

GUIDE COMPLET SUR

L'INTÉGRATION DES DONNÉES



GUIDE COMPLET SUR L'INTÉGRATION DES DONNÉES INTRODUCTION

À l'ère de l'économie de l'information, les données sont devenues les ressources les plus précieuses des entreprises. Quels que soient le secteur ou l'entreprise, les stratégies orientées données constituent un impératif concurrentiel. Pour atteindre leurs objectifs de croissance, de rentabilité et de satisfaction client par exemple, les entreprises s'appuient de plus en plus sur les données pour prendre leurs décisions.

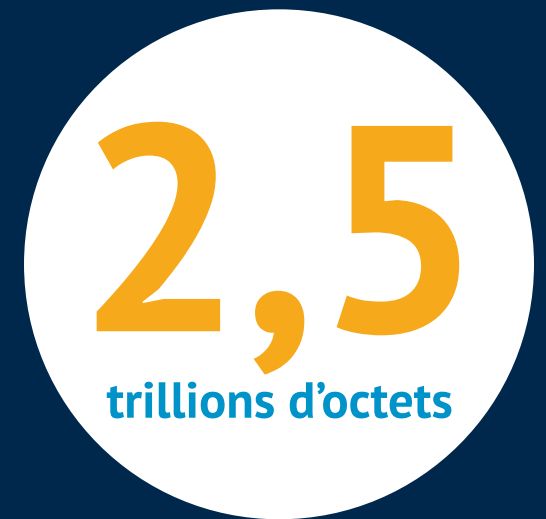
En effet, un [rapport de McKinsey Global Institute](#) indique que les entreprises qui sont vraiment pilotées par les données (qui peuvent collecter, traiter et analyser les données de façon large et intensive) sont mieux armées pour prendre de bonnes décisions. Selon le rapport, ces entreprises :

- Dégagent 93 % de bénéfices en plus ;
- Enregistrent des ventes 82 % plus élevées ;
- Développent leurs ventes 112 % plus rapidement.

Mais, pour disposer des données nécessaires à leur transformation numérique, les entreprises doivent concilier deux paramètres essentiels.

Caractérisée par la rapidité et l'accélération de la commercialisation pour répondre en temps réel aux demandes des équipes commerciales ou offrir des expériences personnalisées aux clients, la transformation numérique exige des données disponibles en temps voulu.

Néanmoins, si la rapidité est essentielle, elle ne fait pas tout. Vous avez besoin de données fiables pour prendre efficacement des décisions et offrir des expériences client exceptionnelles. Et c'est bien cette garantie qui permet aux entreprises de conserver la confiance de leurs clients dans le respect de la réglementation et d'assurer que les bonnes personnes ont accès aux données pertinentes pour prendre des décisions judicieuses. Ceci représente un défi majeur pour les entreprises. Selon la Harvard Business Review, [47 % des enregistrements de données](#), en moyenne, contiennent des erreurs critiques, qui ont des répercussions sur l'activité de l'entreprise.



**ont créés chaque jour
et le volume de données
double chaque année.**

De nos jours, les entreprises ont de nombreuses opportunités de devenir leader du digital. Mais, si on le leur demandait, la plupart admettraient que les nouvelles technologies (cloud, machine learning, Internet des objets et big data, p. ex.) n'ont pas tenu leurs promesses.

La réalité est bien différente :

- 55 % des données d'une entreprise ne peuvent pas être utilisées pour la prise de décision.
- Seuls 45 % des données structurées d'une entreprise sont utilisés de façon active pour la business intelligence et moins d'1 % des données non structurées est analysé ou utilisé.
- Plus de 70 % des salariés ont accès à des données qui devraient leur être interdites.
- Les analystes consacrent 80 % de leur temps à la découverte et préparation des données.
- 47 % des enregistrements de données dont la création est récente comportent au moins 1 erreur critique.
- La mauvaise qualité des données a eu en moyenne un impact financier estimé à 15 millions de dollars par an.
- Les spécialistes de l'information perdent 50 % de leur temps à partir à la chasse aux données, à trouver les erreurs et les corriger et à rechercher des informations pour fiabiliser les données.
- Les data scientists passent 60 % de leur temps à nettoyer et organiser leurs données.

55 %

des données d'une entreprise ne peuvent pas être utilisées pour la prise de décision.

L'écosystème des données prend de l'importance et se complexifie jour après jour en raison de la multiplication des sources de données.

La demande croissante en données toujours plus nombreuses à laquelle sont confrontées les entreprises pèse lourdement sur leur département informatique.

L'INTÉGRATION DES DONNÉES ÉRIGÉE EN STRATÉGIE MÉTIER

L'intégration des données constitue un pilier essentiel de toute stratégie métier et une condition sine qua non de l'obtention de données précises, fiables et exhaustives au rythme nécessaire pour que l'entreprise soit compétitive et puisse se développer. Elle est généralement considérée comme un simple processus technique : les données sont rendues facilement accessibles à ceux qui en ont besoin. Mais, comme de nombreux spécialistes des données ont pu le constater, l'intégration peut facilement devenir le principal frein à l'obtention d'informations.

Gartner estime qu'à l'horizon 2020, **l'intégration représentera 50 % du temps et du coût de la construction d'une plateforme numérique.**

Mais la situation pourrait être tout autre si nous envisagions l'intégration des données sur un plan stratégique et du point de vue de l'entreprise et non pas comme un ensemble de processus techniques. Les entreprises gagnantes sont passées maîtres dans l'art de rendre leur infrastructure de données agile, disponible et facile à utiliser et à mettre à jour. Dans ce guide de référence, nous nous pencherons plus en détail sur les évolutions et défis liés à l'intégration des données et proposerons des mesures concrètes pour élaborer votre stratégie d'intégration des données. Nous aborderons également le choix des outils d'intégration des données et la façon d'organiser votre équipe afin de tirer pleinement profit de vos investissements en matière de données. Nous examinerons comment retirer rapidement une véritable plus-value de vos données, quels que soient la technologie, les plateformes ou les systèmes que vous utilisez.

Il est temps que les données tiennent leur promesse. Et, grâce à une approche stratégique globale de l'intégration des données, c'est enfin possible.

TABLE DES MATIÈRES

CHAPITRE 1 : LA RÉVOLUTION DES DONNÉES EST EN MARCHÉ : QU'EST-CE QUI VOUS RETIENT ?

CHAPITRE 2 : LA DÉPENDANCE CROISSANTE ENVERS LES DONNÉES CONTRAINT LES TECHNOLOGIES DE L'INFORMATION À S'ADAPTER

CHAPITRE 3 : QU'EST-CE QUE L'INTÉGRATION DES DONNÉES ET POURQUOI EST-ELLE IMPORTANTE ?

CHAPITRE 4 : L'INTÉGRATION CLOUD DES DONNÉES : UNE PASSERELLE VERS DES GAINS D'EFFICACITÉ ET DE NOUVELLES INNOVATIONS

CHAPITRE 5 : COMMENT CHOISIR LA STRATÉGIE D'INTÉGRATION DES DONNÉES LA PLUS ADAPTÉE À VOTRE ENTREPRISE

CHAPITRE 6 : LA GRANDE QUESTION : CODAGE MANUEL OU OUTIL D'INTÉGRATION DES DONNÉES ?

CHAPITRE 7 : COMMENT DEVENIR UNE ENTREPRISE ORIENTÉE DONNÉES MODIFIE L'ORGANISATION DE VOTRE ÉQUIPE DE DONNÉES

CHAPITRE 8 : ÉLABOREZ VOTRE STRATÉGIE D'INTÉGRATION DES DONNÉES AVEC DES FONCTIONNALITÉS DE GESTION DES DONNÉES UNIFIÉES

CHAPITRE 9 : API, INTÉGRATION D'APPLICATIONS ET DONNÉES EN TANT QUE SERVICE (DaaS)

CHAPITRE 10 : ÉTUDES DE CAS

CHAPITRE 11 : LA CHECKLIST DE L'INTÉGRATION DES DONNÉES

CHAPITRE 12 : QUE RETENIR ?

CHAPITRE 1

LA RÉVOLUTION DES DONNÉES
EST EN MARCHÉ :
QU'EST-CE QUI VOUS RETIENT ?



LA RÉVOLUTION DES DONNÉES EST EN MARCHÉ – QU'EST-CE QUI VOUS RETIENT ?

Le jour où un scientifique IBM a inventé la base de données relationnelle (en 1970) a bouleversé la façon dont nous utilisons les données. Pour la première fois, les données sont devenues facilement accessibles pour les utilisateurs métiers. Les entreprises ont commencé à exploiter la puissance des données pour prendre des décisions et accélérer leur croissance. Aujourd'hui, toutes les grandes entreprises ont une chose en commun : elles sont pilotées par les données.

Si nous les utilisons à bon escient, les données ont le pouvoir de transformer tout ce que nous faisons dans n'importe quel secteur, que ce soit la finance, la grande distribution ou la santé. Les entreprises gagnantes utilisent leurs données de façon à améliorer l'expérience client et leur logistique et pour en tirer des informations qui les aideront à prendre de meilleures décisions. Mais, dans l'immédiat, nous sommes à un tournant décisif.

Les défis de la révolution des données

Nous sommes actuellement en pleine zone de turbulences concernant les données. Les volumes de données sont en augmentation en raison du coût incroyablement bas des capteurs, de la mise en réseau généralisée, du traitement à bas coût dans le cloud et des ressources informatiques dynamiques. Nous pourrions nous sentir dépassés face à cette accumulation de données mais le machine learning et l'informatique cognitive nous permettent un traitement des données d'une ampleur sans précédent et de trouver des corrélations ne pouvant pas être établies par la matière grise. Les possibilités semblent illimitées quand on envisage d'utiliser les données de manière transformatrice. C'est pourquoi il est impératif, pour les dirigeants d'entreprise, d'exploiter ces données.

En théorie, compte tenu de la technologie, chaque entreprise devrait opter pour le pilotage par les données. Dans la réalité, toutefois, le chemin est pavé d'obstacles :

Prisonniers du cycle traditionnel et d'un budget fixe

La crise des données alimente une demande qui dépasse largement les capacités de la plupart des départements informatiques.

Les budgets sont plafonnés et **n'augmentent que de 3,2% par an**, les entreprises étant souvent confrontées à un choix limité de technologies et de fournisseurs.

De nombreuses équipes IT consacrent toujours la majeure partie de leur budget à tenter de payer les frais courants. Avec le reste, elles essaient de se moderniser et d'innover. Malheureusement, quelques années plus tard, toute cette modernité est devenue archaïque et le cycle se répète. C'est par cette période de souffrance que nous avons dû passer au cours de ces 20 dernières années.

Manque de qualité et d'accessibilité des données

La plupart des données d'entreprise sont de mauvaise qualité : elles sont erronées, incohérentes et inaccessibles, ce qui empêche les entreprises d'en tirer parti. Une étude menée par Harvard Business Review a montré que seuls **3 % des données étudiées** étaient de qualité « acceptable ». C'est pourquoi les **analystes passent 80 % de leur temps à préparer les données** au lieu de faire le travail pour lequel ils sont payés : l'analytique. Si nous sommes incapables de garantir la qualité des données, sans parler d'accéder à celles dont nous avons besoin, comment sommes-nous censés exploiter leur valeur ?

Une étude menée par Harvard Business Review a montré que seuls 3 % des données étudiées étaient de qualité « acceptable ».

Des données de plus en plus exposées aux menaces

La formidable puissance des données va de pair avec la menace liée à leur exploitation. Le piratage et les failles de sécurité sont en hausse ; le coût mondial des répercussions de la cybercriminalité devrait atteindre **6 000 milliards de dollars** d'ici 2021, soit deux fois plus qu'en 2015. Face aux menaces croissantes, les réglementations en matière de sécurité et de confidentialité se multiplient. Compte tenu des problèmes liés à l'intégrité des données, les responsables veulent savoir si leurs données sont exactes et en sécurité. Comment la sécurité de leurs données peut être assurée ?

Demeurer à la pointe de l'innovation

L'univers de la gestion et du traitement des données se réinvente de fond en comble, tandis que les besoins en analytique évoluent. Des mutations importantes telles que l'essor du cloud bouleversent vos systèmes clés, vos processus d'intégration et vos stratégies de gestion. Et ce n'est pas près de s'arrêter : de nouvelles technologies qui, pour certaines, pourraient jouer un rôle déterminant dans votre entreprise, voient le jour. Votre entreprise doit être prête à tirer pleinement parti d'innovations qui lui permettent d'être pilotée par les données et choisir la pile technologique la plus à même de libérer ses données aujourd'hui, mais aussi demain et au cours des prochaines années.

Le coût mondial des répercussions de la cybercriminalité devrait atteindre 6 000 milliards de dollars d'ici 2021, ce qui la rend plus rentable que la totalité du commerce mondial de drogues.

CHAPITRE 2

LA DÉPENDANCE CROISSANTE
ENVERS LES DONNÉES
CONTRAIT LES TECHNOLOGIES
DE L'INFORMATION À S'ADAPTER



LA DÉPENDANCE CROISSANTE ENVERS LES DONNÉES CONTRAIT LES ÉQUIPES INFORMATIQUES À S'ADAPTER

Les données ne cessent de croître à un rythme étourdissant tant en volume, qu'en diversité et vitesse. Mais comment les entreprises vont-elles pouvoir toutes les gérer ? Et garantir un accès rapide à des données fiables ? Bien souvent, cette tâche incombe à l'équipe IT qui doit créer des solutions pour que les données soient accessibles au reste de l'entreprise.

Les départements informatiques doivent opérer une transformation profonde pour pouvoir faire face à l'explosion des données. Il y a cinq ans, le cas d'usage du big data le plus courant était le « transfert ETL » ou l'« optimisation du data warehouse » mais, désormais, de plus en plus de projets exigent de :

- Mettre l'accent sur la vitesse, diversité et véracité des données : p. ex., passer en temps réel, ingérer des données de streaming, prendre en charge davantage de sources de données dans le cloud et garantir la fiabilité des données ;
- Découvrir de nouvelles informations grâce à une utilisation avancée du big data (machine learning et intelligence artificielle) et savoir comment automatiser ce processus ;
- Être compatible avec différentes plateformes cloud utilisant des conteneurs et une informatique sans serveur.

« Dans ce monde en transformation accélérée, comment puis-je m'assurer que je ne suis pas tributaire d'une seule technologie du fait que des [innovations techniques] fleurissent partout dans le monde ? »

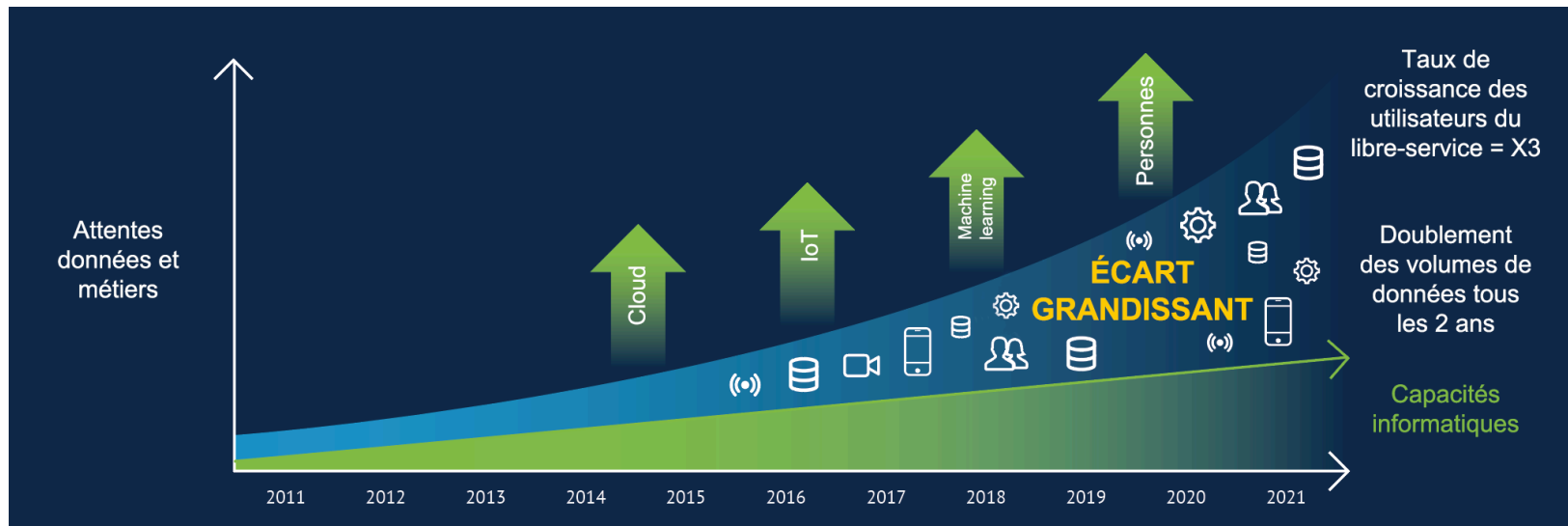
Comment puis-je, à l'avenir, mobiliser toutes ces innovations à l'échelle de l'entreprise ? »

– Mike Tuchen, PDG de Talend

INTÉGRATION DES DONNÉES LA DÉPENDANCE CROISSANTE ENVERS LES DONNÉES CONTRAINT LES ÉQUIPES INFORMATIQUES À S'ADAPTER **CHAPITRE 2**

Les utilisateurs de données se sont multipliés très rapidement, les entreprises s'étant mises à reconnaître l'utilité de leurs données. Chacun veut un accès aux données afin d'obtenir de meilleures informations. De ce fait, de nouveaux rôles (data scientists, ingénieurs de données, data stewards) sont apparus afin d'analyser et de gérer ces données. Selon un rapport récent établi par Gartner, à l'horizon 2020, **le nombre de spécialistes des données et de l'analytique dans les différents départements des entreprises grossira trois fois plus vite que celui des spécialistes des départements informatiques.** En outre, une fois qu'un département découvre l'intérêt d'utiliser l'analytique pour prendre des décisions éclairées, d'autres équipes souhaitent obtenir des informations plus nombreuses et approfondies pour tous les aspects de l'entreprise. Cela signifie que le département informatique doit fournir à tous ces spécialistes dans l'ensemble de l'entreprise un accès à des données provenant de systèmes disparates et rassembler les sources de données de manière accessible.

