

Développement d'un tableau de bord de l'activité du GEOPortail vaudois

Association pour le Système d'Information du Territoire Vaudois

Travail de Bachelor 2012

Lausanne, le 6 août 2012

Gaël Bernard

Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud - HEIG-VD

Département comem+

Filière Ingénierie des Médias – IT Management

Classe MIT38

Professeur

Daniel Rappo
Professeur Dpt C, Comem+
Avenue des Sports 20
CH-1400 Yverdon-les-Bains
www.comem.ch

Mandant

Xavier Mérour
Coordinateur ASIT-VD
Avenue de l'Université 5
CH-1014 Lausanne
www.asitvd.ch/

Management Summary

Le projet

Ce projet consiste en la mise en place d'outils permettant la diffusion de tableaux de bord et de rapports permettant une gestion transversale des activités de l'ASIT-VD. Il concerne les administrateurs de l'association, les fournisseurs de données, les membres ainsi que le grand public. Ces outils font partie de ce qui est communément appelé l'informatique décisionnelle ou la Business Intelligence.

Solutions implémentées

Deux façons d'accéder aux données sources sont identifiées : Un accès direct ou par le biais d'un datawarehouse. La deuxième solution nécessite la mise en place d'un entrepôt de données ainsi qu'un script permettant d'alimenter quotidiennement le datawarehouse. Ce script ETL est réalisé avec Talend.

Ensuite, afin d'offrir un point d'accès unique à ces données, Talend est également utilisé mais cette fois en tant que couche d'accès. Cette couche est ensuite atteignable par les tableaux de bord par le biais d'un web service de type SOAP. Ces tableaux de bord sont réalisés sous la forme d'un composant Joomla permettant une intégration optimale avec le portail actuel. Le composant intègre Bootstrap dans l'optique de la version 3 de Joomla.

Les graphiques présents sur les tableaux de bord sont réalisés avec AmCharts et les rapports dynamiques avec Birt.

Développement d'un tableau de bord de l'activité du GEOPortail vaudois

Contexte: l'ASIT-VD

L'Association pour le Système d'Information du Territoire Vaudois est le point de rencontre entre fournisseurs et utilisateurs de données géographiques. Afin de jouer ce rôle, l'association met à disposition un GEOPortail.

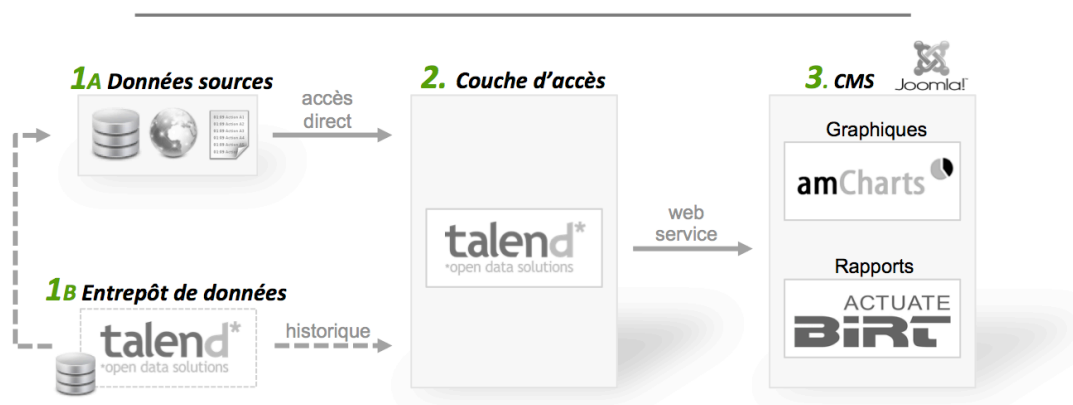
Le but est d'assurer une meilleure coordination aux acteurs oeuvrant pour une bonne gestion du territoire cantonal ainsi que de favoriser l'échange de géodonnées entre partenaires.

Objectif du projet

L'objectif de ce travail est d'intégrer des outils interactifs dans le GEOPortail permettant d'offrir une vision transversale des activités aux différents groupes d'acteurs identifiés.

Exemples d'activités exercées sur le portail:

- Catalogage de ressources géographiques
- Commande de géodonnées



Solution déployée

1a. Données sources

Les données sources sont hétéroclites: bases de données, fichiers de log, web service. La solution est flexible et permet l'ajout de nouvelles sources.

1b. Entrepôt de données

Permet d'assurer la pérenité des informations stockées ainsi qu'un accès rapide en lecture. L'entrepôt est utilisé lorsque la donnée initiale ne permet pas d'accéder à l'historique ou si le temps de traitement est trop lent.

2. Couche d'accès

Accède aux différentes ressources (points 1) permettant de déployer un Web service offrant un point d'accès unique aux données.

3. CMS

Utilise le web service (point 2) pour la diffusion de tableaux de bord. Ce dernier contient des indicateurs et des graphiques ainsi que des liens sur des rapports dynamiques.

Auteur: Gaël Bernard
Répondant externe: Xavier Mérour
Prof. responsable: Daniel Rappo
Sujet proposé par: Xavier Mérour

Remerciements

Je tiens à remercier chaleureusement toute l'équipe de l'ASIT-VD : Annabelle Mas, Xavier Mérour et Yves Blatti. J'ai ressenti un réel soutien durant ce travail. D'ailleurs, l'esprit d'équipe et l'entraide se ressentent très fortement au sein de l'association et je pense que ceci est dû à la très bonne coordination menée par Xavier Mérour. Je le remercie de m'avoir fait confiance en me confiant ce projet passionnant.

J'ai apprécié l'aide technique d'Yves Blatti et également un soutien sur des aspects plus académiques de la part d'Annabelle Mas.

Table des matières

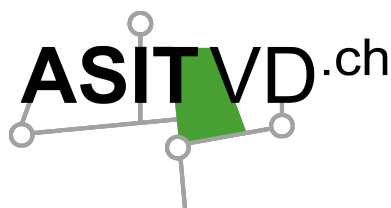
Introduction	4
Mandataire	4
GEOPortail.....	5
Perspectives et évolution.....	6
Objectif du projet	7
Attentes du mandataire:.....	7
Livrables.....	8
Structure du rapport	8
1 Approche théorique de la réalisation de tableaux de bord	9
1.1 Préparer	10
Identification du public cible.....	10
Identification des besoins.....	10
Résultats attendus de la phase de préparation.....	11
1.2 Collecter	12
Accès direct.....	12
Datawarehouse.....	12
Accès direct vs. Datawarehouse.....	14
1.3 Accéder	15
Architecture trois tiers.....	15
1.4 Distribuer	16
Portail d'entreprise.....	16
1.5 Diffuser	17
1.6 Choix des outils	18
1.7 Conclusion de la partie théorique	18
2 Réalisation du TB	19
2.1 Contraintes techniques imposées	20
2.2 Démarche:	21
2.3 Besoins	22
Public cible et groupes d'utilisateurs.....	22
Stratégie de découpe des tableaux de bord.....	23
Besoins.....	24
2.4 Structure de l'entrepôt de données	26
Conception de l'entrepôt de données.....	26
Comparaison accès direct - datawarehouse.....	27
Taille critique du datawarehouse.....	29
2.5 Alimentation de l'entrepôt de données	30
Talend comme outil ETL.....	30
Présentation du job Talend.....	31
Exemples concrets de besoins nécessitant l'entrepôt.....	32
Exemple détaillé : analyse d'un log4j.....	33
Exportation.....	35
2.6 Couche d'accès	36
Aperçu global.....	36
Exportation.....	37
Temps de réponse.....	38

2.7 Composant Joomla	39
Joomla 2.5	39
Structure de l'application.....	39
Résultat obtenu	39
Bootstrap	43
2.8 Reporting	45
Reporting	45
2.9 Graphiques	48
3 Bilan	50
3.1 Objectifs atteints	51
3.2 Divergences avec le travail préparatoire	52
3.3 Problèmes rencontrés	54
Attaquer le web service Google Analytics	54
Problèmes de rendus avec Internet Explorer 7	54
3.4 La suite du projet	54
Intégration du projet au portail actuel.....	54
Suggestions d'amélioration.....	55
3.5 Gestion du temps	57
Diagramme de Gantt	58
Tâches importantes.....	59
Conclusion	60
Ensemble du projet	60
Note personnelle	60
Tables des illustrations	61
Webographie	62
Déclaration formelle	63
Annexes	64
I. Public cible	64
II. Documentation	64
III. Cahier des charges	64
IV. Procès-verbal du bilan intermédiaire	64
V. CD-ROM	64

Introduction

L'objectif de ce travail de Bachelor consiste à mettre en œuvre un environnement permettant de mesurer la synergie créée autour des activités de l'ASIT-VD. Ce projet constitue le travail final d'une formation d'ingénieur des médias.

Mandataire



Ce projet est réalisé pour l'Association pour le Système d'Information du Territoire Vaudois (ASIT-VD). Cette association joue le rôle d'intermédiaire entre les utilisateurs et les fournisseurs de données géographiques. Les fournisseurs désignent entre autres : le canton de Vaud, différentes communes ainsi que certains services industriels et gestionnaires de réseaux. Dans le groupe des utilisateurs on retrouve différents bureaux de géomètres, d'architectes et bien d'autres encore.

Objectifs de l'association :

- *assurer une meilleure coordination aux acteurs oeuvrant pour une bonne gestion du territoire cantonal*
- *favoriser l'échange et le partage de géodonnées*
- *promouvoir l'échange de connaissances entre partenaires*

Source : *asitvd.ch*

Partenaires de l'association :

<i>Administration Cantonale Vaudoise</i>	<i>1</i>
<i>Communes</i>	<i>107</i>
<i>Sociétés, bureaux techniques et association professionnelles</i>	<i>120</i>
<i>Gestionnaires de réseau</i>	<i>16</i>
<i>Ecoles</i>	<i>4</i>
<i>Membre individuels</i>	<i>8</i>
<i>Membres</i>	<i>256</i>
<i>+ Non-Membres</i>	<i>213</i>
<i>Partenaires</i>	<i>469</i>

Source : *assemblée générale statutaire
Etat au 1 janvier 2012*

GEOPortail

Le GEOPortail est un outil mis en place par l'ASIT-VD permettant de faire le lien entre les fournisseurs et les utilisateurs d'informations géoréférencées. Celui-ci permet ainsi d'offrir un point d'entrée homogène aux géodonnées pour le canton de Vaud.

L'illustration 1 permet d'illustrer le rôle central que joue le portail.

Le GEOPortail offre quatre services :

(Les descriptions sont issues du document de l'AG statutaire)

1. GEOCatalogue

Permet de connaître qu'elles sont les données existantes sur le territoire cantonal ainsi que leurs caractéristiques (505 fiches).

2. GEOCommande

Permet la diffusion en ligne de nombreuses données de façon simple et rapide (240 géodonnées).

3. GEOService

Annuaire des outils de consultation de géodonnées qui pointe sur différents guichets cartographiques cantonaux et communaux.

4. GEOGuichet

Environnement d'accès aux GEOService des partenaires qui propose un système de sécurisation et de gestion des droits d'accès à des services web standardisés de type WMS et WFS.

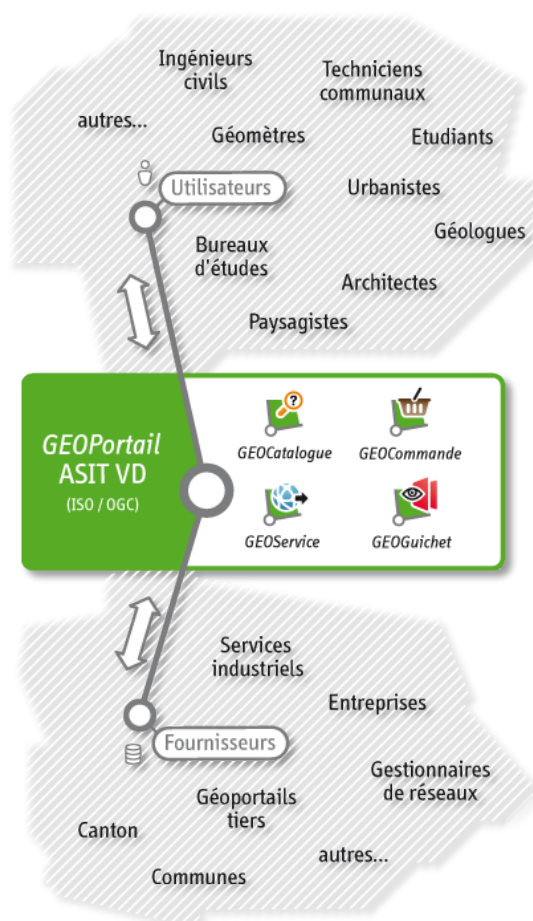
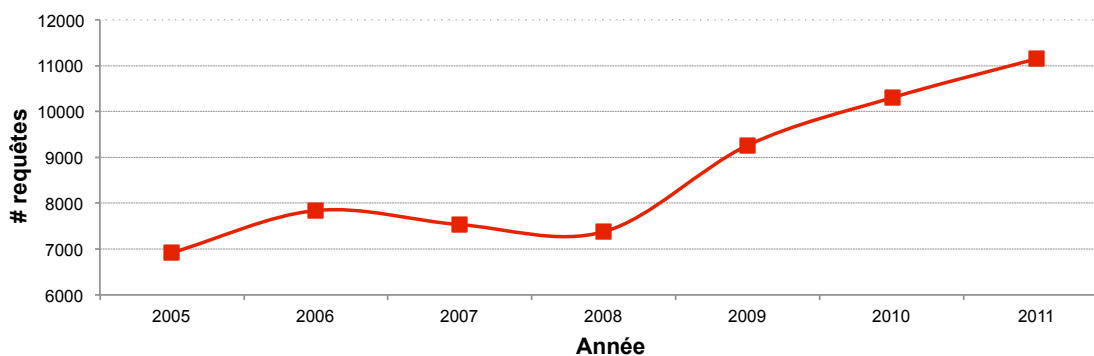


Illustration 1 - Schématisation du GEOPortail
(Source : www.asitvd.ch)

Perspectives et évolution

Entre le 2 août 2011 et le 2 août 2012, **10'240** commandes et **1'524** devis ont été effectués par le biais du GEOPortail. Le catalogue compte actuellement plus de 240 géodonnées. L'illustration ci-dessous démontre que ces chiffres sont d'ailleurs en constantes évolutions depuis quatre ans.

Evolution du nombre de requêtes 2005-2011



*Illustration 2 - Evolution du nombre de requêtes
(Source : présentation de l'AG statutaire 2012)*

L'évolution représentée ci-dessus devrait s'accroître avec l'arrivée de la loi sur la géoinformation. C'est une loi fédérale entrée en vigueur en octobre 2007 et dont le pendant vaudois entre en vigueur maintenant.

« La loi doit notamment garantir que des géodonnées mises à jour, d'un niveau de qualité adéquat et d'un coût raisonnable, couvrant l'intégralité du territoire suisse, seront disponibles durablement afin de permettre une large utilisation. »

Loi fédérale sur la géoinformation : admin.ch

Cette loi renforce le rôle central que joue l'ASIT-VD entre les fournisseurs et les consommateurs de données géographiques. La loi venant tout juste d'entrer en vigueur, il est difficile de se prononcer sur l'impact réel de celle-ci sur l'activité de l'ASIT-VD. Cependant, il est probable que le nombre de partenaires de l'association augmente rapidement ces prochaines années. Le tableau de bord permettra justement d'avoir une meilleure vision de ce type d'augmentation et de mieux suivre l'activité par domaine.

Objectif du projet

L'objectif est de permettre la diffusion de tableaux de bord et de rapports dynamiques sur le GEOPortail de l'association. Il est destiné aux fournisseurs de données, aux administrateurs de la plateforme, aux membres de l'association et également au grand public.

Attentes du mandataire:

- Répondre à des problématiques de suivi et de reporting des activités du GEOPortail
- Assurer un meilleur suivi de diffusion des données des fournisseurs
- Améliorer le suivi des requêtes et services à disposition pour les utilisateurs.

Le projet a une forte composante technique car il implique la mise en place de solutions concrètes permettant de répondre aux différents gains escomptés par l'association. Cependant, la communication est également un aspect important du projet. En effet, l'objectif de ce projet est d'apporter un outil à l'ASIT-VD permettant d'augmenter l'engouement existant autour de la plateforme. Par exemple, la motivation d'un fournisseur pour diffuser de nouveaux produits sera plus grande s'il peut visualiser l'évolution du nombre de produits diffusés ou l'évolution des commandes dans le temps.

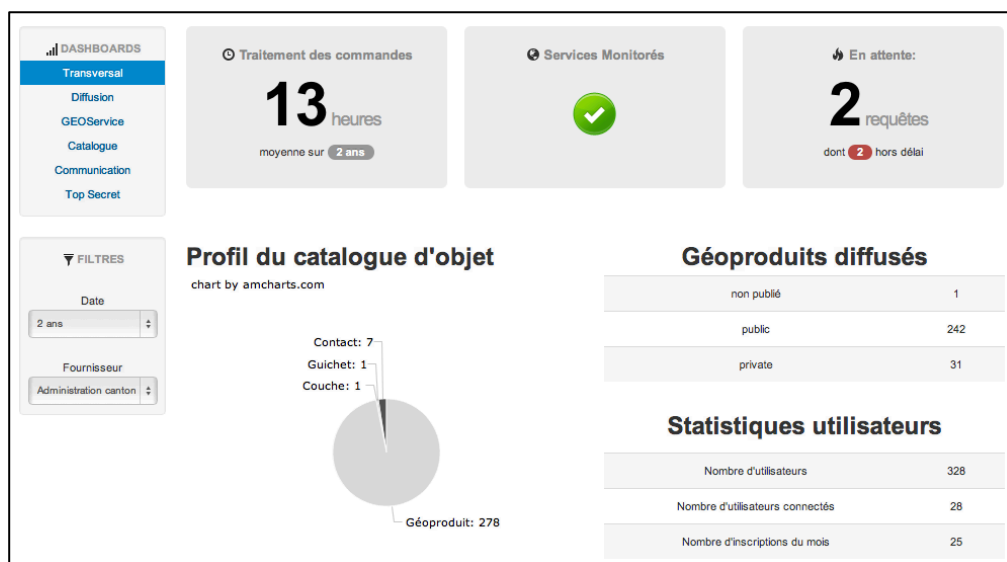


Illustration 3 - Aperçu du tableau de bord

Livrables

Les livrables de ce travail de Bachelor sont :

- Un entrepôt de données
- Un script ETL Talend pour l'alimentation de l'entrepôt de données
- Une couche d'accès Talend exporté sous la forme d'un web service
- Un composant permettant la création de tableaux de bord pour Joomla.
- Un manuel d'utilisation

Structure du rapport

La première partie de ce rapport traite l'approche théorique de la mise en place d'un environnement de Business Intelligence. Elle se veut applicable dans n'importe quel environnement et ne mentionne donc pas de solutions logicielles.

La deuxième partie présente le résultat obtenu, la mise en application de la théorie et les arguments qui ont motivé le choix des différents outils.

1 Approche théorique de la réalisation de tableaux de bord

Le titre de ce travail de Bachelor peut être réducteur. En effet, la réalisation d'un tableau de bord n'est finalement que la pointe de l'iceberg, le résultat désiré. Cependant, ce projet inclut également la mise en place de solutions permettant d'accompagner le processus des données sources aux différentes sorties. Ce processus porte un nom : **la Business Intelligence**

Une solution complète de Business Intelligence permet de traiter des données brutes émanant généralement de différentes sources afin d'en extraire des informations pertinentes pour l'organisation. Ce dernier permet par exemple de générer des tableaux de bord et des rapports. Ci-dessous un schéma représente le concept d'une solution de Business Intelligence. On distingue cinq étapes indépendantes :

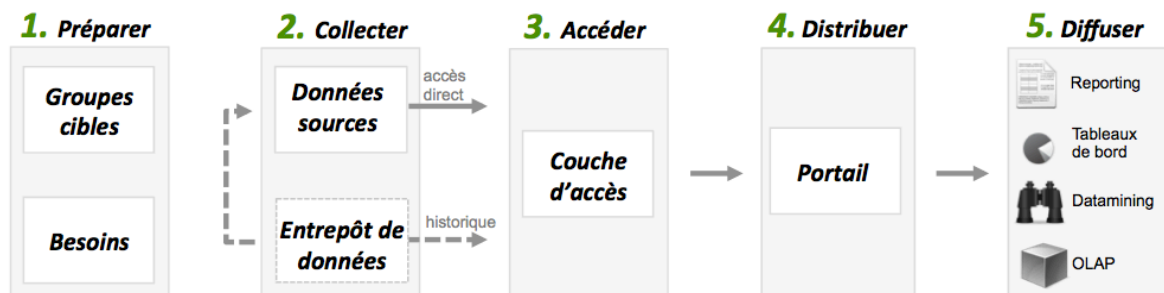


Illustration 4 - BI : les étapes théoriques

(Source : création personnelle inspiré par <http://www.osbi.fr/wp-content/bi.png>)

La première étape consiste à la mise en place d'une stratégie et à l'identification du public cible (1). Ensuite la collecte des données sources permet de définir la structure du datawarehouse et la façon d'accéder aux données (2). Puis vient la couche d'accès qui permet de définir un point d'accès unifié (3). Finalement la partie visible de l'application est traitée avec la gestion du portail de distribution contenant les éléments de diffusion (4+5).

Afin de suivre une structure logique, la partie théorique de ce travail suit scrupuleusement ces cinq étapes. On commence donc avec l'étape numéro une : « préparer ». Celle-ci définit les groupes cibles et les besoins.

1.1 Préparer

1. Préparer

2. Collecter

3. Accéder

4. Distribuer

5. Diffuser

La première étape de mise en place d'un outil de Business Intelligence débute par une opération non technique qui consiste à définir : le profil des utilisateurs, leurs besoins et le type de représentation des données. C'est donc une étape en amont permettant de préparer le terrain afin de mettre en place une solution technique pertinente répondant aux besoins métiers.

On commence par identifier les utilisateurs potentiels. Ceux-ci sont ensuite regroupés par groupe grâce aux besoins.

Identification du public cible

C'est un point clé pour le succès d'un projet de Business Intelligence qui consiste à lister les personnes susceptibles d'utiliser le système à mettre en place. Il est important de prendre du recul : les utilisateurs potentiels ne sont pas nécessairement tous internes à l'organisation. Un outil de Business Intelligence peut également servir d'outil de communication auprès de la clientèle permettant de clarifier les domaines d'activités ou encore être un outil permettant une meilleure coordination avec les partenaires qui font donc, de ce fait, également partie du public cible.

Identification des besoins

Une fois que les acteurs ont été identifiés, l'objectif est de procéder à l'identification de groupes de personnes possédant des besoins similaires. Les domaines d'activités constituent un bon point de départ pour procéder au regroupement. Ensuite, une discussion avec des intervenants de chaque profil permet d'établir une liste de besoins. Cette liste devient ensuite le fil conducteur du projet.

KPI Wheel

Afin de mener à bien la discussion avec les intervenants du projet, il existe un outil très intéressant mis en place par la société de consulting Bright Point nommé la KPI Wheel. Il s'agit d'une roue constituée de quatre questions à ne pas oublier :



Copyright (C) BrightPoint Consulting, Inc. 2005

Illustration 5 – KPI Wheel

(Source : <http://www.brightpointinc.com/>)

Les 4 points de la « KPI Wheel »

1. À quelle question veut-on répondre ?
 - Quel GEOProduit est-il le plus commandé ?
2. Qui cela concerne-t-il ?
 - Les fournisseurs de GEOProduit
3. Pourquoi est-ce important ?
 - Car les produits les plus importants doivent être en « mode automatique »
4. Quelles source(s) de donnée(s) ?
 - La base de données EasySdi

Cet outil permet de s'assurer qu'aucun élément ne soit oublié lors des entretiens permettant d'identifier les besoins. Il prend la forme d'une roue pour indiquer que les réponses n'arrivent pas forcément dans l'ordre. Cela donne ainsi lieu à une discussion ouverte plus propice à des échanges fructueux.

Résultats attendus de la phase de préparation

La phase de préparation est cruciale pour le succès de la mise en place d'un système de Business Intelligence. En effet, la grille des besoins résultant de ce travail est utilisée activement durant tout le projet et, bien que les besoins puissent évoluer durant le projet, cette grille initiale permet de faire des choix stratégiques et technologiques. Par exemple, une grille efficace permet de connaître les principales sources de données, les intervenants et le type de rendu désiré. Il est alors plus facile de sélectionner les outils et les stratégies à mettre en place. La grille des besoins complète est visible en annexe I.

1.2 Collecter

1. Préparer**2. Collecter****3. Accéder****4. Distribuer****5. Diffuser**

La grille des besoins créée durant l'étape précédente permet d'identifier les différentes sources de données possibles. L'étape actuelle « Collecter » consiste à lister ces sources d'informations et à choisir une méthode pour y accéder. Deux choix sont possibles : par un accès direct ou via un datawarehouse. Le schéma ci-dessous illustre ces deux méthodes :

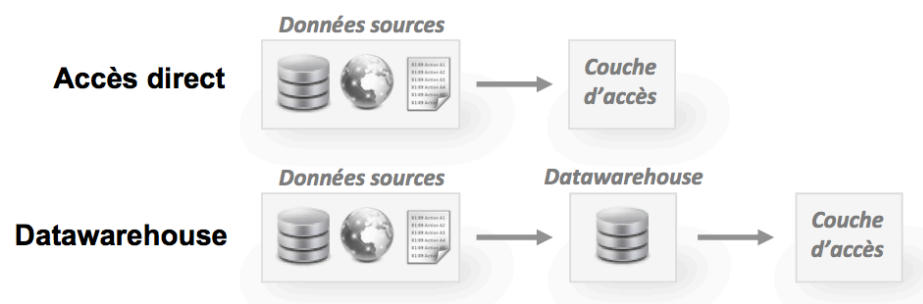


Illustration 6 - Comparaison accès direct et accès par datawarehouse

Un accès direct implique qu'on effectue la requête directement sur le fichier source. La deuxième possibilité est de passer par l'intermédiaire d'un datawarehouse.

Accès direct

Un accès direct permet d'avoir un résultat temps réel tout en évitant la redondance d'informations qu'implique un entrepôt de données. Toutefois, un accès direct n'est pas possible avec toutes les sources de données. Si le temps de traitement est trop lent ou si la source ne gère pas l'historisation des données, il est nécessaire de passer par un datawarehouse.

Datawarehouse

Egalement appelé entrepôt de données, le rôle d'un Datawarehouse est de stocker, à des intervalles définis, des informations prétraitées. Un outil de type ETL (Extract Transform and Load) permet d'effectuer ce travail de prétraitement. Celui-ci extrait puis traite les données afin de les insérer dans une nouvelle base sous forme homogène.

Un datawarehouse diffère en plusieurs points d'une base de données traditionnelle. Les approches top-down et bottom-up et les types d'architecture en sont des exemples. Les chapitres suivants précisent ces particularités.

Approches top-down et bottom-up

Deux visions de la structure architecturale d'un datawarehouse existent. Une approche dite « Top-down » et une approche « Bottom-up ». L'approche top-down stipule qu'il faut d'abord lister les besoins puis structurer le datawarehouse en conséquence. Au contraire, l'approche bottom-up préconise de lister toutes les possibilités puis de construire l'entrepôt en conséquence. La stratégie Bottom-up offre l'avantage d'être plus rapide à mettre en place (car on a pas besoin de lister toutes les possibilités) mais elle offre alors une structure moins flexible pour l'ajout de nouvelles fonctionnalités. Dans la pratique, la démarche est généralement hybride.

Architectures d'un datawarehouse

La particularité d'un datawarehouse est d'identifier des tables de faits qui sont liées à des tables de dimensions. Ce schéma permet un accès rapide aux données et simplifie grandement la construction des requêtes.

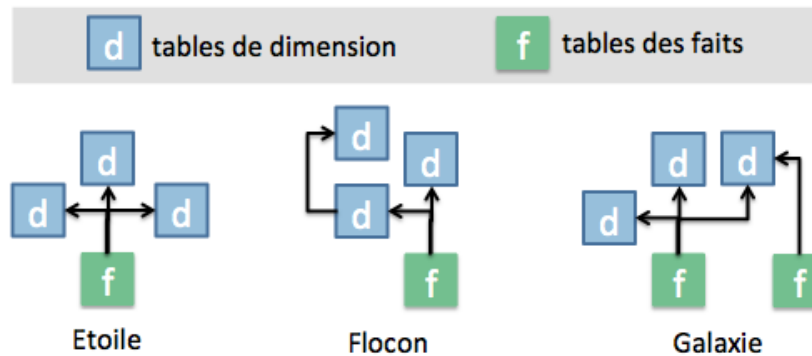


Illustration 7 - Datawarehouse: types d'architectures

(Source : création personnelle inspirée par <http://www.folkstalk.com/2010/01/data-warehouse-dimensional-modelling.html>)

Comme le démontre l'illustration ci-dessus, plusieurs architectures sont possibles. Le schéma en étoile est le plus connu. Il consiste à lier une ou plusieurs dimensions à un fait. Le schéma en flocon se différencie par le fait qu'il accepte que les dimensions soient liées entre elles. Ça offre une plus grande flexibilité si les liens entre les faits et les dimensions sont complexes. Le schéma galaxie indique que les dimensions sont partagées entre plusieurs faits. Ceci réduit le nombre de données et permet éventuellement de corrélérer des faits entre eux (puisque'ils sont liés à la même dimension).

