



**DIRECTION DES SYSTEMES D'INFORMATION
93006 BOBIGNY CEDEX**

INTEGRATION DES FLUX CAF GUIDE TECHNIQUE

Objet	Guide technique des jobs ETL	Rédacteur	Dinah BLIRANDO
Date création	06/08/2010	Date mise à jour	02/02/11

Documents de référence		
Nom du document	Type	Version
INTEGRATION_FLUX_2_0_Architecture	PDF	2.0
INTEGRATION_FLUX_2_0_Annexes	PDF	2.1

SOMMAIRE

SOMMAIRE.....	2
PRESENTATION.....	3
Présentation.....	3
Versions.....	3
Principe général.....	3
ARCHITECTURE DU PROGRAMME.....	5
Définition des jobs.....	5
Arborescence et nomenclature.....	5
Environnement Talend.....	5
Arborescence.....	5
Nomenclature.....	5
Aperçu dans Talend.....	6
Contextes webrsa.....	6
Gestion des erreurs et des rejets.....	8
Distinction entre “rejet” et “erreur”	8
Etapas Extraction et Transformation	8
Rejets	8
Erreurs.....	8
Etape Insertion.....	8
Rejets.....	8
Erreurs.....	8
Gestion des fichiers de log, de rejet et temporaires	9
Fichier de log et de rejets.....	9
Nomenclature des fichiers de logs et de rejets.....	9
Fichiers temporaires.....	9
ENCHAINEMENT DES JOBS.....	10
Les jobs Master	10
Présentation.....	10
Traitement.....	10
Les jobs Maître	10
Présentation.....	10
Traitement.....	10
Les jobs Fils	11
Les jobs d'administration	11
Gestion des erreurs.....	11

PRESENTATION

Présentation

Les données concernant les dossiers de demandes ou de bénéficiaires RSA sont transmises au Conseil Général par la CAF via des flux XML.

Il existe **trois types de flux XML** :

- les flux **INSTRUCTION** provenant de l'application **@RSA**;
- les flux **BENEFICIAIRE** provenant de l'application **Cristal** ;
- les flux **FINANCIER** provenant de l'application **Cristal**.

Ces flux sont ensuite intégrés dans l'application **WebRSA** via des traitements **ETL Talend** qui leur sont propres, mais adoptent un principe général commun.

Ce document a pour but la présentation technique de ces traitements.

Versions

Concernant le traitement Instruction :

- la version **2.0** gère l'intégration des flux de la version **3.3** (virs0201) à **7** (virs0501) de **@RSA** dans **WebRSA 2.0rc9**.
- la version **2.1** gère l'intégration des flux de la version **3.3** (virs0201) à **7** (virs0501) de **@RSA** dans **WebRSA 2.0rc12**.

Concernant les traitements Bénéficiaire et Financier :

- La version **2.0** gère l'intégration des flux des versions **29** (vrsb0301 – vrsf0201) à **32** (vrsb0401 – vrsf0301) de **Cristal** dans **WebRSA 2.0rc9**.
- La version **2.1** gère l'intégration des flux des versions **29** (vrsb0301 – vrsf0201) à **32** (vrsb0401 – vrsf0301) de **Cristal** dans **WebRSA 2.0rc12**.

L'environnement de développement utilisé est **Talend Open Studio 3.2.3**.

Principe général

Chaque traitement est découpé en 3 sous-traitements ou étapes :

- **Etape 1 - Extraction** : extraction des données du fichier XML et intégration dans les tables du schéma *staging* de la base *STAGING_RSA* avec création d'identifiants propres au fichier en cours.
 - Le schéma *staging* est une copie au plus près du fichier XML source.
 - Les données extraites du fichier XML sont stockées en base avec un type *chaîne de caractères* ou *date*.
 - Des identifiants propres au fichier traité sont créés.
 - Des triggers sont installés pour garder en base uniquement la dernière occurrence d'un dossier (numdemrsa) ou d'une personne dans un dossier (infosfoyerrsa, nomnai, prenom, dtnai, rgnai).
 - Il peut y avoir des rejets par rapport à la syntaxe du fichier, à la longueur de la chaîne, ou par rapport au type date.
 - La présence des balises de début et de fin est aussi vérifiée. Le traitement est interrompu en cas de balise absente.

