

**PGDay.fr 2008**

**4 ans de PostgreSQL chez Cityvox**

**Les dessous d'un succès**

Toulouse, le 4 octobre 2008  
Guillaume Smet – Open Wide

- Open Wide : une SSL, 3 métiers
  - OS4I : embarqué,
  - **Accelance : infogérance,**
  - **OWSI : Systèmes d'Informations.**
- Guillaume Smet
  - Chez Open Wide depuis 2003
  - Directeur Technique OWSI
  - Pilotage du projet Cityvox de 2004 à début 2008
  - Utilisateur de PostgreSQL depuis 2001
  - guillaume dot smet at openwide dot fr

- Cityvox en quelques points :
  - Effectif : 50aine de personnes
  - Leader des cityguides en France : 3.2 M de visiteurs uniques, 35 à 40 M de pages vues/mois
  - Plusieurs sites : Cityvox, Webcity, une partie du portail Orange...
  - Une filiale du groupe France Telecom depuis mars 2008
  - 150k lieux, 170k événements, 310k artistes référencés, le tout dans une base PostgreSQL
  - Des pics événementiels : 3 M de pages vues pour la Fête de la Musique 2008



Chercher par nom, par mot clé...

RECHERCHER

sur ☒ Cityvox ☐ le web avec **voila**

Mardi 23 septembre 2008

Constant - Automne



Accueil

Vie locale

Bonnes Adresses

Agenda culturel

Cinéma

Enfants

Annonces

Vidéos

Rencontres

&gt; Cityvox &gt; Toulouse &gt; Accueil Toulouse



## Cityvox Toulouse

Partager la vie d'ici au quotidien.

12 / 24 °C

&gt; Prévisions météo



Autres villes



### Open de Toulouse : complètement green

La 9ème édition de cette manifestation se déroule du 29 septembre au 5 octobre au Golf de Toulouse-Seilh. Plus de 120 joueurs

parmi les meilleurs mondiaux vont se retrouver et s'affronter. Un Open des Personnalités rassemble amateurs et vedettes du monde sportif dans une ambiance conviviale. [Suite...](#)

#### Bonnes adresses

Cinéma

Agenda des sorties

Vie Locale

Enfants

Toulouse : notre sélection d'hôtels pas chers

#### Billetterie & Invitations Toulouse



#### Le Clan des Divorcées

Théâtre Humour

532 Avis ★★★★★

Café Théâtre Les Minimes

A partir de 18.00 €

au lieu de 20.00 € (-10%)

Réserver

Offrir

#### Les derniers avis déposés

Bonnes Adresses

Evénements

#### Le Félibre - 31000 Toulouse

★★★★★ Par : [r3born](#)

Calme et cosy

La Taverne de Maître Kanter Sedel - 31000

Club

Newsletters

Météo

A la télé

E-mail :

Mot de passe :



&gt; Mot de passe oublié ?

&gt; S'inscrire

#### A l'affiche

Partenaires commerciaux



#### Hôtel Saint Claire

Le Temps d'un week-end? Découvrez la formule week-end douceur et ses prix tout doux. . . Atmosphère paisible, décoration entre tradition française et feng shui. Envie d'un shiatsu relaxant?

&gt; En savoir plus



#### Stringfellows

Le restaurant parisien vous reçoit du lundi au samedi de 21h à 2h00 pour un dîner-spectacle à partir de 65€

&gt; En savoir plus



[Accueil](#) [Mon profil](#) [Les membres](#) [Le guide](#)[Se connecter](#) | [S'inscrire](#)

## Restaurants

Avec notre moteur d'idées trouvez ce que vous souhaitez en sélectionnant au fur et à mesure les critères de choix qui vous intéressent. Vous pouvez démarrer votre recherche soit en tapant le texte de votre recherche dans le moteur de ci-dessous, soit en cliquant sur les critères situés à gauche. Vous pouvez à tout instant supprimer l'un des critères que vous avez sélectionné en cliquant sur le lien "enlever".

Entrer un ou plusieurs éléments à rechercher

[Rechercher](#)

Votre sélection : Restaurants

### Affiner la recherche

[Ville](#) [Cuisine](#) [Critère](#) [Prix](#)

Utilisez le champ ci-dessus si votre ville n'apparaît pas dans la liste.

Aix-en-Provence : 510  
Amiens : 233  
Angers : 275  
Besançon : 191  
Bordeaux : 1014  
Brest : 270  
Caen : 413  
Clermont-Ferrand : 309  
Dijon : 461  
Grenoble : 602

### 34597 résultats correspondants à votre recherche

Trier par [Note + avis](#) [Note](#) [Avis](#)

