

LIVRE BLANC

Les ETL Open Source

Une réelle alternative aux solutions propriétaires

D'années en années, l'Open Source n'a cessé de toucher de nouveaux domaines d'applications.

Après la conquête des systèmes d'exploitation (Linux), des bases de données (MySQL, PostgreSQL), de l'internet (serveurs, outils de gestion de contenu web), de nombreuses applications métiers Open Source ont également vu le jour.

Ainsi, des ERP, des outils de gestion de contenu d'entreprise (ECM), de relation client (CRM) ont fait leur apparition. Pour ne citer qu'eux : [Alfresco](#), [Compiere](#), [SugarCRM](#).

Depuis peu, c'est le domaine de la Business Intelligence qui a vu l'apparition des logiciels libres, ceux-ci couvrant tous les aspects du décisionnel: reporting, dashboards, analyse multi-dimensionnelle, data-mining, et bien sûr ETL.

C'est dans un contexte de concentration (*) des grands acteurs commerciaux du domaine et du dynamisme actuel de l'OSBI (**) que nous vous présentons ce livre blanc sur « Les ETL Open Source ».

Sylvain DECLOIX - Responsable Pôle OSBI
Atol Conseils et Développements
www.atolcd.com

(*) rachats en 2007 de Hyperion par Oracle, de Business Objects par SAP, et de Cognos par IBM
(**) OSBI: Open Source Business Intelligence



Ce livre blanc est publié sous licence **Creative Commons** :
« Paternité, Pas de modification, Pas d'utilisation commerciale – 2.0 France »
Détail des conditions: <http://creativecommons.org/licenses/by-nc-nd/2.0/fr>

Table des matières

1	INTRODUCTION.....	3
2	QU'EST-CE QU'UN « ETL » ?.....	4
3	POURQUOI UTILISER UN ETL OPEN SOURCE ?.....	6
4	NOTRE COMPARATIF « TALEND OPEN STUDIO » / « PENTAHO DATA INTEGRATION ».....	7
4.1	Présentation de Talend Open Studio.....	7
4.2	Présentation de Pentaho Data Integration.....	10
4.3	Comparatif des Fonctionnalités.....	15
4.3.1	Accès aux données.....	15
	Accès aux données relationnelles (SGBD).....	15
	Fichiers plats.....	15
	Connecteurs applicatifs.....	15
	Autres.....	15
4.3.2	Déclenchement des processus.....	16
	Déclenchement par message.....	16
	Déclenchement par type de polling.....	16
4.3.3	Traitement des données.....	16
	Transformations et calculs par défaut.....	16
	Transformations manuelles.....	16
4.3.4	Caractéristiques spécifiques.....	17
4.3.5	Développement avancé.....	17
4.3.6	Déploiement/Mise en production.....	17
4.3.7	Administration.....	18
4.3.8	Gestion de la sécurité.....	18
4.4	Comparatif des temps de traitements.....	19
4.4.1	Méthodologie de réalisation des tests.....	19
4.4.2	TEST N°1.....	20
4.4.3	TEST N°2.....	22
4.4.4	TEST N°3.....	24
4.4.5	TEST N°4.....	26
4.4.6	TEST N°5.....	30
4.4.7	TEST N°6.....	33
4.4.8	TEST N°7.....	36
5	CONCLUSION.....	39
6	ANNEXE 1.....	40
7	ANNEXE 2.....	41
7.1	Un Besoin et un Projet client.....	41
7.2	Mise en place technique.....	42
7.3	Résultat sur un extranet	46
7.4	Feedback sur l'utilisation de « Pentaho Data Integration ».....	47
7.5	Retour Client.....	47

1 INTRODUCTION

Depuis quelques années, des ETL Open Source ont fait leur apparition au milieu de la multitude des ETL propriétaires (voir liste en annexe 1).

Ces ETL ont notamment profité de l'émergence de l'OSBI (« *Open Source Business Intelligence* ») et de l'apparition de nouvelles solutions décisionnelles Open Source ambitieuses et déjà matures.

Des sociétés telles que Pentaho (www.pentaho.org), SpagoBI (www.spagobi.org), JasperIntelligence (www.jasperforge.org), Talend (www.talend.com) ont ainsi contribué à l'émergence et la démocratisation de ces ETL Open Source.

Parmi ceux-ci, nous avons voulu effectuer une analyse comparative détaillée de deux d'entre eux :

- « Pentaho Data Integration » (« Kettle » à l'origine)
- « Talend Open Studio »

Ces deux ETL Open Source nous paraissent en effet à l'heure actuelle les plus intéressants en termes :

- de fonctionnalités proposées,
- de maturité,
- de pérennité.

Pentaho Data Integration (« PDI ») et Talend Open Studio (« TOS ») peuvent répondre de façon équivalente à la plupart des ETL propriétaires disponibles sur le marché.

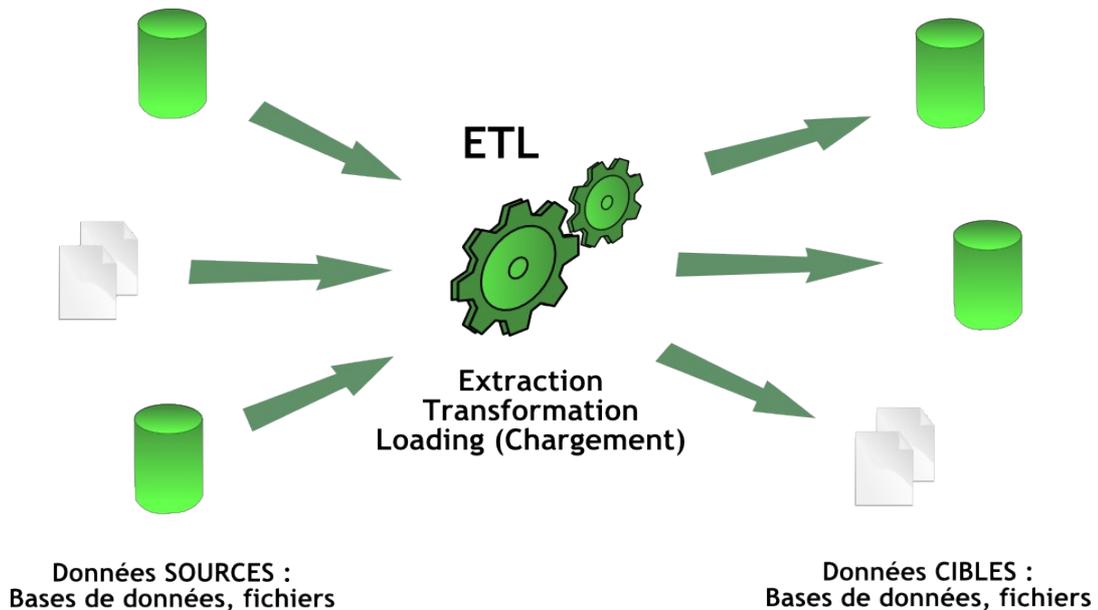
[Atol Conseils et Développements](#) est ainsi en mesure de vous apporter conseil, support et assistance dans la mise en oeuvre de projets spécifiques utilisant l'un de ces 2 ETL Open Source.

Vous trouverez en Annexe 2 un exemple de projet réalisé avec Pentaho Data Integration.

2 QU'EST-CE QU'UN « ETL » ?

ETL est l'acronyme de « **Extract - Transform -Load** ».

Un ETL permet ainsi l' *Extraction*, la *Transformation* et le *Chargement* de données depuis des sources diverses (bases de données, fichiers) vers des cibles préalablement définies.



De nombreux systèmes de gestion de bases de données sont supportés nativement en lecture/écriture (Oracle, MS Sql Server, DB2, Postgresql, MySQL,...).

De nombreux types de fichiers peuvent également être lus ou écrits: Csv, Excel, Txt, Xml, ...

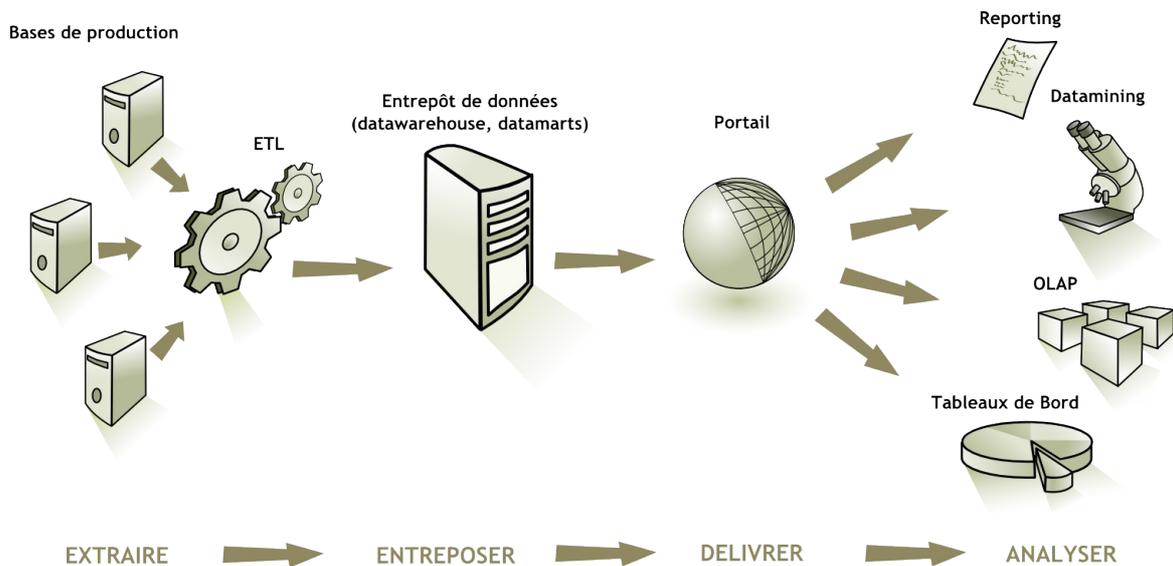
Notons que la plupart des ETL disposent d'une interface graphique permettant l'élaboration des différents scénarios d'intégration.

Le travail des développeurs en est ainsi grandement facilité, tant au niveau de la conception que de la maintenance des traitements de données.

Les ETL sont communément utilisés dans l'informatique décisionnelle afin de permettre l'alimentation des datawarehouses (entrepôts de données).

Ces derniers servent de supports pour l'analyse des données sous plusieurs formes :

- Rapports et états,
- Tableaux de bords (dashboards, balanced scorecard),
- Indicateurs de performance (« KPIs »),
- Analyse multi-dimensionnelle (OLAP) ,
- Analyse exploratoire (Data-Mining).



Ainsi, les ETL proposent dans la plupart des cas des fonctionnalités propres à l'alimentation des datawarehouses :

- Création de clés techniques (« Surrogate keys »)
- Mise à jour de type « dimension Lente » (« Slow Changing Dimension »)
- Agrégation de données
- Alimentation de base multidimensionnelles (tables de faits et dimensions)

Les volumes de données traités sont toujours très importants, ainsi les critères de performance sont primordiaux dans le choix d'un ETL.

3 POURQUOI UTILISER UN ETL OPEN SOURCE ?

Le choix le plus difficile dans tout projet décisionnel ou d'intégration/migration de données consiste à déterminer quelle méthode doit être mise en oeuvre :

1. Faut-il créer du code spécifique (procédures SQL, code Java ou autre) ?
2. Faut-il acheter un ETL propriétaire (Informatica, Oracle Warehouse Builder, BO Data Integrator ou autre) ?

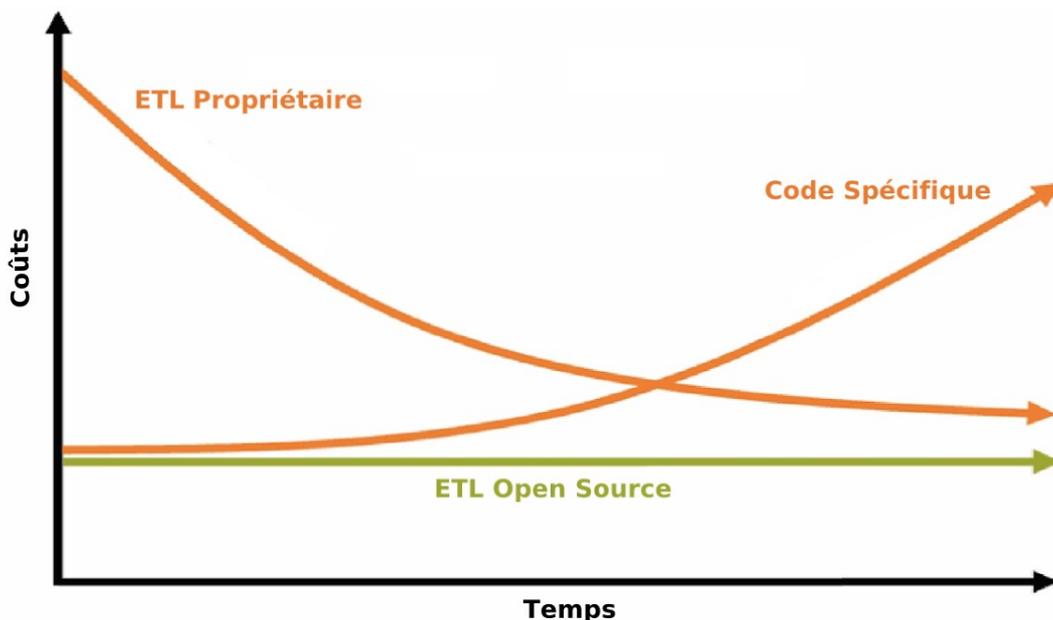
La première solution semble intéressante, car elle permet de rester au plus près des spécificités métiers des données à traiter, tout en s'affranchissant des contraintes liées à l'achat et l'utilisation d'un ETL propriétaire. Cependant, cette solution peut s'avérer coûteuse à long terme, tout simplement car l'évolutivité constante des données métier entraîne une nécessaire adaptation des traitements d'intégration. Celle-ci n'est pas toujours facile à gérer, surtout si les équipes projets évoluent au cours du temps.

La deuxième solution va permettre de mettre en oeuvre très rapidement les traitements d'intégration, avec cependant des coûts élevés (achat des licences, formations,...) et ceci dès la phase de démarrage du projet.

Il existe désormais une solution alternative: **Utiliser un ETL Open Source.**

On bénéficie ainsi des avantages d'un ETL tout en gardant une maîtrise lissée des coûts.

Ces derniers sont en effet réduits aux coûts de formation initiale de l'outil et d'une éventuelle souscription à une hot-line technique. Aucune licence n'est à payer dans ce modèle économique.



(Graphique extrait de la doc. technique Pentaho)

C'est donc dans cette 3ème approche que se positionnent les ETL « Talend Open Studio » et « Pentaho Data Integration »

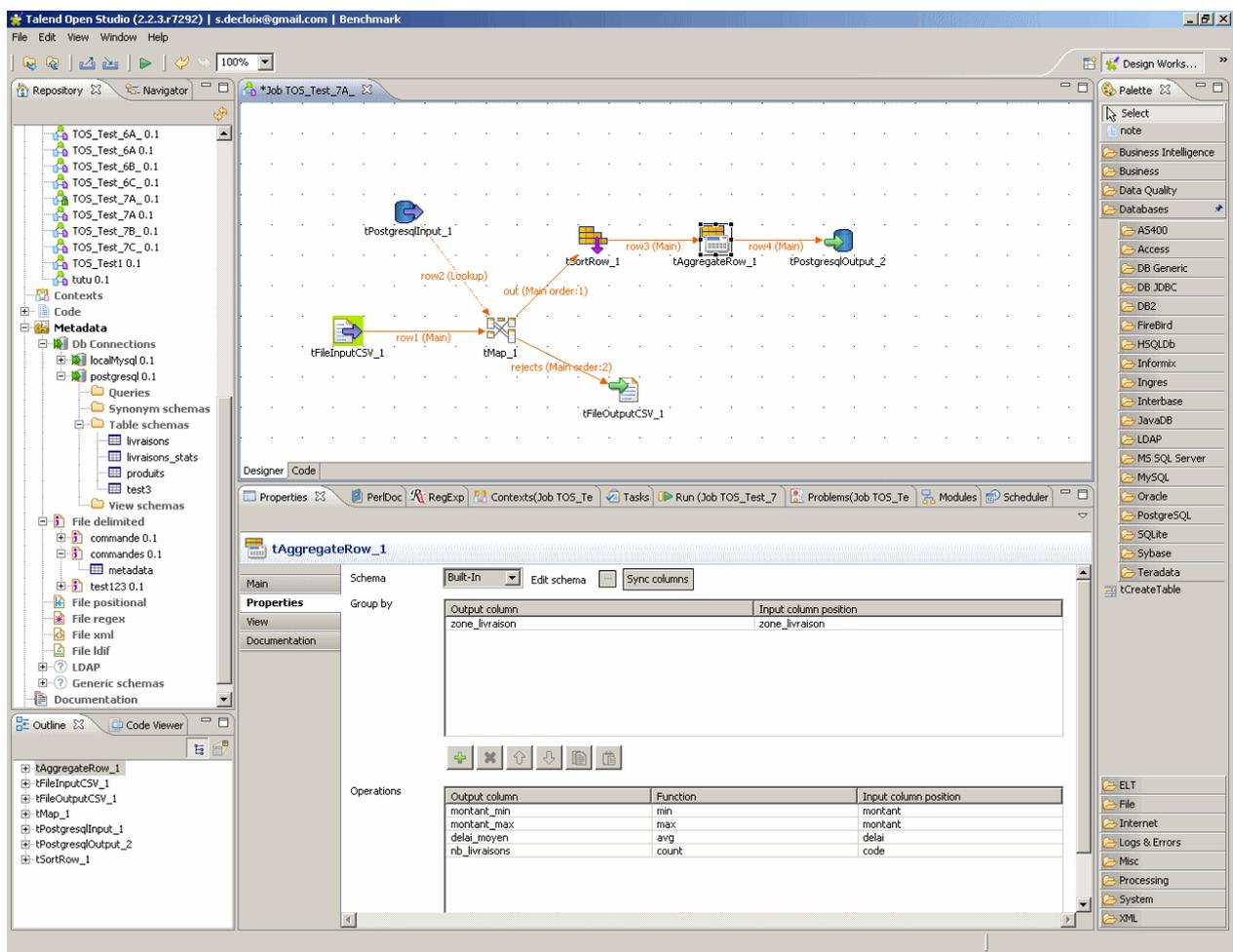
4 NOTRE COMPARATIF « TALEND OPEN STUDIO » / « PENTAHO DATA INTEGRATION »

4.1 PRÉSENTATION DE TALEND OPEN STUDIO

Talend Open Studio est développé par Talend (www.talend.com), une société française dynamique et relativement jeune. La première version de « Talend Open Studio » a vu le jour au 2ème semestre 2006, et la version actuelle est la 2.3.

