



Une architecture de référence pour l'analytique en libre-service

Équilibrer l'agilité et la gouvernance

Par Wayne Eckerson, en collaboration avec Barry Devlin
Septembre 2016 - Extrait du rapport initial

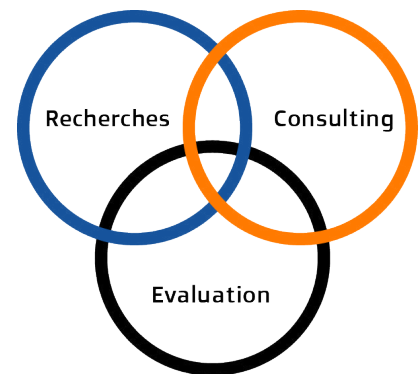
À propos de l'auteur



Wayne Eckerson est un expert éminent dans le domaine du décisionnel et de l'analytique depuis le début des années 90. Il est un consultant recherché, un conférencier réputé et un formateur spécialisé qui allie une pensée critique, une rédaction claire et un exposé probant sur des sujets complexes. Wayne a mené de multiples études de recherche novatrices, présidé de nombreuses conférences et écrit deux livres très lus sur les tableaux de bord de performance et l'analytique. Wayne est le fondateur et le consultant principal d'Eckerson Group.

À propos d'Eckerson Group

Eckerson Group est un cabinet de recherche et de conseil qui aide les entreprises et les professionnels à utiliser les données et la technologie pour favoriser une meilleure analyse et prise de décision. Avec ses rapports et ses services de conseil, le cabinet accompagne les entreprises pour qu'elles optimisent leurs investissements en données. Ses chercheurs et ses consultants possèdent tous une longue expérience en la matière et des compétences exclusives pour aider les entreprises et les directeurs techniques à réussir en termes de décisionnel, d'analytique, de gestion des données, de gouvernance, de gestion de la performance et de science des données.



Sponsors du rapport

Ce rapport n'a été possible que grâce au soutien généreux de nos sponsors : Dundas Data Visualization, Information Builders, Looker Data Sciences, TimeXtender, ThoughtSpot.

Infographie

Ce rapport s'accompagne d'une infographie intitulée « [Analytique en libre-service : six étapes pour réussir](#) ».



Résumé

Malgré sa promesse d'affranchir les utilisateurs de leur dépendance à l'égard de la DSI, l'analytique en libre-service est un objectif ardu. De nombreuses entreprises qui ont déployé l'analytique en libre-service se retrouvent submergées d'un flot de rapports conflictuels, de spreadmarts, de systèmes de reporting déviants et d'autres silos de données. Ces entreprises ont appris que le but du libre-service ne réside pas dans un affranchissement illimité de l'informatique, mais plutôt dans un partenariat qui équilibre la liberté et le contrôle, l'agilité et les normes, la gouvernance et l'autonomie.

Pour réussir dans l'analytique en libre-service, les entreprises ont besoin d'une architecture de référence qui cartographie les utilisateurs fonctionnels, la technologie et les développeurs dans une chaîne de l'information conçue pour transformer les données et l'analyse en action. L'architecture assemble les outils en libre-service dans la chaîne afin que les utilisateurs fonctionnels désignés puissent obtenir leurs propres données et créer ou modifier des rapports pour répondre à des questions immédiates sans attendre que la DSI ne génère des jeux de données ou des rapports personnalisés.

Bien que la révolution du libre-service actuelle soit impossible sans technologie, les clés de son succès sont d'ordre organisationnel. Outre une architecture de libre-service dirigée, les entreprises doivent établir des équipes et des passerelles de gouvernance des données, créer des organisations fédérées avec des développeurs en décisionnel communs et fournir une formation et un support continus.



L'ère du libre-service gouverné : promesse et risques

La promesse

Nous vivons à l'ère du libre-service. Nous devons à présent remplir nous-mêmes nos réservoirs, organiser nos voyages et nous inscrire en ligne pour tout. Grâce à la nouvelle technologie, nous pouvons avec une facilité inédite publier nos propres livres, louer nos maisons, préparer nos impôts et convertir nos voitures en source de revenus, entre autres. Nous avons remplacé les intermédiaires humains par des technologiques, ce qui nous procure plus de contrôle et de commodité à un coût inférieur.

La modernité du libre-service a également envahi le décisionnel (ou la BI). Depuis les années 90, les leaders de la BI ont cherché à doter les utilisateurs fonctionnels d'outils en libre-service pour leur permettre de créer leurs propres rapports et tableaux de bord. Encore récemment, nombre de ces efforts ne portaient pas leurs fruits. Les outils n'offraient pas une facilité d'utilisation suffisante sans assistance pour la plupart du personnel opérationnel. De plus, la DSI (direction des systèmes d'information) se montrait réticente à céder le contrôle aux utilisateurs fonctionnels, par crainte qu'ils ne créent un chaos de données.

Nouvelles technologies. La situation évolue à présent. Les nouvelles technologies de libre-service lancées au cours des cinq dernières années permettent maintenant aux utilisateurs fonctionnels de générer des analyses sans assistance.