- **Etape 2 - Transformation** : traitement des données du schéma *staging* et insertion des données transformées dans le schéma *elementaire* de la base *STAGING_RSA*.
 - Le schéma *elementaire* est une copie au plus près de la base cible *webrsa*.
 - Les données récupérées des tables *staging* sont CASTées aux types correspondants du schéma *elementaire* et donc de la base cible *webrsa*.
 - Il peut y avoir des rejets par rapport à ces types.
- **Etape 3 - Insertion** : insertion des données du schéma *elementaire* de la base *STAGING_RSA* dans la base cible *webrsa* après vérification des contraintes d'intégrité.
 - En fonction des identifiants et des règles de gestion des tables cibles, les données sont insérées ou mises à jour dans les tables cibles *webrsa*, avec quelques fois suppression préalable des données précédentes.
 - Des tables de référence sont utilisées pour avoir la liste des correspondances entre les identifiants propres au fichier et des identifiants techniques de la base *webrsa*.
 - Il peut y avoir des rejets uniquement dans le cas de *NUMDEM RSA* inexistant en base pour le flux FINANCIER.
- Un schéma **administration** est aussi utilisé dans la base *STAGING_RSA* pour les contrôles, les données de référence, les rejets et les rapports d'exécution.

ARCHITECTURE DU PROGRAMME

Définition des jobs

Dans ce document, on parlera de 4 types de jobs :

- les **jobs Master** sont utilisés pour lancer les traitements d'intégration ;
- les **jobs Maître** correspondent à l'exécution d'une étape ;
- les **jobs Fils** correspondent au traitement proprement dit d'une ou plusieurs balises du flux XML traité ou tables cibles ;
- les **jobs d'administration** correspondent à un traitement d'administration (nettoyage des tables de la base *STAGING_RSA*, gestion des rejets, ...).

Arborescence et nomenclature

Environnement Talend

Arborescence

Chaque traitement a son propre projet. Ci-dessous la description hiérarchique de l'arborescence dans *Talend*.

RSA : dossier racine commun utilisé par tous les traitements d'intégration (Instruction, Bénéficiaire, Financier)

A0X_NomFlux : liste des dossiers et des **jobs Master** du flux correspondant

A_Extraction : liste des **jobs Maître et Fils** de l'étape Extraction

B_Transformation : liste des **jobs Maître et Fils** de l'étape Transformation

C_Insertion : liste des **jobs Maître et Fils** de l'étape Insertion

Z_Administration : liste des dossiers d'administration

A_Extraction : liste des **jobs d'administration** de la partie Extraction

B_Transformation : liste des **jobs d'administration** de la partie Transformation

C_Insertion : liste des **jobs d'administration** de la partie Insertion

Nomenclature

Ci-dessous la nomenclature utilisée pour le nommage des différents jobs :

- **job Master (niveau 1) :**
 - Flux_Staging_Master
 - Flux_WebRSA_Master
 - Flux_Master
- **job Maître (niveau 2) :**
 - Etape_Flux_XX_Maitre
- **jobs Fils (niveau 3) :**
 - Etape_Flux_XX_Balises/tables_traitées, chaque job traitant une ou plusieurs tables
- **jobs Administration (niveau 3) :**
 - Etape_Flux_Administration_XX_Titre

