

Services de données essentiels : composants de base pour l'établissement d'une valeur opérationnelle

WHITE PAPER:
QUALITÉ DES DONNÉES

David Loshin



Services de données essentiels : composants de base pour l'établissement d'une valeur opérationnelle

INTRODUCTION

FACE À LA DISPERSION CROISSANTE DES DÉCISIONS PRISES AU SEIN DES ORGANISATIONS ET LA HAUSSE DU VOLUME D'INFORMATIONS TRAITÉES ET DISTRIBUÉES AUPRÈS DES CONSOMMATEURS ET ADMINISTRÉS, UNE REPRÉSENTATION VISUELLE DES DIFFÉRENTES FACETTES DES TRANSACTIONS OPÉRATIONNELLES, TELLES QUE LE TYPE, LA DURÉE ET, PLUS PARTICULIÈREMENT LA LOCALISATION DE CES TRANSACTIONS, EST DE PLUS EN PLUS NÉCESSAIRE. AUGMENTER LA VALEUR ET L'UTILITÉ DES DONNÉES MÉTIERS POUR LES APPLICATIONS OPÉRATIONNELLES ET ANALYTIQUES EXIGE UN ENSEMBLE DE PRATIQUES SAINES DE GESTION DES DONNÉES, TELLES QUE LA GOUVERNANCE, LA QUALITÉ, L'ENRICHISSEMENT ET, DE PLUS EN PLUS, DES CAPACITÉS D'INTELLIGENCE GÉOSPATIALE, COMME LE GÉOCODAGE, LA CARTOGRAPHIE ET LE ROUTAGE.

LA NORMALISATION DES PRATIQUES SAINES DE GESTION DES DONNÉES LORS DU DÉVELOPPEMENT OU DE LA REFORTE DE L'ARCHITECTURE D'UNE ENTREPRISE RÉDUIT LES RISQUES ET RENFORCE LA COHÉRENCE ; LA CONFIANCE ORGANISATIONNELLE DANS LES DONNÉES SIMPLIFIE LE DÉVELOPPEMENT DES APPLICATIONS ET ACCROÎT LA FIABILITÉ DES DONNÉES. BIEN QUE DIFFÉRENTES CONJECTURES EXISTENT QUANT À L'INTÉGRATION DES DONNÉES MAÎTRES (MDM) COMME MÉCANISME D'AMÉLIORATION DE LA QUALITÉ DES DONNÉES, CE DOCUMENT A CHOISI D'EXPLORER LA FAÇON DONT UN ENSEMBLE CONSOLIDÉ DE SERVICES DE DONNÉES À VALEUR AJOUTÉE EST EN MESURE DE RÉDUIRE LES COÛTS OPÉRATIONNELS ET D'AUGMENTER L'EFFICACITÉ TOUT EN AMÉLIORANT LA QUALITÉ ET L'UTILITÉ DES DONNÉES D'ENTREPRISE. UNE SEULE COUCHE DE SERVICES DE DONNÉES EST SUSCEPTIBLE DE RÉDUIRE LA COMPLEXITÉ DU CYCLE DE GESTION DES DONNÉES ; LA NORMALISATION DES MÉTHODES D'UTILISATION DES OUTILS DE GESTION DES DONNÉES AINSI QUE L'APPLICATION DES RÈGLES VEILLE DURABLEMENT À LA FIABILITÉ ET LA VALEUR DES INFORMATIONS.

ALORS QUE DE PLUS EN PLUS D'ORGANISATIONS MIGRENT VERS UNE ARCHITECTURE ORIENTÉE SERVICES, L'INTÉGRATION TRANSPARENTE DES DONNÉES DEVIENT UN ÉLÉMENT FONDAMENTAL DES APPLICATIONS D'ENTREPRISE.

Facteurs de valeur opérationnelle et fiabilité de l'information

Les dirigeants de toutes les organisations reconnaissent la nécessité et l'importance de disposer d'informations de valeur quelles que soient les circonstances. Et certains événements en particulier imposent des changements immédiats en termes d'amélioration de la valeur et de l'utilité des données. En effet, les fusions et acquisitions de sociétés, les modifications dans les impératifs de conformité réglementaire, l'attrition des clients, l'augmentation notable des activités vers les centres d'appel, les migrations de système, l'amélioration des processus de gestion des contacts avec les clients, l'introduction des CRM ou des ERP ou encore les initiatives MDM sont des exemples d'événements dont le succès dépend des ajustements organisationnels effectués au niveau des pratiques de gestion des données.

