Cadre de Référence International
des Pratiques Professionnelles
de l'audit interne

Lignes directrices complémentaires

GTAG®
Global Technology
Audit Guide

Comprendre et auditer le *Big Data*

Table des matières

Synthèse ... 3

Introduction ... 4

Enjeux métier ... 5
- Mise en pratique ... 6

Principes de base du *Big Data* ... 7
- Données structurées et non structurées ... 7
- Stockage de données ... 8
- Les trois « V » ... 9

Composantes d'un programme de *Big Data* ... 10
- Un *business case* explicite ... 10
- Rôles et responsabilités définis ... 10
- Ressources adéquates ... 12
- Indicateurs de performance ... 12
- Outils et technologies ... 12
- Soutien permanent au programme ... 15
- Gouvernance des données ... 16
- Appropriation par le consommateur ... 17
- Analyse et *reporting* ... 17
- Processus cohérents ... 18

Principaux risques ... 19

Le rôle de l'audit interne en matière de *Big Data* ... 22

Annexe A. Principales normes IIA concernées ... 24

Annexe B. Lignes directrices complémentaires concernées ... 28

Annexe C. Glossaire ... 29

Annexe D. Planification d'une mission d'audit du *Big Data* ... 33
- Risques et défis en matière de Big Data ... 33
- Planification de la mission ... 33
- Objectifs de la mission ... 34
- Périmètre d'intervention de la mission et ressources affectées à la mission ... 34
- Programme de travail de la mission ... 34

Contributeurs ... 50

Synthèse

Big Data (ou « mégadonnées ») est un terme communément utilisé pour décrire la croissance exponentielle et la disponibilité de données créées par des personnes, des applications et des appareils dits « intelligents ». Il fait référence à des masses de données complexes qui dépassent les capacités des applications traditionnelles de traitement de données. La prolifération de données structurées et non structurées, à laquelle s'ajoutent les progrès techniques en matière de stockage, de puissance de traitement et d'outils d'analyse, ont fait du *Big Data* un avantage concurrentiel dont se servent les organisations de premier plan pour mieux comprendre leurs opportunités et faire évoluer leurs stratégies. Cependant, les risques et enjeux associés au *Big Data* doivent également être pris en compte.

Une demande accrue, des référentiels encore peu matures, des risques émergents et des opportunités qui ne sont pas clairement compris de tous ni gérés de manière systématique par les organisations font qu'il est nécessaire d'avoir plus de lignes directrices dans ce domaine. Les auditeurs internes doivent en particulier développer de nouvelles compétences et connaître le *Big Data* pour donner l'assurance que les risques sont gérés efficacement et que les avantages sont concrétisés.

Parmi les risques associés au *Big Data* figurent la mauvaise qualité des données, des technologies inadaptées, une sécurité insuffisante et des pratiques de gouvernance des données peu matures. Les auditeurs internes qui s'intéressent au *Big Data* devraient discuter avec le directeur des systèmes d'information (DSI) et les autres responsables concernés pour avoir une meilleure compréhension des risques liés à la collecte, au stockage, à l'analyse, à la sécurité et à la protection des données.

Le présent guide donne une vue d'ensemble sur des sujets tels que l'intérêt du *Big Data,* ses composantes, les stratégies qui peuvent en découler, les considérations relatives à sa mise en œuvre, la gouvernance des données, l'usage qui peut en être fait, le *reporting* correspondant, ainsi que certains risques et enjeux éventuels. Ce guide explique également les rôles et les responsabilités des auditeurs internes dans le cadre de missions d'assurance et de conseil liées à des initiatives de *Big Data*.

Introduction

Ce guide aide les auditeurs internes à acquérir les connaissances nécessaires pour des missions de conseil et d'assurance en matière de *Big Data,* conformément à la Norme 1210 - Compétence et à la Norme 2201 - Considérations relatives à la planification. En donnant un aperçu du *Big Data,* il permet de comprendre les concepts associés et d'ajuster les activités d'audit interne en fonction des initiatives de l'organisation en la matière. Un référentiel des principaux risques et enjeux, ainsi que des exemples de contrôles dont il convient de tenir compte lors de la planification d'une mission dans ce domaine sont également proposés (cf. Norme 2100 - Nature du travail).

Ce document ne traite pas la question de l'exploitation ou de la réalisation d'analyses utilisant le *Big Data* par l'audit interne (cf. GTAG : *« Data analysis technologies »).* Il ne fournit pas non plus une liste de contrôles techniques spécifiques ou de programmes de travail permettant de couvrir tous les aspects du *Big Data,* car ils sont spécifiques à chaque organisation et aux systèmes de *Big Data* utilisés.

Pour parvenir à une couverture optimale des enjeux du *Big Data,* les auditeurs internes devront compléter ce guide avec d'autres GTAG et des programmes de travail plus techniques.

Enjeux métier

L'utilisation de l'analyse de données pour créer de la valeur n'est pas un concept nouveau. En revanche, il est de plus en plus important d'avoir une interprétation rapide de ces données pour une aide à la décision. Les organisations qui acquièrent et exploitent efficacement le *Big Data* peuvent plus rapidement tirer profit des tendances émergentes, des évolutions dans la demande des clients et des opportunités d'amélioration de l'efficacité opérationnelle. Ce qui permet à terme d'améliorer la satisfaction des clients et d'optimiser la réussite de l'organisation.

Comme indiqué dans le livre blanc de l'ISACA (*Big Data:* Impacts and Benefits White Paper, March 2013[1]), « *les organisations qui parviendront à maîtriser la gestion du Big Data peuvent engranger des retombées considérables et se différencier de leurs concurrents* ».

Parmi les avantages du *Big Data* figurent :

- l'avantage compétitif ;
- l'augmentation du chiffre d'affaires ;
- l'innovation et un développement plus rapide des produits et des services ;
- une meilleure prévision de la demande ;
- des décisions opérationnelles mieux informées ;
- l'efficience opérationnelle.

En effet, les programmes de *Big Data* correctement exécutés et contrôlés ont des avantages significatifs. Ils contribuent à l'accumulation et à l'utilisation de gros volumes de données structurées ou non pour des analyses sans précédent et des points de vue en temps opportun. (Ce qui paraissait impossible ou aurait pris des jours voire des semaines devient plus aisé.) En partant des nouvelles perspectives offertes par le *Big Data,* les organisations peuvent améliorer les processus de prise de décision, cibler de nouveaux clients de manière innovante et segmentée, servir les clients existants en adoptant un modèle de prestation amélioré et adapté à chacun, et offrir de nouveaux services et de nouvelles facilités permettant de se distinguer véritablement des concurrents. Les organisations qui tirent profit des opportunités offertes par le *Big Data* sont en mesure de développer un avantage concurrentiel durable.

De plus, les initiatives en matière de *Big Data* peuvent contribuer à améliorer la transparence, les analyses et le *reporting* du management, et à diminuer les coûts dans le cadre des programmes d'amélioration continue. À mesure que les données sont centralisées et consolidées pour un usage stratégique du *Big Data,* le coût marginal de réalisation des analyses utilisant ces données est considérablement réduit. Toute l'organisation peut alors bénéficier de ces d'initiatives.

Cependant, pour exploiter pleinement les avantages opérationnels du *Big Data,* les organisations doivent investir dans la constitution d'un environnement adapté, en recrutant et en maintenant

[1] http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Big-Data-Impacts-and-Benefits.aspx

des collaborateurs qualifiés, en définissant et en mettant en œuvre des processus reproductibles et en déployant des technologies appropriées.

Mise en pratique

Chaque jour, les organisations mettent en œuvre des moyens astucieux et innovants d'utilisation du *Big Data*. Ces derniers sont par exemple mis en pratique dans le secteur de la distribution, dans les organisations publiques, privées et à but non lucratif.

Distribution – Des profils sophistiqués de clients en fonction de l'historique d'achats, de la navigation en ligne, des réactions des clients concernant les produits, de la localisation et de la démographie sont utilisés pour créer une expérience d'achat orientée par les données. Les vendeurs peuvent consulter les achats précédents d'un client pour proposer un service personnalisé et faire des recommandations adaptées au profil du client et à ses messages sur les réseaux sociaux, en les combinant aux dernières données sur les tendances de la mode. Si le client aime un vêtement dont la taille ou la couleur n'est pas disponible en magasin, le vendeur peut utiliser un appareil intelligent pour déterminer instantanément la disponibilité de ce produit dans un autre magasin ou faire en sorte qu'il soit expédié au domicile du client.

Un autre exemple d'utilisation du *Big Data* pour améliorer le service à la clientèle est l'analyse des adresses et des données démographiques des clients ou prospects pour identifier l'emplacement idéal d'un nouveau magasin.

Municipalités – Les municipalités peuvent recueillir des données sur les problèmes de circulation à l'aide de diverses sources (par exemple grâce à des capteurs dispersés dans la ville, dans les autobus et les taxis). L'analyse de ces données permet d'identifier certaines relations, des besoins en termes de transports publics ou d'extension des routes. Elles peuvent également servir pour des sites Internet donnant aux abonnés des informations en temps réel sur le trafic, les aidant ainsi à choisir les itinéraires les plus courts, à éviter les embouteillages et à estimer l'heure d'arrivée.

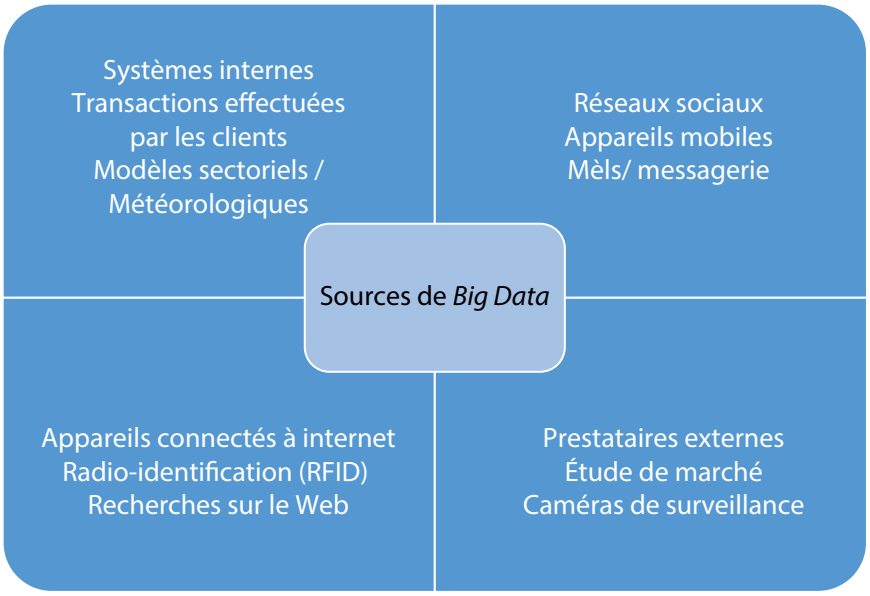
Institutions financières – Les institutions financières peuvent prévoir des tendances, modéliser des options et prédire les résultats afin de développer leur base clients et mieux fidéliser leur clientèle. Par exemple, elles sont en mesure de fournir des services plus personnalisés et des produits sur mesure en utilisant l'historique des opérations et les données démographiques afin de définir des tarifs et des stratégies financières adaptés à chaque individu.

Le *Big Data* permet la collecte de données à partir de sources externes. Par exemple, une banque peut utiliser des données provenant de réseaux sociaux pour identifier les clients qui ne sont pas satisfaits d'une autre institution financière et qui serait prêt à changer. Ce type d'informations peut être utilisé pour créer des messages de *marketing* ciblés et attirer ces clients mécontents en mettant en avant des produits et services personnalisés.

