

POUR L'ÉCOLE
DE LA CONFIANCE



MENJ - RETEX sur la transformation de données

Présentation 2RCE-Cassis 2019

Intervenants :
Raphaël RICORDEAU, Florent BARTH

SOMMAIRE

- Contexte
- Cas d'usage et enjeux
- RETEX axe métier
- RETEX axe système
- La suite ...

POUR L'ÉCOLE
DE LA CONFIANCE

RETEX sur la transformation de données

Contexte

Contexte : Présentation de l'équipe

Equipe à compétence nationale au rectorat de Dijon

▪ Mission d'étude et développement (16 personnes)

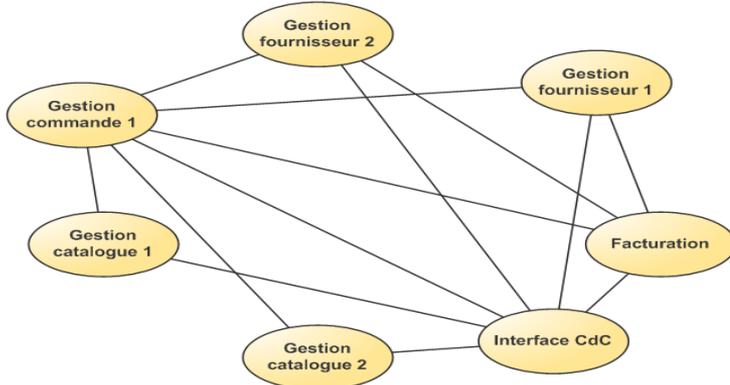
- 2 grands domaines : les applications métier et les systèmes d'échange
- Etude et réalisation de projets nationaux
- Expertise et conseil sur les technologies spécifiques aux échanges (Bus de service, XML, Web services, API, ETL, Orchestrateur, ...)
- Accompagnement des équipes de développement au raccordement technique (Présentations, support aux équipes, mise à dispo d'un environnement de dev).
- Expertise et conseil sur la mise en œuvre des échanges et interconnexion entre SI (internes et/ou externes) : CAU, Ateliers DINSIC, Groupe d'experts, Groupe urbanisation...

▪ Mission d'intégration (4 personnes)

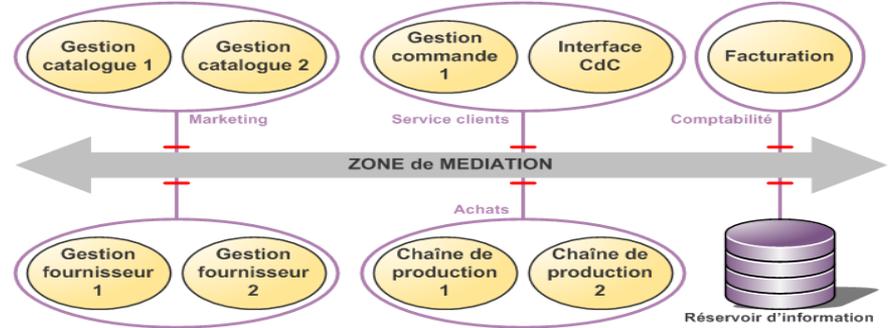
- Ingénieurs systèmes spécialisés dans l'automatisation des installations : conception de socles d'installation
- Qualification technique, Tests de montée en charge / Tests techniques

Contexte : Les systèmes d'échange

NON !



OUI !