Tel est le défi. Les départements informatiques ne peuvent plus faire face aux volumes et types de données désormais disponibles ou au nombre d'utilisateurs demandant à y avoir accès. Pour tirer parti de toutes ces données, les entreprises ont investi plus lourdement chaque année dans des solutions logicielles, leurs infrastructures et leurs équipes IT. Toutefois, une simple augmentation du budget et des ressources informatiques ne constitue pas une stratégie viable, en particulier lorsque les volumes de données continuent à croître et que de plus en plus d'utilisateurs demandent à y avoir accès.



Il est clair que l'économie des données actuelle ne fonctionne pas et que le défi que les entreprises doivent relever au plus vite a l'air particulièrement redoutable face à l'émergence du cloud hybride et à l'arrivée de nouvelles applications.

Si les entreprises veulent créer davantage de valeur à partir d'un volume plus important de données, elles doivent changer de façon de penser dans quatre domaines clés :



Accès en libre-service aux données : permettre aux utilisateurs dans l'ensemble de l'entreprise de disposer d'un accès en libre-service aux données, une fois des règles de gouvernance mises en place, plutôt que de limiter l'accès à quelques groupes.



Qualité des données généralisée : garantir la fiabilité des données en structurant les fonctionnalités de qualité et de gouvernance des données pour obtenir une gouvernance continue et mettre en place des workflows pour permettre à ceux qui connaissent le mieux les données d'assurer leur intendance.



Environnement unifié : réduire le coût total d'acquisition et améliorer la productivité en tirant parti d'un environnement unifié pour automatiser le cycle de vie des données et économiser du temps et des ressources.



Exécution en natif : adopter des solutions exécutées en natif sur les principales plateformes pour permettre à l'ensemble de l'entreprise d'adopter les innovations et de tirer parti du cloud computing, des conteneurs, du machine learning et de toutes les nouveautés à venir.

Les technologies émergentes les plus essentielles à vos projets d'intégration des données

Les entreprises qui sont en mesure d'exploiter rapidement les données pertinentes jouissent d'un avantage concurrentiel certain. Les technologies modernes permettent de libérer vos données et de prospérer dans le monde du machine learning en temps réel, multicloud hybride actuel. Voici trois innovations technologiques ayant un impact sur la gestion des données :

- **Cloud computing** : Le cloud a permis de nouveaux gains d'efficacité et des économies que les entreprises n'auraient jamais cru possibles. Le stockage dans le cloud est distant et s'adapte de façon à n'offrir que la capacité requise. Il évite de consacrer du temps et de l'argent à la maintenance de serveurs sur site et offre aux utilisateurs un accès en libre-service et en temps réel aux données, n'importe où et n'importe quand. Les fournisseurs de services cloud, tels que [Amazon Web Services](#), [Microsoft Azure](#) et [Google Cloud Platform](#), ont contribué à encourager la mise en place de pratiques axées sur les données à l'échelle des entreprises. Ils permettent à des entreprises de toutes tailles de stocker, de traiter et d'explorer un plus grand nombre de données sans les investissements en argent et ressources auparavant requis par les technologies sur site. En outre, la technologie cloud offre des gains de souplesse dans les coûts, de nouvelles fonctionnalités et des gains de productivité avec lesquels les infrastructures sur site ne peuvent pas rivaliser.
- **Conteneurs** : Les conteneurs sont en train de rapidement dépasser les machines virtuelles. Selon une [étude](#) récente, l'adoption des conteneurs d'applications [augmentera de 30 % par an jusqu'en 2022](#). Les machines virtuelles nécessitent des frais généraux coûteux et une maintenance chronophage, l'ensemble du matériel et du système d'exploitation devant être géré. Les conteneurs sont portables, avec peu d'éléments mobiles et de maintenance nécessaires. Une entreprise utilisant la capacité de superposition de la technologie de conteneur déverrouillera un potentiel d'exploitation illimité et des économies considérables pour les disrupteurs de la donnée.
- **Informatique sans serveur** : Le déploiement et la gestion des technologies rendues nécessaires par le big data sont généralement des opérations complexes, coûteuses et qui exigent une expertise difficile à trouver. [Selon une étude réalisée par Gartner](#), « les plates-formes PaaS peuvent améliorer l'efficacité et l'agilité des services cloud, réduire le coût du premier accès au cloud et accélérer la modernisation de l'environnement informatique des entreprises. » L'IT serverless permet aux utilisateurs d'exécuter du code sans avoir à provisionner ou administrer les systèmes sous-jacents ou une infrastructure applicative. Avec ce type de solution, les systèmes évoluent automatiquement en fonction de l'augmentation ou de la réduction des charges de travail, à la demande et en fonction de la disponibilité des données. Les entreprises ne paient que pour ce qu'elles exécutent, mettant ainsi un terme aux gâchis associés aux serveurs sur site. L'avenir sera sans serveur ou ne sera pas, son potentiel de libération des données étant illimité.

CHAPITRE 3

QU'EST-CE QUE L'INTÉGRATION DES DONNÉES ET POURQUOI EST-ELLE IMPORTANTE ?



QU'EST-CE QUE L'INTÉGRATION DES DONNÉES ET POURQUOI EST-ELLE IMPORTANTE ?

« Sans technologie d'intégration des données, nous aurions du mal à fonctionner et à tenir nos engagements commerciaux et nous nous exposerions à des pénalités financières de la part des organismes de réglementation. »

– David Clifton,
architecte des solutions d'entreprise,
Affinity Water

L'intégration des données est le processus qui consiste à combiner des données provenant de différentes sources dans une vue unifiée, ce qui les rend plus exploitables et plus utiles pour les utilisateurs qui les consultent. Une intégration des données réussie allie rapidité et intégrité à grande échelle : vous pouvez répondre aux besoins de votre entreprise dès aujourd'hui à l'aide de données fiables, rester au fait des innovations et suivre la croissance exponentielle des volumes de données.

En matière d'intégration des données, il n'existe pas d'approche universelle ou standard. Toutefois, les solutions d'intégration de données partagent généralement quelques éléments, dont un réseau de sources de données, un serveur maître et des clients qui accèdent aux données à partir de ce serveur maître.

Dans la plupart des processus d'intégration des données, le client envoie une demande de données (requête) au serveur maître. Le serveur maître importe les datasets nécessaires à partir de sources internes et externes. Les données requises sont extraites de ces sources, puis combinées sous une forme cohérente et unifiée. Le résultat est ensuite livré au client.

Les données sont généralement réparties entre différentes sources. Les données de toutes ces sources ont souvent besoin d'être rassemblées à des fins d'analyse.

En l'absence de données unifiées, la génération d'un rapport nécessite généralement une analyse, qui implique elle-même de nombreuses opérations : connexion à plusieurs comptes (souvent sur plusieurs sites), recherche de données dans leurs applications natives, copie des données requises, reformatage et nettoyage de ces données. Ceci demande beaucoup de temps et d'efforts, d'où l'importance de l'intégration des données.

Les processus d'intégration des données présentent de nombreux avantages pour les entreprises :

- **L'intégration des données favorise la collaboration et l'unification des systèmes**
Les membres de tous les départements, même s'ils se trouvent bien souvent sur différents sites, ont de plus en plus besoin d'avoir accès aux données de l'entreprise pour des projets communs et individuels. Pour leur faciliter la tâche, le département informatique doit définir une solution sécurisée pour proposer un accès en libre-service aux données à tous les départements, quelle que soit leur situation géographique. Dans la plupart des départements, les employés génèrent, améliorent et enrichissent des données dont le reste de l'entreprise pourrait profiter.
- **L'intégration des données fait gagner du temps**
Lorsqu'une entreprise prend des mesures pour intégrer ses données, elle réduit considérablement le temps nécessaire à leur analyse. Les employés n'ont plus besoin d'établir des relations de A à Z lorsqu'ils ont besoin de générer un rapport. Par ailleurs, l'utilisation d'outils adaptés [à la place du codage manuel](#) fait gagner encore plus de temps à l'équipe de développement (et permet généralement d'économiser des ressources).
- **L'intégration des données réduit les erreurs (et les besoins de modifications)**
Pour rassembler les données manuellement et s'assurer que leurs datasets seront complets et précis, les employés doivent connaître tous les emplacements et comptes qu'ils pourraient avoir à explorer – et installer tous les logiciels nécessaires avant de lancer leurs recherches. En l'absence d'une solution d'intégration qui synchronise automatiquement les données, les rapports doivent être révisés périodiquement pour tenir compte des changements récents. Avec les mises à jour automatisées des outils d'intégration, les rapports peuvent être générés plus facilement, en temps réel et au moment précis où les utilisateurs en ont besoin.
- **L'intégration de données augmente la valeur des données disponibles**
Sur la durée, les efforts d'intégration de données apportent un avantage complémentaire : ils améliorent la valeur des données de l'entreprise. L'intégration de données dans un système centralisé permet de cerner les problèmes de qualité et d'effectuer les améliorations nécessaires en vue d'obtenir des données plus précises, ce qui est le fondement même de la mise en place d'une business intelligence adéquate.

86 % des entreprises ont vu leur productivité multipliée par deux grâce à l'utilisation d'un outil d'intégration des données plutôt que des intégrations avec codage manuel.

Types de projets d'intégration des données

Dans les entreprises, de nombreuses initiatives nécessitent des projets d'intégration de données, dont les plus courants sont les suivants :

Gestion des tâches ETL ou ELT

Les opérations d'extraction, de transformation et de chargement (ETL) forment un processus dans lequel les données sont extraites du système source, transformées et livrées à un emplacement de destination. Ce processus traditionnel évolue au rythme de la technologie. Par exemple, les processus ELT se généralisent.

Création de data warehouses

Les initiatives d'intégration des données – en particulier dans les grandes sociétés – ont souvent pour but de créer des data warehouses sur site ou dans le cloud qui combinent plusieurs sources de données dans une seule base de données relationnelle. Les data warehouses permettent aux utilisateurs de formuler des requêtes, compiler des rapports, produire des analyses et récupérer des données dans un format cohérent.

Simplification de l'informatique décisionnelle (ou Business Intelligence)

En fournissant une vue unifiée des données provenant de nombreuses sources, l'intégration de données simplifie les processus de business intelligence (BI). Les entreprises peuvent visualiser plus facilement et comprendre plus rapidement les datasets disponibles et récupérer ainsi des informations exploitables sur l'état de l'entreprise. Avec l'intégration des données, les analystes peuvent compiler plus d'informations et obtenir des évaluations plus précises.

Exploitation des data lakes

Les data lakes, des magasins centralisés regroupant des données structurées et non structurées, sont souvent volumineux et très complexes. Plus le nombre de sources de données augmente, plus les entreprises ont accès à des volumes de données importants dont elles peuvent envisager l'exploitation. C'est pour cette raison qu'il est de plus en plus vital pour elles de déployer des efforts poussés pour l'intégration de leurs données.

60 % des clients interrogés par Talend utilisent des outils d'intégration des données pour améliorer le rendement de leur entreprise.

Approches d'intégration des données d'entreprise

Il existe plusieurs façons d'intégrer les données. L'approche adéquate dépendra de la taille de l'entreprise, des besoins à satisfaire et des ressources disponibles.

- **L'intégration manuelle des données** n'est ni plus ni moins que le processus par lequel un utilisateur donné collecte manuellement les données nécessaires à partir de différentes sources (en accédant directement à leurs interfaces), puis les nettoie selon les besoins et les regroupe dans un data warehouse. Cette méthode est vraiment peu efficace et peu cohérente.
- **L'intégration de données par middleware** est une approche dans laquelle une application middleware agit en tant que médiateur et aide les utilisateurs à normaliser les données et à les injecter dans le pool de données de référence. Le middleware intervient chaque fois qu'un système d'intégration de données ne parvient pas à accéder par lui-même aux données d'une application legacy.
- **L'intégration basée sur les applications** est une approche dans laquelle les applications logicielles identifient, récupèrent et intègrent les données requises. Le logiciel d'intégration doit assurer la compatibilité des données provenant de différents systèmes pour pouvoir les transmettre d'une source à une autre.
- **L'intégration avec accès uniforme** se concentre sur la création d'un front-end (interface) qui présente sous une forme cohérente les données accessibles à partir de différentes sources. (En réalité, les données sont conservées dans leur source originale). Avec cette méthode, des systèmes de gestion de base de données orientés objets peuvent être utilisés pour créer une apparence d'uniformité.
- **L'intégration avec stockage commun** est l'approche la plus fréquemment utilisée pour le stockage des données résultant d'opérations d'intégration. Une copie des données de la source originale est conservée dans le système intégré puis traitée pour une vue unifiée. L'approche avec stockage commun est le principe sous-jacent de la solution traditionnelle du data warehousing.

Défis fréquemment posés par l'intégration des données

Regrouper les données de plusieurs sources et les transformer en un dataset unifié et stocké dans une structure unique est un défi technique en soi. Lors de la mise en place de systèmes d'intégration, les entreprises sont souvent confrontées aux défis suivants :



Comment arriver au but : Les entreprises savent généralement ce qu'elles attendent de l'intégration de données : la solution à un défi bien précis. L'entreprise qui décide d'implémenter une solution d'intégration des données doit identifier les types de données à analyser, les sources de ces données, les systèmes qui vont les utiliser, les types d'analyse à effectuer et la fréquence à laquelle les données et rapports devront être mis à jour.



Traiter les données provenant de systèmes legacy : Les efforts d'intégration peuvent exiger l'inclusion des données stockées dans les systèmes legacy. Ces données ne sont pas toujours associées à des marqueurs tels que les heures et dates des activités.



Traiter des données externes : Les données provenant de sources externes peuvent présenter un niveau de détail moins élevé que celles provenant de sources internes, ce qui ne permet pas toujours de les examiner avec la même rigueur. De plus, les termes des contrats signés avec certains fournisseurs externes peuvent rendre difficile le partage des données à l'échelle de l'entreprise.



Suivre l'évolution des technologies : Lorsque le système d'intégration des données est en place et opérationnel, la tâche n'est pas terminée. En effet, il est de la responsabilité de l'équipe chargée des données de maintenir les efforts d'intégration au niveau des meilleures pratiques du secteur tout en répondant aux besoins de l'entreprise et aux exigences réglementaires les plus récentes.

CHAPITRE 4

L'INTÉGRATION CLOUD DES DONNÉES : UNE PASSERELLE VERS DES GAINS D'EFFICACITÉ ET DE NOUVELLES INNOVATIONS



INTÉGRATION DES DONNÉES ET LE CLOUD

Les processus classiques d'intégration des données comme ETL et ELT existent depuis des années. Pour autant, le passage généralisé au cloud computing dans le monde professionnel a une incidence sur le mode d'intégration, son emplacement et sa motivation. Les outils d'intégration des données actuels vont devoir s'adapter aux nouvelles infrastructures cloud et devront aussi s'appuyer sur la puissance et la promesse d'innovation du cloud.