Principes de base du *Big Data*

Le *Big Data* est généré par les organisations (par exemple, des données transactionnelles, des réclamations), les secteurs d'activité (par exemple, le taux d'adoption par catégorie de produits), les collectivités (par exemple, les données enregistrées par les caméras de surveillance du trafic, les données économiques), la nature (par exemple, les données relatives à l'ampleur, l'emplacement et la fréquence des séismes) et de nombreuses autres sources.
Selon les cas, les organisations doivent acheter ces données externes ou peuvent y accéder gratuitement. Quelques exemples de sources figurent dans le **schéma 1.**

Schéma 1 - Exemples de sources

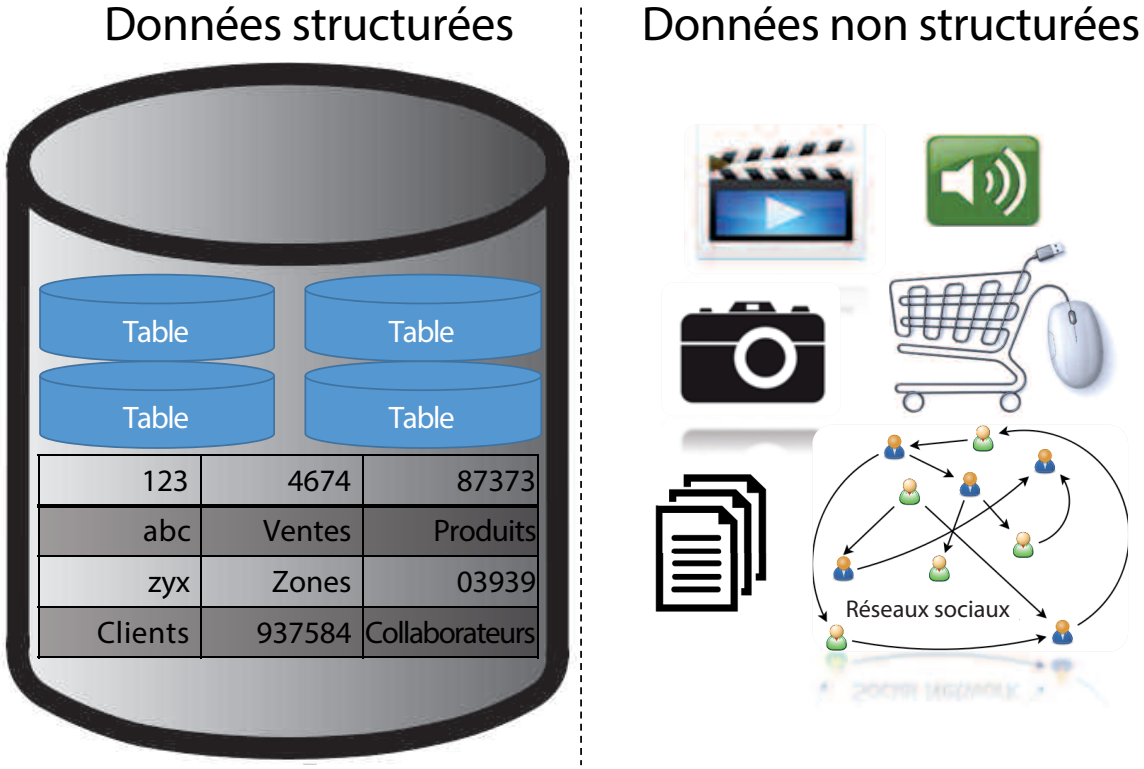


Données structurées et non structurées

Traditionnellement, la majorité des données stockées dans les organisations sont structurées et conservées dans des bases de données relationnelles voire même hiérarchiques ou dans des fichiers à plat. Les données structurées sont organisées et permettent des requêtes répétitives parce que la plupart des données sont dans des tables relationnelles. Il est souvent plus facile de les contrôler par rapport aux données non structurées car leur propriétaire et les prestataires de solutions de gestion de bases de données sont identifiés.

Cependant, l'utilisation des données non structurées progresse et devient une pratique de plus en plus courante. Ce type de données ne se limite pas aux structures de données ou aux contraintes classiques. Elles sont généralement plus difficiles à gérer, en raison de leur nature évolutive et imprévisible mais aussi en raison du nombre et de la diversité de leurs sources souvent externes. Par conséquent, de nouvelles solutions ont été développées pour gérer et analyser ce type de données. Le **schéma 2** illustre la différence entre données structurées et non structurées.

Schéma 2 - Exemples de données structurées et non structurées



Stockage de données

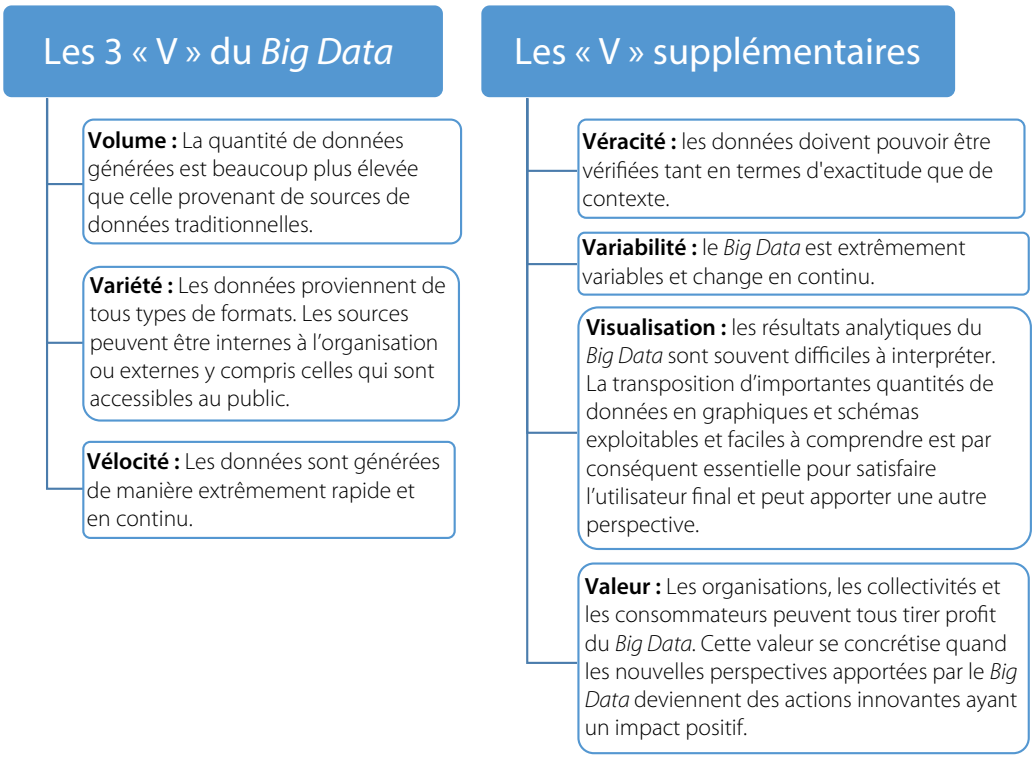
Un énorme répertoire de données concernant l'ensemble de l'organisation est spécialement conçu pour l'analyse et le *reporting* : c'est l'entrepôt de données. Les entrepôts de données sont généralement des bases de données relationnelles dans lesquelles sont stockées des informations provenant de sources multiples. Les données sont téléchargées à partir de systèmes opérationnels qui contiennent des données transactionnelles, vers des entrepôts de données, dans lesquelles sont stockées des informations détaillées sur un ou plusieurs sujets. Les outils ETL (pour *Extract, Transform, and Load* autrement dit des outils d'extraction, de transformation et de chargement) ou ELT (*Extract, Load and Transform* ou extraction, chargement et transformation) sont configurés pour transférer les données du système opérationnel vers l'entrepôt de données. Les données sont chargées dans le format et la structure de l'entrepôt de données, qui est souvent agrégée.

Les lacs de données *(data lakes)* sont une solution de stockage et de restitution de *Big Data* de plus en plus répandue. Les lacs de données sont similaires aux entrepôts de données dans la mesure où ils permettent de stocker d'importantes quantités de données provenant de diverses sources. Mais ils permettent également de stocker des caractéristiques supplémentaires de données à partir des systèmes sources selon un niveau de granularité qui ne pourrait pas être maintenu lors de l'agrégation des données avant leur transfert dans les entrepôts de données. Cela permet de disposer de solution avec tous les éléments de la donnée disponibles à un niveau de granularité suffisant pour effectuer une analyse complète. Ainsi, les organisations ont la flexibilité nécessaire pour résoudre des problèmes imprévus, l'ensemble de la donnée étant conservé dans un format facilement accessible.

Les trois « V »

La gestion du *Big Data* est habituellement caractérisée en se référant aux trois axes que sont **le volume, la vélocité et la variété** (les trois « V »), mais à mesure que les systèmes deviennent plus efficients et que le besoin de traiter rapidement les données augmente, les caractéristiques d'origine de la gestion des données ont évolué pour inclure d'autres dimensions spécifiques aux *Big Data*. Ainsi, Mark van Rijmenam propose quatre dimensions supplémentaires dans l'article publié sur son blog en août 2006 intitulé « *Why The 3Vs Are Not Sufficient To Describe Big Data* »[2]. Ces dimensions supplémentaires sont **la véracité, la variabilité, la visualisation et la valeur.** Le **schéma 3** illustre la gamme de « V » du *Big Data*.

Schéma 3 – Gamme de « V » du *Big Data*



[2] https://datafloq.com/read/3vs-sufficient-describe-big-data/166

Composantes d'un programme de *Big Data*

L'audit interne est souvent sollicité par le Conseil et/ou la direction générale pour obtenir un point de vue et des perspectives sur les programmes de *Big Data* au fur et à mesure de leur mise en œuvre (cf. Norme 2010.C1). Afin de s'engager efficacement auprès du management et d'évaluer ces programmes, les auditeurs internes devront comprendre les différents éléments qui composent un tel programme, ainsi que les rôles et les responsabilités associés (cf. Norme 2100 - Nature du travail, Norme 2200 - Planification de la mission et Norme 2201 - Considérations relatives à la planification, ainsi que leurs guides de mise en œuvre respectifs).

À mesure que les programmes de *Big Data* sont mis en œuvre et modifiés, les auditeurs internes doivent rester impliqués et suivre ces initiatives, conformément à la Norme 2010 - Planification et à la Norme 2030 - Gestion des ressources. Pour ce faire, les auditeurs internes doivent maintenir les connaissances et les compétences requises et modifier de manière dynamique leur évaluation des risques et leur modèle de couverture du périmètre d'audit pour tenir compte des changements (cf. Norme 1210 - Compétence). Le *Big Data* évolue rapidement et continuera à présenter des risques et des opportunités pour les organisations et les auditeurs internes dans un proche avenir.

Un *business case* explicite

Pour réussir un programme de *Big Data*, il est nécessaire de partir d'un *business case* clair qui soit en phase avec la stratégie de l'organisation. Ce programme devra avoir des objectifs et des critères de réussite définis ainsi que des sponsors au niveau de la direction générale. Le *business case* devrait également inclure une analyse coût-bénéfice par rapport aux outils et technologies déjà en place dans l'organisation.

Un fort soutien dans l'organisation est crucial pour obtenir dans la durée les ressources nécessaires et une priorisation adéquate des activités. Le *business case* devrait également prévoir le support technologique, avec une vérification précise des options et des coûts présentés, ainsi qu'une définition claire du propriétaire pour un portage durable du programme après son déploiement. De nombreuses organisations ont désormais un responsable des données *(Chief Data Officer)* chargé de s'assurer que les programmes de *Big Data* bénéficient du soutien nécessaire.

Rôles et responsabilités définis

En définissant clairement les rôles et les responsabilités des ressources clés et des fonctions clés, il est possible d'accélérer et de simplifier le déploiement des programmes de *Big Data* ainsi que le soutien qui leur est nécessaire. Ainsi, lorsqu'une organisation envisage d'analyser les activités et les comportements de ses collaborateurs (par exemple dans le cadre d'investigation de fraudes), il peut s'avérer nécessaire de consulter la direction des ressources humaines et des juristes. L'équipe du marketing peut vouloir mener ou être impliqué dans un test d'analyse de l'engagement client pour jauger la réaction des clients. La direction des systèmes d'information peut avoir des priorités concurrentes qui l'empêchent de venir en soutien des actions planifiées. Les fonctions chargées du management des risques, de la sécurité et de la protection des données peuvent souhaiter examiner et évaluer les contrôles dans le cadre du *Big Data*.

Si les principales parties prenantes au niveau des métiers et les consommateurs ne sont pas convaincus de leurs responsabilités et de leur contribution à l'effort global, les perspectives résultant de l'analyse des données risquent de rester lettre morte après le déploiement du programme. Par conséquent, avant d'investir massivement dans le *Big Data,* les organisations devraient échanger avec toutes les parties prenantes concernées pour confirmer leur engagement, déterminer la valeur ajoutée, se renseigner sur les exigences, sélectionner les priorités, évaluer la marge de manœuvre pour progresser malgré les autres priorités et susciter l'action pendant et après la mise en œuvre du programme de *Big Data*. Avant cette mise en œuvre, une analyse formelle des parties prenantes peut être nécessaire pour identifier toutes celles qui sont concernées. Le **schéma 4** donne des exemples des principales parties prenantes qui peuvent contribuer à orienter ou à soutenir des initiatives en matière de *Big Data*.