Aperçu dans Talend

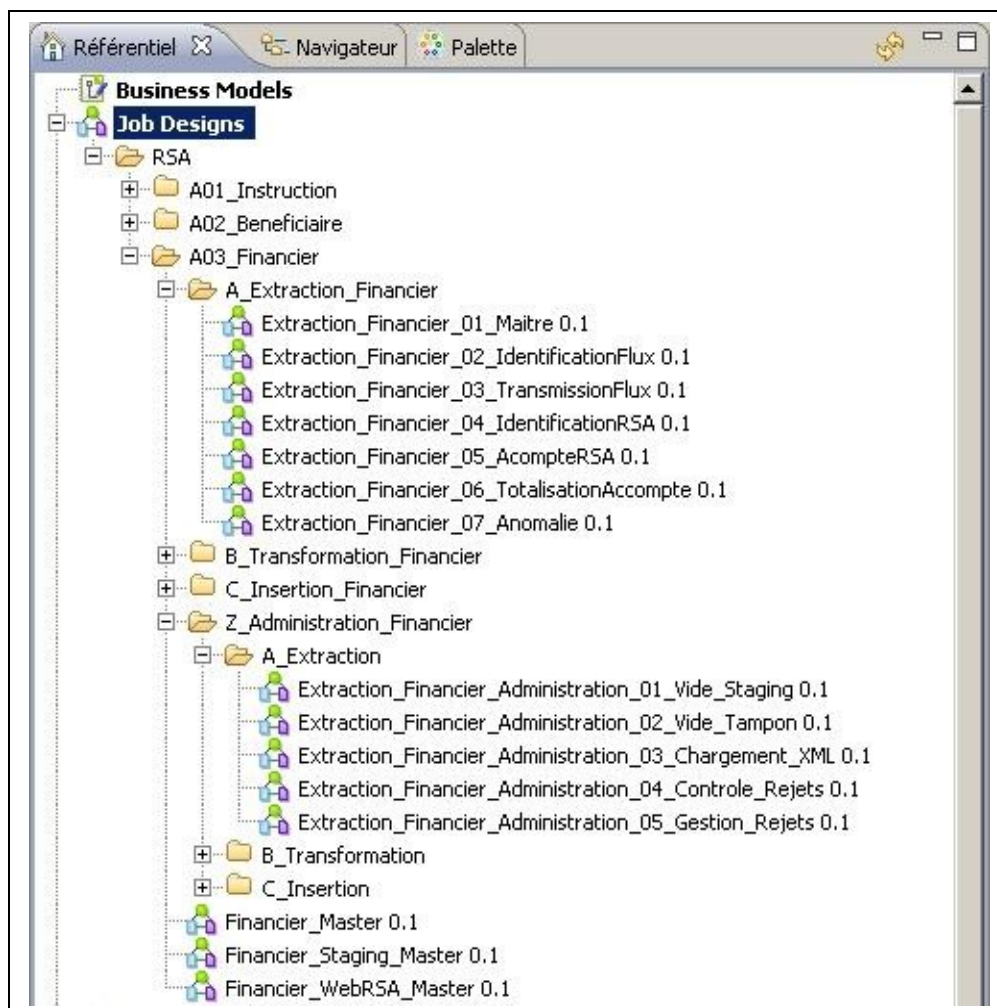


Figure 1 : Arborescence Talend (projet unique)

Contextes webrsa

Les contextes Talend utilisés par les traitements sont les suivants :

- **projet_webrsa** : arborescence sur le serveur ETL
- **webrsa** : paramètres de connexion à la base *webrsa*
- **staging_webrsa** : paramètres de connexion au schéma *staging* de la base de données STAGING_RSA
- **elementaire_webrsa** : paramètres de connexion au schéma *elementaire* de la base de données STAGING_RSA
- **Administration_webrsa** : paramètres de connexion au schéma *administration* de la base de données STAGING_RSA

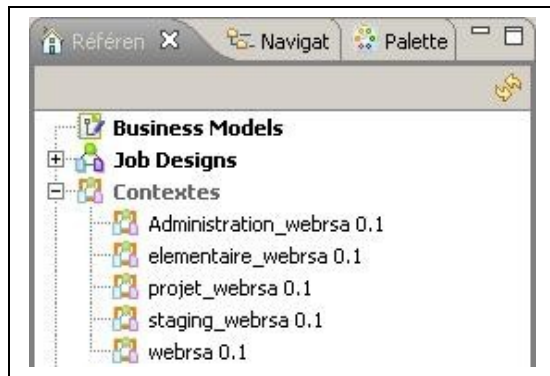


Figure 2 : Contextes Talend

Pour chacun d'eux, quatre environnements d'exécution sont pré configurés : Default, PROD, QUAL, TEST.

Le tableau ci-dessous liste les différentes variables à mettre à jour par contexte.