Chacun de ces événements traduit un besoin d'établissement de processus de gestion solides ainsi que des pratiques de gestion des informations correspondantes en vue d'améliorer l'expérience client tout en maintenant des économies d'échelle. Au lieu de foncer tête baissée dans l'achat de davantage de technologie pour résoudre un problème en espérant que quelque chose se produise, une approche plus réfléchie consiste à rechercher un retour sur les investissements technologiques, ROI qui n'est réalisable que lorsque les pratiques nécessaires de gestion de l'information sont supportées des services de données essentiels, mises en œuvre une seule fois, mais déployées à maintes reprises de façon cohérente.

Cette approche réunit les pratiques fondamentales de la gestion de l'information avec les services qui enrichissent la valeur et l'utilité des données opérationnelles, tels que la qualité des données, l'enrichissement des données et l'intelligence géospatiale. La mise en œuvre de ces services essentiels exige

- Le soutien des processus de production de l'information qui prennent en charge plusieurs consommateurs de données
- La normalisation des outils et des techniques dans toute l'entreprise
- La normalisation de la définition et de l'utilisation des règles

- La normalisation des services d'analyse spatiale et d'intelligence géospatiale associés aux activités et aux transactions

La visualisation ainsi que la normalisation du catalogue des services de données à valeur ajoutée permettent d'éliminer les situations dans lesquelles les mêmes produits ou des produits concurrents sont achetés plusieurs fois ou dans lesquelles différents groupes mettent en œuvre les mêmes techniques sans qu'il y ait concertation. En fin de compte, cette approche permettra de solidifier et de mutualiser l'investissement en outils de gestion des données et de multiplier le rendement des investissements.

Services de données essentiels

Bien que les critères caractérisant un service comme essentiel ne soient pas complètement rigides, il est tout de même possible de les synthétiser selon la façon dont ils supportent les meilleures pratiques de gestion de l'information d'une entreprise. Par exemple, il est entendu que tout service permettant d'enrichir l'interopérabilité des processus de gestion, de réduire la complexité de développement des applications, de normaliser les techniques, d'unifier la définition et l'application des règles ou encore d'améliorer la qualité et l'utilité de l'information client est considéré comme un « service essentiel ».

Toutefois, pour garantir le respect de bonnes pratiques de gestion de l'information, une liste judicieuse doit au moins inclure les services suivants d'intégration des données, de qualité des données et d'intelligence géospatiale :

- Accès aux données et intégration
- Profilage des données
- Analyse syntaxique
- Normalisation
- Résolution d'identité
- Couplage et fusion des enregistrements
- Nettoyage des données
- Enrichissement des données
- Géocodage

Services de données essentiels : composants de base pour l'établissement d'une valeur opérationnelle

- Cartographie et analyse spatiale
- Calculs d'itinéraires
- Inspection et surveillance des données

Accès aux données et intégration

Les techniques employées dans le développement des entrepôts de données sont de nos jours omniprésentes, particulièrement du fait de l'interopérabilité croissante des systèmes. À mesure que davantage d'entreprises procèdent à la refonte de leur système d'informations en vue de faire communiquer, voire de consolider les différents silos de données de l'entreprise, les services d'intégration des données sous-jacentes, rendant possible l'accès transparent aux données depuis différentes sources ainsi que la diffusion des données vers différentes cibles, enregistrent une expansion croissante. La capacité des serveurs distants à accéder aux données depuis leur source favorise la mise en œuvre et le déploiement efficaces des meilleures pratiques de gestion des données à implémenter au moyen des services essentiels présentés dans ce document. Les services d'intégration reposent sur la capacité de connexion aux infrastructures courantes supportant les données parallèlement à l'analyse syntaxique et aux services de normalisation visant à transformer les données extraites en des formats propices à leur utilisation.

Dans l'entreprise moderne, le partage, la fédération et la synchronisation des données sont à présent la norme. Alors que de plus en plus d'organisations migrent vers une architecture orientée services, l'intégration transparente des données devient un élément fondamental des applications d'entreprise.

Profilage des données

Une évaluation objective de la qualité des données est un processus d'analyse et de découverte. Il repose lui-même sur une revue objective des valeurs se trouvant dans le jeu de données cible. Une analyse quantitative statistique permet aux analystes de repérer les données potentiellement défectueuses, notamment les valeurs aberrantes ou manquantes, les éléments de données qui ne se conforment pas aux formats ou aux listes de valeur définies ou qui sont incompatibles par rapport à d'autres éléments.