#### Music Hall Restaurant

Français  
75008 - Paris

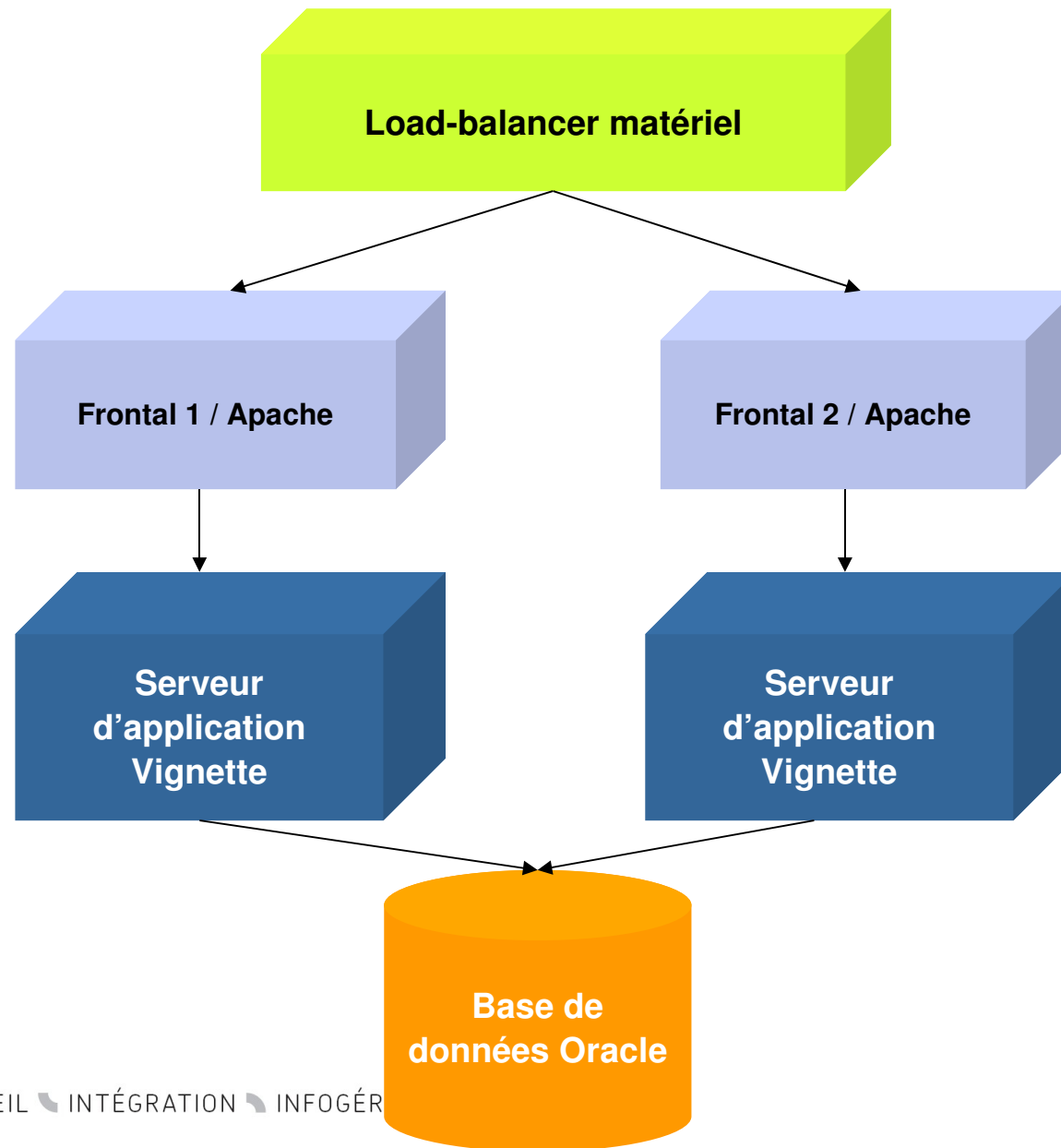
[Show must go on](#)[Présentation et avis](#) 80 avis

#### L'Aromatik

World Food  
75009 - Paris

 77 avis

- En 2004, lors de nos premiers contacts avec Cityvox :
  - Site inaccessible régulièrement le mercredi
  - Affichage des pages très lents lors des pics d'activité quotidiens
  - Peu de garantie de haute disponibilité
- Problèmes de communication avec l'hébergeur -> diagnostic des problèmes de performances difficile
- **Il était impossible pour Cityvox d'assurer sa croissance dans ces conditions**



- Le coût d'ajout d'un serveur était prohibitif :
  - Matériel Sun
  - Licence Solaris
  - Licence Vignette
  - Coût d'hébergement
- Le passage à plus de 10 millions de pages vues par mois demandait :
  - Un changement de licence Vignette pour une licence supérieure
  - La migration vers une nouvelle version de Vignette -> réécriture du site du TCL vers Java



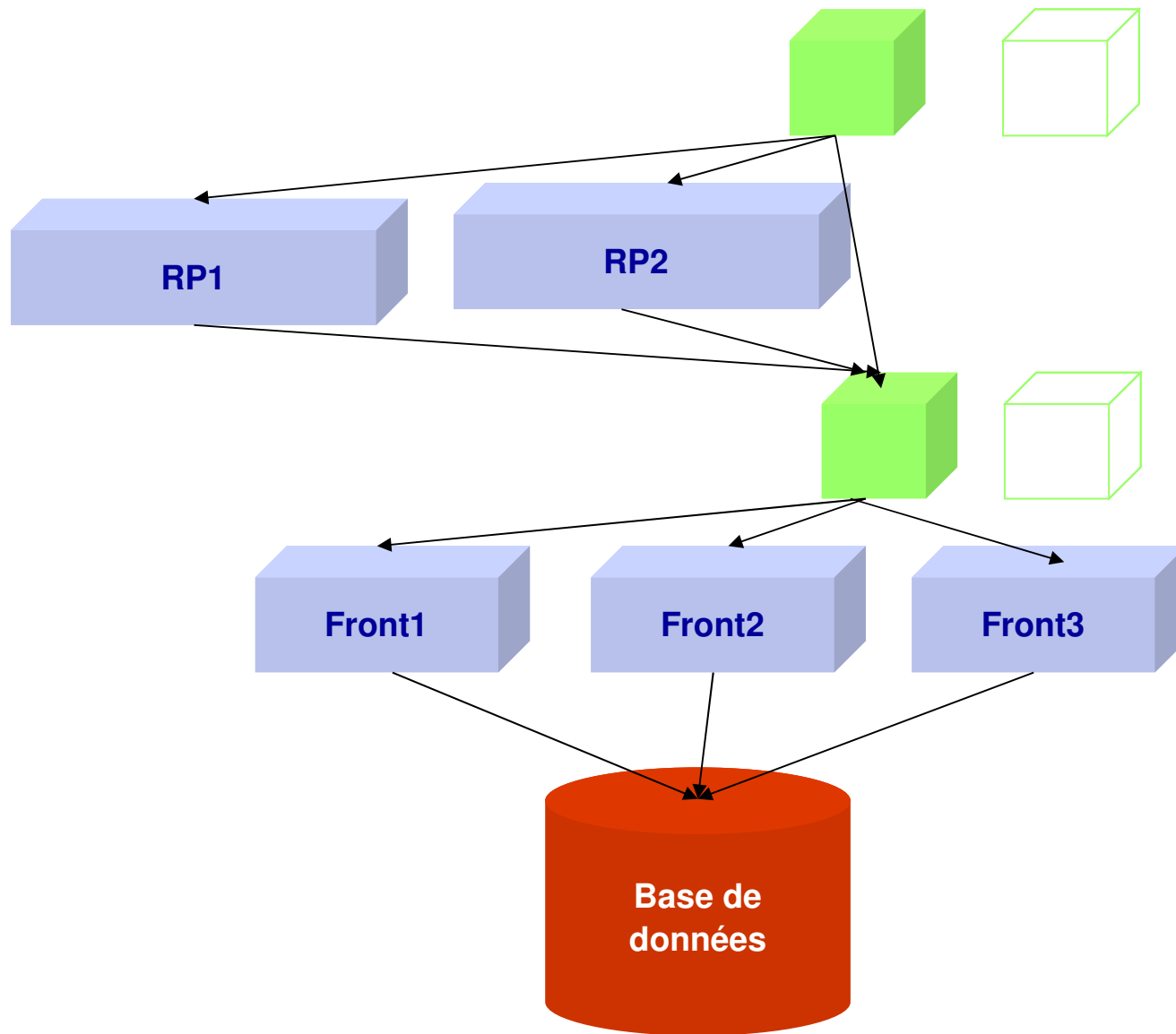
- Passage de Vignette à PHP
- Fourniture d'un environnement de développement (Eclipse, gestion de sources...)
- Conception commune du framework applicatif et de l'infrastructure d'hébergement
- La base de données est un risque majeur
  - Coeur de métier
  - 12 Go, 300 tables, 600 000 photos
  - On met de côté pour plus tard l'éventuelle migration vers un SGBDR Open Source : **Oracle est conservé et on en profite pour migrer vers une nouvelle version (10g).**  
**On évite ainsi l'effet « big bang »**