Par exemple, grâce aux outils d'exploration visuelle, ceux qui ne connaissent ni le langage SQL ni la programmation peuvent plus facilement que jamais accéder à des bases de données, des applications et des fichiers et créer des tableaux de bord riches, interactifs, et des applications analytiques. Les outils de préparation des données et de catalogue permettent aux utilisateurs fonctionnels de trouver, nettoyer, formater et fusionner des données sans avoir à dépendre de la DSI. Enfin, les plateformes open source comme les bases de données Hadoop, Spark et NoSQL, alliées aux nouveaux outils de pipeline de données, permettent aux services et aux unités opérationnelles de créer rapidement des référentiels de données pour répondre aux besoins locaux en analytique.

Ces nouvelles technologies aident les entreprises à concrétiser la promesse de l'analytique en libre-service : les utilisateurs obtiennent ce qu'ils souhaitent, en temps utile et de la manière voulue, et la DSI se libère des tâches pénibles d'élaboration de rapports personnalisés. Cette solution avantageuse pour tous permet d'accélérer l'analyse en évitant l'intermédiaire, la DSI, et offre un accès direct aux données et aux fonctions de création de rapports.

Collaboration entre la DSI et l'opérationnel. La promesse de l'analytique en libre-service consiste non pas à éliminer la DSI de l'équation, mais à favoriser une meilleure collaboration entre elle et les utilisateurs. Plutôt que de présenter les besoins par écrit, l'entreprise utilise des outils en libre-service pour créer des rapports ad hoc et des jeux de données, dont certains peuvent servir de prototypes fonctionnels pour des applications de production destinées à un vaste public. En substance, l'analytique en libre-service instancie les besoins métier, ce qui accélère le déploiement et optimise les résultats.

L'analytique en libre-service modifie le rôle de la DSI, le développement d'applications faisant place à l'organisation des données. Comme la DSI connaît mieux les données que les utilisateurs fonctionnels, elle s'attache à établir une chaîne de l'information dans laquelle ces derniers puisent pour créer des applications personnalisées. Tout comme une raffinerie reçoit en matière première du pétrole brut et le traite en de nombreux produits différents, une chaîne de l'information raffine des données brutes de divers systèmes et les transforme en de l'information exploitable.

La promesse de l'analytique en libre-service consiste non pas à éliminer la DSI de l'équation, mais à favoriser une meilleure collaboration entre elle et les utilisateurs.

En parallèle, le rôle des utilisateurs fonctionnels cesse d'être passif et ils participent activement au processus de développement. Comme ils connaissent mieux leurs besoins que la DSI, ils jouent un rôle moteur dans la création d'applications analytiques. Grâce aux outils en libre-service, ils parcourent rapidement les options et affinent les besoins jusqu'à trouver une solution opérationnelle. La DSI appuie le processus ad hoc en procurant des données organisées, que les utilisateurs fonctionnels enrichissent d'autres données internes ou externes obtenues directement. Le cas échéant, ils peuvent demander à la DSI de convertir leurs rapports ad hoc en applications de production reposant sur une infrastructure de données et de systèmes sécurisée, évolutive et fiable (voir la figure 2 à la page 7).

Les risques

Malgré cette nouvelle génération de technologie, l'analytique en libre-service n'est pas une simple formalité. Trop souvent, l'analytique en libre-service dégrade l'environnement des données, au lieu de l'améliorer.

[Quel problème peut alors se poser ?](#)

Tour de Babel. Parfois, lorsqu'une entreprise déploie des outils en libre-service, quelques utilisateurs fonctionnels s'en emparent avec ardeur et créent une profusion de rapports et de tableaux de bord aux résultats conflictuels. D'autres utilisateurs ne parviennent pas alors à trouver les rapports dont ils ont besoin, et ne comprennent pas que des rapports similaires affichent des résultats différents.

Pire, les créateurs de rapports ne font confiance qu'à leurs propres données. Ils se retranchent dans leur propre silo et refusent d'utiliser d'autres données, pour des raisons d'ordre politique ou pour protéger leur territoire, réputation ou carrière. Cette diffusion de rapports conflictuels crée une tour de Babel où tout le monde parle, mais personne ne communique.

Cette diffusion de rapports conflictuels crée une tour de Babel où tout le monde parle, mais personne ne communique.

Dans ce chaos de données, les dirigeants ne peuvent pas obtenir des réponses directes aux questions simples, comme « Combien de clients avons-nous ? ». De telles questions peuvent déclencher une tempête lorsque chaque unité opérationnelle définit ses propres indicateurs et utilise des données différentes pour répondre aux mêmes questions. Les réunions de comité directeur dégénèrent en débats sur celui qui possède des données exactes. Plutôt que d'utiliser leur temps en prise de décisions productive, les dirigeants se retrouvent à arbitrer des conflits de données au détriment de l'entreprise.

Environnements de reporting. En dernier lieu, beaucoup d'unités opérationnelles échouent dans des environnements de reporting en libre-service dysfonctionnels qui ne peuvent suivre le nombre croissant d'utilisateurs, le volume de données et la complexité des applications. Si des développeurs clés partent, l'unité risque de se retrouver sans les ressources pour corriger, gérer ou améliorer l'environnement de reporting. Les unités opérationnelles peuvent demander à la DSI d'intervenir pour gérer l'environnement avant son implosion.