Le profilage de données consiste en un jeu d'algorithmes permettant l'analyse statistique et l'évaluation de la qualité des données au sein d'un jeu de données ainsi qu'en l'exploration des relations existant entre des éléments de données ou entre des jeux de données. Le profilage rend possible l'identification des données erronées ainsi que la communication de ces cas aux experts chargés de confirmer l'existence de problèmes de données. Un outil de profilage analyse les valeurs d'une colonne et propose une distribution statistique des valeurs en soulignant le type et les utilisations potentielles de chaque colonne. Une analyse inter-colonne est en mesure d'indiquer les dépendances tandis qu'une analyse inter-table explore les jeux de valeurs identiques en vue d'identifier les clés étrangères entre les entités. Le profilage peut permettre de mettre en évidence les règles de gestion incorporées aux jeux de données et susceptibles d'être utilisées dans l'inspection et la surveillance continues.

Pour paraphraser Lord Kelvin, on ne peut pas améliorer ce que l'on ne peut pas mesurer. Le profilage offre un mécanisme concret d'analyse, d'évaluation, de quantification et de mesure de la qualité des données sur plusieurs dimensions, dont l'intégralité, la validité et la rationalité. Le profilage est désormais reconnu comme la première étape dans les projets de migration des données, d'intégration des données ainsi que les projets MDM pour mieux appréhender les problématiques des jeux de données sources et pour faciliter la mise en place des transformations qui permettront de nettoyer les données lors de leur déplacement vers la cible.

Analyse syntaxique

Dans de nombreux cas, il est escompté que les valeurs se conforment aux dimensions et formats prédéfinis. Toutefois, il se peut que des divergences ou des ambiguïtés dans la représentation des valeurs désorientent les applications ainsi que les utilisateurs. Prenons, par exemple, les valeurs suivantes qui peuvent prêter à confusion : {Californie, CA, Calif., US-CA, Cal, 06}. Certaines de ces valeurs se servent de caractères, d'autres de chiffres et d'autres encore d'une ponctuation ou de caractères spéciaux. Dans la plupart des cas, un être humain serait capable de les lire et de se rendre compte qu'elles représentent toutes la même valeur

UN ENSEMBLE CONSOLIDÉ DE SERVICES DE DONNÉES À VALEUR AJOUTÉE PEUT RÉDUIRE LES COÛTS OPÉRATIONNELS ET AMÉLIORER LE RENDEMENT TOUT EN AUGMENTANT LA QUALITÉ ET L'UTILITÉ DES DONNÉES D'ENTREPRISE

conceptuelle, soit l'état de Californie. Toutefois, pour automatiser ce processus d'identification de ces valeurs ou détecter si des enregistrements en doubles existent, ces valeurs doivent être analysées puis transformées en format standard.

La description des différents schémas de composant et de format utilisés pour représenter un objet de donnée (nom de personne, description de produit, etc.) prend en charge le service essentiel d'analyse syntaxique, qui sert à déterminer si une valeur se conforme à un schéma reconnaissable dans le cadre du processus d'évaluation, de correspondance et de nettoyage. L'analyse syntaxique par format permet la reconnaissance automatique et la normalisation ultérieure des composants significatifs, tels que l'indicatif d'un numéro de téléphone ou les différents éléments du nom d'une personne. Ce type de service incorpore, par nécessité, l'aptitude à définir des jeux de formats des listes de valeurs appartenant à des domaines définis ou encore des expressions régulières.

Lorsque la même information est représentée plusieurs fois sous différentes formes, il devient plus difficile de déterminer si plusieurs enregistrements (du même jeu de données ou de jeux de données différents) se rapportent à la même entité. Mais, bien avant de pouvoir appliquer des algorithmes complexes de couplage des enregistrements, il est indispensable d'analyser le texte des champs critiques afin d'identifier, de séparer puis de réordonner les unités lexicales clés porteuses des informations d'identification. De concert avec la normalisation, l'analyse syntaxique permet de réduire les divergences entre les enregistrements et de simplifier la mise en correspondance ainsi que le couplage.