Talend Open Studio est un ETL du type « générateur de code ». Pour chaque traitement d'intégration de données, un code spécifique est généré, ce dernier pouvant être en Java ou en Perl. Les données traitées et les traitements effectués sont donc intimement liés.

Talend Open Studio utilise une interface graphique, le « Job Designer » (basée sur [Eclipse RCP](http://www.eclipse.org/rpc/)) qui permet la création des processus de manipulation de données :



De nombreux types d'étapes sont disponibles pour se connecter aux principaux SGBD (Oracle, DB2, MS SQL Server, PostgreSQL, MySQL,...) ainsi que pour traiter tous les types de fichiers plats (CSV, Excel, XML), aussi bien en lecture qu'en écriture.

Talend facilite la construction des requêtes dans les bases de données en détectant le schéma et les relations entre tables.

Un référentiel permet de stocker les méta-données afin de pouvoir les exploiter dans différents jobs.

Par exemple on peut sauvegarder le type et le format des données d'entrée d'un fichier CSV afin de pouvoir les exploiter ultérieurement.

Une gamme complète de composants :

Le Job Designer intègre une « Component Library »: une palette graphique de composants et connecteurs.

Les processus d'intégration sont construits simplement en déposant des composants et connecteurs sur le diagramme, en dessinant leurs connexions et relations, et en modifiant leurs propriétés.

La plupart de ces propriétés peut être issue des métadonnées déjà définies.

La Component Library inclut plus de 80 composants et connecteurs, fournissant des fonctions basiques telles que des associations, transformations, agrégation et recherches; des fonctions spécialisées comme le filtrage de données, le multiplexage de données...

Cette librairie supporte tous les principaux SGBDR, formats de fichiers, annuaires LDAP...

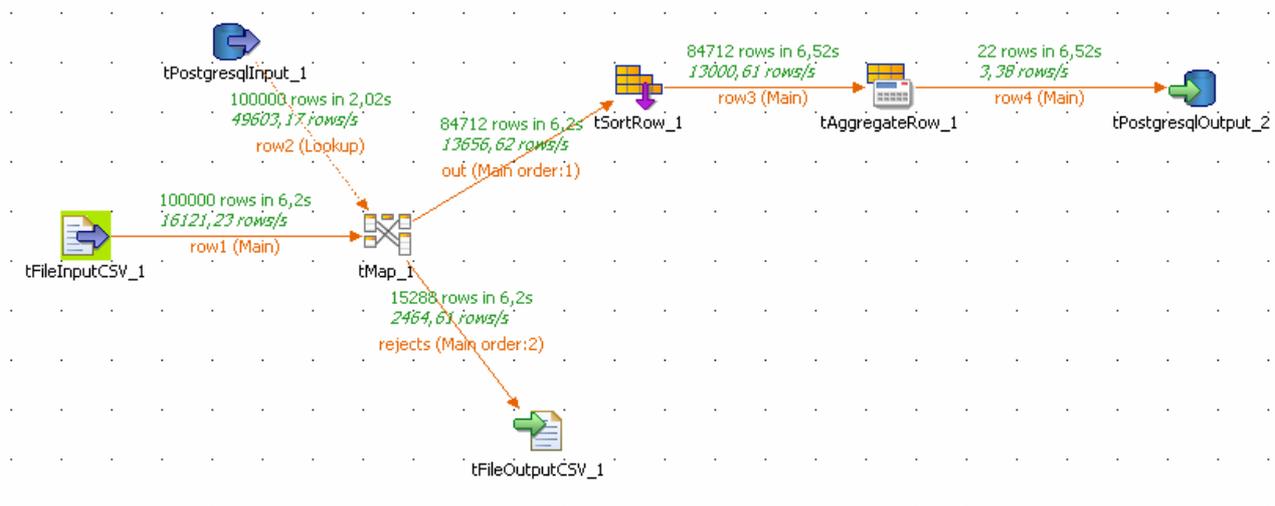
La Component Library peut facilement être complétée en utilisant des langages standards tels que Perl, Java ou SQL.

Des traces et statistiques d'exécution en temps réel :

La conception très visuelle des "jobs" permet de présenter des statistiques d'exécution en temps réel ou encore de tracer les données transitant ligne à ligne dans les composants de la chaîne de traitement.

Quand un job d'intégration est lancé via le Job Designer (en mode graphique), il est possible d'afficher les statistiques de traitement en temps réel, montrant le nombre de lignes traitées et rejetées, ainsi que la vitesse d'exécution (lignes par secondes). On peut ainsi repérer immédiatement les goulots d'étranglement.

Il est aussi possible d'activer un mode de traçage, qui affiche pour chaque ligne le comportement adopté et montre le résultat des transformations. Les fonctionnalités de débogage traditionnelles sont évidemment disponibles.



L'enrichissement des traitements par ajout de code spécifique :

La totalité du code généré par Talend Open Studio, quelque soit le langage cible, est toujours visible et accessible depuis l'environnement de conception.

On peut bien sûr implémenter des spécificités « métiers » propres aux données traitées, ceci en ajoutant de nouvelles « routines ».

Exemple de code généré par Talend Open Studio :

```

1 // =====
2 //
3 // Copyright (c) 2005-2007, Talend Inc.
4 //
5 // This source code has been automatically generated by Talend Open Studio
6 // / JobDesigner (CodeGenerator version 2.2.3.r7292).
7 // You can find more information about Talend products at www.talend.com.
8 // You may distribute this code under the terms of the GNU LGPL license
9 // (http://www.gnu.org/licenses/lgpl.html).
10 //
11 // =====
12 package benchmark.tos_test_7a_;
13
14 import routines.MesRoutines;
15 import routines.DataOperation;
16 import routines.Mathematical;
17 import routines.Numeric;
18 import routines.Relational;
19 import routines.StringHandling;
20 import routines.TalendDate;
21 import routines.TalendString;
22 import routines.system.*;
23
24 import java.text.ParseException;
25 import java.text.SimpleDateFormat;
26 import java.util.Date;
27 import java.util.List;
28
29 /**
30  * Job: TOS_Test_7A_Purpose: <br>
31  * Description: <br>
32  *
33  * @author s.decloix@gmail.com
34  * @version 0.1
35  * @status
36  */
37 public class TOS_Test_7A_ {
38     // create and load default properties
39     private static java.util.Properties defaultProps = new java.util.Properties();
40     // create application properties with default
41     private static java.util.Properties context = new java.util.Properties();
42
43     private static final String jobName = "TOS_Test_7A_";
44     private static final String projectName = "BENCHMARK";
45     public static Integer errorCode = null;
46     private static String currentComponent = "";
47     private static final java.util.Map<String, Long> start_Hash = new java.util.HashMap<String, Long>();
48     private static final java.util.Map<String, Long> end_Hash = new java.util.HashMap<String, Long>();
49     private static final java.util.Map<String, Boolean> ok_Hash = new java.util.HashMap<String, Boolean>();
50     private static final java.util.Map<String, Object> globalMap = new java.util.HashMap<String, Object>();
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

Intégration dans les suites décisionnelles Open Source :

Talend est partenaire des éditeurs des suites décisionnelles SpagoBI et JasperIntelligence.

4.2 PRÉSENTATION DE PENTAHO DATA INTEGRATION

[Pentaho Data Integration](#) est l'ETL de la suite décisionnelle Open Source [Pentaho](#).

Cet ETL, « Kettle » à l'origine, est le fruit du travail de [Matt Casters](#), un consultant BI qui l'a développé à l'origine pour ses propres besoins au début des années 2000.

Courant 2006, Matt Casters rejoint la société Pentaho, et « Kettle » devient « Pentaho Data Integration ».

Les nouvelles versions s'enchaînent alors à un rythme soutenu. La prochaine version est la 3.0.2 qui doit sortir courant Février 2008.

Contrairement à Talend Open Studio, Pentaho Data Integration est un « moteur de transformation » ETL: les données traitées et les traitements à effectuer sont parfaitement séparés. (on parle de « meta-data driven » ETL)

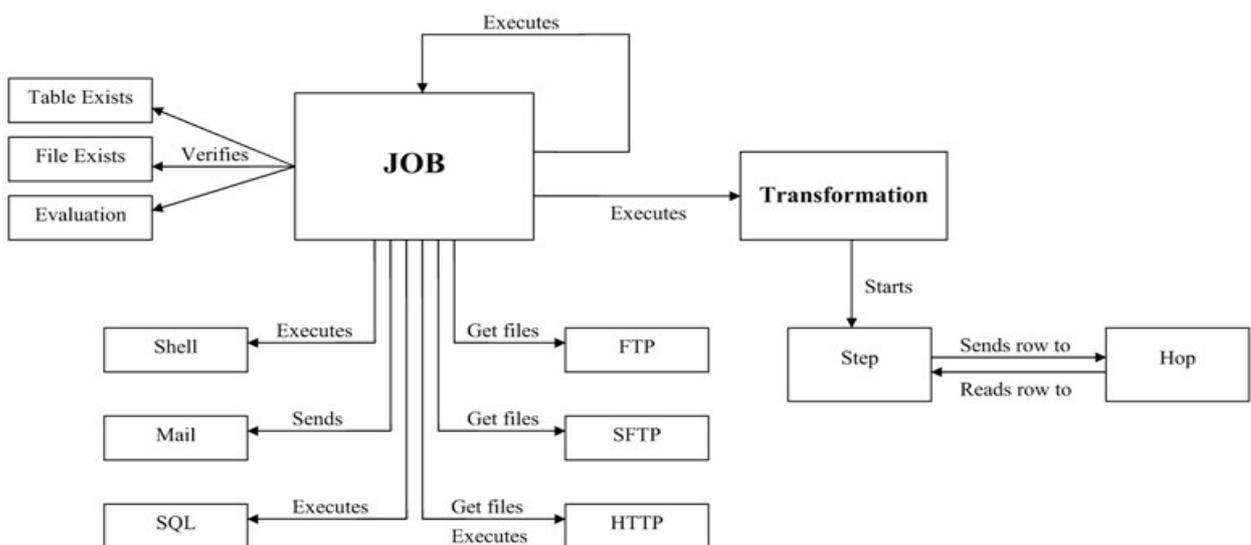
Les traitements sont stockés dans un référentiel (repository) qui peut être soit au format XML (fichiers plats), soit dans une base de données (ce qui permet notamment le partage entre plusieurs designers).

Tout comme Talend Open Studio, de nombreux types de SGBD sont supportés (une trentaine) ainsi que tous les types de fichiers plats (Csv, délimité, Excel, XML).

Pentaho Data Integration dispose d'une interface graphique « Spoon » (basée sur [SWT](#)), depuis laquelle on peut créer deux types de traitements :

- *des transformations* : celles-ci constituent les traitements de base d'intégration de données avec toutes les étapes (steps) nécessaires à l'extraction, la transformation, et le chargement des données.
- *des tâches (jobs)* : ceux-ci permettent le séquençage de plusieurs transformations avec des fonctionnalités plus orientées « EAI » : gestion des erreurs, envoi de mails de notification, transferts FTP/SFTP, exécution de scripts shell ou SQL, etc...

Kettle Conceptual Model



Caractéristiques de Pentaho Data Integration :

- Un produit sous licence GNU disponible sur plusieurs systèmes d'exploitation :

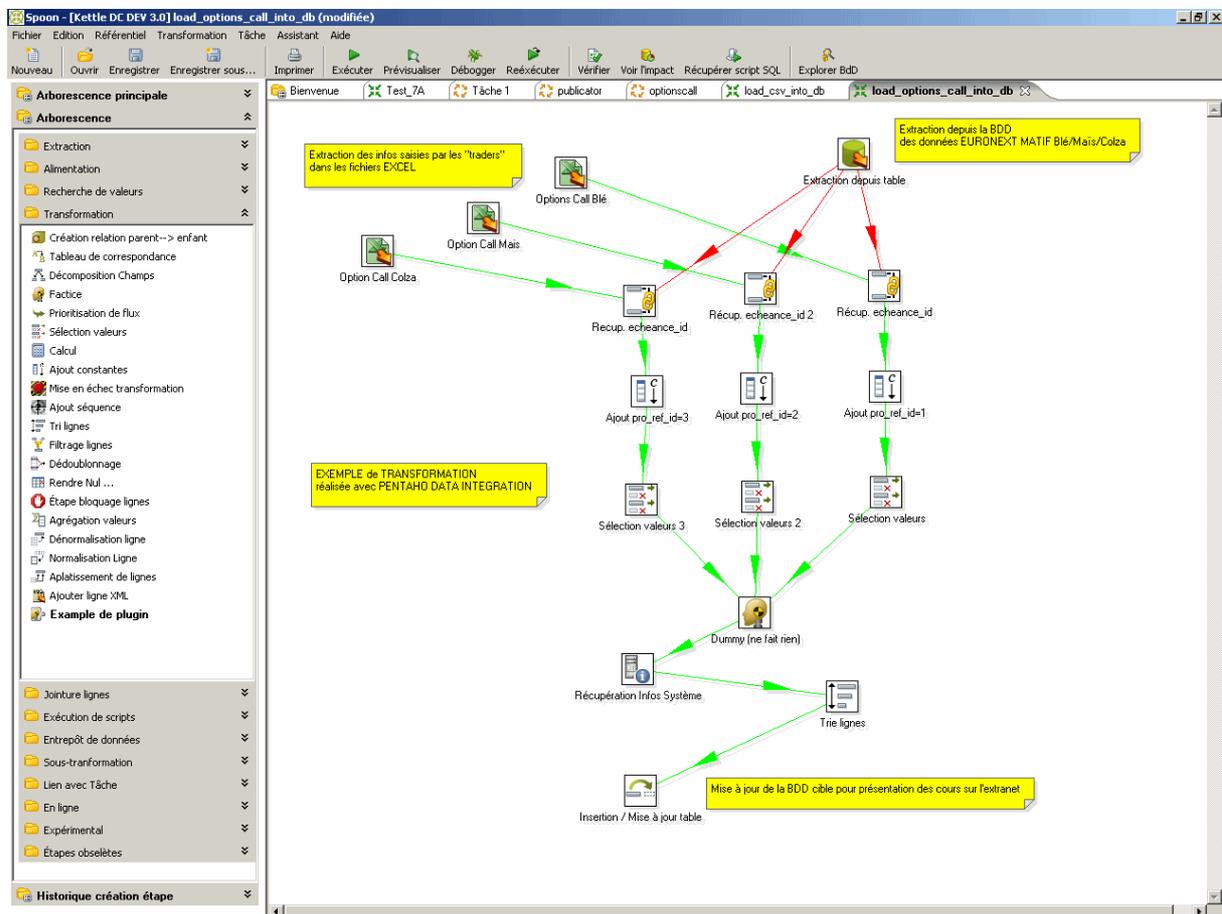
Plates-formes supportées	GUI	Exécution
Windows	✓	✓
Linux	✓	✓
Solaris	✓ *	✓
Apple OSX	✓ *	✓
HP-UX	✓ *	✓
AIX	✓ *	✓

GUI= Interface Graphique (Graphic User Interface)

* Because of the limited GUI capabilities of some of these platforms (such as Motif environments) no guarantees are given to the usability of the Kettle GUI. Possible problem are corrected on a best effort basis.

- Une suite de 3 composants: L'interface graphique « Spoon », les déclencheurs en mode batch pour les transformations (« Pan ») et les jobs (« Kitchen »).
- Une installation simple (un dossier à décompresser) => un environnement d'exécution JAVA 1.5 suffit
- La possibilité de prévisualiser les flux de données traitées, et ceci pour une étape donnée.
- La possibilité d'exécuter les traitements sur le poste local, un serveur distant, ou un ensemble de serveurs (exécution en « grappe »; clustering)
- La possibilité de logger les traitements dans une base de données spécifique.
- L'intégration parfaite avec la plate-forme décisionnelle Pentaho.** Par exemple, les flux de données en provenance de l'ETL peuvent servir à alimenter des rapports ou des dashboards en temps réel.

Spoon, l'interface graphique de création des transformations et jobs :



Briques de traitements de données (liste non exhaustive) :

<i>Step</i>	<i>Etape</i>	<i>Description</i>
Extraction depuis EXCEL		Extraction de données depuis un ou plusieurs onglets d'un fichier XLS
Extraction depuis ACCESS		Extraction de données depuis une table d'une base Access
Extraction depuis une base de données		Extraction de données depuis une table d'un SGBD (Oracle, MySQL, Postgres, SQLServer, DB2, AS400, SAP, etc...) : 30 SGBD supportés via JDBC et/ou ODBC
Extraction depuis un fichier		Extraction de données depuis un fichier CSV ou de type « délimité »
Extraction depuis un fichier XML		Extraction de données depuis un fichier XML
Extraction depuis un annuaire		Extraction de données depuis un annuaire de type LDAP
Alimentation base de données		Insertion ou mise à jour d'une table d'un SGBD (insert/update)
Recherche dans base de données		Recherche des enregistrements dans une base de données selon une liste de valeurs
Recherche dans un flux		Recherche des enregistrements dans un flux de données selon une liste de valeurs
Normalisation Ligne		Normalise des informations
Dénormalisation ligne		Dénormalisation de lignes
Ajout séquence		Ajoute une séquence (calculée ou récupérée depuis une base de données)
Filtrage de ligne		Permet de diriger les flux de données vers 2 cibles différentes selon un ou plusieurs critères
Agrégation de données		Permet de réaliser des calculs d'agrégation sur un ensemble de lignes (somme, moyenne, min, max, count, etc...)
Calculs		Permet de créer des données calculées à partir des données traitées

Stockage des traitements au format XML :

Les jobs et transformations sont stockés dans un meta-langage, qui peut être soit stocké au format XML, soit dans une base de données.

