

Table des matières

<u>1</u>	<u>INTRODUCTION.....</u>	<u>4</u>
<u>2</u>	<u>PRESENTATION DES PROBLEMATIQUES GENERALES.....</u>	<u>5</u>
<u>3</u>	<u>LE DEROULEMENT DE LA REPRISE DES DONNEES</u>	<u>14</u>
<u>4</u>	<u>LA CONVERSION VERS LE MODELE EAV</u>	<u>15</u>
<u>5</u>	<u>LA FORMALISATION DU PROCESSUS</u>	<u>18</u>

www.x2i.fr

1 INTRODUCTION

Ce document présente la méthodologie adoptée pour la reprise de vos données.

Ce document est également l'initialisation de la spécification de reprise de vos données dans le cadre plus global de notre prestation « Refonte de site e-commerce ».

Public concerné

Ce document est essentiellement technique et nécessite, pour sa bonne compréhension, des connaissances dans le domaine des bases de données.

Ce qu'est ce document

Ce document présente une méthodologie.

Ce document est destiné à évoluer vers une spécification de reprise de données.

Ce document est un document applicable du mémoire technique.

Ce que n'est pas ce document

Ce document n'est pas une spécification finalisée.

Ce document n'est pas un contrat commercial.

2 PRESENTATION DES PROBLEMATIQUES GENERALES

La reprise des données est souvent un poste sous-estimé, que ce soit en termes de complexité, de coûts et de compétences nécessaires.

Dans le cas d'une migration vers un site Magento, une problématique supplémentaire vient s'ajouter aux complexités habituelles : la conversion dans un modèle EAV (voir plus loin pour une définition plus précise).

2.1 Identification des données à migrer

La problématique

Lorsque l'on parle de reprise (ou migration) de données, on a souvent tendance à ne penser qu'aux seules données stockées en base.

Pourtant, de nombreuses autres données peuvent être impliquées :

- Les images,
- Les fichiers de type média (fichiers audio, vidéo, flash...)
- Les fichiers de type documentaire (fichier Word, Excel, PDF...)
- Les fichiers de type personnalisation client (fichiers CSV, XML...) qui stockent des informations personnalisées par vos clients en dehors de la base de données.
- ...et d'autres types de fichiers auxquels nous n'avons pas encore pensés...

Ces types de données peuvent ou non être liées ou non aux données de la base.

La réponse Xi Ingénierie

L'une des premières étapes de la spécification détaillée de reprise des données consisté à justement identifier l'ensemble des données qui doivent faire l'objet de cette migration.

Pour chaque donnée identifiée, nous stipulons :

- Son type d'origine : base de données, fichier contenu, fichier de données...

- Son type de destination (idem)
- Son utilisation FO (Front Office) ou BO (Back Office)
- Les relations existantes entre les données
- Sa signification métier : l'utilisation qui en sera faite dans votre site eCommerce
- Les processus d'intégrité qui lui sont associés : par exemple, une adresse est nécessairement liée à une personne physique ou une entité juridique.

2.2 Les problématiques de formalisme en entrée

La problématique

Les données en entrées sont plus ou moins structurées :

- Données sous forme de Base de Données
- Données résultats d'export :
 - Format CSV
 - Format XML
 - Format tabulaire
 - Format binaire
- Données fichiers structurée
 - Arborescence de fichiers
 - Descriptifs de fichiers
- Accès aux données :
 - Données sur des serveurs différents
 - Données protégées par des droits d'accès spécifiques
 - Données nécessitant l'utilisation d'un code logiciel (API) pour leur accès.