Normalisation

L'analyse syntaxique a recours à des formats prédéfinis ou des expressions régulières, le tout géré par un moteur de règles accompagné d'une consultation de tables pour faciliter la distinction entre les valeurs correctes et les valeurs incorrectes. Lorsque des formats sont reconnus, des règles et des actions peuvent être déclenchées en vue de convertir les entrées en une forme utilisable, pour normaliser la représentation, ou afin de corriger les valeurs. Pour reprendre notre exemple, chacune des valeurs

représentant l'état de Californie {Californie, CA, Calif., US-CA, Cal, 06} peut être normalisée en se servant de l'abréviation à deux caractères standard du service postal américain, soit CA.

Le service de normalisation repose sur le service d'analyse syntaxique qui peut être combiné à une bibliothèque de listes de données permettant de séparer les valeurs en plusieurs composants et de réorganiser les composants dans un format normalisé. La normalisation peut, en outre, remplacer des mots entiers en abréviations ou des abréviations en mots entiers, transformer des surnoms en noms normaux, traduire (par exemple, de l'espagnol à l'anglais) et corriger les erreurs d'orthographe courantes.

La normalisation procure deux sortes de valeur. En premier lieu, en mettant en correspondance des représentations variables avec des formats prédéfinis, le processus de normalisation permet de standardiser les différentes représentations et de nettoyer efficacement les données imparfaites. En deuxième lieu, la mise en correspondance d'une valeur avec une représentation normalisée (sans qu'elle fasse nécessairement l'objet d'un nettoyage spécifique) réduit les divergences en termes de structure des valeurs afin d'améliorer le couplage d'enregistrements à des fins de déduplication et de consolidation des données en une donnée unique principale. La normalisation des surnoms (tels que "Bob," "Rob," ou "Bobby" en "Robert") accentue les possibilités de couplage de deux enregistrements représentant la même entité, ce qui améliore de façon significative la résolution unifiée indispensable à la qualité des données et à un programme de MDM.

Résolution d'identité

Du fait que les systèmes opérationnels ont évolué en une suite d'applications d'entreprise, il n'est pas rare que plusieurs instances de données issues de divers systèmes réfèrent différemment à la même entité réelle. En outre, le désir de consolidation et de couplage des données portant sur les mêmes concepts de gestion pourrait laisser penser qu'un enregistrement réel n'existe pas alors qu'en fait il existe bien. Ces deux problèmes représentent en fin de compte le même enjeu : être capable de comparer

Services de données essentiels : composants de base pour l'établissement d'une valeur opérationnelle

les données d'identification au sein d'un couple d'enregistrements pour déterminer les similarités ou pour y distinguer les entités représentées.

Ces deux questions sont prises en charge par un processus appelé résolution d'identification dans lequel le degré de similarité entre deux enregistrements est constaté, la plupart du temps en fonction d'une correspondance approximative pondérée dans un ensemble de valeurs d'attribut. Si la note est au-dessus d'un seuil spécifique, les deux enregistrements sont considérés comme identiques et sont présentés au client final comme représentant la même entité. Grâce à la résolution d'identité, il est possible de distinguer si les divergences constatées indiquent que les enregistrements sont connectés, et donc nettoyables, ou suggèrent que les enregistrements représentent véritablement deux entités distinctes. Du fait que la comparaison de tous les enregistrements entre eux est gourmande en ressources informatiques, les services de données se servent d'algorithmes pour réunir en ensembles plus petits les enregistrements les plus susceptibles de contenir des correspondances et ainsi réduire le temps de calcul.

La résolution d'identité est un composant essentiel de la plupart des applications de qualité des données, de gestion de données permanentes et d'informatique décisionnelle. Le désir de disposer d'une visualisation du client dans sa globalité ou d'un catalogue de produits complet repose sur la capacité à trouver tous les enregistrements porteurs d'informations de chaque entité et à les résoudre dans une vue unifiée. L'analyse syntaxique et la normalisation sont les étapes préliminaires, mais la technique principale réside dans la résolution d'identité.

Couplage, fusion et consolidation des enregistrements en double

La résolution d'identité propose le fondement d'un service de données essentiel plus sophistiqué : l'analyse et l'élimination des enregistrements en double. L'identification d'enregistrements similaires disposant des mêmes données indique qu'il s'agit probablement de doublons pouvant être nettoyés ou supprimés. L'identification d'enregistrements similaires dans différents jeux peut suggérer un couplage entre les jeux, ce qui facilite la fusion des enregistrements à des fins de nettoyage

ainsi que les initiatives de MDM ou d'intégration de données clientes.