- Contacts avec Oracle peu fructueux pour la migration vers la nouvelle version :
  - Règlement de 2 ans de support demandé avant d'accepter de vendre une licence 10g
  - Le coût de la nouvelle licence est prohibitif
  - **Les commerciaux d'Oracle sont souvent les meilleurs commerciaux de PostgreSQL**
- Cityvox veut reconsidérer la possibilité de migrer la base de données vers un SGBDR Open Source
  - Même niveau de fonctionnalités et même fiabilité
  - **Le choix de migrer vers PostgreSQL est entériné**

- Migration pénible car peu d'outils (2004)
- Migration du schéma relativement sereine mais besoin d'automatiser car 300 tables à migrer
- Problèmes lors de la migration des données :
  - Caractères non UTF-8 présents dans la base ;
  - Extraction des 10 go de photos très pénible
  - Processus de conversion très long
  - Optimisation des types (numeric)
- Aujourd'hui :
  - ETL type Talend
  - D'autres outils plus bas niveau comme pgLoader

- Open Wide a réécrit la majorité des requêtes et des procédures stockées
  - Requêtes, pas trop de soucis en dehors de quelques oraclicismes : requêtes complexes mais assez standard
  - Procédures stockées, il faut tout réécrire en pl/pgsql
  - Pas de souci de régression côté performances
- Ce qui a marqué le client au premier abord :
  - La qualité du client psql et notamment la complétion
  - La qualité de la documentation

# Nouvelle architecture 2005



**Load-balancers Linux  
redondés**

**Reverse-proxies  
CentOS 3 / Squid 2.5**

**Load-balancers Linux  
redondés**

**Frontaux  
CentOS 3 / Apache / PHP /  
tthtpd**

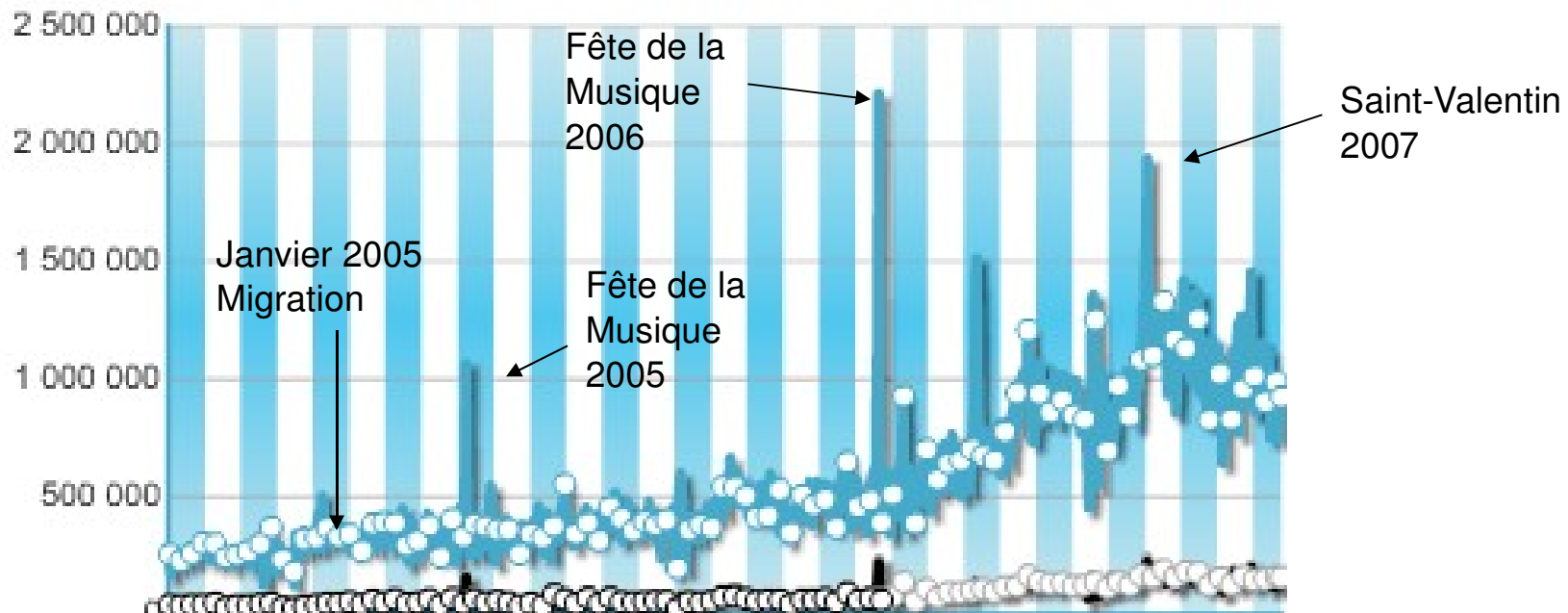
**Base de données  
CentOS 3/ PostgreSQL 7.4**

