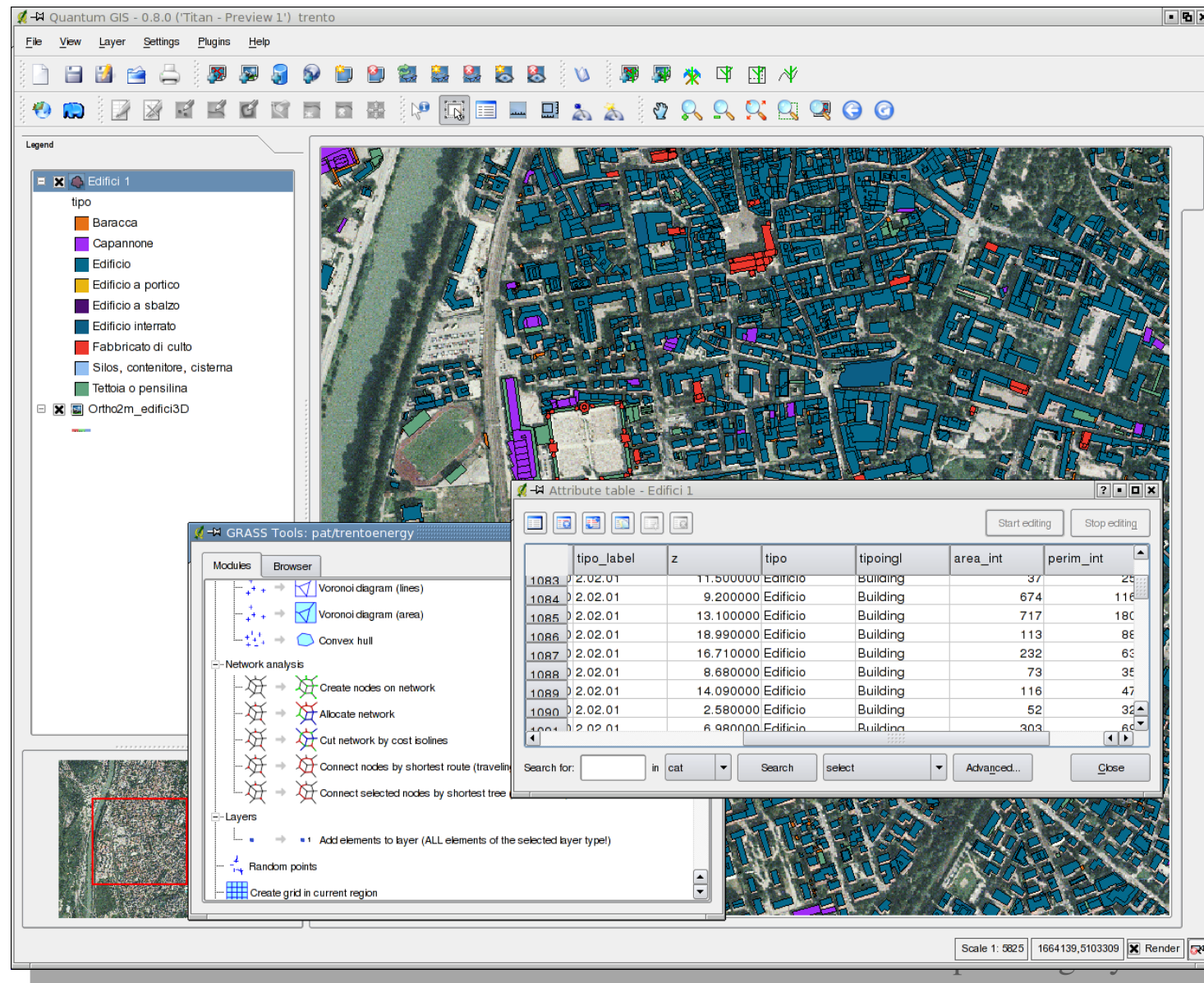
FME & Talend + SDI



- Web
 - Serveurs cartographiques : geoserver, mapserver
 - Serveurs de «features» : FeatureServer
 - Frameworks de développement (geodjango...)
 - Applications de webmapping : Cartoweb, Openlayers (pas d'accès direct).