Les raisons du basculement de l'intégration des données vers le cloud

Par Philip Russom - TDWI, responsable de la recherche sur la gestion des données

Avec l'arrivée à maturité de certaines tendances, les plateformes, technologies et bonnes pratiques en matière d'intégration des données dans le cloud n'ont jamais été aussi pertinentes :

Le cloud est désormais d'usage courant.

Le cloud est devenu la plateforme de données et de calcul privilégiée pour les applications, les données et les pratiques métiers pilotées par les données telles que l'analytique. Le cloud est également en train de s'imposer dans certaines disciplines de gestion des données, y compris pour l'intégration des données ou des domaines tels que la qualité des données, la gestion des données de référence, le data warehousing et le reporting.

De nombreuses entreprises utilisent plusieurs clouds.

Du fait de la prolifération du cloud et parce que, outre les applications d'entreprise traditionnelles exécutées sur leurs systèmes sur site, elles utilisent plusieurs marques de logiciels en tant que service (SaaS) (p. ex., Salesforce et Marketo) et possèdent des applications dans plusieurs environnements cloud publics (p. ex., AWS, Azure et Google Cloud Platform), de nombreuses sociétés ont fait le choix du « multicloud ». Cet environnement complexe présente des défis et des possibilités hybrides et cloud-to-cloud uniques que l'intégration de données dans le Cloud peut aider à relever et à exploiter.

Les données dans le cloud et sur Internet peuvent désormais coexister avec les données d'entreprise.

Pour avoir adopté le big data, les réseaux sociaux, l'Internet des objets, les applications SaaS et le stockage cloud, de nombreuses entreprises disposent de sources et de cibles de données sur site et dans le cloud. Et, afin de faire face à l'environnement hybride qui en résulte, celles qui sont tournées vers l'avenir doivent moderniser et développer leurs infrastructures d'intégration afin de pouvoir pleinement prendre en charge l'Internet et le cloud.

Le développement du cloud a tendance à être agile

Le rythme effréné des entreprises impose un modèle dans lequel les données, sources et produits orientés données (rapports et analyses, p. ex.) doivent être livrés dans des temps minimes. Les plateformes d'intégration cloud dotées d'interfaces agiles peuvent réduire les cycles de développement en intégrant rapidement de nouvelles sources de données et de nouveaux utilisateurs.

Le libre-service, un prérequis.

Les membres de l'entreprise et autres utilisateurs exigent un accès en libre-service aux data lakes et à d'autres datasets cloud destinés à l'analytique afin qu'ils puissent procéder à l'exploration, la préparation et la visualisation autoguidées des données. Autre tendance connexe : offrir aux utilisateurs de données un accès en libre-service à ces dernières afin de s'adapter à leur nombre croissant. Une solution d'intégration des données cloud peut constituer un point central articulé autour des métadonnées pour le partage des données et la collaboration.

Qu'est-ce que l'intégration des données cloud ?

Il s'agit d'une plateforme d'intégration sécurisée orientée cloud qui associe la puissance des fonctionnalités d'intégration des données classiques à une infrastructure d'intégration des données prenant en charge, nativement, le cloud. Elle prend également en charge des fonctions destinées au traitement de la qualité des données, des données de référence, des métadonnées et des événements. Sans oublier le big data, l'Internet des objets ainsi que de nouvelles sources ou cibles de données dans le cloud ou sur Internet.

Quel est le rôle de l'intégration des données cloud ?

L'intégration des données dans le cloud permet aux développeurs de concevoir des solutions unifiées pouvant être exécutées en natif afin d'améliorer les performances ou fonctionnalités au niveau local. Les entreprises étant de plus en plus amenées à disposer de plusieurs applications et datasets dans le cloud, ce type d'intégration devrait également prendre en charge les interfaces, les appels et les modèles de données natifs des applications SaaS, le stockage cloud et les plateformes de data warehouse cloud.

Quel est l'intérêt de l'intégration des données dans le cloud ?

Elle répond à un besoin bien réel : une infrastructure et des solutions d'intégration des données englobant toutes les applications, données et personnes, indépendamment de leur type ou emplacement. Une intégration des données de cette ampleur est essentielle. En effet, les entreprises de toutes tailles et de tous secteurs font face à des environnements de données de plus en plus hybrides et

Les plateformes d'intégration cloud peuvent réduire les cycles de développement en intégrant rapidement de nouvelles sources de données et de nouveaux utilisateurs.

distribués. Certes l'intégration des données dans le cloud leur permet de mettre en place des solutions d'intégration d'entreprise destinées tout particulièrement aux environnements cloud. Toutefois elle peut également permettre aux utilisateurs de migrer vers le cloud des éléments tels que de multiples applications, des datasets et des groupes d'utilisateurs.

Dans quelle mesure le cloud peut-il améliorer l'intégration des données (et inversement) ?

Le cloud offre plusieurs fonctions permettant d'optimiser les solutions d'intégration des données à d'importants égards :

Évolutivité du cloud.

Pour ce qui est de l'intégration des données, les charges de travail s'intensifient rapidement et sollicitent des ressources serveur considérables avant de s'atténuer tout aussi rapidement. Parmi les exemples les plus courants, notons l'ingestion et les transformations de données, et le prétraitement de données avant le chargement des cibles. Dans ces cas, un cloud élastique peut mobiliser automatiquement les ressources nécessaires, puis les réaffecter à l'issue des charges de travail intenses de l'intégration des données.

Centralisation dans le cloud des capacités sémantiques et collaboratives.

En centralisant les ressources et les services partagés, les entreprises rendent la gestion des données cohérente et gérable, tout en augmentant la productivité et la collaboration des développeurs. Ces pratiques initialement mises en place sur site sont désormais également disponibles dans le cloud : ces ressources et services peuvent être encore plus largement partagés entre des collaborateurs et des départements dispersés ou être appliqués à la production entre les nombreuses plateformes des workflows d'intégration des données hybrides.

Profil économique positif du cloud.

Les ressources de serveur et de stockage sont généralement moins coûteuses sur des plateformes cloud qu'avec les ressources traditionnelles sur site. En outre, le prestataire de cloud gère la planification, l'optimisation, les mises à jour et la maintenance des capacités du serveur ; les spécialistes de la gestion des données ne se laissent ainsi plus distraire par ces tâches chronophages. Pour finir, l'utilisation de serveurs et de moyens de stockage dans le cloud évite à ces spécialistes de consacrer du temps à l'intégration du système ou de dilapider leur budget en dépenses d'investissement.

« Le cloud nous permet de développer des applications évolutives, solides et rapides. Nous avons choisi d'utiliser Talend Cloud comme élément clé de notre stratégie cloud car cela nous permet de facilement intégrer nos applications natives cloud et notre environnement de gestion de la relation client d'arrière-plan actuel... »

*— Tom Murphy,
DSI responsable des technologies
et de l'informatique à l'université
de Pennsylvanie*

Il existe deux catégories d'intégration des données dans le cloud :



Intégration des données exécutée en natif dans le cloud

Les processus d'intégration des données exécutés dans le cloud bénéficient de l'évolutivité, de la neutralité et du coût abordable du cloud. En outre, dans cette architecture, l'intégration des données est placée en amont, à proximité des clouds et des autres sources et cibles sur Internet : de nouvelles pratiques (ingestion, triage, temps réel et streaming) deviennent ainsi possibles.



Intégration des données interagissant avec plusieurs clouds

Pour pouvoir maîtriser l'extrême complexité des environnements de données hybrides actuels, les nouvelles technologies (telles que Spark) ainsi que les outils open source, les sources de big data et le stockage dans le cloud doivent être parfaitement pris en charge. Pour garantir des performances et une capacité optimales (et un maximum de fonctionnalités), l'intégration des données dans le cloud doit également prendre en charge des interfaces vers des applications SaaS très répandues. Pour les cas d'usage dans les domaines de l'analytique et du data warehousing, l'intégration des données dans le cloud doit également prendre en charge des interfaces vers des data warehouses et autres bases de données dans le cloud.

Quelques cas d'usage de l'intégration des données dans le cloud

Analytique avancée

L'un des principaux moteurs du changement dans le domaine de la gestion des données aujourd'hui est la demande pour une grande variété d'approches de l'analytique. De nombreuses entreprises optent pour une extension de leur portefeuille de solutions d'analytique avancée grâce à des systèmes cloud. L'intégration des données dans le cloud doit ainsi commencer par réunir de gros volumes de données provenant de sources variées pour les corrélations de sources croisées qu'utilisent la plupart des outils d'analytique avancée.

Data warehousing dans le cloud

Le data warehouse moderne regroupe, par définition, de nombreux ensembles de données. À mesure que les utilisateurs décident du moyen de moderniser leur data warehouse, ils migrent certains ensembles de données depuis leurs bases de données traditionnelles sur site vers des plateformes de données cloud. De plus en plus d'utilisateurs décident de migrer l'ensemble du data warehouse vers le cloud, processus qui n'a jamais été aussi simple grâce au grand nombre de nouvelles solutions de chargement de données dans le cloud disponibles sur le marché. Que cette migration soit totale ou partielle, le data warehouse moderne s'appuie sur une intégration des données dans le cloud pour, dans un premier temps, migrer les données du data warehouse vers le cloud, puis alimenter le data warehouse avec des sources hybrides au cours de la production quotidienne. Face à un tel volume de données opérationnelles de valeur provenant de plusieurs clouds et d'Internet, les rapports opérationnels ne pourraient pas être complets et actualisés sans l'intégration des données dans le cloud.

Synchronisation de données multicloud

Certaines des applications SaaS les plus répandues actuellement, ainsi que les applications et données disponibles dans plusieurs environnements cloud publics (p. ex., AWS, Azure et Google Cloud Platform), automatisent les processus métiers de vente et de marketing et offrent d'autres fonctions touchant directement la clientèle telles que le service client, la facturation et l'expédition. Obtenir des vues clients complètes dans un environnement multicloud et synchroniser les données clients dans l'ensemble des applications orientées clients (sur site ou dans le cloud) demande une intégration des données dans le cloud perfectionnée et très performante.

« Une intégration des données moderne et globale peut être mise en place n'importe où, dans les locaux de l'entreprise ou le cloud, afin de libérer les systèmes en silos et de permettre à l'entreprise de valoriser ses données. »

L'intégration des données dans le cloud permet aux développeurs de concevoir des solutions unifiées pouvant être exécutées en natif afin d'améliorer les performances ou les fonctionnalités au niveau local. »

– Philip Russom - TDWI, responsable de la recherche sur la gestion des données

Cloud et collaboration

Pour les entreprises, les nouvelles infrastructures cloud présentent des défis et opportunités d'adoption de modèles de développement de logiciels agiles destinés à l'intégration de données.

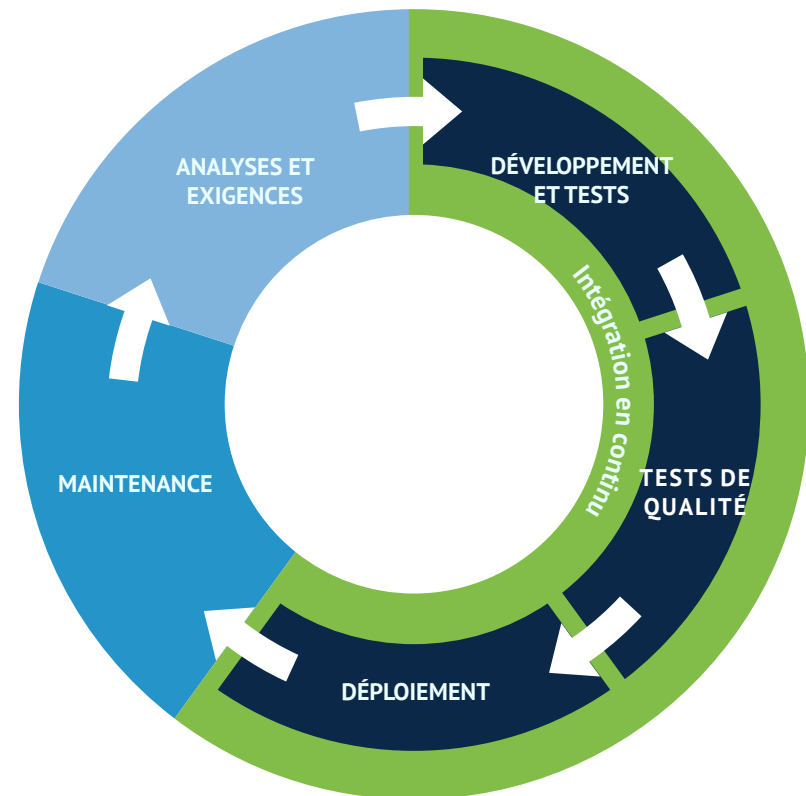
Ces modèles, parfois appelés **DevOps** en matière de développement de logiciels et **DataOps** en matière d'analytique, permettent aux entreprises de développer des applications plus rapides et innovantes, faisant évoluer la façon dont leurs équipes collaborent.

Lorsqu'un logiciel est en cours de développement, plusieurs personnes travaillent généralement sur ce projet et doivent partager le code avec leurs collègues pour éviter tout risque de doublon.

Une fois le logiciel lancé et à l'approche du cycle de maintenance, les anciens et nouveaux développeurs devront savoir où se trouve le code et comment travailler dessus. C'est pourquoi le passage à une infrastructure cloud implique des outils bien conçus afin de faciliter le travail d'équipe et la collaboration.

Par exemple, la livraison en continu (LC) automatise la livraison des logiciels de leur développement à la production. Résultats : moins de défauts, une meilleure qualité du code et des versions d'intégration plus rapides.

En outre, si elle est mise en œuvre correctement et pratiquée régulièrement, l'intégration en continu (IC) permet de réduire les problèmes d'intégration : les jobs, le code et le logiciel peuvent ainsi être livrés plus rapidement.



De plus, en procédant à une intégration régulière, il est possible de détecter rapidement les erreurs et de les localiser plus facilement. L'utilisation des bons outils devrait permettre de réduire le nombre de conflits et de les résoudre facilement pendant l'intégration du code. Le point le plus important est qu'il y a moins de risques que vous détruisiez ce qui existe déjà et, quand bien même cela arriverait, vous pourrez facilement résoudre les problèmes et récupérer les données.

Tout ceci est possible avec un référentiel de logiciels partagé. Ce référentiel est conçu pour regrouper toutes les informations relatives à un projet ainsi que les métadonnées de l'entreprise pour les communiquer à tous afin de faciliter la collaboration et communication, tout au long du développement du logiciel (SDLC).

Il y a d'autres fonctions de collaboration importantes à prendre en compte concernant les outils de développement de logiciel :

- Contrôle de version facilitant la réutilisation des éléments. Il est également possible de revenir à un stade de développement antérieur, au besoin.
- Git, un système de gestion du code source largement utilisé. Git permet un meilleur audit et des référentiels de plus petite taille, p. ex. contrôle de révision et de test des unités automatisé.
- Mise en place d'une sécurité et réutilisation du code dans l'ensemble des projets partagés, des règles et des profils utilisateurs.

« La collaboration fait tomber les barrières technologiques et psychologiques entre les responsables des données de l'entreprise et les utilisateurs de l'information. Ce concept a la faculté de transformer des industries tout entières. »

— Jean-Michel Franco,
responsable principal de la
commercialisation des produits,
Talend

CHAPITRE 5

COMMENT CHOISIR LA STRATÉGIE D'INTÉGRATION DES DONNÉES LA PLUS ADAPTÉE À VOTRE ENTREPRISE



COMMENT CHOISIR LA STRATÉGIE D'INTÉGRATION DES DONNÉES LA PLUS ADAPTÉE À VOTRE ENTREPRISE

Le défi qui consiste à transformer votre entreprise en une organisation pilotée par les données comporte plusieurs volets et le choix de la méthode d'intégration des données la plus adaptée à ses besoins ne constitue que l'un d'entre eux. De nombreux autres facteurs doivent être pris en compte pour savoir comment utiliser vos données afin qu'elles servent davantage vos clients et améliorent le fonctionnement de votre entreprise : où votre infrastructure sera-t-elle hébergée ? Quels types de technologies vos clients utiliseront-ils pour interagir avec vous demain, l'année prochaine ou dans cinq ans ? Qui, au sein de votre entreprise, voudra avoir accès aux données ? Et si la réponse est « tout le monde », comment vous assurerez-vous que chacun obtiendra bien les données qu'il doit analyser ? Comment satisferez-vous non seulement aux réglementations actuelles en matière de protection des données mais aussi à celles qui verront le jour en ligne à l'avenir ?

