



Talend Data Quality

La première solution open source de qualité de données combinant profiling, standardisation, nettoyage, mise en correspondance des données et monitoring de la qualité des données au sein d'une plateforme unifiée.

La qualité de données ne se limite pas seulement à aider les entreprises à charger des données correctes dans leur système d'information ; elle permet également de se débarrasser des données erronées, corrompues ou dupliquées. Le nettoyage des données est une étape essentielle dans l'intégration d'information dans les systèmes car les mauvaises informations prolifèrent très vite, en interne mais aussi aux partenaires commerciaux.

Talend Data Quality est un environnement graphique de gestion de la qualité des données qui inclut toutes les fonctions de qualité de données, dont le profiling, la standardisation/nettoyage, la correspondance, et un monitoring des métriques de qualité de données via des dashboards. En associant cet ensemble complet de fonctionnalités de qualité de données, d'intégration de données et la plateforme MDM, vous pourrez accéder, nettoyer et gérer vos données au sein d'un environnement de développement unique.

Talend Data Quality est développé par Talend, le leader des logiciels open source de gestion de données. Son éventail complet de fonctionnalités répond à toutes les problématiques de qualité de données :

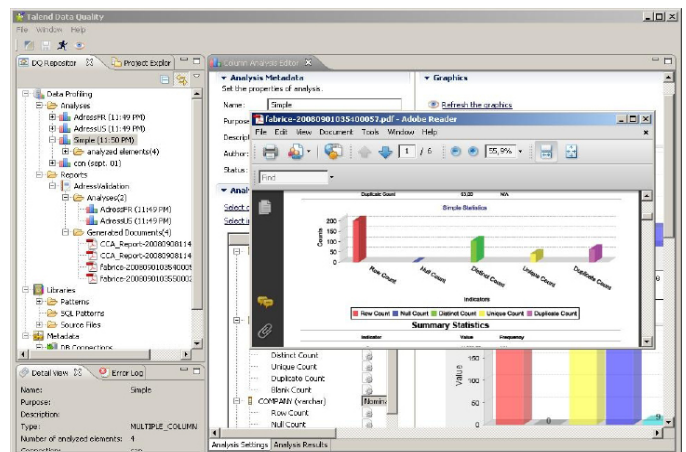
- **Le profiling des données** permet d'évaluer le niveau courant de la qualité des données et d'en mesurer l'évolution dans le temps. Talend Data Quality permet de générer de nombreux rapports d'analyse et d'analyser en profondeur les tables des bases de données ou les fichiers pour résoudre les problèmes immédiatement.
- **Le nettoyage de données et la standardisation** corrigent les données incomplètes ou hétérogènes grâce à de nombreuses fonctionnalités de qualité de données. Ils remettent en forme et normalisent les données pour garantir à vos applications métier une qualité de données irréprochable.
- **L'enrichissement de données** utilise des données de référence commerciales ou libres pour améliorer et compléter vos données, et vous permettre d'atteindre vos cibles commerciales.
- **Le rapprochement de données** fournit des algorithmes puissants et une interface intuitive pour trouver les correspondances partielles dans vos données. La console d'arbitrage de données unique permet de résoudre rapidement les correspondances qui tombent légèrement hors des standards.
- **Le monitoring des données** avec Data Quality Portal donne accès aux métriques clés de qualité de données via une console basée Web, simple d'utilisation. Ce portail permet de sensibiliser une plus large audience à la question de la qualité de données, favorisant l'amélioration de la qualité des données au sein de l'entreprise.

Étant donné que tous les produits Talend font partie d'une même plateforme unifiée, les fonctions de qualité de données sont entièrement intégrées à Talend Integration Suite et Talend

MDM, offrant aux utilisateurs une ergonomie homogène, une prise en main rapide et un niveau de réutilisation élevé.