- Démarrage du projet chez Open Wide en mai 2004
- Livraison du socle applicatif en juillet 2004 :
  - Gestion des pages, des blocs, des URLs sémantiques
  - Système de cache avancé (blocs et pages, gestion des en-têtes HTTP)
  - Couche d'accès aux données existantes
  - Base migrée vers PostgreSQL
- Développement des pages à partir du socle dès ce moment là :
  - Prise en main rapide par Cityvox du socle et du nouvel environnement de développement
  - Accompagnement méthodologique par Open Wide
- Fin décembre 2004 : le nouveau site est prêt
- Mi-janvier 2005 : mise en production



- Une solution entièrement Open Source dont les coûts de licence sont nuls
- Une fiabilité et une disponibilité accrue
- Coût de migration amorti car obligation de migration Vignette
- Une architecture beaucoup plus scalable qu'auparavant
- +20% de pages vues dès le premier mois
- **De 7 M de pages vues/mois à 35 M de pages vues/mois en 2 ans**
- Une capacité importante à absorber les pics événementiels (**2.2 millions de pages** vues en une journée)

- Entre 2005 et 2007 :
  - Croissance du trafic de 10 mbits/s à 40 mbits/s
  - Croissance des visites de 300k pages vues à 1 million de pages vues par jour (en moyenne sur le mois)



- Problème de scalabilité de la version 7.4 (quadri Xeon)
  - Difficile à quantifier jusqu'à la mise en orbite du site pour la St Valentin 2006
  - Seuil des « 4 de charge »
- Explosion du site suite à la lenteur de certaines requêtes : manque de visibilité : quelles requêtes posent problème ?
- Cache du site sous forme de fichiers / serveur : on éteint des serveurs pour mieux tenir la charge...
- SPOF sur la base de données - on se repose sur les sauvegardes

- Scalabilité : passage à PostgreSQL 8.1
- Surveillance de l'activité de la base :
  - Utilisation de PQA au coup par coup (Practical Query Analyzer)
  - Réécriture complète de PQA en pgFouine et industrialisation de la génération des statistiques
  - Formation de Cityvox à l'utilisation du EXPLAIN ANALYZE
- Mise en place de Memcached pour avoir du cache distribué
- Profiling avec Xdebug et KCachegrind
- Mise en place d'un serveur PostgreSQL en standby (merci la 8.1) : plus de SPOF

## Navigation/OngletsTopBlock: dynamique [Cacher](#)

NoCacheCacheProvider::generateContent : 0.1482 s

Clé : /data/var/cityvox-test/cache/main/Navigation/OngletsTop/OngletsTopBlock\_langue.FRA\_ville.TOU\_theme.ACC\_charte.CIT\_kewego\_UrlAbs

```
select tt.themepincipal, t.libelle, tcl.rubrique from vilthemcharte vtc, themchartelang tcl, traduction t, typetheme tt where vtc.codevil = 'TOU' and tcl.codelang = 'FRA' and vtc.codech = 'CIT' and vtc.codech = tcl.codech and vtc.codech = tcl.codech and t.codelang = tcl.codelang and t.nordre = tt.nordretrad and tcl.rubrique = tt.codetypetheme group by tt.themepincipal, t.libelle, tcl.rubrique order by tcl.rubrique [ 0.0015 s ] Backtrace
```

```
SELECT vtc.codech, t.libelle, tcl.rubrique, tcl.ordre, tc.nordretrad FROM vilthemcharte vtc, themchartelang tcl, themcharte tc, traduction t WHERE vtc.codevil = 'TOU' AND vtc.codech = 'CIT' AND vtc.codech = tcl.codech AND vtc.codech = tcl.codech AND t.codelang = tcl.codelang AND t.nordre = tc.nordretrad AND tc.codech = tcl.codech AND tc.codech = tcl.codech AND tcl.ordre > 0 ORDER BY tcl.rubrique, tcl.ordre [ 0.0025 s ] Backtrace
```

```
SELECT codetypevil FROM vilsite where codevil = 'TOU' [ 0.0005 s ] Backtrace
```

```
SELECT distinct e.numleve, e.libeseve, e.nbcrive, e.notmoyeve, e.identifianturl, el.pushbilletterie, l.numlieu, l.nomlieu, e.evemaj, min(r.dateseance) AS dateseance FROM representations r, representationsstock rs, evenement e, lieu l, evelieu el, vilquartier vq WHERE r.numrep = rs.numrep AND r.numleve = e.numleve AND r.codestatut = 'OK' AND rs.complet != 1 AND rs.stockactuel > 0 AND date_trunc('day', r.dateseance) >= date_trunc('day', now()) AND r.numlieu = l.numlieu AND el.numlieu = r.numlieu AND el.numleve = r.numleve AND l.codequar = vq.codequar AND vq.codevil = 'TOU' AND el.pushbilletterie > 0 AND date_trunc('day', el.dfinevelieu) >= date_trunc('day', now()) AND r.dateseance - (r.dmailresa * interval '1 hour') > CURRENT_TIMESTAMP GROUP BY e.numleve, e.libeseve, e.nbcrive, e.notmoyeve, e.identifianturl, el.pushbilletterie, l.numlieu, l.nomlieu, e.evemaj ORDER BY pushbilletterie asc, e.evemaj desc, e.libeseve LIMIT 1 [ 0.1338 s ] Backtrace
```

Total : 4 requêtes [ 0.1383 s ]

## Navigation/ArianeBlock: dynamique

NoCacheCacheProvider::generateContent : 0.0029 s

Clé : /data/var/cityvox-test/cache/main/Navigation/ArianeBlock/ArianeBlock\_langue.FRA\_ville.TOU\_theme.ACC\_charte.CIT\_SpKch\_exep\_UrlAbs

## Navigation/EnteteCentraleBis: dynamique [Requêtes](#)

NoCacheCacheProvider::generateContent : 0.0152 s

Clé : /data/var/cityvox-test/cache/main/Navigation/EnteteBis/EnteteCentraleBis\_langue.FRA\_ville.TOU\_theme.ACC\_charte.CIT\_path\_regfes\_UrlAbs\_dom.test.cityvox.fr