C'est la raison pour laquelle l'intégration des données n'est pas une simple question de choix du logiciel qui convient, ni même du bon projet. Il s'agit d'une stratégie commerciale globale ayant une incidence sur la capacité de votre entreprise à innover et à se développer. Afin de devenir une entreprise pilotée par les données, il est essentiel de comprendre les objectifs commerciaux de votre entreprise, ses besoins, les ressources disponibles, ainsi que l'orientation globale du marché de la gestion des données afin de vous assurer que vous créez bien une stratégie à l'épreuve du temps qui vous met sur la voie du succès.

« L'agilité des données implique d'acquérir, d'intégrer et d'enrichir des données et de les convertir en une véritable valeur commerciale, quelle que soit la plateforme, le lieu et l'heure. »

– Tim Derrico,
directeur de l'analytique mondiale,
Johnson Controls

Les quatre piliers de la stratégie d'intégration des données

Une erreur fréquente faite par les entreprises lorsqu'elles se lancent dans un projet d'intégration des données est de ne pas réfléchir à la façon dont leurs plans d'intégration pourraient dépasser leurs projets d'intégration initiaux.

La stratégie d'intégration des données de votre entreprise est essentielle pour que vous puissiez devenir une organisation pilotée par les données, ce qui, à terme, vous permettra de prendre des décisions sur la base d'informations axées sur des données avant vos concurrents. Les besoins de votre entreprise évoluent au même rythme que vos besoins en données. Vous devez vous assurer que votre stratégie prend en compte cet état de fait.

Au moment d'élaborer votre stratégie d'intégration des données, réfléchissez aux quatre facteurs clés suivants :

1 Quels sont les objectifs à long terme de votre département et de votre entreprise, au-delà du projet d'intégration de données initial ?

De nombreuses entreprises considèrent un projet d'intégration de données comme le premier pas vers quelque chose de beaucoup plus grand. Imaginons que vous cherchiez à transférer vos données Salesforce dans un data warehouse cloud mais, qu'à terme, l'objectif soit de disposer d'une vue à 360° permettant de conserver un « golden record » de tous vos clients. Assurez-vous que les personnes qui choisissent et mettent en œuvre votre technologie d'intégration connaissent suffisamment bien les objectifs suprêmes pour pouvoir prendre des décisions qui seront pertinentes pour l'entreprise, aujourd'hui et demain.

2 Comment adopterez-vous les technologies émergentes (même celles dont vous ne connaissez pas encore l'existence) ?

La plupart des entreprises veulent croire qu'elles sauront tirer parti de toute nouvelle technologie leur apportant des améliorations tangibles. Toutefois, certaines stratégies d'intégration des données (mise en place manuelle de votre pile d'intégration, p. ex.) seront moins en mesure de vous

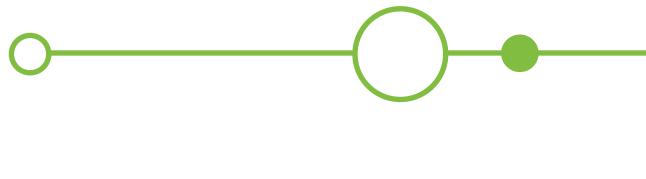
permettre d'utiliser les nouvelles technologies sans devoir consacrer un temps considérable au développement ou acquérir un nouvel outil. Si vous savez d'ores et déjà qu'une flexibilité dans votre environnement de données et les technologies que vous utilisez (pour vous adapter aux applications et infrastructures cloud, p. ex.) sont importantes à vos yeux, vous devrez disposer d'une stratégie d'intégration suffisamment souple pour pouvoir gérer facilement ces changements. L'un des avantages que présente l'utilisation de technologies d'intégration en open source est qu'elles s'adaptent plus facilement aux nouvelles technologies de données. Raison particulière : un grand nombre de ces nouvelles technologies est basé sur des projets open source.

3 Quelle valeur commerciale quantifiable cette stratégie d'intégration des données est-elle censée apporter à votre entreprise ?

Si ce guide met en avant de nombreuses raisons justifiant l'importance qu'a eu une stratégie d'intégration des données pour la plupart des entreprises présentées ici, il est crucial d'expliquer l'incidence qu'une stratégie d'intégration aura précisément sur votre entreprise. Si vous comprenez la valeur commerciale que vous apportera une plateforme d'intégration, vous serez davantage en mesure de hiérarchiser les fonctions les plus importantes à vos yeux au moment d'évaluer les méthodes et fournisseurs d'intégration qui s'offrent à vous. Il est également vital de communiquer vos objectifs commerciaux à vos équipes techniques afin qu'elles puissent créer plus efficacement l'infrastructure de données dont votre entreprise a besoin.

4 Disposez-vous des collaborateurs et ressources techniques nécessaires pour procéder aux changements que vous souhaitez amener dans votre entreprise ?

Bien évidemment, vous aurez besoin de ressources techniques pour développer, tenir à jour et adapter vos projets d'intégration. Toutefois, les stratégies permettant de créer une entreprise pilotée par les données ne se limitent pas à un simple projet technique. Bien au contraire, elles demandent des changements culturels et organisationnels dans l'ensemble des fonctions de l'entreprise et non pas simplement les départements techniques. Si votre projet d'intégration a pour but de transformer le paysage de données de votre entreprise, pensez à collaborer avec d'autres parties prenantes afin de définir ce que ces changements signifient pour le département informatique et les métiers qui utilisent les données.



CHAPITRE 6

LA GRANDE QUESTION : CODAGE MANUEL OU OUTIL D'INTÉGRATION DES DONNÉES ?



LA GRANDE QUESTION : CODAGE MANUEL OU OUTIL D'INTÉGRATION DES DONNÉES ?

De nombreux spécialistes des données se demandent pourquoi ils auraient besoin d'un outil d'intégration des données alors que le **codage manuel** leur permet généralement d'exécuter les tâches requises rapidement et à moindre coût. Tout responsable informatique doit prendre en compte un certain nombre de facteurs lorsqu'il évalue les avantages que présentent le codage manuel et une approche de l'intégration des données basée sur des outils.

Le codage manuel et les outils d'intégration des données vont souvent de pair ; il peut être difficile de déterminer quand chacune de ces méthodes peut être utilisée. La plupart des entreprises les associe lorsqu'elles s'attaquent à des défis techniques.

*« L'attrait des **scripts SQL de codage manuel** destinés à l'intégration des données peut être irrésistible. Mais le codage manuel peut avoir des conséquences imprévisibles à long terme. Ce qui n'était au début qu'un simple script SQL codé manuellement se transforme en douzaines de pages de scripts non documentés et non conformes qui sont difficiles à répéter, auditer, vérifier et valider. »*

– Nick Piette,
Product Marketing Director,
Talend

Les décideurs en matière d'informatique qui envisagent de mettre en place des projets codés de façon personnalisée doivent se pencher sur les points suivants :



Veiller à examiner les coûts à court et long terme : S'il est possible que vos coûts de déploiement soient réduits de 20 % avec une approche codée de façon personnalisée, les coûts de maintenance pour leur part augmenteront de 200 %. Si vous envisagez de mettre en place un processus répétable sur lequel votre entreprise peut s'appuyer, le choix d'un outil d'intégration des données peut s'avérer plus viable.



Le codage manuel se justifie, mais uniquement dans des situations très précises : Le codage personnalisé peut être judicieux pour des projets ponctuels, simples et très ciblés ne nécessitant pas beaucoup de maintenance. Il pourrait également s'avérer nécessaire pour des situations dans lesquelles aucun outil n'est en mesure de faire le travail requis.



Comprendre que les projets d'intégration des données nécessitant plusieurs développeurs bénéficieront d'environnements de conception visuels qu'apportent certains outils : Si plusieurs développeurs travaillent sur votre projet d'intégration, il est important de réfléchir à la façon dont leur travail s'imbriquera. Sans codage sur mesure, rien ne garantit l'uniformité d'un développeur à l'autre, ce qui peut rendre le développement et la maintenance complexes et coûteux. Un outil offrant la possibilité de réutiliser les éléments de développement antérieurs évitera à votre équipe d'intégration de répéter inutilement ses efforts et se traduira par des flux d'intégration des données plus efficaces.



Il est important de bien comprendre les coûts de maintenance et d'assistance qui accompagneront tout projet : Si plusieurs personnes assurent la maintenance et l'assistance du code une fois en production, leur courbe d'apprentissage avec une approche de codage manuel sera élevée. Pire encore, si le code est en production pendant plusieurs années et que la personne qui l'a développé quitte l'entreprise, comprendre comment le job d'intégration fonctionne (et comment résoudre les problèmes) devient beaucoup plus difficile et coûteux.

La checklist du codage manuel :

Compte tenu des avantages et des inconvénients du codage manuel, nous avons élaboré une checklist composée de questions permettant de déterminer le choix adapté à votre situation : le codage manuel ou un outil.

1 Mon équipe de développement possède-t-elle les compétences voulues pour utiliser le codage manuel ?

Si vous utilisez une nouvelle technologie de type Hadoop ou plateforme cloud, qui fera le travail et quel sera le temps de démarrage nécessaire ?

2 Est-ce bien à cela que je souhaite que mes experts en codage manuel consacrent leur temps ?

Les experts en codage manuel constituent généralement un quart d'une équipe de développement, ce qui en fait une denrée rare. Si un non-expert peut faire le même travail à l'aide d'un outil et économiser au passage des heures de travail, ne préféreriez-vous pas que les experts utilisent leurs compétences uniques pour quelque chose exigeant leur savoir-faire ?

3 Existe-t-il un outil me permettant d'accomplir la même tâche plus rapidement et moins cher que le codage manuel ?

On demande à la plupart des équipes du département informatique d'en faire toujours plus avec toujours moins. Une approche basée sur un outil permet souvent de réduire le coût par développeur et d'effectuer plus rapidement le travail.

4 S'agit-il d'un projet ponctuel isolé ou d'un domaine que je prévois de continuer à développer au fil du temps ?

Si vous vous lancez dans un projet utilisant le big data ou une plateforme cloud, il est probable que vous en veniez à utiliser de plus en plus cette plateforme au fil du temps. Dans ce cas, l'utilisation d'experts en codage manuel sera très difficile à mettre en place du fait de leur rareté.

5 Quel sera le degré de portabilité de ce code si je souhaite le réutiliser sur une nouvelle plateforme technologique comme Spark ou Flink ?

Au moment d'établir votre budget et la durée du projet d'intégration, réfléchissez aux efforts supplémentaires requis pour revoir l'ensemble de votre travail, en plus des efforts nécessaires pour ce nouveau développement. Les principaux outils d'intégration de données vous permettent de passer d'un cadre de traitement des données à un autre en toute simplicité et ainsi d'éviter tout recours au code existant.

6 Plusieurs développeurs collaboreront-ils sur ce projet ?

Une approche basée sur des outils présente de nombreux avantages lorsque plusieurs développeurs travaillent de concert : réutilisation et partage du code en toute simplicité, environnements de conception visuels, documentation automatisée et même assistants et experts pour conseiller les développeurs.

7 Pendant combien de temps ce code sera-t-il en production ?

Lorsque vous vous lancez dans un nouveau projet, vous pouvez être tenté de mettre l'accent sur le temps nécessaire au développement et oublier le temps de production. Ce qu'il a fallu six mois pour développer sera souvent en production pendant cinq ans, voire davantage. Si c'est le cas, les coûts d'assistance et de maintenance de ce code subsisteront 10 fois plus longtemps que la durée de développement initial ; il est donc essentiel que vous ayez une idée précise de vos coûts d'assistance et de maintenance.

8 Qui sera responsable de la maintenance de ce code ?

Si vous ne disposez que de quelques développeurs, ils devront assurer la maintenance et l'assistance de leur code. À terme, l'assistance et la maintenance les accapareront à plein temps et ils seront dans l'impossibilité d'accepter de nouveaux projets qui pourraient permettre à votre entreprise de gagner en compétitivité.

9 À quelle fréquence le code devra-t-il être mis à jour pour faire face aux nouveaux besoins de votre entreprise ou aux changements apportés aux sources ou cibles de données ?

Les sources et les cibles de données, ainsi que les besoins de votre entreprise sont en constante évolution. Si on est en droit de s'attendre à ce flux constant de changements, le coût de la maintenance et du support pour garantir l'intégrité et la fiabilité des données en sera sensiblement plus élevé.

Comment choisir l'outil d'intégration de données adapté à vos besoins ?

Une fois que vous aurez décidé de faire l'acquisition d'un outil d'intégration des données et établi s'il est judicieux de migrer votre intégration de données dans le cloud, sachez que la plupart des entreprises auront besoin d'un certain nombre de fonctions pour pouvoir faire face aux défis de leur intégration.

Ce que vous devriez privilégier dans votre outil d'intégration des données :

- Il doit être capable de dialoguer (lire et écrire) avec l'ensemble des sources de données dont vous avez besoin, quel que soit leur emplacement – sur site ou dans le cloud.
- Il doit supporter les processus de transformation de données tels que le tri, le filtrage et l'agrégation.
- Il doit intégrer des fonctionnalités de qualité des données et gouvernance des données telles que la déduplication, les correspondances et le profilage des données.
- Il doit proposer des fonctionnalités de collaboration.
- Avec le passage aux systèmes cloud, la capacité d'intégrer les processus CI/CD (intégration continue/développement continu) est désormais une nécessité.
- Vos outils d'intégration des données doivent être capables de fonctionner dans tous les types d'environnement : sur site, dans le cloud ou dans des infrastructures hybrides.
- Pour les tâches extrêmement personnalisées, aucun outil d'intégration des données ne disposera de toutes les fonctionnalités clés en main. Ce n'est pas un problème. Assurez-vous simplement que votre outil est suffisamment flexible pour vous permettre d'enrichir ses fonctionnalités au moyen de composants personnalisés.

- Il doit être capable de s'adapter facilement aux changements de prestataire. Il est important de disposer d'un outil d'intégration des données compatible avec un environnement multicloud. Il doit également être capable de s'adapter aux changements de prestataire et d'environnement de déploiement en remplaçant simplement quelques composants mais en conservant les algorithmes métier et de transformation.
- Un outil d'intégration des données doit être capable de fonctionner avec les dernières innovations et de s'adapter facilement aux nouvelles technologies. Les outils de qualité sont prêts à s'intégrer avec les technologies sans serveur, Spark, Snowflake, le machine learning, etc. et à s'adapter rapidement aux nouvelles technologies qui n'ont pas encore vu le jour.
- L'évolutivité est très importante dans le choix des outils ETL. Cela implique que l'outil peut traiter facilement des processus simples, ainsi que des intégrations de plus en plus complexes. Cela vous permettra de suivre le rythme de croissance de vos activités analytiques, de réutiliser des objets d'un projet à un autre et ainsi d'économiser du temps et des ressources.
- La portabilité est également une capacité importante mais parfois négligée dans les outils d'intégration de données. Par exemple, la portabilité de l'écosystème Apache Hadoop est incroyablement rapide. En 2014 et 2015, MapReduce était le standard, mais il a été détrôné par Spark à la fin de l'année 2016, et il était impossible de porter les solutions en codage manuel de MapReduce à Spark. Les outils les plus performants permettent cette flexibilité en toute transparence.
- Choisissez un outil ETL adapté aux futurs besoins de votre entreprise. Si un important projet de gestion des données de référence ou de qualité des données est en préparation, assurez-vous de bien choisir un outil qui vous permet d'élargir votre environnement de données. Par exemple, si vous transférez vos données dans un entrepôt de données dans le cloud, votre outil doit être en mesure de le faire rapidement et avec un minimum d'effort.

CHAPITRE 7

COMMENT DEVENIR UNE ENTREPRISE ORIENTÉE DONNÉES MODIFIE L'ORGANISATION DE VOTRE ÉQUIPE DE DONNÉES



COMMENT DEVENIR UNE ENTREPRISE ORIENTÉE DONNÉES MODIFIE L'ORGANISATION DE VOTRE ÉQUIPE DE DONNÉES

Une fois vos outils d'intégration de données sélectionnés, vous estimerez peut-être que votre travail est terminé et que votre entreprise peut désormais récolter les bénéfices qui découlent de son pilotage par les données. Mais les choses ne font que commencer. Le pilotage par les données ne dépend pas simplement du logiciel que l'entreprise a choisi. Il s'agit pour cette dernière d'organiser ses équipes, des utilisateurs individuels à la direction, de façon à ce que les données se trouvent au cœur de l'organisation. C'est ce que l'on appelle ériger l'intégration des données en stratégie, et non pas simplement en processus décisionnel réactif. En élaborant une stratégie d'intégration des données à l'échelle de l'entreprise qui fait des données un sport d'équipe et met en place une gestion collaborative des données, vous serez mieux préparés pour réussir dans l'environnement commercial hyper concurrentiel actuel.