Alimentation du datawarehouse

Il s'agit de mettre à jours les informations dans l'entrepôt de données. Cette opération peut se faire à différentes fréquences : chaque heure, quotidiennement ou même mensuellement. Pour connaître le temps idéal il faut faire un compromis entre les ressources machines à disposition et le besoin en données « fraîches ».

L'alimentation est assurée par un outil dit d'ETL (pour Extract Transform Load). Il permet de puiser l'information dans des sources diverses, éventuellement de les traiter et/ou de les corrélérer à d'autres données puis de les insérer dans une nouvelle base. Ce type d'outil permet de prétraiter l'information.

Accès direct vs. Datawarehouse

Comment savoir si il faut accéder directement à la base de production ou passer par un datawarehouse ? Ce sont des choix stratégiques qui diffèrent selon l'environnement de la société. Voici quelques pistes pour faire un choix.

Le tableau ci-dessous résume les avantages et inconvénients de chacune des méthodes :

	Accès direct	Datawarehouse
Temps traitement de la requête	Suivant le type de fichier source, le temps d'exécution peut être long. Par exemple, il est impensable de s'attaquer directement à un fichier de log de plusieurs MO à cause du temps de traitement que cela engendrerait.	L'information est prétraitée ce qui réduit considérablement le temps de traitement.
Effort de maintenance	Faible, si la structure d'une table change il suffit de modifier une seule requête.	Fort, si la structure change il faut repenser l'architecture du datawarehouse.
Péréemption des données	Données toujours à jour.	La requête reflète le résultat à la date de la dernière mise à jour.

Une société possédant un grand nombre de données et des ressources financières suffisantes pour supporter la maintenance du datawarehouse préfère sans doute tout stocker dans ce dernier dans un souci d'uniformité et de performance. Cependant, l'alimentation en données d'un entrepôt implique un coût de maintenance homme-machine élevé. Une solution hybride est donc parfois préférable. La partie pratique de ce rapport compare les performances entre ces deux accès (p.27 – « Comparaison accès direct - datawarehouse »).

1.3 Accéder

1. Préparer**2. Collecter****3. Accéder****4. Distribuer****5. Diffuser**

Après avoir lister les besoins des différents groupes d'utilisateurs et collecter les données sources nécessaires, il est temps d'accéder à ces données. Suivant les choix effectués précédemment, l'information peut se trouver dans une base de production, un entrepôt de données ou encore par le biais de web services.

Architecture trois tiers

Afin de pouvoir construire un projet qui soit modulaire et évolutif, il est important de séparer l'application en couches. Le modèle logique « architecture 3-tiers » sépare l'application ainsi : couche de présentation, couche métier et couche d'accès aux données. L'illustration de cette page représente cette architecture.

Dans une architecture de type 3-tiers, chaque couche ne communique qu'avec ces voisins immédiats. La couche de présentation n'a donc aucun contact avec la couche d'accès aux données. L'avantage est clair : en cas de modification des sources de données, les rapports, tableaux de bord et autres éléments de diffusion ne subissent aucun changement. Ceci permet ainsi une grande souplesse pour l'ajout de toutes nouvelles technologies.

Cette séparation en couche permet également d'homogénéiser la solution en offrant une porte d'entrée unifiée aux ressources. Ceci permet de réduire la maintenance car toutes interactions avec les données se font au niveau de la couche d'accès.

Cette flexibilité est également vraie pour l'allocation des ressources machines. En effet, grâce à l'indépendance des couches, il est possible de modifier l'allocation des différentes couches en fonction des besoins évoluant tout au long du cycle du projet.

L'étape 3 « Accéder » correspond donc à cette « couche d'accès aux données » du modèle trois tiers. Les deux prochaines étapes « Distribuer » et « Diffuser » correspondent respectivement aux deux niveaux suivants de l'architecture trois tiers « couche métier » et « couche de présentation ».

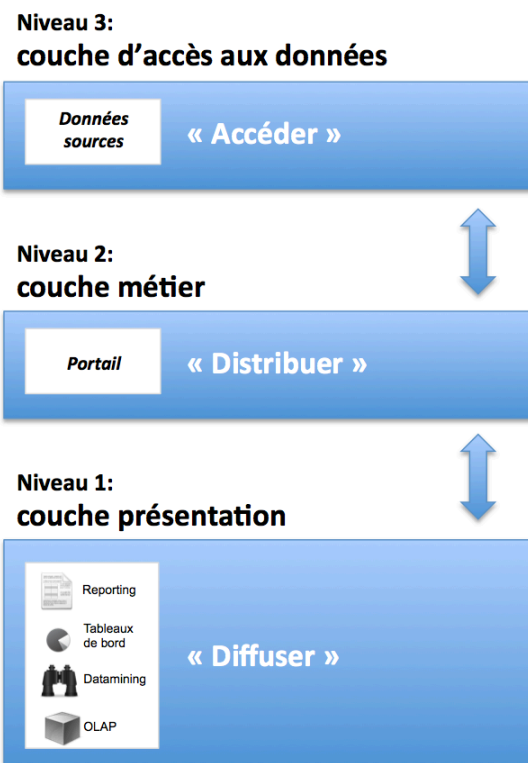


Illustration 8 - Architecture 3-tiers

1.4 Distribuer

1. Préparer

2. Collecter

3. Accéder

4. Distribuer

5. Diffuser

L'étape de distribution et celle qui fait appel à la porte d'entrée créée grâce à la couche d'accès (chapitre précédant). Cette nouvelle couche est responsable des opérations logiques de distribution des informations. C'est la couche métier. Il s'agit d'offrir une interface permettant à une personne de s'authentifier afin d'avoir accès à un environnement qui lui correspond. Son rôle est de banaliser l'accès à la Business Intelligence.

Portail d'entreprise

Egalement abrégé EIP (Enterprise Information Portal), un portail d'entreprise permet d'offrir un point d'accès sécurisé à différentes ressources informatiques mises à disposition par l'entreprise. Ce point d'accès se présente généralement sous la forme d'une application web. Le portail est ainsi facilement accessible à l'aide d'un navigateur Internet. Le tableau ci-dessous résume les différentes fonctionnalités standard offertes par un portail d'entreprise.

Fonctionnalités classiques d'un EIP:

Authentification unique	Permet à un utilisateur de s'authentifier une seule fois pour accéder à de multiples ressources
Intégration	Permet de réunir et de lier des fonctionnalités diverses mises à disposition par l'entreprise.
Fédération	Permet l'intégration de contenu provenant d'autres portails
Personnalisation	Concerne le « look and feel ». Permet également à un utilisateur de mettre en avant les éléments les plus utilisés.
Access Control	Permet de définir des droits et des privilèges pour certains groupes sur des ressources.
Recherche	Offre la possibilité d'effectuer des recherches dans le portail.

Tableau largement inspiré par l'article suivant :
http://en.wikipedia.org/wiki/Enterprise_information_portal

Ces fonctionnalités permettent d'affirmer qu'un portail d'entreprise est un parfait candidat pour jouer le rôle de la distribution des données. La fonctionnalité essentielle est l'« Access Control ». En effet, celle-ci permet de définir des rôles et des permissions qui sont en accord avec la stratégie mise en place durant l'étape 1 : Préparer. On peut ainsi personnaliser l'affichage des tableaux de bord en fonction du rôle de l'utilisateur.

Business Intelligence de deuxième génération

« Afin de dynamiser la réactivité globale, l'information sera largement distribuée auprès de l'ensemble des partenaires. Le portail décisionnel, EIP Enterprise Information Portal, remplit cette fonction essentielle. Désormais, l'essor de la Business Intelligence de 2^{de} génération, la BI 2.0 exploitant pleinement les capacités du web 2.0 pour l'entreprise, démocratise très largement l'accès à l'information décisionnelle. »

Texte intégral : <http://www.piloter.org/business-intelligence/business-intelligence.htm>

La portion d'article ci-dessus colle parfaitement au projet. En effet, les nouvelles technologies web améliorent considérablement les interactions avec les utilisateurs et ceux-ci doivent être utilisées à bon escient afin d'offrir une expérience utilisateur satisfaisante. Les utilisateurs sont maintenant habitués à avoir des informations fluides et dynamiques. Le portail doit donc pouvoir répondre à ces nouvelles exigences.

1.5 Diffuser

1. Préparer

2. Collecter

3. Accéder

4. Distribuer

5. Diffuser

La dernière étape est la génération de graphiques, rapports ou éléments permettant d'analyser les fondations qui ont été bâties durant les quatre premières étapes du projet. On distingue généralement quatre outils permettant l'analyse:

Tableaux de bord :

Les tableaux de bord désignent un affichage virtuel synoptique permettant de rendre compte de la situation de l'ensemble ou d'un domaine d'une entreprise. Il est généralement composé de graphiques et de données empiriques. Grâce à l'évolution des technologies web, il est possible de créer des tableaux de bord permettant à l'utilisateur d'interagir avec les éléments qui le composent.

Reporting :

Les rapports générés sont en principe destinés à l'impression ou à de l'archivage. A la différence des tableaux de bord, les données sont généralement plus fournies.

Datamining :

Il s'agit d'une opération avancée sur un datawarehouse permettant d'analyser automatiquement une grande quantité de données. Elle permet de croiser celles-ci et de détecter les corrélations entre elles.

OLAP :

Le terme Online Analytical Processing désigne une technique permettant d'analyser les données sous forme de cube. L'avantage est d'offrir un accès multidimensionnel aux données sans pour autant nécessiter de connaissances spécifiques dans la manipulation des bases de données. L'utilisateur peut ainsi effectuer une analyse interactive selon ses propres critères.

1.6 Choix des outils

Cette partie théorique n'a volontairement pas évoqué de solution logicielle car c'est un élément qui dépend très fortement de l'architecture déjà en place, des moyens financiers à disposition ainsi que des ressources et compétences humaines et cette introduction théorique se veut applicable dans tout type d'environnements.

Cependant, le fameux institut de conseil Gartner a récemment publié une comparaison entre les principaux outils de Business Intelligence. Cette comparaison énumère 21 solutions. Certaines couvrent l'ensemble des cinq étapes mentionnées et d'autres sont spécifiques à une ou deux étapes. Si un choix doit être effectué, je pense que c'est un article indispensable :

<http://www.gartner.com/technology/reprints.do?id=1-196WFCB&ct=120207>

*Etude sur les solution de BI parue en février 2012
Source : www.gartner.com*

1.7 Conclusion de la partie théorique

La découpe en étape est une grande aide pour la mise en place d'une structure de Business Intelligence. Elle permet de séparer le travail en étape logique et de suivre une évolution des données. En effet, ces étapes permettent de faire la transition entre un simple besoin évoqué lors d'une discussion avec un membre du public cible et une information clé sur un tableau de bord permettant de piloter une société.

Cependant, même si les étapes sont suivies scrupuleusement, la mise en place d'un outil de Business Intelligence se fait toujours de façon itérative. Par exemple, bien que la KPI Wheel soit un excellent outil pour accompagner une discussion, il est fréquent que la mise en place d'un élément permettant de répondre à une attente face émerger de nouveaux besoins. Il est donc conseillé d'utiliser une structure modulaire et évolutive afin qu'au fil des années l'architecture du système de Business Intelligence s'améliore et s'adapte aisément à de nouveaux besoins et à un nouvel environnement.

2 Réalisation du TB

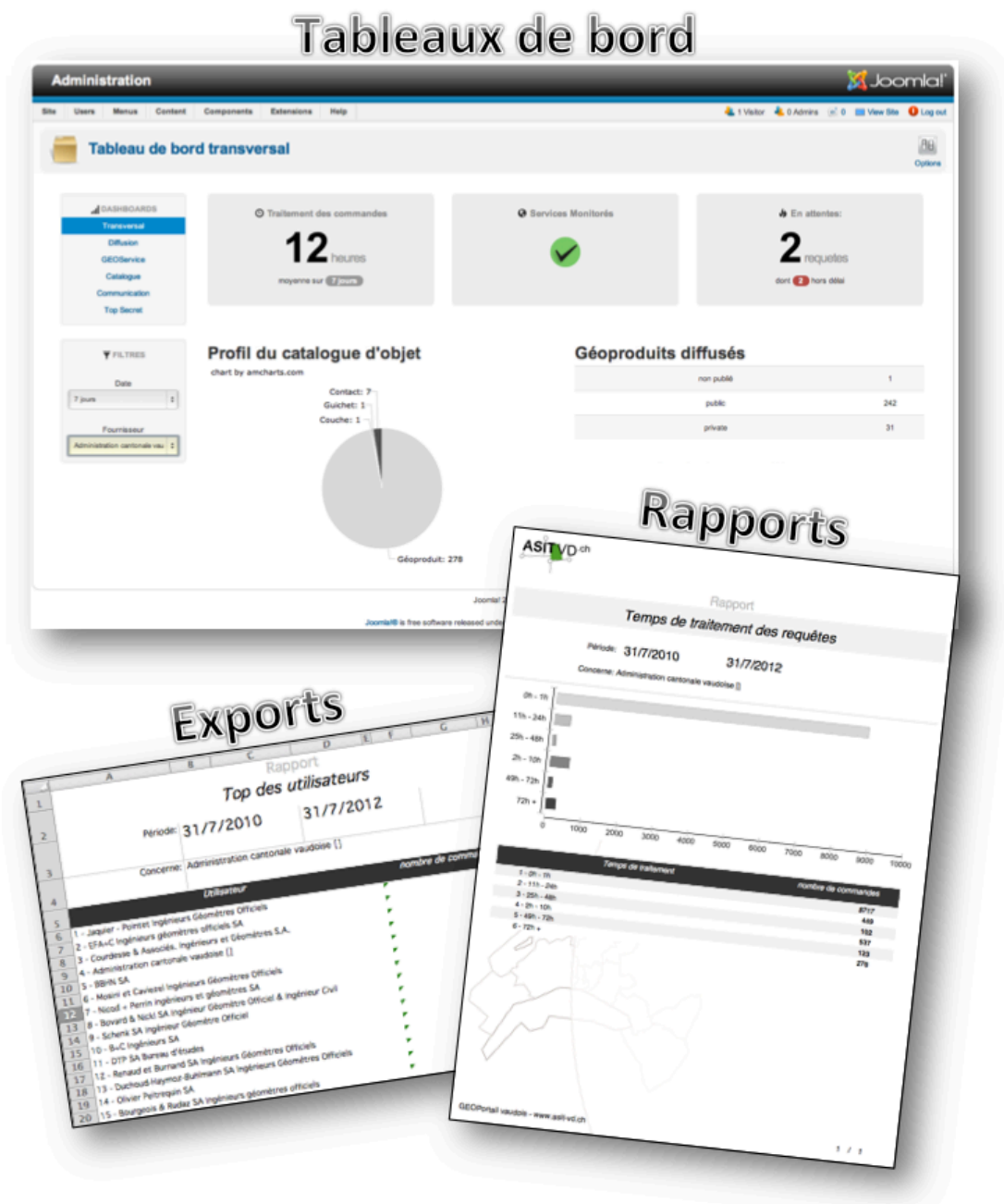


Illustration 9 - aperçu du résultat

Le produit livré est un composant Joomla intégrant différents tableaux de bord regroupés par thématiques. Les tableaux de bord contiennent des graphiques ainsi que des données empiriques et donnent également accès à des rapports dynamiques dont le format de sortie est paramétrable.

2.1 Contraintes techniques imposées

Avant de détailler la réalisation du projet, il est important de mentionner les différentes contraintes techniques de l'environnement de l'ASIT-VD. Celles-ci ont aiguillé les choix qui ont été effectués concernant les outils à mettre en place. Voici les principales contraintes:

Intégration dans Joomla

La première contrainte définit le portail de distribution : Joomla. En effet, Joomla étant déjà en place, c'est sur ce portail que le composant est développé. La version imposée est Joomla 2.5.

Future intégration à EasySDI

Présentation d'EasySDI

EasySDI est le composant Joomla utilisé pour gérer le GEOPortail. C'est un projet OpenSource. L'Association pour le Système d'Information du Territoire Vaudois fait partie du comité de pilotage et est un grand contributeur au projet. Voici d'autres exemples d'administrations utilisant EasySDI :

- Administration du Cadastre et de la Topographie - Luxembourg
- MNHN (Musée National d'Histoire Naturelle de Luxembourg) - Luxembourg
- SIG Pyrénées (Assemble Pyrénéenne d'Economie Montagnarde) - France
- Canton de Berne - Suisse
- CRIGEOS - France
- OPCC - France – Espagne

« EasySDI est une solution simple et prête à l'emploi pour déployer une infrastructure spatiale de données spatiales (IDS), interopérable et basée sur les standards ISO/OGC. Son objectif est de répondre à la directive INSPIRE et aux besoins de toute organisation souhaitant diffuser de l'information géographique. »

Texte intégral easysdi.org

Intégration et contraintes

L'intégration du projet de tableaux de bord à EasySDI est un objectif à long terme. En effet, lorsque le projet de tableaux de bord aura atteint une maturité suffisante, celui-ci pourra être fusionné avec le projet EasySDI afin d'être diffusé à plus grande échelle (esprit Open Source).

Cette future intégration est d'ores et déjà prise en compte lors du choix des outils afin de sélectionner le maximum d'outil Open Source permettant une redistribution totale du code.

Responsive Design

C'est une des exigences du mandataire : le design doit permettre une utilisation sur différents types d'appareils (ordinateurs, mobiles) et avec différentes résolutions. Le projet doit également utiliser les dernières technologies web permettant une interactivité optimum avec l'utilisateur.

Utiliser aux maximums les technologies en place

Actuellement les technologies principalement utilisées sont Java, PHP et Mysql. L'utilisation de ces technologies est donc, si possible, préférée.

2.2 Démarche:

Ce chapitre du rapport décrit les actions effectives qui ont été entreprises afin de mener à bien ce projet. Les étapes suivent scrupuleusement l'approche théorique décrite en première partie. L'illustration ci-dessous permet de visualiser la transition entre les deux parties :

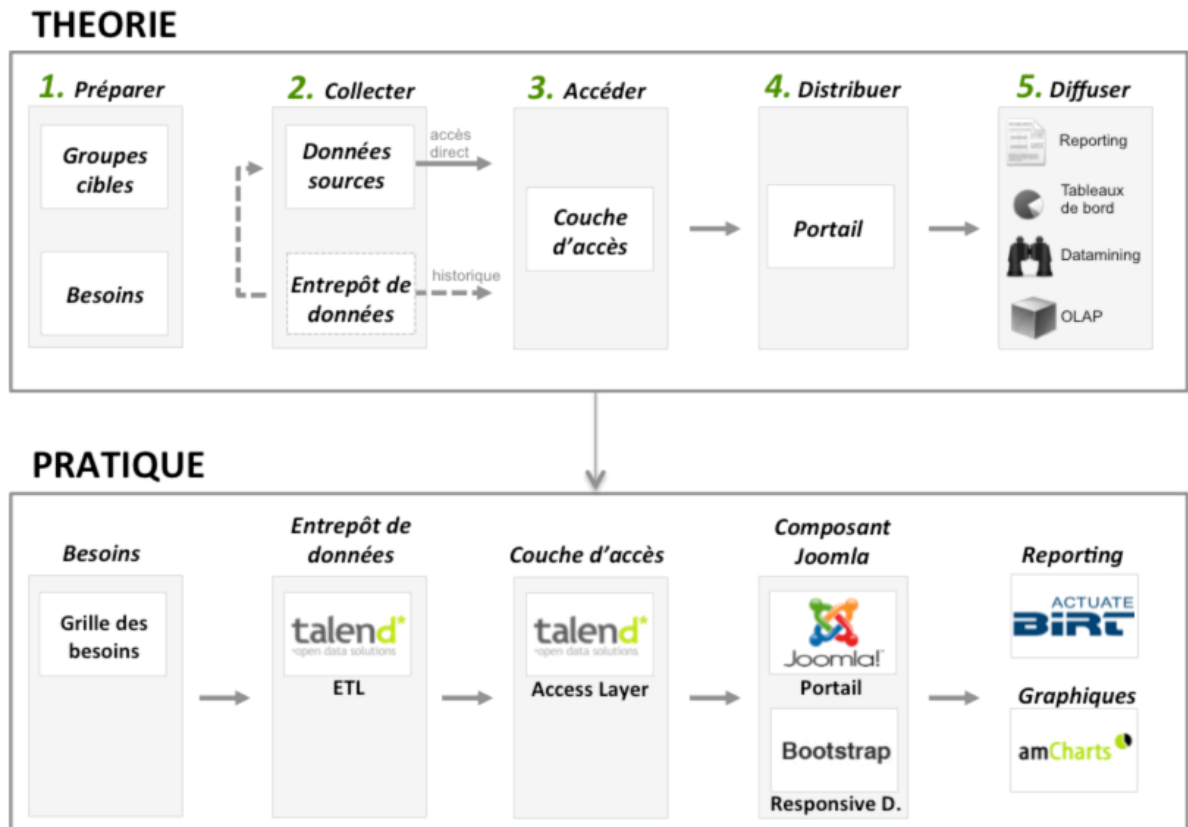


Illustration 10 - Transition entre théorie et pratique

On remarque l'introduction d'outils tels que : Talend, Joomla, Bootstrap, Birt et AmCharts. Ce sont des outils qui ont été sélectionnés en tenant compte des contraintes exprimées durant le chapitre précédent. Ces choix ont été effectués durant le travail préliminaire.

Le prochain chapitre traite de la première étape de l'élaboration d'un outil de Business Intelligence : l'élaboration de la grille des besoins.

2.3 Besoins

Besoins	<i>Entrepôt de données</i>	<i>Couche d'accès</i>	<i>Composant Joomla</i>	<i>Reporting Graphiques</i>
----------------	----------------------------	-----------------------	-------------------------	-----------------------------

L'élaboration de la liste des besoins a été effectuée durant le temps imparti pour le travail préliminaire. Elle a été réalisée dans les locaux de l'ASIT-VD afin de pouvoir s'imprégner de l'univers de l'association et ainsi bien comprendre son fonctionnement. Le fruit de cette recherche a permis de définir:

- Le public cible et les groupes d'utilisateurs
- La stratégie de découpe des tableaux de bord par domaines d'activités
- La liste des besoins

Public cible et groupes d'utilisateurs

Quatre groupes d'utilisateurs sont identifiés. La colonne priorité est définie par le mandataire permettant de préciser avec quels groupes mettre l'accent.

Description du groupe	Taille actuelle	Profils	Rôles / Activités	Priorité
Administrateurs <i>Equipe technique de l'association</i>	3 collaborateurs	Ingénieurs	Coordination, support, formation, maintenance et développement	Haute
Fournisseurs <i>Utilise la plateforme de L'ASIT-VD pour la diffusion de géodonnées</i>	19 entreprises	Multiples	Traitement des commandes, maintenance des données fournies	Haute
Grand public <i>Toute personne non authentifiée qui visite la plateforme</i>	-	Multiples	Recherche d'informations sur l'activité du portail	Moyen
Utilisateurs <i>Utilisateurs enregistrés de la plateforme</i>	940 membres	Architectes Techniciens	Consommateurs	<i>Faible</i>

Stratégie de découpe des tableaux de bord

Le schéma ci-dessous permet de démontrer que la stratégie est définie avant toute activité technique. La phase de planification d'un projet de Business Intelligence est primordiale : c'est elle qui guide les futures choix techniques.

On retrouve sur le schéma ci-dessous trois groupes d'utilisateurs. Des flèches permettent d'illustrer à qui sont destinés les tableaux de bord.

Les administrateurs et les fournisseurs partagent le même contenu sur la partie back-end. La différence réside dans le fait qu'un administrateur peut prendre la vue d'un fournisseur particulier alors que ce dernier n'a accès qu'à sa propre vue. Le grand public a accès à la partie front-end sur un tableau de bord permettant à l'ASIT-VD de faire de la communication sur les activités exercées. Le groupe des utilisateurs a été mis de côté durant le travail préparatoire. C'est le groupe dont la priorité est la plus faible pour l'accès à un tableau de bord.

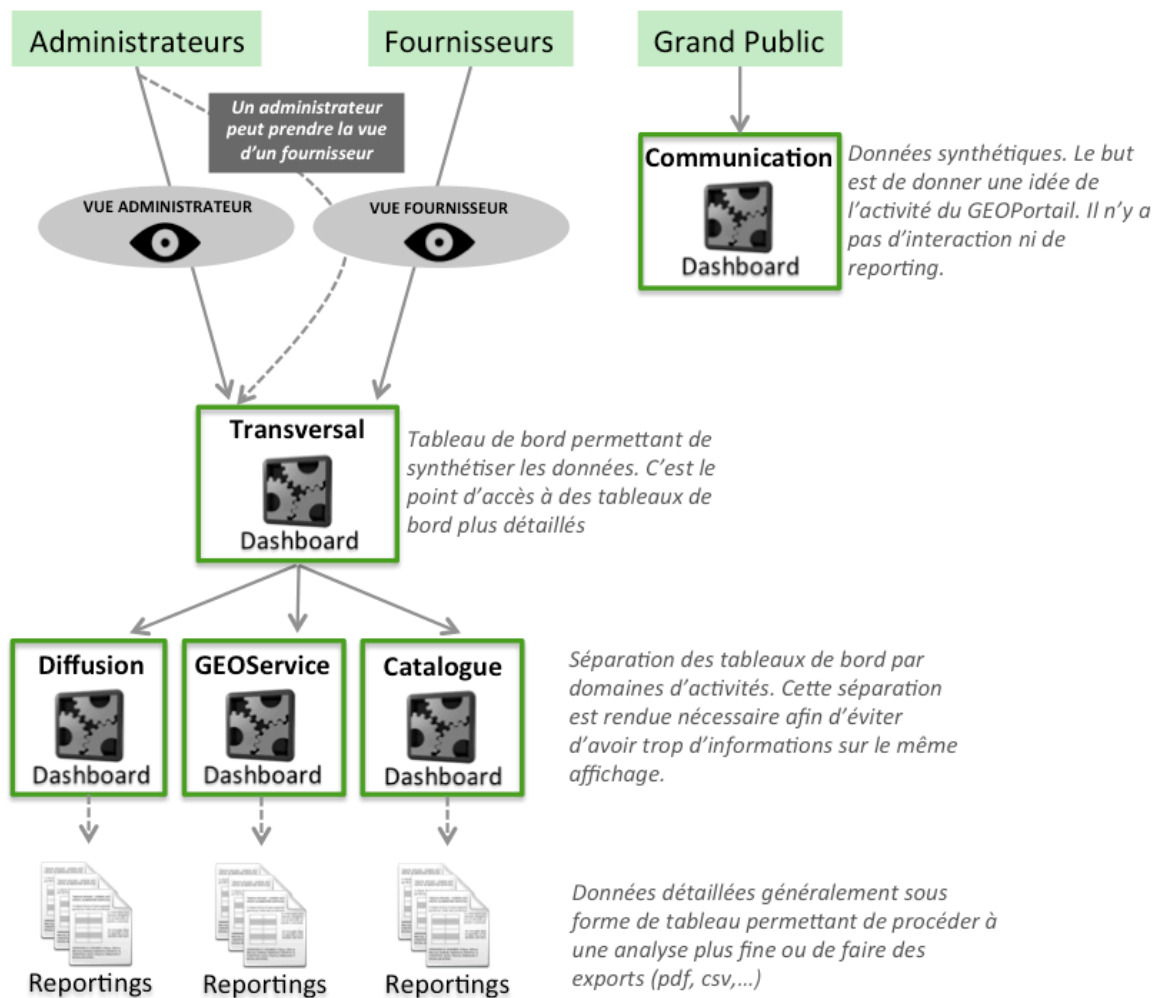


Illustration 11 - Stratégie de découpe des tableaux de bord







Besoins

Après avoir défini les groupes d'utilisateurs et la stratégie de découpe des tableaux de bord, il est temps de s'attaquer aux besoins des différents intervenants.

Stratégie pour la définition des besoins










La stratégie adaptée a été de définir des outils d'aide à la compréhension des besoins. Ces outils consistent en définition de termes permettant d'engager une discussion avec une terminologie prédéfinie. Ceci permet principalement d'éviter les malentendus. La discussion a été menée avec les différents membres de l'association et également avec M. Dinarco Gouveia, responsable de la diffusion des données pour l'OIT (Office de l'information sur le territoire pour le canton de Vaud, fournisseur le plus actif de la plateforme).