```

<?xml version="1.0" encoding="UTF -8" ?>
- <transformation >
+ <info >
  <notepads />
- <connection >
  <name>postgres</name>
  <server>localhost</server>
  <type>POSTGRESQL</type>
  <access>Native</access>
  <database>whitebook</database>
  <port>5432</port>
  <username>postgres</username>
  <password>Encrypted 2be98afc86aa7f2e4bb16bd64d980aac9</password>
  <servername />
  <data_tablespace />
  <index_tablespace />
+ <attributes >
</connection >
- <order>
- <hop>
  <from>Exécution Scripts Java</from>
  <to>insert</to>
  <enabled>Y</enabled>
</hop>
- <hop>
  <from>Recherche dans Base de données</from>
  <to>Exécution Scripts Java</to>
  <enabled>Y</enabled>
</hop>
- <hop>
  <from>Extraction depuis fichier CSV</from>
  <to>Filtrage lignes</to>
  <enabled>Y</enabled>
</hop>
+ <hop>
+ <hop>
+ <hop>
+ <hop>
+ <hop>
</order>
+ <step >
+ <step >
- <step >
  <name>Extraction depuis fichier CSV</name>
  <type>CsvInput</type>
  <description />
  <distributed >Y</distributed >
  <copies >1</copies >
- <partitioning >
  <method>none</method >
  <schema_name />
</partitioning >
  <filename>C:\Pentaho\_input_files\cmde_100000_lignes.csv</filename>
  <separator >;</separator >
  <enclosure >"</enclosure >

```

4.3 COMPARATIF DES FONCTIONNALITÉS

4.3.1 Accès aux données

Accès aux données relationnelles (SGBD)

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Lecture de table complète	✓	✓
Lecture de vue complète	✓	✓
Appel de procédure stockées	✓	Seulement sur certains SGBD
Ajout de clause where/order by	✓	✓
Exécution de requête	✓	✓
Outil de création de requête	✗	✓
Lecture/écriture de tous les types simples de données	✓	✓
Lecture/écriture de types complexes de données	✗	Données cartographiques

Fichiers plats

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
CSV	✓	✓
Fixed/Limité	✓	✓
XML	✓	✓
Excel	✓	✓
Validité des fichiers plats	✗	✓
Validité des fichiers XML	✓	✓

Connecteurs applicatifs

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Quels connecteurs ?	SAP (lecture)	Connecteurs CRM (sugarCRM, Salesforce, ..) Connecteurs ERP

Autres

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Web Services	✓	✓
Cubes OLAP (Mondrian)	✓	✓
Divers	LDAP	RSS, LDAP, MOM, POP, SCP, XMLRPC

4.3.2 Déclenchement des processus

Déclenchement par message

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
CORBA	✘	✘
XML RPC	✘	✓
JMS	✘	✘
MOMS	✘	✓

Déclenchement par type de polling

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Répertoire	✓	✓
POP	✓	✓

4.3.3 Traitement des données

Transformations et calculs par défaut

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Fonctions de transformations des dates et des nbres	✓	✓
Fonctions statistiques de qualité	✘	✓
Permet le transcodage par une table de référence	✘	✘
Jointures hétérogènes	✘	✘
Modes de jointure supportées (BD)	Externe	Jointure de flux seulement
Gestion des requêtes imbriquées	✘	✘

Attention : les étapes marquées « NON » ci-dessus sont seulement pour la prise en charge automatique. Les 2 outils fournissent, bien sûr, un mécanisme de requêtage directement en SQL qui permet de faire toutes les jointures et les requêtes imbriquées...

Transformations manuelles

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Possibilité de traitements par un langage de programmation	✓	✓
Ajout de nouvelles transformations et processus métiers	✓	✓

4.3.4 Caractéristiques spécifiques

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Mapping graphique	✓	✓
« Drag and drop » (Glisser/déposer)	✓	✓
Représentation graphique des flux	✓	✓
Visualisation des données en cours de développement	✓	✓
Outil d'analyse d'impact	✓	✗
Outils de debugging	✓	✓
Génération de documentation technique	✗	✓
Génération de documentation fonctionnelle	✗	✓
Consultation de la documentation à travers le web	✗	✓
Gestion des erreurs d'intégration	Oui, pour certaines étapes	✓

4.3.5 Développement avancé

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Présence d'une API	✓	✓
Intégration de fonctions externes	✓	✓
Mécanisme de reprise sur incident	✗	✗
Paramétrage des buffers/indexes/caches	✓	✓
Gestion du développement en équipe	✓	Oui, mais payant
Versionning	✗	✓

4.3.6 Déploiement/Mise en production

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Compilation des traitements	NON	OUI pour JAVA NON pour PERL
Type de mise en production	Ligne de commande windows ou unix	
Visualisation de l'historique des mise en production	✗	✗

4.3.7 Administration

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Console d'administration	✓	✓
Gestion automatisée des logs	✓	✓
Génération de log spécifique	✗	✗
Interfaçage avec des outils de supervision	✗	✗
Outil de planification des traitements intégré	✗	✓

4.3.8 Gestion de la sécurité

	<i>Pentaho Data Integration</i>	<i>Talend Open Studio</i>
Utilisation des droits d'un annuaire	✗	✗
Type de sécurité	Sécurité du SGBD qui contient le référentiel	Propriétaire
Sécurité sur la création de scénario	✓	✓
Sécurité sur la maj de scénario	✓	✓
Sécurité sur l'accès aux métadonnées	✓	✓
Sécurité sur la console d'administration	✓	✓
Sécurité sur le lancement manuel des tâches	✓	✗

4.4 COMPARATIF DES TEMPS DE TRAITEMENTS

4.4.1 Méthodologie de réalisation des tests

Les performances des temps de traitements sont un critère important dans le choix d'un ETL.

Les résultats des tests qui sont donnés dans les paragraphes suivants correspondent à des cas simples et ne peuvent en aucun cas préjuger des performances réelles en environnement de production.

Seuls des tests poussés sur des traitements d'intégration réels peuvent permettre de qualifier définitivement l'ETL choisi.

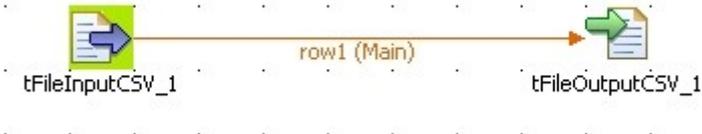
- L'ensemble des tests ont été effectués sur un PC Portable Dell

Les caractéristiques techniques sont les suivantes :

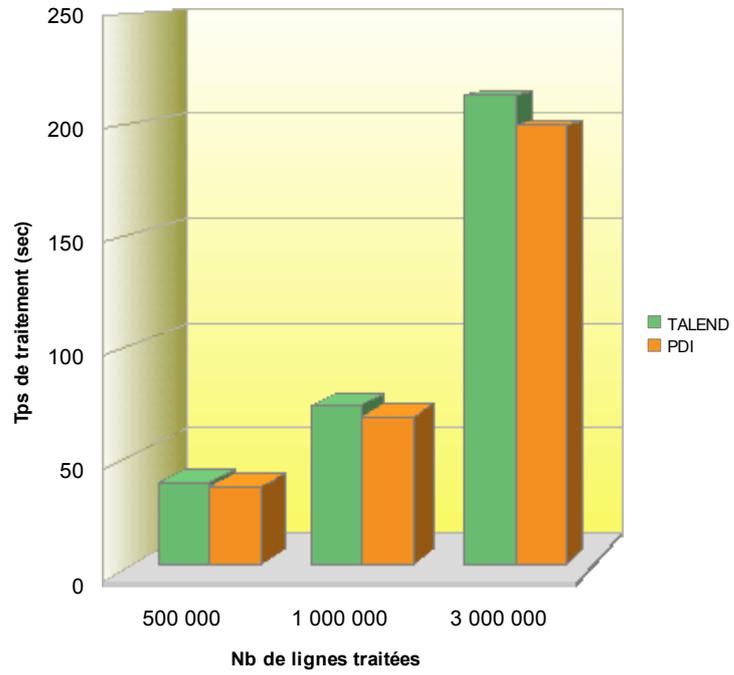
DELL Latitude D520
Intel Core DUO (1,66 Ghz)
1 Go de RAM (980 Mhz)
Windows XP Professional - SP2

- Les versions utilisées des 2 ETL sont les suivantes:
 - Pentaho Data Integration **v3.0.1**
 - Talend Open Studio **v2.2.3**
- Pour chaque test, les résultats présentés sont issu d'une moyenne de trois essais réalisés dans des conditions identiques.
- Tous les traitements sont lancés en ligne de commande.
Ainsi, aucun essai n'est lancé depuis les interfaces graphiques aussi bien pour Pentaho Data Integration que Talend Open Studio.
- Enfin, les essais sont réalisés uniquement en code généré en **Java** pour Talend Open Studio, afin que l'on puisse comparer de façon objective ceux obtenus avec Pentaho Data Integration, dont le moteur est écrit en Java.

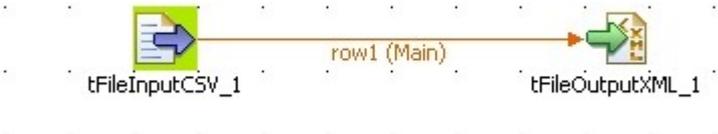
4.4.2 TEST N° 1

Descriptif	1. Extraction des données d'un fichier CSV 2. Chargement des données dans un autre fichier CSV - Le séparateur « ; » du fichier initial est remplacé par le séparateur « , » .														
Détails	Le fichier d'entrée comporte 7 champs typés sequence [integer]; now [datetime]; first [number]; second [string]; third [datetime]; fourth [boolean]; fifth [integer] 0000000001;2007/11/0510:44:43.014;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000002;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000003;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000004;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000005;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000006;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000007;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000008;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000009;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 0000000010;2007/11/0510:44:43.029;12.345;undextroisquatrecinq;0304/12/0500:00:00.000;Y;12345 etc														
Modélisation dans Pentaho Data Integration (PDI)															
Modélisation dans Talend Open Studio															
Résultats du Test (Temps de traitement exprimés en sec.)	<table border="1" data-bbox="555 1675 1339 1845"> <thead> <tr> <th rowspan="2">Nb de lignes traitées</th> <th colspan="2">Tps de traitement (secondes)</th> </tr> <tr> <th>TALEND</th> <th>PDI</th> </tr> </thead> <tbody> <tr> <td>500 000</td> <td>37</td> <td>35</td> </tr> <tr> <td>1 000 000</td> <td>72</td> <td>67</td> </tr> <tr> <td>3 000 000</td> <td>213</td> <td>199</td> </tr> </tbody> </table>	Nb de lignes traitées	Tps de traitement (secondes)		TALEND	PDI	500 000	37	35	1 000 000	72	67	3 000 000	213	199
Nb de lignes traitées	Tps de traitement (secondes)														
	TALEND	PDI													
500 000	37	35													
1 000 000	72	67													
3 000 000	213	199													

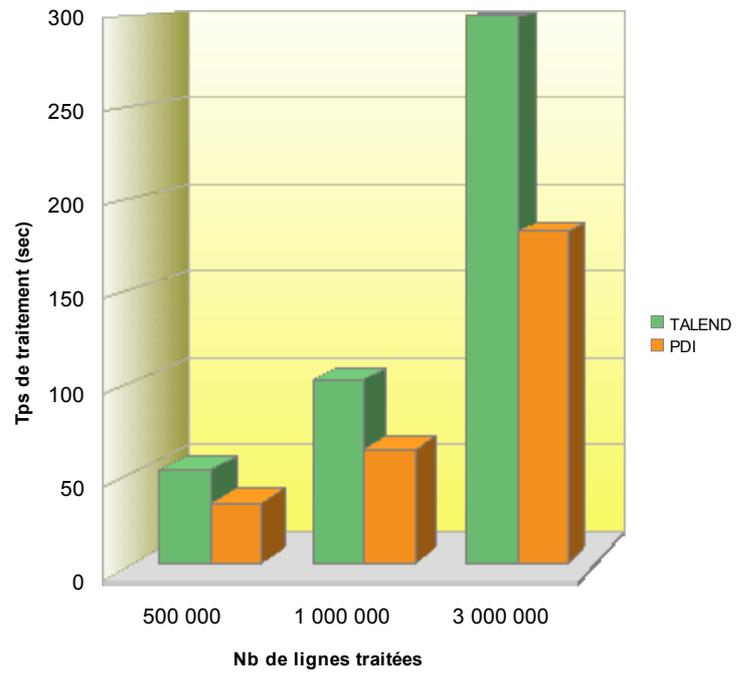
Graphe
Comparatif



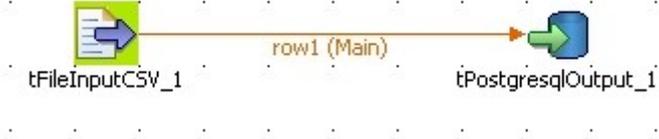
4.4.3 TEST N°2

Descriptif	1. Extraction des données d'un fichier CSV 2. Chargement des données dans un fichier XML														
Détails	<p>Le fichier d'entrée comporte 7 champs typés (même fichier que Test1) Le fichier de sortie est un fichier XML dont la structure est la suivante :</p> <pre> - <Rows> - <Row> <sequence>0000000001</sequence> <now>2007/11/0212:47:19.093</now> <first>12.345</first> <second>undeuxtroisquatrecinq</second> <third>0304/12/0500:00:00.000</third> <fourth>Y</fourth> <fifth>12345</fifth> </Row> - <Row> <sequence>0000000002</sequence> <now>2007/11/0212:47:19.101</now> <first>12.345</first> <second>undeuxtroisquatrecinq</second> <third>0304/12/0500:00:00.000</third> <fourth>Y</fourth> <fifth>12345</fifth> </Row> - <Row> <sequence>0000000003</sequence> <now>2007/11/0212:47:19.101</now> <first>12.345</first> </pre>														
Modélisation dans Pentaho Data Integration (PDI)															
Modélisation dans Talend Open Studio															
Résultats du Test (Temps de traitement exprimés en sec.)	<table border="1" data-bbox="555 1686 1337 1854"> <thead> <tr> <th rowspan="2">Nb de lignes traitées</th> <th colspan="2">Tps de traitement (secondes)</th> </tr> <tr> <th>TALEND</th> <th>PDI</th> </tr> </thead> <tbody> <tr> <td>100 000</td> <td>51</td> <td>33</td> </tr> <tr> <td>500 000</td> <td>100</td> <td>62</td> </tr> <tr> <td>1 000 000</td> <td>300</td> <td>182</td> </tr> </tbody> </table>	Nb de lignes traitées	Tps de traitement (secondes)		TALEND	PDI	100 000	51	33	500 000	100	62	1 000 000	300	182
Nb de lignes traitées	Tps de traitement (secondes)														
	TALEND	PDI													
100 000	51	33													
500 000	100	62													
1 000 000	300	182													

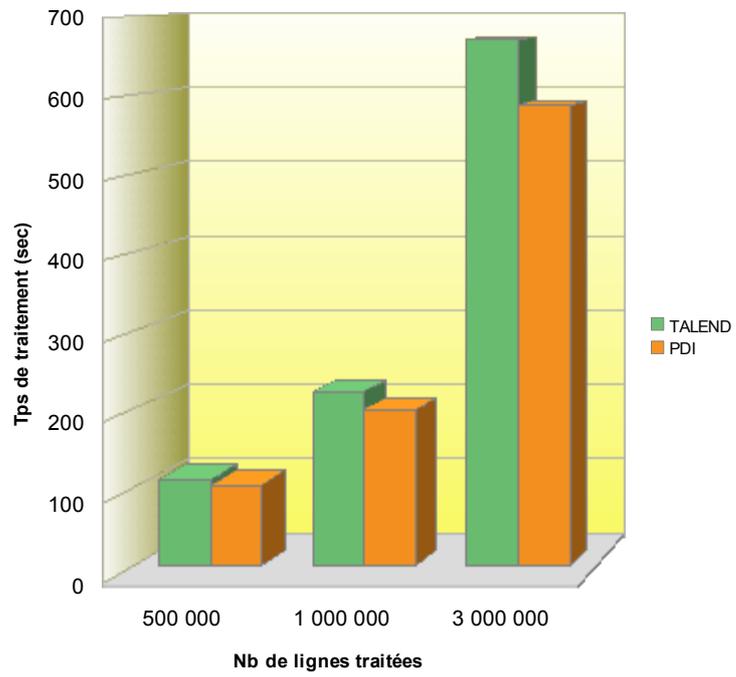
Graphe
Comparatif



4.4.4 TEST N° 3

Descriptif	1. Extraction des données d'un fichier CSV 2. Chargement des données dans une table Postgresql																																																																																										
Détails	<p>Le fichier d'entrée comporte 7 champs typés (même fichier que Test 1 et 2) La table Postgresql chargée comporte les champs suivants :</p> <table border="1" data-bbox="435 562 1257 824"> <thead> <tr> <th>Fields</th> <th>Foreign Keys</th> <th>Checks</th> <th>Indices</th> <th>Triggers</th> <th>Rules</th> <th>Dependencies</th> <th>Data</th> <th>Description</th> <th>DDL</th> </tr> <tr> <th>Field Name</th> <th>Field Type</th> <th>Key</th> <th>Not Null</th> <th>Default</th> <th colspan="5"></th> </tr> </thead> <tbody> <tr> <td>sequence</td> <td>integer</td> <td></td> <td><input type="checkbox"/></td> <td>Null</td> <td colspan="5"></td> </tr> <tr> <td>now</td> <td>date</td> <td></td> <td><input type="checkbox"/></td> <td>Null</td> <td colspan="5"></td> </tr> <tr> <td>first</td> <td>double precision</td> <td></td> <td><input type="checkbox"/></td> <td>Null</td> <td colspan="5"></td> </tr> <tr> <td>second</td> <td>varchar(30)</td> <td></td> <td><input type="checkbox"/></td> <td>Null</td> <td colspan="5"></td> </tr> <tr> <td>third</td> <td>date</td> <td></td> <td><input type="checkbox"/></td> <td>Null</td> <td colspan="5"></td> </tr> <tr> <td>fourth</td> <td>varchar(1)</td> <td></td> <td><input type="checkbox"/></td> <td>Null</td> <td colspan="5"></td> </tr> <tr> <td>fifth</td> <td>integer</td> <td></td> <td><input type="checkbox"/></td> <td>Null</td> <td colspan="5"></td> </tr> </tbody> </table>	Fields	Foreign Keys	Checks	Indices	Triggers	Rules	Dependencies	Data	Description	DDL	Field Name	Field Type	Key	Not Null	Default						sequence	integer		<input type="checkbox"/>	Null						now	date		<input type="checkbox"/>	Null						first	double precision		<input type="checkbox"/>	Null						second	varchar(30)		<input type="checkbox"/>	Null						third	date		<input type="checkbox"/>	Null						fourth	varchar(1)		<input type="checkbox"/>	Null						fifth	integer		<input type="checkbox"/>	Null					
Fields	Foreign Keys	Checks	Indices	Triggers	Rules	Dependencies	Data	Description	DDL																																																																																		
Field Name	Field Type	Key	Not Null	Default																																																																																							
sequence	integer		<input type="checkbox"/>	Null																																																																																							
now	date		<input type="checkbox"/>	Null																																																																																							
