

Matrice de comparaison des fonctionnalités d'intégration de données

Talend Open Studio

vs

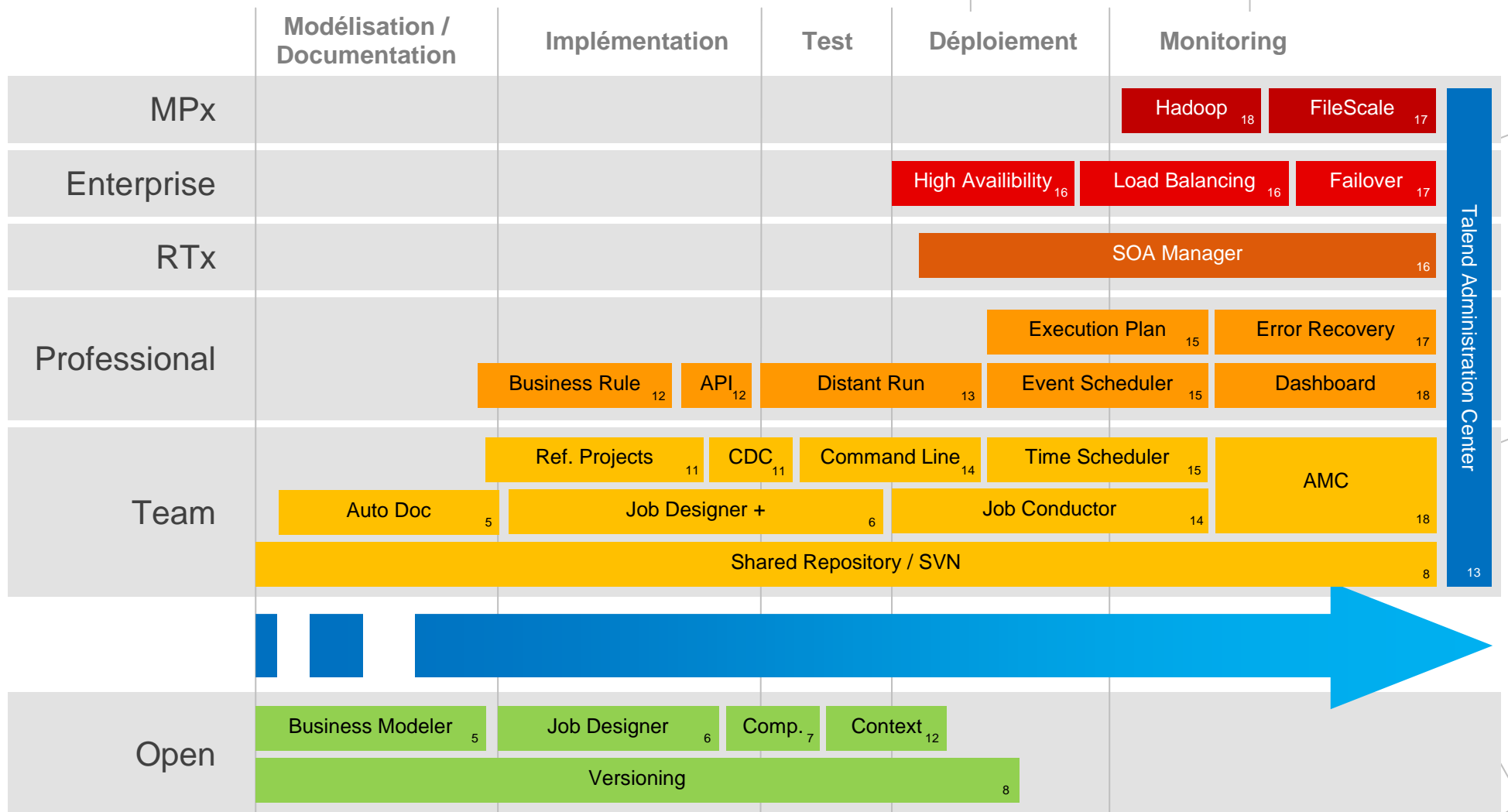
Talend Integration Suite

Team, Professional, RTx, Enterprise & MPx editions

<http://fr.talend.com/products-data-integration/matrix.php>

Q1 2011

Détails des fonctionnalités



Type de licence et Indemnisation

Open Source GPL Licence

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

La licence publique générale GNU, ou GNU General Public License, est une licence qui fixe les conditions légales de distribution des logiciels libres du projet GNU.

L'objectif de la licence GNU GPL est de garantir à l'utilisateur les droits suivants :

- liberté d'exécuter le logiciel pour n'importe quel usage et sans limitation ;
- liberté d'étudier le fonctionnement du logiciel et de l'adapter à ses besoins.

Si l'auteur des modifications du logiciel décide de distribuer son logiciel, il devra le faire sous licence GPL.

L'intégralité du texte de licence GPL peut être consultée ici : <http://www.opensource.org/licenses/gpl-2.0.php>

Accès au code source

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Le code source de Talend Open Studio est disponible ici : <http://www.talendforge.org/trac/tos/>

Dans un souci de totale transparence et de cohérence, Talend fournit également aux clients qui le demandent le code source de tous les outils disponibles en édition commerciale.

Licence de souscription

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Les versions "entreprise" incluent des fonctionnalités à valeur ajoutée (Cf. ci-dessous) et des services qui enrichissent les produits open source ; ces versions sont distribuées sous licence commerciale.

Le modèle de tarification de Talend garantit la transparence et la prévisibilité : le tarif n'est pas lié aux volumes de données ou à d'éventuels besoins supplémentaires en matière de connecteurs ou de CPU mais est indexé sur le nombre de développeurs (Studio), le niveau de fonctionnalités (l'édition choisie) et la durée de l'abonnement.

Cette approche par souscription garantit la rentabilité de votre investissement : le nombre de licences peut être modifié à la hausse ou à la baisse tous les ans, afin de s'adapter à l'évolution de la portée d'un projet ou des effectifs.