La résolution d'identité conduit au processus de fusion et de consolidation en vue d'établir une vue unique du client. Alors que certaines applications de gestion fonctionnent en consultant un ensemble d'enregistrements porteurs d'informations, l'automatisation du processus de fusion permet, quant à lui, de sélectionner les meilleures valeurs figurant dans ces enregistrements et de créer la « meilleure copie » possible des données susceptibles de supporter les processus opérationnels et analytiques.

Nettoyage des données

Après avoir déterminé qu'une valeur de donnée ne se conforme pas aux attentes de l'utilisateur final, l'analyste a le choix entre plusieurs démarches. L'une d'elles consiste à convertir les données dans un format qui correspond au niveau d'acceptabilité de l'utilisateur opérationnel. Ce service de nettoyage des données repose sur l'analyse syntaxique, la normalisation et les services d'enrichissement ainsi que sur la résolution d'identité et le couplage d'enregistrements. En analysant les valeurs et en établissant une détection des formats d'erreur courants, le service de nettoyage des données est en mesure d'appliquer des règles de transformation permettant d'imputer les valeurs de donnée, de rectifier les noms ou les adresses, d'éliminer les données superflues ou sans intérêt et de fusionner les enregistrements en double.

Même si une meilleure pratique en matière de qualité des données consiste à déterminer la source des erreurs et à éliminer les causes fondamentales, l'équipe de qualité des données n'a souvent pas le contrôle administratif nécessaire pour modifier les applications de production existantes. Par conséquent, il peut s'avérer nécessaire de nettoyer les données directement pour fournir des données qui conviennent aux consommateurs de données ; c'est dans ce domaine qu'intervient le service de nettoyage.

Enrichissement des données

Une autre démarche en termes d'amélioration de la valeur des données consiste à enrichir les informations, c'est-à-dire à ajouter des renseignements aux enregistrements existants.

LE PROFILAGE DES DONNÉES EST DÉSORMAIS RECONNU COMME ÉTANT LA PREMIÈRE ÉTAPE DE LA MIGRATION DES DONNÉES, DE L'INTÉGRATION DES DONNÉES ET DES PROJETS DE GESTION DE DONNÉES PERMANENTES

L'enrichissement des données, qui repose sur l'analyse syntaxique, la normalisation et le couplage des données, est un processus d'amélioration des données qui permet d'ajouter des informations provenant de jeux de données tiers (tels que la normalisation des noms, les importations de données démographiques et l'ajout d'informations sur les foyers). Le service d'enrichissement est grandement simplifié lorsque le prestataire se trouve en partenariat avec des fournisseurs dont les données peuvent servir à la fois de référentiel avec lequel les instances de données sont mises en correspondance et de ressource d'enrichissement. Une activité d'enrichissement type, telle que la normalisation et le nettoyage des adresses, repose sur l'analyse syntaxique, la normalisation et la présence d'informations tierces.

Très peu de secteurs ne sont pas en mesure de tirer parti de sources de données supplémentaires pour enrichir les informations de leur entreprise. Qu'il s'agisse, entre autres, de normalisation ou de correction des adresses en vue de réduire les coûts de livraison et d'améliorer les taux de réponse directs, d'ajout d'informations géodémographiques procurant des renseignements sur le profil et la segmentation des clients, de données sur le crédit servant à évaluer et réduire les risques, l'enrichissement des données est un processus qui optimise les décisions dans des contextes opérationnels et analytiques.

Emplacement géographique et géocodage

Le géocodage est un processus d'intelligence géospatiale permettant de faire correspondre un emplacement conceptuel avec des coordonnées spécifiques sur une carte. Une donnée géocodée consiste en un couple de coordonnées correspondant à la latitude et la longitude d'un emplacement. Lorsque le service de géolocalisation reçoit une adresse, il est en mesure de fournir la longitude et la latitude de l'adresse. En les greffant au service d'enrichissement des données, les valeurs peuvent être déterminées (géocodées) à partir d'un nom, d'une rue, d'un code postal et d'une intersection. Ce service fonctionne également dans le sens inverse c'est-à-dire que des coordonnées de latitude et longitude permettent de trouver une adresse.

De nombreux secteurs fondent leurs processus d'évaluation des risques sur la localisation géographique, plus

particulièrement les secteurs financiers et l'assurance. Le secteur du commerce procède à un examen constant des positions géographiques pour déterminer l'emplacement optimal des nouveaux magasins, filiales, etc. En outre, la mobilité croissante et la capacité de diffuser des informations à tout moment et partout suggèrent que les applications de gestion devront être de plus en plus sensibles à la « problématique géographique ».