first	double precision		<input type="checkbox"/>	Null																																																																																							
second	varchar(30)		<input type="checkbox"/>	Null																																																																																							
third	date		<input type="checkbox"/>	Null																																																																																							
fourth	varchar(1)		<input type="checkbox"/>	Null																																																																																							
fifth	integer		<input type="checkbox"/>	Null																																																																																							
Modélisation dans Pentaho Data Integration (PDI)	 <p>Extraction depuis fichier CSV → Insertion dans table</p>																																																																																										
Modélisation dans Talend Open Studio	 <p>tFileInputCSV_1 → row1 (Main) → tPostgresqlOutput_1</p>																																																																																										
Résultats du Test (Temps de traitement exprimés en sec.)	<table border="1" data-bbox="555 1404 1339 1574"> <thead> <tr> <th rowspan="2">Nb de lignes traitées</th> <th colspan="2">Tps de traitement (secondes)</th> </tr> <tr> <th>TALEND</th> <th>PDI</th> </tr> </thead> <tbody> <tr> <td>500 000</td> <td>111</td> <td>103</td> </tr> <tr> <td>1 000 000</td> <td>221</td> <td>198</td> </tr> <tr> <td>3 000 000</td> <td>670</td> <td>587</td> </tr> </tbody> </table>	Nb de lignes traitées	Tps de traitement (secondes)		TALEND	PDI	500 000	111	103	1 000 000	221	198	3 000 000	670	587																																																																												
Nb de lignes traitées	Tps de traitement (secondes)																																																																																										
	TALEND	PDI																																																																																									
500 000	111	103																																																																																									
1 000 000	221	198																																																																																									
3 000 000	670	587																																																																																									

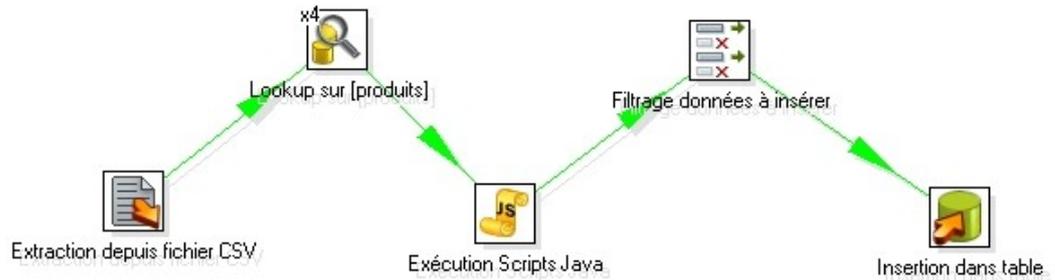
Graphe
Comparatif



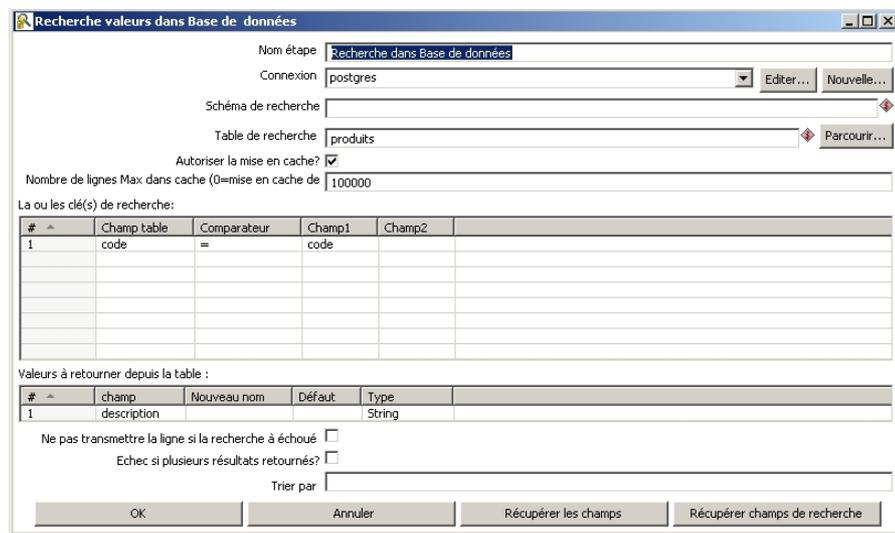
4.4.5 TEST N°4

Descriptif	<p>1. Extraction des données d' un fichier « commandes » (format) CSV</p> <p>2. Chargement des données dans une table [livraisons] avec calcul du délai de livraisons et du montant global de chaque commande. Récupération du nom littéral du produit via un « lookup » sur la table [produits] à partir de son code.</p>																																																																																																																																																		
Détails	<p>Le fichier « commandes.csv » possède la structure suivante. Il existe 26 zone de livraisons (A, B, C, D,, X, Y, Z) :</p> <table border="1" data-bbox="400 571 1402 904"> <thead> <tr> <th>code</th> <th>prix_unit</th> <th>qte</th> <th>zone_livraison</th> <th>date_cmde</th> <th>date_livr</th> </tr> </thead> <tbody> <tr><td>24489</td><td>535,57</td><td>5</td><td>M</td><td>11/11/07</td><td>04/12/07</td></tr> <tr><td>12137</td><td>756,15</td><td>11</td><td>J</td><td>06/11/07</td><td>05/12/07</td></tr> <tr><td>84636</td><td>245,58</td><td>6</td><td>A</td><td>07/11/07</td><td>11/12/07</td></tr> <tr><td>67397</td><td>732,87</td><td>3</td><td>P</td><td>17/11/07</td><td>27/12/07</td></tr> <tr><td>83942</td><td>584,77</td><td>2</td><td>Y</td><td>01/11/07</td><td>01/12/07</td></tr> <tr><td>93948</td><td>1014,06</td><td>8</td><td>J</td><td>14/11/07</td><td>21/12/07</td></tr> <tr><td>90253</td><td>1022,96</td><td>11</td><td>L</td><td>02/11/07</td><td>24/12/07</td></tr> <tr><td>78216</td><td>988,43</td><td>6</td><td>E</td><td>13/11/07</td><td>27/12/07</td></tr> <tr><td>81942</td><td>472,17</td><td>5</td><td>D</td><td>16/11/07</td><td>12/12/07</td></tr> </tbody> </table> <p>La table [produits] possède la structure suivante. Cette table (générée aléatoirement pour les tests) contient le code et la description de 100 000 produits :</p> <table border="1" data-bbox="400 1032 916 1395"> <thead> <tr> <th>code</th> <th>description</th> </tr> </thead> <tbody> <tr><td>1</td><td>PRD-DSEIWA\WG</td></tr> <tr><td>2</td><td>PRD-WOQGM\X\MV</td></tr> <tr><td>3</td><td>PRD-ABUT\FOMB</td></tr> <tr><td>4</td><td>PRD-JWPTZ\VHR</td></tr> <tr><td>5</td><td>PRD-FFHIC\NFD</td></tr> <tr><td>6</td><td>PRD-HNNQ\AVLD</td></tr> <tr><td>7</td><td>PRD-NARQU\DDC</td></tr> <tr><td>8</td><td>PRD-YUGR\WMBL</td></tr> <tr><td>9</td><td>PRD-CUBLU\XJA</td></tr> <tr><td>10</td><td>PRD-IHTOV\CUU</td></tr> <tr><td>11</td><td>PRD-FEJOI\CON</td></tr> <tr><td>12</td><td>PRD-BUXBE\RSF</td></tr> </tbody> </table> <p>La table cible [livraisons] doit contenir les données ci-dessous après traitement :</p> <p>[livraisons].[montant]= {prix_unit} x {qte}</p> <p>[livraisons].[delai]= {date_livr} - {date_cmde}</p> <p>[livraisons].[desc_complete]= Texte comprenant {description produit} et {zone_livraison}</p> <table border="1" data-bbox="384 1644 1428 2000"> <thead> <tr> <th>code</th> <th>montant</th> <th>delai</th> <th>zone_livraison</th> <th>desc_complete</th> </tr> </thead> <tbody> <tr><td>75 698</td><td>175,31</td><td>41</td><td>F</td><td>Produit: PRD-AUAVY\Q\QI - Livraison zone: F</td></tr> <tr><td>40 239</td><td>178,48</td><td>33</td><td>C</td><td>Produit: PRD-USRZT\Y\OD - Livraison zone: C</td></tr> <tr><td>20 213</td><td>194,59</td><td>20</td><td>C</td><td>Produit: PRD-YOIQG\PJ\V - Livraison zone: C</td></tr> <tr><td>50 255</td><td>260,21</td><td>33</td><td>A</td><td>Produit: PRD-AVRNN\CK\H - Livraison zone: A</td></tr> <tr><td>16 762</td><td>287,53</td><td>34</td><td>H</td><td>Produit: PRD-DYLYM\JZR - Livraison zone: H</td></tr> <tr><td>89 091</td><td>292,44</td><td>26</td><td>G</td><td>Produit: PRD-LLTGO\B\FC - Livraison zone: G</td></tr> <tr><td>12 621</td><td>334,35</td><td>12</td><td>G</td><td>Produit: PRD-UORTH\B\OF - Livraison zone: G</td></tr> <tr><td>14 394</td><td>339,19</td><td>18</td><td>C</td><td>Produit: PRD-RRDJM\FOT - Livraison zone: C</td></tr> <tr><td>9 401</td><td>345,3</td><td>51</td><td>Y</td><td>Produit: PRD-AZIVAR\ZJ - Livraison zone: Y</td></tr> <tr><td>10 653</td><td>349,26</td><td>11</td><td>B</td><td>Produit: PRD-WLKEA\X\J - Livraison zone: B</td></tr> <tr><td>78 943</td><td>363,97</td><td>31</td><td>M</td><td>Produit: PRD-MXGSE\IN\M - Livraison zone: M</td></tr> </tbody> </table>	code	prix_unit	qte	zone_livraison	date_cmde	date_livr	24489	535,57	5	M	11/11/07	04/12/07	12137	756,15	11	J	06/11/07	05/12/07	84636	245,58	6	A	07/11/07	11/12/07	67397	732,87	3	P	17/11/07	27/12/07	83942	584,77	2	Y	01/11/07	01/12/07	93948	1014,06	8	J	14/11/07	21/12/07	90253	1022,96	11	L	02/11/07	24/12/07	78216	988,43	6	E	13/11/07	27/12/07	81942	472,17	5	D	16/11/07	12/12/07	code	description	1	PRD-DSEIWA\WG	2	PRD-WOQGM\X\MV	3	PRD-ABUT\FOMB	4	PRD-JWPTZ\VHR	5	PRD-FFHIC\NFD	6	PRD-HNNQ\AVLD	7	PRD-NARQU\DDC	8	PRD-YUGR\WMBL	9	PRD-CUBLU\XJA	10	PRD-IHTOV\CUU	11	PRD-FEJOI\CON	12	PRD-BUXBE\RSF	code	montant	delai	zone_livraison	desc_complete	75 698	175,31	41	F	Produit: PRD-AUAVY\Q\QI - Livraison zone: F	40 239	178,48	33	C	Produit: PRD-USRZT\Y\OD - Livraison zone: C	20 213	194,59	20	C	Produit: PRD-YOIQG\PJ\V - Livraison zone: C	50 255	260,21	33	A	Produit: PRD-AVRNN\CK\H - Livraison zone: A	16 762	287,53	34	H	Produit: PRD-DYLYM\JZR - Livraison zone: H	89 091	292,44	26	G	Produit: PRD-LLTGO\B\FC - Livraison zone: G	12 621	334,35	12	G	Produit: PRD-UORTH\B\OF - Livraison zone: G	14 394	339,19	18	C	Produit: PRD-RRDJM\FOT - Livraison zone: C	9 401	345,3	51	Y	Produit: PRD-AZIVAR\ZJ - Livraison zone: Y	10 653	349,26	11	B	Produit: PRD-WLKEA\X\J - Livraison zone: B	78 943	363,97	31	M	Produit: PRD-MXGSE\IN\M - Livraison zone: M
code	prix_unit	qte	zone_livraison	date_cmde	date_livr																																																																																																																																														
24489	535,57	5	M	11/11/07	04/12/07																																																																																																																																														
12137	756,15	11	J	06/11/07	05/12/07																																																																																																																																														
84636	245,58	6	A	07/11/07	11/12/07																																																																																																																																														
67397	732,87	3	P	17/11/07	27/12/07																																																																																																																																														
83942	584,77	2	Y	01/11/07	01/12/07																																																																																																																																														
93948	1014,06	8	J	14/11/07	21/12/07																																																																																																																																														
90253	1022,96	11	L	02/11/07	24/12/07																																																																																																																																														
78216	988,43	6	E	13/11/07	27/12/07																																																																																																																																														
81942	472,17	5	D	16/11/07	12/12/07																																																																																																																																														
code	description																																																																																																																																																		
1	PRD-DSEIWA\WG																																																																																																																																																		
2	PRD-WOQGM\X\MV																																																																																																																																																		
3	PRD-ABUT\FOMB																																																																																																																																																		
4	PRD-JWPTZ\VHR																																																																																																																																																		
5	PRD-FFHIC\NFD																																																																																																																																																		
6	PRD-HNNQ\AVLD																																																																																																																																																		
7	PRD-NARQU\DDC																																																																																																																																																		
8	PRD-YUGR\WMBL																																																																																																																																																		
9	PRD-CUBLU\XJA																																																																																																																																																		
10	PRD-IHTOV\CUU																																																																																																																																																		
11	PRD-FEJOI\CON																																																																																																																																																		
12	PRD-BUXBE\RSF																																																																																																																																																		
code	montant	delai	zone_livraison	desc_complete																																																																																																																																															
75 698	175,31	41	F	Produit: PRD-AUAVY\Q\QI - Livraison zone: F																																																																																																																																															
40 239	178,48	33	C	Produit: PRD-USRZT\Y\OD - Livraison zone: C																																																																																																																																															
20 213	194,59	20	C	Produit: PRD-YOIQG\PJ\V - Livraison zone: C																																																																																																																																															
50 255	260,21	33	A	Produit: PRD-AVRNN\CK\H - Livraison zone: A																																																																																																																																															
16 762	287,53	34	H	Produit: PRD-DYLYM\JZR - Livraison zone: H																																																																																																																																															
89 091	292,44	26	G	Produit: PRD-LLTGO\B\FC - Livraison zone: G																																																																																																																																															