Source : <http://blog.xebia.fr/>

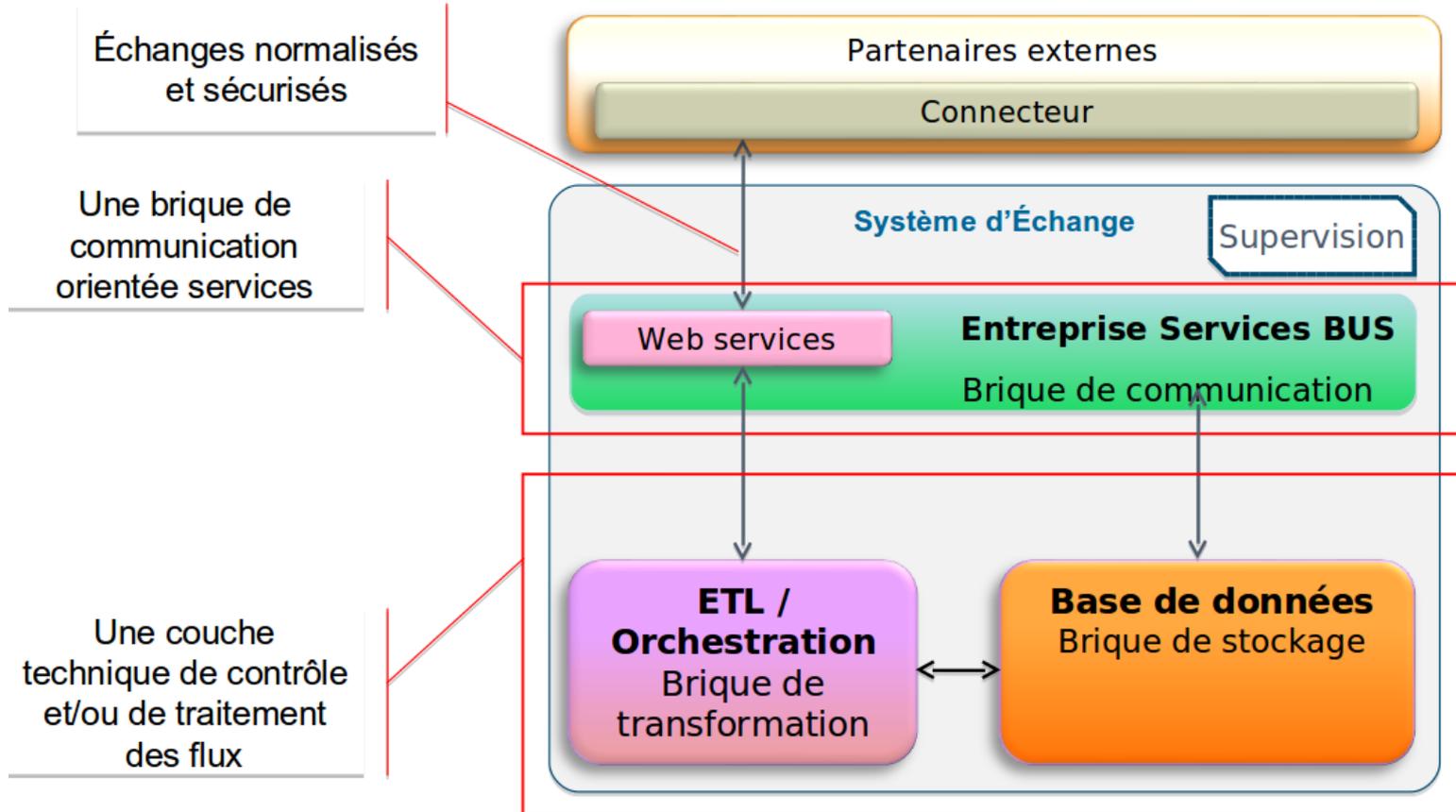
A chaque besoin d'échange entre deux SI hétérogènes, on retrouve toujours deux problématiques :

- Le transport de l'information échangée.
- Le contenu de l'information échangée.

La problématique du transport de l'information est toujours la même : Comment assurer le transport de manière sécurisée, indépendamment de la volumétrie, avec des accusés de réception, une fiabilité et une intégrité des données ?

→ Solutions de médiation

Contexte : nos projets SE



POUR L'ÉCOLE
DE LA CONFIANCE

RETEX sur la transformation de données

Cas d'usage et enjeux



Contexte : Cas d'usage des ETL

- Transformation de formats au sein de nos SE.
- Intégration : Automatisation de tâches répétitives (tests automatisés pour nos systèmes d'échange)
- Exploitation : Rendre autonome nos partenaires (Robot mail).
- Développement : Anonymiser bases de production pour réaliser des environnements de recette.
- Décisionnel (équipe académique) : Fusion de bases, création d'entrepôts, calculs d'indicateur

Contexte : Enjeux

- Développer plus rapidement et profiter des connecteurs embarqués
- Aligner les pratiques et avoir du code homogène
- Permettre à des non développeurs Java de participer à la réalisation de nos projets

2 ETL utilisés :

- Talend Data Integration (**Version communautaire**)
- Oracle Data Integrator (**V10**)

Contexte : Rapide comparatif

ODI	TALEND DI
Outil de développement propriétaire	Variante d'Eclipse
Besoin d'exécuter le code pour voir les erreurs de syntaxe	Possibilité de basculer en JAVA et de compiler avant exécution
Environnement d'exécution nécessitant un environnement d'exécution ODI + JVM	Environnement d'exécution ne nécessitant qu'une JVM (Le jar peut embarquer le runtime + le code)
Traitements ensemblistes (On pense en SQL)	Traitements séquentiels (On pense en JAVA)
Les traitements communs à plusieurs scénarii doivent être copiés d'un scénario à un autre si on veut les réutiliser.	Les traitements communs à plusieurs programmes n'ont pas besoin d'être dupliqués : ils peuvent être appelés. (Sous Job)
Notion d' External tools pour étendre les fonctionnalités	+ de 500 composants et possibilité de développer ses propres composants
Manipulation XML complexe	Manipulation XML via des composants natifs et temps de réponse meilleurs
Cher !	Talend version Open source : 0€ Talend version payante : environ 40 000€ / an (3 dév + 1 env. hors prod + 1 env. prod)