Les solutions Talend sont moins coûteuses à déployer, à maintenir et à supporter ; elles sont 50 à 80% moins chères que les solutions propriétaires équivalentes.

Indemnisation	<i>Cette fonctionnalité est incluse dans :</i>	Open	Team	Professional	RTx	Enterprise	MPx
			●	●	●	●	●

Les logiciels open source étant issus d'un travail de développement collaboratif, le code final n'est autre qu'un mélange de contributions provenant de diverses sources. Si leur intégration dans le code n'est pas gérée et contrôlée avec attention, l'utilisation du logiciel final peut ne pas respecter les droits du contributeur original.

L'utilisateur final pourrait dans ce cas, faire l'objet de poursuites judiciaires et financières pour violations de droits, même si cette violation n'est pas intentionnelle.

Talend fournit une clause d'indemnisation à tous ses clients en souscription. L'utilisateur a donc la garantie d'être juridiquement protégé par Talend, même dans le cas où le code Talend ne respectait pas les droits d'un tiers.

Support & Documentation

Communauté : forums, Bugtracker...	<i>Cette fonctionnalité est incluse dans :</i>	Open	Team	Professional	RTx	Enterprise	MPx
			●	●	●	●	●

Les logiciels open source étant issus d'un travail de développement collaboratif, le code final n'est autre qu'un mélange de contributions provenant de diverses sources. Si leur intégration dans le code n'est pas gérée et contrôlée avec attention, l'utilisation du logiciel final peut ne pas respecter les droits du contributeur original.

L'utilisateur final pourrait dans ce cas, faire l'objet de poursuites judiciaires et financières pour violations de droits, même si cette violation n'est pas intentionnelle.

Talend fournit une clause d'indemnisation à tous ses clients en souscription. L'utilisateur a donc la garantie d'être juridiquement protégé par Talend, même dans le cas où le code Talend ne respectait pas les droits d'un tiers.

Support professionnel avec niveaux de service	<i>Cette fonctionnalité est incluse dans :</i>	Open	Team	Professional	RTx	Enterprise	MPx
			●	●	●	●	●

En souscrivant aux Services de Support de Talend, vous bénéficiez de l'expérience de nos experts techniques internes, qui entretiennent des contacts quotidiens avec notre R&D. Ces services ont été mis en place pour garantir efficacité, sécurité et tranquillité d'esprit à tous nos clients en souscription, et sont déclinés sur trois niveaux : Silver, Gold et Platinum. A chacun de ces niveaux sont associées des garanties liées au temps de prise en compte d'un bug déclaré, au temps de réponse pour fournir un correctif, etc.

<http://www.talend.com/professional-support/support.php>

Guide Utilisateur, Guide de Référence

Cette fonctionnalité est
incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

La documentation de Talend Open Studio est téléchargeable gratuitement. Elle est disponible en anglais et en français au format PDF. Deux manuels, le Guide Utilisateur (276 pages) et le Guide de référence des composants, sont disponibles ici : <http://www.talend.com/resources/documentation.php>

Vous pouvez également acheter une copie papier de ces manuels sur Amazon :

http://www.amazon.com/Talend-Open-Studio-Guide-Version/dp/1453632077/ref=sr_1_1?ie=UTF8&s=books&qid=1296040035&sr=1-1

D'autre part, n'hésitez pas à consulter nos Tutoriels qui sont une véritable mine d'informations !

Ils sont accessibles ici : <http://www.talendforge.org/tutorials/menu.php>

Modélisation / Documentation

Business Modeler

Cette fonctionnalité est
incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Le Business Modeler est un outil non-technique (comme Microsoft Visio). Il permet de structurer, au sein d'un diagramme orienté métier, toute l'information documentaire ainsi que les éléments techniques liés à la mise en place des processus d'intégration de données, permettant ainsi aux différentes équipes (Conception, Dev, Test, Prod...) de travailler à partir d'un modèle commun unique à l'aide d'un outil commun.

Par exemple, les utilisateurs métier peuvent exprimer leurs besoins en intégration de données via ces modèles métier (business model). Les équipes en charge du développement et des opérations peuvent ainsi mieux comprendre leurs besoins métier afin de les traduire en processus techniques (Jobs). A la fin de chaque étape de l'implémentation technique (Jobs), le modèle métier peut facilement être mis à jour, pour indiquer l'état d'avancement du développement à tous les intervenants.

Les administrateurs de bases de données peuvent utiliser ces modèles métier afin de partager les métadonnées de connexion aux base de données et les architectes système peuvent ainsi obtenir une vue d'ensemble des pré-requis en termes d'intégration de données.

La création de modèles métier fait partie intégrante des bonnes pratiques qu'une entreprise devrait adopter le plus tôt possible dans un projet de gestion ou d'intégration de données, afin d'en assurer le succès. L'utilisation de modèles métier permet généralement de détecter et de corriger rapidement les goulets d'étranglement ainsi que les points faibles d'un projet, permettant de réduire les dépassements de budget et/ou l'investissement initial.

Auto Doc

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Cette fonctionnalité permet de générer, à la demande, une documentation technique exhaustive pour tous vos Jobs. Elle rassemble, dans un format interactif facilement exploitable (HTML / XML), les métadonnées du job (auteur, version, statut, date de mise à jour, etc.), la vue graphique du job ainsi que tous les paramètres de tous les composants utilisés.

Cette documentation peut être facilement enrichie avec des commentaires personnalisés.

Auto Doc+

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Grâce à AutoDoc+, la documentation technique (Cf. paragraphe précédent) est automatiquement générée avec chaque version de chaque Job : dès que vous sauvegardez un Job, la documentation est mise à jour et sauvegardée dans le Référentiel ; elle est par conséquent automatiquement partagée et disponible pour tous les utilisateurs.

AutoDoc+ permet en outre de personnaliser l'apparence graphique de cette documentation en ajoutant votre propre logo et le nom de votre entreprise, ou en modifiant les couleurs à l'aide d'une CSS personnalisée.