12 621	334,35	12	G	Produit: PRD-UORTH\B\OF - Livraison zone: G																																																																																																																																															
14 394	339,19	18	C	Produit: PRD-RRDJM\FOT - Livraison zone: C																																																																																																																																															
9 401	345,3	51	Y	Produit: PRD-AZIVAR\ZJ - Livraison zone: Y																																																																																																																																															
10 653	349,26	11	B	Produit: PRD-WLKEA\X\J - Livraison zone: B																																																																																																																																															
78 943	363,97	31	M	Produit: PRD-MXGSE\IN\M - Livraison zone: M																																																																																																																																															

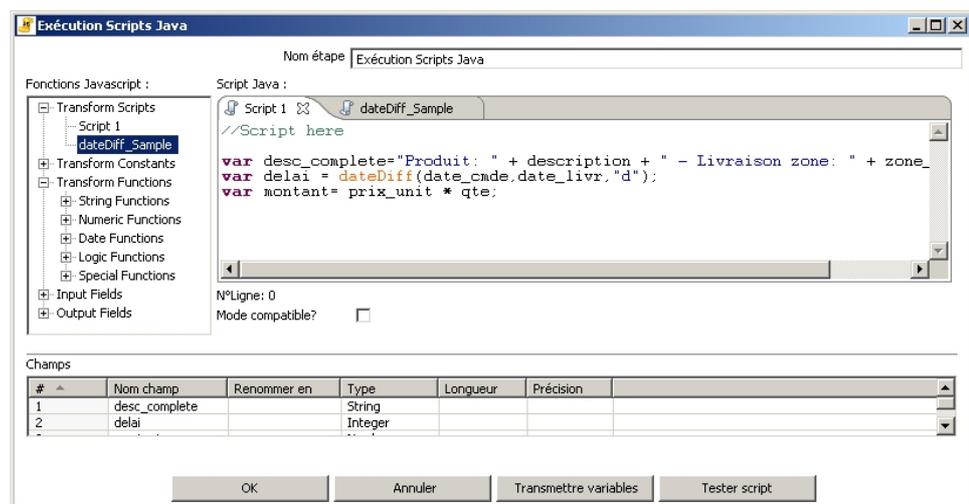
Modélisation dans Pentaho Data Integration (PDI)



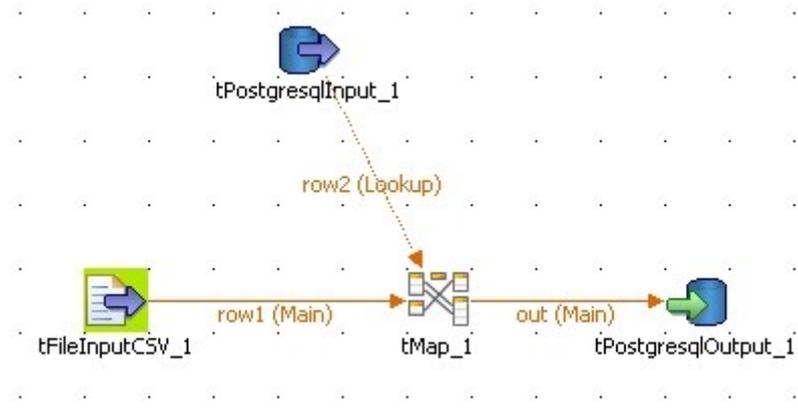
Le lookup sur la table [produits] permet de ramener la description du produit à partir du code produit présent dans le fichier d'entrée :



L'étape « JavaScript » permet la réalisation des calculs :



Modélisation dans Talend Open Studio



Le composant tMap, permet à la fois de réaliser le lookup sur la table [produits] et la réalisation des différents calculs :

The screenshot shows the configuration for the tMap component. It includes the following details:

- row1 (Main) Input Schema:**

Column
code
prtc_unit
qte
zone_livraison
date_cmde
date_livr
- row2 (Lookup) Input Schema:**

Expr. key	Column
row1.code	code
	description
- out (Main) Output Schema:**

Expression	Column
row1.prtc_unit * row1.qte	montant
row1.code	code
MesRoutines.getDateDifference(row1.date_cmde, row1.da...	delai
row1.zone_livraison	zone_livraison
"Produit: " + row2.description + " - Livraison zone: " + row...	desc_complete
- Schema editor (row2):**

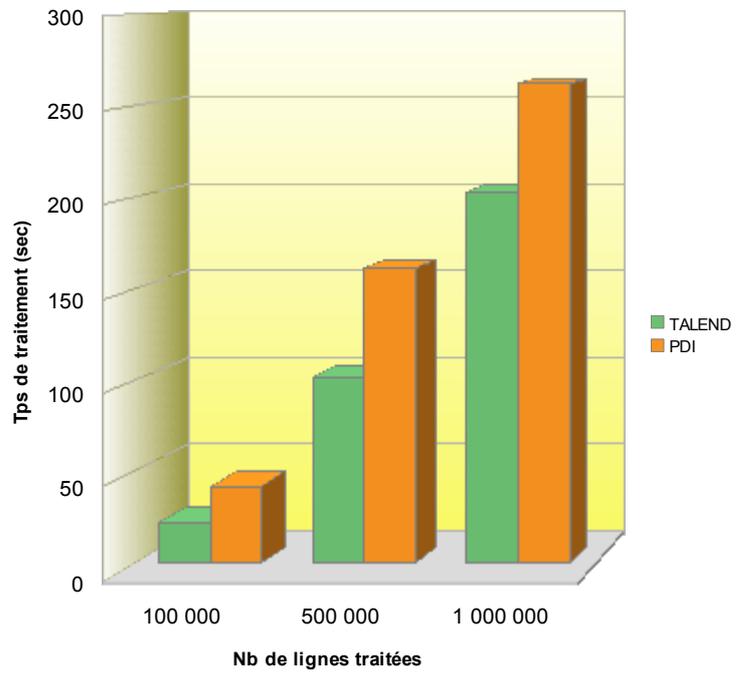
Column	Key	Type	Nullable	Date Pat...	Len...	Pre...	D...	Co...
code	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>		4			
description	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		12			
- Schema editor (out):**

Column	Key	Type	Nullable	Date Pat...	Len...	Pre...	D...	Co...
montant	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>		2			
code	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		6			
delai	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		10			
zone_livraison	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		1			
desc_complete	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		12			

Résultats du Test (Temps de traitement exprimés en sec.)

Nb de lignes traitées	Tps de traitement (secondes)	
	TALEND	PDI
100 000	22	42
500 000	102	161
1 000 000	203	263

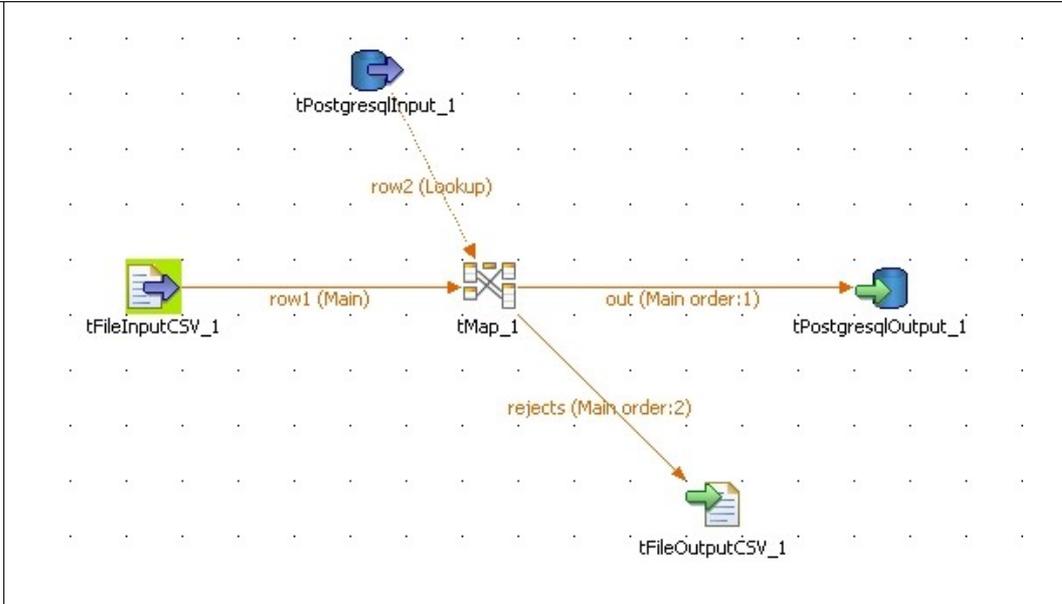
Graphe
Comparatif



4.4.6 TEST N° 5

Descriptif	<p>Test identique au Test 4, sauf qu'on isole au début du traitement toutes les données provenant venant des zones de livraisons codées W, X, Y et Z</p> <p>Ces lignes isolées sont exportées dans un fichier CSV pour traitement ultérieur.</p>
Détails	<p>Le fichier CSV des lignes rejetées contient ainsi le code produit et la zone de livraison rejetée associée (W, X, Y ou Z)</p>
Modélisation dans Pentaho Data Integration (PDI)	<div data-bbox="459 622 1444 981" data-label="Diagram"> </div> <p>Le filtrage des lignes s'effectue de la façon suivante :</p> <div data-bbox="475 1126 1380 1512" data-label="Form"> </div>

Modélisation dans Talend Open Studio



Comme dans le test 4, c'est le composant tMap qui permet d'isoler les lignes à rejeter :

row1

Column	Type	Nullable
code	Inte...	✓
prix_unit	Float	✓
qte	Inte...	✓
zone_livraison	String	✓
date_cmde	Date	✓
date_livr	Date	✓

row2

Expr. key	Column
row1.code	code
	description

out

Expression	Column
row1.prix_unit * row1.qte	montant
row1.code	code
MesRoutines.getDateDifference(row1.date_cmde, row1.date...	delai
row1.zone_livraison	zone_livraison
"Produit: " + row2.description + " - Livraison zone: " + row1....	desc_complete

rejects

Expression	Column
row1.zone_livraison	zone_livraison
row1.code	code

Schema editor

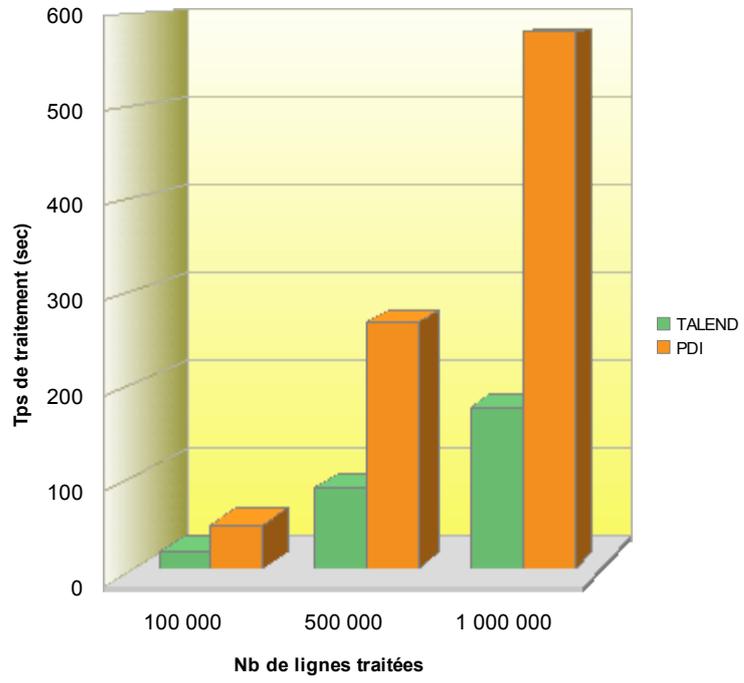
Column	Key	Type	Nullable	Date Patt...	Len...	Pre...	D...	Co...
code		Inte...	✓		6			
prix_unit		Float	✓		7	4		
qte		Inte...	✓		2			
zone_livraison		String	✓		1		==	
date_cmde		Date	✓	"dd/MM/y...	10			
date_livr		Date	✓	"dd/MM/y...	10			

Column	Key	Type	Nullable	Date Patt...	Len...	Pre...	D...	Co...
montant		Float	✓		2			
code		Inte...	✓		6			
delai		Inte...	✓		10			
zone_livraison		String	✓		1		==	
desc_complete		String	✓		12			

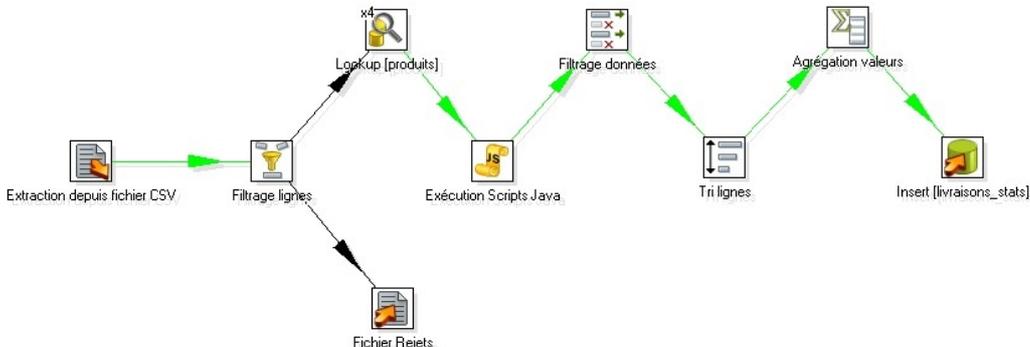
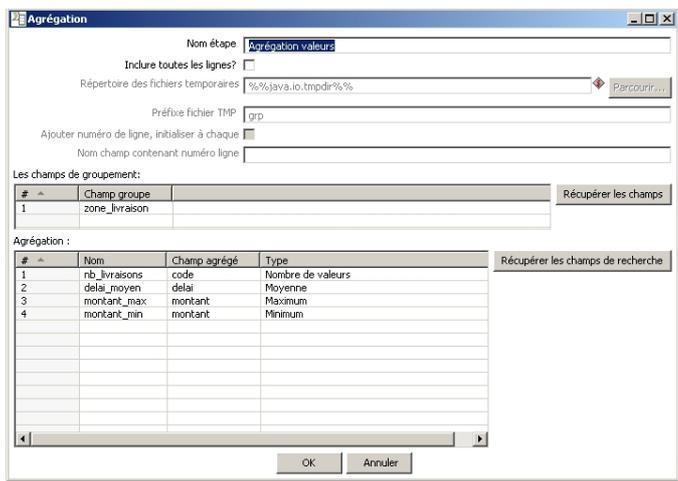
Résultats du Test
(Temps de traitement exprimés en sec.)

Nb de lignes traitées	Tps de traitement (secondes)	
	TALEND	PDI
100 000	19	48
500 000	87	267
1 000 000	174	580

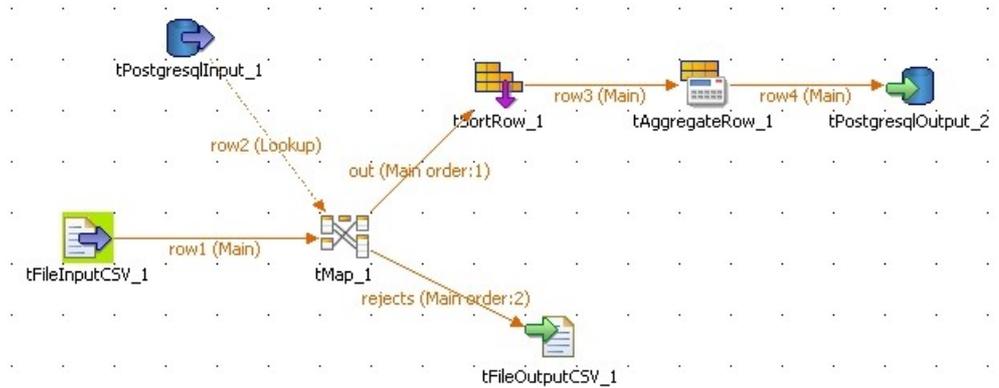
Graphe
Comparatif



4.4.7 TEST N° 6

Descriptif	Test identique au Test 5, avec alimentation d'une table d'agrégation [livraisons_stats]																																																																																																																			
Détails	<p>La table d'agrégation [livraisons_stats] doit contenir les données ci-dessous après les traitements suivants. Le groupement de données est effectué pour chaque zone de livraison.</p> <p>[livraisons_stats].[nb_livraisons]= {nb total de livraisons sur la zone} (count) [livraisons_stats].[montant_max]= {montant maxi d'une commande pour la zone} (max) [livraisons_stats].[montant_min]= {montant mini d'une commande pour la zone} (min) [livraisons_stats].[delai_moyen]= {delai moyen de livraison pour la zone} (avg)</p> <table border="1" data-bbox="427 656 1098 1126"> <thead> <tr> <th>zone_livraison</th> <th>nb_livraisons</th> <th>montant_max</th> <th>montant_min</th> <th>delai_moyen</th> </tr> </thead> <tbody> <tr><td>A</td><td>38 390</td><td>12 642,4</td><td>150,02</td><td>29</td></tr> <tr><td>B</td><td>38 354</td><td>12 647,4</td><td>152,45</td><td>29</td></tr> <tr><td>C</td><td>38 782</td><td>12 638,2</td><td>150,2</td><td>29</td></tr> <tr><td>D</td><td>38 422</td><td>12 632</td><td>150,47</td><td>29</td></tr> <tr><td>E</td><td>38 484</td><td>12 629,3</td><td>151,13</td><td>29</td></tr> <tr><td>F</td><td>38 648</td><td>12 637,3</td><td>151,65</td><td>29</td></tr> <tr><td>G</td><td>38 690</td><td>12 643,4</td><td>152,59</td><td>30</td></tr> <tr><td>H</td><td>38 300</td><td>12 621,3</td><td>150,7</td><td>29</td></tr> <tr><td>I</td><td>38 170</td><td>12 647,8</td><td>150,26</td><td>29</td></tr> <tr><td>J</td><td>38 458</td><td>12 648,6</td><td>150,07</td><td>29</td></tr> <tr><td>K</td><td>38 474</td><td>12 626,9</td><td>150,13</td><td>29</td></tr> <tr><td>L</td><td>38 734</td><td>12 644,1</td><td>153,39</td><td>30</td></tr> <tr><td>M</td><td>38 692</td><td>12 644,2</td><td>150,2</td><td>30</td></tr> <tr><td>N</td><td>38 938</td><td>12 619,4</td><td>152,48</td><td>29</td></tr> <tr><td>O</td><td>38 718</td><td>12 637,3</td><td>150,21</td><td>30</td></tr> <tr><td>P</td><td>38 352</td><td>12 624,5</td><td>150,79</td><td>30</td></tr> <tr><td>Q</td><td>38 292</td><td>12 626,2</td><td>151,51</td><td>30</td></tr> <tr><td>R</td><td>38 164</td><td>12 648,3</td><td>150,78</td><td>30</td></tr> <tr><td>S</td><td>38 520</td><td>12 647,3</td><td>150,16</td><td>30</td></tr> <tr><td>T</td><td>38 436</td><td>12 613,2</td><td>150,25</td><td>29</td></tr> <tr><td>U</td><td>38 144</td><td>12 640</td><td>151,82</td><td>29</td></tr> <tr><td>V</td><td>38 542</td><td>12 645,7</td><td>150,12</td><td>29</td></tr> </tbody> </table>	zone_livraison	nb_livraisons	montant_max	montant_min	delai_moyen	A	38 390	12 642,4	150,02	29	B	38 354	12 647,4	152,45	29	C	38 782	12 638,2	150,2	29	D	38 422	12 632	150,47	29	E	38 484	12 629,3	151,13	29	F	38 648	12 637,3	151,65	29	G	38 690	12 643,4	152,59	30	H	38 300	12 621,3	150,7	29	I	38 170	12 647,8	150,26	29	J	38 458	12 648,6	150,07	29	K	38 474	12 626,9	150,13	29	L	38 734	12 644,1	153,39	30	M	38 692	12 644,2	150,2	30	N	38 938	12 619,4	152,48	29	O	38 718	12 637,3	150,21	30	P	38 352	12 624,5	150,79	30	Q	38 292	12 626,2	151,51	30	R	38 164	12 648,3	150,78	30	S	38 520	12 647,3	150,16	30	T	38 436	12 613,2	150,25	29	U	38 144	12 640	151,82	29	V	38 542	12 645,7	150,12	29
zone_livraison	nb_livraisons	montant_max	montant_min	delai_moyen																																																																																																																
A	38 390	12 642,4	150,02	29																																																																																																																
B	38 354	12 647,4	152,45	29																																																																																																																
C	38 782	12 638,2	150,2	29																																																																																																																
D	38 422	12 632	150,47	29																																																																																																																
E	38 484	12 629,3	151,13	29																																																																																																																
F	38 648	12 637,3	151,65	29																																																																																																																
G	38 690	12 643,4	152,59	30																																																																																																																
H	38 300	12 621,3	150,7	29																																																																																																																
I	38 170	12 647,8	150,26	29																																																																																																																
J	38 458	12 648,6	150,07	29																																																																																																																
K	38 474	12 626,9	150,13	29																																																																																																																
L	38 734	12 644,1	153,39	30																																																																																																																
M	38 692	12 644,2	150,2	30																																																																																																																
N	38 938	12 619,4	152,48	29																																																																																																																
O	38 718	12 637,3	150,21	30																																																																																																																
P	38 352	12 624,5	150,79	30																																																																																																																
Q	38 292	12 626,2	151,51	30																																																																																																																
R	38 164	12 648,3	150,78	30																																																																																																																
S	38 520	12 647,3	150,16	30																																																																																																																
T	38 436	12 613,2	150,25	29																																																																																																																
U	38 144	12 640	151,82	29																																																																																																																
V	38 542	12 645,7	150,12	29																																																																																																																
Modélisation dans Pentaho Data Integration (PDI)	 <p>Pour que les calculs d'agrégation soit corrects, le flux de données doit être trié par zone de livraison. Détail de l'étape « Agrégation de valeurs » :</p> <div data-bbox="622 1624 1300 2105">  </div>																																																																																																																			

Modélisation
dans
Talend Open Studio



Pour que les calculs d'agrégation soit corrects, le flux de données doit être trié par zone de livraison. Détail de l'étape « tAggregateRow » :

Schema: Built-In Edit schema Sync columns

Group by

Output column	Input column position
zone_livraison	zone_livraison

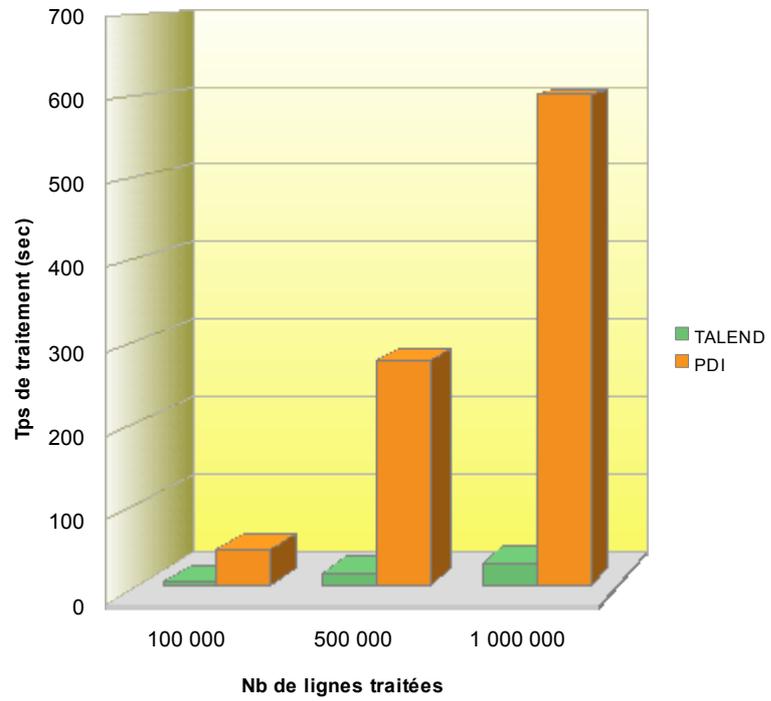
Operations

Output column	Function	Input column position
montant_min	min	montant
montant_max	max	montant
delai_moyen	avg	delai
nb_livraisons	count	code

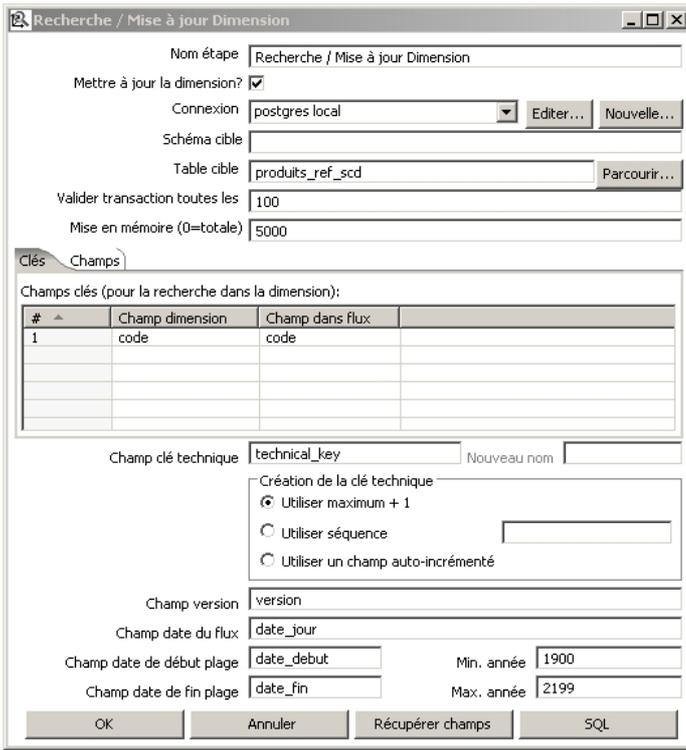
Résultats du Test
(Temps de traitement exprimés en sec.)

Nb de lignes traitées	Tps de traitement (secondes)	
	TALEND	PDI
100 000	4	44
500 000	15	276
1 000 000	28	602

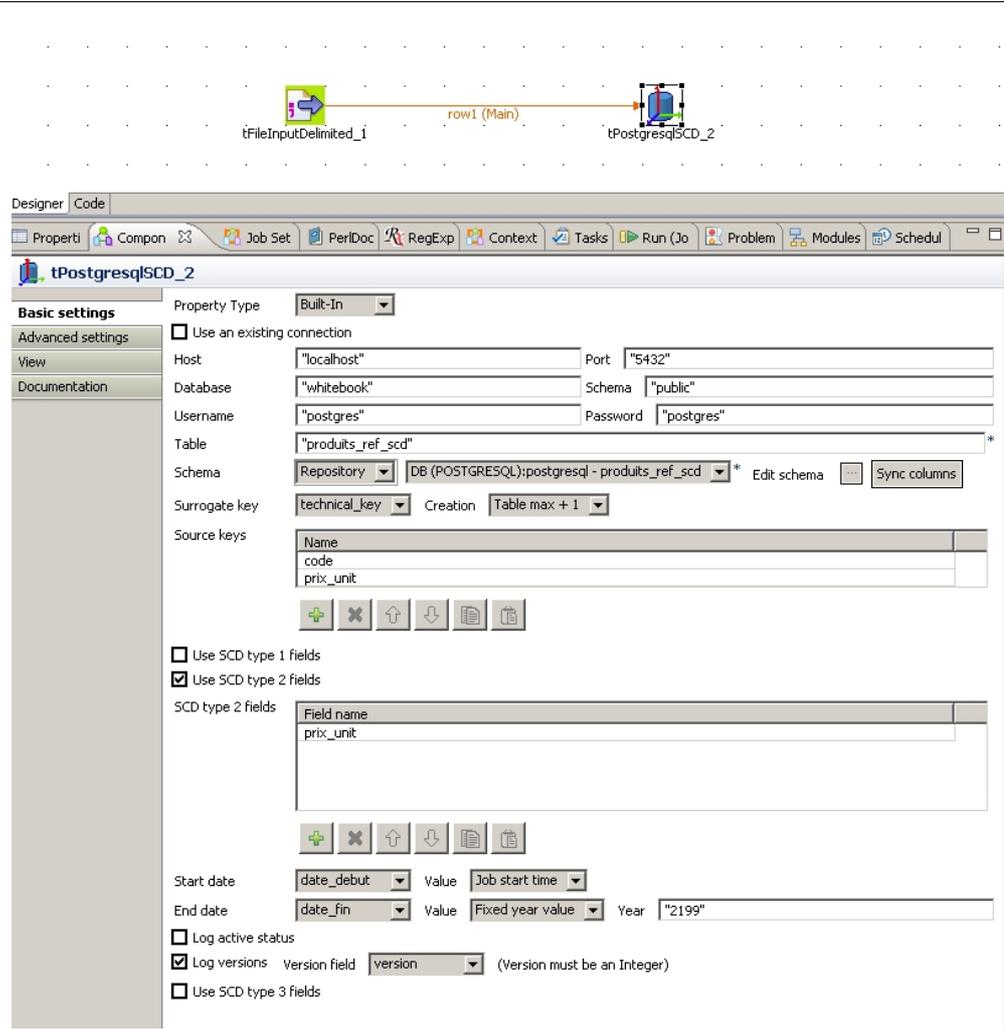
Graphe
Comparatif



4.4.8 TEST N°7

Descriptif	Utilisation d'une étape de traitement « Dimension lente » de type 2 Pour plus d'infos sur les différents types de « Slow Changing Dimension » consulter la page : http://en.wikipedia.org/wiki/Slowly_changing_dimension																																																															