Raisons de la difficulté

Il n'y a pas de solution universelle. Il existe de nombreuses raisons qui expliquent la difficulté à réaliser la promesse de l'analytique en libre-service. L'une des principales réside dans le fait que l'analytique en libre-service n'est pas une entité homogène permettant un déploiement universel. L'analytique en libre-service revêt des significations différentes selon les personnes.

Par exemple, un dirigeant peut définir le libre-service comme la capacité à consulter en ligne un tableau de bord pendant une réunion de bilan opérationnel. Un responsable peut le considérer comme la capacité à explorer, trier et filtrer des tableaux de bord et des rapports. Pour un analyste marketing, il s'agira de la capacité à créer des jeux de données personnalisés pour une campagne de marketing ciblé qui combine des données démographiques internes et externes.

Pour réussir à mettre en œuvre l'analytique en libre-service, les entreprises doivent adapter les résultats à chaque individu dans l'organisation. Le défi s'avère de taille. Il faut réunir une connaissance approfondie des utilisateurs fonctionnels et de leurs besoins en information à une plateforme analytique sophistiquée qui permettent aux administrateurs de définir et de gérer les droits d'accès aux données et aux fonctions d'analyse.

Pour réussir à mettre en œuvre l'analytique en libre-service, les entreprises doivent adapter les résultats à chaque individu dans l'organisation.

Dynamique organisationnelle. En outre, l'analytique en libre-service dresse des forces organisationnelles intrinsèques les unes contre les autres. D'un côté, les dirigeants d'entreprise cherchent une adéquation fonctionnelle, des économies d'échelle et un seul point d'entrée pour le client. De l'autre, les responsables d'unité opérationnelle souhaitent plus d'autonomie pour répondre aux besoins locaux rapidement, efficacement et rentablement. Une partie aspire à une centralisation et une normalisation accrues, l'autre à la décentralisation et la personnalisation. Voir notre rapport 2016 [Exploration de données gouvernée : équilibrer l'agilité et les normes](#).

Ce conflit d'approche descendante et ascendante fait obstacle au déploiement décisionnel. Les dirigeants d'entreprise doivent reconnaître ce conflit et obtenir une trêve. Ils doivent redéfinir les modèles organisationnels pour fédérer la responsabilité de l'analytique entre l'entreprise et les unités opérationnelles. Ils doivent aussi mettre en place une architecture standard et des plateformes de données qui permettent de déployer l'analytique en libre-service sans sacrifier la cohérence et l'adéquation des données.

Facteurs de succès

Pour réussir à obtenir un environnement d'analytique en libre-service gouverné, les entreprises doivent déployer :

1. Une architecture de référence. Les entreprises doivent en premier lieu édifier un environnement de données en libre-service. Diffuser simplement les outils en libre-service à chaque utilisateur qui le souhaite ou permettre aux éditeurs de vendre directement aux unités opérationnelles sans consulter la DSI peut se révéler désastreux. L'architecture idéale cartographie les utilisateurs, la technologie et les développeurs dans une chaîne de l'information.

2. Une organisation fédérée. Les entreprises ont aussi besoin d'un modèle organisationnel qui équilibre la force ascendante (centralisation) et celle descendante (décentralisation). Une organisation fédérée avec une matrice des tâches de reporting, des équipes communes de spécialistes en décisionnel et données, et des experts en action opérationnelle est le meilleur moyen d'équilibrer liberté et contrôle, agilité et gouvernance, libre-service et normes. Créer cet équilibre est crucial pour la réussite du libre-service.

3. Une plateforme de données et d'analytique standard. L'ironie de l'analytique en libre-service réside dans le fait qu'elle exige une normalisation. Outre l'architecture de référence, les entreprises doivent chercher à établir une plateforme de données et d'analytique standard pour tous les utilisateurs. Un environnement unifié permet aux administrateurs d'adapter plus facilement l'analytique aux besoins individuels et aux utilisateurs fonctionnels et aux développeurs de la DSI de collaborer en matière de besoins et de développement, ce qui accélère la création de rapports et de tableaux de bord.

L'ironie de l'analytique en libre-service réside dans le fait qu'elle exige une normalisation.

4. Processus de gouvernance. Finalement, l'analytique en libre-service n'est pas une technologie ou un produit, il s'agit d'un processus. Déterminer qui a le droit d'accéder à des données et des fonctions et de les partager (ou pas) avec certaines personnes nécessite une gestion et un contrôle. L'entreprise (et pas la DSI) doit former un comité d'auto-gouvernance, pour qu'elle puisse optimiser l'utilisation des données et éviter les abus. Elle doit appliquer des moyens de marquage pour différencier les objets de données gouvernés de ceux non gouvernés afin que les utilisateurs fonctionnels comprennent le contexte des données qu'ils exploitent. Voir « [Comment des filigranes peuvent transformer votre organisation décisionnelle](#) ».