Implémentation

Job Designer

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Le Job Designer fournit une vue à la fois graphique et fonctionnelle des processus d'intégration à l'aide d'une palette graphique de composants et de connecteurs.

Les processus d'intégration sont construits via de simples opérations de glisser-déposer des composants et des connecteurs dans l'espace de modélisation graphique. La mise en place des connexions et relations entre ces composants s'effectue d'un simple clic droit et la configuration de leurs propriétés s'effectue via une vue dédiée, sur simple sélection d'un composant.

Les fonctionnalités du Job Designer permettent d'accéder, via une bibliothèque exhaustive de composants, à tous les types de sources et de cibles nécessaires à tous les processus d'intégration, de migration ou de synchronisation de données. Les composants et les connecteurs couvrent tous les types de tâches et d'opérations sur les données ainsi que sur la gestion de données ou l'orchestration du flux de données. Les connecteurs permettent l'accès et la lecture/écriture de tous les systèmes

source et cible au cours de l'intégration, la migration et la synchronisation de données. Les paramètres peuvent être configurés via une vue centralisée, sur simple sélection de chaque composant du Job, ou peuvent être hérités du Metadata Manager (référentiel).

Les composants complexes sont dotés d'interfaces graphiques dédiées et intuitives, ou intègrent des assistants permettant aux utilisateurs de construire leurs Jobs. Pour maintenir la lisibilité d'un Job, le diagramme du Job peut être divisé en sous-jobs, et configuré en Jobs pères et Jobs fils, afin de séquencer leur exécution. Les composants d'orchestration ainsi que différents types de relation permettent à l'utilisateur de séquencer l'exécution des processus. Une console intégrée permet aux utilisateurs de monitorer l'exécution, de vérifier et de suivre les performances directement dans le Studio.

Composants

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Talend fournit en natif des connecteurs open source techniques et métier permettant d'accéder à tous les types d'environnements ou systèmes. Ce large éventail de connecteurs est la clé d'une interopérabilité réussie des applications et des bases de données, permettant de faire communiquer des structures aussi diverses qu'hétérogènes à un niveau de performance inégalé. Cette bibliothèque de composants est alimentée en permanence, enrichissant de nouvelles fonctionnalités les solutions Talend pour l'intégration, la migration et la synchronisation de données.

Plus de 550 composants sont disponibles gratuitement dont 60% ont été créés et développés par la Communauté Talend.

Les connecteurs et les composants développés en externe peuvent être partagés via Talend Exchange (<http://talendforge.org/exchange/>). Un certain nombre des composants soumis par ce biais font l'objet d'une validation et d'une optimisation par l'équipe R&D de Talend, pour être ensuite intégrés et supportés en natif.

Consultez <http://www.talendforge.org/components> pour obtenir la liste exhaustive des connecteurs supportés à ce jour.

Support de l'ETL

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Le mode ETL (Extract, Transform & Load) est le mode par défaut utilisé par les solutions d'intégration de données de Talend. Il consiste à traiter les lignes les unes après les autres en mode flux. Ce mode est particulièrement adapté à un usage en environnement hétérogène et permet d'intégrer n'importe quelles technologies en source et en cible (Webservice, fichiers, bases de données, MOM, application métiers, etc.).

Le mode ETL peut être utilisé tant en batch qu'en temps réel. Les processus ETL peuvent être parallélisés pour accélérer, encore, leur exécution.

L'architecture unique de Talend s'affranchit de tout moteur d'exécution en générant des processus autonomes pouvant être déployés sur n'importe quel serveur (interne ou externe à l'entreprise). Ainsi, les processus ETL peuvent être exécutés au plus près de la donnée minimisant les temps d'accès, la consommation de bande passante et supprimant les goulots d'étranglements.

Dans un même Job, il est possible de combiner cette approche avec l'approche ELT (Cf. paragraphe suivant) pour obtenir le meilleur niveau de performance sans contrainte d'architecture.

Support de l'ELT

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Les solutions d'intégration de données de Talend supportent également le mode ELT (Extract, Load & Transform) qui consiste à traiter les données de manière ensembliste en utilisant le moteur de la base de données cible.

Ce mode est réservé à un usage en environnement homogène (une base de données). Il a l'avantage de tirer pleinement profit des ressources matérielles disponibles et est particulièrement conseillé dans la cadre de traitement de très grosses volumétries de données dans des environnements de type "Data Warehouse Appliance" comme Teradata, Netezza, etc.

Dans un même Job, il est possible de combiner cette approche avec l'approche ETL (Cf. paragraphe précédent) pour obtenir le meilleur niveau de performance sans contrainte d'architecture.

Version

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



La gestion des versions dans le Studio Talend s'effectue facilement grâce à la fonctionnalité native de versionnement manuel.

Une majeure et une mineure de version sont automatiquement définies lors de la création du Job et peuvent être facilement incrémentées dans le temps et au fur et à mesure des mises à jour, via l'onglet Version dédié, disponible directement dans la perspective Designer du Studio Talend.

Tous les éléments créés dans le Studio peuvent être versionnés : Business Models, Jobs, Routines, Métadonnées, Documentation...

Le versionnement fait partie intégrante des bonnes pratiques facilitant la réutilisation des éléments ainsi que le retour à une version antérieure d'un développement lorsque c'est nécessaire.

Référentiel partagé

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Le Référentiel partagé (ou Metadata Manager) est conçu pour consolider toutes les informations projet et les métadonnées d'entreprise dans un référentiel centralisé partagé par tous les intervenants des processus d'intégration.

Côté Studio, les utilisateurs ont accès aux projets selon leurs rôles et autorisations définies dans Talend Administration Center. Ce référentiel partagé permet à toutes les personnes prenant part à un projet d'intégration de travailler en mode collaboratif. Il permet de stocker et partager tous les éléments Talend : Business Models (modèles métier), Jobs (processus), Joblets, Routines, définitions des métadonnées (comme les connexions aux systèmes source/cible)...