Pour que votre entreprise soit réellement pilotée par les données, vous devez effectuer quatre actions tant au niveau de vos collaborateurs que des processus :



Comblent le fossé commercial/IT :

Mettez en place une collaboration intégrée pour les développeurs, les data scientists, les analystes et les opérations. Travaillez de concert tout au long du cycle de vie des données.



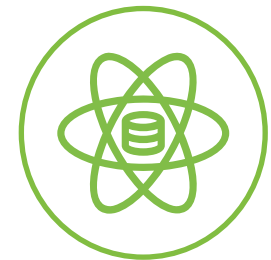
Instaurer une culture axée sur les données :

Mettez les bonnes données entre les bonnes mains au bon moment. Proposez des applications en libre-service adaptées à chaque rôle.



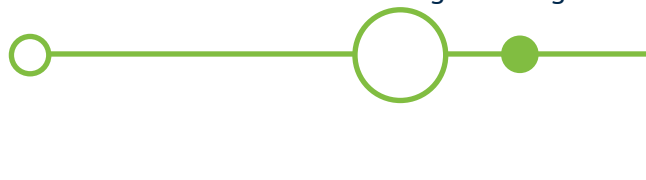
Augmenter radicalement la taille :

Laissez le département informatique accorder à un grand nombre de personnes dans l'ensemble de l'entreprise un accès à des données fiables de façon gérée et gouvernée.



Ne laisser aucune donnée de côté :

Utilisez une seule et même plateforme pour plusieurs cas d'usage : big data, cloud, intégration d'applications, MDM, préparation, intendance.



Le nouveau paradigme des entreprises : la gestion collaborative des données

Aujourd'hui, le département informatique a du mal à répondre aux demandes d'intégration des données dans l'ensemble de l'entreprise. Les utilisateurs veulent accéder aux gisements de données de plus en plus importants que les entreprises ont amassés et aux informations qu'ils sont susceptibles de recéler. Si le département informatique ne dispose pas des ressources nécessaires pour offrir un accès aux informations contenues dans les data lakes de l'entreprise, les collaborateurs de l'entreprise trouveront une solution qui risque de mettre en danger les informations, comme ils le font déjà à l'ère du BYOD.

La meilleure solution est donc de proposer un accès en libre-service aux données à tous les départements. Toutefois, le département informatique doit trouver le bon moyen d'y parvenir afin d'éviter d'exposer les actifs de l'entreprise à des risques inutiles. Il doit adopter un modèle de gestion collaborative des données.

Comment mettre en œuvre une gestion collaborative des données

Le passage d'une gestion et gouvernance autoritaires à une gestion et gouvernance collaboratives peut s'avérer difficile. Toutefois, les départements informatiques ont ici l'occasion de créer un système de confiance autour des magasins de données de l'entreprise, dans lequel le personnel collabore avec ces départements pour tenir à jour ou améliorer la qualité, la gouvernance et la sécurité des données. La bonne nouvelle : les spécialistes de l'informatique disposent pour cela d'un schéma directeur élaboré par les entreprises pionnières de l'utilisation du World Wide Web. Le Web 2.0 s'articulait autour de tendances mettant l'accent sur la collaboration des utilisateurs, le partage de contenus générés par les utilisateurs et les réseaux sociaux. C'est également le cas de la gestion collaborative des données.

La gestion collaborative des données fait tomber les barrières technologiques et psychologiques entre les responsables des données de l'entreprise et les utilisateurs de l'information, ce qui permet à tous les membres de l'entreprise de partager la responsabilité de la sécurisation des données de l'entreprise.

Le Web 2.0, tout comme la gestion collaborative des données, s'articulait autour de tendances mettant l'accent sur la collaboration des utilisateurs, le partage de contenus générés par les utilisateurs et les réseaux sociaux.

INTÉGRATION DES DONNÉES COMMENT DEVENIR UNE ENTREPRISE ORIENTÉE DONNÉES MODIFIE L'ORGANISATION DE VOTRE ÉQUIPE DE DONNÉES **CHAPITRE 7**

Avec ce modèle, le plus grand défi et catalyseur a toujours été la confiance. Les informations étaient conçues et publiées par un très petit nombre de spécialistes des données qui axaient leurs efforts sur les utilisateurs qui les ingéraient.

Aujourd'hui, la prolifération des informations au sein des entreprises est hors de contrôle, tout comme sur le Web. Nous sommes tous confrontés à l'apparition d'un nombre croissant d'applications cloud émanant des départements de la vente, du marketing, des RH, des opérations ou des finances dans le but de compléter les anciennes applications IT conçues au niveau central (ERP, data warehousing ou CRM, p. ex.). Les applications numériques et mobiles connectent les systèmes informatiques au monde extérieur. Pour gérer ces nouveaux flux de données, nous observons au sein des entreprises l'apparition de nouveaux rôles axés sur les données (data analysts, data scientists ou data stewards) qui rend la différence plus floue entre les utilisateurs et les fournisseurs de données d'entreprise.

Au fur et à mesure de l'évolution du Web 2.0, la confiance entre les utilisateurs et leurs prestataires de service a été instaurée par des mécanismes participatifs de notation, classement et réputation numérique (comme Yelp). Les départements informatiques qui adoptent les stratégies mises en œuvre par leurs utilisateurs plus libres peuvent obtenir les mêmes résultats positifs.

Les entreprises IT en pleine évolution disposent de différentes options pour instaurer un système de confiance : l'association d'une gestion collaborative des données et d'un libre-service n'est qu'une parmi tant d'autres. Grâce au libre-service, les utilisateurs des différents secteurs d'activité sont davantage impliqués dans le recueil, le nettoyage et la qualification des données provenant de diverses sources. Ils peuvent ensuite les analyser et les utiliser afin de prendre des décisions plus éclairées.



Grâce à la gestion collaborative des données, le département informatique peut contribuer à garantir la qualité, la sécurité et l'exactitude des informations de l'entreprise dans un environnement en libre-service. Elle permet aux collaborateurs d'une entreprise de corriger, qualifier et nettoyer ces mêmes informations d'entreprise. Le département informatique peut créer des workflows gouvernés afin de fournir des modèles d'intendance collaborative et gouvernée des données. Les enregistrements de données de référence sont par conséquent mis à jour par ceux qui les connaissent le mieux.

En outre, le passage vital à une plus grande participation des utilisateurs de l'entreprise au traitement de ses données cruciales offre de nombreux autres avantages. Par exemple, les utilisateurs économisent du temps et gagnent en productivité s'ils travaillent avec des données fiables. Les départements marketing améliorent leurs campagnes. Les centres d'appels disposent d'informations clients plus fiables et exactes, à la satisfaction de tous. Et l'entreprise peut ainsi mieux contrôler son actif le plus précieux : ses données.

Les 6 conseils pratiques de la gestion collaborative des données :

À faire

Définissez vos attentes dès le départ

Nous avons vu combien il est important pour votre entreprise d'élaborer et de documenter votre stratégie d'intégration des données. Dans cette optique, il est impératif de définir des attentes pour la réalisation de vos objectifs. Quels KPI devriez-vous définir ? À quel point cela aura-t-il une incidence sur les résultats économiques de votre entreprise ? Assurez-vous de répondre à ces questions aussi bien sur le plan commercial que technique. Veillez également à savoir où se trouve votre ligne d'arrivée afin que vous puissiez définir des objectifs et des étapes intermédiaires pour le projet.

Constituez une équipe interdisciplinaire

Vous devrez bien évidemment disposer de l'expertise technique adéquate dans le cadre de vos projets d'intégration des données. Mais vous devrez également impliquer des personnes qui comprennent dans quelle mesure votre stratégie d'intégration des données a une incidence sur votre entreprise et en faire vos relais dans leurs départements respectifs.

Obtenez des résultats rapides

S'il est essentiel de mobiliser les compétences de vos collaborateurs et de définir des objectifs ambitieux, il est également nécessaire de prouver que votre stratégie d'intégration des données et les projets qui lui sont associés auront très rapidement un impact commercial positif. Ne consacrez pas trop de temps à la planification. Vous devez prouver que votre stratégie peut avoir des effets commerciaux et produire des résultats immédiats. Si vous offrez une valorisation de l'information optimisée et plus rapide, vous pourrez acquérir une crédibilité immédiate et le soutien de vos collaborateurs. Une fois que votre projet vous aura assuré crédibilité et confiance, il sera plus facile de demander des ressources supplémentaires lorsque vous présenterez vos succès au reste de l'entreprise. N'oubliez pas : les petites victoires s'additionnent pour en faire une grande.

À éviter

Ne confinez pas l'intégration des données aux équipes techniques

On pense souvent que les projets techniques appellent des réponses techniques. Mais l'intégration des données doit être traitée comme une stratégie commerciale et un défi technique. Pour réussir, vos objectifs doivent être de notoriété publique au sein de votre entreprise. Vous devez contrôler la mise en récit du projet pour éviter qu'une mauvaise communication ne circule dans les différents départements. Vous devez par conséquent trouver la combinaison idéale de savoir-faire et de maîtrise de la communication afin que vos résultats soient connus et communiqués efficacement au sein de votre entreprise.

Ne concevez pas des projets trop complexes

Il est important d'obtenir rapidement des résultats : commencez par des petits projets pour atteindre des résultats remarquables. Ces résultats rapides peuvent parfois être très simples. Par exemple, le détaillant [Carhartt](#) avait du mal à intégrer des données « propres » dans ses systèmes. Le projet initial clé consistait à nettoyer 50 000 enregistrements en une seule journée avec Talend Data Preparation. Il a donné l'impulsion nécessaire pour poursuivre avec d'autres projets d'intégration de données.

Ne vous laissez pas distraire – gardez le cap sur des objectifs clairs pour l'exécution des projets

Pour finir, il est important de fixer des échéances et de les tenir dans la mesure du possible afin d'asseoir votre crédibilité. Votre organisation peut passer à des priorités commerciales à court terme. Ne vous laissez pas distraire ; suivez votre route et ne perdez pas de vue vos objectifs finaux. Veillez à toujours livrer vos projets à temps. Ensuite, lorsqu'une étape d'un projet est terminée, prenez le temps de fêter cela avec votre équipe et l'entreprise.

Le futur de l'intégration : donner aux nouveaux intégrateurs les moyens d'agir sans perte de contrôle

Les entreprises modernes souhaitent être pilotées par les données mais pas question de mettre la pression sur le département informatique pour qu'il reste en phase avec les dernières innovations.

Cela signifie que l'entreprise doit donner les moyens à ses collaborateurs de prendre des décisions à l'aide de données pertinentes, facilement accessibles. En plus de technologies et d'applications en libre-service, des formations et communications internes visant à définir une culture axée sur les données dans divers départements commerciaux doivent être mises en place.

Trois étapes doivent être suivies pour donner aux intégrateurs citoyens les moyens d'agir et de transférer l'intégration des données en dehors du département informatique :

1

Les entreprises doivent garantir un processus dans le cadre duquel les données peuvent être traitées par des utilisateurs IT non centraux, facilement suivies à l'aide d'une gestion du lignage des données et des métadonnées en bonne et due forme et déployées par l'équipe d'ingénierie des données. Sans une plateforme ou un processus gouverné de ce type, l'ouverture de l'accès aux données à des intégrateurs ad hoc ou citoyens ne peut que se traduire par davantage de chaos parmi les données.

2

Lorsque la gestion des données est considérée comme un sport d'équipe, chaque personne traitant des données doit comprendre quel est son rôle. Elle doit savoir comment interagir avec l'équipe et y apporter sa contribution afin d'obtenir les données (et les informations) les plus exactes possible. La mise en place d'un processus convenu au préalable permet aux équipes de travailler de concert.

3

Vous devrez trouver une plateforme d'intégration des données qui vous aidera à donner à tous les membres de votre équipe de données les moyens d'agir et de déployer les processus définis entre les membres de l'équipe, le tout sous la direction du département informatique.

Le défi de la sécurité des données

Quasiment tous les rôles au sein d'une entreprise, du DSI à l'analyste des processus métiers, ont besoin que les données soient à portée de main des utilisateurs. Ces responsables des données doivent avoir accès aux données pour pouvoir mettre en place des stratégies, les exécuter et offrir à l'entreprise les informations les plus pertinentes et à jour possible.

Leur réflexe pour y parvenir peut consister à faire en sorte qu'autant de données que possible soient mises à la disposition du plus grand nombre. Toutefois, cette option n'est pas viable. Avec la mise en place de réglementations telles que le RGPD, les entreprises ont l'obligation de plus en plus lourde de s'assurer que seules les bonnes personnes ont accès à certaines informations. Sinon, elles mettent en péril l'ensemble de l'organisation.

La solution consiste à mettre en place des solutions informatiques en libre-service qui automatisent certaines fonctions telles que les demandes d'accès aux données et la préparation des données. Ceci est fondamental pour que les collaborateurs de l'entreprise puissent accéder plus rapidement aux bonnes données et pour obtenir un lignage clair des personnes ayant eu accès à quelle information et à quel moment. Dans le même temps, l'utilisation d'outils automatisés de préparation des données est essentielle pour soulager l'équipe informatique en lui évitant d'avoir à réaliser des tâches manuelles de nettoyage et formatage. Par ricochet, cela permettra au département informatique de mettre l'accent sur la mise en œuvre de nouvelles technologies plutôt que la résolution de petits problèmes pour le reste de l'entreprise.



59 %

*des salariés emportent
des données d'entreprise
confidentielles lorsqu'ils
quittent l'entreprise.*

CHAPITRE 8

ÉLABOREZ VOTRE STRATÉGIE D'INTÉGRATION DES DONNÉES AVEC DES FONCTIONNALITÉS DE GESTION DES DONNÉES UNIFIÉES



ÉLABOREZ VOTRE STRATÉGIE D'INTÉGRATION DES DONNÉES AVEC DES FONCTIONNALITÉS DE GESTION DES DONNÉES UNIFIÉES

L'un des facteurs les plus importants de toute solution moderne d'intégration de données est qu'elle ne se limite pas à l'intégration de données. De nombreuses autres fonctionnalités doivent être intégrées aux logiciels d'intégration de données, l'idéal étant qu'elles soient conçues avec un seul environnement de conception et un jeu d'outils de gestion. Cela signifie que les développeurs n'ont pas à installer ou apprendre à se servir de nouveaux outils lorsqu'ils veulent s'attaquer à un nouveau type d'intégration de données (batch, temps réel, big data, cloud et sur site). Vous bénéficiez ainsi d'une bonne dose d'agilité et le coût de possession est grandement réduit parce que vous pouvez passer d'un type d'intégration à un autre sans avoir à installer et apprendre à utiliser de nouvelles technologies d'intégration.

Certaines de ces fonctionnalités clés sont la qualité des données, la sécurité et confidentialité des données, l'adoption d'innovations avec des architectures open source, l'automatisation, l'intégration à des plateformes d'analytique et le partage de données via des API.

L'intégration des données est inutile sans qualité des données

Nous avons souligné l'importance d'une business intelligence pilotée par les données pour prospérer dans l'environnement commercial hyper concurrentiel actuel. Mais la qualité des informations qu'une entreprise peut extraire est proportionnelle à la qualité des données. Les mauvaises données peuvent rendre difficile l'extraction d'informations exactes et se traduire par de mauvaises décisions. De nombreux dirigeants sont préoccupés par cette situation. Selon Forbes Insights et l'étude « [2016 Global CEO Outlook](#) » de KPMG, 84 % des PDG s'inquiètent de la qualité des données qu'ils utilisent pour obtenir des informations pertinentes pour l'entreprise.

Les motifs d'inquiétude sont nombreux : intégration de nouvelles sources de données dans leurs systèmes actuels, investissements financiers et pressions concurrentielles nécessaires pour tirer parti de l'ensemble des données d'entreprise disponibles et difficulté d'extraire les données des silos dans lesquels elles se trouvent. Et leurs craintes ne sont pas sans fondement. Une étude menée par MIT Sloan School of Management/Review indique que les données incorrectes peuvent coûter jusqu'à **15 à 25 %** du chiffre d'affaires total d'une entreprise.

Le vieil adage « garbage in, garbage out » selon lequel des données inexactes se traduisent par des résultats erronés reste valable. Les données de mauvaise qualité ont des conséquences néfastes pour toutes les entreprises à de nombreux niveaux, tandis que les données de bonne qualité sont un atout stratégique et un avantage concurrentiel pour elles. Si les données peuvent nourrir votre stratégie commerciale, les données incorrectes peuvent la détruire.