Tableau d'aide à la compréhension des besoins

	Terme	Définition
Temps	Information ponctuelle	Données actuelles <i>Exemple : maintenant !</i>
	Range prédéfini	Données depuis une date fixée à maintenant <i>Exemple : 30 derniers jours</i>
	Range paramétrable	Identique au range prédéfini mais la date de début est paramétrable
	Historique	Choix libres des dates
	Night sync.	La donnée se met à jour la nuit via le datawarehouse
	Live	La donnée est puisée directement dans la base de production
Sources	Log file	Fichier texte généré par le serveur (par ex WMS)
	DB	Base de données (principalement Joomla et Redmine)
	Web Service	Services web SOAP et REST (service monitor, Google Analytics)
Support de sortie	Dashboard	Représentation graphique dont l'objectif est de synthétiser des données brutes pour en ressortir des informations synoptiques. On l'utilise pour donner une vision globale de la situation d'un domaine de l'entreprise.
	Reporting	Un export est plus adapté pour afficher des détails que n'affiche pas un Dashboard. Son format est destiné à de l'archivage ou à de l'impression. Le fichier n'a pas besoin d'être synoptique et peut donc comporter un plus grand nombre d'informations
Données de sortie	Pie	 Permet de visualiser la part de chaque secteur
	Plot	 Son but est de comparer la croissance individuelle d'une série
	Line	 Montre l'évolution au cours du temps
	Gauge	 Permet de contextualiser une valeur en donnant la limite de celle-ci et éventuellement les zones critiques
	Map	 Permet de localiser des valeurs géographiques
	Grid	 Permet de voir en détail les valeurs de l'échantillon. Il peut aussi s'agir d'une simple valeur numérique

Liste des besoins

La liste des besoins est trop conséquente pour être intégrée au rapport. Elle est par contre disponible en annexe I. Cependant, afin de pouvoir mieux comprendre le contexte de l'application, voici quelques extraits de besoins identifiés.

Package	Dashboard sous-jacent	Besoins exprimés	Temps	Source	Output
Transversal  Dashboard Administrateurs Fournisseurs	Diffusion	Temps moyen pour le traitement d'une commande <i>30 derniers jours</i>	 range prédefini  live	 MySQL	
	Géoservice	Connaître l'état des alertes sur les services monitorés	 information ponctuelle  live	 Web service	

Package	Besoins exprimés	Temps	Source	Output
Diffusion  Dashboard Administrateurs Fournisseurs	Nombre de requêtes (manuel/auto) dont le traitement a excédé 48h00	 range paramétrable  night synch.	 MySQL	
	Avoir un top 5 des utilisateurs déposant le plus de requêtes.	 range paramétrable  night synch.	 MySQL	 Reporting

Résumé de la grille des besoins

Nom du tableau de bord	Public cible	Nombre de besoins identifiés	Nombre de rapports
Transversal	Administrateurs, Fournisseurs	12	0
Diffusion	Administrateurs, Fournisseurs	8	5
GEOService	Administrateurs, Fournisseurs	4	3
Catalogue	Administrateurs, Fournisseurs	5	1
Communication	Grand public	9	0

L'objectif du projet n'est pas de pouvoir répondre aux 38 besoins identifiés. Le but est de mettre en place une structure avec suffisamment de cas concrets permettant ensuite à l'ASIT-VD une autonomie totale.

2.4 Structure de l'entrepôt de données

*Besoins***Entrepôt de
données***Couche d'accès**Composant
Joomla**Reporting
Graphiques*

L'entrepôt de données est séparé en deux chapitres. Le premier chapitre décrit la structure de l'entrepôt et le chapitre suivant traite de son alimentation.

Ce chapitre traite d'abord la conception de l'entrepôt et du choix qui a été fait concernant le schéma de modélisation. Ensuite, une comparaison des performances est faite entre un accès direct aux données de la base de production et un accès par un datawarehouse. Finalement la dernière partie de ce chapitre définit la taille limite que peut atteindre un datawarehouse.

Conception de l'entrepôt de données

Le datawarehouse est basé sur une base de données Mysql. Ce choix s'explique car la base de production Joomla utilise également ce SGBD et cet élément faisait donc partie des contraintes fixées initialement. Le moteur utilisé est MyISAM. En effet, ce dernier est plus performant qu'InnoDB pour l'accès en lecture aux données¹.

Le schéma « galaxie » (réf: Architecture p.13) est utilisé pour l'entrepôt. L'avantage est de pouvoir partager certaines tables de dimensions et ainsi minimiser le code permettant la mise à jour de celles-ci.

¹ Les SELECT sont plus rapides. Source : http://www.mikebernat.com/blog/MySQL_-_InnoDB_vs_MyISAM

Voici le schéma de l'entrepôt actuel :

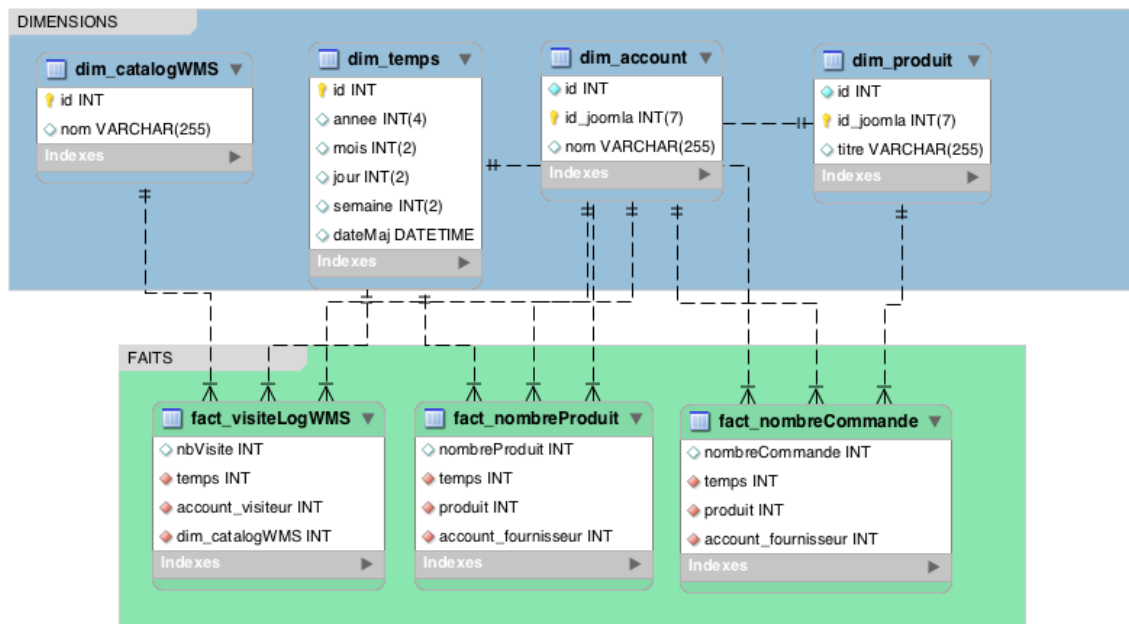


Illustration 12 - Schéma de l'entrepôt de données

Comparaison accès direct - datawarehouse

Est-il mieux d'accéder directement aux données de la base de production ou est-il préférable d'utiliser un datawarehouse ? La question est souvent revenue au cours du projet. Dans quel cas faut-il utiliser l'une ou l'autre solution ?

Parfois, le choix de l'entrepôt de données se fait naturellement. C'est le cas des fichiers de logs. Mentionnée durant le chapitre précédent, l'analyse d'un fichier de log ne prend que 1,7 seconde. Cependant, le serveur WMS crée un fichier quotidiennement. L'analyse des accès sur une année prend donc approximativement 10 minutes. Une opération qui prend moins d'un dixième de seconde lorsqu'on récupère les données prétraitées du datawarehouse.

Dans d'autres cas, le choix est moins évident. Afin d'illustrer ce choix difficile prenons un exemple d'une requête complexe ; on veut obtenir l'information suivante : « Le top 5 des produits de l'administration cantonale vaudoise pour l'année 2011 ».

COMPARAISON DU TYPE D'ACCES « Top 5 des produits de l'administration cantonale vaudoise pour 2011 »		
	Accès Direct	Accès par Datawarehouse
Alimentation du DWH	<p>Aucune action nécessaire</p>	<pre>INSERT INTO fact_nombreCommande (nombreCommande, temps, produit, account_fournisseur) (SELECT count(DISTINCT jos_sdi_order.id) as nbCommande, dim_temps.id, dim_produit.id, dim_account.id_joomla FROM easysdi2.jos_sdi_order #...10 JOINTURES WHERE jos_sdi_order.type_id = 1 AND YEAR(jos_sdi_order.created) = 2011 AND COALESCE(jos_users_with_root.id,jos_users.id) = 22388 GROUP BY jos_sdi_product.id, dim_account.id)</pre> <p style="text-align: center; font-size: 1.2em;">0.22 seconde <i>(moyenne sur 10 requêtes)</i></p>
Requête portail	<pre>SELECT count(DISTINCT jos_sdi_order.id) as nbCommande, jos_sdi_product.name FROM jos_sdi_order #...9 JOINTURES WHERE COALESCE(jos_users_with_root.id,jos_users.id) = 22388 AND jos_sdi_order.type_id = 1 AND jos_sdi_order.created BETWEEN '2011-01-01' AND '2011-12-31' GROUP BY jos_sdi_product.id ORDER BY nbCommande DESC LIMIT 0,5</pre> <p style="text-align: center; font-size: 1.2em;">0.18 seconde <i>(moyenne sur 10 requêtes)</i></p>	<pre>SELECT count(nombreCommande) as nbCommande, dim_produit.titre FROM fact_nombreCommande #...2 JOINTURES WHERE dim_temps.annee = 2011 GROUP BY dim_produit.id ORDER BY nbCommande DESC</pre> <p style="text-align: center; font-size: 1.2em;">0.02 seconde <i>(moyenne sur 10 requêtes)</i></p>

Le temps de réponse de la requête effectuée directement dans la base de production est tout à fait raisonnable. Le choix de l'ASIT-VD est de privilégier l'accès direct aux données donnant une vision « temps réel » du tableau de bord. En effet, l'avantage est de pouvoir suivre l'évolution de l'activité au cours de la journée. Un accès direct est également plus facile à maintenir car il s'agit d'une seule requête, sur une seule source de données.

Taille critique du datawarehouse

Combien d'enregistrements est-il possible d'effectuer dans le datawarehouse ? La question est légitime. Le meilleur moyen d'y répondre est de constater la taille actuelle de la base et de l'extrapoler. Actuellement, après 791 jours simulés, les données sont ainsi :

Table	Taille (Mo)	Nombre d'enregistrements	Poids moyen d'un enregistrement
dim_account	0.12	918	137 octets
dim_catalogWMS	0.01	2	5140 octets*
dim_produit	0.04	450	97 octets
dim_temps	0.03	791	42 octets
fact_nombreCommande	0.04	2466	17 octets
fact_nombreProduit	0.00	215	22 octets
fact_visiteLogWMS	0.03	1504	18 octets
<i>* faussé par le nombre trop faible d'enregistrements (la taille inclut la structure de la table)</i>			

On remarque que les trois tables de faits ont un poids inférieur par enregistrement. Ce n'est pas un hasard car ces tables contiennent uniquement des faits empiriques et des liens sur d'autres tables. Les tables de faits sont justement destinées à contenir un grand nombre d'enregistrements et c'est la raison pour laquelle la structure des datawarehouse est telle qu'elle est.

Basée sur les données générées en l'espace de 791 jours, voici une estimation de ce que serait la situation avec 15 tables de dimensions et 30 tables de faits. Le nombre d'enregistrements et le poids moyen d'un enregistrement sont choisis dans les maximums du tableau ci-dessus.

Scénario

Type	Nombre de tables	Nombre d'enregistrements	Poids moyen d'un enregistrement	Taille (791 jours)	Taille par année (Mo)
Table des dimensions	15	700	200 octets	2 mo	0.9mo*
Table des faits	30	5000	50 octets	7.1 mo	3.3 mo
<i>* Contrairement aux tables des faits, les tables de dimensions n'évoluent que très peu en fonction du temps</i>					

La taille limite avec le moteur MyIsam est de 8To¹. Ce n'est donc pas la limite de Mysql qui va être atteinte. Attention cependant à ne pas dépasser la taille limite d'un fichier imposé par le système d'exploitation. Toutefois, de nos jours, cette taille se situe généralement au delà de 2Go. Suffisant pour alimenter 15 tables de dimensions et 30 tables de faits durant 487 ans.

¹ Source : <http://dev.mysql.com/doc/refman/5.0/fr/table-size.html>

2.5 Alimentation de l'entrepôt de données

*Besoins***Entrepôt de
données***Couche d'accès**Composant
Joomla**Reporting
Graphiques*

Après avoir présenté la structure de l'entrepôt, l'objectif est maintenant d'alimenter la base de données. La partie théorique introduit dans le chapitre « Alimentation du datawarehouse » les outils de type ETL. C'est en effet un ETL qui a été mis en place pour jouer ce rôle. Il s'agit de Talend.

Ce chapitre a pour but d'expliquer pourquoi Talend a été choisi et de présenter le « job » qui a été réalisé avec.

Talend comme outil ETL

Après une analyse effectuée durant le travail préparatoire, le choix de l'outil ETL s'est porté sur Talend car c'est un outil open source, possédant une grande communauté active et avec lequel l'ASIT-VD a déjà de l'expérience.

Il a été préféré à la suite Pentaho car c'est un outil effectuant une tâche précise. A contrario Pentaho propose une solution couvrant l'ensemble des cinq étapes de la Business Intelligence. La première solution a été privilégiée car elle offre une solution plus flexible : chaque outil est indépendant. Plusieurs solutions Talend sont disponibles : celle qui nous intéresse se nomme « Talend Open Studio for Data Integration » (version 5.1.1). Il s'agit d'un environnement d'intégration graphique exécutant du code Java.

Présentation du job Talend

Ci-dessous se trouve le job Talend permettant d'alimenter le datawarehouse. Il est destiné à être lancé quotidiennement afin de populer les tables de faits et les tables de dimensions. On remarque aisément son orientation graphique.

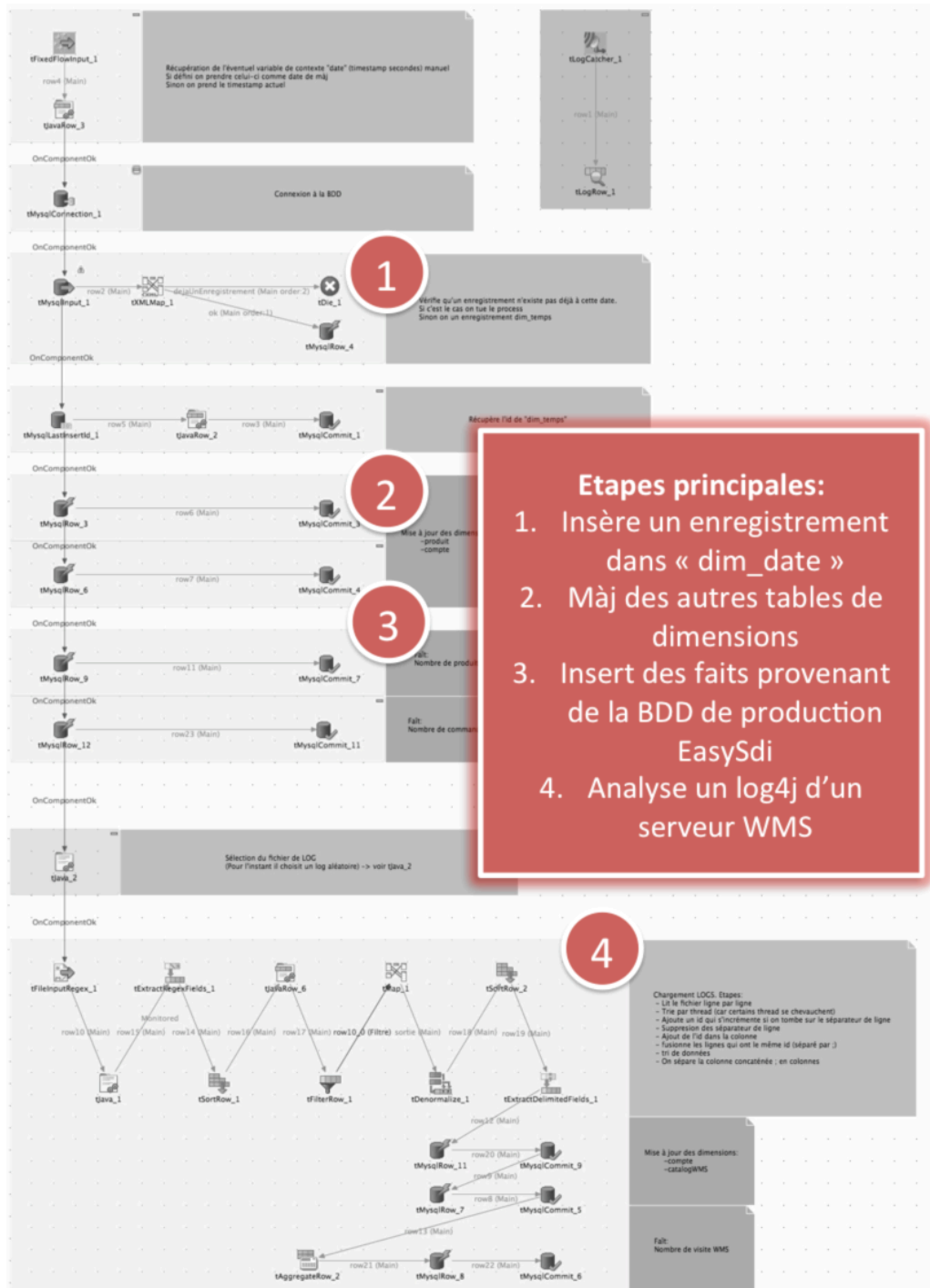


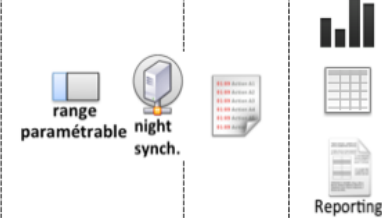


Illustration 13 - job Talend d'alimentation

Exemples concrets de besoins nécessitant l'entrepôt

La liste des besoins permet de connaître les demandes nécessitant un enregistrement dans un entrepôt de données. Voici trois de ces éléments:

<p>1.</p>	<p>Connaître le nombre de commandes et de devis dans le temps</p>	
<p>2.</p>	<p>Connaître l'évolution du nombre de produits dans le temps</p>	
<p>3.</p>	<p>Connaître le nombre d'utilisateurs d'un GEOService.</p>	

Pour les deux premiers il s'agit d'un accès direct à la base de production. Le troisième nécessite la lecture d'un fichier de type Log4j. Pour le cas numéro 1, il était initialement prévu d'utiliser le datawarehouse par crainte d'une requête trop gourmande. Cependant, après la mise en application du script d'alimentation de l'entrepôt, il a été constaté que la requête était rapide et cette dernière a donc été retravaillée afin de puiser l'information directement dans la base de production. Le script d'alimentation du datawarehouse a tout de même été conservé afin d'effectuer une comparaison entre un accès direct et un accès par datawarehouse. Le résultat est visible au chapitre « Comparaison accès direct - datawarehouse » page 27.

Exemple détaillé : analyse d'un log4j

Cet exemple permet de montrer l'utilisation de Talend. Celui-ci analyse un fichier de log (log4j) de plus de 7mo et en retire les informations demandées. Le but est d'obtenir par qui et combien de fois un service WMS a été utilisé. Le processus prend environ 1,7 seconde. Cet exemple est intéressant car il s'agit d'un fichier de log très verbeux qui n'est en principe pas destiné à de l'analyse informatique. Voici le cheminement nécessaire :

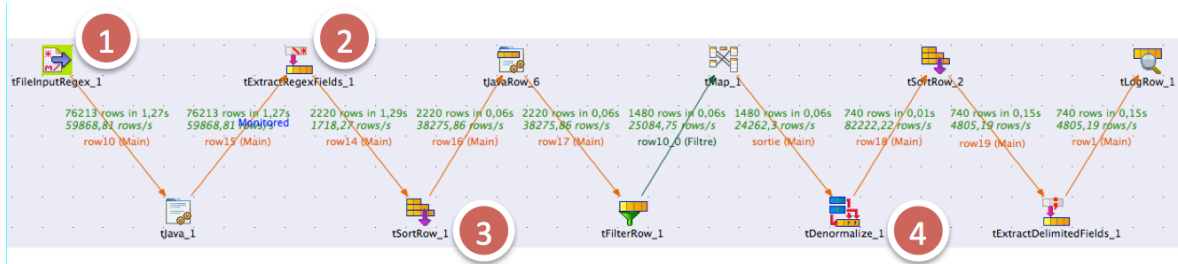


Illustration 14 - analyse d'un log4j

Etape 1 : Ouverture du log et lecture ligne par ligne.

```
[http-8080-Processor20] INFO ProxyLogger:804 - remoteresponseuatetime=22/06/2012 08:18
[http-8080-Processor25] TRACE ProxyLogger:771 - Start - Complete metadata
[http-8080-Processor25] TRACE ProxyLogger:779 - End - Complete metadata
[http-8080-Processor25] TRACE ProxyLogger:208 - Posttreatment file /spring2a2d595e6ed9a0
[http-8080-Processor25] TRACE ProxyLogger:216 - Posttreatment file /202/GetRecordById.xs
[http-8080-Processor21] TRACE ProxyLogger:779 - End - Complete metadata
[http-8080-Processor21] TRACE ProxyLogger:208 - Posttreatment file /spring2a2d595e6ed9a0
[http-8080-Processor21] INFO ProxyLogger:566 - ClientResponseLength=5534
[http-8080-Processor21] INFO ProxyLogger:523 - -----
[http-8080-Processor25] INFO ProxyLogger:563 - ClientResponseDate=22/06/2012 08:18
[http-8080-Processor25] INFO ProxyLogger:566 - ClientResponseLength=5534
[http-8080-Processor24] INFO ProxyLogger:443 - Config=geonetwork
[http-8080-Processor24] INFO ProxyLogger:1308 - Policy=geonetwork
[http-8080-Processor24] INFO ProxyLogger:470 - user=ogc_test
[http-8080-Processor24] INFO ProxyLogger:471 - requestTime=22/06/2012 08:22:19
[http-8080-Processor24] INFO ProxyLogger:472 - RemoteAddr=10.56.2.68
[http-8080-Processor24] INFO ProxyLogger:473 - RemoteUser=ogc_test
[http-8080-Processor24] INFO ProxyLogger:474 - QueryString=null
[http-8080-Processor24] INFO ProxyLogger:475 - RequestURL=http://java.4.cloud4all.ch:8
[http-8080-Processor24] INFO ProxyLogger:822 - Request=<?xml version="1.0" encoding="U
version="2.0.2" resultType="results" outputSchem
xmlns:gml="http://www.opengis.net/gml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-Inst
http://www.opengis.net/cat/csw/2.0.2 http://schemas.opengis.net/csw/2.0.2/CSW-discovery
og:Filter xmlns:ogc="http://www.opengis.net/ogc" xmlns:gml="http://www.opengis.net/gml"
eChar="\`">
<ogc:PropertyName>mainsearch</ogc:Property
<ogc:PropertyName>mainsearch</ogc:PropertyName>
<ogc:PropertyIsLike>
<ogc:PropertyIsEqualTo><ogc:Pro
```

Ces lignes n'ont plus la même structure, car il s'agit du contenu des requêtes (xml)

Aperçu du log brut

Etape 2 : Application d'une expression régulière permettant d'isoler les lignes intéressantes et de les découper en colonnes. On voit que le fichier est déjà nettement plus facile à traiter.

proxyLogger	var	value
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 470	user=	ogc_test
ProxyLogger: 489	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 523	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 523	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 523	-----	

Etape 3 : La colonne des processus est triée permettant de ne pas les mélanger. Les paires config et user sont maintenant bien isolées.

proxyLogger	var	value
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 523	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 523	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 523	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	ogc_test
ProxyLogger: 523	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	ogc_test
ProxyLogger: 523	-----	
ProxyLogger: 443	Config=	geonetwork
ProxyLogger: 504	user=	spring2a2d595e6ed9a0b24f027f2b63b134d6
ProxyLogger: 523	-----	

Etape 4 : Une colonne id est ajoutée. Celle-ci s'incrémente chaque fois qu'on trouve un séparateur de ligne (-----). Ensuite on supprime ces lignes de séparation et on concatène en une seule ligne les éléments possédant les mêmes id. Le résultat est obtenu : chaque ligne contient les informations demandées.

id	value
0	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6
1	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6
2	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6
3	geonetwork ;ogc_test
4	geonetwork ;ogc_test
5	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6
6	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6
7	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6
8	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6
9	geonetwork ;spring2a2d595e6ed9a0b24f027f2b63b134d6

Fin de l'analyse du log4j : Seuls deux éléments ont été isolés. Cependant cette découpe du fichier de log permet aisément d'isoler de nouvelles informations avec la même sortie. La prochaine étape est de mettre à jour les tables de dimensions si on découvre des éléments qui n'y étaient pas encore puis on peuple la table des faits.

Exportation

Un job Talend peut être exporté en tant que job autonome. Il peut ensuite être lancé grâce aux fichiers « .sh » ou « .bat » selon l'OS.

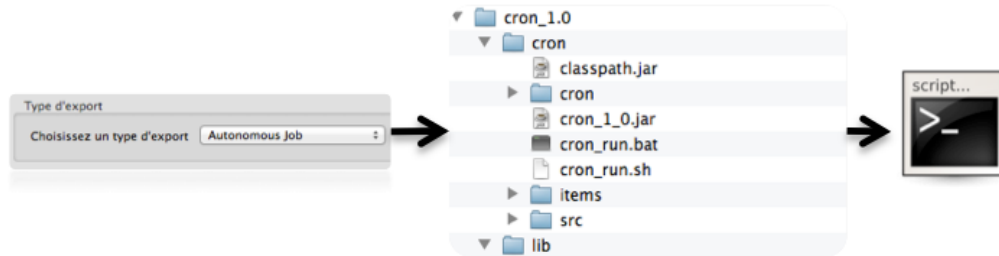


Illustration 15 - Exportation du job d'alimentation

2.6 Couche d'accès



Les données couvrant les besoins sont collectées, il s'agit maintenant d'y accéder grâce à une couche d'accès. Le choix s'est également porté sur Talend permettant ainsi d'homogénéiser la solution. Cependant, le job d'alimentation du datawarehouse et celui de la couche d'accès sont totalement indépendants et différents. Leurs rôles respectifs et le type d'exportation sont différents

Aperçu global

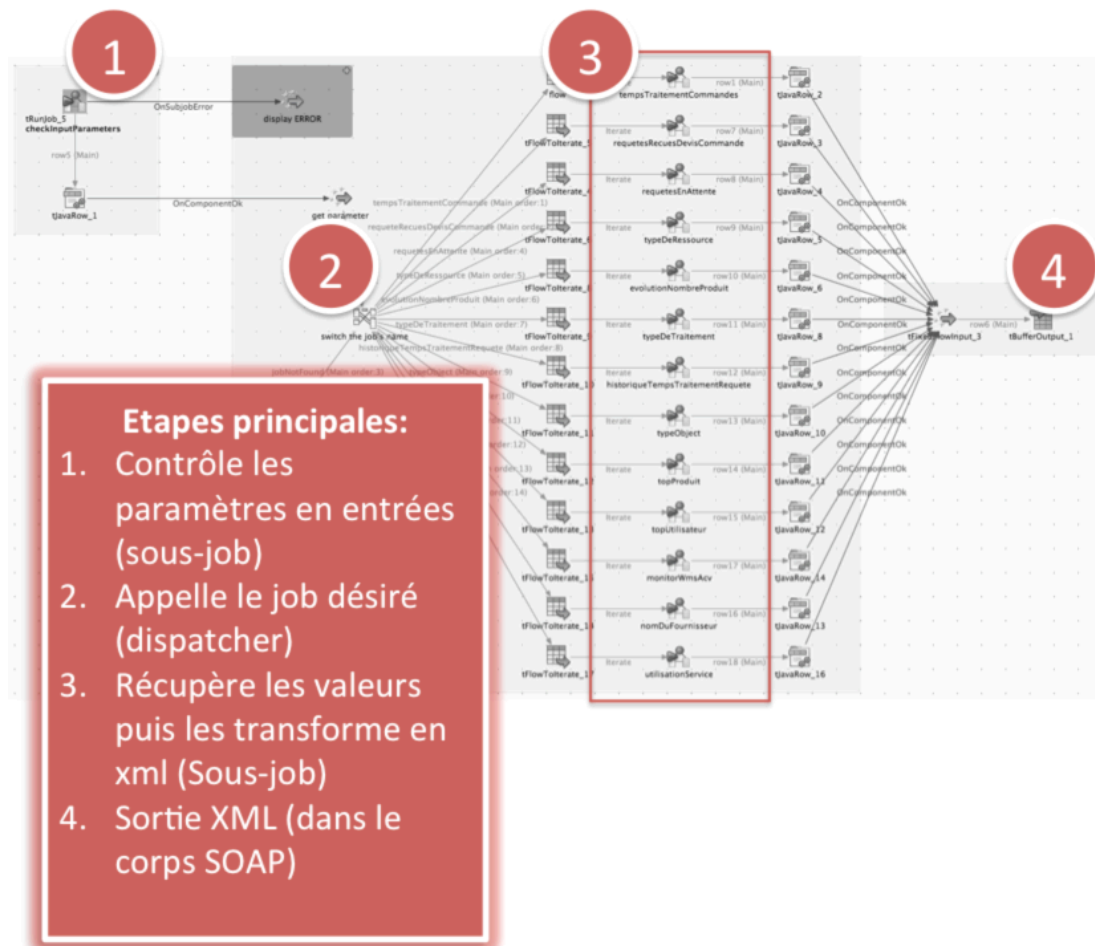


Illustration 16 - Job Talend : Access Layer

L'ajout d'un nouveau besoin est aisé : il suffit de créer un sous-job effectuant le travail de récupération de données puis de le référencer dans le job d'aiguillage numéro 2. L'annexe II « documentation » explique en quelques étapes avec des captures d'écrans comment procéder.