Derrière le référentiel partagé se trouve un système standard de gestion de sources (Subversion) permettant le stockage et la gestion de toutes les versions de tous les éléments.

Un système de verrouillage automatique garantit que le Job en cours de création est bien interdit d'accès et que par conséquent aucun autre utilisateur ne peut modifier le même Job au même moment.

Depuis la version 4.0, Talend exploite tout le potentiel de gestion de versions de Subversion, permettant ainsi de gérer différentes branches, d'effectuer des check-in/check-out, de committer manuellement ou de façon automatique, de faire des commentaires...

Exploration des données

Cette fonctionnalité est incluse dans :

Open

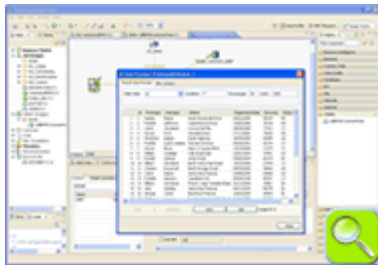
Team

Professional

RTx

Enterprise

MPx



Lorsque vous développez des Jobs avec Talend, vous pouvez avoir besoin de voir le contenu des différents systèmes cible ou source (fichiers, bases de données, etc.). Le Data Viewer vous permet d'explorer les systèmes source/cible quelle que soit l'application généralement utilisée pour son ouverture : Bloc-notes pour les fichiers .txt et .csv, outil de requête SQL pour les tables de bases de données, MS Excel pour les fichiers .xls, navigateur HTML, etc...
Il n'est plus nécessaire d'avoir plusieurs outils différents ni de parcourir les systèmes pour retrouver les données, le Data Viewer utilise les chemins d'accès aux sources/cibles définis dans les composants, pour accéder directement aux données.

de données, CSV...).

Le Data Viewer peut vous permettre de gagner beaucoup de temps, car il est directement accessible dans le Studio, via un simple clic-droit sur n'importe quel composant. C'est un moyen pratique de visualiser les données contenues dans vos systèmes source/cible lorsque vous développez vos processus d'intégration, et ce quel que soit le format de données (Excel, table de base

Assistants

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Schéma dynamique

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Les schémas dynamiques permettent de créer des jobs contenant des schémas dont la structure des colonnes, ainsi que leur nombre, est inconnue. Selon l'option choisie par le développeur, les colonnes dynamiques peuvent être directement mappées en sortie à l'aide du mode transfert (Pass-through).

Cette fonctionnalité est particulièrement utile lors de la mise en place de scénarii de réplication de base de données. Un développeur ayant besoin de migrer l'intégralité d'une base de données Oracle contenant une centaine de tables vers une base MySQL par exemple, peut le faire sans connaître la structure de toutes ces tables, et ce, avec un seul job !

Analyse d'impact

Cette fonctionnalité est incluse dans :

Open

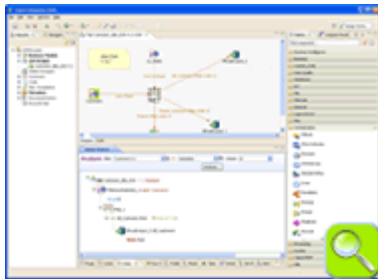
Team

Professional

RTx

Enterprise

MPx



La fonctionnalité d'analyse d'impact vous permet d'identifier les conséquences possibles d'une modification. Cette fonctionnalité est disponible dans le Metadata Manager. Vous pouvez effectuer une analyse d'impact sur n'importe quelle colonne de n'importe quelle métadonnée (base de données, fichier...). Le résultat de l'analyse d'impact s'affiche dans un rapport graphique et interactif à partir duquel vous pouvez suivre l'évolution d'une colonne et ainsi voir toutes les opérations qui ont été effectuées, de la source jusqu'à la cible à travers le Job.

Vous pouvez exporter ce rapport en format HTML.

Lignage de données

Cette fonctionnalité est incluse dans :

Open

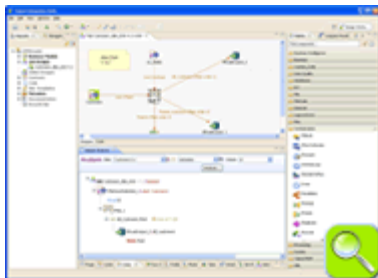
Team

Professional

RTx

Enterprise

MPx



La fonctionnalité de lignage de données vous permet d'identifier où un changement a été effectué.

Cette fonctionnalité est disponible dans le Metadata Manager et peut être effectuée sur n'importe quelle colonne de n'importe quelle métadonnée (base de données, fichier). Le résultat du lignage de données s'affiche dans un rapport dans lequel vous pouvez faire le suivi d'une modification à partir du composant final cible jusqu'à sa source dans le Job.

Vous pouvez exporter ce rapport en format HTML.

Comparaison de Jobs

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



La fonctionnalité de comparaison de Jobs permet d'identifier les différences entre deux versions d'un Job ou entre deux Jobs différents.

La comparaison des Jobs (Job Compare) est entièrement intégrée dans Talend Integration Suite Studio. Le résultat d'une comparaison de Jobs est fourni au sein d'un rapport graphique interactif en format HTML ou XML dans lequel les différences sont mises en évidence.

Dans cet exemple, le rapport de comparaison montre que le champ du séparateur (Delimiter) dans les propriétés du composant tFileInputDelimited n'est pas défini de la même façon dans les deux Jobs comparés : en version 3.2, le séparateur est une tabulation ("t"), alors qu'en version 4.2, le séparateur est un retour à la ligne ("n").

Joblets

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Les Joblets permettent de factoriser une partie d'un Job (ou sous-job) dans un composant Joblet. Sélectionnez simplement les composants formant la partie du Job que vous souhaitez réutiliser ou que vous voulez factoriser, puis cliquez sur l'élément du menu Factoriser en Joblet (« Refactor to Joblet »).