Détails	<p>La table [produits_ref_scd] permet l'historisation des prix d'un catalogue produit.</p> <p>Extrait de la table avant le traitement :</p> <table border="1" data-bbox="405 555 1469 678"> <thead> <tr> <th>code</th> <th>description</th> <th>prix_unit</th> <th>date_debut</th> <th>date_fin</th> <th>version</th> <th>technical_key</th> </tr> </thead> <tbody> <tr> <td>100</td> <td>PRD-JKKZSZDW</td> <td>906,76</td> <td>1899-12-31</td> <td>2199-12-31</td> <td>1</td> <td>100</td> </tr> <tr> <td>101</td> <td>PRD-YCCLNHTN</td> <td>257,84</td> <td>1899-12-31</td> <td>2199-12-31</td> <td>1</td> <td>101</td> </tr> <tr> <td>102</td> <td>PRD-YGOIRRGU</td> <td>586,05</td> <td>1899-12-31</td> <td>2199-12-31</td> <td>1</td> <td>102</td> </tr> </tbody> </table> <p>Le fichier traité indique le code et le prix de l'article à la date du traitement :</p> <p>100; 746.48 101; 528.72</p> <p>Une fois le traitement effectué, la table [produits_ref_scd] stocke l'historique des prix :</p> <table border="1" data-bbox="405 880 1469 1037"> <thead> <tr> <th>code</th> <th>description</th> <th>prix_unit</th> <th>date_debut</th> <th>date_fin</th> <th>version</th> <th>technical_key</th> </tr> </thead> <tbody> <tr> <td>100</td> <td>PRD-JKKZSZDW</td> <td>906,76</td> <td>1899-12-31</td> <td>2008-02-26</td> <td>1</td> <td>100</td> </tr> <tr> <td>100</td> <td></td> <td>746,48</td> <td>2008-02-26</td> <td>2199-12-31</td> <td>2</td> <td>500 098</td> </tr> <tr> <td>101</td> <td>PRD-YCCLNHTN</td> <td>257,84</td> <td>1899-12-31</td> <td>2008-02-26</td> <td>1</td> <td>101</td> </tr> <tr> <td>101</td> <td></td> <td>528,72</td> <td>2008-02-26</td> <td>2199-12-31</td> <td>2</td> <td>500 099</td> </tr> </tbody> </table> <p>Pour un code produit donné, dès que le prix est différent de celui présent en base, un nouvel enregistrement est créée avec un nouveau numéro de version. La période de validité est également mise à jour (date_debut, date_fin)</p>	code	description	prix_unit	date_debut	date_fin	version	technical_key	100	PRD-JKKZSZDW	906,76	1899-12-31	2199-12-31	1	100	101	PRD-YCCLNHTN	257,84	1899-12-31	2199-12-31	1	101	102	PRD-YGOIRRGU	586,05	1899-12-31	2199-12-31	1	102	code	description	prix_unit	date_debut	date_fin	version	technical_key	100	PRD-JKKZSZDW	906,76	1899-12-31	2008-02-26	1	100	100		746,48	2008-02-26	2199-12-31	2	500 098	101	PRD-YCCLNHTN	257,84	1899-12-31	2008-02-26	1	101	101		528,72	2008-02-26	2199-12-31	2	500 099
code	description	prix_unit	date_debut	date_fin	version	technical_key																																																										
100	PRD-JKKZSZDW	906,76	1899-12-31	2199-12-31	1	100																																																										
101	PRD-YCCLNHTN	257,84	1899-12-31	2199-12-31	1	101																																																										
102	PRD-YGOIRRGU	586,05	1899-12-31	2199-12-31	1	102																																																										
code	description	prix_unit	date_debut	date_fin	version	technical_key																																																										
100	PRD-JKKZSZDW	906,76	1899-12-31	2008-02-26	1	100																																																										
100		746,48	2008-02-26	2199-12-31	2	500 098																																																										
101	PRD-YCCLNHTN	257,84	1899-12-31	2008-02-26	1	101																																																										
101		528,72	2008-02-26	2199-12-31	2	500 099																																																										
Modélisation dans Pentaho Data Integration (PDI)	 <p>Extraction depuis fichier → Recherche / Mise à jour Dimension</p>  <p>Recherche / Mise à jour Dimension</p> <p>Nom étape: Recherche / Mise à jour Dimension</p> <p>Mettre à jour la dimension? <input checked="" type="checkbox"/></p> <p>Connexion: postgres local [Modifier... Nouvelle...]</p> <p>Schéma cible: []</p> <p>Table cible: produits_ref_scd [Parcourir...]</p> <p>Valider transaction toutes les: 100</p> <p>Mise en mémoire (0=totale): 5000</p> <p>Clés / Champs</p> <p>Champs clés (pour la recherche dans la dimension):</p> <table border="1" data-bbox="630 1646 1284 1780"> <thead> <tr> <th>#</th> <th>Champ dimension</th> <th>Champ dans flux</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>code</td> <td>code</td> </tr> </tbody> </table> <p>Champ clé technique: technical_key [Nouveau nom:]</p> <p>Création de la clé technique:</p> <p><input checked="" type="radio"/> Utiliser maximum + 1</p> <p><input type="radio"/> Utiliser séquence []</p> <p><input type="radio"/> Utiliser un champ auto-incrémenté</p> <p>Champ version: version</p> <p>Champ date du flux: date_jour</p> <p>Champ date de début plage: date_debut [Min. année: 1900]</p> <p>Champ date de fin plage: date_fin [Max. année: 2199]</p> <p>[OK] [Annuler] [Récupérer champs] [SQL]</p>	#	Champ dimension	Champ dans flux	1	code	code																																																									
#	Champ dimension	Champ dans flux																																																														
1	code	code																																																														

Modélisation dans Talend Open Studio



The screenshot shows the Talend Open Studio interface. At the top, a data flow is visible: **tFileInputDelimited_1** connects to **row1 (Main)**, which in turn connects to **tPostgresqlSCD_2**. Below this, the configuration for **tPostgresqlSCD_2** is shown in the 'Designer' view.

Basic settings: Property Type is set to 'Built-In'.

Advanced settings: 'Use an existing connection' is unchecked. The connection details are: Host: "localhost", Port: "5432", Database: "whitebook", Schema: "public", Username: "postgres", Password: "postgres". The Table is "produits_ref_scd".

Source keys: A table with columns 'Name' and 'code' containing 'prix_unit'.

SCD type 2 fields: A table with columns 'Field name' and 'code' containing 'prix_unit'.

Start date: date_debut, Value: Job start time.

End date: date_fin, Value: Fixed year value, Year: "2199".

Log active status: Unchecked.

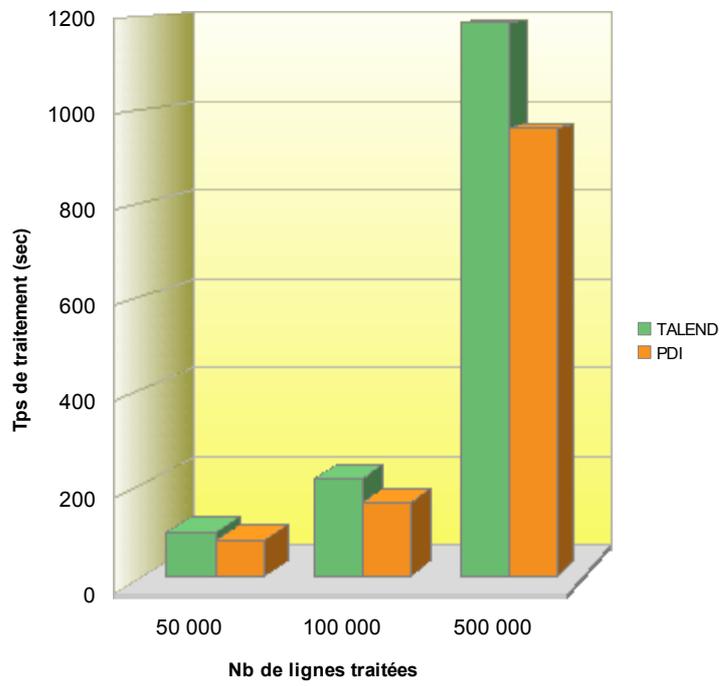
Log versions: Checked. Version field: version. (Version must be an Integer)

Use SCD type 3 fields: Unchecked.

Résultats du Test (Temps de traitement exprimés en sec.)

Nb de lignes traitées	Tps de traitement (secondes)	
	TALEND	PDI
50 000	94	79
100 000	209	158
500 000	1189	959

Graphe
Comparatif



5 CONCLUSION

A l'issue de l'ensemble de notre étude comparative, voici ce que nous pouvons conclure :

Talend Open Studio et Pentaho Data Integration sont des ETL ayant leurs propres spécificités, et sont donc de ce fait plus complémentaires que « concurrents »

Ce sont des **alternatives réelles** à des ETL propriétaires comme Informatica Power Center, Oracle Warehouse Builder, Cognos DecisionStream...

Pourquoi ?

- Ces deux outils proposent l'ensemble des fonctionnalités nécessaires à un ETL.

Même si la modélisation entre Pentaho Data Integration et Talend Open Studio est différente, la conception des traitements d'intégration de données est toujours simple et pragmatique.

- Les temps d'exécution varient suivant la typologie des traitements.

PDI semble moins performant que TOS dans les calculs d'agrégations et les lookups, et meilleur en chargement de données pur ou traitement de dimensions lentes. A noter qu'il est possible avec PDI de faire de la répartition de charge sur plusieurs serveurs lors du traitement de gros volumes de données.

Le choix entre les ETL Talend Open Studio et Pentaho Data Integration dépend donc essentiellement de la typologie du projet à mener
 Une analyse préalable est donc nécessaire pour choisir l'ETL qui semble le plus adapté.
C'est dans ce cadre que Atol Conseils et Développements peut vous accompagner.

Les PLUS de Pentaho Data Integration	Les PLUS de Talend Open Studio
Une interface simple et intuitive qui permet de réaliser très rapidement vos premiers traitements	Un référentiel de métadonnées qui permet la réutilisation des formats de fichiers, des connexions aux bases, ...
Un référentiel de stockage des jobs qui peut être partagé par plusieurs développeurs	La notion de « contexte » qui permet de switcher facilement d'un environnement de test à un environnement de production
Une parfaite intégration à la plate-forme PENTAHO, qui permet éventuellement la récupération de données décisionnelles en temps réel	Un outil de planification (scheduler) intégré
Un forum très riche et dynamique et un wiki bien documenté	Un forum également déjà très riche

6 ANNEXE 1

➔ Liste d' ETL propriétaires et Open Source :

Société	ETL	Commentaire
Abinitio	Abinitio	
Altova	Mapforce	
Attunity	Integration Suite	
Barracuda software	Barracuda Integrator	
Benetl	Benetl	Freeware
Business Objects	Data Integrator	
Cast Iron Systems	Cast Iron Integration Appliance	
CA	Data integrator	
OpenSys	Clover ETL	Open Source
Cognos	Decisionstream	
Corporator	Transformer	
CoSORT	CoSORT ETL tools	
Crossflo Systems	DataExchange	
Datacyr Corporation	DataCyr Transform	
Data Exchanger	Data Exchanger	
DataHabitat	DH ETL	
Datamirror	Datamirror	
D2K	D2K	
Embarcadero	DTStudio	
Enhydra	Octopus	Open Source
ETL Solutions ltd	Transformation Manager	
ETI	ETI Solution	
GammaSoft	data'distribution	
Group 1	Data Flow	
Hummingbird	Genio	
Information Builders	iWay	
IBM	Websphere DataStage	
Ikan	MetaSuite	
Informatica	PowerCenter	Market leader
iSoft	Amadea	
Microsoft	DTS	Ceased to exist
Microsoft	Integration Services	Bundled with SQL server
Oracle	Warehouse Builder	
Oracle	Sunopsis	
Pentaho	Pentaho Data integration	Open Source
Pervasive	Data Junction	
Pervasive	Data Integrator & Business Integrator	
Platinum	Info Pump	Acquired by CA.
Powershift	Manheim	
Sagent Technology	Sagent Solution	Acquired by Group 1
SAS	Data integration	Previously ETL/Studio
Solonde	Warehouse Workbench	Acquired by Sybase
Sterling Commerce	Gentran Integration Suite	
Sunopsis	Data Conductor	Acquired by Oracle
Sybase	Data Integration Suite & Data Federation	
Syncsort	DExpress	
Talend	Talend Open Studio	Open Source
WisdomForce™	FastReader	

7 ANNEXE 2

➔ Retour d'expérience d'ATOL CD sur l'utilisation d'un ETL Open Source dans un projet.

7.1 UN BESOIN ET UN PROJET CLIENT

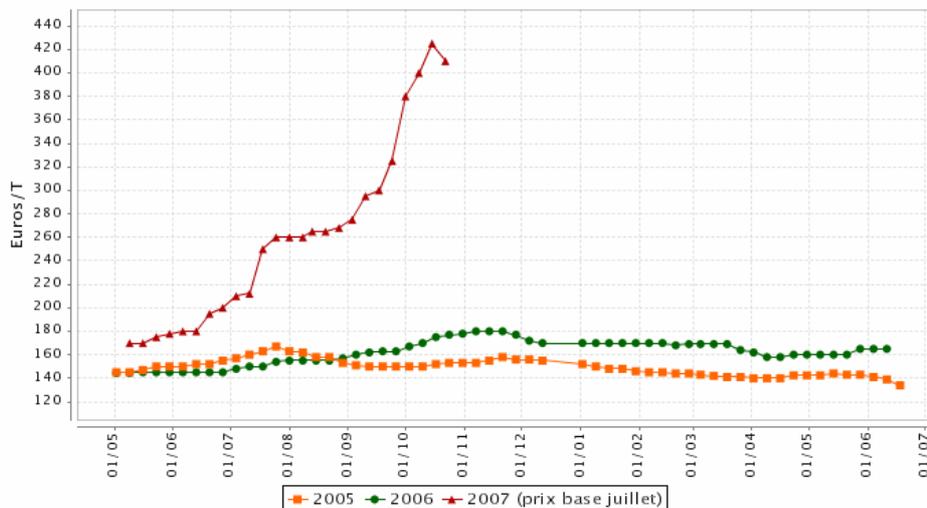
La société « Adhérents.coops » propose un extranet agricole à plus d'une trentaine de coopératives, réparties sur l'ensemble de la France.

Cet extranet est un portail à destination des agriculteurs qui leur permet d'accéder à un ensemble d'informations: alertes techniques, informations agronomiques, suivi de leurs compte adhérent, petites annonces, etc...



En 2007, en lien direct avec l'explosion des cours des céréales, il a été décidé de mettre en place un service en ligne de « Bourse aux céréales » afin de permettre aux céréaliers d'avoir un minimum de garanties concernant les prix de vente et d'achat des céréales.

Blé dur



C'est dans ce contexte que l'ETL **Pentaho Data Integration** a été mis en place pour récupérer les cotations EuroNext MATIF du blé, du colza et du maïs, et ainsi permettre une mise à jour quotidienne des services de bourse disponible sur un trentaine d'extranets.

7.2 MISE EN PLACE TECHNIQUE

- **Step 1 :**

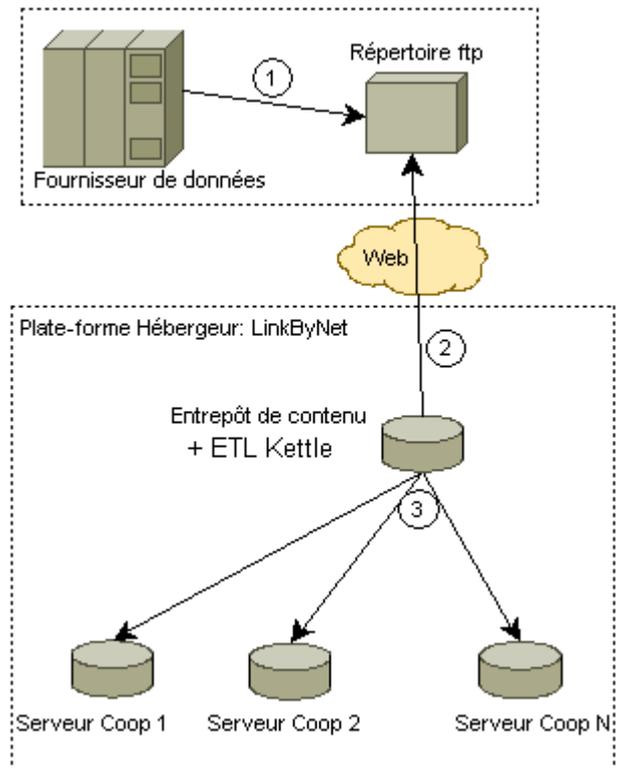
Un prestataire externe délivre chaque jour (du lundi au vendredi) les cotations Euronext MATIF au format XML, sur un serveur FTP. 3 fichiers sont fournis: 1 pour le blé, 1 pour le maïs, 1 pour le colza.

- **Step 2 :**

Ces fichiers XML sont ensuite « parsés » pour l'alimentation d'une base PostgreSQL « tampon », qui est joué le rôle d'entrepôt de contenu centralisé.

- **Step 3 :**

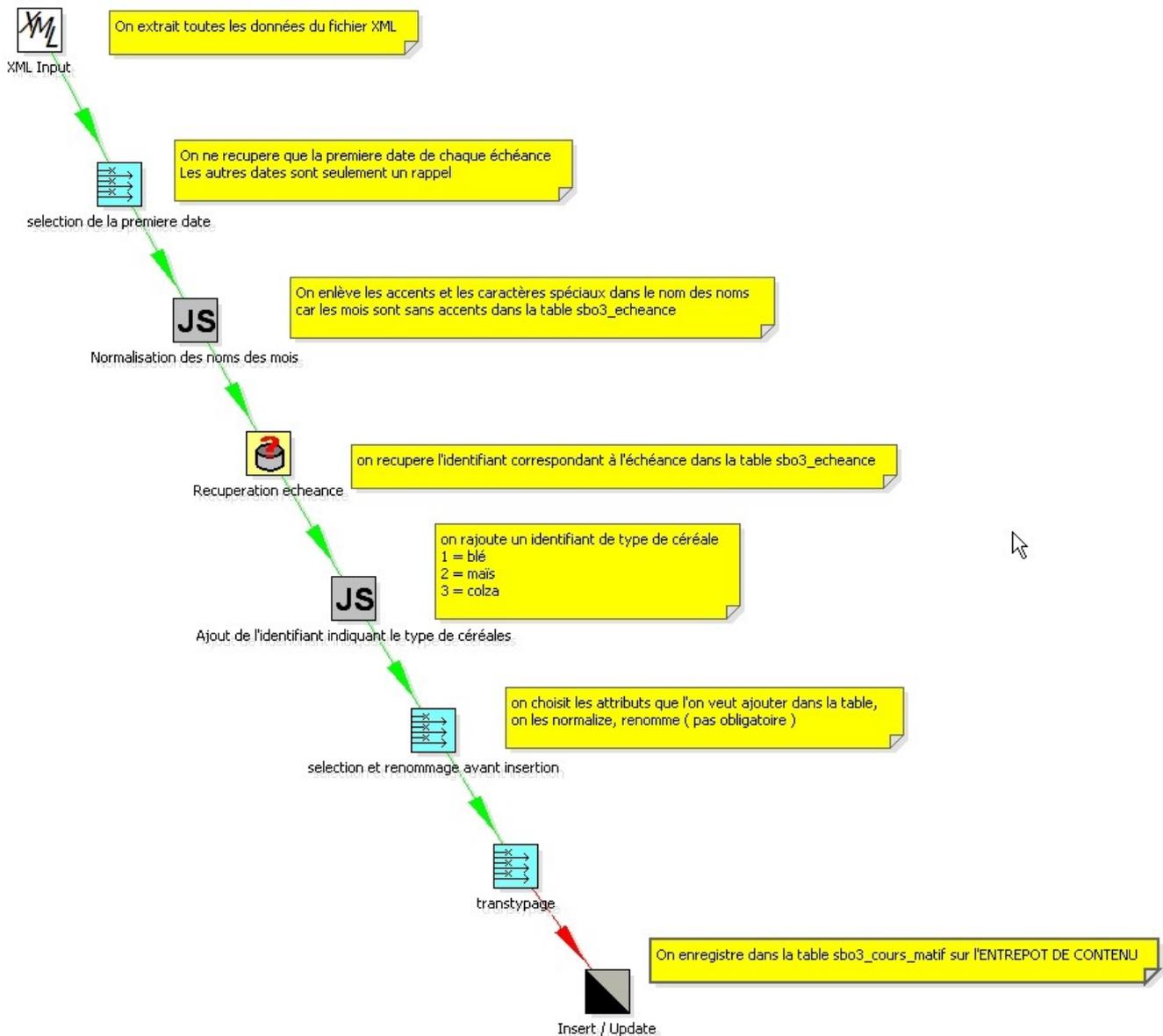
Enfin, les bases de données postgresQL sont alimentées en parallèle pour assurer une mise à jour simultanée des extranets des coopératives.



La syntaxe des fichiers XML à traiter est la suivante :

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <marche nom="Euronext Blé" nomEng="Euronext Wheat">
- <echeance nom="Novembre 2007" id="4454" nomEng="November 2007" nomAbr="Nov 2007" dateClose="12/11/2007">
- <date value="24/10/2007">
  <settle>233.5</settle>
  <openTraite>239.75</openTraite>
  <openInterest>12599</openInterest>
  <high>239.75</high>
  <low>233.5</low>
  <ttvolume>1259</ttvolume>
  <variation>-8.5</variation>
</date>
+ <date value="23/10/2007">
+ <date value="22/10/2007">
+ <date value="19/10/2007">
+ <date value="18/10/2007">
+ <date value="17/10/2007">
+ <date value="16/10/2007">
+ <date value="15/10/2007">
+ <date value="12/10/2007">
+ <date value="11/10/2007">
</echeance>
+ <echeance nom="Janvier 2008" id="4976" nomEng="January 2008" nomAbr="Jan 2008" dateClose="10/01/2008">
+ <echeance nom="Mars 2008" id="4977" nomEng="March 2008" nomAbr="Mars 2008" dateClose="10/03/2008">
+ <echeance nom="Mai 2008" id="4978" nomEng="May 2008" nomAbr="Mai 2008" dateClose="12/05/2008">
+ <echeance nom="Août 2008" id="5233" nomEng="August 2008" nomAbr="Août 2008" dateClose="11/08/2008">
+ <echeance nom="Novembre 2008" id="4980" nomEng="November 2008" nomAbr="Nov 2008" dateClose="10/11/2008">
+ <echeance nom="Janvier 2009" id="4981" nomEng="January 2009" nomAbr="Jan 2009" dateClose="12/01/2009">
+ <echeance nom="Mars 2009" id="4982" nomEng="March 2009" nomAbr="Mars 2009" dateClose="10/03/2009">
</marche>
```