## Publicite/ExpendBlock: pas de cache par bloc mais cache sur reverse proxy

Clé : /data/var/cityvox-test/cache/main/Publicite/ExpendBlock/ExpendBlock\_langue.FRA\_ville.TOU\_theme.ACC\_page.1

Généré en : 0.0000 s

## News/NewsFlashBlock: pas de cache par bloc mais cache sur reverse proxy

Clé : /data/var/cityvox-test/cache/main/NewsFlashBlock\_langue.FRA

Généré en : 0.0001 s

Navigation/OngletsTopBlock: **dynamique**

NoCacheCacheProvider::generateContent : **0.1482 s**

Clé : /data/var/cityvox-test/cache/main/Navigation/OngletsTop/OngletsTopBlock\_langue.FRA\_ville.TOU\_theme.ACC\_charte.CIT\_kewego\_UrlAbs

```
select tt.themepincipal, t.libelle, tcl.rubrique from vilthemcharte vtc, themchartelang tcl, traduction t, typetheme tt where vtc.codevil = 'TOU' and tcl.codelang = 'FRA' and vtc.codech = 'CIT' and vtc.codech = tcl.codech and vtc.codeth = tcl.codeth and t.codelangue = tcl.codelang and t.nordre = tt.nordretrad and tcl.rubrique = tt.codetypetheme group by tt.themepincipal, t.libelle, tcl.rubrique order by tcl.rubrique [ 0.0015 s ] 
```

/data/services/web/cityvox-test/include/data/ville/Ville.class.php à la ligne 962

Database->query( " select tt.them ..." )

/data/services/web/cityvox-test/sites/default/blocks/Navigation/OngletsTopBlock.class.php à la ligne 35

Ville->getMetaRubriques( "CIT", "FRA" )

/data/services/web/cityvox-test/include/layout/Block.class.php à la ligne 305

OngletsTopBlock->buildContent()

/data/services/web/cityvox-test/include/cache/NoCacheCacheProvider.class.php à la ligne 29

Block->generateContent()

/data/services/web/cityvox-test/include/layout/Block.class.php à la ligne 275

NoCacheCacheProvider->getContent( false )

/data/services/web/cityvox-test/include/cache/CacheManager.class.php à la ligne 50

Block->getContent( false )

/data/services/web/cityvox-test/include/layout/Page.class.php à la ligne 90

CacheManager->buildContent( object[CityHome], array[3], array[3], array[3] )

/data/services/web/cityvox-test/www/index.php à la ligne 366

Page->buildOutput()

```
SELECT vtc.codeth, t.libelle, tcl.rubrique, tcl.ordre, tc.nordretrad FROM vilthemcharte vtc, themchartelang tcl, themcharte tc, traduction t WHERE vtc.codevil = 'TOU' AND vtc.codech = 'CIT' AND vtc.codech = tcl.codech AND vtc.codeth = tcl.codeth AND tcl.codelang = 'FRA' AND t.codelangue = tcl.codelang AND t.nordre = tc.nordretrad AND tc.codeth = tcl.codeth AND tc.codech = tcl.codech AND tcl.ordre > 0 ORDER BY tcl.rubrique, tcl.ordre [ 0.0025 s ] 
```

```
SELECT codetypevil FROM vilsite where codevil = 'TOU' [ 0.0005 s ] 
```



# Outils de diagnostic : la trace (prod)

## Blocs

**Navigation/ControleOuvert** : en cache | généré le : 23/09/2008 18:36:30 | expire le : 24/09/2008 18:36:30  
*MemCachedCacheProvider::getValidCachedValue : 0.0000 s*  
Clé : /data/var/cityvox-prod/cache/main/Navigation/Control/ControleOuvert\_langue.FRA\_ville.TOU\_charte.CIT  
Généré en : 0.0011 s | Serveur : 192.168.41.22

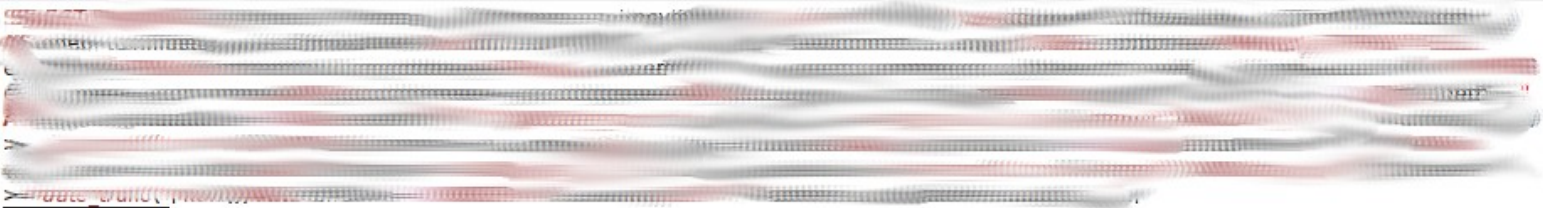
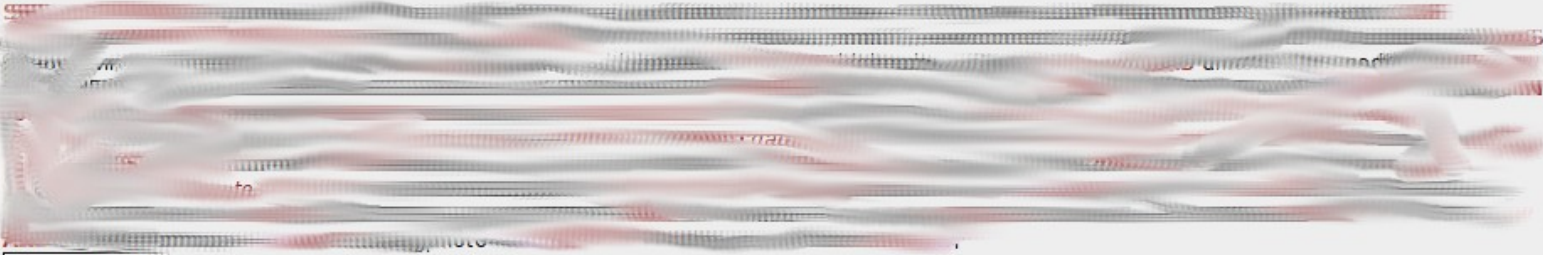
**Navigation/MetaTags** : en cache | généré le : 24/09/2008 00:13:15 | expire le : 25/09/2008 00:13:15  
*MemCachedCacheProvider::getValidCachedValue : 0.0000 s*  
Clé : /data/var/cityvox-prod/cache/main/Navigation/Meta/MetaTags\_langue.FRA\_vil.TOU\_the.ACC\_dom.cityvox.fr\_rst\_pp.CityHome\_aut\_NVX.sans\_GMP.-\_URL\_ch  
Généré en : 0.0095 s | Serveur : 192.168.41.21