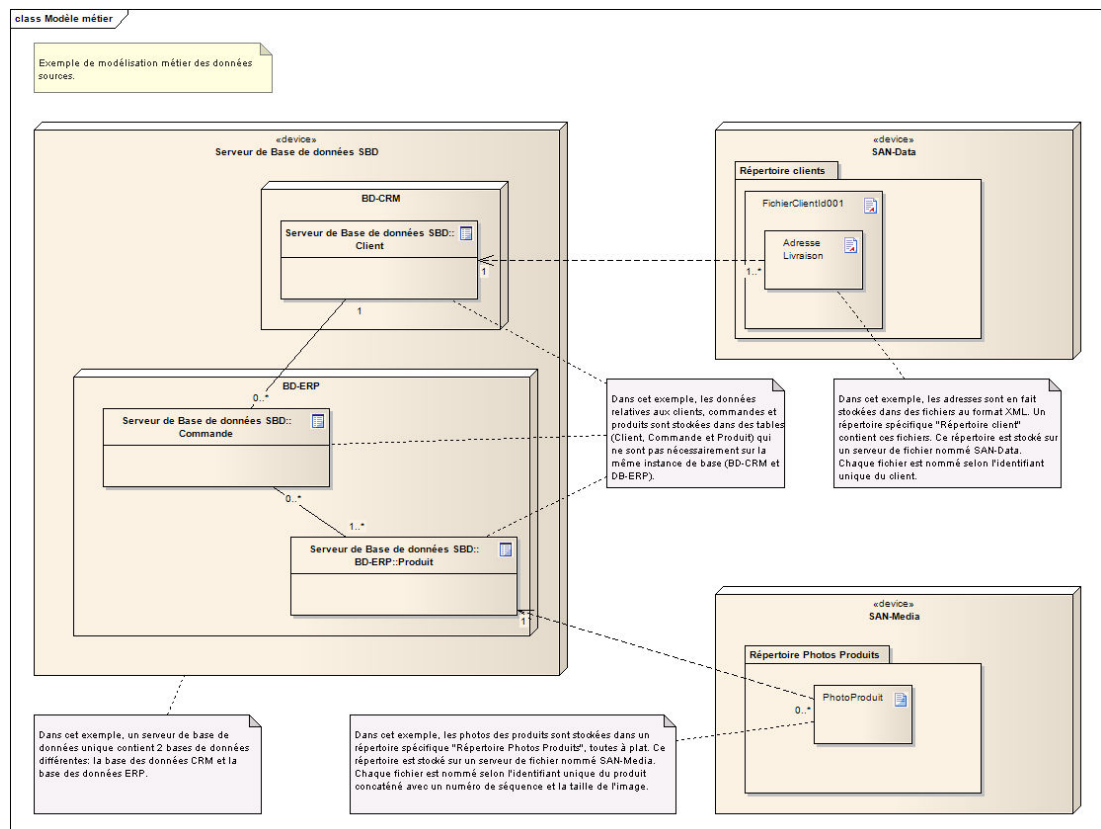
Les formats des données sont donc de façon générale assez divers et en général multiples.

La réponse Xi Ingénierie

La phase de recueil des besoins comporte donc nécessairement l'analyse de ces formats ainsi que la localisation des données et des droits d'accès nécessaires.

Cet aspect est parfois critique : en effet, le jour « J » de la migration, l'ensemble des sources de données doit être accessible depuis un seul point focal afin de que les données puissent être accédées et migrées.

Ce recueil d'information est formalisé dans un modèle UML dit modèle métier comme dans l'exemple ci-dessous :



2.3 Les problématiques de volumétrie

La problématique

L'un des aspects de la migration de données consiste, selon les cas, à devoir manipuler de gros volumes d'information.

Ces volumes peuvent être des volumes de base de données ou de fichiers. Il est important, dès lors que l'on parle de volumétrie, de choisir des outils adaptés et habilité à manipuler de grandes quantités de données : les fichiers Excel ou base de données légère (comme Access par exemple) ne peuvent plus suffire.