Le Job est automatiquement simplifié, puisque les composants sélectionnés sont réduits à un seul composant. Ce composant Joblet peut être partagé via le dossier Joblets dédié dans la Palette des composants et est facilement réutilisable dans tout autre Job.

Les Joblets simplifient considérablement la maintenance des Jobs complexes et redondants. De plus, un mécanisme d'« analyse d'impact » permet de trouver quels Jobs utilisent un Joblet défini.

Projets de référence

Cette fonctionnalité est incluse dans :

Open

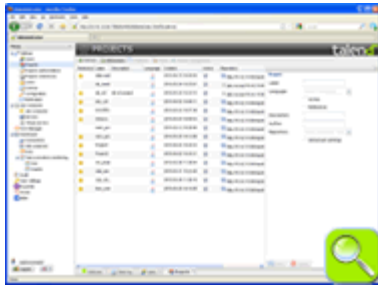
Team

Professional

RTx

Enterprise

MPx



Les projets de référence permettent d'éviter la duplication (copier-coller) d'éléments (Jobs, Routines, Documentation, Métadonnées...) entre les projets.

Les projets « esclaves » sont liés à un (ou plusieurs) projets « maîtres » par référence et héritent donc des éléments de ce/ces projet(s) père(s).

Les ressources provenant du projet maître apparaissent dans le projet esclave en lecture seule : elles sont disponibles à la réutilisation et à l'exécution mais ne peuvent pas être modifiées.

Parce qu'un lien fort est ainsi établi entre les projets maître et esclave, dès qu'une personne modifie un élément dans le projet maître, tous les projets esclaves sont mis à jour.

Les projets de référence partagent tous les éléments redondants d'un projet (Jobs, modèles, métadonnées) afin de les rendre disponibles aux autres projets. Cette fonctionnalité permet de tirer profit de ces éléments et de réutiliser les 30% d'éléments généralement communs à tous les projets d'intégration de données, réduisant ainsi considérablement la maintenance associée.

Capture de données modifiées

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



La mise en place d'un data warehouse comprend l'extraction et le transfert des données d'une ou plusieurs base(s) de données vers un ou plusieurs système(s) cible, pour analyse. Cela implique l'extraction et le transfert d'énormes volumes de données généralement très consommateurs de ressources et de temps.

La possibilité de ne capturer que les données modifiées en temps réel est connue sous le nom de Change Data Capture (CDC).

Ainsi la capture des modifications seulement réduit le trafic des données entre les systèmes et permet de réduire le temps d'ETL.

L'architecture CDC Talend est construite sur le modèle éditeur/souscripteur. L'éditeur capture les modifications des données et les rend disponibles aux souscripteurs (Jobs Talend). Les souscripteurs utilisent les modifications de données obtenues de l'éditeur.

Cette fonctionnalité détecte les enregistrements modifiés en temps réel, permettant aux données d'être immédiatement envoyées aux Jobs souscripteurs, réduisant ainsi le temps nécessaire au chargement et à la mise à jour des données durant les opérations d'ETL ou d'intégration de données opérationnelles.

La fonctionnalité de capture de données modifiées de Talend propose les modes les plus couramment utilisés : Trigger et Redo logs. Le mode disponible dépend du type de base de données traitée.

Règles métier

Cette fonctionnalité est incluse dans :

Open

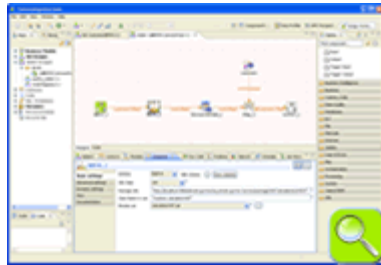
Team

Professional

RTx

Enterprise

MPx



Les règles métier (Business rules) sont généralement définies par les utilisateurs métier dans un cahier des charges qui doit être interprété et implémenté par l'équipe technique.

Talend Integration Suite embarque un moteur de règles métier qui permet aux utilisateurs de configurer leurs propres règles métier. Les utilisateurs peuvent ainsi définir des critères de segmentation de marché (par âge, par région...) et paramétrer leurs règles métier via une feuille de calcul Excel ou via l'interface Drools Guvnor accessible directement à partir de l'interface basée Web de Talend Administration Center.

Grâce à l'interface Drools Guvnor, les experts métier disposent d'un éditeur graphique leur permettant de créer et éditer rapidement et directement des règles, de contrôler l'accès aux règles et aux autres fonctionnalités, de gérer les versions et la modification des règles au fil du temps. Les règles peuvent être testées et appelées dans des Jobs développés (voir diapo suivante)

API

Cette fonctionnalité est incluse dans :

Open

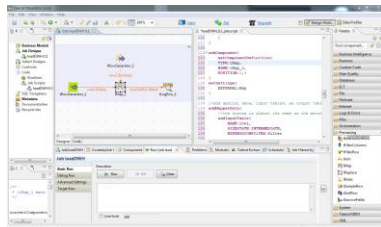
Team

Professional

RTx

Enterprise

MPx



Sophisticated but simple to grasp scripting language which enables programmatic job generation

Fully featured dedicated editor including syntax highlight, code completion aid & warning system

Automatic reverse engineering from an existing Job design to produce the corresponding script

Test

Gestion des contextes

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Les contextes permettent d'externaliser presque tous les paramètres de composants / Jobs. Les utilisateurs peuvent ainsi configurer les paramètres à la volée lors de l'exécution ou d'utiliser des paramètres différents pour le test et pour la production.

Les contextes peuvent être définis comme requis dans tous les environnements (Développement, Test, Production...) sans limite en termes de nombre de contextes créés.

Les utilisateurs peuvent changer de contexte à n'importe quel moment, que ce soit à la création ou à l'exécution d'un Job, afin d'utiliser les paramètres définis.

Les valeurs de paramètre peuvent également être modifiées via une boîte de dialogue à la conception comme lors des tests d'exécution d'un Job. En outre, un composant de chargement de paramètres dédié peut être utilisé pour remplacer dynamiquement toute valeur.