Une approche proactive de la qualité des données vous permet de vérifier et mesurer la « propreté » de vos données avant qu'elles ne soient intégrées dans vos systèmes de base. L'accès à ces données et leur surveillance dans des applications internes, cloud, web, et mobiles constituent un défi de taille. Ce type de suivi ne peut être mis en place dans l'ensemble de ces systèmes qu'à l'aide de processus d'intégration des données.

Avec les bons outils, vous pouvez créer des lanceurs d'alerte qui peuvent détecter certaines des causes profondes des problèmes globaux de qualité des données à mesure que vos données sont traitées et intégrées. Vous devrez ensuite suivre les données dans l'ensemble de vos applications et systèmes. Ceci vous permettra d'analyser, de normaliser et de faire correspondre les données en temps réel. Vous pouvez configurer le contrôle et la correction des données au besoin.

Comment limiter les risques en matière de sécurité des données

Une violation des données aux conséquences catastrophiques est l'un des plus grands risques auquel les entreprises sont confrontées dans le paysage d'affaires moderne. Des chefs de la sécurité, des SIS et même des PDG ont perdu leur poste suite à une violation des données qui a abouti à la divulgation à des parties externes de données sensibles appartenant à leurs clients. Mais les conséquences ne se limitent pas à une personne ou un département. Une importante faille de sécurité peut coûter à une entreprise des millions voire des milliards de dollars d'amendes, sans oublier une perte de la confiance du public, une dégradation de l'image de marque et une importante perte d'activité.

La confidentialité et la protection des données retiennent de plus en plus l'attention des chefs d'entreprise, des citoyens, des forces de l'ordre et des gouvernements. Les réglementations sur les données qui se limitaient auparavant à des secteurs fortement réglementés comme la banque, l'assurance, les soins de santé ou les sciences de la vie, sont en plein essor dans tous les pays et s'appliquent à toute entreprise de n'importe quelle taille ou n'importe quel secteur.

Ces réglementations ont une incidence sur la protection des données et sont modifiées par des réglementations gouvernementales relatives à la confidentialité, au stockage, au traitement et aux transferts des données au-delà des frontières. Ces lois sont devenues un obstacle au stockage de données dans le cloud et doivent être parfaitement comprises et prises en compte lorsque des informations sont créées dans un pays puis transférées dans un autre en vue d'une analyse ou d'un traitement.

Avec l'accélération de l'économie numérique, la législation sur la protection des données et la cybersécurité se répand à travers le monde, obligeant les entreprises à sécuriser les informations personnelles, à gérer les risques liés à leurs données et à assumer leurs responsabilités légales.

Les entreprises doivent s'efforcer de faciliter l'harmonisation des équipes juridiques, de conformité, de protection des données personnelles et de gestion des données d'entreprise afin qu'elles réutilisent les artefacts de gouvernance des données existants pour faire appliquer la conformité des données. Les entreprises doivent notamment définir les éléments de données concernés par la souveraineté des données et associer ces attributs à des applications dans le référentiel de métadonnées.

Pour en savoir plus sur la façon de configurer les contrôles de sécurité et confidentialité des données au sein de votre entreprise, en conformité avec les réglementations de protection des données actuelles, consultez le [plan de gouvernance des données](#) étape par étape de Talend concernant la conformité avec le RGPD.

Innovations en open source

Les plateformes d'intégration des données enracinées dans l'open source vous donnent accès aux dernières innovations avant vos concurrents car de nombreux systèmes intégrés sont eux-mêmes en open source.

Les infrastructures de données se complexifient de plus en plus. Une plateforme d'intégration des données doit donc pouvoir être exécutée sur chacune d'entre elles : sur site, cloud, hybride ou multicloud. La portabilité doit être universelle ; vous devriez pouvoir créer vos jobs une seule fois et les exécuter n'importe où, quel que soit le prestataire.

En outre, il est important de générer un code natif optimisé pour le déploiement que vous choisirez. Aucun élément propriétaire ne devrait être installé ou géré sur le cluster. Vous pourrez ainsi voir le code et le déboguer plus facilement. Vous pouvez concevoir vos flux d'intégration une fois, puis les redéployer sur d'autres plateformes technologiques au rythme de l'évolution de vos besoins. Ceci vous permet de tirer pleinement profit de vos investissements informatiques actuels. En effet, vous pouvez les réutiliser à l'infini au fur et à mesure que vos besoins changent. Votre stratégie d'intégration peut évoluer avec vous.

Pour finir, il va sans dire que votre plateforme d'intégration des données doit comprendre des centaines de connecteurs prêts à l'emploi. Votre entreprise doit se connecter à de nombreux systèmes. Les connecteurs préinstallés peuvent accélérer de façon significative vos projets d'intégration de données et la productivité de vos développeurs.

« L'intégration de données en ligne et hors ligne nous aide à trouver de nouveaux moyens de communication avec nos clients sur tous les canaux. Ce type d'interaction fidélise nos clients. »

– Matt Steell, directeur de l'architecture d'information et d'intégration,
Office Depot

Mise en place de la stratégie par le biais de l'automatisation

Les plateformes d'intégration de données comportent une fonctionnalité importante : l'incorporation de l'automatisation et du machine learning. Pour mettre en place la stratégie d'intégration des données, soulager le département IT et faire des données un sport d'équipe, il est nécessaire d'automatiser le cycle de vie des données : recueillir, gérer, transformer et partager.

La suppression du codage manuel et de toute intervention humaine dans ces processus vous permet de gagner en vitesse et d'extraire davantage de données des silos pour les mettre à la disposition d'analystes qui sont en mesure de les utiliser. En outre, le machine learning a amélioré la qualité des données, promesse d'un enrichissement et de meilleures informations.

Cela signifie que, pour la première fois, les départements informatiques peuvent offrir à un très grand nombre de personnes un accès à des données fiables, améliorant au passage la productivité, la vitesse et la qualité des données.

Un seul et même outil doté de nombreuses fonctionnalités

Talend, disponible dans le cloud ou sur site, améliore la qualité, gouvernance et protection des données au sein du processus d'intégration afin que votre équipe puisse mettre des données fiables à la disposition de tous les utilisateurs. Grâce à la facilité d'intégration, aux contrôles de qualité intégrés et à la gestion des règles, les données sont enrichies, protégées et disponibles à partir d'un environnement unique et unifié.

Talend propose une solution complète d'intégration, de gouvernance et de qualité des données avec connectivité, profilage, nettoyage, mise en correspondance et suivi intégrés des données pour répondre à tous vos besoins en matière de gouvernance et de qualité des données. Avec Talend, vous pouvez supprimer toutes les données incohérentes, mettre en place des règles métiers et créer des informations cohérentes grâce à la normalisation. Les équipes chargées de la gouvernance des données gagnent en productivité en collaborant pour accéder, comprendre et normaliser les données correspondant à certains domaines de données (données de nom, adresse ou produit, p. ex.).

Une aide et une assistance automatisées sont nécessaires pour traiter les analyses de données. Cela vaut particulièrement pour les importants volumes de données non structurées car cela permet aux utilisateurs d'accélérer l'analyse et la valorisation des informations provenant de différents formats et sources de données. Le machine learning joue un rôle clé dans ce processus collaboratif.

INTÉGRATION DES DONNÉES ÉLABOREZ VOTRE STRATÉGIE D'INTÉGRATION DES DONNÉES AVEC DES FONCTIONNALITÉS DE GESTION DES DONNÉES UNIFIÉES **CHAPITRE 8**

Talend combine la sémantique intelligente et le machine learning pour transformer plus rapidement les données en informations. La sémantique intelligente capture automatiquement les empreintes numériques dans un pipeline de données afin d'accélérer la découverte et la liaison des données, ainsi que la gestion de la qualité. Le machine learning suggère la meilleure démarche à adopter pour le pipeline de données ou enregistre les connaissances tacites des utilisateurs de la plateforme Talend (développeur dans Talend Studio ou steward dans Talend Data Stewardship, p. ex.) et automatise les processus en fonction des besoins.

Tous ces outils et fonctionnalités sont réunis sur une seule et même plateforme. Tout ce dont vous avez besoin pour que votre stratégie d'intégration des données puisse se développer en même temps que votre entreprise se trouve à portée de mains.



CHAPITRE 9

API, INTÉGRATION D'APPLICATIONS ET DONNÉES EN TANT QUE SERVICE (DaaS)



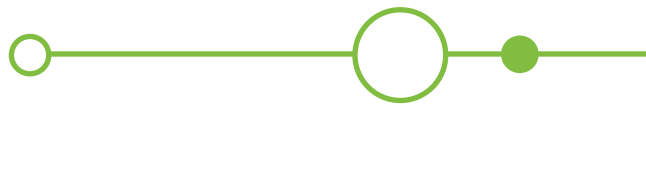
API, INTÉGRATION D'APPLICATIONS ET DONNÉES EN TANT QUE SERVICE (DaaS)

Alors que les données se retrouvent au cœur de la prise de décision des entreprises, les convertir en ressources exploitables reste l'un des défis les plus difficiles auxquels sont confrontées les entreprises aujourd'hui. Elles continuent à recueillir et générer plus de données que jamais mais, dans la pratique, les utiliser s'avère difficile. Selon une étude récente d'Accenture, **80 % des entreprises sont assises sur des données non structurées et inaccessibles.**

Le partage de ces données aux niveaux interne et externe présente un avantage stratégique et ouvre la voie à de nouveaux produits et services. Mais quelle est la meilleure façon de partager des données sans compromettre leur sécurité et leur qualité ? Comment les mettre en commun avec des partenaires aussi bien internes qu'externes ?

Pour résoudre ce défi de l'intégration, les équipes ont généralement dû mettre en œuvre des intégrations point à point au sein de leur entreprise. Mais, à mesure que la demande en données augmente, il devient difficile de maintenir à jour les intégrations existantes et de créer rapidement de nouvelles connexions.

Il existe un meilleur moyen de partager des données fiables dans l'ensemble de l'entreprise, dans toutes les applications qui en ont besoin, au moment où elles en ont besoin : les données en tant que service.



Que sont les données en tant que service ?

Les données en tant que service visent à accélérer la collaboration en matière de données et à abolir les silos de données en créant un flux de données toujours prêt, toujours actif et en permanence accessible par les abonnés. L'expert de la transformation numérique, Daniel Newman, explique : « L'utilisation de données en tant que service (DaaS) permet aux entreprises d'accéder à des flux de données en temps réel depuis n'importe quel endroit dans le monde. Ces données suppriment les contraintes liées aux sources de données internes. Presque toutes les entreprises modernes ont adopté les données comme outil de prise de décision mais rares sont celles qui disposent du personnel et des ressources internes nécessaires pour tirer pleinement profit des données qu'elles collectent. »

Les entreprises qui exploitent les DaaS mettent toutefois l'accent sur le recueil des données et leur compilation au sein de flux pertinents. Les abonnés accèdent ensuite aux flux dont ils ont besoin, quand ils en ont besoin. Il s'agit fondamentalement d'éliminer la nécessité d'un investissement interne dans les données et de permettre aux entreprises d'opérer avec une plus grande agilité car elles peuvent obtenir de façon transparente et sans effort les données dont elles ont précisément besoin. » Les architectures qui utilisent des services de données peuvent éviter toute redondance et rationaliser les coûts grâce à un hébergement des données critiques en un seul emplacement, ce qui permet à plusieurs utilisateurs d'y accéder tout en garantissant un point de mise à jour unique.

Il est intéressant de noter que les DaaS pourraient aussi bien être utilisées en interne qu'aider les entreprises à rentabiliser leurs données en les vendant à l'extérieur à des clients DaaS.

80 %

*des entreprises sont assises
sur des données non
structurées et inaccessibles.*

Quels avantages les données en tant que service présentent-elles ?

Les avantages d'un flux de données en temps réel toujours prêt, toujours actif sont évidents pour toute entreprise qui traite des données, notamment celles qui souhaitent les rentabiliser. De nombreuses entreprises ont trouvé d'autres avantages liés à la mise en œuvre d'un modèle de DaaS :



Agilité :

Les utilisateurs de données peuvent rapidement faire preuve d'innovation grâce à la simplicité de l'accès et au fait qu'ils n'ont pas besoin d'avoir une connaissance étendue des données sous-jacentes.



Qualité et gouvernance améliorées :

Grâce à un accès par le biais d'une série de services gérés, les entreprises pourront plus facilement améliorer la qualité et le lignage de leurs données. Les entreprises doivent orienter leur choix vers des plateformes qui supportent l'intégration, la qualité et le développement de services afin de s'assurer que la fourniture de la couche du service de données demeure agile.



Réduction des coûts :

Les équipes techniques peuvent concevoir les services de base avec les spécialistes des données et, ce faisant, abstraire la complexité et réduire les coûts. Il est possible de ne réaliser les mises à jour qu'une seule fois, ce qui permet aux spécialistes des données d'être plus productifs.

Quelle est la première étape pour devenir une entreprise qui utilise les DaaS ? Adoption des API

Lors de la mise en place de données en tant que service, certains éléments d'infrastructure doivent être créés. Les DaaS devront fournir, à l'insu de l'utilisateur, un accès à des services de données et middleware, des moteurs d'exécution, des services de gestion d'interface et d'autres services dont les applications peuvent avoir besoin pour être alimentées en données. Il s'agit d'une infrastructure compliquée composée de nombreux éléments mobiles qui peuvent changer rapidement. Les intégrations point à point pouvant être utilisées pour créer l'infrastructure DaaS ont une faiblesse importante : leur fragilité. Si une intégration s'écroule, elle risque de détruire l'ensemble du maillage et de mettre en péril l'entreprise. On doit pouvoir isoler les intégrations des différentes applications et les mettre à jour de façon dynamique, à mesure que chaque mise à jour d'un point de terminaison individuel arrive dans l'API.

La plupart des API web sont des services REST qui offrent un accès à certaines fonctionnalités à granularité fine ou grossière des clouds privés ou publics. Ainsi, les API supportent de nombreux modèles de système, notamment :

- Services utilitaires, tels que stockage et calcul ;
- Services d'application, tels que gestion des inventaires ou enregistrement d'une vente ;
- Services de base de données, tels que transfert de données vers une base de données et extraction de données d'une base de données ;
- Services analytiques, tels qu'analytique prédictive afin de déterminer la configuration future des ventes en utilisant la plupart des données d'une entreprise ;
- Services middleware, tels qu'envoi de données dans une file d'attente et extraction de données d'une file d'attente.

Les API permettent à ceux qui créent et déploient des applications ou des intégrations de données à l'aide de plateformes cloud de ne pas avoir à s'occuper des services natifs sous-jacents. Au lieu de cela, ils utilisent une API pour gérer l'interaction avec ces services. Les développeurs n'ont pas besoin de réinventer la roue chaque fois qu'ils exécutent des services simples (accès aux informations d'une base de données, mise en service d'un serveur, envoi d'un message, etc.). Vous définissez l'API une fois selon vos préférences, et elle restera disponible pour fournir ce service.

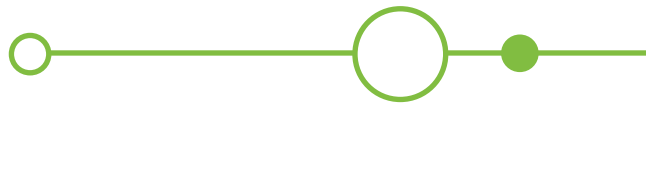
Quelles technologies sont utilisées dans une API ?

Les API utilisent un ensemble de technologies spécifiques ; elles sont ainsi facilement comprises par un large éventail de développeurs. Parce qu'elles sont axées sur la simplicité, les API peuvent fonctionner avec n'importe quel langage de programmation courant et être comprises par tous les programmeurs, même ceux qui ont peu ou pas d'expérience de la technologie API.

- ▶ **REST** - L'approche la plus populaire pour créer des API web est la technologie REST (Representational State Transfer). Cette méthode de conception d'API tire parti des mécanismes d'Internet utilisés pour consulter des pages web normales. Elle offre donc l'avantage de permettre des implémentations plus rapides et est plus facile à comprendre et à utiliser pour les développeurs. Les API REST vous permettent de prendre des informations et des fonctionnalités déjà disponibles sur votre site web et de les rendre rapidement disponibles via une API de programmation. Les applications aussi bien web que mobiles peuvent la réutiliser, étendant ainsi de façon radicale la présence de votre entreprise sur de nouveaux canaux, et ce sans grand travail supplémentaire.
- ▶ **JSON** - JavaScript Object Notation (JSON) est un mode d'échange d'informations destiné aux programmes. Les API permettent à des programmes de communiquer mais, comme elles n'ont pas de voix, elles ont besoin d'un moyen pour décrire les données et les informations qui sont échangées. JSON utilise différents symboles (crochets, guillemets, deux points et virgules) pour séparer les données. Les informations disposent ainsi d'une structure pertinente permettant aux ordinateurs de faire la différence entre un prénom et un nom de famille ou toute autre information susceptible de décrire des données. JSON est devenu l'un des moyens privilégiés par les programmeurs pour permettre la communication des API. Il offre une méthode simple et légère d'échanger des données sur Internet, tout en conservant la structure et la signification de ces données.