Exportation

Le job de la couche d'accès n'est pas exporté en tant que job autonome mais en tant que webservice. Ceci est une grande force de Talend : un job peut être exporté en tant que fichier « .war » prêt à être déployé sur un serveur Java.

Le schéma ci-dessous illustre le déploiement d'un job Talend en tant que web service. La flèche entre Talend et le serveur Tomcat est volontairement fine pour indiquer que le lien entre Talend et Tomcat est faible. En effet, mis à part le fichier compressé « .war » à déposer dans le dossier WebApps de Tomcat, aucun élément ne les lie. Ce qui implique qu'il y a une grande flexibilité entre les outils.

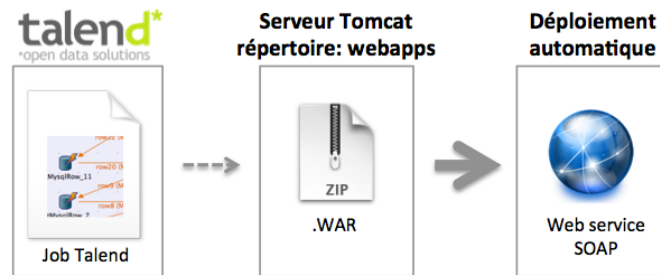


Illustration 17 - Exportation access layer

Tomcat déploie automatiquement le fichier « .war » sur le serveur. Cela génère ensuite un répertoire qui est l'entrée de notre web service.

Pour accéder au web service, on appelle l'url suivante :

```
http://[serveurTomcat]/accessLayer_0.4/services/accessLayer?method=runJob
&arg1=--context_param%20job=topUtilisateur
&arg2=--context_param%20beginDate=0
&arg3=--context_param%20endDate=1343684565
&arg4=--context_param%20idFournisseur=22388
```

L'url ci-dessus retourne un résultat semblable à celui-ci :

```
<?xml version="1.0" encoding="UTF-8"?> <graph><label>Top utilisateurs</label><serie><data>1075</data><label>Jaqui
Officiels</label></serie><serie><data>969</data><label>EFA+C Ingénieurs géomètres officiels SA</label></serie><se
& Amp; Associés. Ingénieurs et Géomètres S.A.</label></serie><serie><data>878</data><label>Administration cantonal
<data>874</data><label>BBHN SA</label></serie><serie><data>821</data><label>Mosini et Caviezel Ingénieurs Géomètr
<data>650</data><label>Schenk SA Ingénieur Géomètre Officiel</label></serie><serie><data>639</data><label>Nicod +
SA</label></serie><serie><data>622</data><label>Bovard & Nickl SA Ingénieur Géomètre Officiel & Ingénieur
<data>614</data><label>B+C Ingénieurs SA</label></serie><serie><data>570</data><label>DTP SA Bureau d'études</lab
<label>Renaud et Burnand SA Ingénieurs Géomètres Officiels</label></serie><serie><data>548</data><label>Duchoud-H
```

On remarque que la réponse contient d'abord une enveloppe SOAP, le résultat à récupérer se trouve ensuite dans le corps de la réponse. Il s'agit d'un fichier XML. Un proxy dans le composant Joomla permet d'accéder au web service, d'analyser l'enveloppe SOAP puis d'extraire les informations nécessaires.

Temps de réponse

Il est intéressant de connaître le temps de réponse de Talend. Le meilleur moyen d'y parvenir est probablement de « sniffer » les appels asynchrones aux proxy. Voici le résultat :

Name	Size	Time	Timeline
proxy.php	757B	1.14s	
proxy.php	426B	554ms	
proxy.php	1.02KB	508ms	
proxy.php	547B	1.27s	
proxy.php	682B	2.01s	
proxy.php	442B	388ms	
proxy.php	397B	934ms	

Illustration 18 - temps de réponse du proxy

L'illustration ci-dessus concerne le tableau de bord de la diffusion. Lors du chargement de la page on fait appel à sept reprises à la couche d'accès Talend. Les résultats sont bons, la totalité des éléments est chargée en moins de 3 secondes. L'avantage des appels avec Ajax et de voir les résultats apparaître au fur et à mesure que le temps progresse. Ces tests ont été réalisés en local avec deux serveurs Java et PHP virtuels.

A noter qu'en premier lieu, le proxy interrogeait la couche d'accès grâce à la classe SoapClient mais chaque appel prenait plus de 3 secondes. Finalement la méthode file_get_contents s'est montrée bien plus performante avec moins d'un cinquième de seconde pour interroger le web service déployé grâce à Talend.

2.7 Composant Joomla

[Besoins](#)[Entrepôt de données](#)[Couche d'accès](#)**Composant Joomla**[Reporting](#)
[Graphiques](#)

La couche d'accès est maintenant accessible par le biais d'un web service. La partie 4 définit le portail par lequel les utilisateurs ont accès à ce web service.

Joomla 2.5

L'ASIT-VD utilise actuellement Joomla afin de permettre aux fournisseurs de distribuer leurs données et aux utilisateurs de les commander. Pour l'instant, la version du portail mis en œuvre est la 1.5. Cependant une migration vers la plateforme 2.5 sera faite d'ici cet automne et c'est donc sur cette dernière version que le développement est réalisé.

Structure de l'application

Joomla comporte une partie front-end et une partie back-end. Afin d'éviter de la redondance de code, le comité de pilotage du projet EasySdi a pris le parti de donner un accès back-end aux fournisseurs. On peut voir dans le tableau ci-dessous que les deux groupes de public cible dont la priorité est la plus haute ont un accès back-end à l'application. Le focus est donc porté sur cette partie de Joomla.

Résultat obtenu

Deux tableaux de bord sont complétés afin d'illustrer les différentes possibilités d'affichage. Les captures d'écran ont été prises sur une période de 2 ans dans la peau de l'administration cantonale Vaudoise (le plus gros fournisseur de données actuelles).

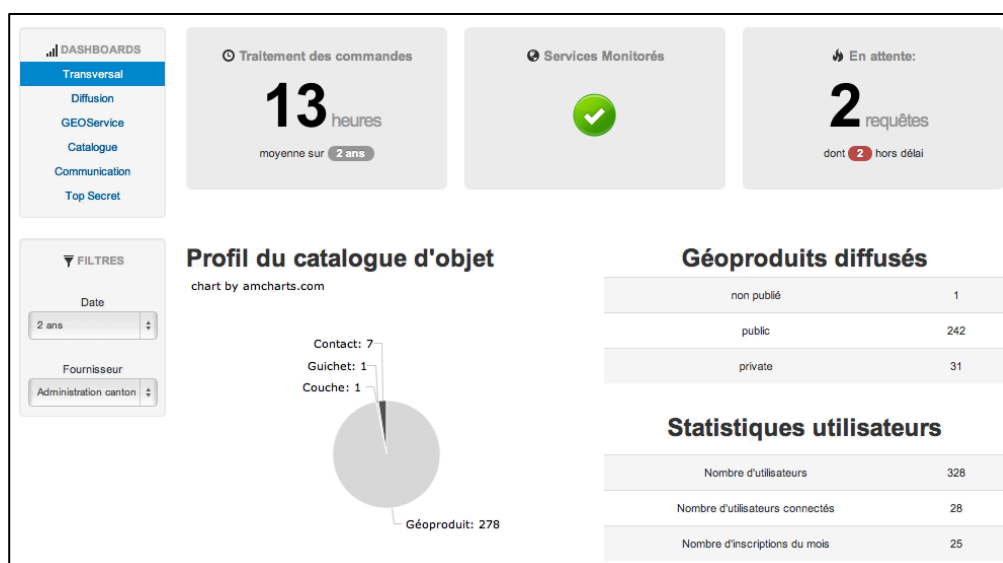


Illustration 19 - Tableau de bord transversal

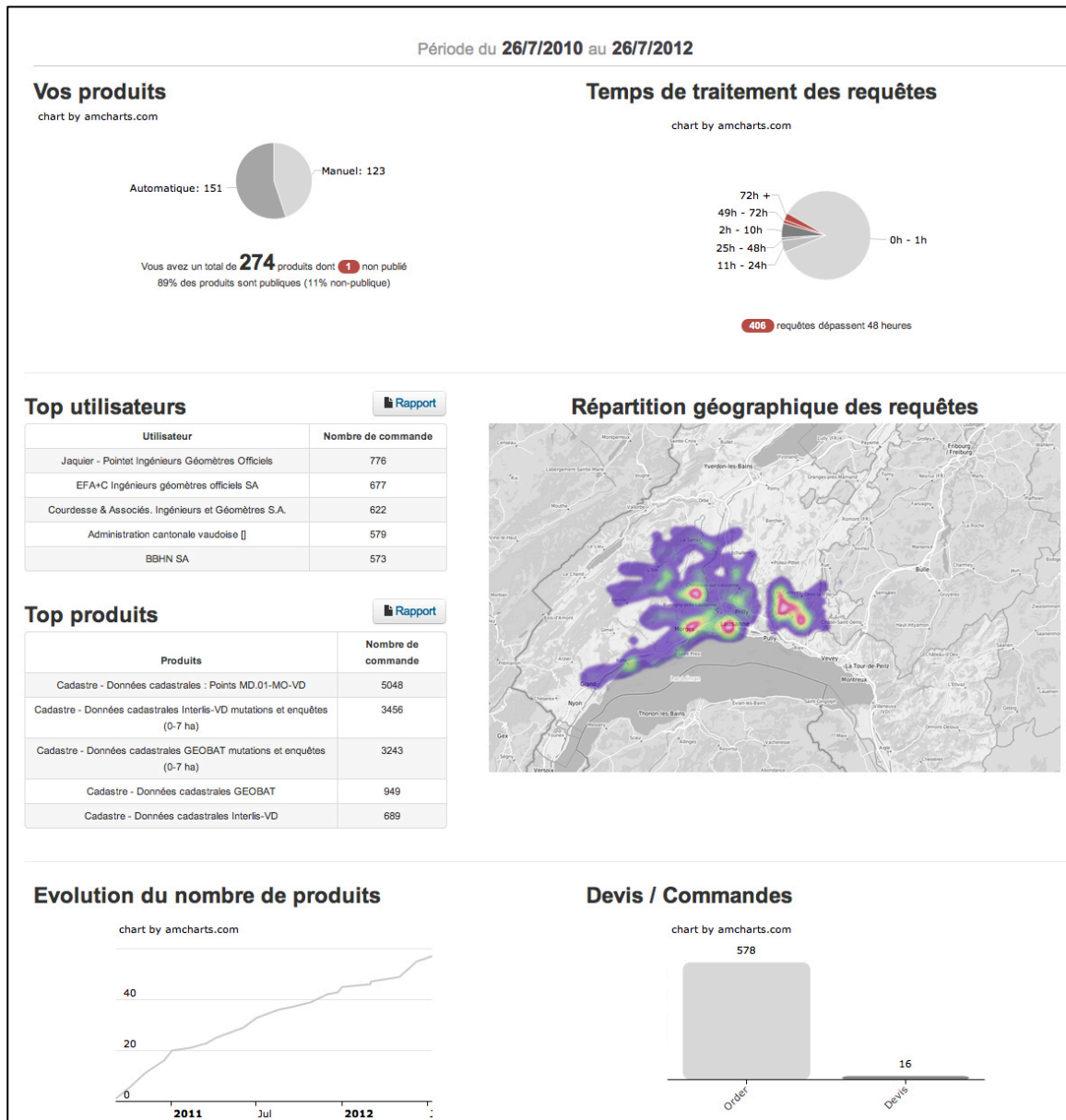
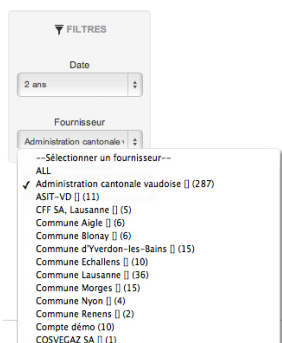


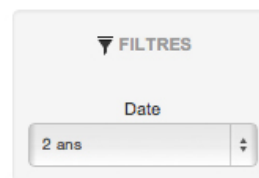
Illustration 20 - Tableau de bord de la diffusion (sans le menu)

Fonctionnalité multi-vue

La fonctionnalité multi-vue permet à un utilisateur de prendre la vue d'un fournisseur de données. C'est une fonctionnalité qui est en principe destinée à l'ASIT-VD. Cela permet à l'association de voir l'ensemble des données (de tous les fournisseurs) ou de prendre la vue d'un fournisseur particulier. Cette fonctionnalité permet de répondre au besoin « Un administrateur peut prendre la vue d'un fournisseur ».



Utilisateur avec la fonctionnalité « multi-vue »

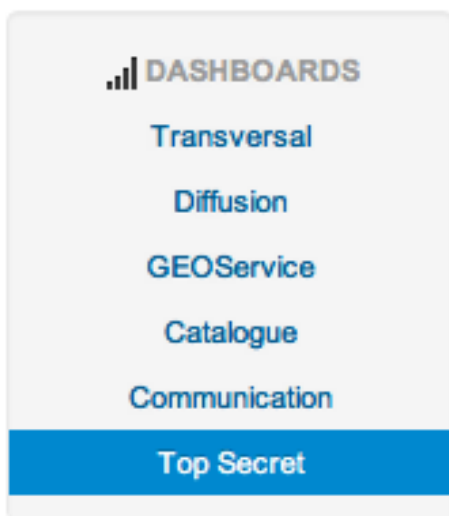


Utilisateur normal. Il ne voit que les données qui le concerne

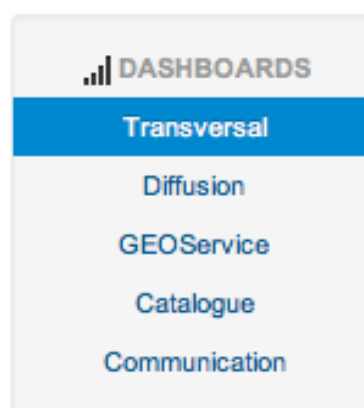
Afin de choisir qui a le droit à la multi-vue, on utilise les ACL (Access Control List) de Joomla.



Illustration 21 - Utilisateur à vue multiple

Fonctionnalité tableaux de bord cachés

Utilisateur ayant accès à
une vue cachée



Utilisateur normal. Il ne voit que
les tableaux de bord publiques

```
1 • SELECT displayName, x52dk_dashboard.ordering, x52dk_viewlevels.title
2 FROM joomla.x52dk_dashboard
3 INNER JOIN x52dk_viewlevels ON x52dk_viewlevels.id = x52dk_dashboard.access
```

displayName	ordering	title
Transversal	0	Public
Diffusion	1	Public
GEOService	2	Public
Catalogue	3	Public
Communication	4	Public
Top Secret	5	asit-vd admin access level

Illustration 22 - Les vues en base de données

La capture d'écran ci-dessus montre l'utilisation des niveaux de vues de Joomla afin de gérer l'affichage ou non d'un tableau de bord. Les vues peuvent facilement être gérées dans le back-end de Joomla. Cependant, l'attribution d'une vue à un élément du menu se fait pour l'instant dans la base de données du composant.

Cette fonctionnalité est utile dans la situation suivante : un fournisseur a un besoin particulier qui ne concerne que lui (par exemple des types de produits que lui seul possède) et il désire pouvoir les visualiser sur un tableau de bord. Dans ce type de situation il suffit de créer une vue pour ce fournisseur et de créer un tableau de bord dont la colonne « access » correspond à cette nouvelle vue.

Bootstrap



Bootstrap, from twitter

Bootstrap est un ensemble d'éléments HTML, CSS et Javascript permettant d'obtenir une interface moderne, fonctionnelle et surtout de type « responsive design ». En effet, celle-ci s'adapte parfaitement à toutes les résolutions (y compris pour une utilisation mobile). Bootstrap a été initialement développé pour Twitter. Cependant son succès est tel qu'il a ensuite été distribué indépendamment sous licence GPL.

Pourquoi Bootstrap? Car en plus d'être extrêmement efficace, l'équipe de développement Joomla a annoncé en mars de cette année que Bootstrap sera intégré par défaut dans Joomla. Voici un article très intéressant à ce sujet.

« Joomla 3 sort en septembre et l'équipe a travaillé dur afin de mettre sur place un nouvel air Joomla.

Le changement le plus important concerne l'ensemble du design Joomla. L'ensemble de la partie front-end et back-end sont en train d'être entièrement repensé.

La prochaine révolution web concernera les mobiles. De plus en plus de personnes se rendent sur le web et un grand nombre d'entre eux n'utiliseront plus que les appareils mobiles pour le faire. Notre objectif est de rendre Joomla « mobile-friendly » tout en préservant l'identité de Joomla.

Kyle Ledbetter (architecte des interfaces utilisateurs, développeur influent de Joomla) et son équipe travaillent sur la refonte de Joomla en y intégrant Bootstrap de Twitter. Twitter décrit Bootstrap comme étant : « une extension front-end pour développer sites et applications ». Utilisé Bootstrap permet à Joomla de développer un outil mobile-friendly basé sur un framework populaire d'ores et déjà apprécié par de nombreux développeurs et designers. Plus d'informations sur la page officielle du projet Bootstrap pour Joomla : <http://twitter.github.com/bootstrap>

Il reste encore cinq mois avant la sortie de Joomla 3.0 et du travail doit encore être effectué mais ceci est la solution vers laquelle on tend. Vous pouvez trouver sur : <http://ux.joomla.org> l'état d'avancement du projet et de plus amples informations. »

Article datant du 16 avril 2012 traduit par mes soins

Article original : <http://community.joomla.org/blogs/community/1568-30-first-look.html>

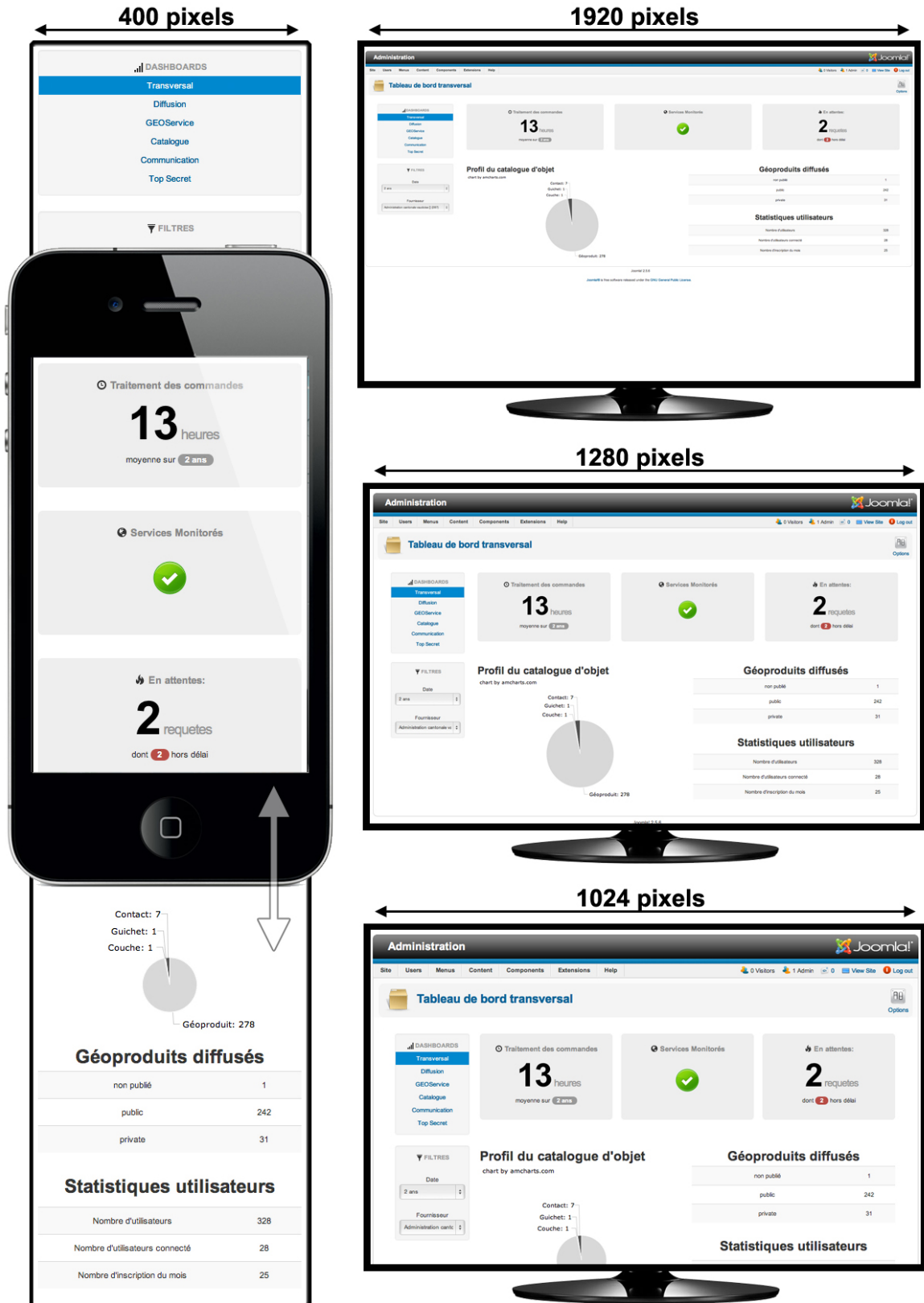


Illustration 23 - Responsive design en action

2.8 Reporting

Besoins

Entrepôt de
données

Couche d'accès

Composant
Joomla

Reporting

Graphiques

Quatre sorties standards ont été identifiées dans la partie théorique. Pour rappel : Reporting, Tableau de bord, Datamining, OLAP. Cependant, seuls les deux premiers outils ont été privilégiés pour ce travail. En effet, l'objectif de ce travail est de fournir des outils permettant la création de tableaux de bord et de rapports. Toutefois, des outils de type OLAP ou Datamining sont des évolutions possibles.

Reporting



Pour le reporting, l'outil retenu est Birt. C'est un projet Open Source basé sur Java. Birt est composé de deux outils. D'une part, le report designer permet de configurer le rapport grâce à une interface graphique intuitive et d'autre part le report Engine permet d'interpréter les fichiers précédemment créés.

Structure

Le report designer crée des fichiers de type .rptdesign (format XML). Le deuxième outil est l'engine Report qui doit être installé sur un serveur Apache. Ce dernier permet d'interpréter les fichiers rptdesign et d'en générer des fichiers html, pdf, doc, xls, odt ou encore ods.

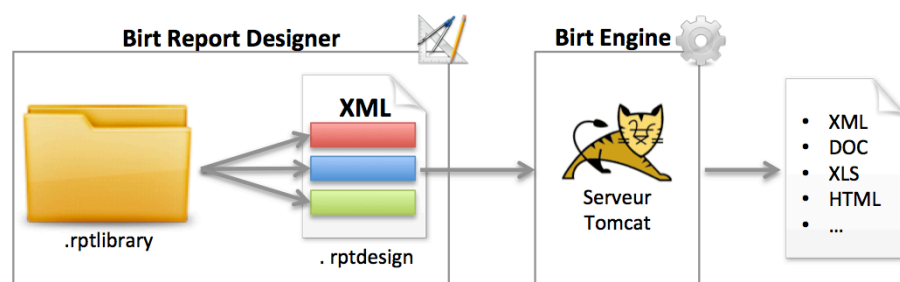


Illustration 24 - structure des rapports Birt

Le schéma ci-dessus illustre qu'un fichier rptdesign inclut des éléments puisés dans un rptlibrary. Ce n'est pas obligatoire mais ça permet de pouvoir réutiliser aisément des éléments (mise en page, graphique, style,...) afin de simplifier la création de futures rapports et garder une ligne graphique similaire sur l'ensemble des rapports.

Exemples de rapports

Temps de traitement des requêtes au format PDF :

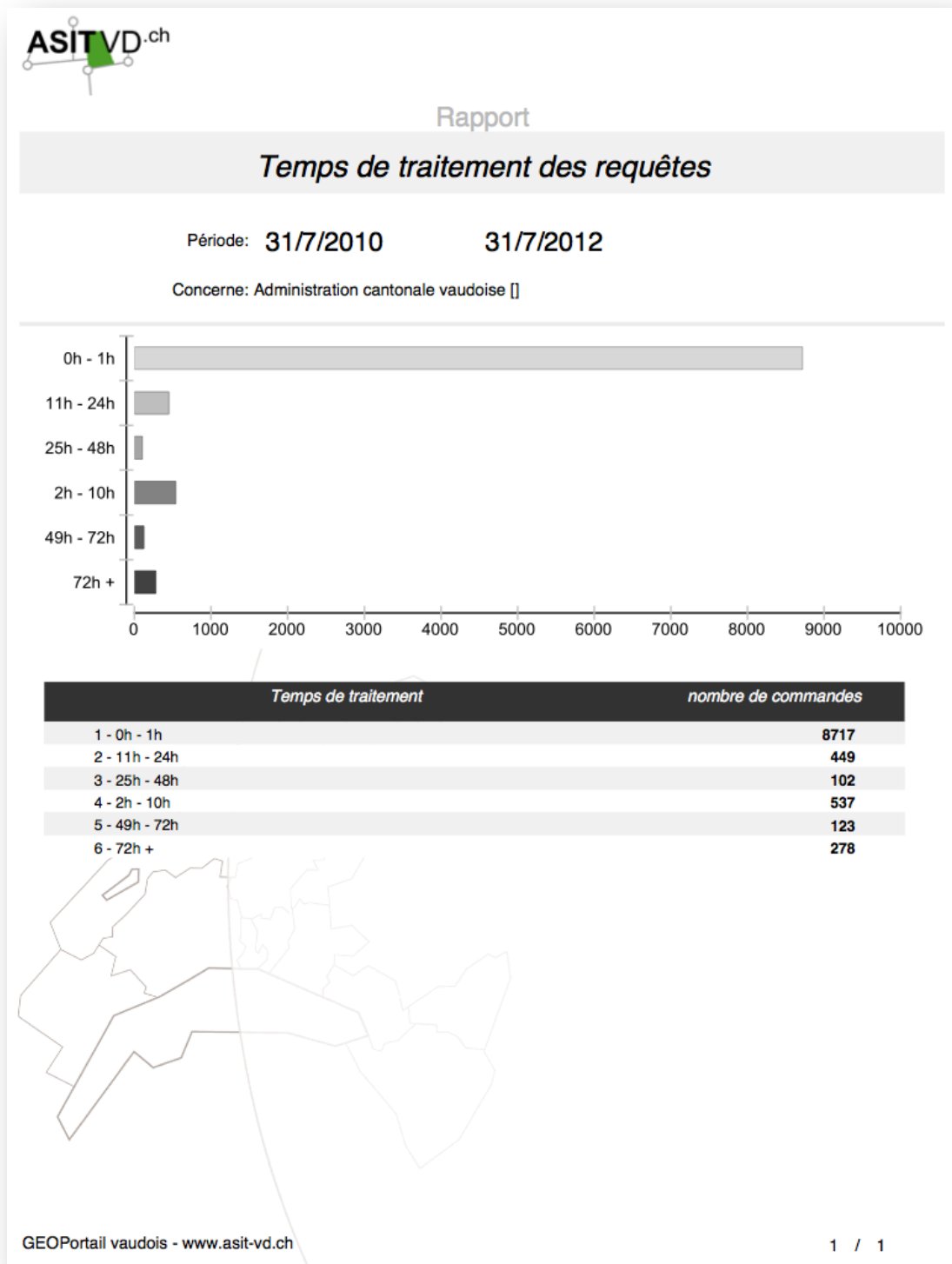


Illustration 25 - rapport Birt

Top des utilisateurs au format PDF :

	A	B	C	D	E	F	G	H	I
1	Rapport								
2	Top des utilisateurs								
3	Période:		31/7/2010	31/7/2012					
4	Concerne: Administration cantonale vaudoise []								
5	Utilisateur								nombre de commandes
6	1 - Jaquier - Pointet Ingénieurs Géomètres Officiels								763
7	2 - EFA+C Ingénieurs géomètres officiels SA								666
8	3 - Courdesse & Associés. Ingénieurs et Géomètres S.A.								618
9	4 - Administration cantonale vaudoise []								567
10	5 - BBHN SA								563
11	6 - Mosini et Caviezel Ingénieurs Géomètres Officiels								554
12	7 - Nicod + Perrin ingénieurs et géomètres SA								478
13	8 - Bovard & Nickl SA Ingénieur Géomètre Officiel & Ingénieur Civil								471
14	9 - Schenk SA Ingénieur Géomètre Officiel								431
15	10 - B+C Ingénieurs SA								394
16	11 - DTP SA Bureau d'études								366
17	12 - Renaud et Burnand SA Ingénieurs Géomètres Officiels								362
18	13 - Duchoud-Haymoz-Buhlmann SA Ingénieurs Géomètres Officiels								342
19	14 - Olivier Peitrequin SA								313
20	15 - Bourgeois & Rudaz SA Ingénieurs géomètres officiels								309
21	16 - Parisod Ingénieur Géomètre Officiel								297
22	17 - Jomini-Van Buel SA								289
23	18 - Richard & Cardinaux SA Ingénieurs Géomètres Officiels								280
24	19 - Rossier Ingénieur Géomètre Officiel								280
25	20 - ORCEF SA								274
26	21 - Jourdan Ingénieur Géomètre Officiel								255
27	22 - Boss Philippe Bureau de géomètre officiel								185
28	23 - Gasser Jean-Claude Ingénieur géomètre officiel								177
29	24 - Commune Lausanne []								175
30	25 - Bureau technique Stuby & Völlmy Sàrl								172
31	26 - Thorens SA Bureau d'études techniques								167
32	27 - Lehmann Ingénieur géomètre officiel								162
33	28 - Romande Energie SA []								148
	...								
200	195 - PRONA SA								1
201	196 - Natura biologie appliquée Sàrl								1
202	197 - Bureau d'études biologiques Delarze								1
203	198 - Fabienne Gaiani								1
204	199 - IJB Ingenieur-Unternehmung AG								1
205	200 - GAZNAT SA []								1
206	201 - Commune Echallens []								1
207	202 - Compte démo								1
208									
209	GEOPortail vaudois - www.asit-vd.ch								1 / 1

Illustration 26 - Export XLS Birt

On remarque que les colonnes sont respectées et ainsi les données peuvent facilement être extraites par l'utilisateur.