Schéma 4 - Exemples de parties prenantes du *Big Data*

Sponsor du projet
Responsable, de niveau direction générale, qui mobilise les soutiens et les ressources nécessaires

Métiers / propriétaires de données
Les propriétaires de données qui s'engagent dans la consolidation et l'intégration des données dans une solution unique pour l'atteinte des objectifs de l'organisation

Analystes opérationnels
Les personnes qui assurent le maintien des connaissances concernant les besoins métiers et les capacités technologiques afin de transformer les exigences opérationnelles en solutions de *Big Data*

Les consommateurs (par exemple, le marketing)
Toute fonction au sein de l'organisation qui consomme des données et / ou utilise les résultats des analyses effectuées

Directeurs des systèmes d'information
Responsable, de niveau direction générale, qui est chargé de la fourniture de la solution technologique et de la relation avec les prestataires en cas de sous-traitance

Responsable de la protection des données personnelles / responsable de la sécurité des systèmes d'information
Responsables, de niveau direction générale, qu'il convient de consulter à propos des contrôles liés à la sécurité, la protection et l'utilisation des données et des résultats des analyses

Responsable des données
Responsable, de niveau direction générale, qui pilote la gouvernance des données à l'échelle de l'organisation

Techniciens chargés de l'analyse des données / analystes de données
Par exemple : les administrateurs de bases de données, les développeurs de logiciels, des administrateurs d'outils techniques et des programmeurs

Data scientist
Un expert de l'analyse de données qui comprend les technologies et les processus métier, et peut développer et soutenir des analyses innovantes (par exemple, des analyses prédictives) pour stimule la création de valeur au niveau opérationnel

Ressources adéquates

Les organisations doivent disposer de ressources suffisantes pour soutenir la mise en œuvre du *Big Data*, ce qui implique davantage que les fonds nécessaires à l'achat du matériel pour stocker les données et traiter des travaux d'envergure, complexes et analytiques. L'utilisation de ressources et/ou de solutions tierces est à prendre en compte lors de l'élaboration de l'approche globale du programme, car en fonction de la taille et de l'ampleur du programme, la sous-traitance peut constituer une option plus rapide, évolutive et rentable.

La direction des ressources humaines contribue souvent au recrutement et à la sélection des profils appropriés pour les postes concernant le *Big Data*, à la définition des rémunérations et des avantages et aux programmes de formation et de développement des compétences en matière de *Big Data*. Les organisations font face à de nombreux défis lors de ces recrutements, en raison de la complexité technique des postes et de la rareté des talents sur le marché du travail.

Indicateurs de performance

Les critères de réussite établis lors de la conception du programme devront faire l'objet d'un suivi avec les indicateurs de performance pré-définis. Ils doivent permettre d'évaluer de manière équilibrée les performances opérationnelles et organisationnelles. Ils devraient également fournir au management un aperçu du coût, du niveau d'acceptation, de la disponibilité et de l'utilisation de la solution de *Big Data* à l'échelle de l'organisation.

L'évaluation d'initiatives ou de projets pilotes en matière de *Big Data* par des indicateurs opérationnels pertinents (par exemple, réduction des coûts d'attraction et de fidélisation des clients, croissance de la part de marché et augmentation des taux de clics) apportent une preuve tangible du retour sur investissement. Cependant, les avantages relatifs à la qualité devraient faire également l'objet d'un suivi, car ils permettent de mettre en exergue la valeur du programme.

Des risques surviennent souvent lorsque les indicateurs ne sont pas entièrement et explicitement conçus en fonction du *business case*. Lorsque la définition des mesures de performance est inadéquate ou qu'elles sont utilisées pour les rémunérations, certains aspects du programme de *Big Data* peuvent être privilégiés ou faire l'objet d'un surinvestissement au détriment d'autres domaines. La direction doit être extrêmement attentive lors de l'établissement des critères de succès et de soutien de tout programme clé, y compris les initiatives de *Big Data*, car ces indicateurs peuvent aussi bien induire des comportements appropriés que des comportements inappropriés.

Outils et technologies

Il est essentiel pour les organisations d'identifier les outils et les technologies les plus appropriés pour répondre à leurs besoins actuels et futurs. Ce type d'outils devrait permettre à l'organisation d'obtenir, de traiter, d'analyser et d'utiliser des quantités croissantes de données structurées et non structurées provenant de sources diverses. Par exemple, les outils de *reporting*, d'analyse visuelle, de surveillance et d'automatisation des tâches peuvent améliorer l'expérience de l'utilisateur et réduire le temps consacré à maintenir les environnements. Un stockage adéquat est également

indispensable car le volume de données peut se développer rapidement et de manière significative au fur et à mesure de l'enrichissement du programme. Les solutions technologiques doivent être cohérentes avec les exigences opérationnelles définies au préalable en fonction de plusieurs variables, y compris les utilisations planifiées et futures, la taille de l'organisation et bien d'autres facteurs.

Solutions de stockage

Le défi majeur que pose depuis toujours la quantité de données en constante augmentation, est le besoin croissant de stockage supplémentaire et de capacités suffisantes de sauvegarde. Compte tenu de la diversification des sources de données et de la demande croissante de données, des outils et des technologies nouveaux et plus puissants sont devenus une nécessité. Des données provenant de nouvelles sources et générées par des machines plutôt que par des êtres humains bousculent les technologies et outils traditionnels car ces données nécessitent de nouvelles capacités et solutions permettant de les stocker et de les rendre facilement accessibles.

Environnement sur site ou environnement *cloud*

Les plates-formes et les outils dont disposent les organisations évoluent au fil du temps. Cependant, les principales considérations technologiques restent souvent les mêmes quelles que soient les solutions. Par exemple, les organisations doivent choisir entre les environnements de *Big Data* sur site ou dans le *cloud* et prendre en considération les différentes modalités d'obtention des ressources nécessaires au développement des analyses. Les solutions sur site nécessitent une installation capable d'héberger un grand nombre de serveurs et une équipe informatique de maintenance de l'infrastructure. L'installation devrait être d'une taille suffisamment importante pour prendre en charge l'évolutivité de l'utilisation du *Big Data*. Le *cloud* constitue une autre option certes plus économique et plus rapide mais qui ne va pas sans des risques supplémentaires qui doivent être évalués dans leur ensemble avant de décider de recourir à cette solution. Les solutions « *Big Data* en tant que service » dans le *Cloud* (BDaaS- *Big Data as a Service*) offrent une très grande flexibilité. Les solutions BDaaS peuvent se décliner en :

- IaaS – Infrastructure en tant que service (par exemple, le matériel, les périphériques de stockage et les composants de réseau pour le *Big Data*).
- PaaS – Plate-forme en tant que service (par exemple, les systèmes d'exploitation, les plates-formes de développement et le *middleware*).
- SaaS[3] – Logiciel en tant que service (par exemple, des applications de traitement de *Big Data* ou d'analyses et de *reporting*).

Les organisations peuvent choisir d'utiliser un service dans le cloud (IaaS, PaaS ou SaaS) pour compléter les systèmes internes ou intégrer les trois services cloud pour développer l'ensemble de la solution *Big Data*.

[3] Pour de plus amples informations sur le cloud computing (« informatique dans les nuages »), reportez-vous à la publication spéciale SP800-145 du National Institute of Standards and Technology (NIST) "The NIST Definition of Cloud Computing", http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Outils d'exploration de données

La grande variété et le volume de données disponibles aujourd'hui sont le résultat d'un stockage de données exponentiellement moins coûteux et d'une collecte omniprésente, grâce à la croissance rapide des plates-formes des réseaux sociaux, des capteurs et du multimédia. Malheureusement, les entrepôts de données traditionnels et les solutions de business intelligence n'ont pas été conçus pour satisfaire aux exigences grandissantes en matière de données et ont été dépassés par la vague d'exigences relatives au stockage et au traitement des données. Par conséquent, les organisations qui continuent d'utiliser des entrepôts de données risquent de fonctionner et de prendre des décisions à partir de données incomplètes.

L'exploration du *Big Data* repose sur trois éléments principaux : (1) comprendre les données disponibles ; (2) les obtenir ; et (3) les mettre à profit pour développer des perspectives pertinentes conduisant à des pistes d'action. Les niveaux de maturité des organisations diffèrent quant aux capacités de gestion et de compréhension des données structurées internes. De nombreuses organisations ont énormément de mal à gérer les données non structurées ou les données externes. C'est dans ce domaine que les technologies de *Big Data* et les organisations dotées de programmes de *Big Data* efficaces prospèrent. L'identification et l'obtention de ces données requièrent une réflexion novatrice, un développement ou une configuration d'interfaces de programmes d'applications (API) et d'éventuels frais d'abonnement auprès de fournisseurs de données. L'obtention de l'ensemble des données disponibles est l'une des approches, mais il est préférable pour les organisations ayant des ressources limitées, de commencer par une utilisation pilote spécifique et de développer progressivement le programme.

Le traitement des données distribuées[4] et un apprentissage machine[5] amélioré augmentent la valeur du *Big Data*. Ces avancées informatiques peuvent aider les organisations à identifier des schémas méconnus par les humains et les applications à faible capacité. En outre, de nouveaux outils de visualisation de données sont inclus dans les solutions de *Big Data* pour plus de flexibilité, d'interaction et d'analyse ad hoc.

Outils de suivi

Lors de la mise en œuvre, il est important de définir les indicateurs clés de performance (KPI) pour les systèmes de *Big Data* et les analyses afin de permettre un suivi continu de la production. Les outils de suivi devraient être utilisés pour générer des rapports sur l'état et le fonctionnement de l'environnement du *Big Data* et fournir les informations nécessaires pour identifier de manière proactive et atténuer les risques opérationnels associés au *Big Data*. Les outils de suivi devraient permettre de générer des rapports sur les anomalies concernant différents aspects de la plateforme de *Big Data*, ainsi que sur le traitement des tâches. Comme indiqué précédemment, les KPI devraient être conçus pour rendre compte de l'efficacité et de la performance des systèmes de *Big Data*.

[4] Le traitement des données distribuées fait référence à plusieurs ordinateurs situés au même endroit ou dispersés et qui partagent leur capacité de traitement pour accélérer des traitements informatiques.

[5] *Machine learning* ou « apprentissage machine » fait référence aux programmes informatiques capables d'apprendre des algorithmes sans avoir besoin d'interaction humaine pour la programmation.

Obtention des logiciels

Les activités de développement ou d'achat et de personnalisation de logiciels de *Big Data* sont très différentes des systèmes traditionnels. Des technologies *open source* peuvent être gratuitement téléchargées à partir de plusieurs emplacements. Des éléments supplémentaires sont également disponibles gratuitement ou à l'achat auprès de fournisseurs. Bien qu'elles puissent être attrayantes, les solutions téléchargées gratuitement ne bénéficient généralement pas d'un support produit ou technique.

Les offres proposées sont difficiles à comprendre et à différencier en raison de la diversité des fonctionnalités et des caractéristiques des offres et des nombreuses personnalisations proposées par les fournisseurs des différentes plates-formes. Par exemple, les composants logiciels de requête structurés ne sont pas systématiques, et certains fournisseurs proposent de meilleures fonctions de sécurité que d'autres. Afin de faire son choix il est impératif de comprendre ces différences au regard des exigences du programme de *Big Data*. En mettant en œuvre une solution sur site ou dans le cloud, les DSI devraient évaluer attentivement les exigences en matière de *Big Data* et éviter d'acheter des logiciels, une puissance de traitement et une capacité de stockage inutiles.

Bien que le matériel lié au *Big Data* soit banalisé aux fins du traitement distribué, la complexité des logiciels sous-jacents augmente l'importance de la phase de conception et de développement de la solution. Des logiciels supplémentaires sont presque toujours intégrés dans les plates-formes de *Big Data*. Ces modules logiciels supplémentaires fournissent des fonctionnalités étendues concernant la façon de gérer, d'interagir et d'analyser les données ainsi que de présenter les résultats. Les programmes de *Big Data* comportent de plus en plus souvent un logiciel de visualisation de données spécialisé pour présenter les résultats sous la forme de tableaux de bord.

Les fournisseurs de *Big Data* aident les organisations à naviguer dans l'environnement technique, à gérer la personnalisation, l'abondance des outils logiciels, les nombreuses interfaces de données et à modéliser les données complexes. Malgré tout, les organisations devront identifier des collaborateurs en interne (par exemple, les gestionnaires de programmes de *Big Data*) qui disposent de connaissances suffisantes pour travailler avec les fournisseurs de *Big Data* tout au long du cycle de développement. Souvent, des *data scientists* sont recrutés pour aider dans le développement des modèles analytiques.

Soutien permanent au programme

Les solutions de *Big Data* ne sont pas vouées à rester statiques une fois développées et leur coût de production ne doit pas rester élevé. Pourtant, comme c'est le cas de nombreuses technologies de transformation *open source*, le rythme soutenu de changements dans le paysage du *Big Data* crée des défis qui dépassent souvent la capacité des architectes du *Big Data* à suivre des dizaines de nouveaux outils, des *plug-ins* et le rythme soutenu de lancement de nouvelles versions de produits.

Par conséquent, le soutien continu des ressources internes ou des fournisseurs est nécessaire pour assurer le succès permanent du programme. Ce soutien comprend les opérations informatiques

traditionnelles, telles que la planification de la capacité (c'est-à-dire la flexibilité de mise à l'échelle), la surveillance de la production et la planification de la reprise après sinistre. En outre, les sources de données internes et externes sont régulièrement ajoutées, supprimées ou modifiées. Les infrastructures support de stockage des données et les intégrations de données associées doivent être évaluées et ajustées en conséquence. Les méthodes standard de changement d'application et de gestion des correctifs s'appliquent également (cf. GTAG « Contrôles de la gestion du changement et des patchs : Un facteur clé de la réussite pour toute organisation », 2ème édition). Enfin, les modèles analytiques eux-mêmes doivent faire l'objet d'un suivi et d'une maintenance.