Exécution Distante

Cette fonctionnalité est incluse dans :

Open

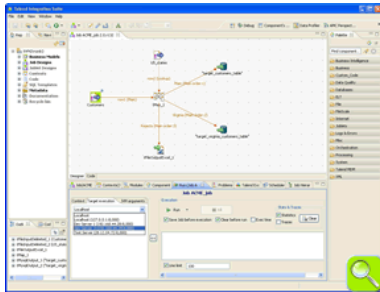
Team

Professional

RTx

Enterprise

MPx



La fonctionnalité d'exécution distante permet l'exécution de Jobs sur n'importe quel serveur distant, directement à partir du Studio.

Cela peut être extrêmement pratique lorsque vous avez besoin de tester des Jobs, par exemple :

- dans une configuration similaire à l'environnement de production
- sur différents systèmes d'exploitation
- sur demande dans des systèmes spécifiques

En effet, cette fonctionnalité évite d'avoir à passer par des procédures complexes de déploiement.

Le système cible peut être sélectionné dynamiquement à l'exécution, directement à partir du Studio. Les options de débogage, traces et statistiques temps réel restent disponibles dans ce mode d'exécution distante.

Déploiement

Talend Administration Center

Cette fonctionnalité est incluse dans :

Open

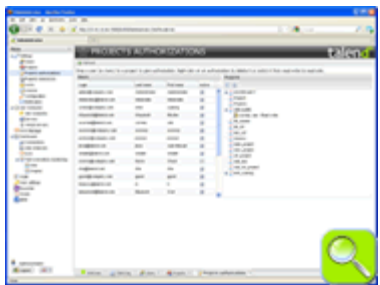
Team

Professional

RTx

Enterprise

MPx



Toutes les offres de souscription incluent au moins un Studio (ou plus selon le nombre d'utilisateurs) ainsi qu'un module logiciel qui peut être installé sur un serveur et administré via l'interface Web du Talend Administration Center.

Le ou les Studios ne sont donc plus en mode local mais connectés à distance aux projets définis dans Talend Administration Center.

Talend Administration Center est une application légère (dans un navigateur, aucun déploiement requis) permettant aux personnes en charge de la gestion des projets d'intégration d'administrer les utilisateurs, les projets, les droits utilisateurs, la licence...

Les autorisations d'accès aux projets peuvent facilement être assignées pour chacun des utilisateurs (support de l'annuaire LDAP). Et les utilisateurs accèdent aux projets selon leur rôle : aucun accès, lecture seule, lecture écriture...

Les utilisateurs peuvent partager des éléments du référentiel (Jobs, Business Models, Métadonnées de connexion aux bases de données...) avec d'autres utilisateurs, directement dans leur Studio, pour les projets dont ils ont les droits d'accès. Plus d'informations concernant le référentiel partagé ci-après.

Selon l'édition de Talend Integration Suite à laquelle vous avez souscrit, de nombreux plug-ins supplémentaires sont disponibles dans le panneau de navigation situé à gauche (Dashboard, SOA Manager, gestionnaire de serveur...).

Job Conductor

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Le Job Conductor coordonne l'exécution des Jobs d'intégration de données. Il fournit une interface centralisée d'exécution à partir de laquelle l'exécution des Jobs peut être lancée à la demande, ou programmée à une date définie (à partir de Team Edition), ou selon un événement paramétré (à partir de Professional Edition).

Le module Job Conductor repose sur des « JobServers » ou agents, qui sont des petites applications installées sur chacun des serveurs d'exécution.

Une fois les agents configurés, le Job Conductor vous permet de monitorer, en temps réel, toutes vos ressources matérielles (CPU disponible, RAM, HD...), permettant la répartition des exécutions de Jobs sur le grid, selon la disponibilité des serveurs. Le support natif de JMX vous permet de monitorer plus de 40 indicateurs. Chaque Job peut alors être déployé sur n'importe quel serveur en un seul clic !

Command Line

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Les processus d'intégration développés dans le Job Designer peuvent être déployés, mis à jour et exécutés à l'extérieur du Studio, à l'aide du module Command Line.

Le Command Line de Talend fournit un ensemble d'options en ligne de commande permettant aux développeurs et aux administrateurs d'effectuer facilement des opérations groupées.

Presque toutes les fonctions de gestion des Jobs disponibles dans le Studio Talend et dans Talend Administration Center sont également disponibles via le Command Line. Cela inclut par exemple les fonctions telles que : la mise à jour des propriétés des Jobs, la mise en production des projets, l'export/l'import de Jobs ou d'ensembles de Jobs, etc.

La fonctionnalité de Command Line permet de lancer facilement et rapidement de nombreux déploiements et exécutions de Jobs même complexes, tout en prenant en compte leurs dépendances et leurs métadonnées d'exécution.

La fonction native Help répertorie de façon exhaustive toutes les commandes disponibles en Command Line et fournit une courte description de ces commandes.

Time Scheduler (Ordonnanceur temporel)

Cette fonctionnalité est incluse dans :

Open

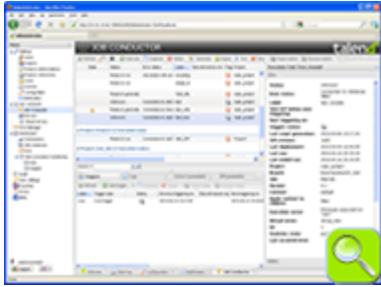
Team

Professional

RTx

Enterprise

MPx



La planification temporelle vous permet de lancer une exécution de Job à une date et à un moment précis (premier lundi du mois, chaque mardi...) ou de manière régulière, sur une période de temps. un système de « Tâches » est utilisé pour centraliser toutes les informations nécessaires à l'exécution du Job (nom du projet, nom du Job, version du Job, serveur...).

La tâche est déclenchée selon les paramètres de planification et le Job est ainsi déployé et exécuté automatiquement sur le serveur défini au moment défini. Un statut système très pratique vous permet de monitorer l'état du déclenchement et le succès/l'échec du lancement de l'exécution directement dans le Job Conductor.