2.9 Graphiques

Besoins

Entrepôt de
données

Couche d'accès

Composant
Joomla

Reporting

Graphiques

Choix de l'outil

Quel outil choisir pour la création de graphique ? Le flash étant une technologie sur le déclin (l'incompatibilité avec les mobiles Apple et l'utilisation abusive du processeur explique en partie ce déclin), le choix s'est naturellement tourné vers la technologie Javascript et HTML5 pour générer des graphiques SVG/VML. Cependant le nombre de librairies disponibles est vertigineux. Afin d'effectuer un choix, les principales librairies open source ont été intégrées et testées. Le résultat est visible sur <http://gbernard.ch/documents/prototype/> (le site est construit avec Bootstrap pour un test optimal). Résultat du test :

	Graph gauge*	Auto size**	Qualité Visuelle	Commentaire
Highcharts	✓	✓	✓✓✓	Une librairie parfaite. Cependant le prix de la licence est contraignant : 360\$ pour une utilisation commerciale.
Flot	✗	✓	✓	Dernière mäj de mars 2011. C'est beaucoup dans cet environnement qui évolue beaucoup.
Jqplot	✓	✗	✗	Le redimensionnement est disponible mais il est vraiment problématique : le graphique est souvent illisible à cause de chevauchements d'éléments
Flotr2	✗	✗	✓	Très bonne librairie, dommage que le redimensionnement ne soit pas disponible
Elycharts	✗	✗	✓	Idem que pour flot2
Extjs	✓	✓	✓✓	Extjs est très abouti. Cependant, mettre en place un tel framework pour n'utiliser que les graphiques semblent lourds. L'ensemble de la solution Ext JS aurait plutôt remplacé la solution Bootstrap.
AmCharts	✗	✓	✓✓	Gratuit si on laisse le lien sur le site AmCharts (également pour une utilisation commerciale). Sinon la licence est 100€.
	→ choix retenu			

* Graph gauge : Permet de contextualiser une valeur en donnant la limite de celle-ci et éventuellement les zones critiques. Ce n'est pas un critère rédhibitoire car il existe des librairies JS spécialisée dans ce type de graphiques (Jgauge, JsGauge entre autres).

** Auto size : indique si le graphique à la capacité de s'adapter dynamiquement à la taille qu'on lui alloue (généralement sur la largeur). C'est un critère important du fait de la mise en place du responsive design avec Bootstrap.

Après une analyse approfondie, c'est amCharts qui a finalement été sélectionné. Le choix a été difficile entre amCharts et Highcharts.

Highcharts est très populaire, son rendu est très soigné et ses options sont multiples. Cependant, le prix de la licence est contraignant car, pour rappel, le but ultime du projet est d'être intégré à EasySdi. Highcharts exige une licence pour chaque utilisation commerciale (incluant les administrations mais pas les associations donc se serait gratuit pour l'ASIT-VD). L'intégration d'un outil payant dans EasySdi risque donc de générer un rebus de l'outil par la communauté. C'est cet argument qui a finalement porté le choix sur AmCharts.

La licence AmCharts est bien plus souple, l'outil est gratuit avec un lien sur le site amCharts et payant si on désire retirer le lien. Voici un aperçu du lien (dans le coin supérieur gauche) :

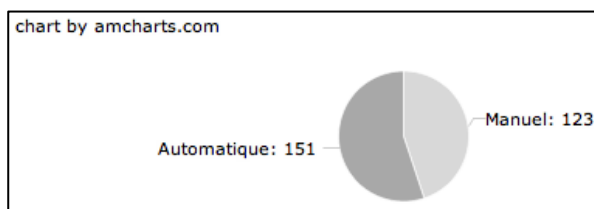


Illustration 27 – Lien dans le graphique amCharts

AmCharts

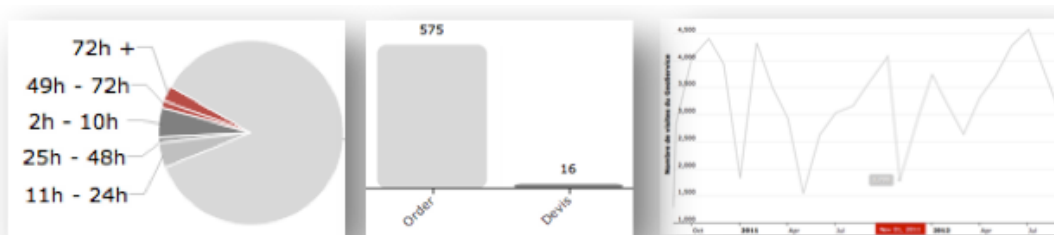


Illustration 28 - Exemples de rapport Birt

Voici les trois principaux graphiques générés par la librairie. Pour les couleurs, le choix s'est porté sur des tons de gris et sur du bordeaux pour les valeurs à mettre en évidence (généralement une valeur critique).

3 Bilan

L'ensemble des cinq étapes de la réalisation du travail a permis d'obtenir un système permettant aisément la réalisation de tableaux de bord, de graphiques et de rapports. Deux tableaux de bord ont été entièrement réalisés : le tableau transversal et le tableau de la diffusion. Ceux-ci ont permis de couvrir l'ensemble des situations : accès direct dans la base de production, accès par un datawarehouse, analyse d'un fichier de log, interrogation d'un web service, création de rapports, affichages de graphiques et visualisation de données empiriques. La réalisation des autres tableaux de bord peut donc s'effectuer aisément en se basant entièrement sur ce projet.

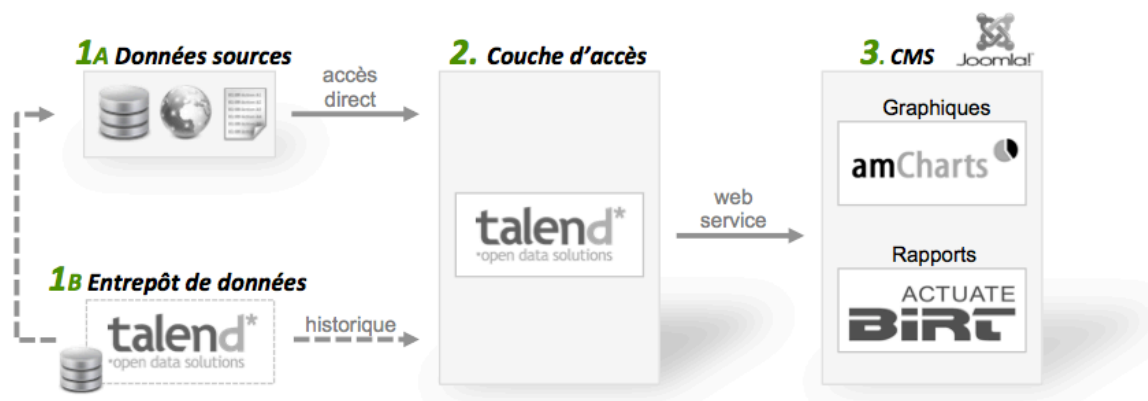


Illustration 29 - Schéma final

L'illustration ci-dessus résume bien le projet : une couche d'accès (2) récupère des données provenant directement de la base de production (1a) ou d'un entrepôt de données (1b). Un composant Joomla (3) accède ensuite à cette couche d'accès pour réaliser des tableaux de bord et des rapports.

3.1 Objectifs atteints

Bilan des objectifs fixés dans le cahier des charges :

Objectifs fixés	Réponses
Rendre un tableau de bord permettant de mesurer l'activité du GEOPortail vaudois.	<i>Un tableau de bord transversal donne une vision d'ensemble de l'activité. Quatre tableaux de bord permettent de donner des détails concernant des domaines d'activités</i>
Proposer un outil permettant la création de reporting	<i>Trois rapports Birt incluent textes et graphiques. Ceux-ci peuvent être exportés en pdf, doc et xml.</i>
Offrir à chaque fournisseur un outil de monitoring d'utilisation des données fournies	<i>Les différents tableaux de bord affichent des données qui dépendent de l'utilisateur enregistré</i>
Intégrer le tableau de bord au portail existant	<i>La plateforme de distribution est sous la forme d'un composant Joomla parfaitement intégrable au portail existant.</i>
Rendre un code source dont la structure est modulaire et évolutive	<i>Les différents outils mis en place sont parfaitement indépendants et ceci offre une grande flexibilité pour l'évolutivité du projet.</i>

Le tableau ci-dessus permet d'affirmer que les objectifs fixés dans le cahier des charges sont atteints.

3.2 Divergences avec le travail préparatoire

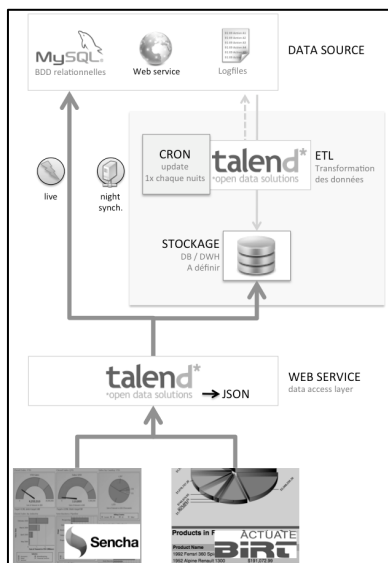


Illustration 30 - Schéma du travail préparatoire

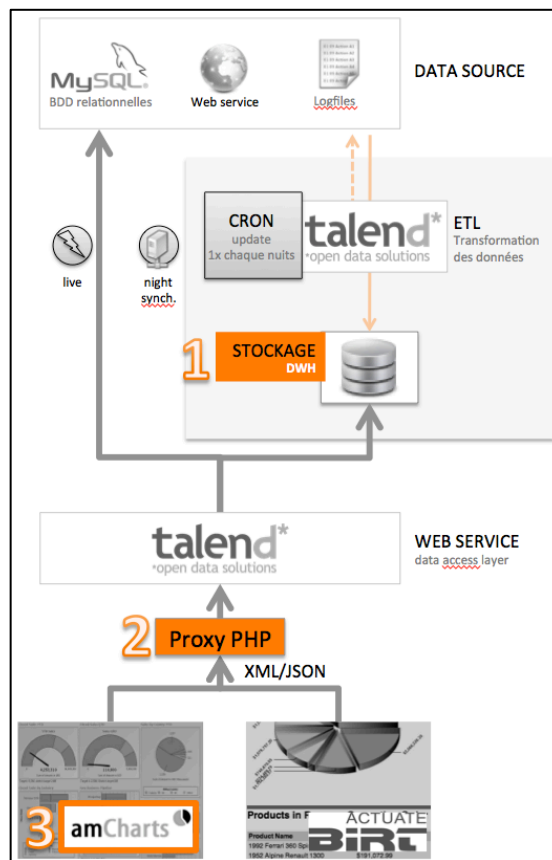


Illustration 31 - Divergences avec le TPB

Le schéma de gauche est le fruit du travail préparatoire. Il faisait office de conclusion du rapport. Le modèle de droite est la solution qui est actuellement en place. Les trois éléments oranges mettent en évidence les différences qui sont les suivantes :

1. Stockage dans un datawarehouse

Durant le travail préparatoire, il n'était pas clair pour moi que les données traitées par un ETL soient forcément destinées à être insérées dans un datawarehouse ou si on peut simplement les insérer dans une nouvelle base. Ma conclusion est que cette nouvelle base répond aux critères de ce qu'on appelle entrepôt de données.

2. Mise en place d'un proxy

Mon souhait était de générer du JSON en sortie de Talend. Cependant l'outil « Talend for data Intergration » génère un web service SOAP (donc forcément avec une enveloppe XML) et Talend offre plus de fonctionnalités pour le XML. Le choix a donc été de mettre en place un proxy analysant la réponse XML de Talend et générant ensuite au choix du XML ou du JSON. Le format JSON est plus adapté pour du javascript et le XML est disponible par défaut pour Birt (ce qui n'est pas le cas de JSON).

3. Utilisation d'amCharts

Initialement il était prévu d'utiliser Ext JS de Sencha pour générer les graphiques. Cependant, après quelques tests, il s'est avéré qu'Ext JS est excellent lorsqu'il est utilisé avec l'ensemble de son framework (avec les conteneurs) mais que l'unique utilisation des graphiques ne justifiait pas la mise en place d'un framework complet pour cause de lourdeur. La décision a donc été prise de sélectionner une librairie ayant un focus exclusif sur la création de graphiques. Après quelques recherches, c'est amCharts qui a été sélectionné.

Des divergences existent donc entre la solution planifiée durant le travail préparatoire et la solution finale. Cependant ces différences sont mineures. Les jours supplémentaires pris pour le travail préparatoire (voir p. 57 « 3.5 Gestion du temps ») ont donc été bénéfiques car ils ont permis de faire une planification proche de la solution finale.

3.3 Problèmes rencontrés

Quelques problèmes ont été rencontrés durant ce travail :

Attaquer le web service Google Analytics

Bien que cette tâche apparaissait être un travail facilement réalisable je me suis confronté à un sérieux problème pour attaquer le web service de Google Analytics. Ce dernier utilise le protocole oauth 2.0 pour assurer une connexion sécurisée. Google n'est d'ailleurs pas le seul : Facebook, Flickr, LinkedIn, Yahoo et Twitter utilisent également ce protocole.

Cependant, malgré la popularité de oauth 2.0, je n'ai pas réussi à trouver de composant Talend permettant d'accéder à ce type de web service. Beaucoup de tutoriaux existent mais exigent une interaction avec l'utilisateur ce qui n'est pas adéquat avec un ETL.

C'est donc avec regret que j'ai renoncé à atteindre ce service. Cependant l'appel au web service Monitor d'EasySDI s'est fait très facilement.

Problèmes de rendus avec Internet Explorer 7

Quelques problèmes de rendus et de navigation sont visibles avec Internet Explorer 7 rendant l'utilisation du composant Joomla un peu difficile. Ceci peut certainement être corrigé avec quelques modifications. Cependant le manque de temps ne m'a pas permis de le faire.

Aucun problème n'a été constaté avec Safari, Chrome, Firefox et les versions d'internet explorer supérieures à la 7.

3.4 La suite du projet

Intégration du projet au portail actuel

Pour l'instant le composant est totalement indépendant. L'ASIT-VD implémentera cet automne la version 2.5 de Joomla. Le projet étant développé avec cette version, il peut donc théoriquement être intégré au portail à ce moment là. Cependant, le souhait du mandant est de passer relativement rapidement à la version 3.5 qui implémentera par défaut Bootstrap. Un travail devra donc être effectué afin d'adapter le composant pour cette nouvelle interface 3.5 et il faut compter tout de même environ un à deux mois pour implémenter la liste de tous les besoins. Le projet sera donc certainement implémenté d'ici une année.

Suggestions d'amélioration

Sécurité

Un travail reste à effectuer au niveau de la sécurité. Lors de l'implémentation de la couche d'accès, il faut s'assurer que seul le proxy puisse accéder à cette couche. Pour l'instant, un utilisateur expérimenté peut analyser le trafic web et aisément consulter les données d'un autre fournisseur. Le risque existe donc au niveau de la confidentialité. Cependant, en accord avec le mandant, ce point a été mis de côté afin de se concentrer sur le reste du projet.

Gestion multilingue

Pour l'instant le composant est développé entièrement en français et sans utiliser les fichiers de langues par défaut de Joomla. Ceci est donc un élément qui pourrait être corrigé.

Ajout d'options dans le menu

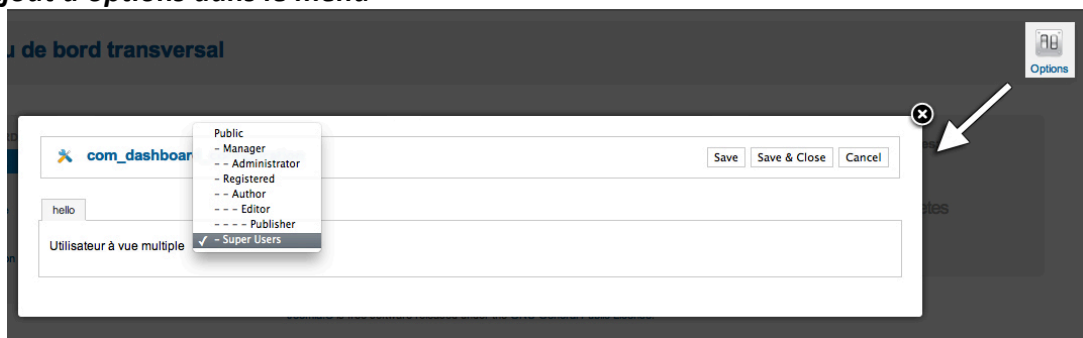
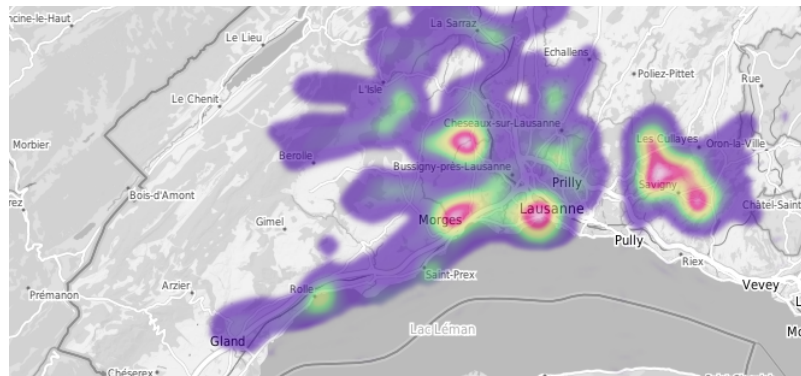


Illustration 32 – Options d'un composant Joomla

Pour l'instant, comme le démontre la figure ci-dessus, seule la gestion de la vue multiple est possible dans les options du composant. Certaines variables définies dans le fichier PHP contenant les variables statiques pourraient être déplacées dans ces options afin de permettre à l'utilisateur de définir plus d'options facilement.

Le chapitre « Fonctionnalité multi-vue » p.41 mentionne qu'il est nécessaire de modifier un champs dans la base de données afin d'attribuer une vue à un tableau de bord. Cette fonctionnalité pourrait par exemple figurer dans les options.

Heatmap*Illustration 33 - Heatmap statique*

Lors d'une présentation du projet, un fournisseur a rapidement flashé sur le heatmap présent sur le tableau de bord de la diffusion. Il a été déçu d'apprendre que ce n'était qu'une image statique car, pour l'instant, aucun outil en place ne permet de visualiser les zones d'activités.

Ajout des analyses OLAP

Pour l'instant, seules les sorties de type tableau de bord et rapport sont mises en place. Cependant il serait possible d'ajouter une possibilité de faire des analyses de type OLAP. Celles-ci offre une grande liberté aux utilisateurs pour faire des analyses avancées. De plus, l'outil mis en place pour les rapports, Birt offre ce type de fonctionnalité.

3.5 Gestion du temps

Le projet a duré 57 jours à raison de 8 heures par jour. La durée initialement allouée pour la réalisation du travail de Bachelor est de 51 jours.

Les jours excédentaires sont dûs au fait que le travail préparatoire a débuté un peu avant. En effet, l'univers de la Business Intelligence étant nouveau pour moi, j'ai pris l'initiative de faire durer le travail préparatoire un peu plus longtemps afin de pouvoir me documenter sur cette discipline.

Le tableau ci-dessous permet d'avoir une vue d'ensemble sur la durée de chacune des phases du projet. Il est intéressant de noter que la phase d'analyse s'est déroulée durant le travail préparatoire de Bachelor (TPB) et les phases de conception, réalisation et test durant le travail de Bachelor (TB)

DEBUT		FIN	DUREE	
07.05.12	TPB	29.05.12	18 jours 3 heures 40 m 10 tâches	+8 jours
07.05.12	Analyse	29.05.12	18 jours 3 heures 40 m 10 tâches	
11.06.12	TB	03.08.12	37 jours 21 heures 40 m 22 tâches	-2 jours
11.06.12	Conception	22.06.12	8 jours 22 heures 40 m 5 tâches	
22.06.12	Réalisation	26.07.12	24 jours 0 heures 30 m 10 tâches	
26.07.12	Test	03.08.12	5 jours 21 heures 15 m 7 tâches	

Illustration 34 - Durée des phases du projet
(réalisé avec ADN Projector 2.4.3)

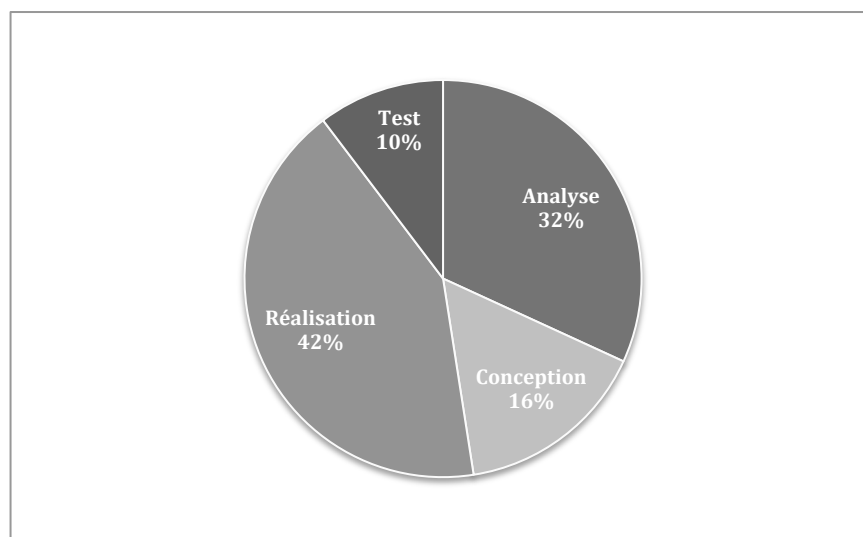


Illustration 35 - Répartition temporelle des phases du projet

Diagramme de Gantt

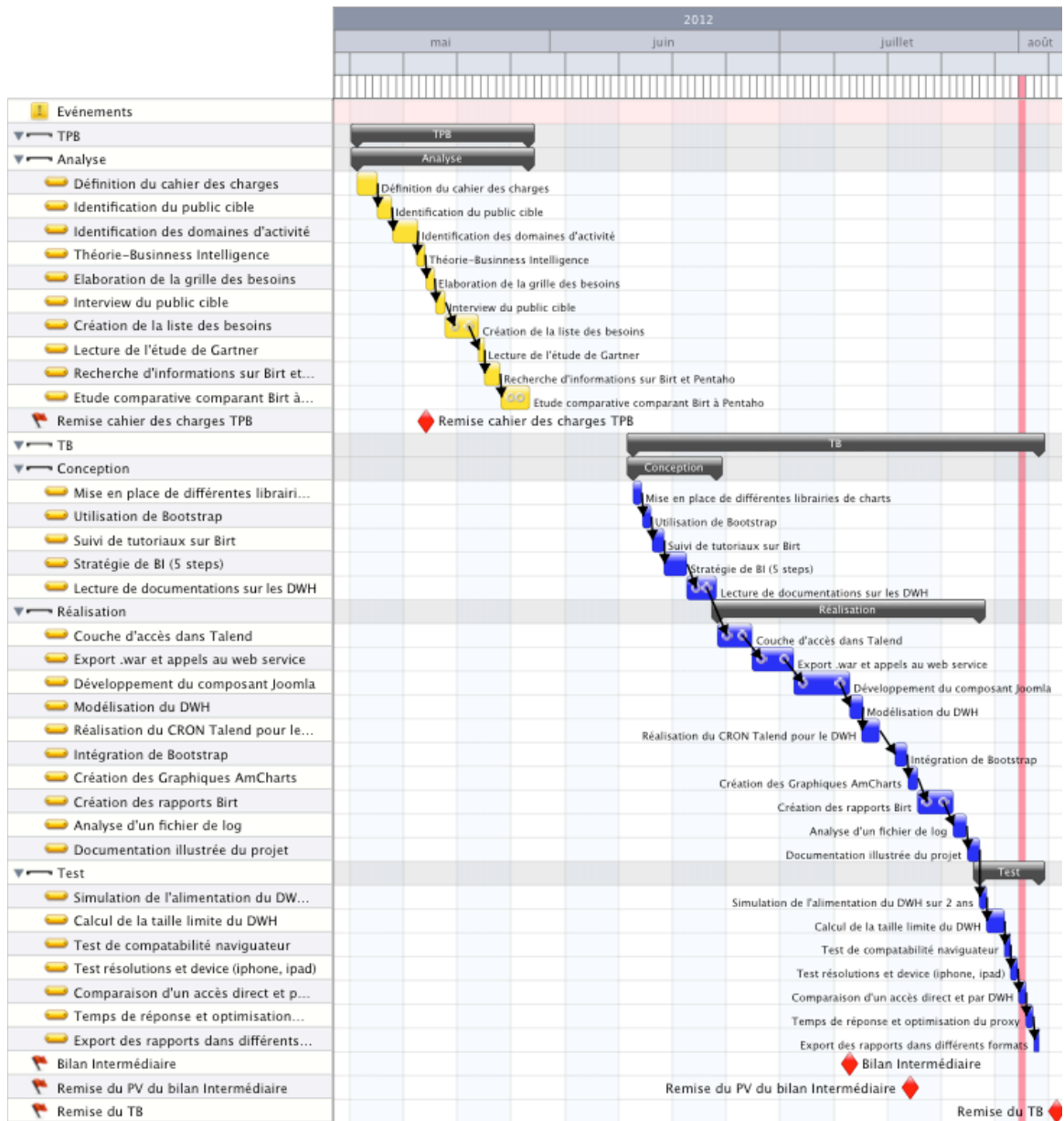


Illustration 36 - Diagramme de Gantt
(réalisé avec ADN Projector 2.4.3)

Tâches importantes

Il est intéressant de connaître les tâches qui ont demandé le plus de temps :

Durée	Tâche	Phase
146 h	Développement du composant Joomla	Réalisation
79 h	Lecture de documentations sur les DWH	Conception
76 h	Export .war et appels au web service	Réalisation
54 h	Création des rapports Birt	Réalisation

C'est sans surprise que la tâche « Développement du composant Joomla » arrive en première position. En effet, la prise en main du modèle MVC de Joomla n'est pas des plus aisées. Cependant j'ai été surpris que la phase de documentation sur les entrepôts de données ait été si longue.

Conclusion

Ensemble du projet

Le choix a été fait durant le travail préparatoire de privilégier des outils indépendants permettant de répondre à chacune des étapes de façon autonome. Ce choix a été payant car l'outil en place est flexible et évolutif. Les tutoriaux de la documentation (annexe II) sont un moyen de le prouver.

Le travail effectué en grande partie dans les locaux de l'ASIT-VD a permis d'être au plus proche des besoins évoqués par le mandataire. De plus l'approche itérative proche du mandataire et du responsable technique a permis de rapidement faire évoluer le projet tout en s'assurant de répondre aux objectifs.

La première version du cahier des charges a été créée le 1er mars 2012. Cette version a ensuite été confirmée sans changement durant le bilan intermédiaire du 10 juillet 2012. Je pense que c'est un signe positif concernant la gestion de ce travail.

Note personnelle

Ce projet était très stimulant pour moi car celui-ci était varié. En effet, il contenait une partie de théorie sur les systèmes de Business Intelligence avec lesquels je n'avais aucune expérience et une partie pratique touchant de multiples outils. Joomla, Talend, Bootstrap et Birt sont tous des outils que je n'avais jamais utilisés. Ce projet m'a donc apporté une grande satisfaction et des connaissances que je pourrais certainement faire valoir par la suite.

Je dois avouer que Talend m'a particulièrement impressionné. Les nombreux tutoriaux, la réactivité de la communauté et la qualité de l'outil ont fait de l'utilisation de Talend une expérience agréable et enrichissante.

La remarque suivante avait déjà été faite durant le travail préparatoire mais je tiens à la réitérer en guise de note finale de ce projet car elle s'est confirmée durant la phase de réalisation du projet:

Comment transformer des données brutes en une information graphique compréhensible par tous les acteurs de l'association ?