La réponse Xi Ingénierie

Magento propose en standard un « loader » qui permet par exemple d'importer directement des données issues d'une base OsCommerce (non modifiée). Toutefois, les mécanismes proposés, quels qu'ils soient, ne sont absolument plus adaptés dès lors que la volumétrie des données devient significative où que le(s) modèle(s) de données sources sont exprimés dans un standard différents (OsCommerce modifié ou autre source).

Xi Ingénierie a donc opté pour la mise en œuvre du produit Talend (voir en annexe pour plus de précisions sur Talend).

Talend apporte de nombreux points positifs et entre autre une capacité à traiter efficacement et rapidement de gros volumes de données.

2.4 Les problématiques de performances

La problématique

La problématique des performances est induite par 2 points forts :

- Le temps de migration est conditionné par la volumétrie des données à migrer
- La criticité du site eCommerce considéré, particulièrement en termes de CA, est dimensionnante du point de vue de l'indisponibilité.

La réponse Xi Ingénierie

De la même façon, Talend nous permet d'adresser la problématique des performances.

2.5 Les problématiques de test

La problématique

Vous disposez de données en entrée, sur votre site actuel ou dans votre système d'information. Comment s'assurer qu'à l'issue de la migration vos données vont effectivement bien se retrouver dans votre site eCommerce ?

La réponse Xi Ingénierie

La réponse Xi Ingénierie sur les aspects tests est en fait déclinée tout au long du projet :

- La migration des données fait l'objet d'une spécification précise, au formalisme UML : dans le même temps, la génération du plan de test de validation est effectuée

- La migration des données est outillée avec Talend : Xi met en œuvre des vérifications systématiques des processus d'importation de façon unitaire pour chaque type de données identifié
- Pour chaque type de donnée, nous vous demandons un jeu de données étalon qui permet de procéder à la recette « manuelle » de la migration.
- L'outillage des processus de migration permet, grâce à la répétabilité des migrations, de garantir qu'aucun problème de non régression ne surviendra lors des évolutions qui seront apportées
- Comme nous le verrons plus loin, une bonne partie de la migration est déjà amenée en standard et ne nécessite plus de validation.

2.6 Les problématiques de répétabilité

La problématique

Le nombre de données à migrer est en général important. La migration des données peut s'avérer extrêmement complexe. Par ailleurs, la mise en place de votre site nécessite que nous puissions tester au plus vite les aspects volumétrie.

De ce fait, la migration est généralement déroulée de façon itérative : tous les éléments de votre système source ne sont considérés en même temps, et au sein de ces éléments, toutes les données ne sont pas nécessairement considérées dès le départ.

Concrètement, nous allons dans un premier temps d'abord nous attacher à migrer vos produits afin de pouvoir valider et tester le modèle de catalogue que nous avons mis au point. Pour autant, nous ne nous attacherons pas aux commandes dans cette première phase.

La réponse Xi Ingénierie

Ce processus peut vous interpeler : nous migrons les produits, mais que ce passera-t-il quand je basculerai mon site en production : j'aurai certainement ajouté de nouveaux produits...

En fait, cela ne posera aucun problème : nous rejouerons le script de migration Talend pour les produits mis au point dans la première phase.

L'un des avantages principaux de Talend est la répétabilité des processus : une fois que le processus de migration est validé (ou une partie du processus est validée), vous êtes certain de pouvoir le rejouer ultérieurement sans avoir à le valider de nouveau.

Ainsi, nous modéliserons et testerons la migration de vos données, étape par

étape. Chaque étape ou type de données validé ne sera plus remise en cause et pourra être joué le jour du basculement opérationnel.

2.7 Les migrations en 2 temps

La problématique

Nos clients nous demandent souvent s'il est possible de migrer leurs données vers leur site en 2 temps. De façon résumée, les raisons en sont les suivantes :

- La volumétrie est conséquente et une migration en 2 temps serait rassurante
- Faire fonctionner en doublons un ancien site et un nouveau site serait également un gage de sécurité.