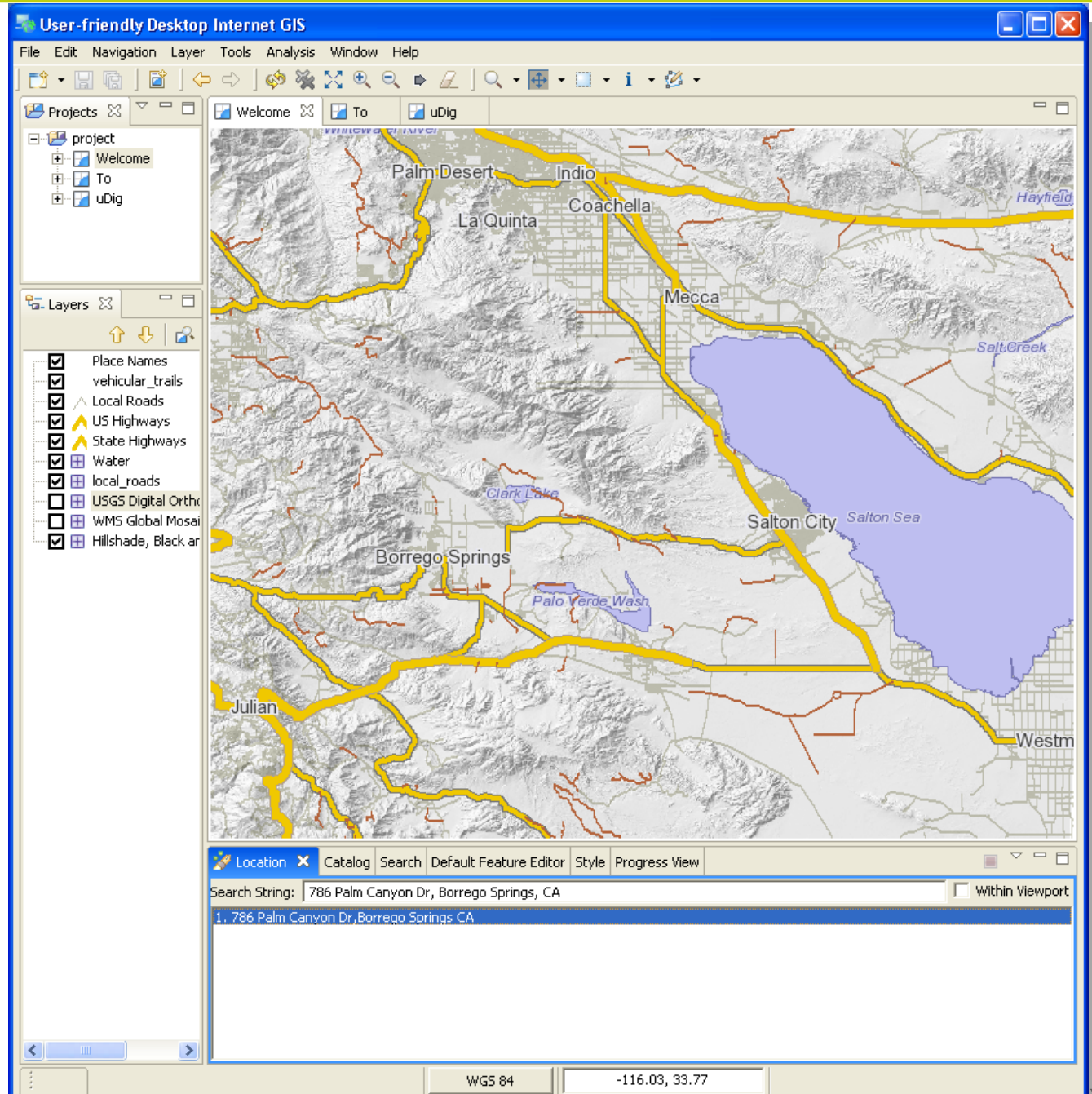
- Clients desktop : SIG «traditionnel»
- QGIS,
- Udig,
- GvSIG
- ...