- Une première transformation « Kettle » permet la réalisation du step 2.
Elle utilise entre autre l'étape de « parsing XML » de Kettle v2.4

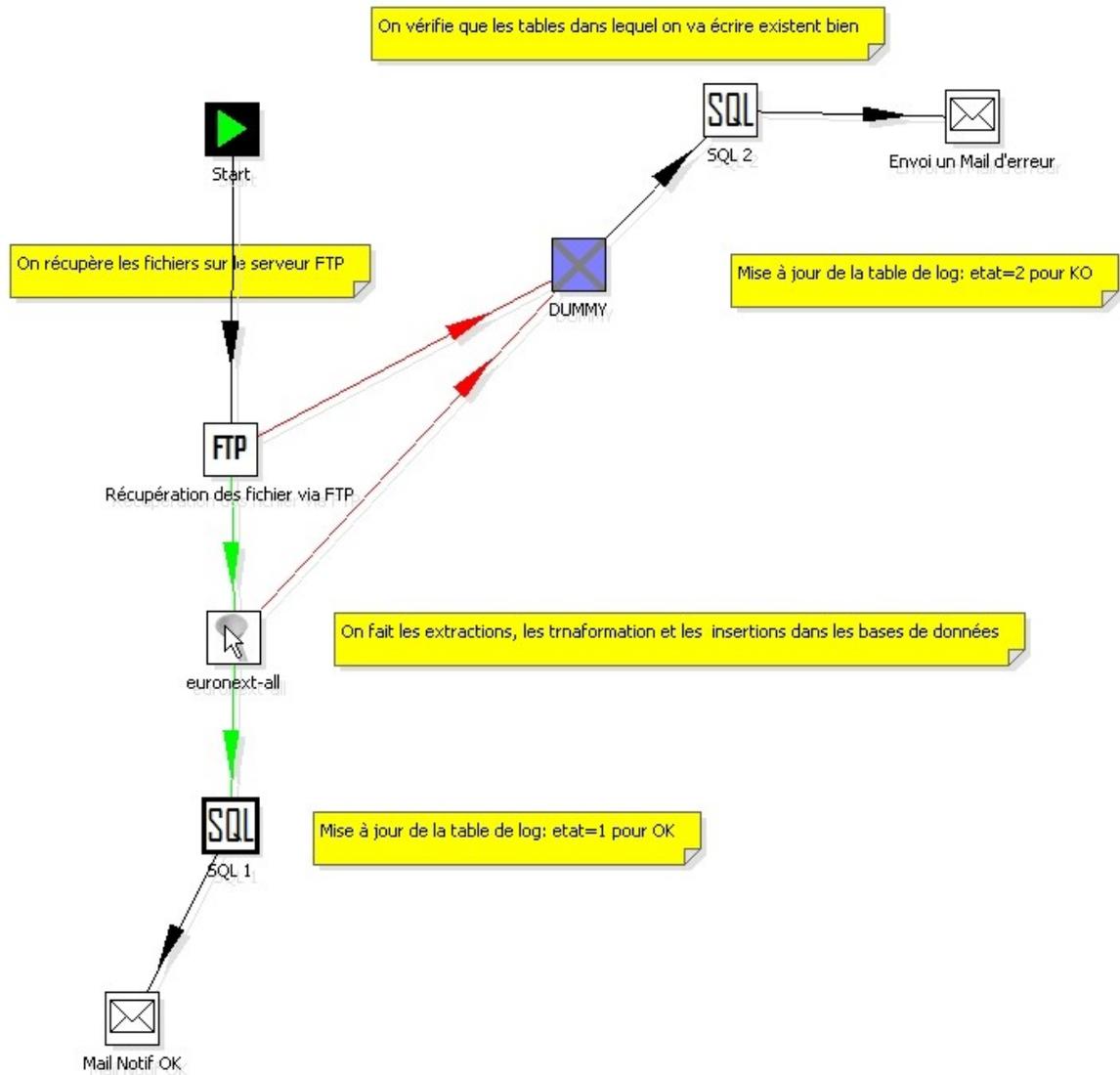


- Un job « Kettle » permet l'exécution de la transformation et la gestion des erreurs.

