

Cas d'Utilisation de l'ETL Spatial Data Integrator au sein du CEN-LR



ATEN / Forum TIC
Montpellier - 2 et 3 juin 2010



par
Yoann Perrot

SOMMAIRE

- Qu'est ce qu'un ETL?
- de Talend Open Studio (TOS) à Spatial Data Integrator (SDI)
- Cas d'utilisation: « Mise en ligne d'un lot de données catalogué sur CARMEN »
- Bilan et perspectives des utilisations

Qu'est ce qu'un ETL

- ETL = **E**xtract - **T**ransform - **L**oad
- Technologie informatique intergicielle permettant d'effectuer des ***synchronisations massives d'information***
- ***Extraction*** de données de différents formats
- ***Manipulation*** des données sources et ***mise en correspondance*** avec les données cibles (mappage)
- ***Chargement*** des données résultantes

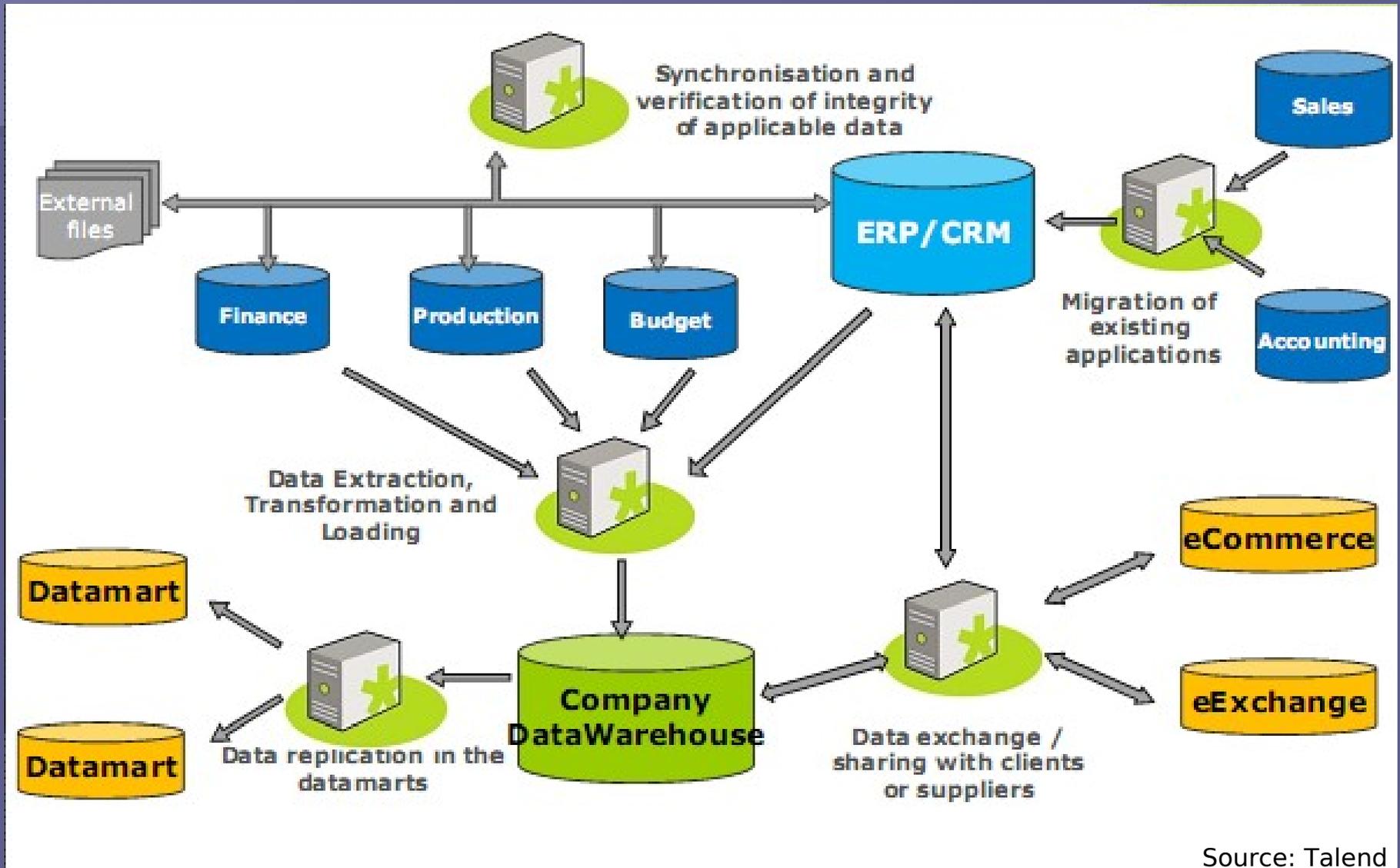
Sommaire

ETL ?