La réponse Xi Ingénierie

Notre réponse est systématiquement négative quand elle adresse la première raison : il est largement préférable de mettre l'accent sur une optimisation de la migration, quitte à rendre indisponible le site pendant 1 à 2 heures.

Dans le second cas, la réponse est également en général négative mais n'est pas formulée par nos soins. En effet, faire fonctionner 2 sites en parallèle est possible mais nécessite la mise en place de script de migration fonctionnant dans le sens inverse (nouveau site vers ancien site) et qui fonctionne pratiquement en temps réel. La mise en place d'une telle solution est tout à fait possible d'un point de vue technique mais engendre des coûts similaires ou supérieurs aux coûts de la migration des données. Il s'agit donc en général d'un arbitrage budgétaire.

2.8 Modèle orthogonal et modèle EAV

La problématique

Le principe du modèle dit « orthogonal » est relativement simple à comprendre. Il s'agit d'une représentation en base de données de type objet, c'est-à-dire proche du modèle sémantique des données. En clair et avec un exemple, pour modéliser un client, on utilisera 3 tables :

- Une table « Client » comprenant les champs spécifiques d'un client (nom, prénom, email, téléphone...) ainsi qu'une liste d'identifiant pointant vers la table Adresse (voir ci-après)
- Une table « Adresse » contenant les informations d'adresse d'un

client : Numéro, type de voie, nom de la voie...

- Une table commande contenant les informations de la commande : date, montant, statut...

A contrario, un modèle EAV (Entity / Attribute / Value) approche les données d'une façon toute différente. En schématisant à l'extrême, un modèle EAV, pour décrire les mêmes objets clients, adresses et commandes, ne comportera que 2 tables :

- Table descripteurs :
 - Identifiant de l'objet ou attribut
 - Nom
 - Type
- Table valeurs :
 - Identifiant de l'objet
 - Identifiant de l'attribut
 - Valeur de l'attribut

En fait, des tables complémentaires de paramétrage annexes sont nécessaires pour fournir des descriptions des attributs et permettre des contrôles d'intégrité. Mais la différence réside en cette explication simple :

- Dans un modèle orthogonal, une table par objet manipulé est nécessaire
- Dans un modèle EAV, le nombre de table ne dépend pas du nombre d'objet manipulés (schématiquement on utilise seulement 2 tables)

L'explication que nous fournissons ici est simplifiée à l'extrême et sa mise en œuvre réelle est beaucoup plus complexe. Toutefois, les principes restent les mêmes.

Les tableaux ci-dessous expliquent comment, dans un cas simplifié, les données sont prises en compte dans une optique orthogonale et dans une optique EAV.

Exemple de modélisation orthogonale

Représentation schématique d'un modèle orthogonal					
Client					
Colonne	Id	Nom	Prenom	Email	Telephone
Type	INT	CHAR	CHAR	CHAR	CHAR
Valeurs	1	BENOIT	Marc	mbenoit@test.com	01.01.01.02.03
	2	VINCENT	Michel	mvincent@truc.com	02.01.01.02.03
	3	PAUL	Jean	jean@coment.com	03.01.01.02.03
Adresse					
Colonne	Id	IdClient	Numero	TypeVoie	NomVoie
Type	INT	INT	INT	CHAR	CHAR
Valeurs	1	1	12	Rue	Marcel PROUST
	2	1	32	Avenue	Marcel AYME
	3	2	11	Rue	PREBOST
	4	3	786	Impasse	des Aynats
Commande					
Colonne	Id	IdClient	Date	Montant	Statut
Type	INT	INT	DATE	CHAR	CHAR
Valeurs	1	1	01/01/2009	42,00 €	Close
	2	2	01/02/2009	546,00 €	Close
	3	3	01/03/2009	75,00 €	Expédiée
	9	3	01/04/2009	89,00 €	En validation