A partir de l'édition Professional Edition, une fonction supplémentaire de planification sur événement/fichier est disponible. (Consultez la diapo Planification basée événements.).

Event Scheduler (Planification sur événements)

Cette fonctionnalité est incluse dans :

Open

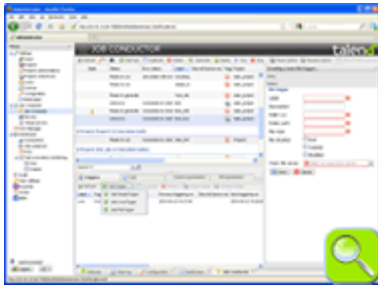
Team

Professional

RTx

Enterprise

MPx



L'Event Scheduler étend les fonctionnalités de planification temporelle à l'intégration en temps réel.

L'Event Listener permet de déclencher une exécution à la demande, ou bien selon un événement donné.

Les événements sur fichier incluent notamment : l'apparition, la disparition d'un fichier d'un emplacement donné, ou la modification de ce fichier. Les événements sur condition utilisent la condition SQL « wait for ». Une fois l'événement attendu identifié, la tâche d'exécution est déclenchée et le déploiement et le lancement du Job sont effectués.

Vous pouvez facilement ajouter de nouveaux déclencheurs d'événements à une tâche quelle qu'elle soit, afin d'industrialiser l'exécution automatique.

Plan d'exécution

Cette fonctionnalité est incluse dans :

Open

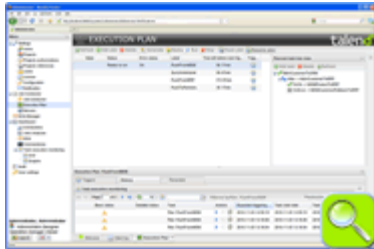
Team

Professional

RTx

Enterprise

MPx



La fonctionnalité Execution Plan (Plan d'exécution) vous permet de séquencer et d'orchestrer les différentes exécutions de Jobs ainsi que de faciliter la reprise sur erreur, directement à partir du Job Conductor. Le plan d'exécution est une fonctionnalité basée sur un système de tâches qui met en avant les dépendances entre les différentes tâches permettant d'orchestrer le séquençement de l'exécution.

Les dépendances des tâches sont définies via une vue hiérarchique composée de tâches principales et secondaires, dans laquelle chaque tâche peut avoir une tâche subordonnée.

A partir d'une vue unique dans le Job Conductor, les plans d'exécutions peuvent être planifiés, déclenchés et paramétrés pour utiliser les paramètres d'exécution (contexte) d'un environnement particulier.

SOA Manager / Services Web

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Le SOA Manager fournit un environnement entièrement graphique basé Web qui permet d'exposer un ou plusieurs Job(s) d'intégration de données en tant que services (Services Web). Leur déploiement automatique s'effectue de façon automatique au sein et à travers des applications et des systèmes hétérogènes, à l'aide de liaisons SOAP (RPC ou documents). Un assistant WSDL dédié vous permet de générer des descripteurs WSDL pour exposer des Jobs dans un environnement de Services Web et trouver des entrées UDDI correspondantes lors de l'utilisation de Services Web.

Le SOA Manager comprend une fonctionnalité avancée de gestion des requêtes entrantes basée sur un système optimisé de pool et de files d'attente. Le pool de services actifs défini par l'utilisateur traite un certain nombre de requêtes en temps réel, tandis qu'un gestionnaire de files d'attente prend en charge les requêtes supplémentaires, permettant de contrôler le débit des requêtes pour un traitement asynchrone.

Le SOA Manager permet d'exposer des Jobs créés à l'aide de Talend Integration Suite Studio dans un environnement orienté Services Web et d'assurer l'interopérabilité entre plusieurs applications et systèmes.

Répartition de charge (Load Balancing)

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Le module Grid Conductor (accessible via le Job Conductor) optimise l'extensibilité et la disponibilité des processus d'intégration en assurant un usage optimal de la grille d'exécution.



Le Grid Conductor repose sur la définition de serveurs virtuels qui regroupent les ressources disponibles, quel que soit le type de système (processeur, système d'exploitation, etc.).

Les tâches sont assignées aux serveurs virtuels du Grid Conductor plutôt qu'à un serveur d'exécution unique.

En assurant le suivi constant des ressources disponibles sur les serveurs d'exécution, le Grid Conductor garantit une exécution sans problème de tous les Jobs au moment du déclenchement, et optimise ainsi l'exploitation des ressources disponibles, en supprimant les goulets d'étranglement créés par l'approche traditionnelle de serveur unique.

Le Grid conductor permet de réduire les problèmes liés à l'utilisation concurrente des ressources lorsqu'un grand nombre de Jobs s'exécutent simultanément, ou lorsque des serveurs non dédiés sont utilisés. Le Grid Conductor permet également le fail-over d'un Job dans l'éventualité où une ressource n'est plus disponible.

Haute disponibilité

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



La Haute disponibilité est disponible grâce à la possibilité de déploiement multiple de Job Conductors et de serveurs d'exécution.

En outre, la mise en cluster des bases de données assure également le fail over et évite les interruptions d'exécution.

Failover

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

FileScale

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx

Benchmark Results

Configuration

- 2 CPU Xeon 5520 Quad Core / 2,26GHz / 24 GB RAM
- Sun StorageTek 2540 Fiber Channel: 12 HD SAS 136Gb

Results

	1	10	100	300	550
# rows	6 million	60 million	600 million	1.8 billion	3.3 billion
data volume	740 MB	7.4 GB	74 GB	225 GB	415 GB
sorting time	6 sec	1 min 02 sec	21 min 30 sec	2 hr 01 min	5 hr 17 min

Independent benchmark made by Sun Microsystems

Talend Integration Suite MPx met à profit les architectures multi-serveurs, multi-processeurs, et multi-core, dans lesquelles le code et les sous-processus séparés peuvent être exécutés en parallèle afin d'exploiter au maximum l'architecture. Cette fonctionnalité de parallélisation massive optimise l'utilisation des capacités des serveurs d'entreprise ainsi que du nombre de processeurs disponibles, améliorant considérablement le temps de traitement.