Cartographie et analyse spatiale

Les applications qui effectuent un suivi des activités sont souvent supposées pouvoir noter la position des informations spatiales (par exemple, emplacement, point d'intérêt itinéraire ou adresse) sur une carte. Le service cartographique permet la modification, l'analyse spatiale et l'affichage des informations spatiales sur une carte. En outre, il se peut que l'utilisateur final souhaite consulter davantage d'emplacements ou points d'intérêts à proximité des positions identifiées. Un service d'annuaire fait correspondre l'analyse spatiale avec les données de référence fournies pour effectuer des recherches dans les catégories définies (guide des restaurants ou liste des parcs publics, par exemple) et renvoyer une liste des résultats les plus pertinents.

L'analyse spatiale, grâce à un examen statistique et quantitatif des phénomènes et des activités dans l'espace, permet de déduire des connaissances sur le contexte géographique des informations existantes. Savoir où habite un individu ou connaître l'emplacement d'un événement fournit un vecteur de données ; l'analyse spatiale permet de répondre à des questions plus intéressantes telles que :

- Combien de clients similaires se trouvent dans le rayon spécifié ?
- Quel est le taux de mortalité par maladie dans différentes régions du pays ?
- Dans quels endroits, les distributeurs sont-ils plus à même de mieux servir une région ?
- Combien de fermetures y a-t-il eu précédemment dans un quartier particulier ?
- Combien de clients se trouvent dans chaque code postal ?

Services de données essentiels : composants de base pour l'établissement d'une valeur opérationnelle

- Le revenu médian d'une région a-t-il changé au cours d'une période donnée ?
Chaque événement ou transaction est relié à un emplacement géographique. L'analyse spatiale permet de rechercher des tendances basées sur la localisation et d'utiliser les résultats pour améliorer le rendement.

Calculs d'itinéraires

Le routage est un service d'intelligence géospatiale permettant de planifier un itinéraire entre deux points, de déterminer le chemin optimal entre plusieurs points et de créer des matrices et polygones de temps passé au volant. Le service de routage est en mesure de calculer un itinéraire en fonction de différents critères, tels que la distance et la durée, et de fournir des résultats optimisés en termes de distance parcourue ou de temps passé au volant. Grâce au service de routage, les entreprises sont à même de maintenir de hauts niveaux de service, un bon rapport coût-efficacité en matière d'expédition, d'identifier les zones de chalandise et d'améliorer les prises de décision.

D'un point de vue opérationnel, le service de routage permet d'optimiser la logistique et la livraison et d'améliorer le service clients. La détection de l'itinéraire le plus adapté réduit les coûts énergétiques et peut permettre d'améliorer le planning afin d'augmenter les cadences de livraison. Parallèlement, des délais de livraison plus courts renforcent le service clients et procurent plus de satisfaction.

Inspection et surveillance

Du fait que le profilage des données met en lumière les règles utilisées au sein et entre les différents processus de gestion, l'analyse des données est en mesure de documenter les règles et de confirmer leur importance de concert avec les experts en la matière. Ces règles décrivent essentiellement les attentes de l'utilisateur final et peuvent servir à mesurer et surveiller la validité des données. L'audit et la surveillance des règles définies dans le domaine de la qualité des données offrent une évaluation proactive de la conformité aux attentes. Les résultats de ces audits viennent alimenter la métrologie qualitative permettant de renseigner les tableaux de bord et les cartes de notation.

Le service d'audit et de surveillance repose sur les services d'analyse syntaxique et de profilage en vue de valider les données par rapport à un ensemble de règles prédéfinies (ou constatées). Cette inspection des données permet de notifier les analystes lorsque des instances de données ne se conforment pas aux attentes de qualité prédéfinies ainsi que de référencer les mesures en regard desquelles les inspections sont comparées.

Nous avons déjà vu que le profilage des données permet l'analyse et la constatation des problèmes potentiels susceptibles d'affecter la réussite opérationnelle. Mais, en prenant en charge l'inspection et la surveillance des données, le profilage remplit une condition de gestion critique, c'est-à-dire l'assurance continue de la qualité des données. Le service de profilage des données encourage la gouvernance des données opérationnelles et fournit aux responsables de données les défaillances pouvant être examinées et rectifiées avant toute répercussion négative sur le métier.