Découvrez combien il est facile de mapper JSON dans ce [webinaire](#).

- ▶ **Webhooks** - Les webhooks sont une forme de notifications Push pouvant être déclenchée par une action précise. Une fois déclenchés, ils envoient des informations à une adresse de site web externe. Les webhooks permettent aux développeurs de choisir l'action, l'URL et les champs associés au push. Lorsqu'un webhook est déclenché, il transmet toutes les informations associées à un emplacement où le développeur peut les traiter. Les webhooks sont un excellent moyen de réduire les interrogations constantes sur une API parce qu'ils n'envoient des données aux développeurs de l'API que lorsque l'action requise est déclenchée. Avec les webhooks, l'API devient une voie à deux sens : les développeurs peuvent non seulement extraire des données des plateformes d'API mais aussi en recevoir en temps réel à mesure que des événements surviennent.



Comment puis-je m'assurer que mes API sont sécurisées ?

La sécurité est la principale préoccupation en matière de déploiement d'API web. Les entreprises qui rendent de précieuses données accessibles sur Internet ont besoin de méthodes solides pour sécuriser ces ressources. Au cours des 10 dernières années, le mouvement des API web modernes a développé des techniques éprouvées pour sécuriser le nombre croissant de ressources d'API :

- ▶ **Clés** - L'une des méthodes les plus courantes pour sécuriser l'accès des applications aux API est l'utilisation d'une « clé ». Il s'agit souvent d'une combinaison de caractères alphanumériques que chaque développeur peut obtenir par le biais d'une simple procédure d'inscription d'API. Chaque utilisateur de l'API doit inclure une clé à chaque appel vers une API, ce qui permet aux fournisseurs de cette API de vérifier l'utilisation et d'émettre ou de révoquer des clés, s'ils le jugent nécessaire.
- ▶ **Autorisation de base** - Une autorisation de base est un moyen de fournir des informations d'identification sous la forme d'un nom d'utilisateur et d'un mot de passe. L'autorisation de base étant intégrée au protocole HTTP, c'est la façon la plus simple pour les utilisateurs de s'authentifier auprès d'une API web. L'autorisation de base est facile à utiliser. Toutefois, si SSL n'est pas utilisé, le nom d'utilisateur et le mot de passe ne sont pas transmis en texte brut et peuvent être facilement interceptés sur l'Internet ouvert. L'autorisation de base est parfaitement adaptée aux API destinées à un plus large public mais n'offre pas d'accès à des informations trop sensibles.
- ▶ **Autorisation ouverte** - L'autorisation ouverte (OAuth) est une norme ouverte d'autorisation qui permet aux utilisateurs de partager leurs ressources privées avec une autre application sans avoir à donner leurs informations d'identification personnelles (en général leurs nom d'utilisateur et mot de passe). En lieu et place de ces informations d'identification personnelles, OAuth permet aux utilisateurs de remettre des jetons uniques leur donnant accès à un ensemble spécifique de ressources. OAuth est un service complémentaire mais distinct d'OpenID, qui implique de définir l'identité. OAuth nécessite également de définir de l'identité mais aussi d'établir les ressources auxquelles la personne a accès. Cette autorisation constitue une excellente option pour la sécurité des API.

REST, JSON, OAuth et les webhooks sont devenus les technologies privilégiées par les fournisseurs et les utilisateurs d'API car ils respectent les principes clés de simplicité, de sécurité, de mise à disposition des données et des ressources et d'intégration en toute simplicité dans les applications web et mobiles. Cette suite de technologies pour API web n'a pas été désignée par un organisme de normalisation ou une entreprise unique. Elle s'est mise en place au cours des 13 dernières années au travers des bonnes pratiques des principaux fournisseurs d'API existants en réponse aux demandes des développeurs.

Quelles questions se poser avant de mettre en place une infrastructure d'API ?

Si les données en tant que service intéressent votre entreprise, vous devrez les activer en élaborant une approche « API-first » consistant à construire une API au centre de votre système d'information. Mais avant de vous lancer dans la création d'un système d'API, vous devez commencer par un peu de planification. Si vous envisagez de rendre vos données disponibles via des API, voici les questions auxquelles vous devez répondre :

- Dans quel format vos données sont-elles stockées ?
- Quels format et structure souhaitez-vous adopter pour vos données lorsque vous les partagez ?
- Où sont-elles stockées ?
- Quelles données ont la plus haute valeur ajoutée au sein de votre entreprise ?
- Où les données les plus facilement accessibles se trouvent-elles dans votre entreprise ?

En intégrant ces sources de données dans une vision centralisée des API à l'aide d'outils d'API courants et d'une stratégie uniforme de déploiement, de gestion et de découverte de ces API, vos données sont prêtes à l'usage dans le cadre d'intégrations de systèmes publics et privés, web, mobiles et tiers. En résumé, la carte que vous élaborerez pour définir votre stratégie d'API simplifiera grandement la création de votre infrastructure de DaaS. Une conception efficace des API est construite spécialement et facile à utiliser, et aide votre entreprise à comprendre et utiliser rapidement les ressources disponibles. La mise en place d'une couche d'accès aux API peut réduire le temps nécessaire pour intégrer vos données et accroît la probabilité d'une valeur ajoutée et d'un engagement accru.

Une approche d'API peut en outre faire des merveilles pour favoriser l'adoption de la mise en œuvre d'une architecture de données en tant que service. L'utilisation d'API peut **réduire les coûts informatiques traditionnels de 41 % et multiplier par 12 le nombre de nouvelles versions.** »

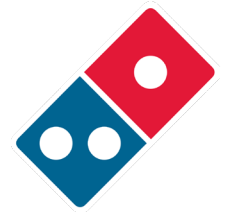
Découvrez comment mettre en place une architecture d'API dans votre entreprise.

« L'utilisation d'API peut réduire les coûts informatiques traditionnels de 41 % et multiplier par 12 le nombre de nouvelles versions. »

CHAPITRE 10

ÉTUDES DE CAS





Aliments et boissons Domino's Pizza

Fondée en 1960, Domino's Pizza, est la plus grande société de distribution de pizzas au monde. Une grande partie de son activité est axée sur les pizzas livrées et à emporter. Elle fait partie des plus grandes marques mondiales de restaurants publics avec près de 15 000 points de vente sur plus de 85 marchés.

Domino's AnyWare est le nom de la fonctionnalité permettant à ses clients de commander des pizzas sur des montres intelligentes, des téléphones, des systèmes multimédia embarqués et des plateformes de réseaux sociaux. Domino's souhaitait intégrer les informations venant de tous ces canaux, soit 85 000 sources de données structurées et non structurées, de façon à avoir une vue d'ensemble de ses données clients et de ses opérations mondiales. Malheureusement, l'architecture informatique en place chez Domino's l'empêchait d'atteindre ces objectifs.

« Nous ne disposions pas d'un moyen standard d'ingérer les données ; en fait, nous en avons trois : nous faisons beaucoup de développement personnalisé et rencontrons des problèmes de qualité. », déclare Dan Djuric, le vice-président en charge de la gestion des infrastructures et informations d'entreprise mondiales chez Domino's Pizza, Inc. « Nous avons également plus de 11 000 utilisateurs métiers, 35 data scientists, ainsi que des agences marketing. . . ». Ajoutant encore à la complexité, l'entreprise utilisait plusieurs outils de capture de données et devait gérer 17 To de données.

Domino's a choisi Talend Data Fabric, une plateforme unifiée d'intégration de données, de big data, de gestion des données de référence, de qualité des données, etc. Domino's a combiné la plateforme Talend à Hadoop (Hortonworks) et à l'analytique de MicroStrategy dans le cadre de son cadre de gestion des informations d'entreprise. Les raisons clés ayant motivé le choix de Talend étaient une plus grande liberté pour s'adapter à une architecture plus agile, la souplesse de l'open source, une mise en œuvre rapide, une gestion des licences rentable et intelligible et la prévisibilité. « Talend est désormais notre plateforme de mobilisation des données, », annonce M. Djuric. « Tout ce qui se passe dans notre écosystème débute avec Talend. Il capture les données, les nettoie, les normalise, les enrichit, les stocke et permet à plusieurs équipes de les utiliser. »

Avec Talend, Domino's a construit un traceur de données qui recueille les données auprès de tous ses systèmes de points de vente, de 26 centres dédiés à la chaîne d'approvisionnement et à travers des canaux aussi variés que les SMS, Twitter, Pebble, Android et Amazon Echo. Les données sont ensuite combinées à des données d'enrichissement provenant d'un grand nombre de sources tierces, ainsi que d'informations de géocodage et de données démographiques et sur la concurrence.

Grâce à sa plateforme de données moderne, Domino's dispose désormais d'une source unique d'informations fiables qu'il peut utiliser pour améliorer ses performances métiers, de la logistique aux prévisions financières, tout en permettant une expérience d'achat individuelle dans plusieurs points de contact.

« Nous sommes aujourd'hui une entreprise d'e-commerce qui vend des pizzas. Talend nous a accompagnés dans notre transformation numérique. »

– Dan Djuric, vice-président en charge de la gestion des infrastructures et des informations d'entreprise mondiales, Domino's Pizza, Inc.

Énergie et électricité : Uniper

Uniper produit, négocie et commercialise de l'énergie à grande échelle et compte parmi les plus grands producteurs d'électricité au monde. L'entreprise achète, stocke, transporte et distribue des produits de base tels que le gaz naturel, le gaz naturel liquéfié (GNL), le charbon et de nombreux autres produits liés à l'énergie.

« Nous évoluons dans un monde de plus en plus complexe dont les marchés et les technologies ne cessent d'évoluer. » a déclaré René Greiner, responsable de l'intégration des données chez Uniper SE. Nous produisons de l'énergie. Nous achetons et vendons de l'énergie sur les marchés. Quelle quantité de charbon et de gaz devons-nous produire aujourd'hui ? Et à l'avenir ? Le marché va-t-il prendre une tout autre direction ? Comment allons-nous développer nos positions sur le marché ? Comment pouvons-nous optimiser nos profits et réduire nos pertes ?

« Après avoir formulé l'idée d'une stratégie de données à l'échelle de l'entreprise, nous avons décidé d'opter pour une solution cloud public pour des raisons de coût et de capacité d'évolution. Nous sommes également parvenus à la conclusion que les solutions Talend seraient les mieux adaptées pour une architecture cloud de ce type. » explique M. Greiner. « Les solutions Talend sont essentielles à notre stratégie cloud car elles nous permettent d'accéder à l'ensemble des données de pratiquement n'importe quel endroit, et elles éliminent les restrictions de format et de volume des données ; « [et] la capacité des solutions Talend à se connecter à un grand nombre de sources et la conception modulaire de leurs produits ont également été des facteurs décisifs. »

Uniper a utilisé Talend pour centraliser plus de 120 sources de données internes et externes dans un data lake Snowflake hébergé par le cloud Microsoft Azure. « Sur 40 jours, se souvient M. Greiner, nous avons pu implémenter 30 % des cas d'usage prévus dans notre feuille de route et en ajouter 25 % au stade de prototypes. Et nous avons réduit nos coûts d'intégration de 80 % ! »

« La gouvernance des données assurée par les solutions Talend est essentielle à la réussite de notre data lake. », ajoute M. Greiner. Si vous voulez vous appuyer sur une plate-forme telle qu'un data lake qui contient des données provenant de sources disparates, vous devez implémenter une solution efficace de catalogage des données (ou un bon mécanisme de marquage des données), et vous devez affecter des propriétaires aux différents ensembles de données. Talend Data Catalog fournit ces capacités et nous aide à établir l'historique (lignage) des données et à garantir le niveau de sécurité dont nous avons besoin pour nous conformer à la réglementation du RGPD. »

Talend a également changé la donne en nous permettant d'accéder aux données dix fois plus vite et dix fois moins cher. Selon M. Greiner : « Nous ne voulons pas que l'IT devienne un point de ralentissement, mais bien au contraire un facilitateur du libre-service. Nous sommes confrontés à des volumes considérables de données, et nous voulons que les utilisateurs disposent de moyens autonomes pour consulter et exploiter ces données, ce qui est désormais possible grâce aux solutions Talend. Et cette ubiquité du libre-service nous donne un avantage très net en termes de rapidité de mise sur le marché. »

Négociant avec efficacité et gérant les risques, Talend travaille à l'échelle de l'entreprise pour aider Uniper dans des activités telles que l'analyse des marchés, le commerce du gaz, la gestion des ressources et l'administration post-marché.

« La disruption est aujourd'hui le mot-clé du secteur de l'énergie. La gestion de l'information est donc essentielle. Grâce à notre nouvelle plateforme d'analyse de données qui repose sur des solutions Talend, nous pouvons désormais mieux comprendre l'évolution du marché, ce qui nous aide à optimiser le trading énergétique tout en gérant les risques et en respectant les réglementations. »

– René Greiner, responsable des solutions d'intégration de données chez Uniper SE

Haute technologie : Lenovo

Lenovo, une entreprise de technologie grand public qui vaut 46 milliards de dollars, est le plus important fabricant de PC et le 4^e producteur de smartphones au monde et compte des clients dans plus de 160 pays.

« Les attentes des clients ont évolué au fil des années. », déclare Marc Gallman, le directeur de Lenovo Analytics and Data Platform.
« Nous avons dû répondre aux questions typiques suivantes : Quelles options amènent le plus les clients à choisir nos ordinateurs ? Quel type de disque dur serait préférable pour nos clients ? » Pour y répondre, Lenovo exploite différents gros ensembles de données. Mais l'émergence de technologies sur site, cloud et SaaS a créé un problème de connectivité des données.

« Nous avons décidé de créer une architecture hybride big data associant Amazon Web Services (AWS) et nos propres serveurs Lenovo. L'idée était d'assurer la confidentialité et sécurité de nos données ainsi que de bénéficier des avantages du cloud. Talend est rapidement devenu un composant clé de cette architecture. »

Avec Talend, Lenovo a créé une plateforme hybride-cloud élastique afin d'analyser plus de 22 milliards d'informations clients par an. Chaque année, les départements de Lenovo capturent plus de 250 téraoctets de données et 60 types de sources de données différents. 8 300 rapports offrant des tableaux de bord en temps réel, des flux de données API et des analyses de données sont délivrés tous les ans à plus de 615 utilisateurs de Lenovo.

« Avec Talend, près de 300 processus d'intégration de nombreux types et sources de données sont exécutés simultanément et nous nous attendons à ce que ce chiffre continue de grimper à mesure que nous développerons cette approche. », annonce M. Gallman.
« Nous avons pu augmenter notre chiffre d'affaires par département de 11 %. Le taux de vente de la gamme d'ordinateurs portables ThinkPad est également en hausse de 18 %. »

Avec Talend, Lenovo a économisé environ 140 000 dollars rien que sur ses coûts de migration initiaux. Les coûts opérationnels (frais de personnel) de Lenovo ont diminué de plus de 1 million de dollars en 1 an (34 %) tandis que la productivité a été multipliée par 2-3. En outre, Talend a contribué à améliorer les rapports de performance et à réduire les délais de traitement de quelques heures à seulement quelques minutes. La simplicité d'utilisation de la plateforme Talend permet à Lenovo de pouvoir répondre aux attentes de ses clients concernant l'augmentation de la vitesse d'acquisition des données. Les délais de mise sur le marché pour plus de 95 % des requêtes est de 14 jours.

« Notre objectif est d'améliorer la satisfaction client via des tactiques marketing. En combinant toutes nos données, nous avons réussi à mieux connaître nos clients et à mieux les satisfaire. »

– Marc Gallman, directeur,
Lenovo Analytics and
Data Platform

Santé – AstraZeneca

AstraZeneca plc est un groupe biopharmaceutique international guidé par la science, basé à Cambridge au Royaume-Uni. Septième plus grand groupe pharmaceutique au monde, il opère dans plus de 100 pays.

Les données d'AstraZeneca sont réparties dans l'ensemble de l'entreprise, au sein d'un large éventail de sources et de référentiels. Les données devant être extraites de systèmes financiers, de CRM, de RH et de plusieurs versions différentes des systèmes SAP ERP, certains projets de reporting et d'analyse cruciaux étaient retardés.