Répondre à cette question était un travail passionnant. En effet, pour ce faire, cela a nécessité la mise en application de plusieurs domaines clefs de la profession d'un ingénieur des médias : gestion d'entreprise, communication et technique. Ce travail a suscité un réel intérêt de ma part pour la Business Intelligence et c'est une direction que j'aimerai bien prendre pour la suite de ma carrière professionnelle.

Tables des illustrations

Illustration 1 - Schématisation du GEOPortail (Source : www.asitvd.ch)	5
Illustration 2 - Evolution du nombre de requêtes (Source : présentation de l'AG statutaire 2012).....	6
Illustration 3 - Aperçu du tableau de bord.....	7
Illustration 4 - BI : les étapes théoriques (Source : création perso. inspirée par http://www.osbi.fr/wp-content/bi.png).....	9
Illustration 5 - KPI Wheel (Source : http://www.brightpointinc.com/).....	11
Illustration 6 - Comparaison accès direct et accès par datawarehouse	12
Illustration 7 - Datawarehouse: types d'architectures (Source : création perso. inspirée par http://www.folkstalk.com/2010/01/data-warehouse-dimensional-modelling.html)	13
Illustration 8 - Architecture 3-tiers	15
Illustration 9 - aperçu du résultat	19
Illustration 10 - Transition entre théorie et pratique	21
Illustration 11 - Stratégie de découpe des tableaux de bord.....	23
Illustration 12 - Schéma de l'entrepôt de données.....	27
Illustration 13 - job Talend d'alimentation	31
Illustration 14 - analyse d'un log4j	33
Illustration 15 - Exportation du job d'alimentation	35
Illustration 16 - Job Talend : Access Layer	36
Illustration 17 - Exportation access layer	37
Illustration 18 - temps de réponse du proxy.....	38
Illustration 19 - Tableau de bord transversal.....	39
Illustration 20 - Tableau de bord de la diffusion (sans le menu)	40
Illustration 21 - Utilisateur à vue multiple.....	41
Illustration 22 - Les vues en base de données	42
Illustration 23 - Responsive design en action	44
Illustration 24 - structure des rapports Birt.....	45
Illustration 25 - rapport Birt	46
Illustration 26 - Export XLS Birt	47
Illustration 27 - Lien dans le graphique amCharts	49
Illustration 28 - Exemples de rapport Birt	49
Illustration 29 - Schéma final	50
Illustration 30 - Schéma du travail préparatoire.....	52
Illustration 31 - Divergences avec le TPB	52
Illustration 32 - Options d'un composant Joomla.....	55
Illustration 33 - Heatmap statique.....	56
Illustration 34 - Durée des phases du projet (réalisé avec ADN Projector 2.4.3)	57
Illustration 35 - Répartition temporelle des phases du projet.....	57
Illustration 36 - Diagramme de Gantt (réalisé avec ADN Projector 2.4.3)	58

Les illustrations ne citant aucune source sont des réalisations personnelles.

Webographie

Introduction

- asitvd.ch Présentation de l'association
http://www.asitvd.ch/index.php?option=com_content&view=article&id=91&Itemid=52&lang=fr
- admin.ch Loi fédérale sur la géoinformation
<http://www.admin.ch/ch/f/rs/5/510.62.fr.pdf>

Théorie

- dashboardinsight.com Key Performance Indicators & Metrics
Article très intéressant sur la terminologie propre aux Dashboards
<http://www.dashboardinsight.com/articles/digital-dashboards/fundamentals/key-performance-indicators-and-metrics-choosing-the-right-data-to-display.aspx>
- osbi.fr Concepts de base
Introduction aux BI. L'ensemble du blog est très intéressant.
http://www.osbi.fr/?page_id=112
- business-process-it.com Business Intelligence Data Mining
Conseils clés pour débuter un système BI
<http://www.business-process-it.com/data-mining.html>
- brightpointinc.com KPI Wheel
Société de conseils aillant mis en place la KPI Wheel
<http://www.brightpointinc.com/Articles.asp?File=Dashboard%20Design%20Metrics%20and%20KPIs.htm>
- folkstalk.com Schéma des entrepôts de données
Théorie sur les différents types d'entrepôt de données
<http://www.folkstalk.com/2010/01/data-warehouse-dimensional-modelling.html>
- piloter.org Piloter la performance
Théorie sur les tableaux de bord
<http://www.piloter.org/business-intelligence/business-intelligence.htm>
- gartner.com Magic Quadrant for Business Intelligence Platforms
Etude comparative très complète et d'actualité sur les outils BI actuelles
<http://www.gartner.com/technology/reprints.do?id=1-196WFCB&ct=120207&st=sb>
- mikebernat.com Web-Development, Design, Usability
Théorie sur types de moteurs MySQL
<http://www.mikebernat.com/blog/MySQL - InnoDB vs MyISAM>
- dev.mysql.com Quelles tailles peuvent atteindre les tables MySql
<http://dev.mysql.com/doc/refman/5.0/fr/table-size.html>

Outils

eclipse.org	Birt Project <i>Business Intelligence and Reporting Tools</i> http://www.eclipse.org/birt/phoenix/
talend.com	Talend Open Studio for Data Integration http://fr.talend.com/products/open-studio-di.php
joomla.org	Site officiel Joomla
community.joomla.org	Portail de la communauté
docs.joomla.org	Developing a MVC Component for Joomla!2.5 <i>Tutorial très complet sur la création d'un composant</i> http://docs.joomla.org/Developing_a_Model-View-Controller_(MVC)_Component_for_Joomla!2.5
twitter.github.com	Bootstrap, from Twitter http://twitter.github.com/bootstrap/
easysdi.org	An open source solution for spatial data infrastructure http://easysdi.org
amcharts.com	amCharts <i>Outil de génération de graphiques Javascript</i> http://www.amcharts.com/javascript/

Déclaration formelle

Le soussigné déclare avoir écrit seul le présent rapport et ne pas avoir utilisé d'autres sources que celles indiquées.







































Lieu, date, signature : *Lausanne, le 6 août 2012,* _____

Gaël Bernard


































Annexes

- I. Public cible**
- II. Documentation**
- III. Cahier des charges**
- IV. Procès-verbal du bilan intermédiaire**
- V. CD-ROM**



















Annexe I : Public cible (1/5)

Package	Dashboard sous-jacent	Besoins exprimés	Temps	Source	Output
Transversal  Dashboard <i>Administrateurs</i> <i>Fournisseurs</i>	Diffusion	Temps moyen pour le traitement d'une commande <i>30 derniers jours</i>	 	MySQL	
	Diffusion	Nombre de requête en attente (manuel/auto) excèdent 48h00	 	MySQL	
	Diffusion	Nombre de requêtes en attente (manuel/auto)	 	MySQL	
	Diffusion	Connaître le nombre de GEOproduits diffusés, en attente de diffusion et la forme de	 	MySQL	
	Diffusion	Nombre de requêtes reçues (devis/commande) <i>30 derniers jours</i>	 	MySQL	
	Géoservice	Connaître l'état des alertes sur les services monitorés	 	Web service	
	Géoservice	Connaître le nombre de GEOServices disponibles	 		
	Catalogue	Statistiques des types de ressources cataloguées	 	MySQL	
	Catalogue	Connaître l'évolution du nombre de méta dans le temps	 	MySQL	
	Utilisateur	Afficher le nombre total d'utilisateurs du GEOPortail	 	MySQL	
	Utilisateur	Afficher le nombre d'utilisateurs connectés	 	MySQL	
	Utilisateur	Nombre de visites <i>30 derniers jours</i>	 	Web service	






















Annexe I : Public cible (2/5)

Package	Besoins exprimés	Temps	Source	Output
<p>Diffusion</p>  <p>Dashboard</p> <p>Administrateurs Fournisseurs</p> <p>Range par défaut : 6 derniers mois</p>	<p>Connaître le nombre de GEOproduits diffusés, en attente de diffusion et la forme de diffusion (téléchargement, manuel, auto), public/privé</p>	 		
	<p>Nombre de requêtes (manuel/auto) dont le traitement a excédé 48h00</p>	 		
	<p>Temps moyen pour le traitement d'une commande + répartition du pie par classes suivantes (inférieur à 1h, 1h à 8h, 8h à 24h, 24h à 48h, 48h à 72h, 72h et plus)</p>	 		
	<p>Avoir un top 5 des commandes</p> <ul style="list-style-type: none"> - Titre du produit - Lien sur le produit - Nombre de commandes - % sur total de commandes <p>Pour le reporting, on affiche toutes les commandes.</p> <p>Dimension supplémentaire : Par commande, par devis ou les deux</p>	 		 <p>Reporting</p>
	<p>Avoir un top 5 des utilisateurs déposant le plus de requêtes.</p>	 		 <p>Reporting</p>
	<p>Connaître le nombre de requêtes avec un tiers + % sur requêtes totales</p>	 		 <p>Reporting</p>
	<p>Connaître le nombre de commandes et de devis dans le temps</p>	 		 <p>Reporting</p>
	<p>Connaître l'évolution du nombre de produits dans le temps</p>	 		 <p>Reporting</p>








































Annexe I : Public cible (3/5)

Package	Besoins exprimés	Temps	Source	Output
GEOService  Dashboard <i>Administrateurs</i> <i>Fournisseurs</i> Range par défaut : 6 derniers mois	Connaître le nombre de GEOServices (par type) et leurs proportions + liste complète + états : - fourni par l'ASIT-VD - externe	 information ponctuelle  night synch.	 MySQL +  Web service	
	Connaître le nombre des alertes sur les services monitorés et le type d'erreur (http, etc...)	 range paramétrable  night synch.		 Reporting
	Connaître le nombre d'utilisateurs d'un GEOService.	 range paramétrable  night synch.		 Reporting
	Connaître le temps de réponse des serveurs WMS (liste par serveur)	 range paramétrable  night synch.		 Reporting

Annexe I : Public cible (4/5)

Package	Besoins exprimés	Temps	Source	Output
Catalogue  Dashboard Administrateurs Fournisseurs Range par défaut : 6 derniers mois	Statistiques des types de ressources cataloguées (produit, couche, guichet, etc...) et état (public/privé/publié)	 		
	Connaître l'évolution du nombre de métadonnées dans le temps	 		
	Nombre total de consultations	 		
	Avoir un top 5 des métadonnées les plus consultées permettant de savoir : - Titre de la métadonnée - Lien sur la métadonnée - Nombre de hits - % sur total de hits Pour le reporting, le nombre d'occurrences n'est pas limité.	 		 Reporting
	Avoir un top ten des critères de recherche sur les métadonnées	 		

Annexe I : Public cible (5/5)

Package	Besoins exprimés	Temps	Source	Output
Communication  Dashboard <i>Grand public</i> Range par défaut : dernière année	Afficher le nombre de métadonnées gérées dans le catalogue, par type et par statut (public/privé)	 information ponctuelle  live	 MySQL	 +DATA
	Afficher l'évolution du nombre de métadonnées gérées dans le catalogue, par type (sans distinction des statuts)	 historique  night	 MySQL	
	Afficher le nombre de GEOproduits diffusés via la GEOCommande dans le temps + moyenne par mois	 historique  night	 MySQL	 + DATA
	Nombre de fournisseurs sur la plateforme	 historique  night	 MySQL	
	Nombre d'utilisateurs de la plateforme : - nombre total d'utilisateurs (sans comptes affiliés/avec comptes affiliés) - profil des utilisateurs (administrations, écoles, bureaux)	 historique  night synch. 365j	 MySQL +  Web service	 
	Afficher le nombre de GEOproduits par thématique (ISO)	 historique  night	 MySQL	
	Afficher un heatmap des zones de commande de GEOproduits	 historique  night	 MySQL	
	Afficher le nombre de requêtes (commande/devis) réalisées par mois	 historique  night	 MySQL	
	Afficher le temps moyen pour le traitement d'une commande + répartition du pie par classes suivantes (inférieur à 1h, 1h à 8h, 8h à 24h, 24h à 48h, 48h à 72h, 72h et plus)	 historique  night synch. 365j	 MySQL	

Annexe II : Documentation (1/28)

Réalisation d'un tableau de bord pour le GEOPortail de l'ASIT-VD

Travail de Bachelor 2012

DOCUMENTATION



Sujets traités :

Structure de l'entrepôt de données
Création d'un script d'alimentation de l'entrepôt avec Talend
Création d'une couche d'accès avec Talend
Présentation du composant Joomla
Génération de graphiques avec AmCharts
Génération de rapports Birt

Gaël Bernard

Août 2012

Annexe II : Documentation (2/28)

Réalisation d'un tableau de bord pour le GEOPortail de l'ASIT-VD

Documentation

SOMMAIRE

1	INTRODUCTION.....	3
1.1	OBTENIR LE PROJET.....	3
1.2	CONTACTS.....	3
2	STRUCTURE DE L'ENTREPOT DE DONNEES.....	4
3	CREATION D'UN SCRIPT D'ALIMENTATION DE L'ENTREPOT AVEC TALEND.....	5
3.1	VERSION DE TALEND.....	5
3.2	IMPORTATION DU JOB.....	5
3.3	PRESENTATION DU JOB.....	6
3.4	EXPORT DU JOB.....	10
4	CREATION D'UNE COUCHE D'ACCES AVEC TALEND.....	10
4.1	VERSION DE TALEND.....	10
4.2	IMPORTATION DU JOB.....	10
4.3	PRESENTATION DU JOB.....	11
4.4	EXPORT DU JOB.....	14
4.5	TUTORIAUX.....	15
5	PRESENTATION DU COMPOSANT JOOMLA.....	18
5.1	VERSION DE JOOMLA.....	18
5.2	INSTALLATION DU COMPOSANT.....	18
5.3	PRESENTATION DU COMPOSANT.....	19
5.4	PROXY.....	20
5.5	INTEGRATION A BOOTSTRAP.....	21
5.6	TUTORIAUX.....	21
6	GENERATION DE GRAPHIQUES AVEC AMCHARTS.....	24
6.1	CHOIX DES COULEURS.....	25
7	GENERATION DE RAPPORTS BIRT.....	27
7.1	VERSIONS.....	27
7.2	INSTALLATION DE L'ENGINE REPORT.....	27
7.3	IMPORT DE L'ENVIRONNEMENT.....	27
7.4	ASIT.RPTLIBRARY.....	28

Annexe II : Documentation (3/28)

1 Introduction

La mise en place de solutions permettant la création de tableaux de bord pour le GEOPortail est le contenu d'un projet de Bachelor pour l'obtention d'un titre d'« Ingénieur des médias » orientation IT.

Ce document a pour but de démontrer, en grande partie grâce à des captures d'écran, la mise en place et l'utilisation des différents outils déployés pour le projet. Celui-ci est téléchargeable sur <http://dashboard.gbernard.ch> (sans les bases de données).

Ce document ne présente pas le contexte du projet. Afin d'en savoir plus sur la nature du projet et obtenir le rapport du travail de Bachelor, veuillez contacter M. Xavier Mérour, coordinateur de l'association ou moi-même.

1.1 Obtenir le projet

Cette documentation fait toujours référence au « projet zip ». Il s'agit d'un zip contenant les livrables du projet. Celui-ci est en partie disponible (sans les dump mysql) sur :

<http://dashboard.gbernard.ch>

Pour obtenir les dump mysql veuillez contacter le coordinateur de l'ASIT-VD.

1.2 Contacts

Il est toujours difficile de consacrer le temps nécessaire à la mise en place d'une documentation exhaustive. Cependant, les nombreux commentaires dans le code devraient aider. Si ce n'est pas le cas, je répons volontiers aux questions.

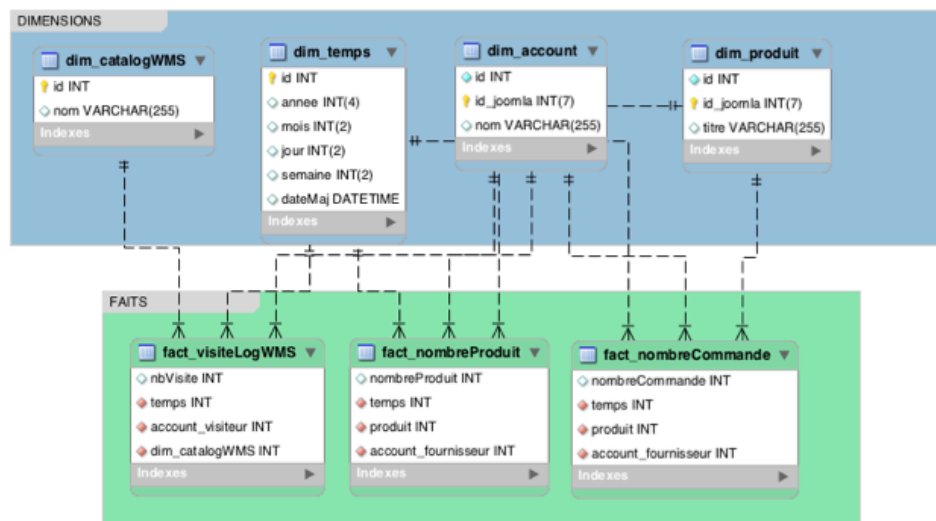
<p><i>Pour me contacter :</i> Gaël Bernard contact@gbernard.ch www.gbernard.ch</p>	<p><i>Coordinateur de l'ASIT-VD</i> Xavier Mérour xavier.merour@asit.vd.ch www.asitvd.ch</p>
--	---

Annexe II : Documentation (4/28)

2 Structure de l'entrepôt de données

projet zip	
« projet zip » /01_datawarehouse/DB/modélisation.mwb	Contient la modélisation de la table ci-dessous
« projet zip » /01_datawarehouse/DB/export_table.sql	Script sql de création des tables
« projet zip » /01_datawarehouse/DB/export_row.sql	Contient les enregistrements (pour des raisons de confidentialité, cette partie n'est pas sur le site)

L'entrepôt de données n'est finalement qu'une base de données standard mais avec quelques spécificités. Celui-ci a été créé avec Mysql.



On remarque d'abord que le schéma est séparé en deux parties : les tables de dimensions et les tables de faits. Les tables de dimensions possèdent les descriptions (nom du produit, nom du fournisseur, numéro de la semaine) et les tables de faits possèdent les données factuelles (nombre de visite, nombre de produit vendus,...). Les tables de faits sont destinées à contenir un grand nombre d'enregistrements et, puisque ces tables ne contiennent que des données empiriques et des liens sur d'autres tables, elle reste légère.

Une deuxième spécificité d'un entrepôt de données réside dans des liens simples entre les tables. En effet, on évite au maximum de lier une table de dimension à une autre table de dimension (c'est faisable avec le schéma en flocon, mais c'est utile seulement si on peut économiser des milliers d'enregistrements). La redondance des informations n'est pas rédhibitoire si on peut améliorer la vitesse d'accès à l'information (et c'est une grosse différence avec une modélisation classique).

Encore une spécificité réside dans le type d'accès. Dans une base de données traditionnels, on utilise tout type de requête (insert, delete, update,...). Dans un Datawarehouse, on fait en général un INSERT pour populer les tables et un SELECT lorsqu'on veut obtenir des statistiques. C'est d'ailleurs la raison pour laquelle le moteur utilisé est MyISAM.

Annexe II : Documentation (5/28)

3 Création d'un script d'alimentation de l'entrepôt avec Talend

« projet zip »*	
/01_datawarehouse/talend/logs_tests/	Contient des exemples de logs qui ont été testés
/01_datawarehouse/talend/resultat_job_export/	Résultat du job exporté en autonomous job puis décompressé. On y trouve les script .sh et .bat à lancer
/01_datawarehouse/talend/talend.zip	Le projet source à importer

* Référence au chapitre 1.1

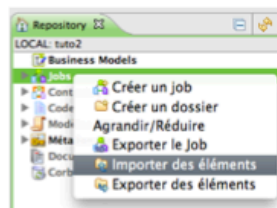
3.1 Version de Talend

Le job a été réalisé avec "Talend Open Studio for Data Integration". Version: 5.1.1.

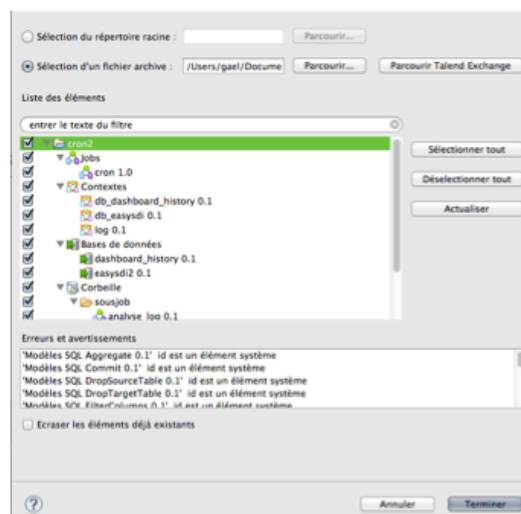
<http://fr.talend.com/products/open-studio-di.php>

3.2 Importation du job

1. La première étape consiste à créer un nouveau job vide.
2. Ensuite, suivez les captures d'écran :



Clique droit sur Jobs puis importer des éléments



On sélectionne le fichier d'archive /01_datawarehouse/talend/talend.zip de « projet zip ».
Sélectionnez tous les éléments !

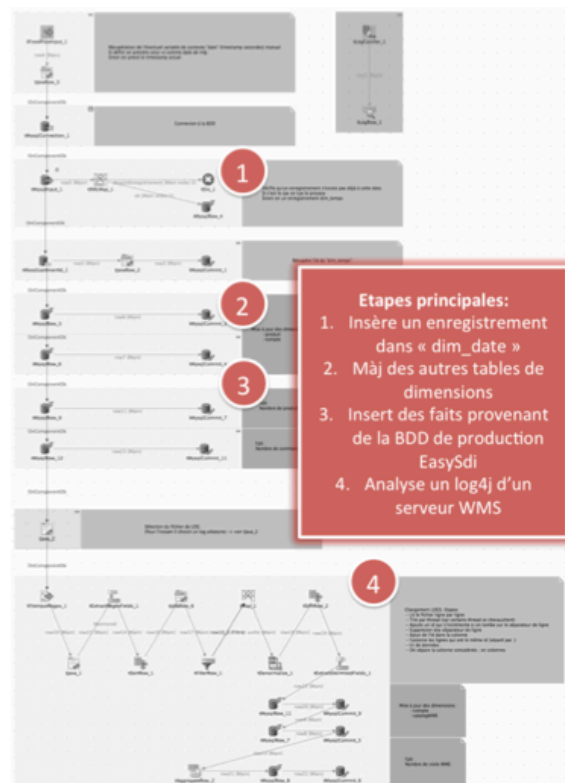
Annexe II : Documentation (6/28)

3.3 Présentation du job

Le projet d'alimentation du DWH ne contient qu'un seul job

Les variables de contextes sont essentielles. Les deux premiers 'db_' contiennent les informations de connexion à la DB. D'ailleurs, vous devez probablement les modifier pour les faire correspondre à l'environnement. log contient lui les informations sur le chemin des logs à analyser

Informations de connexion à la DB. Fait simplement référence aux contextes ci-dessus.



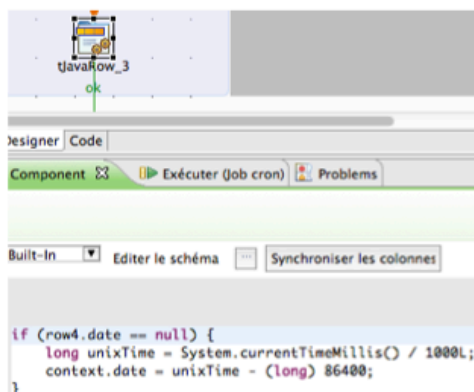
Les étapes suivantes suivent scrupuleusement les pastilles rouges de l'image précédente, veuillez vous y référer.

Annexe II : Documentation (7/28)

3.3.1 Etape 1 : enregistrement dans « dim_dates »

La première étape consiste à choisir la date à laquelle on désire faire l'enregistrement. Ici le choix a été fait de prendre le jour précédant durant lequel est lancé le job. Avec cette configuration, le job devrait être lancé après minuit.

C'est l'élément tjavaRow_3 qui définit le timestamp :



On remarque qu'il prend le timestamp actuel et qu'il retire 86400 sec (1 journée). La seconde exact n'est pas importante car après il sélectionne le jour en cours de cette manière :

```
WHERE YEAR(FROM_UNIXTIME("+context.date+")) = YEAR(dateMaj)
AND MONTH(FROM_UNIXTIME("+context.date+")) = MONTH(dateMaj)
AND DAY(FROM_UNIXTIME("+context.date+")) = DAY(dateMaj)
```

3.3.2 Etape 1 : enregistrement dans « dim_account »

Le but de cette étape est de mettre à jour la table de dimension dim_account (par ex : détecter de nouveaux clients)

Mis à jour des comptes utilisateurs

```
INSERT INTO dim_account (id_joomla,nom)
SELECT id, name
FROM easysdi2_jos_users
ON DUPLICATE KEY UPDATE nom= name
```

La requête ci-dessus est un peu particulière mais très pratique. On insère dans la table de dimension « dim_account » l'id_joomla et le nom des personnes présent dans la table « jos_users » d'easysdi. Cependant, on lance ce job quotidiennement donc on imagine facilement le nombre de doublon possible. La dernière opération « ON DUPLICATE KEY » permet justement d'éviter les doublons. Le « KEY » de « ON DUPLICATE KEY » est la clé primaire de la table « dim_account » donc c'est la colonne « id_joomla ». Donc si il détecte que la clé id_joomla existe déjà, il met à jour le nom. Cette simple opération permet donc de d'insérer les nouveaux clients (aucun doublon) et de mettre à jour les clients qui auraient changer de nom (duplicate key). On fait cette opération avec toutes les tables.

Annexe II : Documentation (8/28)

3.3.3 Etape 3 : Insère les faits

L'objectif est maintenant d'alimenter la table des faits. C'est une opération un peu plus complexe que pour les tables de dimensions. L'exemple ci-dessous compte le nombre de commandes effectuées par fournisseur.

Fait : nombre de commandes du jour d'un fournisseur

```
INSERT INTO fact_nombreCommande (nombreCommande, temps, produit, account_fournisseur)
SELECT count(DISTINCT jos_sdi_order.id) as nbCommande, dim_temps.id, dim_produit.id, dim_account.id_joomla
FROM "+context.easysdi2_Database+".jos_sdi_order
INNER JOIN dim_temps ON dim_temps.id = "+context.last_insert_id+"
INNER JOIN "+context.easysdi2_Database+".jos_sdi_order_product ON jos_sdi_order_product.order_id = jos_sdi_order.id
INNER JOIN "+context.easysdi2_Database+".jos_sdi_product ON jos_sdi_order_product.product_id = jos_sdi_product.id
INNER JOIN "+context.easysdi2_Database+".jos_sdi_objectversion ON jos_sdi_objectversion.id =
jos_sdi_product.objectversion_id
INNER JOIN "+context.easysdi2_Database+".jos_sdi_object ON jos_sdi_object.id = jos_sdi_objectversion.object_id
INNER JOIN dim_produit ON dim_produit.id_joomla = "+context.easysdi2_Database+".jos_sdi_product.id
INNER JOIN "+context.easysdi2_Database+".jos_sdi_account ON jos_sdi_account.id = jos_sdi_object.account_id
INNER JOIN "+context.easysdi2_Database+".jos_users ON jos_users.id = jos_sdi_account.user_id
LEFT JOIN "+context.easysdi2_Database+".jos_sdi_account as jos_sdi_account_with_root ON jos_sdi_account_with_root.id =
jos_sdi_account.root_id
LEFT JOIN "+context.easysdi2_Database+".jos_users as jos_users_with_root ON jos_users_with_root.id =
jos_sdi_account_with_root.user_id
INNER JOIN dim_account on dim_account.id_joomla = COALESCE(jos_users_with_root.id,jos_users.id)
WHERE jos_sdi_order.type_id = 1
AND YEAR(jos_sdi_order.created) = dim_temps.annee
AND MONTH(jos_sdi_order.created) = dim_temps.mois
AND DAY(jos_sdi_order.created) = dim_temps.jour
GROUP BY jos_sdi_product.id, dim_account.id
```

Count(Distinct jos_sdi_order.id) permet de compter le nombre de commandes différentes. Il est important de grouper ensuite avec les dimensions utilisées (sauf le temps). On compte alors le nombre de commande par produit et par account.

INNER JOIN dim_temps ON dim_temps.id = '+context.last_insert_id»+' permet de faire correspondre l'enregistrement effectué dans dim_temps et ainsi sélectionner les enregistrement qui date du jour.

3.3.4 Etape 4 : Analyse d'un log4j.

L'analyse d'un log4j n'a pas été une tâche facile. En effet, ce type de log n'est pas destiné à de l'analyse informatique. Il est possible de configurer le fichier de config d'un log4j afin de lui ordonner de créer une sortie XML beaucoup plus facile à analyser avec Talend. Cependant il a été décidé de ne pas créer ce type de fichier afin de ne pas augmenter la charge serveur.