Gouvernance des données

L'adoption du *Big Data* dans une organisation nécessite une gestion renforcée des données afin de garantir que l'information demeure exacte, cohérente et accessible. Il existe plusieurs domaines clés où la gouvernance du *Big Data* est essentielle ; Il s'agit notamment de la gestion de la méta-donnée (c'est-à-dire des données sur les données), de la sécurité, de la protection, de l'intégration et de la qualité des données ainsi que de la gestion des données des fichiers maîtres. Les principales activités classiques de gouvernance des données correspondant à ces domaines comprennent l'identification des propriétaires de données, des consommateurs, des éléments de données critiques (CDE - *Critical Data Elements)*, des exigences particulières de gestion, des lignées, des données des fichiers maîtres et des sources de données autorisées. Les organisations doivent mettre en place des contrôles appropriés pour s'assurer que toutes les dimensions nécessaires à la qualité des données (par exemple, l'exhaustivité, la validité, l'unicité) sont correctement maintenues et que ces contrôles protègent de la même manière l'ensemble des CDE. Certains exemples de processus de contrôle liés à la qualité des données et à la protection des CDE incluent l'identification des défauts de données et la prévention des pertes et des fuites de données (DLP - *data loss/Leakage prevention).*

La principale spécificité de la gouvernance des données dans un contexte de *Big Data,* par rapport aux programmes traditionnels, concerne l'agilité dont les organisations doivent faire preuve tout au long du cycle de vie des données pour répondre aux demandes d'analyse. Les risques associés à cette exigence d'agilité sont aggravés par des caractéristiques ou des exigences opérationnelles propres à chaque série de données, notamment en matière de protection des données et de sécurité ou pour répondre à un besoin particulier des métiers.

Les propriétaires des données devraient assumer la responsabilité de la qualité et de la sécurité de leurs données, en mettant l'accent sur les éléments de données qui présentent le plus de risque. Ces éléments devraient être déterminés à l'issue d'un processus d'évaluation des risques, qui peut être dirigé par les propriétaires de données ou d'autres fonctions au sein de l'organisation (par exemple, la sécurité des systèmes d'information). Les propriétaires des données doivent s'assurer que les systèmes d'enregistrement sont correctement définis et que les processus de mise à jour des CDE sont clairs. Ils devraient notamment inclure l'identification des sources de données autorisées (c'est-à-dire celles qui définissent les données système qui ont la priorité lorsque les éléments de données varient ou entrent en conflit entre deux systèmes ou plus). Certaines organisations désignent des « gestionnaires de données » afin de soutenir ces efforts et d'autres aspects de la gouvernance de données.

En fin de compte, l'objectif est de permettre le transfert rapide d'informations tout en conservant un niveau élevé de qualité et de sécurité. Cela nécessite une gouvernance agile des données qui assure la mise en place de contrôles appropriés pour soutenir la durabilité et la proposition de valeur de ces programmes.

Appropriation par le consommateur

L'analyse du *Big Data* peut être mise en œuvre pour une utilisation interne (pour l'aide à des décisions liées aux opérations, au marketing, aux ressources humaines ou aux technologies de l'information) ou pour répondre aux besoins des clients (par exemple, analyser l'historique des comportements d'achat peut permettre de recommander de nouveaux produits ou services lorsque les clients visitent le site Internet de l'organisation). Quoi qu'il en soit, l'objectif d'une solution d'analyse du *Big Data* devrait être de fournir des informations significatives pour les consommateurs de données en interne (c'est-à-dire, les collaborateurs) et à l'extérieur (c'est-à-dire, les clients ou les fournisseurs) et d'améliorer l'aide à la décision. Les solutions de *Big Data* qui créent de la valeur seront durablement adoptées dans l'organisation.

Comme pour tout autre projet informatique, ce processus doit commencer par une identification exhaustive des besoins afin de comprendre les questions auxquelles le consommateur essaie de répondre ou les problèmes que le consommateur essaie d'anticiper. L'implication des consommateurs internes dans le processus de conception et de test les aide à s'approprier la solution. Toutes les réactions post-déploiement devront également être traitées rapidement afin de soutenir et d'améliorer l'appropriation. Les organisations devraient planifier des campagnes de promotion et de formation pour partager les expériences réussies et former les consommateurs internes sur le potentiel de l'analyse du *Big Data*. Les enquêtes auprès des parties prenantes sont un outil efficace pour obtenir des commentaires et des retours d'expérience afin d'améliorer les processus de développement des futurs projets.

Analyse et *reporting*

Les paramètres d'entrée des rapports devraient être suffisamment flexibles (par exemple, les dates de début et de fin, les segments de clients et les produits) pour permettre aux consommateurs d'affiner ou d'élargir leur analyse. Cette flexibilité permet aux consommateurs de poser des questions qui n'ont peut-être pas été anticipées au cours de la phase initiale de développement et facilite l'adoption en permettant aux consommateurs de disposer de capacités en libre-service, ce qui permet de minimiser le cycle de développement de rapports traditionnellement lents et coûteux. La granularité des rapports pour répondre à des besoins de *reporting* standard ou détaillés des consommateurs devrait être équilibrée en fonction des exigences du consommateur, des capacités de traitement et des questions de protection des données.

Les outils en libre-service sont importants pour les activités qui concernent des clients, des fournisseurs ou des collaborateurs qui doivent prendre rapidement des décisions. Par exemple, un représentant du service à la clientèle peut utiliser le *Big Data* et une application de *reporting* en libre-service pour afficher l'historique de distribution multicanal des produits et des services sur

un seul écran. Le nombre d'appels pour répondre à ces différentes requêtes est ainsi réduit. Les problèmes de protection des données et de sécurité peuvent être résolus en restreignant l'accès aux champs de données sensibles aux consommateurs qui ont un besoin valable de les visualiser.

De nombreux utilisateurs sont habitués au concept de l'analyse prédictive, qui tente d'expliquer ce qui va se produire en fonction des données historiques. Par exemple, les hôpitaux utilisent des analyses prédictives pour déterminer quels patients peuvent être admis pour un traitement supplémentaire. Les *data scientists* peuvent utiliser des algorithmes d'analyse de survie pour aider les départements des ressources humaines à prédire l'insatisfaction des collaborateurs et cette information peut être utilisée pour gérer les effectifs et planifier les activités.

Les rapports analytiques peuvent également être générés à partir d'alertes, afin d'aider les consommateurs à identifier les actions nécessaires pour résoudre une situation particulière. Par exemple, les techniques d'analyse des sentiments peuvent être utilisées pour déterminer la satisfaction d'un client concernant un produit ou un service, à partir d'informations partagées par les clients sur les réseaux sociaux. Des taux de satisfaction élevés peuvent générer de nouvelles stratégies de distribution, alors que des indicateurs médiocres peuvent nécessiter des mesures correctives immédiates pour fidéliser la clientèle et renforcer la réputation de l'organisation.

D'autres techniques d'analyse avancées, telles que l'analyse des liens, peuvent être utilisées par les auditeurs internes et les investigateurs de fraude pour mieux comprendre les relations entre les employés, les fournisseurs et les clients, détecter le blanchiment d'argent ou d'autres schémas de fraude. Les scénarios d'analyse sont illimités et offrent d'innombrables possibilités de création de valeur pour le consommateur final grâce à différents canaux et options de *reporting*.

Processus cohérents

Des processus définis et contrôlés sont nécessaires pour assurer le succès continu d'un programme de *Big Data*. Sans un processus défini et cohérent, les environnements peuvent rapidement devenir instables et provoquer la perte de confiance dans les données sous-jacentes. Lorsque les rapports d'analyse conçus pour fournir des informations stratégiques utilisent des données inexactes ou incomplètes, leur valeur est immédiatement compromise. Si ce type de risques n'est pas pris en compte, le financement des programmes de *Big Data* peut rapidement cesser et être voués à disparaître. Par conséquent, au fur et à mesure du déploiement des personnes et de la technologie, une attention particulière devrait également être accordée aux processus sous-jacents et aux contrôles connexes qui soutiennent le programme de *Big Data*.

Principaux risques

Les risques liés au Big Data peuvent résulter de nombreux facteurs internes et externes à l'organisation. Les principaux domaines de risque correspondent aux catégories suivantes :

- Gouvernance du programme
- Disponibilité et performance de la technologie
- Sécurité et protection des données
- Qualité, gestion et *reporting* des données

Domaine	Principaux risques	Activités de contrôle
Gouvernance du programme	L'insuffisance du soutien du management, du financement et / ou de la gouvernance d'un programme de *Big Data* peut avoir pour effet d'exposer l'organisation à un risque excessif ou empêcher la réalisation des objectifs stratégiques.	• Le financement devrait être suffisant pour répondre aux besoins des métiers • Les objectifs du programme doivent accompagner les initiatives stratégiques à l'échelle de l'organisation • Le management devrait recevoir des indicateurs témoignant de la réalisation des objectifs • L'organisation devrait mettre en place une entité responsable de la gestion du *Big Data* • Des contrats de niveau de service (SLA- *Service Level Agreement)* permettant de définir et de mesurer les attentes en matière de performance devraient être conclus entre les métiers et la DSI • Les exigences opérationnelles et techniques doivent être documentées, analysées et approuvées • La direction générale devrait développer une stratégie de *Big Data* qui fournit des solutions pour l'ensemble de l'organisation • Avant d'approuver le *business case,* la direction devrait démontrer la preuve de l'intérêt du concept au regard des objectifs stratégiques • Les rôles et les responsabilités doivent être clairs et bien définis • L'organisation devrait fournir les ressources nécessaires pour déployer et maintenir la stratégie de *Big Data* • Les bonnes pratiques de gestion des prestations externalisées devraient être utilisées dans les relations avec les fournisseurs du *Big Data* • La gouvernance des données devrait faire partie de la gouvernance générale de l'organisation afin de s'assurer que les objectifs en matière de *Big Data* sont en phase avec les objectifs stratégiques de l'organisation (cf. Norme 2110 - Gouvernance)

Domaine	Principaux risques	Activités de contrôle
Disponibilité et performance de la technologie	Des solutions et / ou des configurations technologiques inefficaces risquent d'entraîner une expérience client négative, de réduire la disponibilité système et/ou de dégrader les performances.	• Les processus informatiques devraient être structurés de manière à permettre la prise en charge des attentes de niveau de service relatif au *Big Data* • Les politiques et les procédures relatifs au cycle de vie des données devraient être documentées et faire l'objet d'un suivi • Les systèmes de *Big Data* devraient être intégrés dans la stratégie de maintenance • Les systèmes de *Big Data* devraient être intégrés dans la stratégie de gestion du changement • Les systèmes de *Big Data* devraient être intégrés dans la stratégie de gestion des correctifs • Les systèmes de *Big Data* devraient être acquis, développés et/ou configurés en fonction de la complexité et des exigences documentées dans le *business case* • Les systèmes et les outils de support devraient être configurés pour fournir des notifications automatiques aux services support • Les outils de *reporting* doivent être configurés pour être flexibles, intuitifs et faciles à utiliser • Les systèmes de *Big Data* devraient être configurés pour permettre une flexibilité et une évolutivité sans compromettre les performances • Des tests périodiques de performance devraient être effectués en vue de remédier aux faiblesses • Le cycle de vie des systèmes de *Big Data* devrait être correctement géré • Les contrôles généraux informatiques devraient être évalués périodiquement
Sécurité et protection des données	Des normes et des configurations de sécurité d'information inefficaces peuvent entraîner un accès non autorisé aux données et le vol de celles-ci, des modifications inappropriées des données et des violations de la conformité réglementaire.	• La gestion de la sécurité de l'information devrait être intégrée dans la stratégie de *Big Data* • La gestion de la sécurité de données devrait être intégrée dans la stratégie de *Big Data* • L'accès de tiers doit être géré de manière adéquate • La protection des données devrait être intégrée dans la stratégie de *Big Data*

Domaine	Principaux risques	Activités de contrôle
Qualité, gestion et *reporting* des données	Les problèmes de qualité des données et/ou l'inexactitude des rapports peuvent entraîner l'inexactitude des *reportings* effectués par le management et fausser le processus de prise de décision.	• Il convient d'établir des politiques et des procédures visant à garantir la qualité des données • Des politiques et des procédures devraient être établies pour s'assurer que les données obtenues auprès de tiers respectent les normes de qualité des données • Il convient d'établir des politiques et des procédures visant à garantir l'exactitude des rapports • L'accès aux rapports devrait être accordé en fonction des besoins opérationnels • Les outils et procédures de *reporting* devraient permettre la flexibilité et l'élaboration de rapports *ad hoc* • Les utilisateurs devraient être formés périodiquement pour optimiser l'utilité des rapports • La sélection des fournisseurs de produits et de services de *reporting* devrait refléter les besoins opérationnels

Le rôle de l'audit interne en matière de *Big Data*

L'audit interne devrait tenir compte du rôle du *Big Data* dans les organisations lors de l'évaluation des risques et de la planification des missions (cf. Norme 2010 - Planification et Norme 2010.A1). Lorsque les risques sont significatifs, l'audit interne peut établir un plan de couverture approprié des risques et des contrôles relatifs au *Big Data*. Ce faisant, l'audit interne saisit l'opportunité de sensibiliser le Conseil aux initiatives de *Big Data* de l'organisation, aux risques et défis qui en résultent, ainsi qu'à l'importance des opportunités et avantages qui en découlent. En règle générale, l'audit interne assure une couverture plus importante des enjeux du *Big Data* en effectuant plusieurs missions plutôt qu'une mission unique dédiée à ce sujet.

