

Audit Technique Geod'Air

Thierry Caminel

Architecte Senior,
Communauté Scientifique d'Atos

07-11-2016

1

Contexte

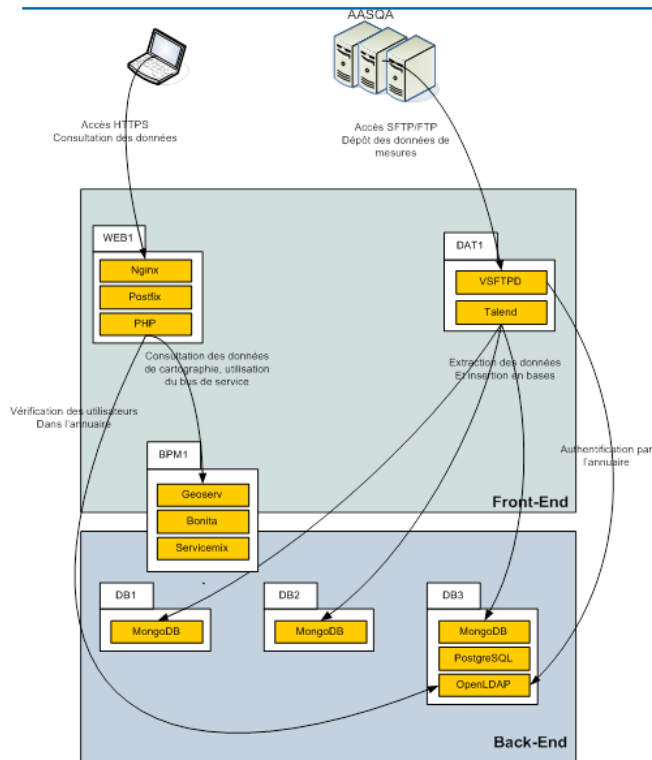
Contexte

Dans le cadre de la préparation d'un plan d'action pour le projet Geod'Air, un audit technique a été réalisé en Novembre 2016.

Il est particulièrement destiné à analyser le problème de performance sur la génération de rapport.

Ce document présente en synthèse les résultats de cet audit et les préconisations qui en ont découlé.

Rappel Architecture / 1



Nom	Description
Nginx	Serveur web front-end (alternative à Apache).
Talend	ETL extraction, la transformation et le traitement d'ensembles de données volumineux et hétérogènes.
ServiceMix	ESB (Bus de services d'Entreprise). Fuse est le nom de la version sous licence payante.
PostgreSQL	Serveur de base de données pour le BPM.
MongoDB	Serveur de base de données NoSQL.
GeoServer	Server pour partager et modifier des données géographiques.
GeoWebCache	Mise en cache des tuiles cartographiques générées par GeoServer.
JQuery & JQuery-UI	Bibliothèque JavaScript.
OpenLayers	Bibliothèque de fonctions JavaScript assurant un noyau de fonctionnalités orientées vers la mise en place d'applications clientes Web cartographiques fluides.
GeoExt	Bibliothèque de fonctions et objets JavaScript combinant ExtJS et OpenLayers.
Tomcat	Conteneur d'application
Symfony	Framework PHP

The diagram illustrates the data architecture. It starts with 'Fichiers Auto' (Automatic Files) and 'Fichiers Manu' (Manual Files) on the left. These feed into a 'talend' logo at the top. The data then flows into a 'mongoDB' database (green cylinder). From 'mongoDB', a curved arrow labeled 'Calculs statistiques' (Statistical Calculations) points to a 'PostgreSQL Référence' database (blue cylinder). Finally, both the 'mongoDB' and 'PostgreSQL Référence' databases feed into an 'IHM Web' (Web Interface) represented by an orange box on the right.