Le job permet :

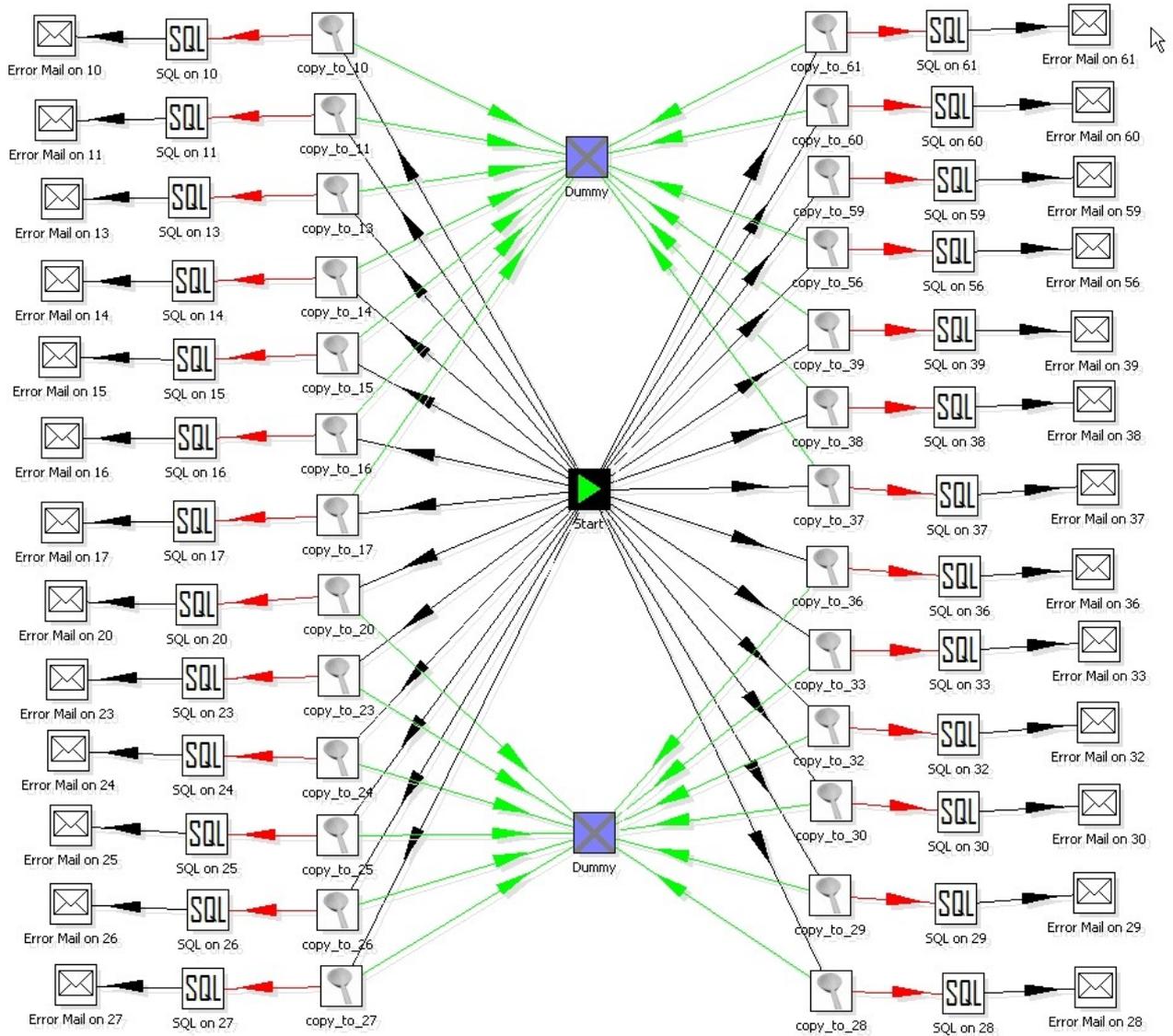
- de récupérer par FTP les fichiers XML
- de lancer le traitement et l'intégration des données dans l'entrepôt central
- de notifier par mail le bon déroulement (ou non) des traitements
- d'exécuter des ordres SQL pour indiquer à l'application l'heure des traitements et le bon déroulement de ceux-ci.

- **Détail du job du step 2 :**



- **Détail du job du step 3 :**

Mise à jour en parallèle des 30 serveurs extranets :



7.3 RÉSULTAT SUR UN EXTRANET

Le résultat du traitement d'intégration de données permet de visualiser quotidiennement à partir de 19h45 les cours de clôture du MATIF sur chacun des extranets.

Les données sont présentées de façon tabulaire et graphiques :

RECHERCHER | AIDE | PLAN DU SITE | CONTACT



Ma coopérative
Céréales
Techniques
Agrofournitures
Mes activités

MON PROFIL
ME DECONNECTER
APPROUVER
PREVISUALISER

Infos marchés

- Commentaires
- Analyses
- ADMIN Analyses
- Les parités
- Les options Matif
- ADMIN Les options Matif

Les contrats

- Notre offre
- ADMIN Notre offre
- Conditions générales
- ADMIN Conditions générales

Commercialisation

- Cours MATIF
- Cotations MATIF
- ADMIN Cotations MATIF
- Cotations Prix du Jour
- ADMIN Cotations Prix du Jour
- Bourse aux Céréales
- Bourse aux Céréales OS
- Bourse aux céréales Finance
- ADMIN Bourse aux céréales
- Finance
- Les formulaires
- ADMIN Les formulaires

Liste des cours Matif du 23/10/2007

Dernière mise à jour : Le 23/10/2007 à 19:45:10

Production :

Cotations Matif

	Cours de compensation	Variation	+ bas	+ haut	Nb de lots (PO)	Lots traités	Prix CERENA	Prendre position
Novembre 2007	371.00	0.25	371.00	373.00	4875	725		
Aout 2008	373.25	1.25	373.00	374.50	6608	104		

Historique des cours

	22/10/2007	19/10/2007	17/10/2007	16/10/2007	15/10/2007	12/10/2007
Novembre 2007	370.75	372.00	369.75	368.00	367.00	362.50
Aout 2008	372.00	374.75	370.00	368.50	368.00	363.50

Nb de jours :

Sélection des échéances :

Novembre 2007 Aout 2008

Historique des cours

Date	Novembre 2007	Aout 2008
16/09	350	344
17/09	350	344
18/09	348	343
19/09	346	340
20/09	350	344
21/09	354	348
22/09	354	348
23/09	354	348
24/09	358	352
25/09	358	352
26/09	358	352
27/09	362	356
28/09	362	356
29/09	362	356
30/09	362	356
01/10	364	356
02/10	360	354
03/10	356	352
04/10	356	352
05/10	358	354
06/10	358	354
07/10	354	350
08/10	354	350
09/10	358	354
10/10	362	358
11/10	362	358
12/10	366	362
13/10	366	362
14/10	368	364
15/10	368	364
16/10	368	364
17/10	370	366
18/10	370	366
19/10	372	368
20/10	372	368
21/10	370	366
22/10	370	366
23/10	370	366

Copyright © 2007-2008, CERENA. Tous droits réservés.

[Mentions légales](#)

Page 46 / 49

7.4 FEEDBACK SUR L'UTILISATION DE « PENTAHO DATA INTEGRATION »

Dans ce contexte métier très précis, voici les éléments que l'on peut retirer de l'utilisation d'un **ETL Open Source** :

- **Phase de Conception/Développement :**

L'utilisation de Pentaho Data Integration est très simple et intuitive, il n'y a pratiquement aucun code à écrire pour la conception et la définition des traitements d'intégration.

Il y a donc des gains conséquents en terme de développement, malheureusement on ne peut pas s'affranchir de l'analyse (!)

La documentation, les « samples » fournis avec l'outil et le site de Pentaho Data Integration (kettle.pentaho.org) permettent bien souvent de trouver la solution à un point de blocage.

Les fonctions de prévisualisation de l'ETL permettent de tester efficacement les traitements lors de la phase de développement.

- **Phase d'installation :**

Il suffit juste d'installer Pentaho Data Integration sur le serveur cible (décompression d'un zip, installation de java 1.5 si besoin), puis ensuite de planifier les traitements batch (via crontab sous Linux).

Le déploiement des jobs peut se faire par recopie du « repository » de DEV sur le serveur de PROD (remontée du dump de la bdd du repository kettle)

- **Phase de production :**

Chaque jour, les mails de notification permettent d'être informés du bon déroulement des traitements.

A ce jour, aucun « plantage » n'a été constaté depuis la mise en production effectuée le 23 Mai 2007.

- **Maintenance :**

La maintenance évolutive ou correctrice est grandement simplifiée du fait de la vision graphique des traitements et du découpage de ceux-ci en briques élémentaires.

Si jamais un jour il faut aller chercher les données via webservice, il suffira de remplacer l'étape « parsing XML » par un « webservice call »

- **Coûts de licences :**

Aucun

7.5 RETOUR CLIENT

- **Adhérents.coop** communique chaque année via la rapport annuel de la [FFCAT](#) (Fédération Française des Coopératives Agricoles de Collecte d'Approvisionnement et de Transformation).

Les 2 pages qui suivent illustrent la satisfaction des coopératives quant à la mise en place des services de bourse en ligne, qui sont alimentés en Back-End par un ETL Open Source !

[Sylvain Decloix](#) - Février 2008

Union Adhérents.Coop : Toujours plus !



Qui a dit que les coopératives ne savaient pas prendre le virage des nouvelles technologies ? Certainement pas celles qui, regroupées au sein d'Adhérents.Coop, œuvrent, depuis plus de cinq ans pour certaines, au développement des outils extranets spécifiques aux métiers de la coopérative.

En quelques années, Adhérents.Coop est devenu la plateforme de référence des portails extranet des coopératives. Aujourd'hui, l'audience cumulée des sites, s'appuyant sur le kit d'Adhérents.Coop, dépassent les 100 000 visites mensuelles. Cette audience double chaque année depuis 3 ans, signe que les agriculteurs utilisent de plus en plus ces nouveaux outils qui allient performances, efficacité et disponibilité.

Pour encore mieux appréhender les attentes de ses membres et des coopératives sur les nouvelles technologies, Adhérents.Coop a commandé une enquête exclusive. Cette dernière, menée auprès de 70 coopératives de différents secteurs, est riche d'enseignements et montre l'intérêt grandissant que portent les coopératives vis-à-vis des outils internet.

Mettre en place un extranet : de plus en plus fréquent et bientôt indispensable

Avec 62 % de coopératives équipées, la présence des sites internet (ouverts sans restriction à tout public) est fréquente et constitue souvent la première étape dans l'appropriation des outils web. Ces sites destinés à présenter l'entreprise contiennent rarement des informations métiers destinées aux agriculteurs.

Us nous font confiance :

- 01 – Cérégrain
- 02 – Cerena
- 10 – Nouricia
- 10 – Scara
- 14 – Agrial
- 21 – Dijon Céréales
- 28 – Agralys
- 28 – Interface
- 28 – Cabep
- 28 – Scael
- 40 – Maisadour
- 45 – Caproga
- 51 – Champagne Céréales
- 51 – Cohesis
- 55 – Emc2
- 57 – Lorca
- 60 – Océal
- 62 – Unéal
- 63 – Limagrain
- 67 – Comptoir Agricole
- 68 – Cac Colmar
- 75 – Arvalis – Institut du végétal
- 75 – COOP DE FRANCE
métiers du grain
- 75 – Invivo
- 75 – Sofiprotéol
- 76 – Capseine
- 77 – Terres Bocage Gâtinais
- 80 – Noriap
- 89 – 110 Bourgogne
- 91 – Ile de France Sud

Liste des cours Matif du 25/10/2007

Dernière mise à jour : Le 25/10/2007 à 19:45:09
Production :

Cotations Matif

	Cours de compensation	Variation	+ bas	+ haut	Nb de lots (PO)	Lots traités	Prendre position
Novembre 2007	202.75	-1.75	192.50	202.75	2095	240	
Janvier 2008	207.00	-2.00	194.50	207.00	4412	160	fixer
Mars 2008	208.75	-1.75	200.00	210.00	2311	375	
Juin 2008	212.25	-1.75	210.00	210.00	509	10	
Aout 2008	209.50	-1.75	0.00	0.00	26	0	
Novembre 2008	0.00	0.00	0.00	0.00	0	0	

[Cliquez ici pour acheter des options Call](#)

Historique des cours

	24/10/2007	23/10/2007	22/10/2007	19/10/2007	17/10/2007	16/10/2007
Novembre 2007	204.50	213.25	213.00	213.50	210.00	213.25
Janvier 2008	209.00	217.25	216.50	217.50	215.00	217.75
Mars 2008	210.50	218.75	218.50	218.50	215.00	218.50
Juin 2008	214.00	220.00	219.50	219.50	217.00	220.50
Aout 2008	211.25	217.25	216.75	216.75	214.25	217.75
Novembre 2008	0.00	0.00	0.00	0.00	0.00	0.00

Rejoignez-les !

Si vous souhaitez bénéficier d'une solution éprouvée et facile à mettre en œuvre, rejoignez Adhérent.coop !

Contact :
Sébastien Gaborit



La commercialisation des céréales en ligne s'envole !

Pour mieux accompagner les coopératives qui souhaitent mettre en œuvre des outils permettant aux agriculteurs de maîtriser la commercialisation de leurs productions, Adhérents.Coop vient de mettre à disposition tout un ensemble de modules de commercialisation via internet : Prix indexé Matif, Prix du jour, Prix sur contrats. L'accueil très favorable fait à ces nouveaux outils par les agriculteurs et les responsables de coopératives nous renforce dans l'idée que demain les extranets, pour peu qu'ils apportent un service simple et efficace, deviendront incontournables.

La mise en place des extranets est plus limitée (47 % des coopératives interrogées) mais 80 % des responsables interrogés déclarent que d'ici quelques années, il sera indispensable ou important que sa structure en possède un.

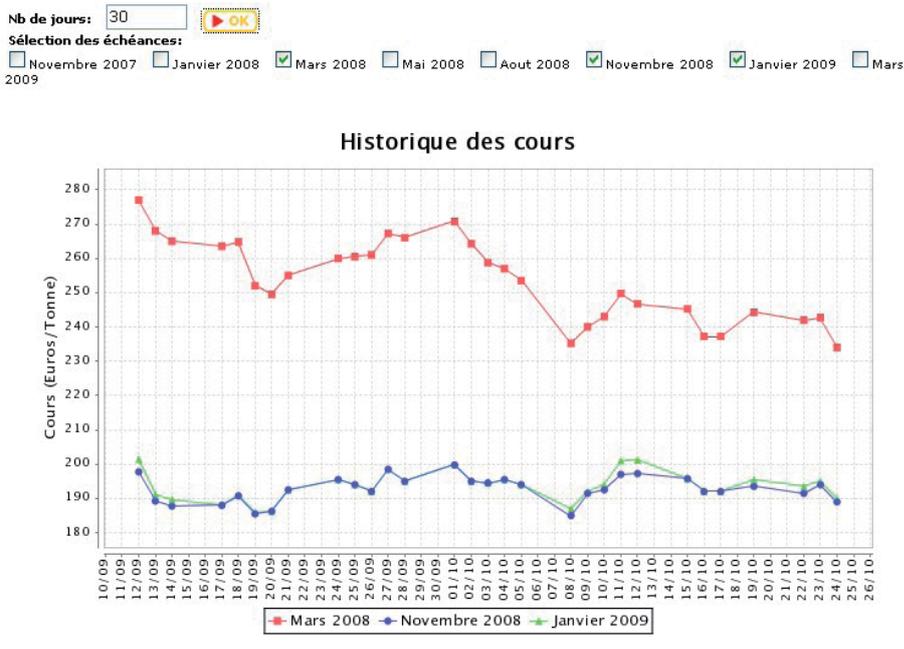
Face à ces attentes, Adhérents.Coop poursuit ses actions pour accompagner les coopératives dans la maîtrise de ces outils. Avec son kit, Adhérents.Coop permet :

De développer un outil indépendant

Avec le kit, chaque coopérative garde la maîtrise de ses informations et de leur distribution.

De déployer rapidement un extranet personnalisé

Fort de son expérience et de sa connaissance, Adhérents.Coop vous



permet de déployer un kit en quelques mois sans mobiliser systématiquement les équipes informatiques.

De bénéficier de nouveaux services

Chaque année, Adhérents.Coop développe de nouveaux services qui répondent aux mieux à vos préoccupations de distributeur.

De maîtriser son budget

Les coûts de fonctionnement annuel vous permettent de disposer de l'ensemble des modules développés, sans coûts d'investissements supplémentaires.

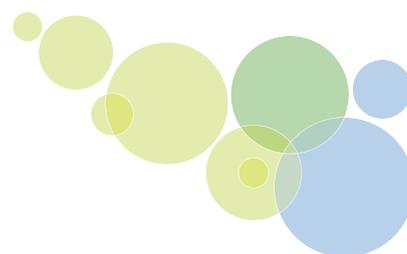
Une nouvelle identité visuelle pour de nouvelles ambitions

Pour mieux affirmer son attachement à COOP DE FRANCE et faciliter sa reconnaissance, Adhérents.Coop s'est doté d'une identité visuelle qui rappelle

le tout à la fois, sa volonté d'accueillir le plus grand nombre de coopératives



venant d'horizon et de métiers différents et sa capacité à offrir des outils modulaires.



Rapport d'activité 2007

Directeur de la Publication : Vincent MAGDELAINE

Entreprises : Hubert LEGUILLETTE

Economie : Vincent MAGDELAINE

Environnement : Jacques SALÉ

Services : Rémi PLÉAU

Coordination - Réalisation : Yves BOULAY



11, rue des Halles - 75039 Paris cedex 01
Tél. 01 53 00 90 00 • Fax. 01 40 26 64 64
metiersdugrain@coopdefrance.coop
www.metiersdugrain.coopdefrance.coop