Comme dans toute initiative d'envergure, au fur et à mesure de la mise en œuvre des programmes de *Big Data*, l'audit interne devrait envisager d'y contribuer par des évaluations formelles et / ou informelles. Il peut notamment s'agir de missions de conseil, de revues pré- ou post-déploiement ou d'une participation adéquate aux comités de gouvernance et de pilotage. Comme indiqué dans la Norme 2130 - Contrôle, *« L'audit interne doit aider l'organisation à maintenir un dispositif de contrôle approprié en évaluant son efficacité ainsi que son efficience et en encourageant son amélioration continue »*. Ainsi, l'audit interne devrait évaluer les processus et les contrôles technologiques. Il devrait également particulièrement s'intéresser à la manière dont les données sont consommées et aux mesures prises par l'organisation en fonction des résultats de l'analyse du *Big Data*. Les auditeurs internes devraient jouer un rôle essentiel dans les initiatives de *Big Data* d'une organisation, et ce rôle peut être ajusté au fur et à mesure de la mise en œuvre, de l'affinement et de l'évolution des solutions (cf. Norme 2201 - Considérations relatives à la planification).

Les auditeurs internes peuvent également exploiter les solutions de *Big Data* pour leurs propres analyses de données. L'organisation ayant déjà obtenu, consolidé et intégré des données dans un entrepôt ou un lac de données, l'audit interne peut générer d'importants gains d'efficience en les utilisant plutôt qu'en ciblant de nombreux systèmes sources.

Les programmes d'audit de *Big Data* varient selon l'organisation et les usages. La gouvernance des programmes est un élément clé des programmes d'audit de *Big Data*. Les auditeurs internes doivent vérifier que les objectifs d'un programme de *Big Data* sont cohérents avec la stratégie de l'organisation. En outre, les auditeurs internes devraient effectuer des tests pour s'assurer que le programme de *Big Data* crée de la valeur et qu'il bénéficie d'un soutien approprié de la part de la direction compétente au sein de l'organisation. Bien que la technologie et le niveau de recours à des prestataires de solutions de *Big Data* varient d'une organisation à l'autre, les auditeurs internes devraient s'assurer que la confidentialité, l'intégrité, la disponibilité et la performance des systèmes de *Big Data* correspondent aux exigences et aux besoins opérationnels de la direction.

Les systèmes de *Big Data* requièrent d'importantes quantités de données à analyser, les programmes d'audit devraient donc inclure des étapes de test pour s'assurer de la qualité, de la sécurité et de la confidentialité des données utilisées pour les analyses, ainsi que des résultats desdites analyses. Étant donné que le *Big Data* utilise des données provenant de nombreuses

sources afin de fournir l'aperçu le plus complet d'un sujet, les programmes d'audit devraient fournir une assurance raisonnable que les données sont protégées contre toute modification non autorisée et que leur consultation est réservée uniquement aux personnes autorisées. Les programmes d'audit devraient également servir à tester les contrôles de la qualité des données d'entrées, des résultats et des rapports du système ; ces travaux peuvent inclure des tests de corroboration concernant la qualité des principales données du système et le *reporting*.

L'annexe D fournit un exemple de programme de travail qui aidera les auditeurs internes à lancer une revue du *Big Data* au sein de leur organisation. La liste d'activités de revue proposées n'est pas exhaustive car les programmes de travail doivent être personnalisés pour répondre aux besoins spécifiques de chaque organisation.

Les systèmes de *Big Data* et les documents de référence évoluent fréquemment. Par conséquent, l'audit interne devra veiller à utiliser les lignes directrices les plus appropriées lors de l'élaboration et de l'exécution de plans d'audit et de modèles de couverture relatifs au *Big Data*.

Annexe A. Principales normes IIA concernées

La sélection des *Normes internationales pour la pratique professionnelle de l'audit interne (les Normes)* figurant ci-après est pertinente en matière de *Big Data*. Les informations figurant dans chaque norme devront être complétées par le Guide de mise en œuvre correspondant.

Norme 1210 – Compétence

Les auditeurs internes doivent posséder les connaissances, les savoir-faire et les autres compétences nécessaires à l'exercice de leurs responsabilités individuelles. L'équipe d'audit interne doit collectivement posséder ou acquérir les connaissances, les savoir-faire et les autres compétences nécessaires à l'exercice de ses responsabilités.

Interprétation :
Compétence est un terme générique qui fait référence aux connaissances, aux savoir-faire et autres aptitudes dont les auditeurs internes doivent disposer pour pouvoir exercer efficacement leurs responsabilités professionnelles. La compétence concerne les activités existantes, les tendances et les problématiques émergentes. Elle permet de fournir des conseils et des recommandations pertinents. Les auditeurs internes sont encouragés à démontrer leurs compétences en obtenant des certifications et qualifications professionnelles appropriées telles que le CIA (Certified Internal Auditor) et tout autre diplôme promu par l'IIA (The Institute of Internal Auditors) ou par d'autres organisations professionnelles appropriées.

> **1210.A3 –** Les auditeurs internes doivent posséder des connaissances suffisantes des principaux risques et contrôles relatifs aux systèmes d'information, et des techniques d'audit informatisées susceptibles d'être mises en œuvre dans le cadre des travaux qui leur sont confiés. Toutefois, tous les auditeurs internes ne sont pas censés posséder l'expertise d'un auditeur dont la responsabilité première est l'audit informatique.

Norme 2010 – Planification

Le responsable de l'audit interne doit établir un plan d'audit fondé sur une approche par les risques afin de définir des priorités cohérentes avec les objectifs de l'organisation.

Interprétation :
Pour établir le plan d'audit fondé sur une approche par les risques, le responsable de l'audit interne consulte la direction générale et le Conseil et prend connaissance de la stratégie, des principaux objectifs opérationnels, des risques associés ainsi que des processus de management des risques.
Le responsable de l'audit interne doit, le cas échéant, réviser et ajuster le plan afin de répondre aux changements dans les activités, les risques, les opérations, les programmes, les systèmes et les contrôles de l'organisation.

2010.A1 – Le plan d'audit interne doit s'appuyer sur une évaluation des risques documentée et réalisée au moins une fois par an. Les points de vue de la direction générale et du Conseil doivent être pris en compte dans ce processus.

2010.A2 – Le responsable de l'audit interne doit identifier et prendre en considération les attentes de la direction générale, du Conseil et des autres parties prenantes concernant les opinions et d'autres conclusions de l'audit interne.

2010.C1 – Lorsqu'on lui propose une mission de conseil, le responsable de l'audit interne, avant de l'accepter, devrait considérer dans quelle mesure elle est susceptible de créer de la valeur ajoutée, d'améliorer le management des risques et le fonctionnement de l'organisation. Les missions de conseil qui ont été acceptées doivent être intégrées dans le plan d'audit.

Norme 2030 – Gestion des ressources

Le responsable de l'audit interne doit veiller à ce que les ressources affectées à l'audit interne soient adéquates, suffisantes et mises en œuvre de manière efficace pour réaliser le plan d'audit approuvé.

Interprétation :
On entend par ressources adéquates, la combinaison de connaissances, les savoir-faire et autres compétences nécessaires à la réalisation du plan d'audit. On entend par ressources suffisantes, la quantité de ressources nécessaires à la réalisation du plan d'audit. Les ressources sont mises en œuvre efficacement quand elles sont utilisées de manière à optimiser la réalisation du plan d'audit.

Norme 2100 – Nature du travail

L'audit interne doit évaluer les processus de gouvernance de l'organisation, de management des risques et de contrôle, et contribuer à leur amélioration sur la base d'une approche systématique, méthodique et fondée sur une approche par les risques. La crédibilité et la valeur de l'audit interne sont renforcées lorsque les auditeurs internes sont proactifs, que leurs évaluations offrent de nouveaux points de vue et prennent en considération les impacts futurs.

Norme 2110 – Gouvernance

L'audit interne doit évaluer et formuler des recommandations appropriées en vue d'améliorer les processus de gouvernance de l'organisation pour :

- les prises de décisions stratégiques et opérationnelles ;
- la surveillance des processus de management des risques et de contrôle ;
- La promotion des règles d'éthique et des valeurs appropriées au sein de l'organisation ;
- Une gestion efficace des performances de l'organisation, assortie d'une obligation de rendre compte ;
- La communication aux services concernés des informations relatives aux risques et aux contrôles ;
- La coordination des activités et la communication des informations entre le Conseil, les auditeurs externes, les auditeurs internes, les autres prestataires d'assurance, et le management.

2110.A2 – L'audit interne doit évaluer si la gouvernance des systèmes d'information de l'organisation soutient la stratégie et les objectifs de l'organisation.

Norme 2130 – Contrôle

L'audit interne doit aider l'organisation à maintenir un dispositif de contrôle approprié en évaluant son efficacité ainsi que son efficience et en encourageant son amélioration continue.

Norme 2200 – Planification de la mission

Les auditeurs internes doivent concevoir et documenter un plan pour chaque mission. Ce plan de mission précise les objectifs, le périmètre d'intervention, le calendrier de la mission, ainsi que les ressources allouées. Ce plan doit prendre en considération la stratégie, les objectifs et les risques de l'organisation pertinents au regard de la mission.

Norme 2201 – Considérations relatives à la planification

Lors de la planification de la mission, les auditeurs internes doivent prendre en compte :

- la stratégie, les objectifs de l'activité auditée et la manière dont elle pilote sa performance ;
- les risques significatifs liés aux objectifs de l'activité, à ses ressources et à ses opérations, ainsi que les moyens par lesquels l'impact potentiel des risques est maintenu à un niveau acceptable ;
- la pertinence et l'efficacité des processus de gouvernance, de management des risques et de contrôle de l'activité, en référence à un cadre ou un modèle approprié ;
- les opportunités d'améliorer de manière significative les processus de gouvernance, de management des risques et de contrôle de l'activité.

2201.C1 – Les auditeurs internes doivent conclure avec le client donneur d'ordre un accord sur les objectifs et le périmètre d'intervention de la mission de conseil, les responsabilités de chacun et plus généralement sur les attentes du client donneur d'ordre. Pour les missions importantes, cet accord doit être formalisé.

Norme 2210 – Objectifs de la mission

Les objectifs doivent être précisés pour chaque mission.

2210.A1 – Les auditeurs internes doivent procéder à une évaluation préliminaire des risques liés à l'activité soumise à l'audit. Les objectifs de la mission doivent être déterminés en fonction des résultats de cette évaluation.

2210.A2 – En déterminant les objectifs de la mission, les auditeurs internes doivent tenir compte de la probabilité qu'il existe des erreurs significatives, des cas de fraudes ou de non-conformités et d'autres risques importants.

2210.C1 – Les objectifs d'une mission de conseil doivent porter sur les processus de gouvernance, de management des risques et de contrôle dans la limite convenue avec le client.

2210.C2 – Les objectifs de la mission de conseil doivent être en cohérence avec les valeurs, la stratégie et les objectifs de l'organisation.

Norme 2220 – Périmètre d'intervention de la mission

Le périmètre d'intervention doit être suffisant pour atteindre les objectifs de la mission.

2220.A1 – Le périmètre d'intervention de la mission doit couvrir les systèmes, les données, les personnes et les biens concernés, y compris ceux qui se trouvent sous le contrôle de tiers.

Norme 2230 – Ressources affectées à la mission

Les auditeurs internes doivent déterminer les ressources appropriées et suffisantes pour atteindre les objectifs de la mission. Ils s'appuient sur une évaluation de la nature et de la complexité de chaque mission, des contraintes de temps et des ressources disponibles.

Interprétation :
On entend par ressources appropriées, la combinaison de connaissances, savoir-faire et autres compétences nécessaires à la réalisation de la mission. On entend par ressources suffisantes, la quantité de ressources nécessaires pour la réalisation de la mission avec conscience professionnelle.

Norme 2240 – Programme de travail de la mission

Les auditeurs internes doivent élaborer et documenter un programme de travail permettant d'atteindre les objectifs de la mission.