Structuration en couches des services d'information sur les activités

De plus en plus, les entreprises sont à la recherche d'opportunités visant à réduire les coûts inhérents à la conception, le développement et la maintenance de la technologie développée en interne comme acquise en externe. L'alliance de services de données essentiels avec des produits d'informations agrégées dans des services normalisés d'information sur les activités permet de prendre en charge différentes applications au travers de l'entreprise. Ces services répondent généralement aux besoins qui se recoupent sur un ensemble d'applications de gestion différentes. L'apport de versions normalisées permettra de réduire les coûts tout en introduisant un bon rendement opérationnel. Voici quelques exemples courants :

- **Validation de la qualité des données** – Lorsque des données fréquemment utilisées sont acquises ou créées par plusieurs canaux (par exemple, données clients saisies sur un site Web ou transcrites par un agent de centre d'appel), des divergences risquent d'être présentes au niveau de l'orthographe et de l'ordre des noms ou il risque de manquer des informations (absence de désignation ou de titre honorifique). Une application incohérente des règles peut entraîner la duplication des données ou empêcher l'identification des enregistrements dans les jeux de données existants.

Pour faire face aux incohérences potentielles, un service de validation de la qualité des données peut être formé à l'aide des services d'analyse syntaxique, de normalisation, de nettoyage et de profilage. Ce service de validation est en mesure d'appliquer des jeux de normalisation de données ainsi que des règles de validation à plusieurs canaux d'acquisition de données au travers de l'entreprise. La normalisation des contraintes de qualité et des transformations au niveau de tous les points de

PITNEY BOWES BUSINESS INSIGHT PROPOSE UNE SOLUTION D'ENTREPRISE QUI REND POSSIBLE LE CONTRÔLE CENTRALISÉ ET AMÉLIORE LES PERFORMANCES

saisie et d'échange des données réduit significativement les risques de duplication ou de discordance et permet de disposer de données cohérentes dans toute l'entreprise.

- **Évaluation des risques en matière d'emplacement foncier** – Tout secteur qui dépend de l'évaluation des risques afférents aux emplacements et aux biens immobiliers pourrait bénéficier d'un tel service. Par exemple, les sociétés d'assurance évaluent les risques associés aux souscriptions de police d'assurance maison, notamment les risques de catastrophes naturelles telles qu'ouragans et tremblements de terre, la criminalité ainsi que la disponibilité et l'efficacité des services d'urgences. Les souscripteurs de prêts s'intéressent plutôt à la classification (maison vs appartement), l'occupation du logement par le propriétaire, les biens immobiliers comparables ainsi que les inconvénients du quartier. Toutefois, les différentes applications de gestion au sein de la même entreprise sont susceptibles d'appliquer leurs règles d'évaluation des risques de façon dissemblable, au moyen de techniques et d'outils distincts.

Un service d'évaluation des risques en matière d'emplacement foncier permettrait d'incorporer l'analyse syntaxique et l'enrichissement des données afin de fournir une adresse normalisée. Cette adresse normalisée peut être soumise aux services de géocodage et de cartographie puis combinée aux jeux de données agrégées contenant les risques portant sur les biens, les zones de catastrophe naturelle et autres attributs relatifs à l'emplacement géocodé.

- **Gestion des données** – Dans de nombreuses entreprises, on ressent un intérêt croissant pour l'opérationnalisation des politiques et des procédures de gouvernance des données. Cette tendance se traduit par l'introduction d'un rôle d'intendant des données chargé de veiller à ce que les informations répondent bien aux besoins des employés. Les attentes en matière de qualité des données sont définies par les utilisateurs finaux de l'information, mais puisque les jeux de données créés dans un contexte précis sont utilisés par plusieurs processus de gestion, les besoins collectés doivent être imposés au niveau de toutes les étapes de traitement au travers de plusieurs processus de bout en bout.

Les services de gestion des données peuvent servir à instituer des inspections et une surveillance appliquées de façon cohérente à la qualité des données au sein des processus de gestion permanents. Les règles de qualité des données peuvent être définies une seule fois puis déployées au moyen des services de profilage en vue de valider des enregistrements individuels ou des jeux d'enregistrement, où ils se trouvent ou lors de leur transmission dans le cadre des activités. Un service d'intendance commun inspecte automatiquement la validité des données et notifie l'intendant lorsque les attentes ne sont pas remplies. En outre, des statistiques cumulatives peuvent être collectées et mises à disposition au moyen de la carte de notation et du tableau de bord.