Contexte	Nom de la variable	Description
projet_webrsa	repertoireSource	répertoire source contenant les fichiers XML à traiter
	fichierParametre	chemin complet du fichier de propriété
	directoryRejects	répertoire contenant les fichiers de log et de rejet
	directoryArchives	répertoire où sont archivés les fichiers CAF après traitement
	directoryTemp	répertoire contenant les fichiers temporaires
	fileRejects	nom des fichiers de rejets
	fileLogsReject	nom des fichiers de log des rejets
	fileLogsAvertissement	nom du fichier de log d'exécution
webrsa	webrsaServer	Nom du serveur
	webrsaPort	Port d'écoute
	webrsaSid	Nom de la base de données
	webrsaSchema	Nom du schéma
	webrsaLogin	Nom d'utilisateur
	webrsaPassword	Mot de passe
staging_webrsa	staging_webrsa_Server	Nom du serveur
	staging_webrsa_Port	Port d'écoute
	staging_webrsa_Database	Nom de la base de données
	staging_webrsa_Schema	Nom du schéma
	staging_webrsa_Login	Nom d'utilisateur
	staging_webrsa_Password	Mot de passe
elementaire_webrsa	elementaire_webrsa_Server	Nom du serveur
	elementaire_webrsa_Port	Port d'écoute
	elementaire_webrsa_Database	Nom de la base de données
	elementaire_webrsa_Schema	Nom du schéma
	elementaire_webrsa_Login	Nom d'utilisateur
	elementaire_webrsa_Password	Mot de passe
Administration_webrsa	Administration_webrsa_Server	Nom du serveur
	Administration_webrsa_Port	Port d'écoute
	Administration_webrsa_Database	Nom de la base de données
	Administration_webrsa_Schema	Nom du schéma
	Administration_webrsa_Login	Nom d'utilisateur
	Administration_webrsa_Password	Mot de passe

Gestion des erreurs et des rejets

Distinction entre “rejet” et “erreur”

Les **rejets** sont des erreurs SQL non bloquantes concernant un dossier RSA. Ils sont logués dans la table *administration.rejet* et archivés dans la table *administration.rejet_historique*.

Les **erreurs** sont les autres erreurs (SQL, java, *tDie*) qui sont bloquantes et interrompent le traitement.

Etapas Extraction et Transformation

Rejets

Sont traités comme des rejets :

- les erreurs SQL lors de l'insertion de données d'un dossier RSA dans une table du schéma *staging* ou *elementaire* de la base *STAGING_RSA*.

Erreurs

Sont traités comme des erreurs :

- les erreurs java ;
- les *tDie* ;
- les erreurs SQL non traitées comme des rejets (suppressions, schéma administration, traitement des données propres au flux entier - *IdentificationFlux*, *TransmissionFlux*, *TotalisationAcompte*, ...).

Etape Insertion

Rejets

Sont traités comme des rejets, *uniquement* pour le flux Financier :

- l'absence d'un numéro de dossier RSA (*NUMDEM RSA*) en base.

Erreurs

Sont traitées comme des erreurs :

- les erreurs java ;
- les *tDie* ;
- les erreurs SQL non traitées comme des rejets, à savoir, toutes les erreurs exception faite du *NUMDEM RSA* inexistant en base pour le flux Financier.

Gestion des fichiers de log, de rejet et temporaires

Fichier de log et de rejets

Trois types de fichiers de log et de rejet peuvent être générés par les traitements. Chaque flux XML traité a ses propres fichiers.

- **un fichier de log d'exécution - *fileLogsAvertissement* :**
 - Il s'agit d'un fichier enregistrant les dates de début et de fin du traitement étape par étape.
 - En cas d'erreur, les informations suivantes seront aussi loguées :
 - la date ;
 - le message d'erreur ;
 - le job en cours ;
 - le contexte en cours.
- **0, un ou plusieurs fichiers de log des rejets - *fileLogRejects* :**
 - Il s'agit d'un fichier enregistrant le message de rejet de dossier, extrait de la table *administration.rejet*. Il y a un fichier par NUMDEM RSA.
 - Les informations suivantes y sont recensées :
 - l'identifiant propre au fichier du dossier rejeté ;
 - le type de flux ;
 - l'étape ;
 - la table en erreur ;
 - le message d'erreur ;
 - le NUMDEM RSA du dossier rejeté ;
 - le MATRICULE du dossier rejeté.
- **0, un ou plusieurs fichiers de rejets - *fileRejects* :**
 - Il s'agit de l'extraction du dossier rejeté à corriger et à rejouer. Il y a un fichier par NUMDEM RSA.