Norme 2310 - Identification des informations

Les auditeurs internes doivent identifier des informations suffisantes, fiables, pertinentes et utiles pour atteindre les objectifs de la mission.

Interprétation :
Une information suffisante est factuelle, adéquate et probante, de sorte qu'une personne prudente et informée pourrait parvenir aux mêmes conclusions que l'auditeur. Une information fiable est une information concluante et facilement accessible par l'utilisation de techniques d'audit appropriées. Une information pertinente conforte les constatations et recommandations de l'audit, et répond aux objectifs de la mission. Une information utile aide l'organisation à atteindre ses objectifs.

Annexe B. Lignes directrices complémentaires concernées

GTAG : Audit continu : Coordonner l'audit et le contrôle en continu pour fournir une assurance continue, 2ème édition

GTAG : Évaluer le risque de cybersécurité – Le rôle des trois lignes de maîtrise

GTAG : Audit des contrôles applicatifs

GTAG : Contrôles de la gestion du changement et des patchs : Un facteur clé de la réussite pour toute organisation, 2ème édition

GTAG : *Data Analysis Technologies*

GTAG : Élaboration d'un plan d'audit des SI

GTAG : Gestion des identités et des accès

GTAG : L'infogérance, 2ème édition

GTAG : Les risques et les contrôles des systèmes d'information, 2ème édition

Guide pratique : L'audit des risques d'atteinte à la vie privée, 2ème édition

Annexe C. Glossaire

Les termes marqués d'un astérisque (*) sont extraits du « Glossaire » des *Normes internationales pour la pratique professionnelle de l'audit interne*, édition 2017.

Analyse prédictive – Le processus consistant à essayer de prédire l'avenir à partir de l'analyse de données concernant des événements passés. Le processus peut faire appel à plusieurs techniques prédictives (par exemple, l'exploration de données, les statistiques, la modélisation et l'apprentissage machine).

Conseil* – Le niveau le plus élevé des organes de gouvernance (par exemple, un conseil d'administration, un conseil de surveillance ou un organe délibérant) responsable du pilotage, et/ou de la surveillance des activités de l'organisation et, envers lequel la direction générale a un devoir de rendre compte.
Bien que les mécanismes de gouvernance varient selon les juridictions et les secteurs, le Conseil comprend généralement des membres qui ne font pas partie du management. Si un tel Conseil n'existe pas, le mot « Conseil » dans les *Normes* correspond à un groupe ou à une personne en charge de la gouvernance de l'organisation. En outre, « Conseil » dans les *Normes* peut renvoyer à un comité ou une autre instance auquel l'organe de gouvernance a délégué certaines fonctions (par exemple, un comité d'audit).

Contrat de niveau de service (SLA- *Service level agreement)* – Un accord formalisé entre un fournisseur de services ou de produits et l'organisation qui décrit les objectifs minimaux de performance, le mécanisme permettant de mesurer les performances et les conséquences de performances en deçà des objectifs.

Contrôle* – Toute mesure prise par le management, le Conseil et d'autres parties afin de gérer les risques et d'accroitre la probabilité que les buts et objectifs fixés seront atteints. Le management planifie, organise et dirige la mise en œuvre de mesures nécessaires pour donner une assurance raisonnable que les buts et objectifs seront atteints.

Dimensions de la qualité des données – Le comité de la gestion des données *(Entreprise data management council)* a défini sept principales dimensions concernant la qualité des données : exhaustivité, couverture, conformité, cohérence, exactitude, réplication, rapidité.[6]

Données du fichier maître – L'ensemble des attributs utilisés pour identifier et décrire les données.

Données en continu *(streaming)* – Les données générées en continu par des milliers de sources de données, qui envoient généralement de manière simultanée les enregistrements de données.[7]

[6] http://www.edmcouncil.org/dataquality

[7] https://aws.amazon.com/streaming-data/

Données non structurées – Ce type de données ne se limite pas aux structures de données ou aux contraintes classiques. Elles sont généralement plus difficiles à gérer, en raison de leur nature évolutive et imprévisible mais aussi en raison du nombre et de la diversité de leurs sources souvent externes. Par conséquent, de nouvelles solutions ont été développées pour gérer et analyser ce type de données.

Données structurées – Traditionnellement, la majorité des données stockées dans les organisations sont structurées et conservées dans des bases de données relationnelles voire même hiérarchiques ou dans des fichiers à plat. Les données structurées sont organisées et permettent des requêtes répétitives parce que la plupart des données sont dans des tables relationnelles. Il est souvent plus facile de les contrôler par rapport aux données non structurées car leur propriétaire et les prestataires de solutions de gestion de bases de données sont identifiés.

Éléments de données – La plus petite unité de données qui transmet des informations significatives (par exemple, nom, numéro de compte).

Éléments de données critiques (CDE- *Critical data elements)* – Éléments de données essentiels pour permettre aux utilisateurs ou aux systèmes d'effectuer des calculs ou de mener des activités.

Environnement de contrôle* – L'attitude et les actions du Conseil et du management au regard de l'importance du dispositif de contrôle dans l'organisation. L'environnement de contrôle constitue le cadre et la structure nécessaires à la réalisation des objectifs primordiaux du système de contrôle interne. L'environnement de contrôle englobe les éléments suivants :

- intégrité et valeurs éthiques ;
- philosophie et style de direction ;
- structure organisationnelle ;
- attribution des pouvoirs et responsabilités ;
- politiques et pratiques relatives aux ressources humaines ;
- compétence des collaborateurs.

Extraire, charger et transformer (ELT- *Extract, load and transform)* – Le processus de transfert des données du système source vers un entrepôt de données. ELT fait référence à l'ordre dans lequel les étapes de mise en scène se produisent. Ce procédé utilise le système cible pour la transformation.

Extraire, transformer et charger (ETL- *Extract, transform and load)* – Le processus de transfert des données du système source vers un entrepôt de données. ELT fait référence à l'ordre dans lequel les étapes de mise en scène se produisent.

Evolutivité – La capacité des systèmes à changer de taille pour tenir compte de l'évolution de la demande de traitement.

Fraude* – Tout acte illégal caractérisé par la tromperie, la dissimulation ou la violation de la confiance sans qu'il y ait eu violence ou menace de violence. Les fraudes sont perpétrées par des personnes et des organisations afin d'obtenir de l'argent, des biens ou des services, ou de s'assurer un avantage personnel ou lié à leur activité.

Gouvernance* – Le dispositif comprenant les processus et les structures mis en place par le Conseil afin d'informer, de diriger, de gérer et de piloter les activités de l'organisation en vue de réaliser ses objectifs.

Gouvernance des données – L'exercice de l'autorité, du contrôle et de la prise de décision partagée (planification, suivi et exécution) dans la gestion des actifs informationnels. La gouvernance des données est le niveau le plus élevé de planification et de contrôle de la gestion des données.[8] La gouvernance des données est nécessaire à une gestion optimale des données tout en maintenant la qualité et l'exactitude.

Identifiant de fréquence radio (RFID - *Radio frequence ID)* – Petits périphériques dotés d'une antenne qui utilisent des champs électromagnétiques pour marquer et suivre les éléments et transmettre des informations aux appareils capables de lire leurs signaux.

Indicateur clé de performance (KPIs) – Une mesure permettant de déterminer dans quelle mesure la performance d'un processus permet d'atteindre un objectif particulier.

Interfaces de programmation d'applications (API- *Application programming interface)* – Un ensemble de routines, de protocoles et d'outils utilisés dans le développement de logiciels.

Management des risques* – Processus visant à identifier, évaluer, gérer et contrôler les évènements ou les situations potentiels pour fournir une assurance raisonnable quant à la réalisation des objectifs de l'organisation.

Mission d'assurance* – Il s'agit d'un examen objectif d'éléments probants effectué en vue de fournir à l'organisation une évaluation indépendante des processus de gouvernance, de management des risques et de contrôle. Par exemple, des audits financiers, de performance, de conformité, de sécurité des systèmes ou de *due diligence*.

Mission de conseil* – Conseils et services y afférents rendus au client donneur d'ordre, dont la nature et le périmètre d'intervention sont convenus au préalable avec lui. Ces activités ont pour objectifs de créer de la valeur ajoutée et d'améliorer les processus de gouvernance, de management des risques et de contrôle d'une organisation sans que l'auditeur interne n'assume aucune responsabilité de management. Quelques exemples : avis, conseil, assistance et formation.

Objectifs de la mission* – Enoncés généraux conçus par les auditeurs internes et définissant ce qu'il est prévu de réaliser pendant la mission.

[8] Le Guide de l'Association internationale de gestion des données (DAMA), *Data Management Body of Knowledge* (DAMA-DMBOK), 1ère Edition 2009, p.4

Opinion relative à une mission* – La note, la conclusion et/ou toute autre description des résultats d'une mission d'audit interne donnée, relative aux éléments rentrant dans le cadre des objectifs et du périmètre de la mission.

Prestataire externe* – Une personne ou une entreprise, extérieure à l'organisation, qui possède des connaissances, des savoir-faire et une expérience spécifiques dans une discipline donnée.

Prévention des pertes de données (DLP- *Data loss prevention)* – La série de contrôles mis en place pour s'assurer que les utilisateurs ne diffusent pas d'informations critiques en dehors de l'organisation.

Programme de travail de la mission* – Un document qui recense les procédures à mettre en œuvre, conçu pour réaliser le plan de mission.

Propriétaire des données – Personne ou entité responsable de la sauvegarde et de la classification des données ainsi que de l'octroi ou du refus de l'accès aux données.

Responsable de l'audit interne* – « Responsable de l'audit interne » désigne une personne, occupant un poste hiérarchique de haut niveau, qui a la responsabilité de diriger efficacement la fonction d'audit interne conformément à la charte d'audit interne et aux dispositions obligatoires du Cadre de référence international des pratiques professionnelles de l'audit interne. Le responsable de l'audit interne ou des personnes qui lui sont rattachées de l'audit interne devront disposer des certifications et des qualifications professionnelles appropriées. L'intitulé exact du poste et/ou de la fonction du responsable de l'audit interne varie selon les organisations.

Responsable des données (CDO- *Chief data officer)* – Poste de niveau cadre chargé de la gouvernance et de la gestion des données à l'échelle de l'organisation.

Risque* – Possibilité que se produise un évènement qui aura un impact sur la réalisation des objectifs. Le risque se mesure en termes de conséquences et de probabilité.

Système de référencement (SOR- *System of record)* – La source faisant autorité concernant un élément de données particulier.

Sources de données autorisées (ADS- *Authoritative data sources)* – Sources qui fournissent des données « officielles » à d'autres systèmes.

Traçabilité des données – La visualisation du traitement des données de bout en bout. La traçabilité formalise l'origine, la manipulation, le calcul, la transformation et la destination finale des données.

Valeur ajoutée* – L'audit interne apporte de la valeur ajoutée à l'organisation et à ses parties prenantes lorsqu'il fournit une assurance objective et pertinente et qu'il contribue à l'efficience ainsi qu'à l'efficacité des processus de gouvernance, de management des risques et de contrôle.

Annexe D. Planification d'une mission d'audit du *Big Data*

Pour réussir leur mission d'assurance ou donner des conseils sur les programmes de *Big Data,* les auditeurs internes doivent comprendre les risques et les défis connexes (cf. Norme 2200 - Planification de la mission et 2201 - Considérations relatives à la planification). La gravité du risque varie selon l'organisation et dépend du but stratégique et du déploiement opérationnel de ces initiatives (cf. Norme 2210 - Objectifs de la mission et Norme 2220 - Périmètre d'intervention de la mission). Les paragraphes suivants fournissent des détails supplémentaires sur les principaux domaines de risque, ainsi que des exemples d'éléments d'appréciation des risques et des contrôles à utiliser dans l'élaboration d'un programme de travail d'audit du *Big Data* (cf. Norme 2240 - Programme de travail de la mission).

Risques et défis en matière de *Big Data*

Les avantages potentiels de la mise en œuvre d'un programme de *Big Data* impliquent d'importants risques et défis. L'audit interne doit permettre de s'assurer que les risques de l'organisation sont identifiés, compris et traités de manière appropriée. La gestion des niveaux acceptables de risques en matière de *Big Data* augmente la probabilité d'atteindre les objectifs opérationnels planifiés et de bénéficier des avantages potentiels du programme de *Big Data*.

Comme décrit ci-dessus, les principaux domaines de risque touchant le *Big Data* sont :

- Gouvernance du programme
- Disponibilité et performance de la technologie
- Sécurité et protection des données personnelles
- Qualité, gestion et *reporting* des données

Les programmes de *Big Data* et les environnements devraient également être soumis à des contrôles généraux informatiques (cf. GTAG « Les risques et les contrôles des systèmes d'information », 2ème édition pour plus d'informations sur les risques et les défis liés aux contrôles généraux informatiques).