Talend Integration Suite MPx repose notamment sur la technologie unique **FileScale**, qui exploite au mieux l'architecture matérielle du serveur d'exécution et utilise au maximum les performances des algorithmes de tri de bas niveau. La technologie FileScale fonctionne en mode bulk sur des fichiers (très) volumineux. Elle tire pleinement profit de l'architecture d'exécution, puisqu'elle n'est pas restreinte par la JVM ou par les limitations du moteur d'exécution, typiques des architectures d'intégration de données traditionnelles.

La technologie FileScale trie et transforme des données à l'aide d'algorithmes de traitement de données mathématiques innovants et très performants. Elle utilise l'architecture MapReduce pour décomposer n'importe quel processus traitement des données en de multiples processus granulaires, atteignant ainsi des niveaux de performance inégalés. Consultez le workbench de Sun Microsystems : http://blogs.sun.com/aja/entry/talend_s_new_data_processing.

Hadoop

Cette fonctionnalité est incluse dans :

Open

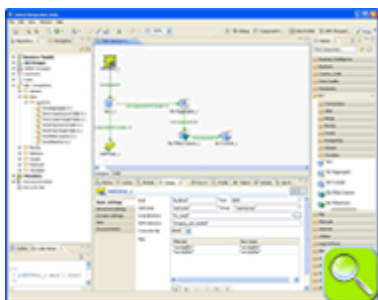
Team

Professional

RTx

Enterprise

MPx



Hadoop est un framework logiciel open source Apache supportant la gestion de très grands volumes de données de façon distribuée. Il s'inspire de l'architecture MapReduce et permet aux applications de gérer des milliers de nœuds et des pétaoctets de données, à l'aide d'un large griddes serveurs peu coûteux. Le support natif d'Hadoop dans Talend Integration Suite MPx permet la gestion évolutive à tous les niveaux de volumes et quelle que soit la complexité des données. Les entreprises peuvent ainsi tirer profit de leurs clusters Hadoop pour les pics de volumes de données et leurs transformations complexes.

Un ensemble de composants dédiés accessible depuis la Palette des composants permet de lire et d'écrire dans les systèmes HDFS et Hive, et comprend des fonctionnalités d'ELT et des modèles SQL.

Monitoring

AMC

Cette fonctionnalité est incluse dans :

Open

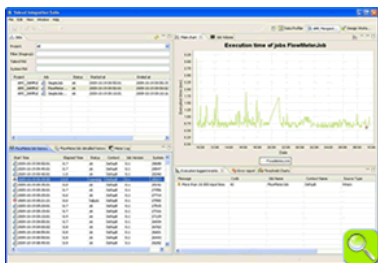
Team

Professional

RTx

Enterprise

MPx



Talend Activity Monitoring Console est une interface graphique et un outil de contrôle centralisé.

Talend Activity Monitoring Console fournit des fonctionnalités de contrôle pouvant être utilisées pour consolider les informations de log collectées, comprendre les interactions de Jobs sous-jacentes, empêcher la génération non désirée d'erreurs, et supporter les décisions de gestion du système.

Talend Activity Monitoring Console fournit une console unique et centralisée de surveillance des événements liés aux Jobs (succès, échecs, avertissements, etc.), des temps d'exécution et des volumes de données.

Cet outil est disponible en standalone ou peut être entièrement intégré dans le Studio.

Dashboard

Cette fonctionnalité est incluse dans :

Open

Team

Professional

RTx

Enterprise

MPx



Le Dashboard est une version Web de Talend Activity Monitoring Console facilement accessible via un navigateur.

Le Dashboard fournit des graphiques de performance d'exécution et des indicateurs d'état, permettant ainsi aux utilisateurs y ayant accès de consulter à la fois l'état courant et l'historique d'exécution de n'importe quel processus d'intégration.

Il fournit également des fonctionnalités de monitoring permettant de consolider les informations de log collectées, de comprendre les interactions sous-jacentes des composants et des Jobs, d'obtenir rapidement les informations d'exécution des tâches, d'empêcher la survenue d'erreurs inattendues et de fournir un support décisionnel dans la gestion système.

Reprise sur erreurs

Cette fonctionnalité est incluse dans :

Open

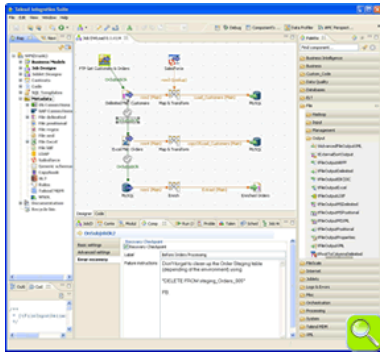
Team

Professional

RTx

Enterprise

MPx



Les processus d'exécution de Jobs ainsi que les opérations de sauvegarde et de restauration prennent généralement beaucoup de temps.

Talend Integration Suite Studio permet de mettre en place des points de validation en cas de reprise sur erreur, qui peuvent être configurés dès la conception du Job.

En cas d'erreur, les processus peuvent être redémarrés à partir de l'un des points de validation. Les développeurs de Jobs peuvent également concevoir et intégrer une gestion d'erreurs correspondant à des conditions spécifiques à l'aide de la fonction « on-failure » des points de validation.

Les points de reprise sur erreur peuvent être initialisés à des intervalles spécifiques du flux de données (connexions Trigger), le but étant de réduire le temps et les efforts nécessaires lorsqu'un processus d'exécution d'un Job doit être redémarré, à la suite d'un échec.

Grâce à la fonctionnalité de points de reprise, le process peut être redémarré à partir du dernier point de validation précédant l'échec (ou tout autre point de validation avant que l'échec ne se produise), plutôt que de recommencer entièrement le processus d'exécution du Job.