**Publicite/PubliciteMegaBanBlock** : pas de cache par bloc mais cache sur reverse proxy  
Clé : /data/var/cityvox-prod/cache/main/PubliciteMegaBanBlock\_langue.FRA  
Généré en : 0.0002 s

**Navigation/HeaderBlockBis** : en cache | généré le : 24/09/2008 00:01:04 | expire le : 25/09/2008 00:00:00  
*MemCachedCacheProvider::getValidCachedValue : 0.0000 s*  
Clé : /data/var/cityvox-prod/cache/main/Navigation/Head/HeaderBlockBis\_langue.FRA\_ville.TOU\_theme.ACC\_UriAbs\_kewego\_dom.cityvox.fr  
Généré en : 0.0280 s | Serveur : 192.168.41.23

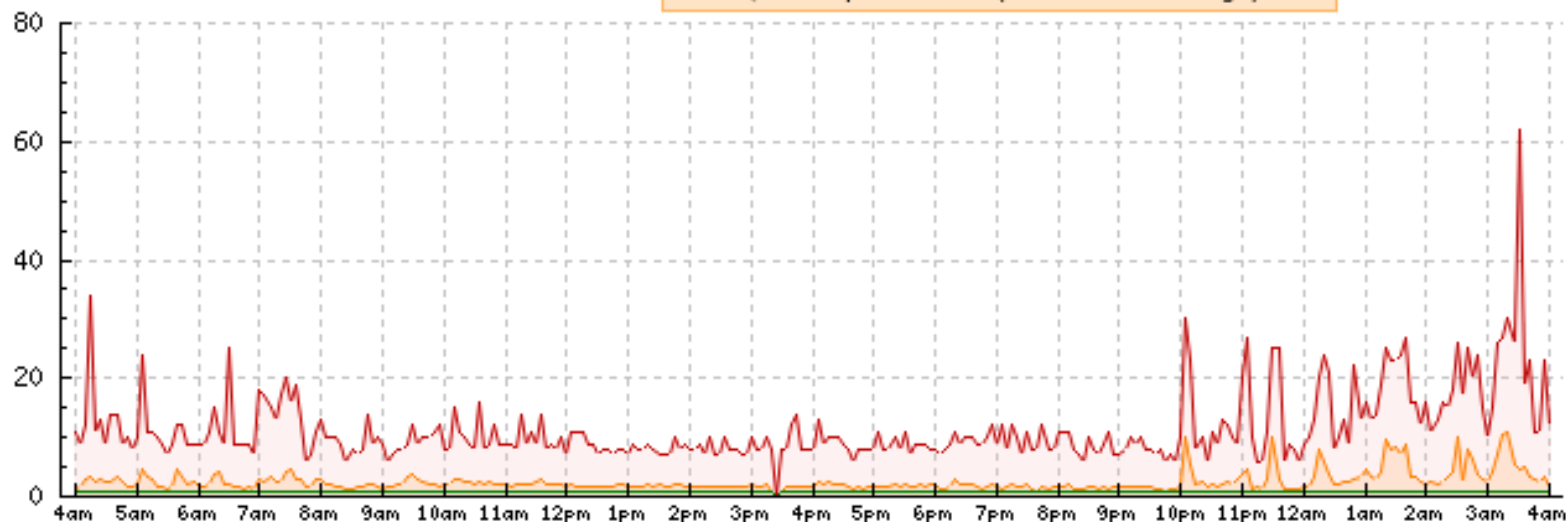
# Outils de diagnostic : pgFouine

Queries that took up the most time (N) ^

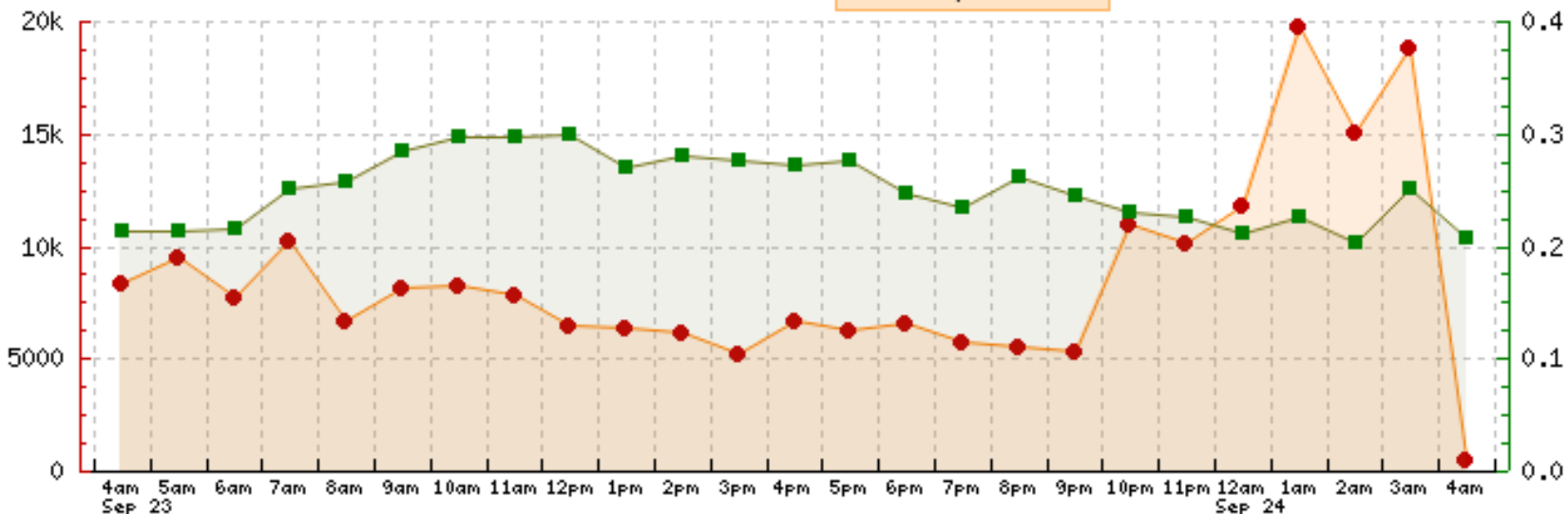
Rank	Total duration	Times executed	Av. duration (s)	Query
1	3h55m28s	43,683	0.32	 <a href="#">Show examples</a>
2	1h07m31s	26,408	0.15	 <a href="#">Show examples</a>