5. Formation continue. En dernier lieu, les entreprises doivent intensifier la formation, aussi paradoxal que cela puisse sembler. Kevin Sonsky, directeur principal du décisionnel chez Citrix Systems, pense que l'analytique en libre-service exige un sérieux accompagnement. Le fait est que même les utilisateurs chevronnés ne seront productifs ou constructifs avec des outils en libre-service qu'à condition de recevoir une solide formation formelle et informelle et le soutien de leurs collègues.

« **L'analytique en libre-service exige un accompagnement soutenu.** »

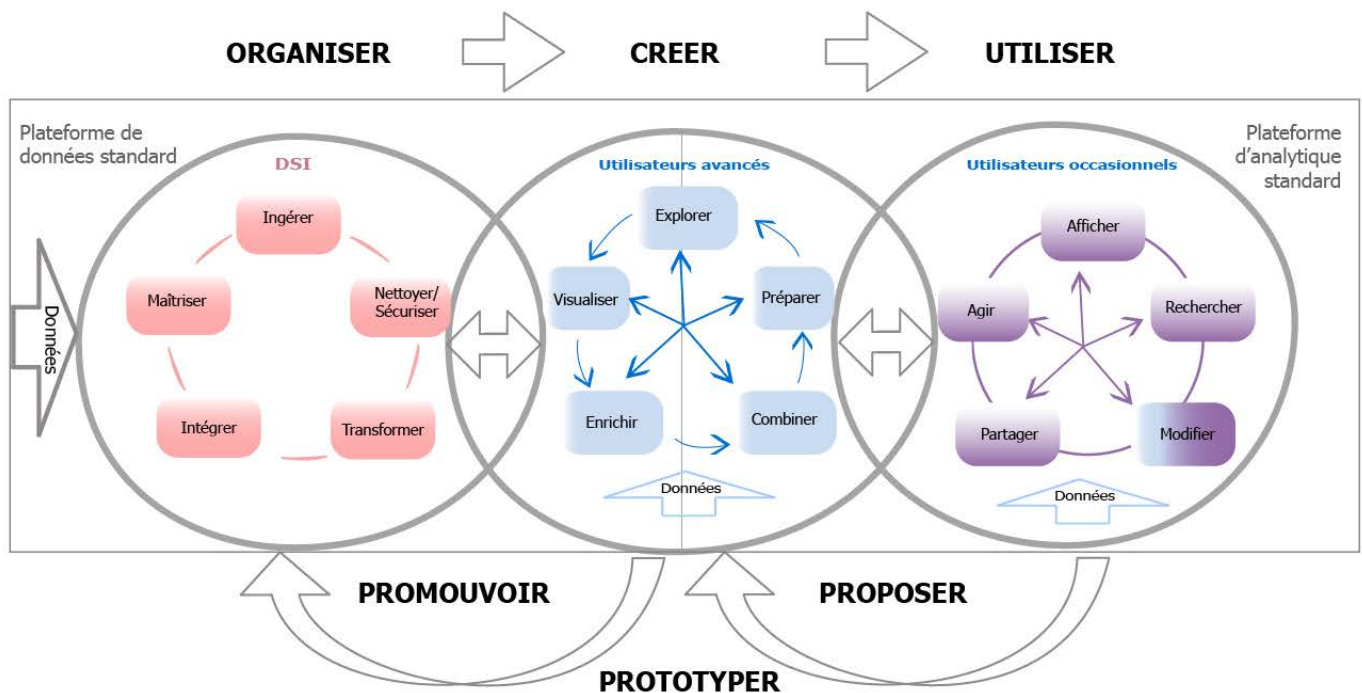
— **Kevin Sonsky, directeur principal du décisionnel chez Citrix Systems**

Workflows en libre-service

De concert, les utilisateurs fonctionnels et les développeurs créent un environnement d'analytique en libre-service qui affine les données pour l'exploitation. L'environnement consiste en des workflows bidirectionnels et itératifs qui permettent aux utilisateurs fonctionnels d'affiner et d'enrichir des données organisées pour répondre rapidement à leurs besoins. Il s'agit également de promouvoir des prototypes et des exigences pour élargir successivement les limites de l'environnement de données. Ces workflows créent un environnement de données dynamique, synergique, qui accélère la connaissance et la productivité des utilisateurs (voir la figure 2).

Workflow de gauche à droite. Dans la plupart des entreprises, les données circulent de la source vers la cible. Elles deviennent plus affinées et organisées au fur et à mesure qu'elles se rapprochent des utilisateurs fonctionnels. Habituellement, la DSI gère le processus d'organisation des données et le processus de création des rapports. En conséquence, la DSI devient souvent un point de congestion qui se dresse entre les utilisateurs fonctionnels et les données.