POUR L'ÉCOLE
DE LA CONFIANCE

RETEX sur la transformation de données

RETEX Métier



RETEX Métier : Inconvénients

- Effet boîte noire pour les développeurs : difficulté à diagnostiquer les problèmes notamment les problèmes de performance
- Concepts et ensemblisme difficiles à appréhender. Besoin d'un accompagnement
- Pour ODI v10, manipulation de XML pas natif
- Pour Talend, des manques dans la version communautaire (travail collaboratif, manque l'opérateur pour le suivi des log...)
- Pour ODI, versionning du code très lourd (en base de données)

POUR L'ÉCOLE
DE LA CONFIANCE

RETEX sur la transformation de données

RETEX Système



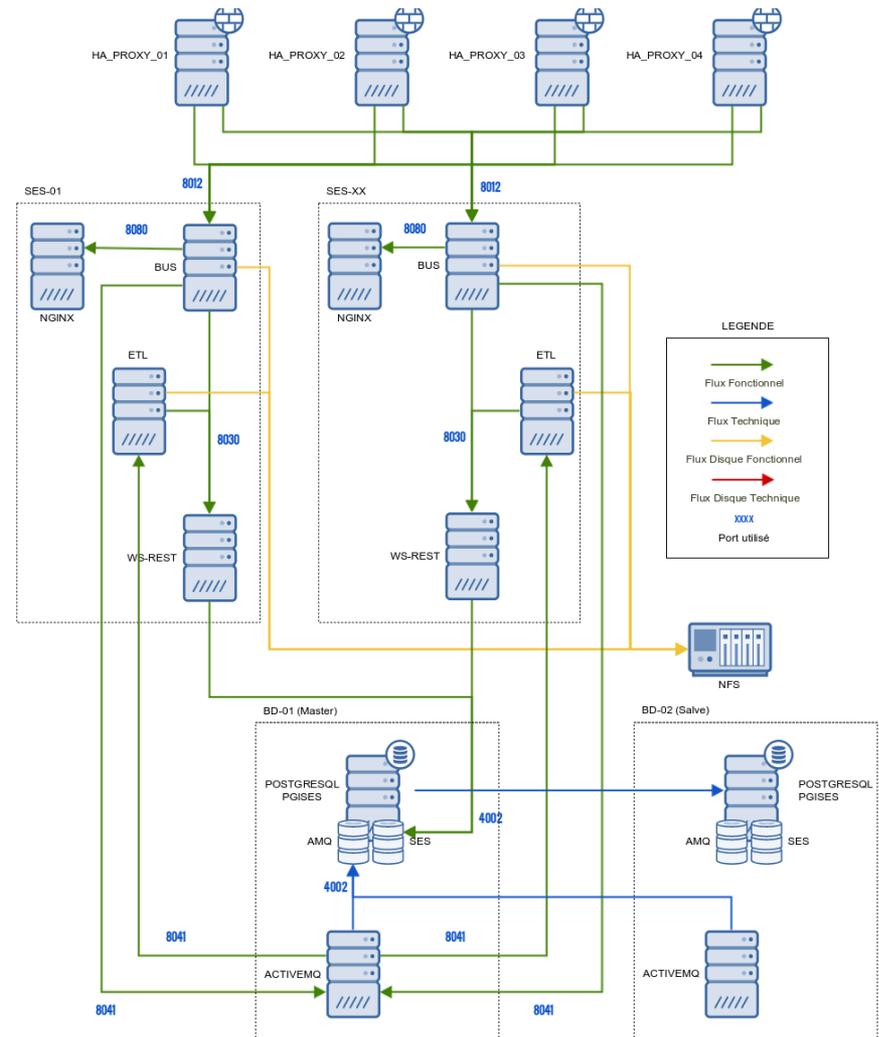
RETEX Système : Avantages

- Déploiement middleware rapide pour TALEND (tar.gz à décompresser + fichier xml à éditer)
- Automatisation complète de l'installation possible via le composant ssh de TALEND.
- Support éditeur pour ODI et TALEND Entreprise
- Quelques métriques JMX

RETEX Système : Inconvénients

- Montée de versions difficiles sur ODI (pas de doc de migration)
- Pas de mode Cluster sur ODI et à faire soi-même avec TALEND DI communautaire
- ODI, pas de fail over côté BDD
- TALEND : énormément de composants (Karaf, Jvm, ActiveMQ,...)
- Pas de métriques intégrées dans les outils pour diagnostic des performances

Exemple d'implémentation



POUR L'ÉCOLE
DE LA CONFIANCE

RETEX sur la transformation de données

La suite ...