Nomenclature des fichiers de logs et de rejets

- **fileLogAvertissement**
 - nomFichier.log / nomFichier_Staging.log / nomFichier_WebRSA.log
- **fileLogsReject**
 - etapeX_nomFichier_NUMDEM RSA.ERR
- **fileRejects**
 - etapeX_nomFichier_NUMDEM RSA.RCV

Fichiers temporaires

Quatre types de fichiers temporaires peuvent être générés :

- des fichiers de référence des rejets :
 - il s'agit d'un fichier référençant les étapes et identifiants propres au fichier traité des dossiers rejetés. Il y a un fichier par fichier XML traité.
- un fichier flag correspondant à l'étape en cours lorsque survient une erreur ;
- des fichiers temporaires utilisés par les *tMap* lors des *lookup* (*temp.csv*, *tempX.csv*) ;
- des fichiers temporaires utilisés par les *tBulkExec*. Ils sont automatiquement supprimés après traitement.

ENCHAÎNEMENT DES JOBS

Les jobs Master

Présentation

Les jobs Master sont de niveau 1. Ce sont les classes principales appelées par les exécutables de chaque traitement d'intégration.

Un job Master correspond à l'exécution d'une ou de plusieurs étapes du traitement d'intégration :

- **Flux_Staging_Master** : Lance la partie intégration dans la base *STAGING_RSA* (étapes A et B)
- **Flux_WebRSA_Master** : Lance la partie intégration dans *webrsa* (étape C).
- **Flux_Master** : Lance les trois étapes, de l'extraction à l'intégration dans *webrsa* (étapes A, B et C).

Attention : seul le job Master *Flux_Master* permet le traitement d'intégration intégral de tous les fichiers du type de flux traité présents dans le répertoire source.

Les deux autres jobs Master permettent le traitement **d'un seul fichier**. En effet, la base *STAGING_RSA* ne peut traiter qu'un seul fichier de chaque type. A chaque nouveau fichier, la base est purgée avant traitement.

Traitement

Chaque job Master instancie la variable du nom du fichier traité et la transfert, ainsi que les contextes, à chaque job Maître appelé.

Les variables transférées sont :

- les différents contextes (**Administration_webrsa**, **elementaire_webrsa**, **staging_webrsa**, **webrsa**, **projet_webrsa**) ;
- la nouvelle valeur de **nomFichier**.

Les jobs Maître

Présentation

Les jobs Maître sont de niveau 2. Ils sont appelés par les jobs Master.

Un job Maître correspond à l'exécution d'une étape :

- **Flux_Extraction_XX_Maitre** : étape Extraction ;
- **Flux_Transformation_XX_Maitre** : étape Transformation ;
- **Flux_Insertion_XX_Maitre** : étape Insertion.

Traitement

Les jobs Maître gèrent les connexions partagées, instancient les variables de contexte correspondant aux noms des fichiers (logs, erreur..) et les transmettent à leurs sous-jobs ainsi que les contextes reçus de leurs jobs Master. Les variables transférées sont :

- les différents contextes (**Administration_webrsa**, **elementaire_webrsa**, **staging_webrsa**, **webrsa**, **projet_webrsa**) ;
- **fileLogsAvertissement** ;
- **fileLogsReject** * (job administration de contrôle des rejets) ;
- **fileRejects** * (job d'administration de gestion des rejets) ;
- **nomFichier** * (jobs utilisant la visionneuse dans l'étape Insertion – *DossiersRSA*, *personnes*, *DSP*...) ;
- **identificationsflux_id** * (job *TransmissionFlux* de l'étape Insertion).

Les jobs Fils

Les jobs Fils sont de niveau 3. Ils sont appelés par les jobs Maître.

Un job Fils correspond au traitement d'une ou de plusieurs balises du flux XML ou de tables cibles.
Les jobs Fils sont chargés des traitements proprement dits (voir chapitre « Principe général »).

Les jobs d'administration

Les jobs d'administration sont de niveau 3. Ils sont appelés par les jobs Maître.

Ce sont des sous-jobs utilisés pour les traitements d'administration :

- `Etape_Flux_Administration_XX_Vide_Tampon/Staging/Elementaire/Reference` ;
- `Etape_Flux_Administration_XX_Chargement_XML` ;
- `Etape_Flux_Administration_XX_NomFichier` ;
- `Etape_Flux_Administration_XX_Controlle/Gestion/Historique_Rejets`.

Gestion des erreurs

En cas d'erreur dans l'un des jobs, le traitement est interrompu en remontant au job père. Un fichier flag est créé, correspondant à l'étape en cours lors de l'erreur :

- **null / 0** : avant la première étape ;
- **1** : étape extraction ou transformation ;
- **2** : étape insertion.