Fig. 2 - Workflows bidirectionnels en libre-service



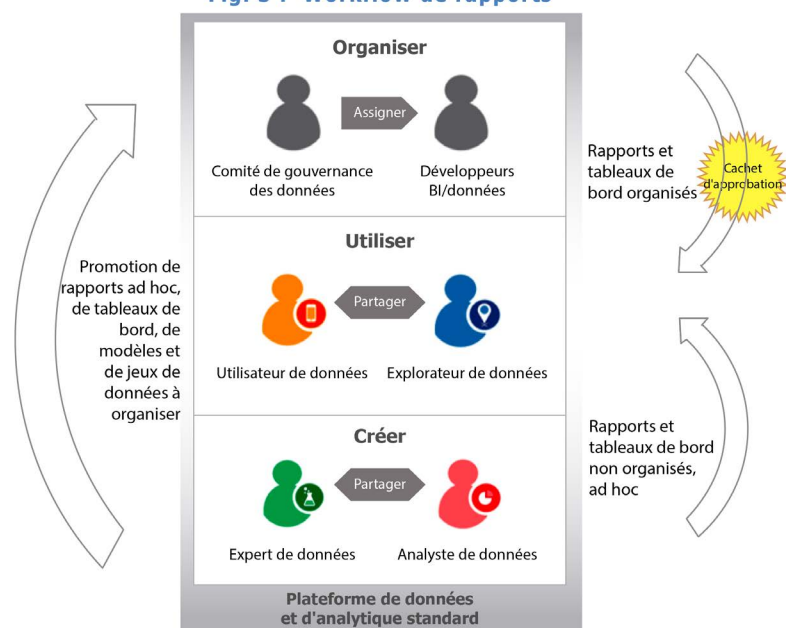
Toutefois, les outils d'analytique en libre-service permettent aux utilisateurs avancés de participer au processus d'organisation. Grâce à des outils de préparation, ils peuvent transformer, combiner et enrichir les données d'entreprise et locales de manière itérative. Grâce à des outils d'exploration visuelle, ils peuvent afficher et analyser des données, puis partager leurs résultats avec des collègues ou responsables. Ensuite, ces utilisateurs occasionnels consultent, recherchent, modifient et commentent les rapports des utilisateurs avancés et prennent des mesures en fonction des résultats.

Les outils d'analytique en libre-service permettent aux utilisateurs avancés de participer au processus d'organisation.

Workflow de droite à gauche. Un workflow inverse est tout aussi important pour que les utilisateurs fonctionnels intègrent leurs besoins aux processus d'organisation et de création. Après l'analyse des données dans les rapports et les tableaux de bord, les utilisateurs occasionnels comprennent mieux leurs besoins et peuvent les proposer aux utilisateurs avancés de leur service. Ces derniers peuvent alors intégrer ces besoins à un rapport ad hoc que les utilisateurs occasionnels vont examiner et affiner.

En cas de rapport général, les utilisateurs avancés peuvent demander à une équipe de gouvernance de l'étudier. Le rapport sert de prototype pour en produire un nouveau. Le comité contrôle les critères du rapport, les définitions de données et la disposition de l'interface pour garantir le respect des normes de l'entreprise. Enfin, la DSI détermine comment industrialiser le rapport ad hoc avec l'évolutivité, la sécurité et la fiabilité adéquates (voir la figure 3).

Fig. 3 : Workflow de rapports



Filigrane. Si la proposition est validée, le rapport porte ensuite un cachet ou un tampon d'approbation. Cette marque indique aux utilisateurs fonctionnels que le rapport est conforme pour la prise de décision. Le filigrane aide les utilisateurs à distinguer les données organisées des données brutes. Les utilisateurs fonctionnels peuvent alors refuser des rapports qui ne portent pas le cachet d'approbation. Ce processus crée une culture de gouvernance à la base. Plutôt que de travailler indépendamment, les utilisateurs avancés réalisent l'intérêt d'adopter des canaux formels pour publier des rapports afin d'élargir la diffusion.

Le filigrane aide les utilisateurs à distinguer les données organisées de données brutes.

Comprendre les workflows de données et de rapports, ainsi que les rôles des divers utilisateurs fonctionnels, permet aux entreprises d'accélérer la diffusion des informations et d'éliminer l'engorgement décisionnel. Il en résulte un environnement en libre-service géré qui équilibre les exigences de rapidité, d'agilité et de liberté d'une entreprise avec celles de normalisation, de contrôle et de gouvernance.

Chaîne de l'information

L'épine dorsale de l'architecture de référence réside dans la chaîne de l'information, à savoir le flux de données des systèmes sources aux utilisateurs fonctionnels (voir la figure 4). Une chaîne de l'information est cruciale pour accélérer le flux des données gérées et minimiser l'activité en libre-service.

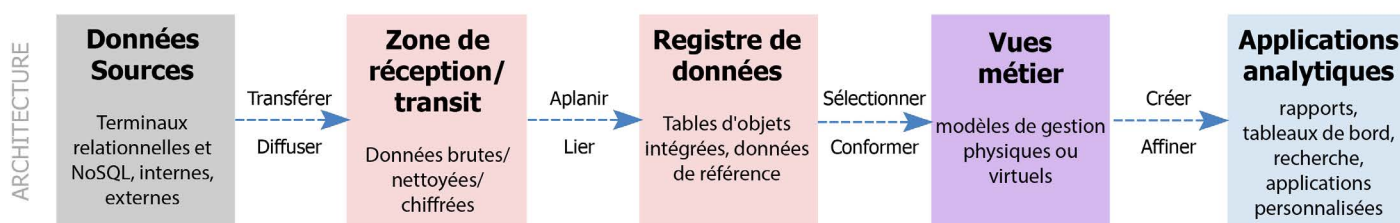
Une chaîne de l'information bien conçue et agile rend les utilisateurs fonctionnels plus productifs en mettant les données dont ils ont besoin à leur disposition.