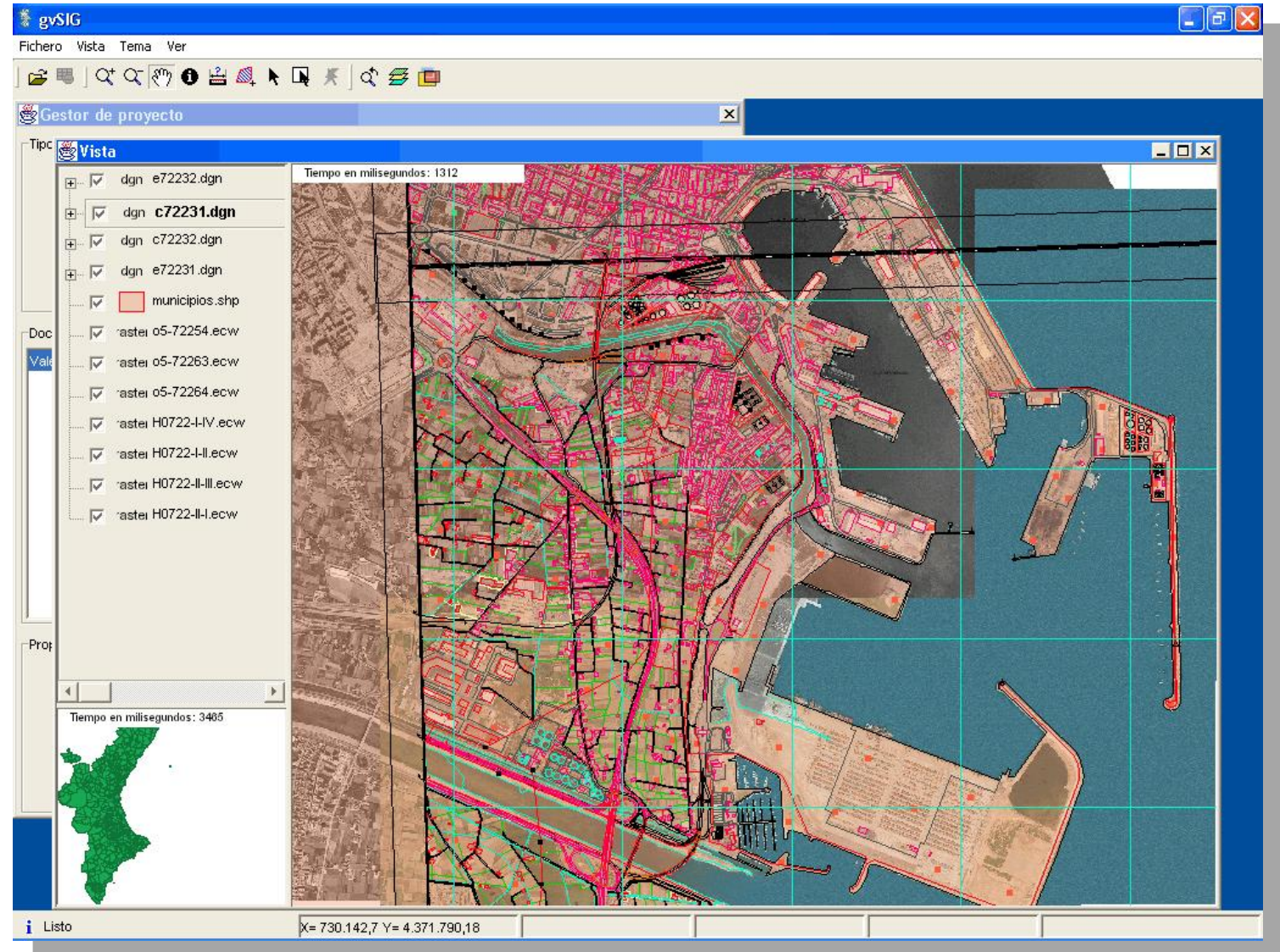




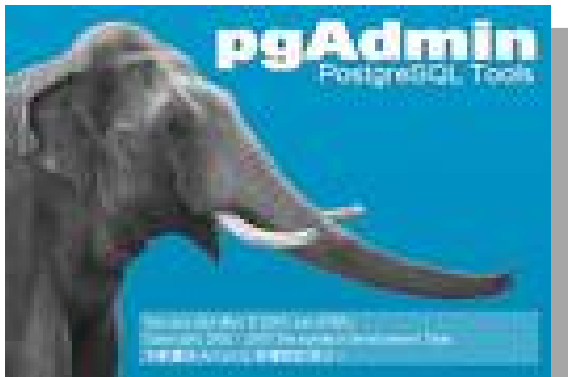
MAKINA
CORPUS

Outils clients SIG desktop





- PgAdmin III
- Scripts maison : python + psycopg2
- Django + geodjango



- Casser le modèle relationnel → performances
 - Les requêtes sont fixées par avance
 - Importance des prétraitements
- Indexation géographique
- CLUSTER

- De plus en plus d'édition web de données
 - Attention à la performance
- Volumes de plus en plus importants
- Problèmes de synchronisation des bases
- Processus d'échanges de données entre les bases
- Gestion des métadonnées
- Support de PostGIS dans les applications clientes classiques (ESRI, Mapinfo...)
- Remplacement des BdD propriétaires