Exemple de modélisation EAV

Représentation schématique d'un modèle EAV			
Descripteurs			
Colonne	Id	Nom	Type
Type	INT	CHAR	CHAR
Valeurs	1	Nom	CHAR
	2	Prenom	CHAR
	3	Email	CHAR
	4	Telephone	CHAR
	5	LienClientAdresse	INT
	6	Numero	INT
	7	TypeVoie	CHAR
	8	NomVoie	CHAR
	9	LienClientCmde	INT
	10	Date	DATE
	11	Montant	INT
	12	Statut	CHAR

Valeurs					
Colonne	IdObjet	IdAttribut	Val_INT	Val_Date	Val_CHAR
Type	INT	INT	INT	DATE	CHAR
Valeurs	1	1			BENOIT
	1	2			Marc
	1	3			mbenoit@test.com
	1	4			01.01.01.02.03
	1	5	200		
	200	6	12		
	200	7			Rue
	200	8			Marcel PROUST
	1	5	300		
	300	6	32		
	300	7			Avenue
	300	8			Marcel AYME
	1	9	400		
	400	10		01/01/2009	
	400	11	42		
	400	12			Close
	2	1			VINCENT
	2	2			Michel
	2	3			mvincent@truc.com
	2	4			02.01.01.02.03
	2	5	500		
	500	6	11		
	500	7			Rue
	500	8			PREBOST
	2	9	600		
	600	10		01/02/2009	
	600	11	546		
	600	12			Close
	2	1			PAUL
	2	2			Jean
	2	3			jean@coment.com
	2	4			03.01.01.02.03
	2	5	700		
	700	6	786		
	700	7			Impasse
	700	8			des Aynats
	2	9	800		
	800	10		01/03/2009	
	800	11	75		
	800	12			Expédiée
	2	9	900		
	900	10		01/04/2009	
	900	11	89		
	900	12			En validation

La réponse Xi Ingénierie

Le chapitre « La conversion vers le modèle EAV » explique de façon détaillée comment est prise en compte la complexité de la transformation du modèle orthogonal fourni en entrée vers le modèle EAV de Magento.

3 LE DEROULEMENT DE LA REPRISE DES DONNEES

La reprise des données est un projet à part entière. Son déroulement suit donc un processus projet relativement classique :

- Spécification
- Conception
- Réalisation
- Validation
- Migration opérationnelle

3.1.1 Spécification

La phase de spécification, comme pour tous les projets, est la phase la plus critique. Elle permet d'identifier :

- L'ensemble des données à migrer (base de données, fichiers de données, fichiers et répertoires)
- Le format des données à migrer
- La destination, dans le système cible (entre autre Magento) de chacune des données, champ par champ pour les données de base ou de fichiers
- Les moyens de vérification de la migration : plan de test de migration
- Les temps de migration acceptables

3.1.2 Conception

La phase de conception est en générale transparente pour nos clients et consiste essentiellement à paramétrer (ou développer, en Java sous Talend) les processus de migration à partir de Talend. Cette phase s'attachera essentiellement à identifier et modéliser les processus qui sortent du cadre standard (voir plus loin le chapitre « La conversion vers le modèle EAV »).

3.1.3 Réalisation

La réalisation consiste à implémenter les scripts complémentaire, à jouer la migration sur des jeux de tests.

3.1.4 Validation

Cette étape est jouée de façon itérative et peut être appliquée sur des parties indépendantes de la migration (par exemple, on valide la migration des comptes clients dans un premier temps). Elle consiste à jouer la migration et à vérifier que son résultat est bien conforme, notamment en passant le plan

de test et en jouant la migration des jeux de données témoins. On prendra particulièrement soin de valider les temps de migration à cette occasion.

3.1.5 Migration opérationnelle

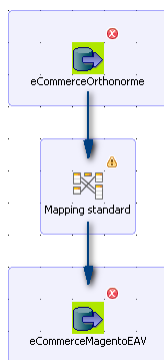
C'est l'étape finale. Elle ne peut avoir lieu qu'une fois que la validation de la migration ait été prononcée pour la globalité de la migration. Cette étape consiste à jouer la migration le jour du basculement en production du nouveau site sous Magento.