Voici un copier-coller d'une partie du rapport expliquant étape par étape l'analyse de ce log (veuillez consulter le rapport pour les tailles originales):

Exemple détaillé : analyse d'un log4j

Cet exemple permet de montrer l'utilisation de Talend. Celui-ci analyse un fichier de log (log4j) de plus de 7mo et en retire les informations demandées. Le but est d'obtenir par qui et combien de fois un service WMS a été utilisé. Le processus prend environ 1,7 seconde. Cet exemple est intéressant car il s'agit d'un fichier de log très verbeux qui n'est en principe pas destiné à de l'analyse informatique. Voici le cheminement nécessaire :



Etape 1 : Ouverture du log et lecture ligne par ligne.

Annexe II : Documentation (9/28)



Etape 2 : Application d'une expression régulière permettant d'isoler les lignes intéressantes et de les découper en colonnes. On voit que le fichier est déjà nettement plus facile à traiter.

Processus	user	id	ip
Processus: 4431	Config		
Processus: 4711	user		
Processus: 4811	user		
Processus: 4821	Config		
Processus: 504	user		
Processus: 5221	user		
Processus: 5221	Config		
Processus: 504	user		
Processus: 5221	user		
Processus: 504	user		
Processus: 5221	user		
Processus: 5221	user		

Etape 3 : La colonne des processus est triée permettant de ne pas les mélanger. Les paires config et user sont maintenant bien isolées.

Processus	user	id	ip
Processus: 4431	Config		
Processus: 4811	user		
Processus: 4821	Config		
Processus: 504	user		
Processus: 5221	user		
Processus: 4431	Config		
Processus: 4811	user		
Processus: 4821	Config		
Processus: 504	user		
Processus: 5221	user		
Processus: 4431	Config		
Processus: 4811	user		
Processus: 4821	Config		
Processus: 504	user		
Processus: 5221	user		

Etape 4 : Une colonne id est ajoutée. Celle-ci s'incrémente chaque fois qu'on trouve un séparateur de ligne (-----). Ensuite on supprime ces lignes de séparation et on concatène en une seule ligne les éléments possédant les mêmes id. Le résultat est obtenu : chaque ligne contient les informations demandées.

id	[ip] ip
0	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
1	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
2	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
3	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
4	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
5	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
6	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
7	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
8	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6
9	ipnetwork ,ipr3ng0a24595e6ed9a0b24f027f2b63b134d6

Fin de l'analyse du log4j : Seuls deux éléments ont été isolés. Cependant cette découpe du fichier de log permet aisément d'isoler de nouvelles informations avec la même sortie. La prochaine étape est de mettre à jour les tables de dimensions si on découvre des éléments qui n'y étaient pas encore puis on popule la table des faits.

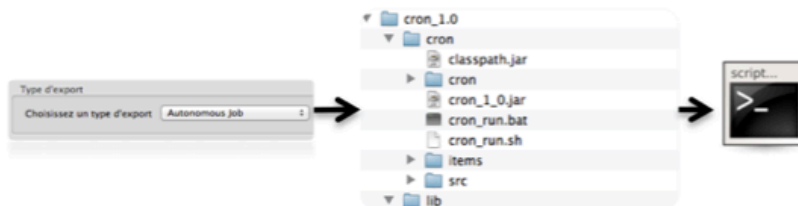
Comme on peut le voir sur les brèves explications ci-dessus, l'analyse d'un fichier de log4j n'est pas aussi facile qu'on aurait pu l'imaginer. Cependant cette démarche ci-dessus fonctionne plutôt bien. Après avoir analysé tous les logs fournis « /01_datawarehouse/talend/logs_tests/ » aucun n'a posé de problème.

Il y a juste une particularité qu'il faut savoir au niveau de la table de dimension « dim_account ». Le fichier de log ne contient que le nom correspondant à l'account. Si on retrouve ce nom dans la table « dim_account » on utilise l'id correspondant à cette enregistrement. Cependant, il y a aussi des comptes machines par ex : « spring2a2d595e6ed9a0b24f027f2b63b134d6 » qui ne correspondent à rien dans la table « dim_account ». Il a donc été décidé que si un compte ne correspond à aucun enregistrement, on l'insère dans la table dim_account mais avec un id négatif pour id_joomla.

Annexe II : Documentation (10/28)

3.4 Export du job

Ce job est exporté en tant que job autonome. Il peut ensuite être lancé grâce aux fichiers « .sh » ou « .bat » selon l'OS.



Le cron peut ensuite être lancé comme suit :

Popule le DWH pour la veille.

```
sh cron_run.sh ;
```

Popule le DWH pour le 4/8/2012

```
sh cron_run.sh --context_param date=1344071785
```

4 Création d'une couche d'accès avec Talend

« projet zip »*	
/02_accessLayer /talend/resultat_job_export/	Résultat du job exporté en tant que web service (.war)
/02_accessLayer /talend/talend.zip	Le projet source à importer

** Référence au chapitre 1.1*

4.1 Version de Talend

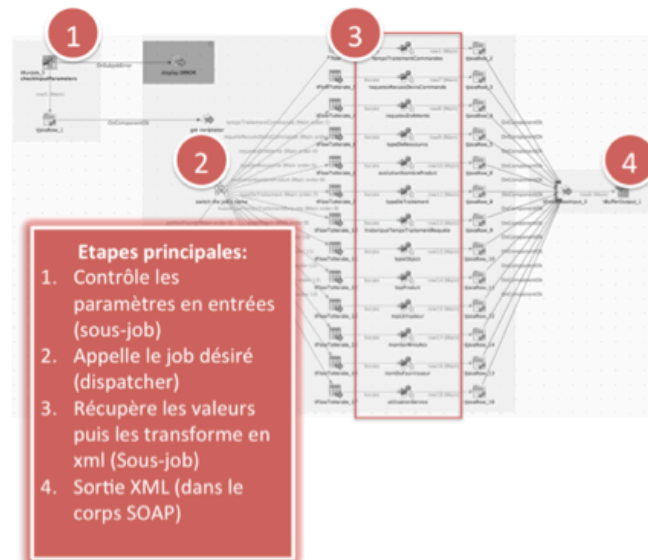
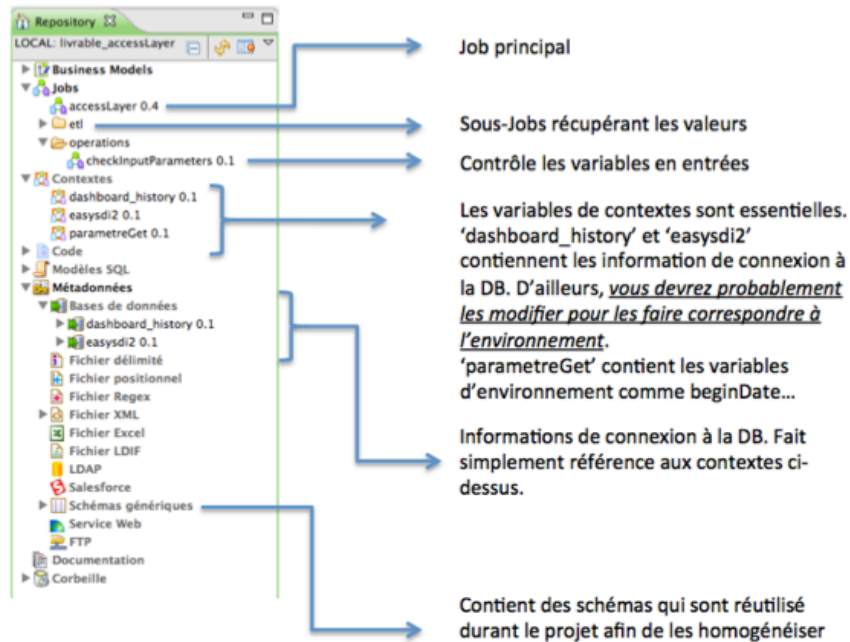
Identique au chapitre 3.1

4.2 Importation du job

Identique au chapitre 3.2

Annexe II : Documentation (11/28)

4.3 Présentation du job



Les étapes suivantes sont en liens avec l'illustration précédente (pastille rouge). Veuillez vous y référer.

Annexe II : Documentation (12/28)

4.3.1 Contrôle les paramètres en entrées

C'est la première opération qui est effectuée. Tout job passe donc par cette étape. Elle permet de contrôler les paramètres en entrées : par exemple, si la limite n'est pas définie, on récupère la limite par défaut définie dans le contexte de l'application

The screenshot shows the Talend Open Studio interface. At the top, a job diagram is visible with components like tFixedFlowInput_1, tJavaRow_4, tJavaRow_1, tJavaRow_3, tLogRow_4, and tBufferOutput_2. Below the diagram, the code editor shows the following code for the tJavaRow_1 component:

```
//Code généré selon le schéma d'entrée et de sortie
output_row.beginDate = input_row.beginDate;
output_row.endDate = input_row.endDate;
output_row.limit = input_row.limit;
output_row.job = input_row.job;
output_row.idFournisseur = input_row.idFournisseur;

if (input_row.limit != null) {
    output_row.limit = input_row.limit;
}

//Si la limite n'est pas définie on prend celle par défaut
else {
    output_row.limit = context.defaultLimit;
}
```

Nom	Type	Script code
beginDate	int Integer	context.beginDate
defaultLimit	int Integer	context.defaultLimit
endDate	int Integer	context.endDate
idFournisseur	int Integer	context.idFournisseur
job	String	context.job
limit	int Integer	context.limit
xmlOutput	String	context.xmlOutput

contexte : parametreGet

4.3.2 Dispatcher

Le dispatcher peut être considéré comme un Switch classique en programmation. On test si le paramètre « job » (String) correspond à un job existant. La sortie jobNotFound à le paramètre « Catch output reject ===true » donc si aucun job ne correspond au texte envoyé, on renvoie sur la sortie jobNotFound.

The screenshot shows a dispatcher configuration in Talend Open Studio. On the left, a table named 'row3' is defined with columns: beginDate, endDate, limit, job, and idFournisseur. On the right, a 'jobNotFound' component is configured with the following properties:

- Property: Catch output reject, Value: true
- Property: Catch lookup inner job, Value: false
- Property: Schema Tvoe, Value: Built-In
- Property: row3.job.equals('tempsTraitementC', Value: true

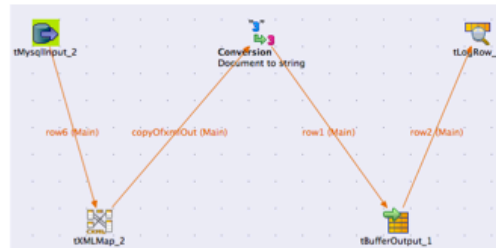
Below these properties, an expression is defined: 'Job "" + row3.job + "" has not message'. At the bottom, a table shows the mapping of input columns to output columns:

Expression	Column
row3.beginDate	beginDate
row3.endDate	endDate
row3.limit	limit
row3.job	job

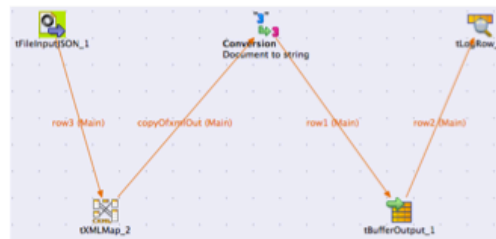
Annexe II : Documentation (13/28)

4.3.3 Récupère les valeurs

Un job de récupération de valeur aura toujours la même forme : On effectue l'opération (requête sql, lecture d'un fichier flat, appel à un webservice), on récupère les colonnes qu'on transforme en XML grâce à TXMLMap, on transforme en texte et on renvoie le tout en sortie grâce au buffer.



Interrogation d'une BDD (job tempsTraitementCommande)



Appel à un web service (job monitorWmsAcv)

4.3.4 Sortie XML

Cette dernière opération récupère le texte (xml transformé en string) et l'envoie en sortie.

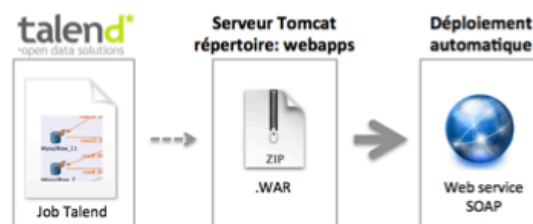
Annexe II : Documentation (14/28)

4.4 Export du job

Extrait du rapport

Le job de la couche d'accès n'est pas exporté en tant que job autonome mais en tant que webservice. Ceci est une grande force de Talend : un job peut être exporté en tant que fichier « .war » prêt à être déployé sur un serveur java.

Le schéma ci-dessous illustre le déploiement d'un job Talend en tant que web service. La flèche entre Talend et le serveur Tomcat est volontairement fine pour indiquer que le lien entre Talend et le serveur Tomcat est faible. En effet, mis à part le fichier compressé « .war » à déposer dans le dossier WebApps de Tomcat, aucun élément ne les lie. Ce qui implique qu'il y a une grande flexibilité entre les outils.



Tomcat déploie automatiquement le fichier « .war » sur le serveur. Cela génère ensuite un répertoire qui est l'entrée de notre web service.

Pour accéder au web service, on appelle l'url suivante :

```
http://[serveurTomcat]/accessLayer_0.4/services/accessLayer?method=runJob
&arg1=--context_param%20job=topUtilisateur
&arg2=--context_param%20beginDate=0
&arg3=--context_param%20endDate=1343684565
&arg4=--context_param%20idFournisseur=22388
```

L'url ci-dessus retourne un résultat semblable à celui-ci :

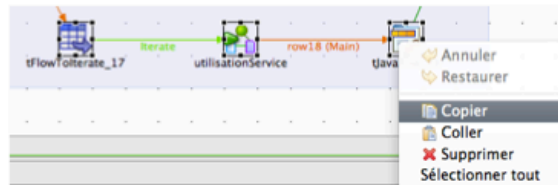
```
<?xml version="1.0" encoding="UTF-8"?> <graph><label>Top utilisateurs</label><serie><data>1075</data><label>Jaqui
Officiels</label></serie><serie><data>969</data><label>EFA+C Ingénieurs géomètres officiels SA</label></serie><se
& amp; Associés. Ingénieurs et Géomètres S.A.</label></serie><serie><data>878</data><label>Administration cantonal
<data>874</data><label>BBHS SA</label></serie><serie><data>821</data><label>Mosini et Caviesel Ingénieurs Géomètre
<data>650</data><label>Schenk SA Ingénieur Géomètre Officiel</label></serie><serie><data>639</data><label>Nicod &
SA</label></serie><serie><data>622</data><label>Bovard & amp; Nickl SA Ingénieur Géomètre Officiel & amp; Ingénieur
<data>614</data><label>B+C Ingénieurs SA</label></serie><serie><data>570</data><label>DTP SA Bureau d'études</lab
<label>Renaud et Burnand SA Ingénieurs Géomètres Officiels</label></serie><serie><data>548</data><label>Duchoud-H
```

Afin d'en savoir plus l'utilisation de ce web service, je vous invite à découvrir la partie proxy du composant Joomla. C'est lui qui appelle le web service et qui évalue le résultat.

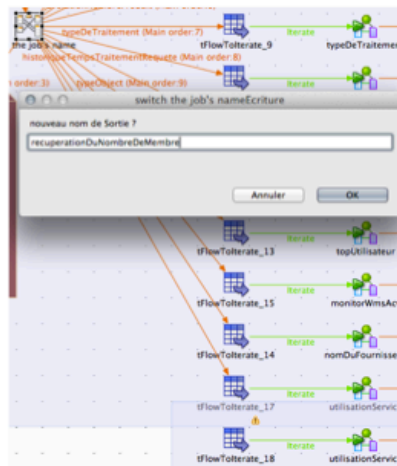
Annexe II : Documentation (15/28)

4.5 Tutoriaux

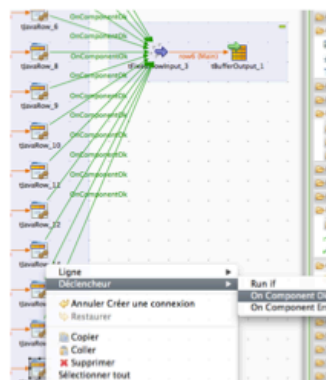
4.5.1 Ajout d'un nouveau job



Copiez-collez les 3 éléments de la dernière ligne dans le job accessLayer
(Point 3 lors de la présentation du job dans ce chapitre)

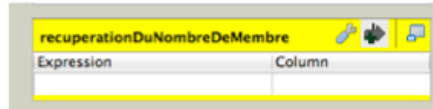


Cliquez droit sur le tmap « switch the job'name » puis sélectionnez « new Output ». Une flèche apparaît. Relié là au TFlowTolterate fraîchement collé. Un prompt demande le nom de sortie. Insérer le nom du job (copier-coller le nom du job, il sera utile à maintes reprises)

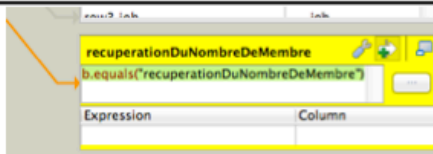


Cliquez droit sur le tjavaRow fraîchement collé.
Sélectionné Déclencheur -> On Component OK puis relié la flèche au TfixedflowInput

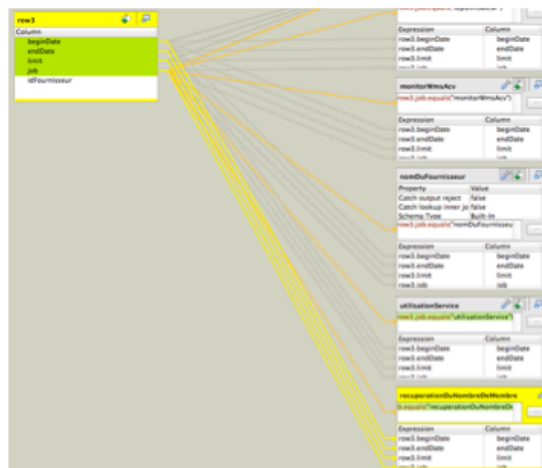
Annexe II : Documentation (16/28)



Double-cliquez sur le tmap « switch the job name ». Une fenêtre s'ouvre. Descendez l'ascenseur pour voir l'élément ci-dessus (le seul élément qui n'est pas relié). Cliquez sur la flèche pointant sur la droite (capture d'écran)



Tapez « row3.job.equals("recuperationDuNombreDeMembre") » dans le champ texte.
Il faut évidemment modifier recuperationDuNombreDeMembre par le nom donné à votre sortie.

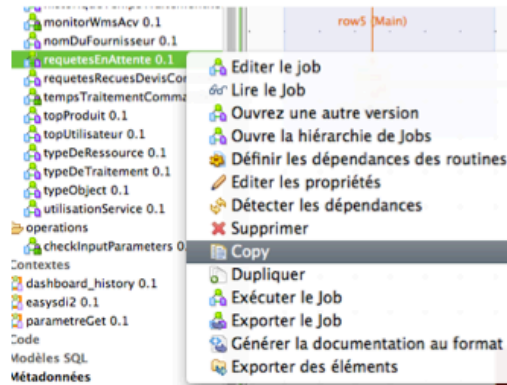


Faites glissez les quatre champs beginDate, endDate, limit et job dans la sortie nouvellement créé. Ceci n'est pas obligatoire mais c'est utile pour faire du débogage. Lorsqu'on appelle le job, on peut ainsi voir les valeurs de ces paramètres. Honnêtement je suis en train de me demander si ce lien n'est pas superflu. Malheureusement je n'ai pas le temps de faire des tests (dossier à rendre demain). A voir...

Annexe II : Documentation (17/28)

Réalisation d'un tableau de bord pour le GEOPortail de l'ASIT-VD

Documentation



Copiez-collez maintenant un job du dossier ETL qui fait une opération semblable à celle que vous voulez faire. Pour info : monitorWmsAcv interroge un web service, evolutionNombreProduit interroge le DWH et les autres job interrogent la base de production EasySdi.. Après copiez-collez le job, renommez-le par le nom donné à votre sortie (dans mon cas *recuperationDuNombreDeMembre*). Ouvrez la job et modifié la requête ou l'appel au web service pour récupérer les infos souhaitées.



Revenez sur le job AccessLayer.
Faîtes un **simple** click sur le job (ici utilisation service)



Dans l'onglet « Component » cliquez sur le bouton {...} et sélectionner le job que vous venez de créer.

Voilà, on a maintenant une nouvelle sortie dans le TMap. Si on appelle le web service avec le bon nom de job, TMap redirige sur le job nouvellement créé et envoie en sortie le XML. Ce n'est pas forcément simple d'expliquer par écrit mais une fois que vous avez effectué l'opération en entier, vous remarquerez que c'est finalement très simple.

Annexe II : Documentation (18/28)

5 Présentation du composant Joomla

« projet zip »*	
/03_cms /joomla/resultat_job_export/com_easysdi_dashboard.zip	Composant à installer

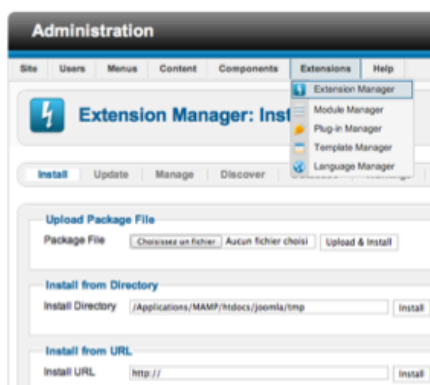
**Référence au chapitre 1.1*

5.1 Version de Joomla

Le composant a été développé pour Joomla 2.5 mais avec une intégration de bootstrap en vue d'une intégration à Joomla 3.5 (qui était malheureusement en train d'être conçu durant ce travail).

5.2 Installation du composant

C'est une installation tout à fait standard de Joomla.

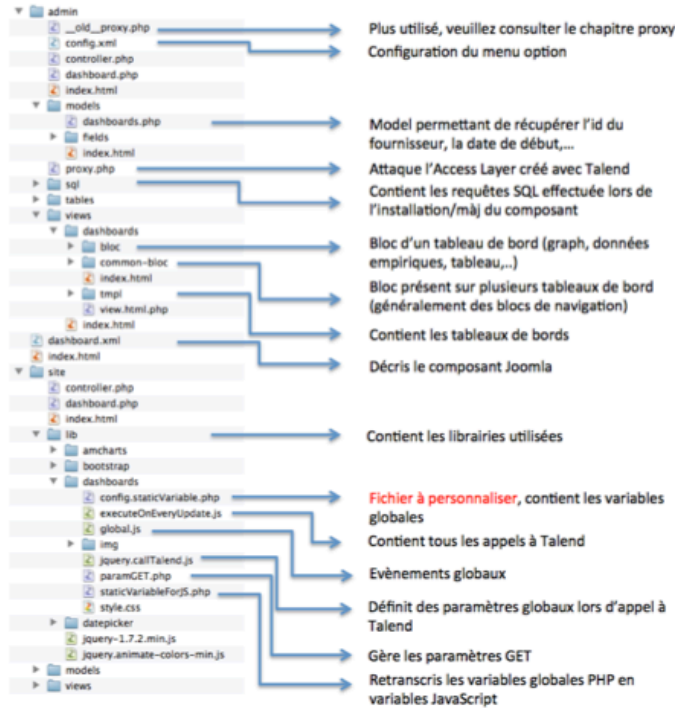


Dans l'extension Manager choisissez le fichier
/03_cms /joomla/resultat_job_export/com_easysdi_dashboard.zip

Cliquez ensuite sur upload & install. C'est fait !

Annexe II : Documentation (19/28)

5.3 Présentation du composant



Pour l'instant, le composant a été développé pour la partie Back-end. La partie front-end peut facilement être créé en se basant sur la partie développée. Il y a certainement quelques informations à modifier dans le fichier `config.staticVariable.php`.

5.3.1 Base de données liées

Une seule table est créée lors de l'installation du composant. Il s'agit de la table `__SUFIXE__dashboard`

id	name	displayName	description	ordering	access
1	transversal	Transversal	Tableaux de b...	0	1
2	diffusion	Diffusion	Tableaux de b...	1	1
3	geoservice	GEOService	Tableaux de b...	2	1
4	catalogue	Catalogue	Tableaux de b...	3	1
5	communication	Communication	Tableaux de b...	4	1
6	topSecret	Top Secret	Tableaux de b...	5	4

Annexe II : Documentation (20/28)

5.4 Proxy

Le proxy est un élément important. C'est lui qui fait appel au web service déployé grâce à la couche d'accès. Cette couche d'accès retourne un résultat SOAP. Et le xml se trouve dans le body de cette requête.

Sans proxy

Par exemple :

http://localhost:8080/accessLayer_0.4/services/accessLayer?method=runJob&arg1=--context_param%20number=4&arg2=--context_param%20job=topUtilisateur&arg3=--context_param%20beginDate=0&arg4=--context_param%20endDate=1343684565&arg5=--context_param%20idFournisseur=22388

Retourne :

```
This XML file does not appear to have any style information associated with it. The document tree is shown below.
<?xml version="1.0" encoding="UTF-8"?>
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <soapenv:body>
    <runJobReturn xmlns="http://talend.org" xmlns:xsi="http://talend.org" xsi:type="xsd:string">
      <xml:comment xsi:type="xsd:string">
        <?xml version="1.0" encoding="UTF-8"?>
        <graph>
          <label>Top utilisateurs</label>
          <serie>
            <data>1575</data>
            <label>Jaquier
              Gémètres Officiels</label>
            </serie>
            <serie>
            <data>989</data>
            <label>EER</label>
            </serie>
            <serie>
            <data>915</data>
            <label>Courdesse
              & Associés. Ingénieurs et Gémètres S.A.</label>
            </serie>
            <serie>
            <data>878</data>
            <label>Administration cantonale vaudoise
              [</label>
            </serie>
            <serie>
            <data>878</data>
            <label>BBHN SA</label>
            </serie>
            <serie>
            <data>852</data>
            <label>Mosini et Caviezol
              Ingénieurs Gémètres Officiels</label>
            </serie>
            <serie>
            <data>852</data>
            <label>Schenk SA II
              Officiel</label>
            </serie>
            <serie>
            <data>439</data>
            <label>Nicod + Perrin
              Ingénieurs et Gémètres SA</label>
            </serie>
            <serie>
            <data>414</data>
            <label>Bovard & Nickl SA
              Ingénieur Gémètre Officiel & Ingénieur Civil</label>
            </serie>
            <serie>
            <data>414</data>
            <label>E&G
              Ingénieurs SA</label>
            </serie>
            <serie>
            <data>370</data>
            <label>DTP SA
              Bureau d'Etudes</label>
            </serie>
            <serie>
            <data>366</data>
            <label>Bureau
              de gémètres Officiels</label>
            </serie>
            <serie>
            <data>357</data>
            <label>Duchoud-Haymoz-Buhlmann SA
              Ingénieurs Gémètres Officiels</label>
            </serie>
            <serie>
            <data>357</data>
            <label>Renaud et Burnand SA
              Ingénieurs Gémètres Officiels</label>
            </serie>
            <serie>
            <data>357</data>
            <label>Richard & Cardinaux SA
              Ingénieurs Gémètres Off
             iciels</label>
            </serie>
            <serie>
            <data>299</data>
            <label>Par
              & Jomini-Van Buel SA </label>
            </serie>
            <serie>
            <data>280</data>
            <label>Richard & Cardinaux SA
              Ingénieurs Gémètres Officiel</label>
            </serie>
            <serie>
            <data>245</data>
            <label>Commune Lausanne
              [</label>
            </serie>
            <serie>
            <data>225</data>
            <label>Bureau
              technique Stuby & Vu00ff
              d'Etudes techniques</label>
            </serie>
            <serie>
            <data>196</data>
            <label>Lehmann Ingénieur
              Gémètre Officiel</label>
            </serie>
            <serie>
            <data>162</data>
            <label>Belotti Ingénieur
              Gémètre Officiel</label>
            </serie>
            <serie>
            <data>154</data>
            <label>Compagnie
              Commerciale du Gas SA
              (CICG)</label>
            </serie>
            <serie>
            <data>147</data>
            <label>Wautier Jacques
              Ingénieur gémètre offic
              </label>
            </serie>
            <serie>
            <data>129</data>
            <label>Perrini Gémétrique
              s&lt;/label>
            </serie>
            <serie>
            <data>119</data>
            <label>Marblan Ingénie
              r</label>
            </serie>
            <serie>
            <data>111</data>
            <label>Delacrétas Bernard
              Ingénieur gémètre officiel</label>
            </serie>
            <serie>
            <data>97</data>
            <label>C&D
              INGENIEURS SA</label>
            </serie>
            <serie>
            <data>99</data>
            <label>Claude Thurier
              Ingénieurs SA</label>
            </serie>
            <serie>
            <data>77</data>
            <label>C&D
              Architecture SA</label>
            </serie>
            <serie>
            <data>72</data>
            <label>CFP SA,
              Lausanne [</label>
            </serie>
            <serie>
            <data>59</data>
            <label>W. & C.
              Pillonel</label>
            </serie>
            <serie>
            <data>47</data>
            <label>Berter &
              Wisemann Ingénieurs
              Consoils SA</label>
            </serie>
            <serie>
            <data>45</data>
            <label>
          </graph>
        </xml:comment>
      </runJobReturn>
    </soapenv:body>
  </soapenv:Envelope>
</pre>
```

Avec Proxy

Par exemple :

http://localhost:8888/joomla/administrator/components/com_dashboard/proxy.php?limit=1000&beginDate=1280872800&idFournisseur=22388&job=topUtilisateur

Retourne :

```
{ "label": "Top utilisateurs", "serie": [{"data": "756", "label": "Jaquier - Pointet Ing'u00e9nieurs Gu00e9om'u00c9mètres officiels SA"}, {"data": "617", "label": "Courdesse & Associu00e9s. Ing'u00e9nieurs et cantonale vaudoise []"}, {"data": "558", "label": "BBHN SA"}, {"data": "552", "label": "Mosini et Caviezol Ing"}, {"data": "477", "label": "Nicod + Perrin Ing'u00e9nieurs et g'u00e9om'u00e9mètres SA"}, {"data": "467", "label": "Ing'u00e9nieur Civil"}, {"data": "431", "label": "Schenk SA Ing'u00e9nieur Gu00e9om'u00e9mètres Officiel"}, {"data": "366", "label": "DTP SA Bureau d'u00e9tudes"}, {"data": "357", "label": "Renaud et Burnand SA Ing"}, {"data": "338", "label": "Duchoud-Haymoz-Buhlmann SA Ing'u00e9nieurs Gu00e9om'u00e9mètres Officiels"}, {"data": "308", "label": "Olivier Peitrequin SA"}, {"data": "294", "label": "Par"}, {"data": "288", "label": "Jomini-Van Buel SA"}, {"data": "280", "label": "Richard & Cardinaux SA Ing'u00e9nieur Gu00e9om'u00e9mètre Officiel"}, {"data": "272", "label": "ORCEF SA"}, {"data": "254", "label": "Boss Philippe Bureau de g'u00e9om'u00e9mètre officiel"}, {"data": "177", "label": "G&V"}, {"data": "175", "label": "Commune Lausanne []"}, {"data": "171", "label": "Bureau technique Stuby & Vu00ff d'u00e9tudes techniques"}, {"data": "162", "label": "Lehmann Ing'u00e9nieur g'u00e9om'u00e9mètre officiel"}]
```

Le résultat avec le proxy est bien plus digeste pour du Javascript car il s'agit du JSON (plus léger) et dans l'enveloppe de l'appel Soap. Notez qu'il est également possible d'obtenir du XML en ajoutant &format=xml à l'url (Birt préfère le XML).