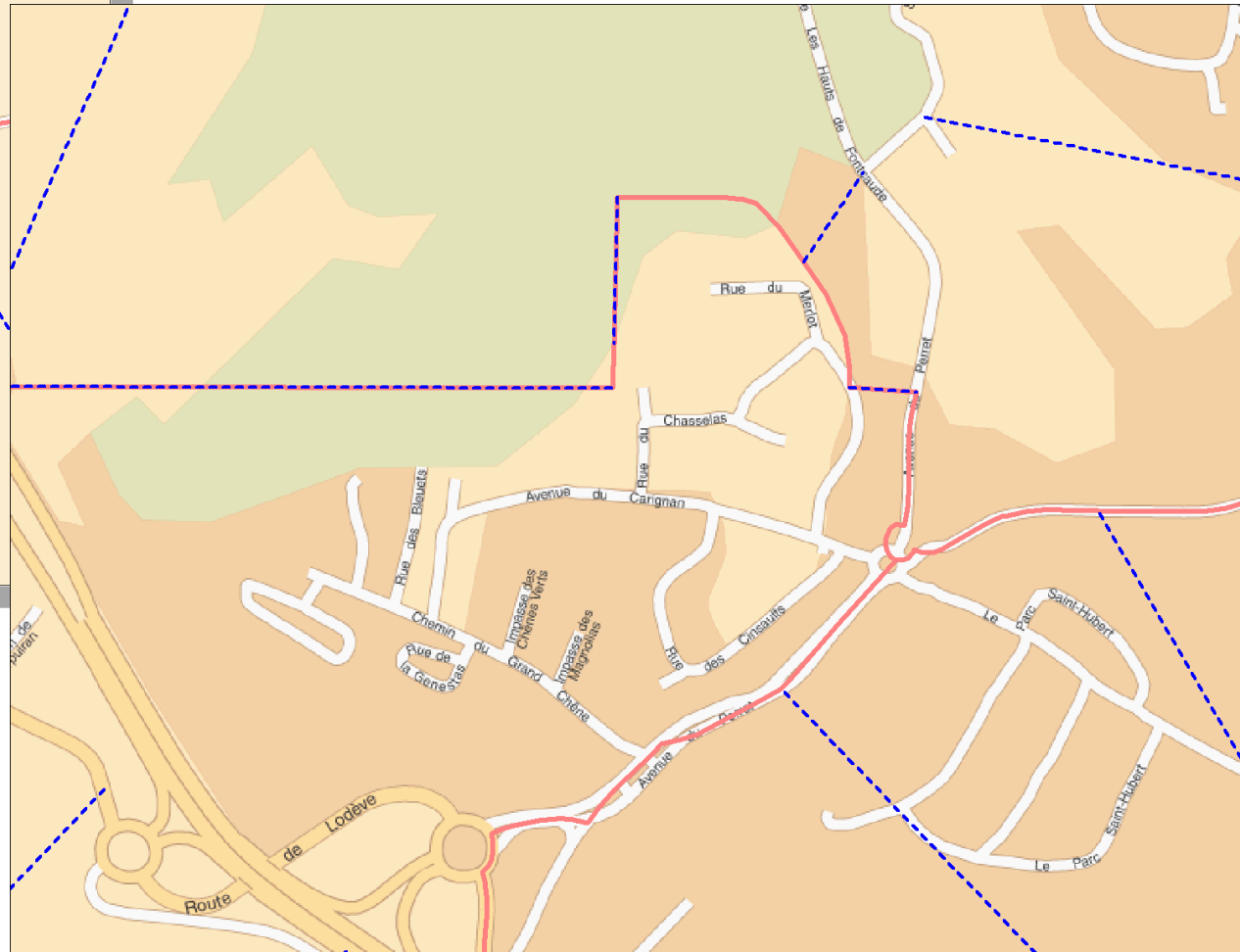
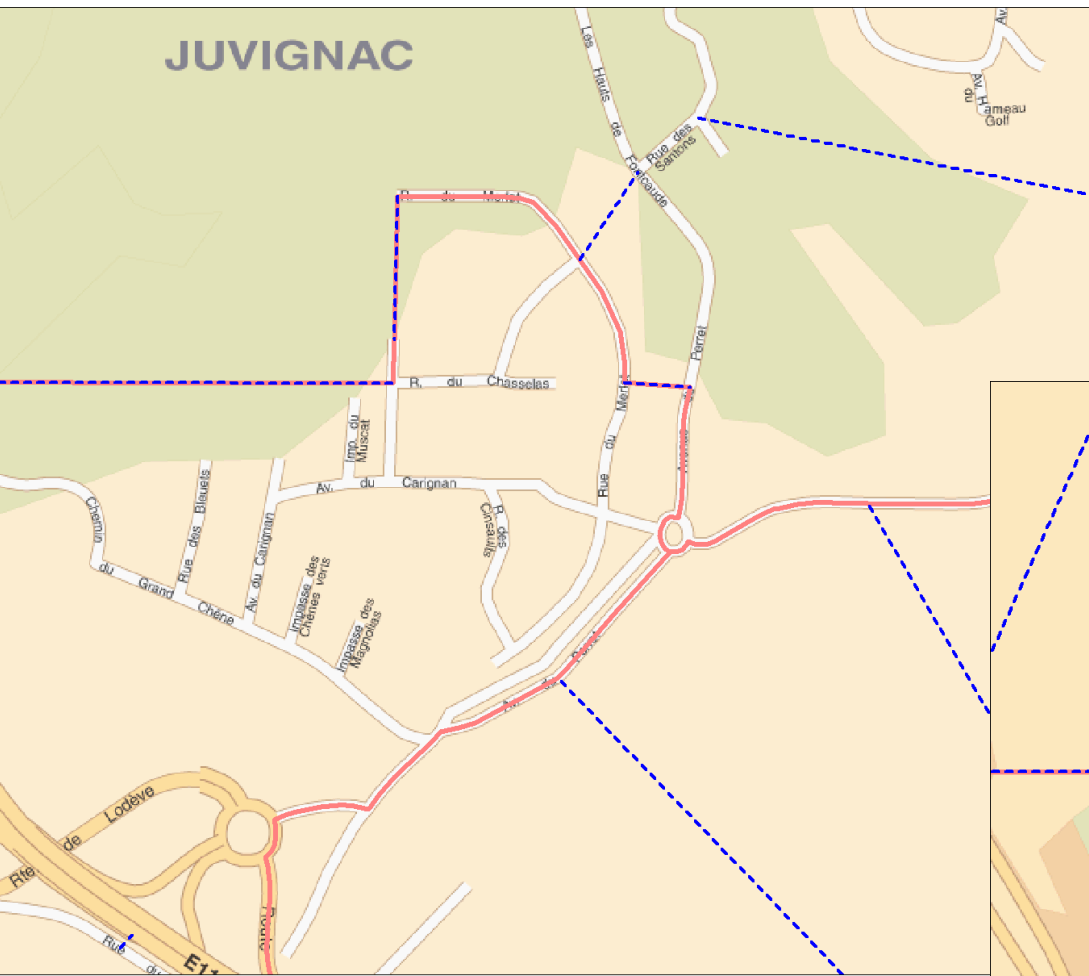
Un projet SIG particulier sous PG graphes de réseaux