Cette chaîne descendante dirigée est gérée par la DSI, qui supervise les déploiements d'entrepôt de données et Hadoop d'entreprise. Les entreprises peuvent mettre en place la chaîne de l'information au moyen de plusieurs technologies. Les bases de données Hadoop et relationnelles sont à présent les plateformes de données privilégiées, alors que bases de données Spark, NoSQL et Cloud représentent les nouveaux concurrents. Voir notre rapport 2015 « [Sélection d'une plateforme Big Data : jeter les bases de l'avenir](#) ».

Les entreprises peuvent utiliser la chaîne pour déployer ou alimenter un entrepôt de données. Inversement, un entrepôt de données peut alimenter une chaîne de l'information déployée sous Hadoop. Par essence, la chaîne de l'information est un ensemble de structures de données en cascade qui permet de satisfaire divers besoins de traitement des données et d'analytique.

Une chaîne de l'information complète et bien conçue minimise le besoin pour les utilisateurs fonctionnels d'obtenir leurs propres données en dehors des canaux dirigés. Toutefois, en réalité, les utilisateurs avancés demandent souvent une masse de données dépassant celle que la DSI peut placer dans un référentiel de données d'entreprise. La plupart des entreprises doivent donc, délibérément ou pas, compléter la chaîne de l'information par des workflows en libre-service (voir la figure 4).

Fig. 4 : Chaîne d'information gouvernée



Composants

Voici une brève description de chaque composant de la chaîne de l'information.

- **Données sources.** Il s'agit de données transactionnelles, structurées, semi-structurées ou non structurées, générées automatiquement ou manuellement. En général, les entreprises transfèrent les données opérationnelles structurées par lot (ou par micro-lot en quasi temps réel) et capturent les données modifiées dans une zone de réception. Elles envoient de gros volumes de données de médias sociaux générées automatiquement vers la zone de réception à l'aide d'outils de streaming comme Kafka, Flume, Storm et Spark Streaming.

- **Zone de réception.** Les processus par lot ou de streaming placent des données brutes de tout type dans une zone de réception où elles sont stockées de manière permanente sous leur forme d'origine jusqu'à ce qu'elles ne soient plus nécessaires. La plupart des données sont chiffrées à leur réception, les plus sensibles pouvant également être masquées ou tokenisées. Hadoop est devenu une zone de réception courante grâce à sa structure de type fichier et à son faible coût. Les entreprises peuvent aussi utiliser une base de données relationnelle comme zone de réception, mais cela n'a de sens que si leurs données sont structurées et peu volumineuses. Seuls les administrateurs informatiques touchent aux données dans la zone de réception.

- **Zone de transit.** Les administrateurs informatiques effectuent des contrôles de qualité des données et des processus de nettoyage pour corriger les erreurs avant de déplacer les données de la zone de réception vers celle de transit. Ils normalisent également les conventions de dénomination, en créant des objets de base ou des éléments constitutifs au plus bas niveau de granularité possible. Ils appliquent un masquage ou une tokenisation pour protéger les champs sensibles, le cas échéant. Les experts de données accèdent souvent à ce niveau, car ils préfèrent utiliser des données brutes avant qu'elles ne soient fortement transformées ou agrégées.

- **Registre de données.** Dans le registre de données, les administrateurs préparent les données pour une utilisation générale en aval. Ils résumet et joignent les données, créent des critères pertinents et fusionnent des objets de base dans des tables d'espace d'objets liées. Par exemple, ils peuvent fusionner toutes les données d'un compte client dans une vue stockée dans une vaste table plate des clients. Un registre de données peut alimenter un entrepôt de données ou un outil d'exploration ainsi que des banques de données opérationnelles ou des applications plus transactionnelles par nature.

- **Vues Métier.** Les vues métier transforment des données du registre en objets de données pertinents pour un groupe d'utilisateurs particulier. Créées par des analystes fonctionnels ou des développeurs orientés gestion, ces vues sont instanciées physiquement dans une base de données ou virtualisées dans une couche d'abstraction ou sémantique qui affiche les données dans une perspective métier. Les développeurs peuvent créer des vues métier avec une gamme de technologies : outils décisionnels, outils de virtualisation des données, cubes OLAP et vues de base de données. Les vues métier externes de chaque outil sont souvent appelées des « services de données ». De nombreuses sociétés de services financiers en utilisent pour fournir aux utilisateurs et aux applications un accès transparent à plusieurs systèmes dorsaux et sources externes qui seraient trop coûteux ou impossibles à fédérer dans un entrepôt de données. Ces vues ou services offrent aux administrateurs informatiques une meilleure agilité pour modifier des plateformes de données dorsales et des flux sans impact sur les utilisateurs ou les applications en aval.