SDI / TOS

Cas d'utilisation

Bilan et perspectives



Source: Talend

de Talend Open Studio (TOS) à Spatial Data Integrator (SDI)

- SDI est une évolution spatiale de TOS
- 2 logiciels ETL open source (licence GPL)

Talend Open Studio (TOS)

par 

- ETL open source et gratuit (version supérieur en souscription)
- 1^{er} logiciel open source spécialisé dans l'intégration de donnée
- Plus de 300 composants/connecteurs
- 300 000 utilisateurs

Spatial Data Integrator (SDI)

par

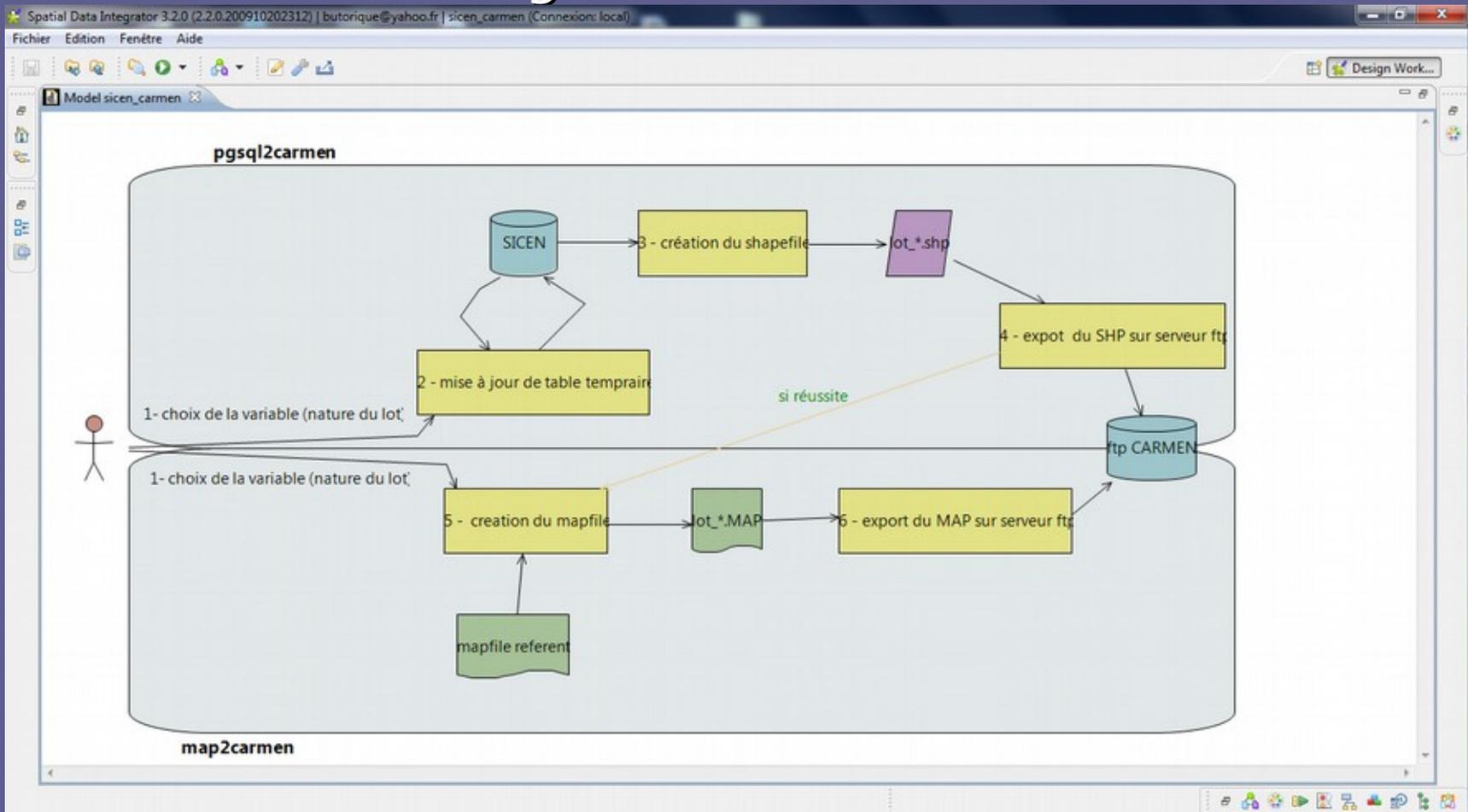


- TOS + composantes spatiales
 - Lecture/écriture de formats SIG
 - Transformations géo-spatiaux
 - Création et publication de métadonnées
- Plus de 50 composants/connecteurs