« Pour pouvoir analyser facilement ces données, nous savions que nous avions besoin de mettre en place une architecture nous permettant de réaliser une consolidation de masse et de regrouper les données au cœur d'un seul et même référentiel. », déclare Simon Bradford, ingénieur en chef analytique et données chez AstraZeneca. « Nous voulions tout regrouper et disposer d'un seul ensemble de métriques global afin de pouvoir suivre les activités de l'ensemble des divisions et marchés et établir des comparaisons impossibles jusque-là. »

« Nous avons commencé à utiliser le data lake et chaque semaine voit naître de nouvelles questions métiers de la part de nos responsables. Nous n'avons fait qu'effleurer la surface. »

– Simon Bradford, senior data and analytics engineer, AstraZeneca

AstraZeneca a opté pour la mise en place sur AWS d'un data lake destiné au stockage de données provenant d'un large éventail de systèmes sources. Le groupe a choisi Talend pour la saisie des données. Andy McPhee, science and enabling units data and analytics engineering lead, explique : « Talend est chargé de recueillir, déplacer, transformer et remettre nos données dans le cloud. Ces données sont extraites de multiples sources, puis transférées dans Amazon S3. Les jobs Talend sont créés puis exécutés dans AWS Elastic Beanstalk. Après transformation, Talend charge en masse les données dans Amazon Redshift où elles seront analysées. Talend est également utilisé pour la connexion à AWS Aurora. »

Le data lake créé par AstraZeneca montre l'intérêt d'une stratégie d'intégration des données basée sur une infrastructure réutilisable. « Le data lake nous permet d'extraire d'importants volumes de données précieuses provenant de systèmes disparates et de les mettre à la disposition de l'ensemble de nos divisions. », déclare Mr McPhee.

Grande distribution : Office Depot

Office Depot Europe opère dans 14 pays sous ses deux principales marques, Office Depot et Viking, et est désormais le plus grand revendeur de produits et services destinés au lieu de travail. Le secteur traditionnel de la grande distribution est en pleine évolution : une nouvelle forme de concurrence qui cherche à s'appropriier les clients de l'entreprise arrive sur le marché. Les distributeurs de fournitures de bureau font ainsi face à une menace directe qui les prive d'une partie de leur cœur de clientèle.

« Nous misons sur une combinaison d'activités en ligne/hors ligne et de services technologiques pour augmenter notre chiffre d'affaires. », explique Matt Steell, le directeur de l'architecture d'information et d'intégration chez Office Depot Europe. « Mais certaines questions restaient sans réponse : pourquoi avons-nous perdu des clients dans certains segments et en avons-nous attiré dans d'autres ? Que se passerait-il si nous accroissions nos dépenses sur certains canaux ? Nous avons besoin d'être plus prédictifs et de vérifier différentes hypothèses. »

La difficulté pour nous résidait dans le fait que les données d'Office Depot Europe étaient stockées dans des silos et gérées séparément par chaque fonction ; la direction était donc dans l'impossibilité de se faire une idée globale de ses clients, opérations ou finances. « Pour pouvoir apporter des réponses à des questions de haut niveau touchant différentes fonctions telles que la chaîne d'approvisionnement, le commerce électronique ou le marketing, nous devons extraire des données de différents systèmes se trouvant dans divers pays. L'opération était beaucoup trop chronophage. », déclare Mr Steell. « Nous savions que nous devions centraliser nos données et en attribuer la gestion à un seul groupe afin de pouvoir disposer d'une perspective d'ensemble de nos opérations et nos clients et baser nos décisions sur ces informations et non plus sur notre intuition. »

L'approche adoptée par Office Depot pour résoudre ce problème a été la mise en place d'un data hub intégré. « Grâce à notre Enterprise Data Hub qui repose sur des solutions Talend et Hadoop, Office Depot est passé d'une entreprise orientée application à une organisation réellement pilotée par les données. L'ensemble de nos collaborateurs, des directeurs aux analystes métiers, ont accès aux bonnes informations au bon moment, ce qui leur permet de prendre de meilleures décisions opérationnelles. », indique Mr Steell

Office Depot utilise également Talend Data Quality pour réaliser des contrôles et vérifier la qualité des données avant de les intégrer dans le data lake du hub. « Nous disposons d'un cadre de qualité des données pour toutes les données traitées par le data hub. », déclare Mr Steell. « Nous tenons à nous assurer que nous disposons de données fiables de qualité car elles seront utilisées pour l'analytique prédictive et dans la prise de décisions commerciales essentielles. »

Talend et la nouvelle architecture de data hub offrent de nombreux avantages à Office Depot, notamment un rendement accru et une réduction des coûts.

« Nous sommes très satisfaits de Talend. Ces solutions nous permettent d'apporter des informations à nos utilisateurs plus rapidement et à moindre coût. »

– Matt Steell, directeur de l'architecture d'information et d'intégration, Office Depot Europe

CHAPITRE 11

LA CHECKLIST DE L'INTÉGRATION DES DONNÉES



LA CHECKLIST DE L'INTÉGRATION DES DONNÉES

Toutes les entreprises, qu'elles soient petites ou grandes, ont besoin d'outils d'intégration des données. Qu'il s'agisse d'extraire des données de feuilles de calcul, d'applications packagées, de bases de données, de réseaux de capteurs ou de flux de réseaux sociaux, remplacer les processus en double et les silos d'information par un partage et une réutilisation de l'information présente un avantage considérable. Il est également important de choisir une solution pouvant répondre à tous vos besoins en matière d'intégration de données : intégration des données, migration des données, intégration de big data, intégration de data warehouses ou intégration dans des systèmes de business intelligence.

Cette checklist constitue un cahier des charges fonctionnel de la mise en œuvre et du déploiement d'une intégration des données dans un environnement d'entreprise. Utilisez-la pour valider et hiérarchiser vos besoins.

	Inclus	Description
Connexion et fourniture		
Connexion à des sources de données traditionnelles		Connexion à des données stockées dans des bases de données relationnelles, des applications OLAP, des structures non relationnelles comme des fichiers plats, XML, des applications packagées courantes comme SAP, des applications cloud comme salesforce.com, des données semi structurées (p. ex. Excel), des données non structurées (p. ex. audio, vidéo) et des systèmes de messagerie. Conformité avec les normes du secteur (EDI).
Connexion au big data et à NoSQL		Intégration aux technologies big data (p. ex. Hadoop, Hbase, Hive), aux plateformes big data (p. ex. Cloudera, Hortonworks, MapR) et aux bases de données NoSQL (p. ex. MongoDB, Cassandra).
Connexion au cloud		Connexion à des environnements cloud (AWS, Microsoft Azure, Google Cloud Platform, etc.) et des entrepôts de données cloud (Amazon Redshift, Snowflake, Azure SQL, etc.).
Déplacement de données		Possibilité pour les utilisateurs des données de recevoir les données de multiples façons. Prise en charge du développement de données en masse, des services de données, de la fédération de données, de la capture des données de modification et de la réplique directe de données entre différentes sources de données.
Synchronisation des données		Prise en charge d'ETL (Extract, Transform, and Load) et d'ELT (Extract, Load, and Transform), de la livraison en temps réel et de la livraison basée sur les événements (données de déclencheur ou modifiées).

	Inclus	Description
Transformation		
Transformations simples		Calculs, conversions de types de données, manipulations de chaînes, agrégations, opérations automatiques de recherche et de remplacement.
Transformations avancées		Dimensions à évolution lente, normalisation des données, fonctionnalités avancées d'analyse et transformation en normes complexes (p. ex., EDIFACT, HL7, etc.).
Transformations personnalisées		Capacité à créer de nouvelles transformations personnalisées et à étendre les transformations existantes.
Enrichissement		La solution devrait pouvoir utiliser des données d'enrichissement provenant d'un large éventail de sources. Les données d'enrichissement peuvent avoir différents formats et schémas de fichiers internes et externes. Elles peuvent provenir de sources en ligne via des API de service, de partenaires commerciaux ou de fournisseurs de données.
Développement et modélisation de données		
Outils graphiques		Outils graphiques par glisser-déposer faciles à utiliser permettant de créer des processus et des transformations et de concevoir des modèles de données, des métadonnées et des flux de données. Représentation graphique des objets et connecteurs. Assistants d'automatisation des tâches courantes.
Outils de modèle d'entreprise		Outil non technique permettant une collaboration entre les utilisateurs techniques et métiers afin de structurer l'ensemble de la documentation pertinente et les éléments techniques appuyant le processus d'intégration des données.
Création et gestion des modèles de données		Possibilité de créer et tenir à jour des modèles de données. Utilisation d'outils graphiques pour définir les relations.

	Inclus	Description
Développement et modélisation de données		
Gestion des métadonnées		Découverte automatique des métadonnées. Possibilité de rechercher des métadonnées dans plusieurs sources et d'afficher leur lignage. Utilisation d'un référentiel de métadonnées unique dans toutes les fonctionnalités du produit, avec possibilité de partager les métadonnées et de les synchroniser de manière fluide entre des outils d'intégration des données et d'autres outils (p. ex. qualité des données, profilage des données et gestion des données de référence).
Règles métier et workflows		Possibilité de définir et gérer des règles métier et des flux d'exécution. L'exécution des processus peut être programmée pour être lancée immédiatement, à un moment précis ou en fonction d'un événement.
Versionnage		Les développeurs peuvent facilement contrôler la version des métadonnées, des routines, des processus, des transformations ou de tout autre objet utilisé dans le processus d'intégration. Ils ont ensuite la possibilité de consulter les modifications et de revenir à une version antérieure, au besoin.
Collaboration		Ensemble d'outils défini en fonction de chaque utilisateur (p. ex. utilisateurs métiers, développeurs et personnel chargé des opérations informatiques) et référentiel partagé regroupant l'ensemble des informations de projet et des métadonnées d'entreprise partagé par toutes les parties prenantes.
Test, débogage et optimisation		Outils destinés au test des processus avec des données dans l'outil graphique, puis au débogage interactif et à l'optimisation des performances.
Analyse d'impact		Outils graphiques destinés à comparer les processus, évaluer l'impact des modifications et afficher le lignage des données en cas de modifications.
Conformité aux normes		Pour faciliter le temps de montée en charge et permettre l'exploitation des ressources existantes, les produits doivent adopter des normes telles qu'Eclipse, Java, JDBC, ODBC, et des services Web.
Réutilisation		Possibilité de réutiliser des projets, des processus de métadonnées, les routines de nettoyage, de validation, d'enrichissement et autres routines fréquemment utilisées de façon rapide et simple.
Personnalisable		Possibilité de personnaliser les artefacts générés pour une souplesse optimale. Possibilité de créer des composants personnalisés. Simplicité de personnalisation et d'extension des transformations.

	Inclus	Description
Gouvernance des données		
Intégration à des outils de qualité des données		Fonctionnalité intégrée avec des outils assurant le profilage et le nettoyage des données, l'analyse et la normalisation des données, la mise en correspondance, la fusion et l'identification des enregistrements en double, puis leur rationalisation en fonction des exigences. Possibilité de définir des règles métier pour les données.
Intégration à des outils de profilage des données		Fonctionnalité intégrée avec des outils réalisant des analyses par colonne, des analyses de dépendances, des analyses de tendances et des analyses personnalisées.
Intégration à des outils MDM		Fonctionnalité intégrée ou intégration prédéfinie avec des outils créant une vue unifiée des informations et gérant cette vue de référence au fil du temps.
Rapports et tableaux de bord		Rapports prédéfinis et personnalisables affichant des indicateurs de qualité des données clés au fil du temps. Possibilité d'exporter les résultats dans divers formats y compris XML, PDF, HTML, etc. Système de reporting des métriques de qualité des données par tableaux de bord (Web) et fourniture de métadonnées aux systèmes de business intelligence (BI).
Déploiement		
Prise en charge du runtime multi-plateforme		Possibilité de déploiement fluide sur des systèmes Unix, Linux et Windows. Possibilité d'exécution sur site, dans le cloud et dans des environnements de virtualisation. Possibilité d'exécution dans des environnements de traitement distribué big data (MapReduce). Génération du code pour garantir la portabilité et les performances.
Répartition de charge et évolutivité		Fonctionnalités de clustering permettant de répartir la charge du serveur sur plusieurs machines. Possibilité de traiter des volumes de données très importants grâce au big data et à des data warehouses de plusieurs téraoctets.
Basculement		Possibilité de restaurer une transaction et de poursuivre le traitement en cas de défaillance du serveur, sans perte de données.
Exécution à distance		Possibilité d'exécuter les processus à distance sur divers systèmes d'exploitation à l'aide de la même configuration.

	Inclus	Description
Déploiement		
Services d'intégration des données		Possibilité de déployer tous les aspects de la fonctionnalité de runtime en tant que services au sein d'une architecture orientée services.
Compatibilité avec des middleware		Fonctionnalité intégrée avec systèmes MOM et ESB.
Prise en charge d'Hadoop		Déploiement de jobs MapReduce natifs directement dans un cluster sans avoir à installer d'outils ou de logiciels supplémentaires dans le cluster. Possibilité d'adapter des processus MapReduce au cluster sans modifier le code.
Supervision et administration		
Administration centralisée		Possibilité de contrôler et gérer tous les déploiements et ressources depuis un seul et même emplacement.
Supervision sur le Web		Possibilité de contrôler les ressources et déploiements depuis n'importe quel navigateur.
Rapports et tableaux de bord		Rapports prédéfinis et personnalisables affichant des indicateurs d'intégration des données clés. Le tableau de bord présente des informations des statistiques au fil du temps [p. ex. performances, volumes de chargement, indicateurs individuels pour chaque tâche subordonnée (p. ex. lecture et droits de base de données, temps de réponse des services d'enrichissement)].
Reporting et gestion des exceptions		Possibilité de définir, signaler et gérer des exceptions. Possibilité d'invoquer des processus spéciaux en cas de violations des règles d'intégration des données. Exemples : alerte par e-mail, SMS ou arrêt d'un processus.
Contrôles de sécurité		Mécanisme de sécurisation des messages en vol entre applications, ainsi que sécurité en fonction des utilisateurs/des rôles dans l'outil même, LDAP.
Interaction entre les utilisateurs métiers		La solution devrait offrir un environnement facile à utiliser pour les utilisateurs métiers leur permettant de respecter les KPI correspondant à l'intégration des données, p. ex. rapports PDF et portails Web.
Prise en charge du cloud		Possibilité de configurer, de déployer et d'arrêter une instance cloud, p. ex. Amazon EC2. Possibilité d'augmenter la capacité de calcul des processus d'intégration.

	Inclus	Description
Support		
Support complet		Vous disposez du support dont vous avez besoin, quand vous en avez besoin (p. ex., forums communautaires, base de connaissances Web, support par e-mail et par téléphone). Support technique stratégique 24 h/24, 7 j/7. Contrats de niveau de service pour le délai de réponse, la correction des bugs et les mises à jour de maintenance.
Formations		Formations publiques, en ligne et sur demande pour les débutants, les développeurs avancés et les administrateurs.
Services professionnels		Le fournisseur propose un large éventail de services de conseil : évaluation, stratégie et architecture, démarrage rapide, conception et développement, réglage, audit technique et offres personnalisées.

CHAPITRE 12

QUE RETENIR ?



QUE RETENIR ?

Le moment est venu pour votre entreprise de tirer parti de ses données d'entreprise et d'utiliser l'intégration des données pour obtenir des informations et des analyses stratégiques. Une stratégie d'intégration des données établie avec les bons outils, la bonne équipe et les moyens adéquats vous permettra d'atteindre cet objectif.

Talend vous propose pour cela une plateforme complète de gestion des données.

Contactez-nous dès aujourd'hui pour en savoir plus sur les solutions Talend et essayer Talend Cloud GRATUITEMENT pour avoir accès à tous les outils présentés dans ce guide.



Contactez-nous

<https://fr.talend.com/contact/>



Support technique

Consultez la communauté Talend



En savoir plus

fr.talend.com

Talend (NASDAQ : TLND), leader de l'intégration cloud, libère les données des infrastructures existantes et permet d'exploiter les données plus efficacement. La plate-forme Talend Cloud assure l'intégration des données dans les environnements sur site et cloud (public, privé ou hybride) et permet une meilleure collaboration entre les équipes IT et les équipes métiers. Son architecture native, ouverte et extensible lui permet de s'adapter rapidement aux innovations et de répondre à moindre coût aux demandes sans cesse croissantes des volumes de données, des utilisateurs et des cas d'usage.

Plus de 3 000 entreprises du monde entier ont confié leurs données à Talend, notamment GE, HP Inc. et Domino's. Talend a été reconnu leader dans son domaine par les principaux cabinets d'analyse et par les publications du secteur, y compris Forbes, InfoWorld et SD Times. Pour plus d'informations, rendez-vous sur le site www.talend.com.