# Outils de diagnostic : pgFouine

Queries per second (5 minutes average)



All queries



- Grosses lenteurs sur le site sans croissance du nombre de visiteurs (stats)
  - Aucun problème de lenteur spécifique
  - Difficulté à analyser les logs de par leur taille et la diversité des URLs (AWStats et Webalizer ont du mal)
  - Google génère un trafic énorme notamment sur de nombreuses pages pas en cache (jusque 12 Mb/s...)
- Mise en place d'un serveur web et base de données dédié à Google : capacité à limiter la pollution générée par Google
- Dommage qu'on n'ait pas pu accéder en lecture au serveur en standby (travail en cours pour la 8.4 ou la 8.5 !)

- Requêtes géographiques : cube et earthdistance
  - Pas totalement satisfaisant car répartition très inégale et statistiques constantes
- 
- Recherche plein texte dans le backoffice avec des %mot % : quelques essais sont faits avec tsearch2 mais :
  - À l'époque, complexe à maintenir car pas dans le core (les joies des migrations de contrib/) : c'est maintenant le cas en 8.3
  - Manques fonctionnels : pas de recherche par préfixe (sera probablement présent en 8.4), pas de recherche approximative

- De la garantie des serveurs...
- De performances, principalement au niveau de la base de données – base plus grosse, beaucoup de requêtes très complexes et serveur vieillissant
- De flexibilité et de capacité à maintenir une plate-forme qui a beaucoup évolué en 3 ans
- Bon timing pour le passage en 8.3 qui promet beaucoup

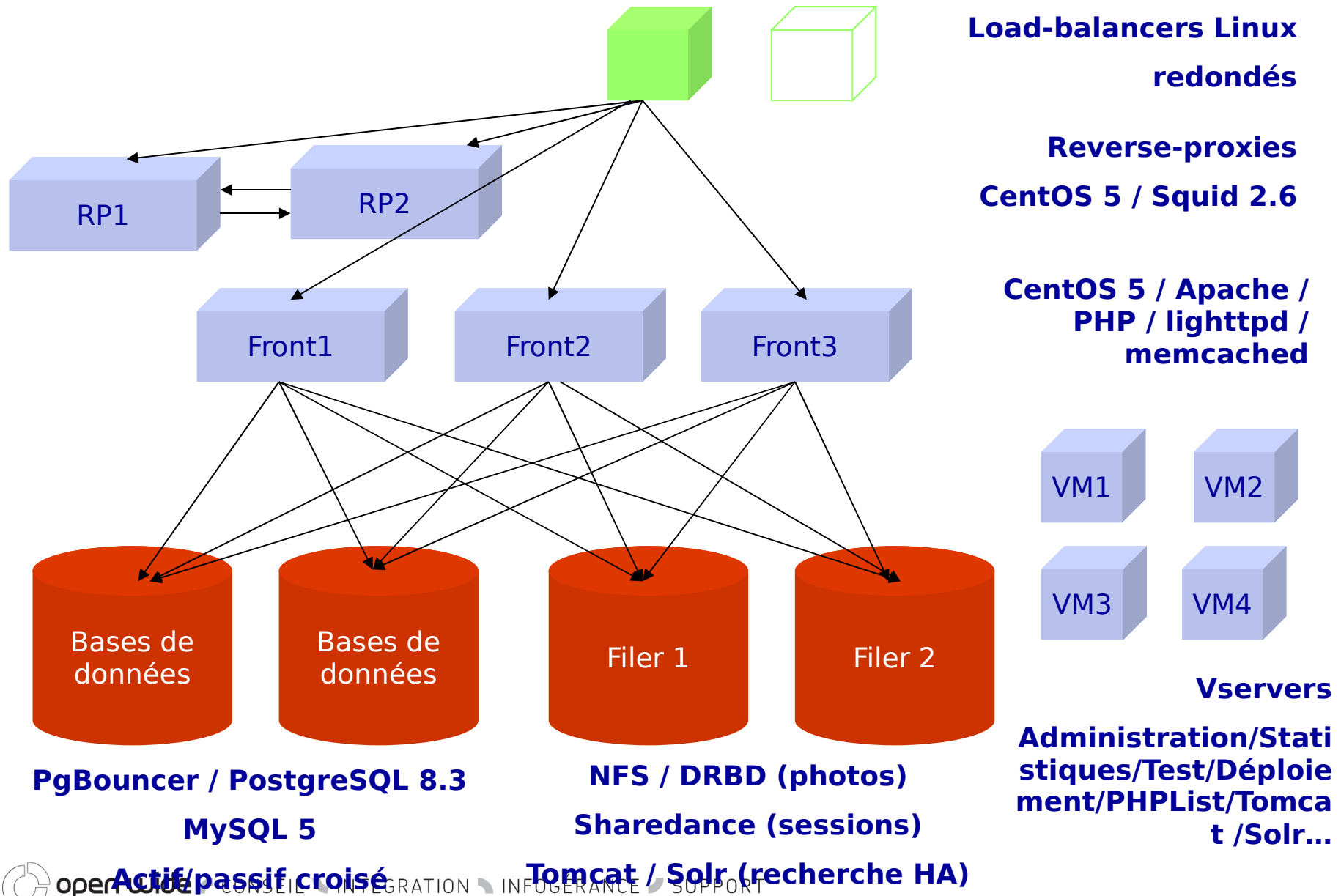


- Travail sur tous les problèmes structurels
  - Delay\_pools de Squid 2.6 pour canaliser les robots d'indexation : plus de serveur dédié et de base dédiée
  - Centralisation des photos sur NFS + DRBD (gestion trop lourde en mode distribué)
  - Mise en place de Sharedance pour les sessions
- Centralisation de tous les sites sur les serveurs frontaux
  - Plus de fiabilité mais aussi plus de connexions parallèles (connexions persistantes)
- Utilisation de vservers (applications périphériques)
- Mise en place de toutes les bonnes pratiques PostgreSQL du moment
- Mise en place d'instances Solr en HA