La suite ... : Le MDM

■ La transformation de données c'est bien mais la mise en qualité c'est mieux !

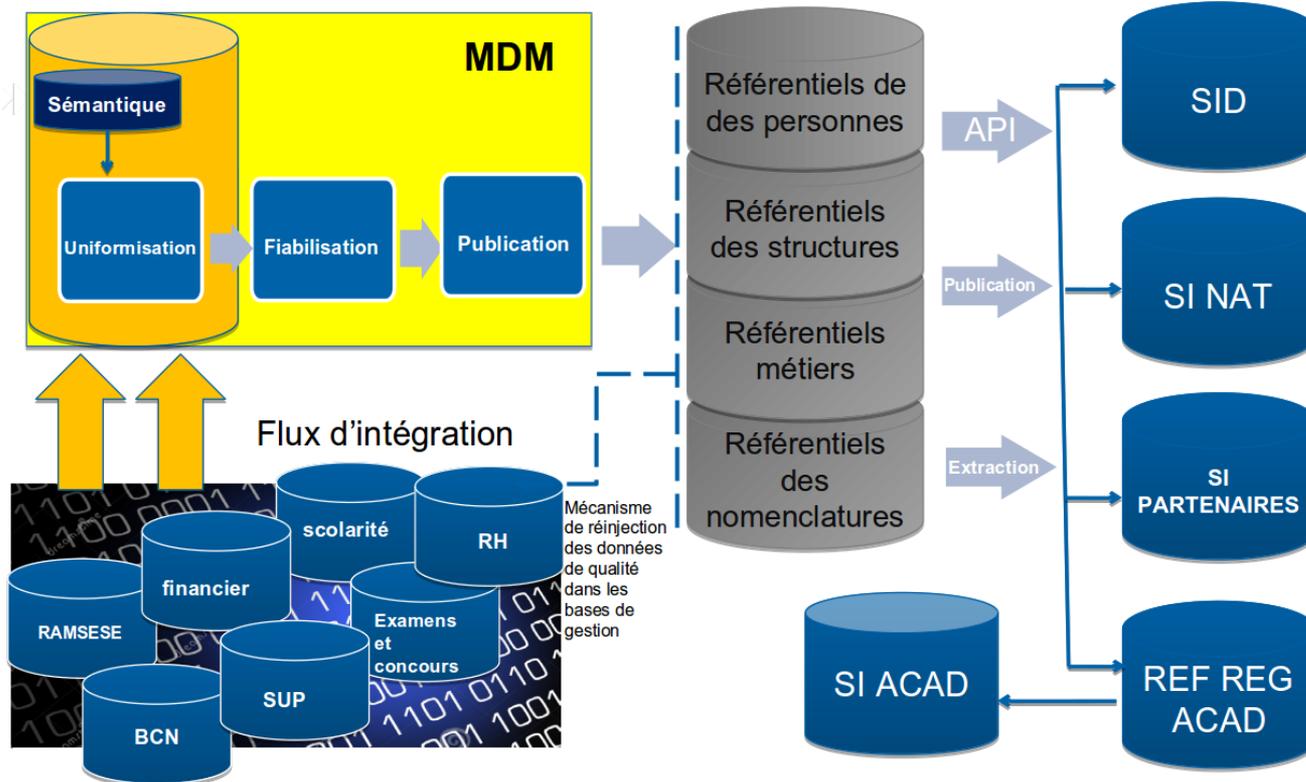
• Problèmes rencontrés lors de la transformation :

- Doublons (élèves, responsables légaux, ...)
- Données manquantes (notamment dans des nomenclatures)
- Mauvais formattage ou données erronées (date de naissance en 0003, adresses...)

• La solution ? Le Master Data Management

- Construction d'un référentiel de données
- Module de mise en qualité
- Tableau de bord
- Exposition via API

La suite ... : Le MDM



POUR L'ÉCOLE
DE LA CONFIANCE

RETEX sur la transformation de données

Echanges / Questions ?



POUR L'ÉCOLE DE LA CONFIANCE

Contact :

Equipe nationale de Dijon

Florent BARTH

florent.barth@ac-dijon.fr

Raphaël RICORDEAU

raphael.ricordeau@ac-dijon.fr