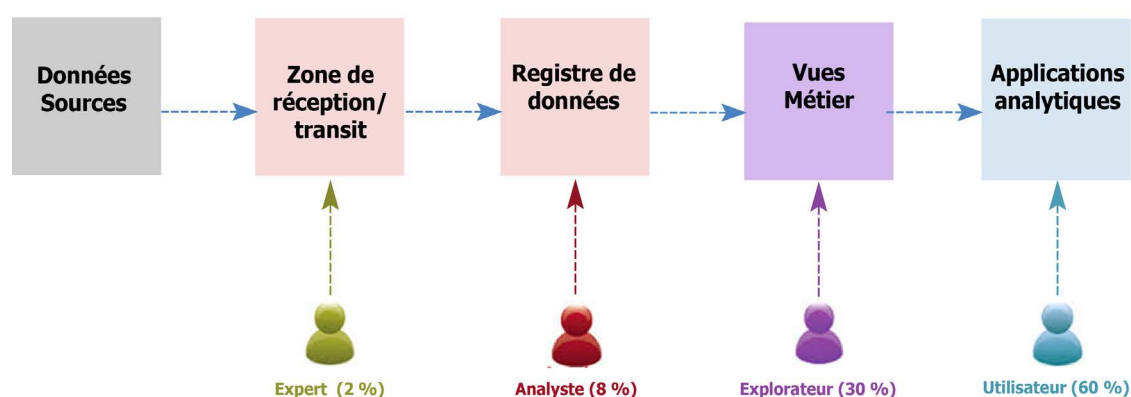
- **Applications analytiques.** Finalement, les développeurs en décisionnel utilisent les vues métier pour créer des rapports, des tableaux de bord et des applications analytiques personnalisées. Ces applications utilisent souvent un sous-ensemble de la vue métier, publié pour les explorateurs de données qui souhaitent modifier des rapports ou des tableaux de bord en ajoutant de nouveaux critères ou d'autres dimensions provenant des champs de base dans le modèle. De nombreux outils décisionnels offrent maintenant de riches interfaces de programmation d'applications (API) qui permettent aux développeurs connaissant le langage SQL et JavaScript de créer des interfaces graphiques et des workflows très personnalisés.

Cartographie des utilisateurs dans la chaîne de l'information

L'aspect le plus marquant de cette architecture de référence en libre-service est qu'elle cartographie les types d'utilisateurs fonctionnels dans la chaîne de l'information (voir la figure 5). Auparavant, la DSI verrouillait en grande partie la chaîne de l'information, en permettant uniquement aux utilisateurs d'accéder aux outils décisionnels ou aux entrepôts de données. Quelques individus se voyaient accorder un accès direct au registre de données (ou à l'entrepôt de données) et peu d'utilisateurs, voire aucun, pouvaient interroger les données dans la zone de réception ou celle de transit. Une chaîne de l'information bien définie offre aux administrateurs informatiques une confiance accrue pour accorder à différents types d'utilisateurs fonctionnels un accès aux données.

Notre architecture de référence permet aux utilisateurs d'accéder à la chaîne de l'information en fonction de leurs rôles et besoins. Pour avoir accès à la zone de réception/transit et au registre de données, les individus de certaines entreprises doivent passer un examen pour prouver qu'ils possèdent les compétences et connaissances afin d'exploiter ces données correctement.

Fig. 5 : Type d'Utilisateur Métier



1. **Experts de données et zone de transit.** Les experts de données (et les statisticiens) ont besoin de données brutes, non agrégées, pour créer des modèles prédictifs. Plutôt que d'extraire des données des systèmes opérationnels, ils doivent disposer d'un accès à des données brutes dans la zone de transit, qui ont au moins été nettoyées et chiffrées. Les entreprises doivent leur fournir des catalogues de données pour faciliter la recherche, ainsi que des outils de préparation pour formater, transformer et combiner les données et des outils d'exploration visuelle pour afficher ou analyser les données.

2. Analystes de données et registre de données. Un véritable analyste de données souhaite extraire les données des vastes tables spécifiques du registre de données pour alimenter un outil d'exploration visuelle. Les entreprises pré-alimentent de plus en plus les outils d'exploration avec ces données, pour éviter aux analystes (et aux explorateurs de données) la tâche consistant à corrélérer, explorer et ingérer les données. Toutefois, les analystes préfèrent souvent explorer les données derrière les outils. Comme les experts de données, ils doivent disposer de catalogues, d'outils de préparation et d'outils d'exploration visuelle des données pour faciliter leur analyse.

3. Explorateurs de données et vues métier. Les explorateurs de données utilisent des vues métier pour modifier et compléter les rapports existants ou en créer des simples ex nihilo. De nombreux systèmes décisionnels intègrent maintenant des outils légers de préparation de données qui permettent aux explorateurs de fusionner le contenu d'un rapport avec d'autres jeux de données absents de la vue métier. Certains intègrent également des marchés de données pour aider les explorateurs (notamment) à englober d'autres données internes et externes dans leurs rapports.