- En 3 ans, PostgreSQL s'est très industrialisé
  - La 8.3 est pour bientôt (prévue pour décembre 2007)
  - Mise en place de l'autovacuum
  - Mise en place de phpPgAdmin
  - Mise en place de pgBouncer (merci Skype)
  - Mise en place de walmgr (Skytools, merci Skype)
  - Mise en place de PostGIS pour les requêtes géographiques  
-> rapidité et statistiques fiables
- L'ensemble en packaging RPM sur CentOS 5

- A mener avec beaucoup de prudence vu la diversité et la complexité des requêtes
  - Etalée sur plusieurs mois en amont de décembre (date de sortie prévue)
  - Détection de régressions au cours du cycle de test, régressions très vite corrigées :
    - pg\_restore beaucoup plus long à cause du nouvel autovacuum (de 15 à 50 minutes)
    - Requête simples sur des clés de type varchar 8% plus lentes qu'en 8.2 (après correction, la 8.3 devient 45% plus rapide que la 8.2 dans ce cas précis)
- Comme toujours, réactivité exemplaire : les problèmes ont été résolus en quelques jours

- Le retard de la sortie de la 8.3 ne nous arrange pas
  - Le nouveau mode de développement mis en oeuvre pour la 8.4 (commit fests) a été mis en place pour éviter que cela se reproduise
- Mise en production le jour de la sortie de la 8.3 :
  - Sécurisée par toute la phase de test amont
  - Validée par le client : la confiance est importante
- Importance de cette démarche de test au cours de la phase de développement :
  - Pour certaines régressions, il aurait été trop tard une fois la 8.3 sortie car les corrections ont demandé des modifications importantes

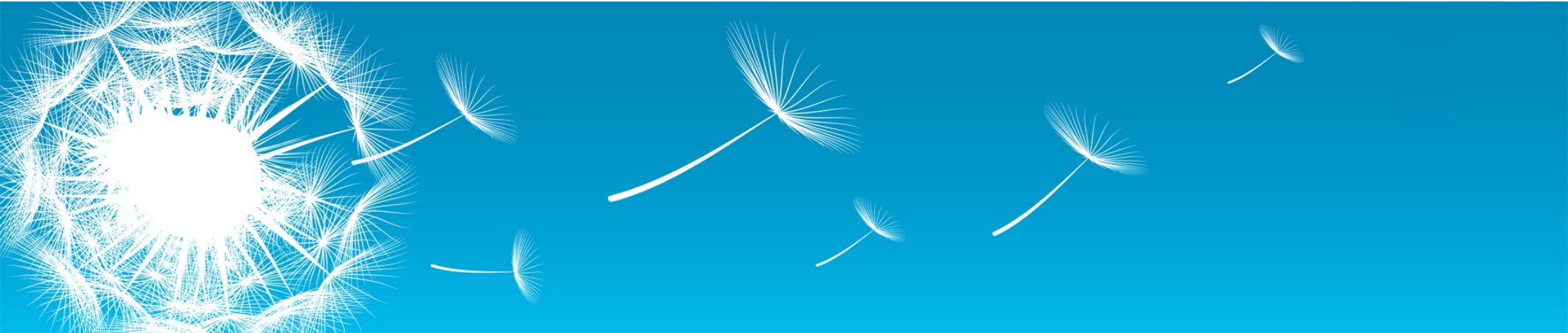


- En production depuis début février 2008
- Grande réussite en terme de :
  - Performances,
  - Fiabilité,
  - « Administrabilité » de la plate-forme.
- PostgreSQL 8.3 est une grande version : le risque (relatif et maîtrisé) pris a été payant.



- Mise en place de Solr (serveur de recherche basé sur Lucene interrogeable en HTTP) pour la recherche full text et moyens de navigation différents
  - Moins intégré à la BDD que tsearch2
  - Etat de l'art en ce qui concerne les fonctionnalités de recherche (faceting, recherche approximative...) : tsearch2 ne propose pas encore toutes les fonctionnalités de Lucene/Solr
  - Nouvelles manières de rechercher et naviguer
  - Ne remet évidemment pas en cause le choix de PostgreSQL comme élément clé de la plate-forme !

- PostgreSQL a été le bon choix : tout au long des 4 années, les nouvelles versions ont permis d'absorber la croissance du trafic
- La confiance du client a été un des facteurs clés de réussite (choix initiaux, confiance dans la mise en production de la 8.3...)
- La communauté PostgreSQL est une richesse : implication, disponibilité, qualité des échanges...
- Elle a produit en 3 ans énormément d'outils de qualité permettant de répondre aux besoins concrets des infrastructures basées sur PostgreSQL



**Vos questions techniques, anecdotiques,...  
sont les bienvenues !**

- En savoir plus sur Cityvox : <http://www.cityvox.fr/> et <http://www.cityvox.net/>
- En savoir plus sur Open Wide : <http://www.openwide.fr/>
- pgFouine : <http://pgfouine.projects.postgresql.org/>
- SkyTools :  
<https://developer.skype.com/SkypeGarage/DbProjects/SkyTools/>
- PgBouncer :  
<https://developer.skype.com/SkypeGarage/DbProjects/PgBouncer/>
- Repository Yum : <http://yum.pgsqlrpms.org/>
- PhpPgAdmin : <http://phppgadmin.sourceforge.net/>
- Solr : <http://lucene.apache.org/solr/>