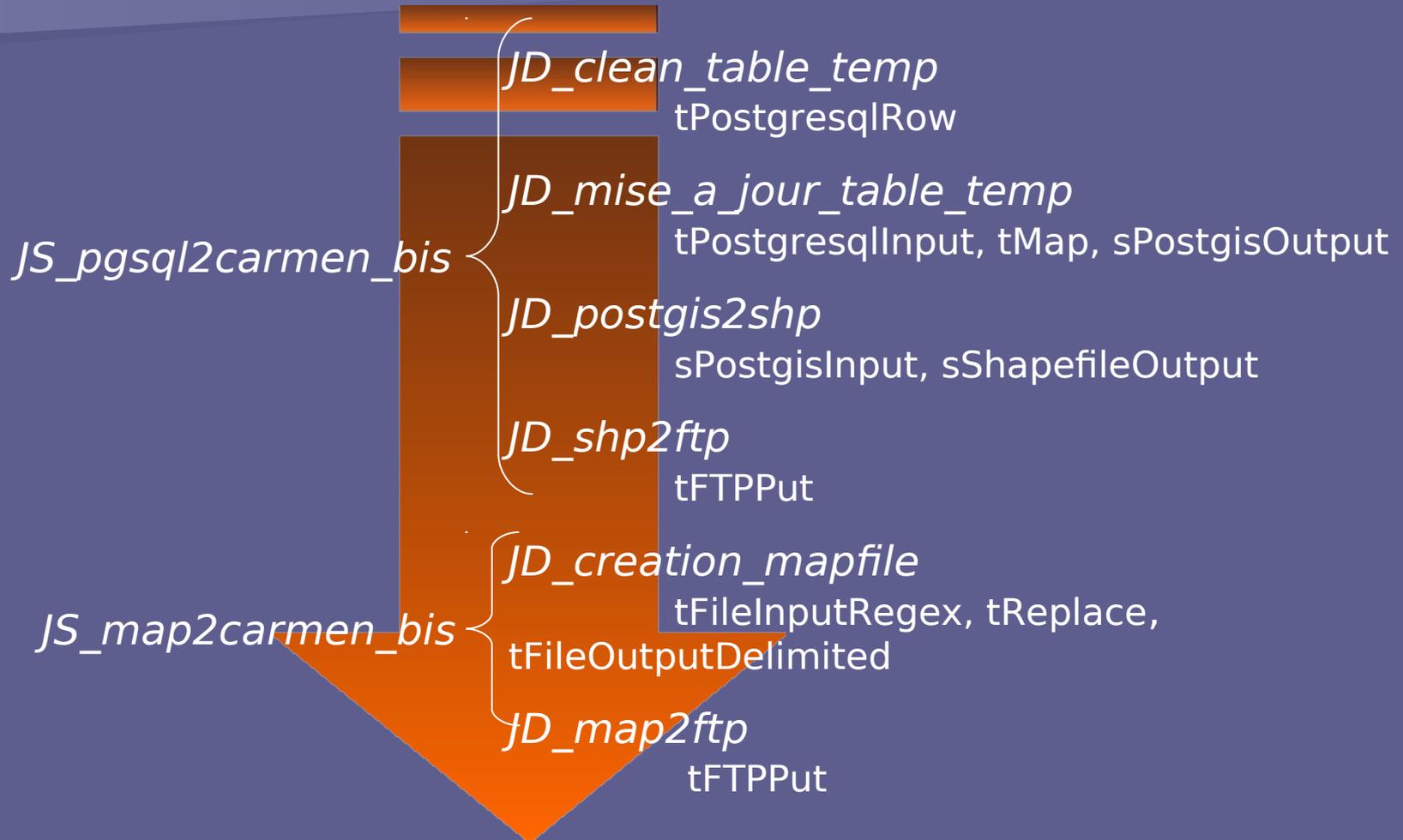
Cas d'utilisation par le CEN-LR

- **Contexte CEN-LR** : *Evolution de la chaine de traitement de la donnée produite*
 - 1 - phase terrain
 - 2 - saisie / validation / intégration des données à la bdd spatiale - *outil en cours de développement*
 - 3 - publication des métadonnées (en partie calculées dans la bdd et avec SDI) - *outil en cours de développement*
 - 4 - diffusion du lot (via SDI) - *présenté ci-après*

- Cas d'utilisation: « Mise en ligne d'un lot de données catalogué sur CARMEN »



- Processus et composants du job *sicen_carmen*



Bilan et perspectives

■ BILAN

- Outils open source
- Données spatiales ou non
- SDI, basé sur la version gratuite de Talend (TOS), offre une large gamme de fonctionnalités
- Création de routines, d'exécutables réutilisables, d'automatisations de tâches sans passer par l'écriture de scripts
- Automatisation de chaînes de traitements complexes difficiles à réaliser dans un script
- Communauté relativement active et en augmentation
- Nécessite une formation pour la prise en main

■ PERSPECTIVES

- Contexte SINP

- Transformation/formatage/normalisation des données
- Diffusion des données produites
 - vers des partenaires (échange de fichiers normalisés)
 - mise en ligne sur serveur cartographique (par exemple sur carmen)
 - production d'un flux rss permettant aux structures de maintenir sur leur site web la liste de données mises en ligne

■ PERSPECTIVES

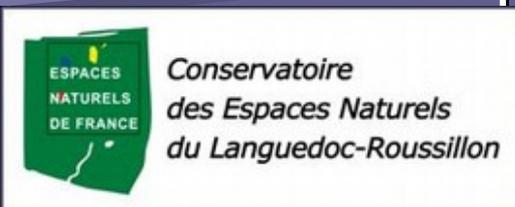
– Contexte **INSPIRE**

- Transformation de données
- Tests à venir dans le cadre du projet NatureSDIplus

Cas d'Utilisation de l'ETL Spatial Data Integrator au sein du CEN-LR



ATEN / Forum TIC
Montpellier - 2 et 3 juin 2010



par
Yoann Perrot