4 LA CONVERSION VERS LE MODELE EAV

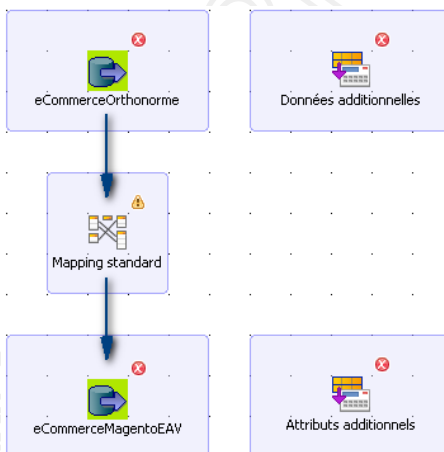
Notre retour d'expérience sur plusieurs migrations nous a conduit à finalement opter pour une approche très spécifique :

- Nous travaillons à partir d'un mapping standard : une base de données type, modèle orthogonal, est migrée vers une base de données type de Magento (modèle EAV)
- Nous procédons au paramétrage (de façon itérative) des attributs du modèle EAV sous Magento pour incorporer les spécificités de vos données et ajoutons dans le modèle orthonormé les spécificités de vos données
- Nous mappons les spécificités de vos données avec le modèle cible Magento
- Nous implémentons les processus spécifiques de migration complémentaires :
 - Processus de migration de votre modèle de données vers le modèle orthonormé standard
 - Processus des spécificités de votre modèle vers le modèle EAV.

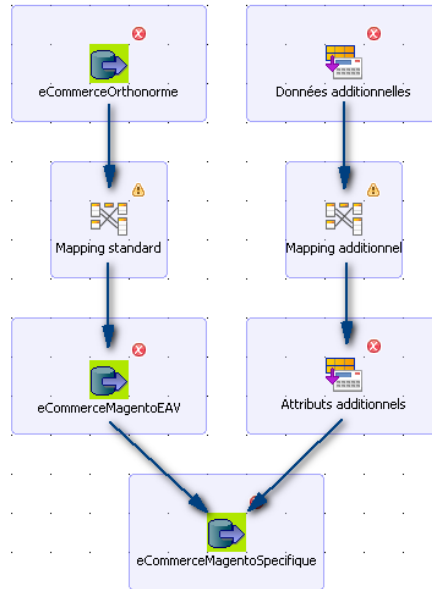
Migration standard



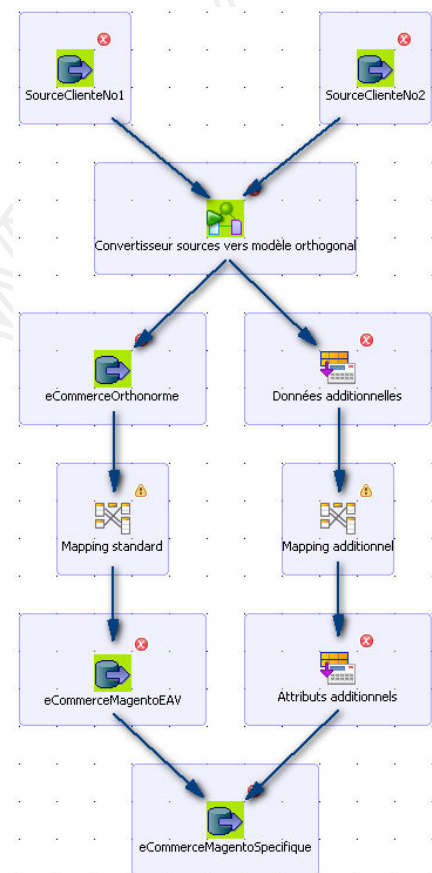
Paramétrage spécifique



Mapping du spécifique



Migration des sources vers le modèle orthogonal



5 LA FORMALISATION DU PROCESSUS

La formalisation de l'ensemble du processus de migration se fait dans un document de spécification. Ce document est principalement généré à partir de :