Planification de la mission

La Norme 2200 – Planification de la mission, prévoit que les auditeurs internes doivent concevoir et documenter un plan pour chaque mission. Ce plan de mission précise les objectifs, le périmètre d'intervention, le calendrier de la mission, ainsi que les ressources allouées. L'un des aspects les plus importants que l'audit interne doit déterminer lors de la planification de la mission est de savoir si l'organisation dispose d'une structure de gouvernance unifiée et cohérente, y compris des politiques et des procédures, à partir desquelles il est possible de tracer des lignes directrices claires et cohérentes applicables à l'ensemble de l'organisation. Un modèle de gouvernance solide fournira des politiques, des processus et des outils nécessaires permettant de gérer l'environnement de manière cohérente et de contrôler les risques liés au *Big Data*, ce qui est essentiel pour une protection adéquate de l'information détenue par l'organisation. Bien que plusieurs fonctions puissent être responsables d'une partie de la stratégie de *Big Data,* la clé de la réussite de l'audit du *Big Data*

consiste à identifier un groupe d'acteurs-clé qui soit à même de fournir les informations nécessaires en vue de minimiser les perturbations des activités et d'optimiser les ressources opérationnelles et d'audit.

Objectifs de la mission

Conformément à la Norme 2210 - Objectifs de la mission, les auditeurs internes doivent établir des objectifs de la mission pour faire face aux risques associés à l'activité auditée. Une évaluation des risques devrait être effectuée pour permettre de définir les objectifs initiaux et d'identifier d'autres sujets de préoccupation.

L'objectif d'audit concernant le *Big Data* peut être défini de différentes façons. Par exemple, l'objectif peut être défini au niveau du plan d'audit annuel, en fonction des activités de management des risques à l'échelle de l'entreprise, des résultats des missions antérieures, des exigences réglementaires ou des besoins d'assurance spécifiques exprimés par le Conseil ou le comité d'audit.

Périmètre d'intervention de la mission et ressources affectées à la mission

Les procédures à suivre et le périmètre (nature, calendrier et étendue) de la mission devraient être déterminés une fois que les risques ont été identifiés. Conformément à la Norme 2220.A1, *« le périmètre d'intervention de la mission doit couvrir les systèmes, les données, les personnes et les biens concernés, y compris ceux qui se trouvent sous le contrôle de tiers. »*

La mission d'audit devrait concerner la stratégie et la gouvernance (y compris les politiques, les normes et les procédures), la sensibilisation des employés et la formation. L'audit interne doit déterminer les compétences nécessaires à la réalisation de la mission d'audit et le nombre total de ressources requises. L'équipe d'audit interne doit avoir le niveau approprié d'expertise, de connaissances et de compétences pour mener à bien la mission d'audit, ou à défaut, il sera fait appel à des ressources externes dotées des compétences requises.

Dans la mesure où il peut être difficile d'évaluer l'ensemble du programme de *Big Data,* le périmètre d'intervention peut être défini par unité opérationnelle, emplacement, objectif stratégique ou tout autre critère pertinent pour l'organisation.

Programme de travail de la mission

Conformément à la Norme 2240.A1 *« Les programmes de travail doivent faire référence aux procédures à appliquer pour identifier, analyser, évaluer et documenter les informations lors de la mission. »*

Vous trouverez ci-dessous un exemple de programme de travail relatif au *Big Data.* Les auditeurs internes peuvent l'utiliser comme point de départ pour créer un programme d'audit spécifique qui réponde aux besoins de leur organisation.

Objectif 1 : Planifier et déterminer le périmètre d'intervention de la mission		
Activités d'audit		**Observations**
1.1 Définir les objectifs de la mission (Norme 2210) Les objectifs d'audit sont ambitieux et décrivent les objectifs globaux de la mission.		
1.2 Identifier et évaluer les risques (Norme 2210.A1) L'évaluation des risques est nécessaire pour évaluer les priorités sur lesquelles les auditeurs internes doivent se concentrer.		
1.2.1	Identifier les risques opérationnels associés au *Big Data* qui intéressent les responsables opérationnels et les principales parties prenantes.	
1.2.2	Vérifier que les risques opérationnels et les risques SI sont cohérents.	
1.2.3	Évaluer le facteur de risque global en vue de la réalisation de la mission.	
1.2.4	Sur la base de l'évaluation des risques, identifier d'éventuels changements dans le périmètre.	
1.2.5	Discuter des risques avec le management et ajuster l'évaluation des risques.	
1.2.6	Sur la base de l'évaluation des risques, revoir le périmètre.	
1.3 Définir le périmètre d'intervention de la mission (Norme 2220) Le périmètre d'intervention de la mission doit être défini. L'auditeur interne devrait comprendre l'infrastructure de *Big Data,* les processus et les applications, ainsi que les risques associés pour l'organisation.		
1.3.1	Obtenir une liste de documents en rapport avec le *Big Data* qui peuvent aider à définir le périmètre. Par exemple : • La liste des endroits où le *Big Data* est utilisé • La liste des utilisateurs • La liste des rapports générés à l'aide du *Big Data* • La liste des processus opérationnels qui dépendent du *Big Data* pour des prises de décisions stratégiques	
1.3.2	Déterminer le périmètre d'intervention de la mission.	
1.4 Définir les critères de réussite de la mission Les facteurs de succès doivent être identifiés et acceptés.		
1.4.1	Identifier les leviers de succès de la mission.	
1.4.2	Communiquer les critères de réussite au propriétaire du processus ou à la partie prenante pour aboutir à un accord.	
1.5 Définir les ressources nécessaires à la réalisation d'une mission d'audit interne (Norme 2230) Dans la plupart des organisations, les ressources d'audit interne sont gérées en privilégiant les domaines prioritaires.		
1.5.1	Déterminer les compétences d'audit nécessaires pour la mission.	
1.5.2	Évaluer le volume total de ressource (en nombre d'heures) et le calendrier (dates de début et de fin) nécessaires pour la mission.	

Objectif 1 : Planifier et déterminer le périmètre d'intervention de la mission		
Activités d'audit		**Observations**
1.6 Définir les livrables. (Norme 2410) Les livrables ne se limitent pas au rapport final. La communication entre les équipes d'audit et le propriétaire du processus est essentielle à la réussite de la mission.		
1.6.1	Déterminer les livrables intermédiaires, notamment les premiers constats, les rapports d'étape, les projets de rapport, les dates d'échéance pour les réponses ou les réunions ainsi que le rapport final.	
1.7 Communiquer le processus. (Norme 2201) Le processus d'audit doit être clairement communiqué au client ou au commanditaire.		
1.7.1	Tenir une réunion de lancement afin de : • discuter du périmètre et des objectifs avec les parties prenantes ; • obtenir des documents et des ressources en matière de sécurité de l'information nécessaires à une réalisation efficace de la mission ; • communiquer sur le calendrier et les livrables.	

Objectif 2 : Identifier et obtenir des documents (Norme 2310)		
Activités d'audit		**Observations**
2.1	Examiner les politiques et les normes régissant le *Big Data*.	
2.2	Examiner la documentation de l'infrastructure informatique et identifier les systèmes qui prennent en charge le *Big Data*.	
2.3	Passer en revue les documents de conception du système.	
2.4	Examiner le diagramme des interfaces et identifier les systèmes qui partagent des données avec les systèmes de *Big Data*.	
2.5	Consulter la liste des utilisateurs internes et / ou externes.	
2.6	Examiner les contrats avec les fournisseurs de services.	
2.7	Examiner les contrats de niveau de service (SLA).	
2.8	Examiner les mesures de performance et les plans de remédiation.	
2.9	Examiner le plan de reprise après sinistre et les résultats des tests.	
2.10	Examiner le plan de continuité des activités et les résultats des tests.	

Risques relatifs à la gouvernance du programme

Pour réussir le déploiement du programme de *Big Data,* les organisations doivent assurer une gouvernance appropriée des personnes, des processus et de la technologie nécessaires. Sans une gouvernance adéquate des programmes, la mise en œuvre du *Big Data* peut exposer l'organisation à un risque injustifié, allant de l'échec et de l'adoption limitée aux problèmes de sécurité et de protection des données. Les organisations ont également des difficultés à concevoir des indicateurs permettant de mesurer le coût et la valeur des programmes de *Big Data*. La direction générale peut décider d'arrêter le financement d'un programme de *Big Data* si sa valeur ne peut pas être démontrée et communiquée de manière adéquate.

Trouver des personnes qualifiées et compétentes pour parrainer, gérer et mettre en œuvre un programme de *Big Data* dans un contexte technologique en constante évolution est un défi auquel toutes les organisations sont confrontées. Le *McKinsey Global Institute* prédit une importante pénurie d'analystes de données ainsi que des gestionnaires et des analystes dotés de la capacité d'exploiter le *Big Data* en vue d'assurer l'efficacité du processus de prise de décision.[9] Les écoles et les universités ont également du mal à ajuster les programmes aux besoins opérationnels qui ne cessent d'évoluer rapidement pour créer des filières avec des compétences pertinentes.

L'évolution technologique met la pression sur les organisations qui doive prendre la bonne décision quant au développement interne ou à l'acquisition de solutions et services de *Big Data*. Les organisations qui sous-traitent une partie ou la totalité de leurs services de *Big Data* font face à des risques supplémentaires de gestion de fournisseurs tiers, de sécurité du cloud et de protection des données personnelles.

Même si des personnes qualifiées et des technologies sont mobilisées, les entreprises doivent disposer de processus de gouvernance et de gestion de données suffisants pour s'assurer que les différentes dimensions de la qualité des données permettent la prise de décision. Les données de l'organisation sont souvent conservées en silos, ce qui augmente la complexité de l'identification et de l'inventaire des bases de données critiques, des éléments de données et de la lignée de données.

[9] http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Objectif 3 : Comprendre la gouvernance du programme de *Big Data*	
Objectif de contrôle	**Description**
3.1 Le budget devrait être suffisant pour répondre aux besoins de l'entreprise.	Les modèles de financement sont choisis pour soutenir la conception initiale et le déploiement, les activités permanentes (par exemple, les ressources de support de production durable et la maintenance de la technologie tout au long du cycle de vie) et les projets recommandés résultant de la mise en œuvre d'un programme de *Big Data*.
3.2 Les objectifs du programme doivent soutenir les initiatives stratégiques à l'échelle de l'organisation.	Les objectifs du programme et le *business case* répondent à la stratégie et aux initiatives de l'organisation afin de s'assurer que l'analyse coûts-avantages tient compte de la nécessité d'établir un programme de *Big Data*.
3.3 Les indicateurs témoignant de l'atteinte des objectifs devraient être portés à la connaissance du management.	Les indicateurs – à la fois quantitatifs et qualitatifs – sont conçus, mis en œuvre et surveillés en vue de démontrer la valeur du programme.
3.4 L'organisation devrait mettre en place une instance de gouvernance responsable de la gestion du *Big Data*.	Une structure transverse de gouvernance a pour but de prioriser les activités relatives au *Big Data* (par exemple, l'ordre des intégrations du système source, la sélection des analyses, l'élaboration du rapport) pour arbitrer les priorités concurrentes.
3.5 Des SLA permettant de définir et de mesurer les attentes en matière de performance devraient être conclus entre les métiers et les DSI.	Les SLA sont conçus et mis en œuvre pour s'assurer que les attentes des consommateurs sont gérées de manière proactive (par exemple, le calendrier de la disponibilité des rapports, la fréquence de rafraîchissement des données, les plages d'indisponibilité).
3.6 Les exigences opérationnelles et techniques doivent être documentées, analysées et approuvées.	Les exigences professionnelles et techniques sont recueillies et analysées pour soutenir la décision de développer ou d'acquérir (par exemple, par intervalles sur le cloud) un environnement de *Big Data* et de soutenir la sélection ultime d'un prestataire de solutions ou de technologies.
3.7 La direction générale devrait développer une stratégie de *Big Data* qui fournit des solutions applicables dans l'ensemble de l'organisation.	Le programme de *Big Data* englobe tous les principaux domaines d'activité concernés pour limiter les doubles emplois et les environnements technologiques redondants au sein de l'organisation.
3.8 Avant d'approuver le *business case,* la direction devrait démontrer la preuve de l'intérêt du concept afin de vérifier que la conception des systèmes est cohérente avec les objectifs stratégiques.	Un projet pilote (dont les opportunités sont connues et qui est dépourvu de tout complexité significative) a été sélectionné, et les domaines prioritaires sont identifiés sous l'angle des avantages qui plaident en faveur du déploiement du programme.