Profiling de données

La première étape d'un projet de qualité des données d'entreprise est de « profiler » ces données et les évaluer. Le data profiler est une application sophistiquée mais simple d'utilisation qui n'exige pas de connaissance particulière des moteurs de bases de données ou de la structure de fichiers. Les utilisateurs métier ou les équipes en charge de la gestion des données peuvent ainsi effectuer toutes sortes d'analyses sur ces données.



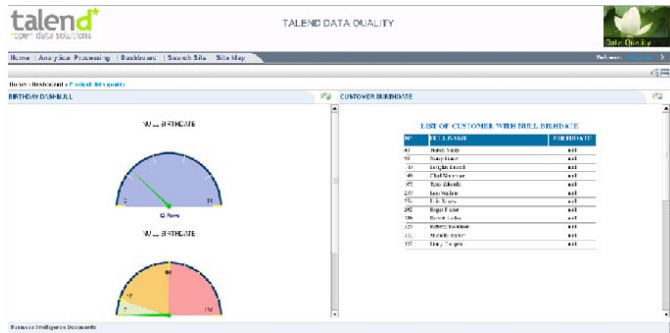
Ces analyses peuvent porter sur des statistiques simples telles que le nombre de lignes, le nombre de valeurs nulles ou les doublons, mais des analyses plus avancées sont disponibles telles que la recherche basée sur des modèles, la fréquence de soundex, l'analyse des chaînes de textes ou numériques, la distribution des enregistrements et l'analyse des « clés étrangères ». Ces analyses sont personnalisables pour s'adapter à vos règles de qualité de données. En vérifiant régulièrement ces métriques et en observant leur tendance et leur évolution, une entreprise peut constater l'amélioration ou la dégradation de la qualité de ses données. Talend Data Quality inclut également d'autres fonctionnalités de profiling et reporting :

- Historisation des analyses
- Analyse en mode batch
- Personnalisation des feuilles de style des rapports
- Formats de rapport : PDF, HTML et XML

Nettoyage des données

Une fois les problèmes identifiés, les données doivent être corrigées. Talend Data Quality dispose d'outils puissants pour redresser et réparer toutes les données non conformes à vos normes. Elle vous permet d'utiliser des données de référence

pour paramétrer les normes de valeur, des expressions régulières pour les normes de format et de taille. Vous pouvez paramétrer des processus de nettoyage grâce à une gamme exhaustive de composants d'intégration et de qualité de données. Par exemple, il est possible de normaliser les pseudonymes et adresses en utilisant les composants dédiés.



Enrichissement de données

L'enrichissement de données permet de compléter les données manquantes de vos données pour atteindre vos objectifs de qualité. La variété de ces informations est quasi illimitée, cela peut par exemple être des informations sur une société, les points de fidélité d'un consommateur, la longitude et la latitude d'une adresse pour optimiser les itinéraires de livraison ou la collecte de données de recensement pour définir des catégories démographiques ou socioprofessionnelles. Tirez avantage des données de référence fournies par les services postaux, les administrations et les fournisseurs de données libres du monde entier. L'environnement de développement intuitif permet de développer dans un environnement unique des processus homogènes pour consolider, fusionner ou simplement insérer des données dans tous les systèmes cibles.

Rapprochement des données

Les composants utilisent des algorithmes de correspondance qui permettent de trouver et réparer les doublons et quasi doublons contenus dans vos données. Contrairement à d'autres solutions de qualité de données, le rapprochement se configure directement dans le composant Talend. Pas de manipulation de fichiers de règles complexes ou d'édition fastidieuse de fichiers textes souvent associée aux outils de qualité de données commerciales. Il s'agit d'un ensemble de fonctionnalités de rapprochement robuste et facile d'utilisation.