- La modélisation UML des données (voir exemple plus haut)
- La modélisation sous Talend des processus de migration (voir exemple plus haut)

www.x2i.fr

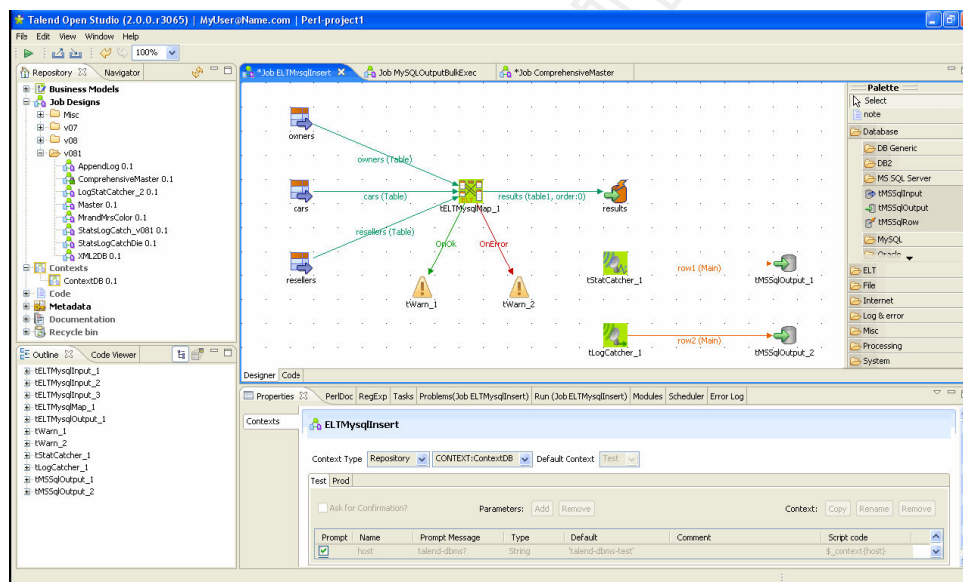
Annexe A: Présentation de Talend

Talend, le premier fournisseur de solutions d'intégration de données Open Source, tire partie du modèle Open Source pour mettre l'intégration de données à la portée d'entreprises de toutes tailles, de tous niveaux d'expertise technique, et de tous les budgets.

Les solutions de Talend se connectent à toutes les sources et cibles et peuvent être téléchargées librement et gratuitement.

Talend offre la possibilité de réaliser l'ensemble du processus de migration :

- Conception
- Développement
- Exécution



Les atouts de Talend sont décisifs :

- Une solution **Open Source** accompagnée par une communauté très dynamique.
- Un **environnement de développement entièrement graphique** qui améliore de manière significative la productivité, facilite l'exécution de galops d'essais et permet la réutilisation des mappings de données et des transformations pour les processus de synchronisation (si applicable).
- Une plateforme performante et fortement évolutive avec un déploiement en mode Grid et exploitant les architectures ETL et

ELT.

- Une **connectivité étendue** qui supporte tous les systèmes source et cible : progiciels (ERP, CRM, etc.), bases de données, serveurs centraux, fichiers, Web Services.
- Une **maintenance facilitée** par la centralisation des étapes du processus dans un même outil.
- Une **excellente performance**. Les batchs sont optimisés.
- Un **monitoring du processus** (statistiques, notifications d'erreurs, traces d'exécution) qui permet de s'assurer du bon déroulement des processus.
- Un **déploiement ouvert** de la solution. Les jobs sont exportables en programmes Java.

www.x2i.fr