4. Utilisateur de données et applications analytiques. Les utilisateurs de données explorent et analysent des données dans un rapport ou un tableau de bord. Ils n'obtiennent jamais de nouvelles données, mais souhaitent souvent enregistrer un instantané d'un rapport ou d'un tableau de bord pour pouvoir le consulter ultérieurement avec des données à jour. Il s'agit d'un moyen facile de créer un rapport ou un tableau de bord personnalisé.

Technologie

Ce rapport a mentionné différentes technologies nécessaires pour créer et gérer un environnement d'analytique en libre-service. Cette section décrit plus en détail ces technologies de base et les rôles qu'elles jouent pour étayer un environnement en libre-service gouverné. Le prochain rapport d'Eckerson Group (dont la publication est prévue en décembre 2016) va approfondir l'étude de beaucoup de ces technologies.

Pipeline de données

Les outils de pipeline des données forment une classe de produits de gestion des données pour la gouvernance du flux entre la source et la cible dans un environnement Big Data. Ils ingèrent, valident, nettoient, transforment, fusionnent, sécurisent et formatent les données à des fins analytiques et gèrent les processus requis pour exploiter un environnement de production. La plupart permettent des mises à jour par lot, en quasi temps réel (micro-lots) et en temps réel (streaming), ainsi que la capture de données modifiées pour ne charger qu'un delta et minimiser les volumes de données. Surtout, ils collectent des métadonnées à chaque étape du processus. Les utilisateurs techniques et fonctionnels peuvent vérifier la traçabilité des données, comprendre comment et quand les données ont été transformées et évaluer l'impact d'une modification sur les applications en aval.

Il existe plusieurs types d'outils de pipeline des données. Certains s'adressent à la DSI et d'autres aux utilisateurs avancés.

Outils de pipeline orientés DSI

- **Extraction, transformation et chargement (ETL).** Les outils ETL sont la base des entrepôts de données. Ils permettent aux développeurs de cartographier des données sources et cibles et d'exécuter des transformations. Conçus à l'origine comme des moteurs d'exécution externes, beaucoup exploitent maintenant les systèmes de traitement des données sous-jacents (des bases de données relationnelles ou Hadoop) pour gérer d'autres processus d'extraction, de chargement et de transformation, les rendant plus comparables aux outils de gestion des lacs de données et d'automatisation d'entrepôts de données.

- **Moteurs de streaming.** Ces outils se spécialisent dans la capture de données en temps réel à partir de bus de messages, la transformation de données à la volée et la réception de données en temps réel avec une vérifiabilité et une gestion des files d'attente.

- **Automatisation d'entrepôts de données.** Ces outils orientés métadonnées automatisent les processus pour créer, alimenter, exploiter et gérer des entrepôts de données relationnels et des mini-entrepôts. Ils facilitent le développement agile et gèrent maintenant directement les données dans des bases de données Hadoop et NoSQL.
- **Gestion des lacs de données.** Ces produits fonctionnent habituellement en natif sur des plateformes Hadoop et permettent de gérer le flux bidirectionnel des gros volumes de données entrants et sortants. Nombre d'entre eux englobent des environnements d'exploration à disposition des utilisateurs fonctionnels.
- **Virtualisation des données.** Ces outils génèrent une vue virtuelle, orientée métier, des ressources de données dorsales. Ces vues, aussi désignées en tant que couche sémantique ou métadonnées commerciales, sont générées par des outils décisionnels, des plateformes analytiques et des produits de virtualisation des données indépendants. Comme les produits de gestion des lacs de données, ces outils rapprochent l'entreprise et la DSI.

Conclusion

Bien que complexe visuellement, l'architecture de référence d'Eckerson Group pour l'analytique en libre-service cartographie les utilisateurs fonctionnels, les développeurs et les technologies dans une chaîne de l'information qui sert de base au traitement analytique.

L'architecture de référence permet des workflows de données itératifs dans lesquels la DSI, les utilisateurs avancés et les utilisateurs occasionnels collaborent pour créer de l'information, des rapports et des tableaux de bord qui répondent rapidement aux besoins de gestion. Elle définit également des zones en libre-service et le rôle de la DSI, des développeurs et de la technologie dans chaque zone. Elle indique comment fusionner les méthodes de gouvernance descendante et ascendante et les workflows de reporting pour garantir que les utilisateurs fonctionnels obtiennent rapidement l'information nécessaire et puissent différencier le contenu dirigé et non dirigé.

Il n'est pas facile de réussir en matière d'analytique en libre-service. Outre l'architecture de libre-service dirigée, il faut déployer des processus de gouvernance bien conçus, une plateforme de données et d'analytique standard, une structure organisationnelle fédérée avec des équipes communes de développeurs en décisionnel, ainsi qu'une formation et un support continus. Ce rapport pose les bases de la réussite en définissant une architecture de référence pour étayer l'analytique en libre-service.

Avez-vous besoin d'aide en termes de stratégie d'analytique opérationnelle, de gestion des données ou de gouvernance ? Souhaitez-vous découvrir les dernières tendances en matière d'analytique opérationnelle et d'outils décisionnels ?