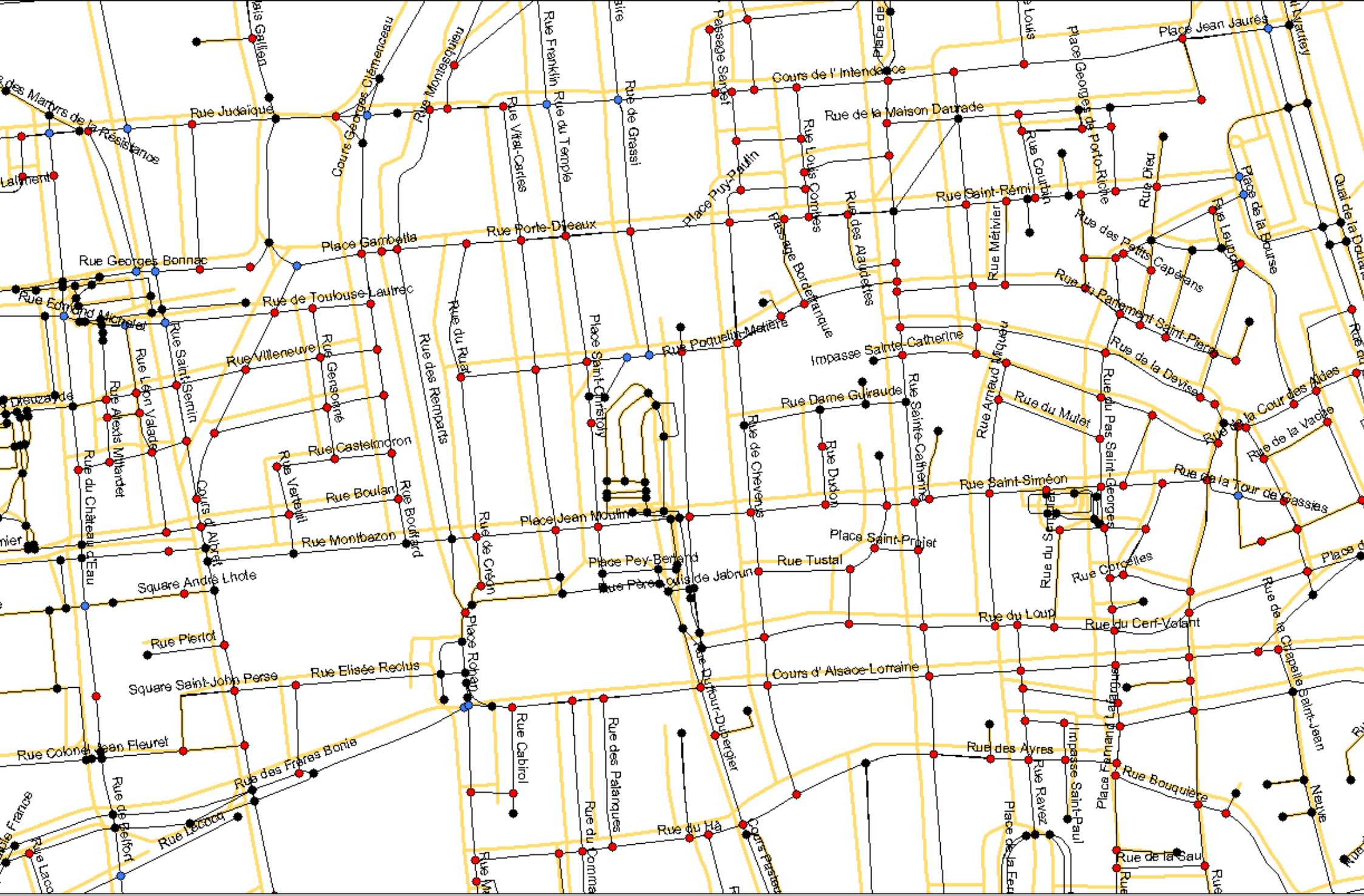
- Client : grand besoin info géographique
 - Graphes de réseaux routiers
 - Données métier associées à ces graphes
 - Données initiales 2004
 - Évolution en parallèle pendant 4 ans
- Désynchronisation des données



Contexte



- Resynchroniser les données
 - «recoller» les graphes
 - Réassocier et/ou reconstruire les données métier
 - Construire un processus de mise à jour des données