1

Audit d'architecture

Appréciation d'ensemble

- ▶ Architecture globale J2EE satisfaisante
 - Usage correct d'un ETL, bus de services, patterns J2EE pour XML et DAO
- ▶ Bon niveau d'industrialisation
 - Intégration continue, déploiement successifs, ..
 - Test unitaires globalement OK.
- ▶ Problèmes sur les IHM maîtrisés
 - Solution 3-tier, PHP + WS
 - robuste, mais vieillissante et simplifiable
- ▶ Principal problème: performance génération d'un rapport, et retraitements

Problème performance sur Rapports type 'E'

- ▶ Cause racine: mauvais usage de MongoDB pour requêtage complexe
- ▶ Détails
 - Les informations statistiques sont insérées dans une 'table' MongoDB
 - Le référentiel est dans PostgreSQL (122 tables)
 - Requêtes imbriquées
 - Requête d'agrégation avec MongoDB
 - Pour chaque élément, requête de filtrage sur MongoDB
 - » Pour chaque élément, requête sur PostgreSQL
 - Mauvais indexage de MongoDB
 - Vieille version de MongoDB. Problème reconnu sur l'agrégation
- ▶ Résultat: plusieurs heures de traitements, crash, difficultés à maintenir, ...

Solutions possibles

	Amélioration	Simplification	Meilleure combinaison
Description	<ul style="list-style-type: none">• Mise à niveau MongoDB• Optimisation des index• (recodage de l'algo)	<ul style="list-style-type: none">• Remplacement de MongoDB par PostgreSQL	<ul style="list-style-type: none">• Déplacement des données STATS de MongoDB vers PostgreSQL
Avantages/ Inconvénients	<ul style="list-style-type: none">+ Changement minimal architecture et code- Performance non optimales (mais peut être satisfaisante)- Evolutivité faible	<ul style="list-style-type: none">+ Meilleures performances et maintenabilité+ Réduction TCO- Changements significatifs architecture et code	<ul style="list-style-type: none">+ Modification réduites de l'architecture et du code+ Plus de problèmes de performances sur génération de rapport+ Meilleure évolutivité de la partie requêtage

2

Recommendations

Stratégie d'évolution

Déplacer le stockage des données statistiques

➤ Modifications:

- 1) Insertion des enregistrements de stats vers PostgreSQL (simple)
 - 2) Ajout de 1 job ETL Talend (simple)
 - 3) Mise à jours requêtes
 - Calcul du rapport 'E' (complexe)
 - Calcul de dépassement (simple)
 - 'IHM 'accès aux données/ stats' (simple)
- 200 à 400 lignes de codes à modifier + code SQL
- Peux se faire par ajout à l'existante
 - évolution progressive 1) + 2) \Rightarrow 3)

Avantage de l'approche

- ▶ Modifications réduites
 - Ce qui marche bien n'est pas modifié
- ▶ Architecture mieux adaptée au BI
 - Meilleures performances pour les rapports
 - Possibilité d'utiliser des outils de BI pour de nouvelles fonctionnalités
 - Pas nécessaire de connaître MongoDB pour faire évoluer la partie 'métier'
- ▶ Risques réduits

Suggestions moyen terme : BI

1. Passer les tables statistiques sur PostgreSQL
 2. Intégrer un outils de BI (ex: Tableau, QlickView, Spotfire, BIRT)
 - Génération des rapports
 - 'Dashboards' complexes avec graphes, cartes, ...
 3. Stockage 'brut' des informations des fichiers ISO en base de données
 4. Utiliser progressivement l'outils de BI pour simplifier le processus d'ETL
 - Vérifications, calculs moyennes, dépassements,
 - Traitement 'à la demande', plutôt que sur le flux
 5. Séparation de la solution en 2
 - Gestion du référentiel (basé sur la solution actuelle)
 - Reporting interactif ou batch (basé sur un outils BI)
- Bénéfice:
- Simplification, maintenabilité
 - Moins de problèmes liés au référentiel non à jours

Autres recommandations

- ▶ Regrouper bases de données opérationnelle et historique
 - mais gérer explicitement la disponibilité (redondance, sauvegardes,..)
- ▶ Améliorer les requêtes et index PostgreSQL (revue par un DBA)
- ▶ Sécuriser Mongo / PostgreSQL

Thanks

For more information please contact:

T+ 33 1 98765432

F+ 33 1 88888888

M+ 33 6 44445678

firstname.lastname@atos.net

Atos, the Atos logo, Atos Codex, Atos Consulting, Atos Worldgrid, Worldline, BlueKiwi, Bull, Canopy the Open Cloud Company, Unify, Yunano, Zero Email, Zero Email Certified and The Zero Email Company are registered trademarks of the Atos group. April 2016. © 2016 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.



Données et Traitements