- **Profilage démographique** – L'un des aspects du marketing ciblé fait intervenir une combinaison de données fiables sur les individus et de données géodémographiques. Pour reprendre le proverbe « Qui se ressemble s'assemble », la compréhension des caractéristiques générales associées à l'emplacement d'un prospect instruit les approches marketing en terme de ciblage.

Suivant le nom et l'adresse d'un prospect, un service de profilage démographique est en mesure d'analyser et de normaliser le nom et d'employer la résolution d'identité pour établir une correspondance avec des clients connus ou encore pour identifier des entités semblables dans un jeu de données client connu. L'adresse est ensuite analysée puis normalisée. Les données agrégées associées à l'emplacement cartographié peuvent servir à ajouter des données démographiques portant sur la région de l'emplacement. La combinaison des données démographiques et de l'enregistrement d'un prospect offre à l'équipe commerciale des recommandations de ciblage basées sur le profil matérialisé.

- **Gestion des impôts juridictionnels** – Un autre aspect de la gestion de la conformité et des risques fait intervenir la détermination des taux d'imposition par juridiction. Ce type de service est important non seulement pour l'analyse des risques fonciers, mais aussi pour les entreprises dont les responsabilités fiscales sont liées à plusieurs emplacements géographiques.

Services de données essentiels : composants de base pour l'établissement d'une valeur opérationnelle

Ce type de service combine les services de normalisation et de nettoyage, les services de géocodage et d'intelligence géospatiale ainsi que les données agrégées afin de fournir les renseignements fiscaux régionaux, départementaux, municipaux et de l'arrondissement associés à une adresse.

de conception ou d'achat. En d'autres termes, les critères d'évaluation des solutions doivent incorporer le degré de prise en charge des services essentiels de qualité des données, les services d'intelligence géospatiale et l'intégration avec les jeux de données à valeur ajoutée.

Considérations

Les exemples suivants illustrent les services de gestion pouvant être conçus au moyen des services de données essentiels. Il est à noter que les trois caractéristiques dont ils font preuve semblent être un sous-produit de la combinaison des services de données essentiels :

- L'analyse syntaxique, la normalisation et les services d'inspection et de surveillance pour normaliser et nettoyer les données
- La combinaison de données nettoyées et d'intelligence géospatiale pour préparer l'enrichissement des données
- L'enrichissement et l'amélioration des valeurs en y adjoignant des informations provenant de jeux de données agrégées

S'il est clair que des pratiques de gestion saines, telles que la gouvernance des données, l'assurance de la qualité des données et l'enrichissement des données, sont en mesure d'améliorer les résultats, alors il est important de considérer les approches en matière de déploiement des services d'informations normalisées au moyen d'une démarche orientée vers les services de données essentiels. Une évaluation des besoins en informations opérationnelles permet de mettre en évidence les services qui, si normalisés au niveau de toute l'entreprise, permettraient d'améliorer la qualité et la cohérence tout en diminuant les coûts opérationnels.

Une dernière pensée – il peut y avoir de nombreuses façons de concevoir et de déployer les services essentiels : conception interne, achat d'une suite de produit ou externalisation. Toutefois, s'il est clair que les services d'information critiques reposent sur une combinaison de qualité des données, d'intelligence géospatiale et de jeux de données à valeur ajoutée, la qualité de ces facettes doit alors être incorporée dans la décision, qu'elle soit

PITNEY BOWES BUSINESS INSIGHT : RÉPONSES ET SOLUTIONS

ÉTATS-UNIS

One Global View
Troy, NY 12180
Tél. : 518.285.6000
800.327.8627
www.pbinsight.com
pbbi.sales@pb.com

France

Immeuble Le Dièse
134, rue Danton
92593 Levallois-Perret Cedex
+33 1 46 17 54 00
www.pbinsight.fr
info.pbbifrance@pb.com

EUROPE/ROYAUME-UNI

Minton Place, Victoria Street
Windsor, Berkshire SL4 1EG
Royaume-Uni
+44.1753.848.200
www.pbinsight.co.uk
pbbi.europe@pb.com

ASIE PACIFIQUE/AUSTRALIE

Level 7, 1 Elizabeth Plaza
North Sydney NSW 2060
Australie
+61.2.9437.6255
pbbi.australia@pb.com
pbbi.singapore@pb.com
pbbi.china@pb.com