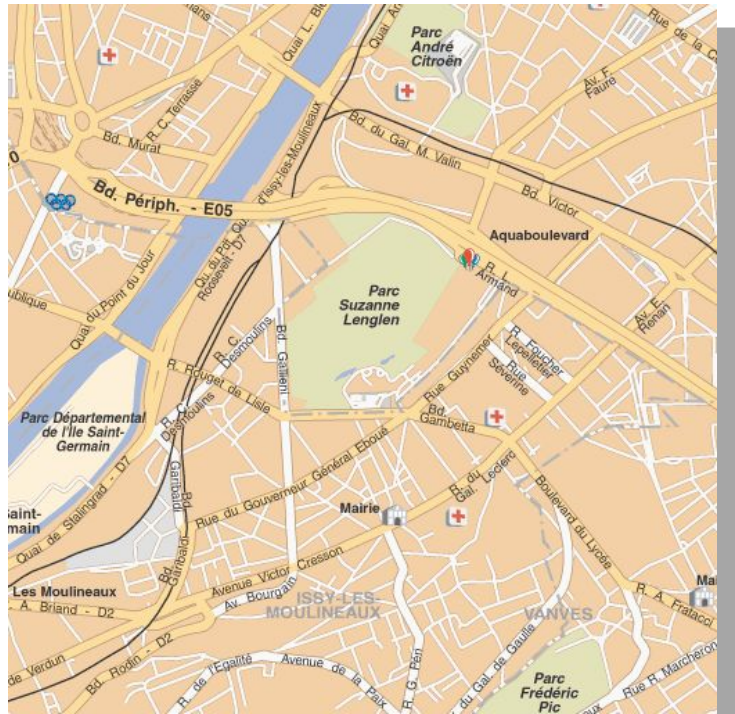


- Méthodologie des traitements effectués
 - (1) Chargement données de référence
 - (2) Chargement données métier
 - (3) Création du modèle d'entrée d'appariement
 - (4) Appariement des graphes
(noeuds, voies, tronçons)
 - (5) Correspondances des données métier
 - (6) Recréation de données métier
 - (7) Création de modèle de sortie
 - (8) Export du modèle de sortie

- Sémantiques
 - Basées sur la typonymie
 - Traitement de texte
- Topologiques
 - Basées sur la structure des graphes
 - Calcul matriciel ou parcours de graphes
- Géométriques
 - Basées sur les objets géométriques
 - Le moins possible
 - Traitements géométriques (Postgis)

- «ELT» : Extract, Load, Transform
- Traitements longs et lourds
 - Chaine de production globale ~ 1 mois
- Traitements dans Postgres + Postgis + quelques outils externes
- Non utilisés :
 - Gestion de droits
 - Accès concurrentiels
 - Réplication





Edit Data - s92s08260 [s92s08260:5432] - TEST_PROD_08_07 - nav_noeud_app

Fichier Édition Affichage Aide

100 lignes.

	id [PK] bigint	id_nav bigint	insee character var	the_geom geometry
1	1	13297	29019	0101000020BE68000011801
2	2	13429	29019	0101000020BE68000063A1A
3	3	13481	29019	0101000020BE6800000091A
4	4	13525	29019	0101000020BE680000EB50E
5	5	13539	29019	0101000020BE6800000700C
6	6	13609	29019	0101000020BE68000060ADC
7	7	13629	29019	0101000020BE680000C6126
8	8	13682	29019	0101000020BE680000DBB85
9	9	13697	29019	0101000020BE68000086617
10	10	13704	29019	0101000020BE680000338CF
11	11	13708	29019	0101000020BE680000358B1
12	12	13711	29019	0101000020BE680000D2833
13	13	13719	29019	0101000020BE68000014216
14	14	13721	29019	0101000020BE6800005D505

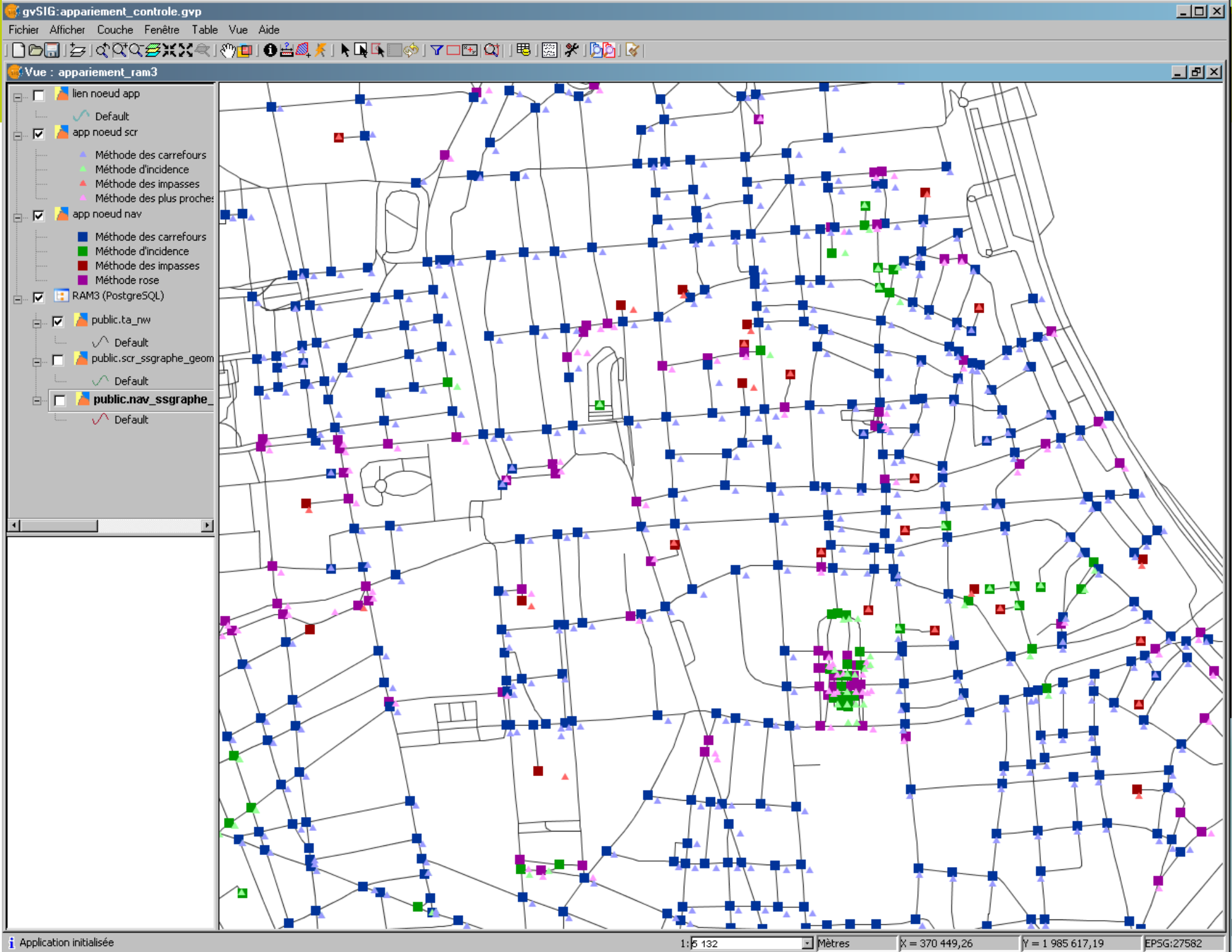
Edit Data - s92s08260 [s92s08260:5432] - TEST_PROD_08_07 - nav_troncon_app