Le rôle du proxy est donc d'interpréter le résultat de l'appel SOAP et d'éliminer les éléments d'en-tête non nécessaire.

Annexe II : Documentation (21/28)

5.4.1 Différence `__old_proxy.php` et `proxy.php`

Initialement, l'appel au web service Talend se faisait grâce à la classe SoapClient. C'était une façon très propre de procéder car on utilisait le protocole SOAP. Cependant, chaque appel prenait presque 3 secondes.

Le test avec un simple `get_file_content()` à donné des résultats nettement meilleures.

La deuxième solution est donc préférée pour une meilleure performance. `__old_proxy.php` pourrait être supprimé mais j'ai pensé que celui-ci pourrait être intéressant si un développeur veut découvrir les raisons de cette lenteur.

5.5 Intégration à Bootstrap

Joomla intégrera Bootstrap de Twitter dans la partie back-end de l'application. Celle-ci n'étant encore qu'à un stade d'ébauche à l'heure où j'écris ces lignes, Bootstrap a simplement été intégré au composant 2.5. Cependant les choses n'ont pas été forcément simple dû à des conflits avec les styles CSS.

Je vous invite à ouvrir le fichier `site/lib/bootstrap/css/warningCustomBootstrap.css`.

Cette feuille de style CSS est simplement la feuille standard disponible avec Bootstrap mais avec quelques « hack » afin d'éviter les conflits CSS.

```
/*
 * Pas mal de collisions de styles entre Joomla et bootstrap fait que ce fichier a été édité
 * manuellement. Ces modifications concernent:
 * - Suppression de toutes les classes ".modal2" qui posait un problème lors de l'ajout d'un bouton configuration Joomla
 * - Modification de la class icon
 *
 * Pour plus de commentaires sur ces changements, faites une recherche avec XXX
 */
```

Ce chapitre a pour but de vous rendre attentif que si vous désirez mettre à jour Bootstrap, il faudra certainement garder cette feuille de style ou appliquer les mêmes hacks. Ce genre de problème devrait disparaître avec Joomla 3 car l'équipe Joomla travaille justement pour résoudre ce genre de bug.

Les solutions de hack ont été trouvées grâce à ce post sur le forum :

<http://ux.joomla.org/forum/User-Research/662-Bootstrap-and-Joomla-156725---a-common-solution-needed>

5.6 Tutoriaux

5.6.1 Création d'un nouveau tableau de bord

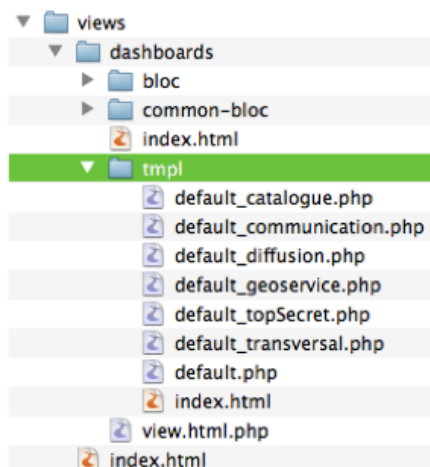
La création d'un nouveau tableau de bord est facile. Il faut d'abords insérer un nouvel enregistrement dans la table `dashboard`. La requête ressemble à l'INSERT ci-dessous.

Créer un nouveau tableau de bord

```
INSERT INTO `joomla`.`x52dk_dashboard` (`id`,`name`,`displayName`,`description`,`ordering`,`access`) VALUES (7, 'nouveau Dashboard', 'new', 'new', 6, 1);
```

La colonne `ordering` permet de choisir dans quel ordre sont affichés les éléments dans le menu. Pour la colonne `access`, veuillez consulter le tutorial suivant. Lorsque l'enregistrement dans la BDD est fait, le tableau de bord apparaît dans le menu. Il faut cependant encore créer le template qui lui correspond

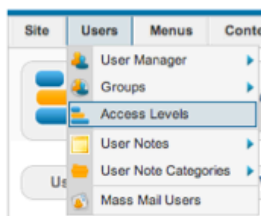
Annexe II : Documentation (22/28)



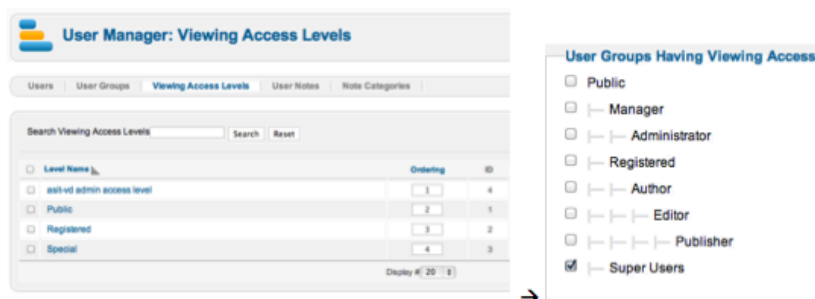
Il faut copier-coller un template existant et le renommer « default_[dashboard.name].php ». dashboard.name doit donc correspondre à la colonne name de la table dashboard. Il reste plus qu'à personnaliser ce dashboard en y insérant des blocs qui lui sont propres.

5.6.2 Cacher des tableaux de bords à certains utilisateurs

Afin de cacher des tableaux de bord qui sont propres à un seul fournisseur ou groupe d'utilisateur, on utilise les « Access Vue » de Joomla.



Rendez-vous dans Joomla sur Access Levels.



Sur la fenêtre à gauche on voit les différents groupe de vues. La partie de droite permet de définir pour un groupe de vues quelles sont les groupes qui y ont accès. Un utilisateur fait donc partie d'un groupe et ce groupe possède éventuellement des autorisations sur une vue.

On récupère ensuite l'id du groupe de vue (visible sur la fenêtre de gauche) et on effectue un opération sql qui ressemble à ceci :

Annexe II : Documentation (23/28)

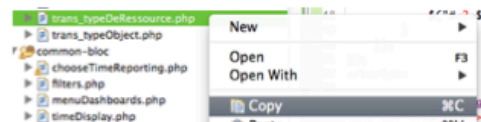
Modification de l'accès à un tableau de bord

```
UPDATE `joomla`.`x52dk_dashboard` SET `access`=4 WHERE `id`=7;
```

Pour résumer, si vous devez rendre un tableau de bord accessible qu'à un seul groupe de personne, créer un nouvelle Access Vue via l'interface de Joomla. Configurer cet Access Vue pour n'autoriser que ce groupe de personne Récupérer l'id de l'Access Vue (visible dans l'interface ou dans la BDD) et modifier la colonne « access » de la table « dashboard » pour les faire correspondre.

5.6.3 Faire appel à un job Talend

L'appel à un job Talend se fait depuis un bloc. Un bloc contient généralement un graphique, une donnée empirique ou encore un tableau de valeurs.



La première étape consiste à copier-coller un bloc existant. Ici trans_typeDeRessources.php et dupliqué en trans_test.php. Notez la nomenclature, le fichier commence par trans pour indiquer que le bloc est intégré au tableau de bord transversal. Ce n'est pas une obligation mais ça rend l'ensemble plus lisible.

```
<?php
//Attention cette id doit être unique parmi les autres blocs
//Une bonne stratégie est de le nommer en fonction du nom du job
$idDiv = 'trans_typeDeRessource';

//Nom du job Talend
$jobName = 'typeDeRessource';
?>
```

On définit toujours en haut d'un bloc l'id du div et le nom du job. Je vous laisser lire les commentaires sur la capture d'écran ci-dessus. Le nom du job correspond à celui qui a été choisi dans le tutorial de l'access Layer « Ajout d'un nouveau job ».

On ajoute la fonction au tableau des fonctions

```
<script type="text/javascript">
talendexecutedOnDate.push(function() {
//callTalent gère les paramètres globaux aux appels ajax
var callTalent = new $.fn.callTalent();

//Récupère les type de géoproduits diffusés (non public, public, private)
$.ajax({
url : callTalent.url,
dataType: callTalent.dataType,
data : callTalent.dataToSend+"&id=$jobName"&";
beforeSend: function() {
//Affichage / message de div
$("#"+$idDiv+"-result-success").hide();
$("#"+$idDiv+"-no-result").hide();
$("#"+$idDiv+"-waiting-for-result").show();
},
success: function(json) {
$("#"+$idDiv+"-body").empty();
$.each(json.result, function(key,value) {
$("#"+$idDiv+"-body last").append("<tr>");
$("#"+$idDiv+"-body tr last").append("<td style='width:100px;'><div style='width:100px;'><div style='width:100px;'>";
});
//Affichage / message de div
$("#"+$idDiv+"-result-success").fadeIn("slow");
$("#"+$idDiv+"-no-result").hide();
$("#"+$idDiv+"-waiting-for-result").hide();
},
error: function(error,ajaxOption, errorThrown) {
callTalent.errorHandler(error.status+"<throttling>"<error_responseText>);
//Affichage / message de div
$("#"+$idDiv+"-result-success").hide();
$("#"+$idDiv+"-no-result").show();
$("#"+$idDiv+"-waiting-for-result").hide();
}
});
});
</script>
```

Fait référence au fichier jquery.callTalent.js Affiche/masque div Récupération des valeurs

Partie Javascript du code

Annexe II : Documentation (24/28)

```
<h2 style="text-align:center;">Géoproduits diffusés</h2>
<div id="<?=$idDiv?>">
  <table class="table table-striped result-success">
    <tbody>
    </tbody>
  </table>

  <div class="well waiting-for-result">
    <div class="progress progress-striped active">
      <div class="bar"></div>
    </div>
  </div>
  <div class="well no-result">
    <span class="no-data">Aucune donnée</span>
  </div>
</div>
```

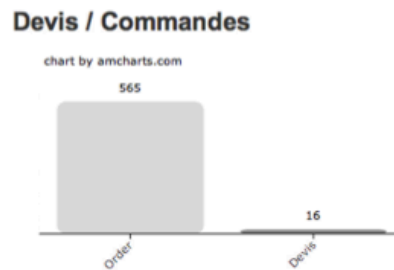
partie HTML

L'appel à un job Talend se fait donc par un appel au proxy grâce à une requête ajax.

6 Génération de graphiques avec AmCharts

AmCharts se trouve dans les bibliothèques utilisées directement dans le composant Joomla (chapitre précédent).

Afin de comprendre le façon dont sont intégrés les graphiques dans les tableaux de bord, je vous invite à ouvrir le fichier suivant : Diffusion_devisCommande.php qui se trouve dans le composant Joomla sous /admin/views/dashboards/bloc/. Celui-ci créer un graphique comme celui-ci :



Annexe II : Documentation (25/28)

6.1 Choix des couleurs

AmCharts n'offre pas la possibilité de choisir une palette de couleur. Un petit script a donc été développé. Veuillez lire les commentaires ci-dessous.

```
/*
 * COULEUR
 * Ce script permet de définir un tableau de couleur (longueur variable)
 * Le script choisit des couleurs équi-distante en fonction du nombre de série
 * Il repart à zéro si la taille du tableau est dépassée
 * Merci Yves pour le prototype du script
 */

//color est définit dans le fichier config.staticVariable.php

function getColorGrid(nombreCouleur) {
    var selectedcolors = [];
    length = COLOR.length;

    //Distance entre chaque couleur
    dist = Math.round(length/nombreCouleur);
    if (dist<0) dist = 0;

    for (i=0;i<(nombreCouleur*dist);i=i+dist) {
        //Le modulo permet de 'repartir à zéro' si la taille du tableau
        //color est inférieur aux nombre de série
        selectedcolors.push(COLOR[i % length]);
    }

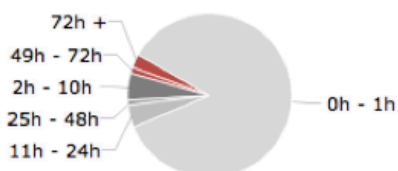
    return selectedcolors;
}
*/
```

Ce script est visible dans site/lib/dashboards/global.js

```
define ("COLOR", serialize (
    array (
        "#D8D8D8",
        "#D8D8D8",
        "#D0D0D0",
        "#C8C8C8",
        "#C0C0C0",
        "#B8B8B8",
        "#B0B0B0",
        "#A8A8A8",
        "#A0A0A0",
        "#989898",
        "#909090",
        "#888888",
        "#808080",
        "#787878",
        "#707070",
        "#686868",
        "#606060",
        "#585858",
        "#505050",
        "#484848",
        "#404040",
        "#383838",
    )
));
```

Les couleurs sont définies dans site/lib/dashboards/config.staticVariable.php
Il s'agit ici d'une palette de ton de gris.

Annexe II : Documentation (26/28)



Cependant, comme vous pouvez le constater sur le graphique ci-dessus, il est possible de définir certaines valeurs en bordeaux afin de les faire ressortir.

```
/**
 * SUR-BRILLANCE ROUGE
 * Call Talend a une fonction permettant de mettre en rouge certains éléments du graphique
 * Dans le cas du délai de traitement de commande on utilise le Label pour faire correspondre les portions à mettre en rouge
 * On stocke donc dans ce tableau les éléments que l'on souhaite mettre en surbrillance
 * C'est également utilisé pour faire la somme des requête qui ont dépassé 48 heures
 */
var moreThan48hLabel = [];
moreThan48hLabel.push('49h - 72h');
moreThan48hLabel.push('72h +');
callTalend.setAlertValue('inArray',moreThan48hLabel);
```

On définit dans le bloc quelles valeurs mettre en surbrillance. Ce script est visible dans `/admin/views/dashboards/bloc/diffusion_historiquesTempsTraitementCommande.php`

```
/*
 * Gestion des VALEURS CRITIQUES
 * Parfois, on veut qu'une couleur soit mis en exergue car la valeur qui l'accompagne est critique
 *
 * Pour ce faire il faut définir la variable defineAlert à true
 * et appelé la fonction setAlertValue
 */
var defineAlert = false;
var labelArray = [];

/*
 * Pour l'instant, il n'y a que le type inArray
 * inArray signifie qu'on envoie un tableau de string et si un label
 * correspond à une de ces valeurs, on met l'élément du graph correspondant en exergue
 *
 * On pourrait imaginer d'autres façon de faire:
 * Par exemple définir une valeur critique à partir d'une certaine valeur
 * ou inférieur
 */
this.setAlertValue = function(type,value) {
  if (type = 'inArray') {
    defineAlert = true;
    labelArray = value;
  }
};
```

C'est ensuite dans le fichier `jquery.callTalend.js` qu'on gère l'attribution des couleurs. Veuillez lire les commentaires

Annexe II : Documentation (27/28)

7 Génération de rapports Birt

« projet zip »*	
/04_reporting/birt/export.zip	Export de l'environnement du report designer
	* Référence au chapitre 1.1

7.1 Versions

Birt est constitué de deux outils : le report designer permet de générer les rptdesign et le « engine report » permet de les interpréter sur le serveur Tomcat.

Les deux outils sont à la version 4.2.0

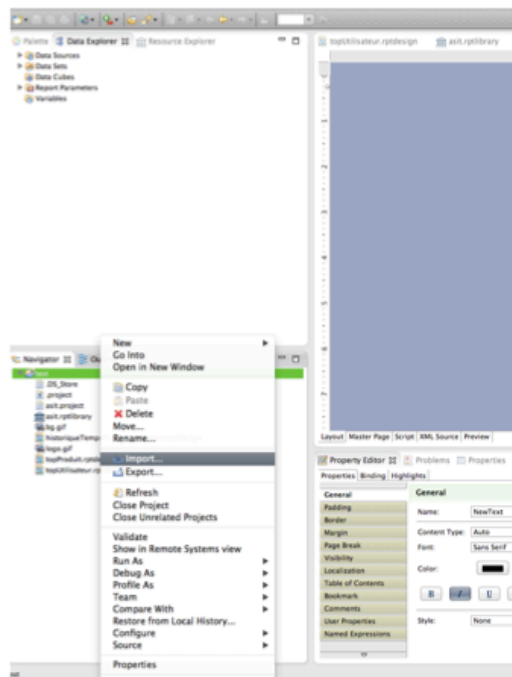
7.2 Installation de l'Engine report

L'installation est aisée. Veuillez vous référer au guide suivant : <http://www.eclipse.org/birt/phenix/deploy/viewerSetup.php>

7.3 Import de l'environnement

Ouvrez le report designer. Assurez-vous d'être dans la perspective « report design ». Pour ceci cliquez sur « Window » -> « Open Perspective » -> « Other » -> « Report Design ».

Une fois faits, vous devriez obtenir l'environnement ci-dessous :



Annexe II : Documentation (28/28)

Faites un clic droit sur le dossier root dans l'onglet « navigator » (capture d'écran) et choisissez « Import » puis importer le fichier d'archive « /04_reporting/birt/export.zip ». Vous devriez obtenir l'environnement ci-dessous :



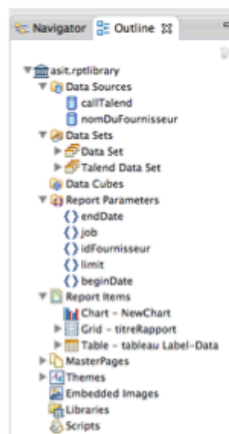
La clé des rapports réside dans le fichier asit.rptlibrary

7.4 Asit.rptlibrary

Le fichier rptlibrary est un projet standard ressemblant fortement à un rptdesign. Cependant, celui-ci ne sera jamais ouvert par un utilisateur standard. Une librairie rptLibrary permet de définir des éléments qui sont communs à tous les rapports. Elle permet en quelque sorte de créer une boîte d'outils permettant de générer des fichiers rptdesign. Les éléments qu'on peut définir dans un rptlibrary sont très variés : Source de données, data set, graphiques, trame de fond, pied de page, style css. Finalement chaque élément peut être défini dans le rptlibrary.

Mais pourquoi le définir dans le rptlibrary et pas directement dans le rptdesign ? C'est simple. Ça permet de n'avoir qu'à modifier cette librairie pour que les effets soient produits sur tous les rapports.

Double-cliquez sur le fichier asit.rptlibrary. La page n'affiche rien de particulier. Recherchez l'onglet « Outline ».



On voit ci-dessus les différents éléments qui sont définis dans le rptdesign. Double-cliquez sur un élément pour le modifier.

Annexe III : Cahier des charges (1/3)

heig-vd **comem+**
Haute Ecole d'Ingénierie et de Gestion
du Canton de Vaud

Hes
Haute Ecole de
Sion

Diplôme / Diplom 2012
Travail de Bachelor
Bachelorarbeit

Département comem+
Ingénierie des Médias - IT
Gaël Bernard

01.03.2012

Développement d'un tableau de bord de l'activité du
GEOPortail vaudois

Mandant / Externer Ansprechpartner

Xavier Mérour
Coordinateur
Asit-VD
Avenue de l'Université 5
1014 Lausanne
xavier.merour@asitvd.ch
021 316 70 24

Lieu du travail / Ort der Arbeit

Association pour le Système d'Information du
Territoire Vaudois (ASIT VD)
Mérour Xavier
Coordinateur
Avenue de l'Université 5
1014 Lausanne
xavier.merour@asitvd.ch
021 316 70 24

Le candidat


Gaël Bernard

Le professeur responsable


Daniel Rappo

Le mandant


Xavier Mérour

Le chef du département comem+


Andreas Sidler

Annexe III : Cahier des charges (2/3)

heig-vd comern+

Haute Ecole d'Ingénierie et de Gestion
du Canton de Vaud

Hes-so
Haute Ecole Spécialisée
de Suisse occidentale

Diplôme / Diplom 2012

Travail de Bachelor
Bachelorarbeit

Développement d'un tableau de bord de l'activité du GEOPortail vaudois

Explication générale (motivation, définition stricte du sujet)

Allgemeine Angaben (Motivation, genaue Definition des Projektes)

Asit-VD

L'Association pour le Système d'Information du Territoire Vaudois est le point de rencontre entre fournisseurs et utilisateurs de données géographiques. Il s'agit d'une association basée sur un partenariat public-privé composée de quelque 350 partenaires dont l'Administration cantonale vaudoise.

Le but de l'association est d'assurer une meilleure coordination aux acteurs œuvrant pour une bonne gestion du territoire cantonal ainsi que de favoriser l'échange de géodonnées et de connaissances entre partenaires.

Le GEOPortail

Afin de remplir son rôle de point d'entrée des services de l'association, le GEOPortail met à disposition des outils en ligne ainsi que les géodonnées à sa disposition. Le portail est composé de quatre sections :

GEOCatalogue : Recense les géodonnées à disposition

GEOCommande : Permet de commander les géodonnées

GEOService : Permet d'afficher des couches thématiques publiées par le fournisseur et d'y superposer ses propres données

GEOGuichet : Permet de visualiser les géodonnées sous forme de cartes interactives

Les fournisseurs de données

Les fournisseurs de données mettent à disposition des couches thématiques se superposant au guichet cartographique utilisé par les partenaires. Les fournisseurs sont actuellement au nombre de 15 mais ce chiffre est en constante augmentation.

Définition du sujet

Tableau de bord de l'activité du GEOPortail vaudois

A ce jour, le GEOPortail est totalement opérationnel mais aucun outil centralisé ne permet de mesurer et de monitorer son utilisation. Le but serait de palier à ce manque en mettant à disposition des administrateurs de Asit-VD, aux fournisseurs de données et aux partenaires un tableau de bord permettant de mesurer en temps réel l'utilisation de celui-ci et également d'en tirer des statistiques historiques sur les services les plus utilisés.

Cadrage du sujet (problématique, hypothèses de travail, méthodologie, ressources, etc.)
Spezifische Angaben zum Projekt (Problematik, Arbeitshypothese, Methodik, Ressourcen, usw.)

Le tableau de bord se présente sous la forme d'un module s'intégrant dans le système de gestion de

Annexe III : Cahier des charges (3/3)

heig-vd comem⁺

Haute Ecole d'Ingénierie et de Gestion
du Canton de Vaud

Hes-so
Haute Ecole Spécialisée
de Suisse occidentale

Diplôme / Diplom 2012

Travail de Bachelor
Bachelorarbeit

contenu Joomla actuellement en place.

Problématique

Les données permettant d'effectuer le tableau de bord émanent de différentes sources (par exemple : bases de données Mysql de Joomla, API Google, ...) et sous différents formats de données (XML, bases de données relationnelles, textes bruts, JSON,...) Le but est donc d'unifier ces données afin de pouvoir les interpréter et les représenter sous forme homogène.

D'ailleurs, la transformation de ces données brutes en illustrations graphiques attractives est également un défi pour ce projet. Il s'agit de créer une interface synoptique interactive et ergonomique avec des données qui ne le sont pas aux premiers abords.

Il est également important d'offrir une standardisation permettant d'adapter le tableau de bord pour un futur service avec un effort minime.

Hypothèse de travail

Afin de répondre au besoin d'uniformisation des données, une couche de transformation puis d'abstraction de celles-ci sera mise en place.

Le besoin de standardisation pourra quand à lui être traité grâce à la mise en place d'un patron de conception « Modèle-Vue-Contrôleur » permettant une maintenance et une évolutivité simplifiée.

Objectifs (mesurables) à atteindre (e.g. livrables)

Angestrebte (messbare) Ziele (z. B. zu liefernde Ergebnisse)

- 1) Rendre un tableau de bord permettant de mesurer l'activité du GEOPortail vaudois.
- 2) Proposer un outil permettant la création de reporting
- 3) Offrir à chaque fournisseur un outil de monitoring d'utilisation des données fournies
- 4) Intégrer le tableau de bord au portail existant
- 5) Rendre un code source dont la structure est modulaire et évolutive

Liste des tâches à réaliser

Liste der auszuführenden Arbeiten

- 1) Recenser les données disponibles et nécessaires pour le tableau de bord
- 2) Traduire les données disponibles en indicateurs graphiques.
- 3) Concevoir une interface graphique attractive et ergonomique
- 4) Identifier les outils existants (framework, API, librairie) pouvant s'intégrer dans le cadre du projet
- 5) Concevoir une architecture modulaire et évolutive
- 6) Intégrer le tableau de bord au portail existant sous Joomla

Annexe IV : Procès-verbal du bilan intermédiaire

Bilan intermédiaire PV		
[Sélectionner la date]	14:15	Bureau ASIT-VD - Lausanne
Réunion organisée par	Gaël Bernard	
Mandant	Xavier Mérour	
Professeur Responsable	Daniel Rappo	
Présentation de l'avancement du projet (Développement d'un tableau de bord de l'activité du GEOPortail vaudois)		
60 minutes	Gaël Bernard	
<p>Discussion : Dans l'ordre :</p> <ol style="list-style-type: none">La première partie de la discussion consistait à présenter le projet.<ul style="list-style-type: none">- La discussion est introduite par le schéma technique du projet.- Explication des différentes sources de données utilisées. Daniel Rappo demande des explications sur le rôle du portail et plus généralement sur l'activité de l'ASIT-VD.- Présentation du datawarehouse- Présentation de la couche d'accès de Talend- Justification du choix de amCharts. Présentation des tests effectués avec les autres plateformes.- Présentation du composant Joomla (tableau de bord transversal et de diffusion).<p>Remarque personnelle : La présentation du projet aurait été plus pertinente en présentant d'abords le résultat puis ensuite les étapes pour y arriver.</p>Explications sur ce qu'il reste à faire.<ul style="list-style-type: none">- Utiliser des web services comme source de données (Monitor, ou Google analytics)- Utiliser des fichiers de type « log » (fichier texte) comme source de données- Générer des fichiers de template pour Birt et réaliser quelques rapports d'exemples- Créer une documentation pour le projetRelecture des objectifs du cahier des charges initial par Daniel Rappo. Le cahier des charges est approuvé et signé par tous les intervenants.Présentation du poster. OKPrésentation du résumé. Ok		
Résumé		
Le cahier des charges initial correspond au projet actuel. Il a été signé. Le poster et le résumé ont été approuvés.		

Annexe V : CD-ROM



Contenu du CD-ROM

▼	livrables	41.5 Mo
▶	01_datawarehouse	35.2 Mo
▶	02_accessLayer	5.1 Mo
▶	03_cms	955 Ko
▶	04_reporting	207 Ko
▼	pdf	37 Mo
	01_rapport_final.pdf	27.5 Mo
	02_cahierDesCharges.pdf	692 Ko
	03_documentation.pdf	4.4 Mo
	04_poster.pdf	233 Ko
	05_pv_billant_intermédiaire.pdf	47 Ko
	06_resume_projet.pdf	32 Ko
	07_dossier_travail_preparatoire.pdf	3 Mo
	08_directives.pdf	1.1 Mo

Résumé

Le projet consiste à réaliser un tableau de bord interactif du GEOPortail Vaudois qui est la porte d'entrée de l'infrastructure cantonale de données géographiques. Ce portail est géré par l'Association pour le Système d'Information du Territoire Vaudois.

L'outil mis en place propose une interface web permettant d'offrir une vision transversale des activités de catalogage et de diffusion de géodonnées. Il s'adresse à trois publics : les administrateurs du portail, les gestionnaires de données et le grand public.

On trouve au cœur de ce projet :

- La mise en place d'un entrepôt de données afin d'assurer la pérennité des informations et offrir un accès rapide à celles-ci*
- La réalisation d'une couche d'accès unifiée à des données hétéroclites*
- La représentation des données sous forme de graphiques*
- La génération de rapports dynamiques*