Arbitrage des données

Les algorithmes de rapprochement sont des outils puissants, mais ils ne peuvent pas toujours déterminer avec certitude si deux enregistrements correspondent parfaitement. Une intervention humaine est parfois nécessaire. Talend Data Stewardship Console, disponible en tant que module optionnel pour Talend Data Quality, permet une médiation humaine des correspondances incertaines au sein d'un environnement point-and-click. Les problèmes de correspondance des données sont ainsi facilement traités et rapidement résolus pour atteindre la « golden data » que vous pouvez utiliser pour créer un hub de données maître ou que vous pouvez intégrer à toute application MDM.

Monitoring de la qualité de données et Reporting

Data Quality Portal fournit des indicateurs clé de qualité (KQI – key quality indicators) via un portail Web dédié sur lequel les équipes peuvent collaborer à l'amélioration de la qualité des données à travers toute l'entreprise. Ce portail inclut

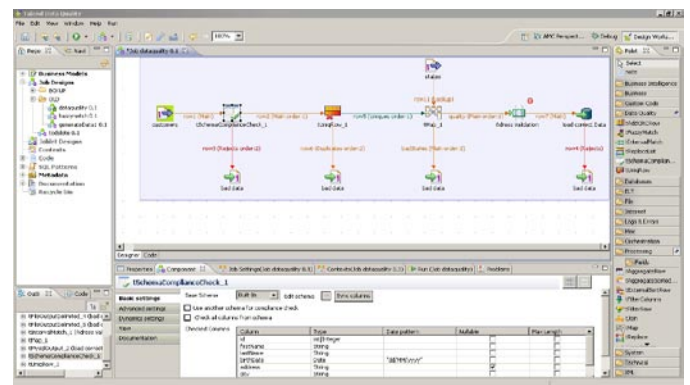
notamment des fonctions de reporting PDF, de dashboards personnalisables, de requête Ad-hoc et de monitoring des KQI dans le temps. Il fournit également un accès à un ensemble prédéfini de rapports et de jauges de qualité globale qui signalent les violations des seuils de qualité.

Plateforme unifiée

Étant donné que tous les produits Talend font partie d'une même plateforme unifiée, les fonctions de qualité de données sont entièrement intégrées à Talend Integration Suite et Talend MDM, offrant aux utilisateurs une ergonomie homogène, une prise en main rapide et un niveau de réutilisation élevé. Cette solution apporte des améliorations sans précédent en termes d'optimisation des ressources, de simplicité d'utilisation et d'homogénéité des projets. Talend Data Quality et les autres produits de gestion de données de Talend partagent un référentiel de métadonnées commun. Cela améliore le partage des informations vitales telles que les données utilisateurs, les métadonnées d'applications, les modèles métier, les règles métier, les règles de transformation et validation, les connecteurs, la validation des données et les workflows

Interface graphique et conception des processus

Talend Data Quality, offrant à la fois une vue graphique et fonctionnelle des processus d'intégration, il permet une approche accessible et non technique des données. Les systèmes, connexions, étapes et pré-requis d'un workflow sont représentés par des symboles graphiques intuitifs. Les intervenants fonctionnels peuvent prendre part à la création des processus de données et suivre les développements. L'environnement de développement unifié, commun à toutes les solutions de gestion de données Talend, offre des performances sans précédent en termes d'optimisation des ressources, de simplicité d'utilisation et d'homogénéité des projets.



Processus versatiles et extensibles

Talend Data Quality fournit en natif des connecteurs techniques et métier pour tous les environnements. Ces connecteurs, enrichis en permanence, permettent un accès performant aux structures les plus diverses et hétérogènes. En outre, Talend Data Quality utilise la même interface de développement que Talend Integration Suite. Cette interface permet notamment aux membres de la communauté de développer leurs propres composants de qualité de données pour se connecter, par exemple, à des sources de données métier spécifiques ou pour incorporer des routines personnalisées de nettoyage des données. Tous les composants personnalisés peuvent être partagés et utilisés par la communauté via la plateforme Talend Exchange.

Plus d'informations : www.talend.com | info@talend.com