Fichier Édition Affichage Aide

Facteur remplissage

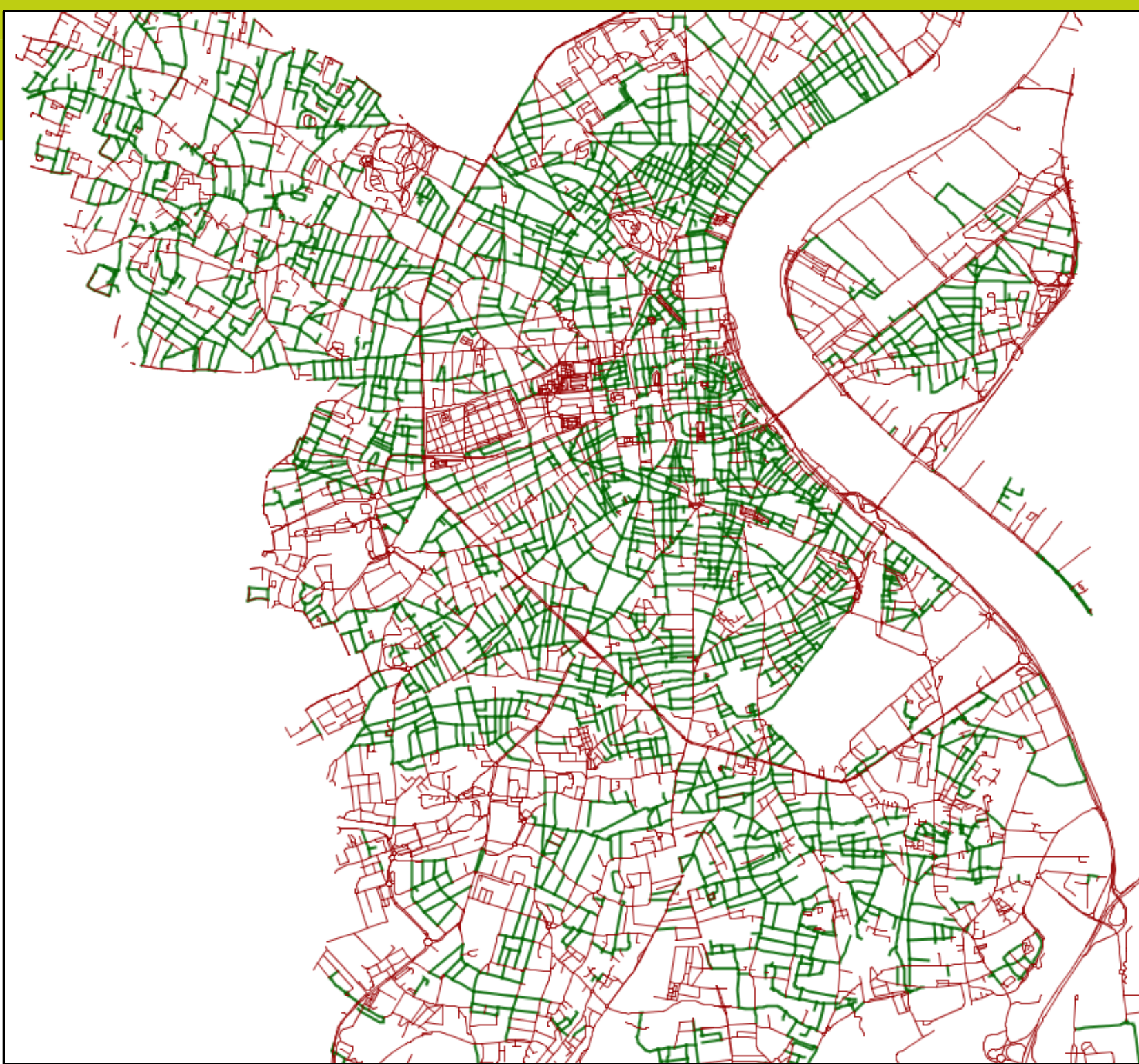
100 lignes.

	id [PK] bigint	id_nav bigint	id_noeud_fro bigint	id_noeud_to bigint	id_voie bigint	insee character var	longueur integer
1	1	95886	74	77	2028	29019	128
2	2	95905	464	473	2715	29019	62
3	3	95910	435	430	2130	29019	33
4	4	95922	967	981	2001	29019	43
5	5	95974	3455	3416	1905	29019	58
6	6	95980	1039	1040	2740	29019	15
7	7	95982	372	337	3477	29019	82
8	8	95988	103	104	781	29019	15
9	9	96000	1013	1054	3000	29019	45
10	10	96009	428	400	3447	29019	107
11	11	96010	843	790	4341	29019	47
12	12	96015	1673	1663	3160	29019	63





MAKINA
CORPUS



- ~ 50 Tables
 - + tables temporaires
 - + tables de controle et stats
 - Entre quelques Mo et quelques Go
 - Entre 1500 et > 20M enregistrements
- ~ 30 champs de géométrie
- Volume total : ~ 100Go, 150Go en fin de traitement
- En entrée ~ 15 tables
- En sortie ~ 10 tables

- Serveurs windows
 - Deux serveurs de test et prod
 - Dimensionnement large
- Clients windows
 - Avantage libre : serveurs de dev en local
- Développement
 - PgAdmin
 - Éditeurs classiques
 - SVN
 - Python + psycopg2

- Talend SDI
 - Données métier
 - Sybase vers Postgres
 - Postgres vers Sybase
- Shp2pgsql, pgsq2shp & ogr2ogr
 - Intégration des données géographiques de référence
 - Exports de la géographie en fin de process



-

- Organisation du code
- Extension en C sous windows
- Paramétrage serveur
- Monitoring et statistique d'exécution
- Intégration d'outils externes peu adaptés
- Environnement Windows
- Optimisation des requêtes

- Organisation du code
- Mettre le plus de code coté serveur : plpgsql
- Utilisation de fonctions, regroupées par modules
- Requêtes EXECUTE
- Utilisation ELT
 - transactions longues
 - mécanisme de cache inefficace
 - problèmes avec le cache lors de changement de la structure de la base (résolus en 8.3 ?)
 - Problème d'accès au log des requêtes unitaires
- Compromis modularité / généricité / performances

- Extensions en C sous windows
- Environnement de développement complexe à mettre en place (DevC++, bibliothèques...)
- Problèmes avec certaines bibliothèques
- Manque de documentation
- Les modules C déstabilisent le serveur

- Paramétrages serveur
- Manque de connaissance admin système
- Fortement lié au matériel (baie SAN...)
- Quels paramètres pour quels usages ?
- Paramètres nécessaires dépendant des requêtes effectuées
- Nécessité de benchmarker, long

- Monitoring serveur et statistiques d'exécution
 - Monitoring serveur : admin sys
 - Statistiques d'exécution des traitements
 - Requêtes longues
 - Pgfoine : pas forcément adapté au traitements type ELT
 - Problème des requêtes unitaires EXECUTE
- Instrumentation dans le code plpgsql et python
- Utilisation d'astuces du type
`SELECT nextval('compteur'), * FROM ...`

- Intégration d'outils externes peu adaptés
 - Programmes de traitements en C++ windows (avec interface)
 - Tournent en partie coté client
 - Mauvaise intégration dans la chaine de traitement
 - transferts réseau
 - conversion de formats de données
- Recoder les applications en modules pg

- Environnement windows
- Manque de connaissance admin sys
- Passage à l'échelle sur serveur gonflé problématique
 - Utilisation de > 3Go de mémoire
 - Stabilité du serveur
- Monter en compétence
- Passer sous Unix (Linux, *BSD)
- Les modules C empêchent la migration

- Optimisation des requêtes
- Validation des algos et requêtes sur échantillons
- Tests grandeur réelle difficiles car longs
- Obligation de gérer manuellement les index et la structure
- Problèmes des updates massifs (MVCC)
- Vacuum et autovacuum
- Loi de Pareto (80-20)

- Développement
- Optimisation
- Rationalisation du code
- Passage sous environnement UNIX
- Intégration des outils externes sous forme de modules
- Généralisation de l'utilisation de Postgresql

OUI !

- **PostgreSQL est bien soluble dans les SIG....**
- **... ou l'inverse !**
- **Expériences positives**
- **Montée en compétence des équipes**
- **Volonté de généraliser l'utilisation de PG**
- **Marché très vaste**
- **Manque de support des applications (SIG)**





**MAKINA
CORPUS**

Questions ?

Vincent Picavet

vincent.picavet@makina-corpus.com

Makina Corpus

www.makina-corpus.com

contact@makina-corpus.com

Tel : 01 44 82 00 80