Objectif 3 : Comprendre la gouvernance du programme de *Big Data*	
Objectif de contrôle	**Description**
3.9 Les rôles et les responsabilités doivent être clairs et bien définis.	Un sponsor de niveau direction générale qui apporte un soutien et un appui solides au programme de *Big Data* a été identifié.
	Les rôles et les responsabilités des propriétaires de données, des gestionnaires et des experts du domaine ont été établis et définis.
	Une matrice d'attribution de responsabilités (RACI- *Responsible, Accountable, Consulted, Informed)* a été documentée et maintenue en place concernant les rôles et les responsabilités en matière de gouvernance de données à l'échelle de l'organisation.
	Les fonctions et les responsabilités des partenaires au niveau des métiers qui utilisent les solutions et le *reporting* de *Big Data* ont été établis et définis, y compris les contrôles qu'ils doivent mettre en œuvre pour arriver à consommer des données extraites de ces environnements.
3.10 L'organisation devrait fournir les ressources nécessaires pour déployer et maintenir la stratégie de *Big Data*.	L'organisation a identifié et budgété les postes critiques pour soutenir le programme de *Big Data* et a recruté des talents appropriés dotés des compétences et de l'expérience requises pour assurer la réussite du programme. L'évaluation des compétences est réalisée périodiquement en fonction des besoins changeants de l'organisation et de la technologie utilisée.
3.11 Les bonnes pratiques de gestion des prestations externalisées devraient être utilisées pour gérer les fournisseurs de *Big Data*.	Les contrats doivent prévoir des clauses adéquates en matière de sécurité, de disponibilité, de type de support, de prix, etc. Les réactions appropriées de la part du service juridique et de conformité sont intégrées dans l'accord avant la signature.
	Les contrats des fournisseurs tiers qui hébergent et/ou qui bénéficient de l'accès aux environnements de *Big Data* sont particulièrement suivis. Ces contrats reflètent de manière appropriée l'évolution dynamique de l'environnement pour soutenir une demande accrue ou réduite.
	Les rôles et les responsabilités des fournisseurs sont documentés et approuvés dans un contrat de services générique.
	Les SLA sont documentés, approuvés et font l'objet d'un suivi pour s'assurer que les tierces parties respectent les niveaux minimaux de performance. Des pénalités adéquates sont définies et appliquées en cas de non-respect des SLA. La direction a mis en place des routines de gouvernance des fournisseurs pour évaluer formellement les résultats du SLA et prendre des mesures appropriées.
	La transition et la résiliation ont été prises en compte dans le contrat, de même qu'une évaluation globale de la solution (par exemple, qu'est-ce qui arrive aux données à l'expiration du contrat ? Combien de temps faudra-t-il pour transmettre des analyses critiques à l'organisation ?).

Objectif 3 : Comprendre la gouvernance du programme de *Big Data*	
Objectif de contrôle	**Description**
	Veuillez vous reporter au GTAG « L'infogérance », 2ème édition pour bénéficier des conseils supplémentaires sur les considérations relatives au management des risques des fournisseurs tiers.
3.12 La gouvernance des données devrait faire partie de la gouvernance de l'organisation.	Les politiques de gestion des données et les procédures connexes (par exemple, la propriété, le partage, la confidentialité et les exigences de conservation, etc.) sont définis, documentés et partagés.
	L'organisation a identifié un responsable des données ou un rôle équivalent responsable de la gestion des données de l'organisation.
	Un inventaire des bases de données critiques, des tableaux, des éléments de données et de la lignée, ainsi que d'autres métadonnées est créé et mis à jour.
	Les normes en matière de données de l'organisation sont définies et communiquées à tous les employés concernés dans le cadre des exigences annuelles de formation.
	Les contrôles de qualité des données sont mis en place concernant les éléments de données critiques, ainsi qu'une gouvernance adéquate pour être en mesure de remédier aux problèmes de données identifiés.
	Les systèmes de référencement (SOR- *System of record)* et les sources de données autorisées (ADS- *Authoritative data sources)* ont été identifiés et les données sont convenablement rapprochées entre le SOR et l'entrepôt de données.
	L'équipe de gestion des données de l'organisation suit et gère la conformité de l'organisation avec les politiques de gouvernance et de gestion des données.

Disponibilité et performance de la technologie

Lorsque les plates-formes de *Big Data* ne sont pas évolutives et ne suivent pas les quantités croissantes de données structurées et non structurées nécessaires à l'analyse, les performances de l'analyse risquent de se dégrader et les résultats risquent d'être inexacts ou inopportuns. Le nombre de données extraites et chargées dans les systèmes de *Big Data* étant en constante augmentation, les hypothèses du modèle analytique devront peut-être être revues pour déterminer si le modèle et les hypothèses sous-jacentes sont toujours exactes.

La performance et la disponibilité des systèmes de *Big Data* deviennent de plus en plus importantes, car la dépendance de l'organisation à l'égard de ces systèmes est plus forte pour l'aide à la

décision et l'augmentation des revenus. Les systèmes de *Big Data* ne peuvent pas fournir de résultats analytiques lorsque les systèmes ou les flux de données connexes ne sont pas disponibles. Cela peut nuire à la rapidité des décisions de gestion, créer une expérience client négative et/ou entraîner des pertes de revenus. Les solutions de haute disponibilité et / ou de reprise après sinistre peuvent atténuer le risque d'inactivité du système, mais leur coût est plus élevé.

Les programmes de *Big Data* peuvent impliquer des exigences de stockage et de conservation de données sensiblement variables. Certaines données en continu, par exemple, peuvent ne jamais être sauvegardées ou peuvent être sauvegardées dans les systèmes sans la possibilité d'être récupérées si elles ne sont pas traitées et analysées immédiatement. D'autres applications de *Big Data,* cependant, peuvent exiger un historique important pour comprendre et identifier des modèles ou des comportements sur de longues périodes de temps.

Les lacs de données, qu'ils soient internes à l'organisation ou dans le cloud, présentent également des risques importants. Les lacs de données sont une consolidation de données détaillées provenant de nombreux systèmes organisationnels différents réunies à un emplacement centralisé, qui pourrait servir de point unique d'exposition de données. Étant donné que les lacs de données acceptent toutes les données dans leur format natif, des métadonnées bien définies peuvent faire défaut. Ces scénarios peuvent entraîner une confusion quant au véritable système d'enregistrement des données qu'il est permis d'utiliser dans les analyses.

Les politiques et les procédures de sauvegarde, de stockage et de conservation doivent être adaptées aux exigences d'analyse et à la valeur des données pour éviter toute perte de données. Les programmes tels que les tâches de sauvegarde et d'analyse de production devraient être configurés pour des notifications automatiques de réussite d'une tâche en temps opportun aux fonctions support de la production, ou pour des informations par lots ou en temps réel concernant les échecs. En l'absence d'une résolution rapide des défaillances, les données risquent d'être perdues et/ou les résultats de l'analyse risquent d'être inexacts, ce qui est particulièrement critique pour la transmission de données.

Des configurations technologiques inefficaces risquent d'entraîner une expérience client négative, réduire la disponibilité système et/ou dégrader les performances. Les outils ETL/ELT qui ne sont pas configurés de manière appropriée peuvent entraîner la perte d'importants volumes de données. La bande passante du réseau peut inhiber le transfert de nombreuses quantités de données lorsque les configurations ne sont pas optimisées. L'analyse des impacts en amont et en aval des mises à niveau et de la maintenance de la plate-forme peut être difficile en raison de l'évolution technologique rapide du matériel et des logiciels de *Big Data*.

Objectif 4 : Comprendre les risques en matière de disponibilité et de performance de la technologie	
Objectif de contrôle	**Description**
4.1 Les processus SI devraient être structurés de manière à prendre en charge les attentes en matière de production rapide de *Big Data*.	Les modèles de support de production sont définis et convenus avec les partenaires concernés dans les métiers pour assurer un support adéquat (par exemple, la nécessité d'un support 24 heures sur 24 et 7 jours sur 7, le délai de réponse aux problèmes ou questions des utilisateurs).
	Des solutions de haute disponibilité et / ou de reprise après sinistre sont mises en place pour les systèmes de *Big Data* en vue de répondre aux besoins de disponibilité du système et minimiser les temps d'arrêt en cas de panne.
	Des procédures sont en place aux fins de l'exécution, de la surveillance et de la récupération des sauvegardes de données pour les systèmes de *Big Data*.
	Un processus documenté de soutien à la production est en place afin d'effectuer le suivi des tâches en temps réel et par lot, alerter les collaborateurs concernés et suivre les incidents et les problèmes de la notification à la résolution.
4.2 Les politiques et les procédures relatives au cycle de vie des données devraient être documentées et faire l'objet d'un suivi.	Les exigences et les procédures de stockage et de conservation des données sont documentées et mises en place pour assurer le stockage, la conservation et la destruction appropriés des données.
4.3 Les systèmes de *Big Data* devraient être intégrés dans la stratégie de gestion des correctifs.	Le service informatique surveille et corrige les applications et les bases de données *Big Data* pour s'assurer qu'elles sont conservées et supportées par les fournisseurs.
4.4 Les systèmes de *Big Data* devraient être intégrés dans la stratégie de maintenance.	Un processus documenté de mise à niveau et de maintenance est en place pour s'assurer que les environnements de *Big Data* sont améliorés et maintenus conformément aux attentes de l'organisation et aux accords d'assistance des fournisseurs.
4.5 Les systèmes de *Big Data* devraient être acquis, développés et configurés en fonction de la complexité et des besoins documentées dans le *business case*.	Conformément à une évaluation documentée, l'organisation déploie un environnement ou une solution avec le matériel et les performances nécessaires pour répondre aux besoins de l'organisation en fonction de la taille et de la complexité des analyses prévues.
4.6 Les systèmes et les outils de support doivent être configurés pour fournir des notifications automatiques aux fonctions de support.	Les outils ETL / ELT sont configurés pour fournir des notifications automatisées d'achèvement et d'échecs des tâches. Le service informatique examine le nombre d'enregistrements et contrôle la vraisemblance des totaux des tâches achevées. Le service informatique résout les échecs des tâches avec une rapidité raisonnable et documente la raison de chaque échec et la solution apportée.

Objectif 4 : Comprendre les risques en matière de disponibilité et de performance de la technologie	
Objectif de contrôle	**Description**
4.7 Les systèmes de *Big Data* devraient être intégrés dans la stratégie de gestion du changement.	Lorsque les systèmes en amont et en aval sont modifiés, les modifications apportées aux flux de données impactés sont intégrées de manière appropriée dans le projet afin de limiter les ruptures dans le flux de données et/ou les problèmes de l'environnement de *Big Data*.
	Avant de mettre à niveau ou de changer les technologies de *Big Data,* le service informatique analyse l'impact sur les systèmes interfacés en amont et en aval pour assurer le fonctionnement continu des systèmes de *Big Data*.
	Les exigences de la couche réseau concernant les transmissions de *Big Data* sont documentées et analysées pour s'assurer que les analyses peuvent être diffusées avec succès entre les différents appareils.
4.8 Les outils de *reporting* doivent être configurés pour être flexibles, intuitifs et faciles à utiliser. Les supports de formation devraient être fournis.	Les outils et technologies de *reporting* sont correctement configurés et intégrés à l'environnement de *Big Data* pour répondre aux besoins des utilisateurs finaux. Une formation suffisante est fournie sur ces outils destinés aux utilisateurs finaux.
4.9 Les systèmes de *Big Data* devraient être configurés pour permettre une flexibilité et une évolutivité sans compromettre les performances.	Les systèmes de *Big Data* sont conçus et mis en œuvre pour permettre une évolutivité et s'assurer que les niveaux de transmission, de stockage, de disponibilité et de performance nécessaires sont maintenus au fur et à mesure de l'utilisation accrue du système.
4.10 Des tests périodiques de performance devraient être effectués en vue de remédier aux faiblesses.	Les tests de performance sont effectués sur des analyses nouvelles et modifiées pour s'assurer que les résultats sont produits dans des délais raisonnables tels que prévus.
4.11 Le cycle de vie des systèmes de *Big Data* devrait être correctement géré.	La direction réévalue en permanence la solution en en fonction des changements dans le secteur d'activité, la technologie et le contexte extérieur, afin de maintenir des services compétitifs et performants (par exemple, un changement d'architecture nouvellement recommandé peut améliorer considérablement les performances en fonction de connaissances sectorielles nouvelles).
4.12 Le grand modèle d'analyse de données devrait faire partie de la stratégie de maintenance.	La maintenance du modèle analytique est effectuée régulièrement pour assurer une précision et une fiabilité continues de l'analyse du *Big Data*.
4.13 Les contrôles généraux informatiques doivent être évalués périodiquement.	Veuillez vous reporter au GTAG « Les risques et les contrôles des systèmes d'information », 2ème édition pour des informations sur le contrôles des accès, les opérations informatiques, le cycle de vie du développement du système (SDLC- *system development life cycle)* et les considérations relatives à la gestion des changements. Ces contrôles sont essentiels à la réussite, à la durabilité et à la sécurité des programmes de *Big Data*.

Sécurité et protection des données personnelles

Les risques de sécurité et de protection des données personnelles sont facilement identifiables et compréhensibles par la plupart des personnes. Il est possible d'avoir des perspectives incroyables sur la vie d'une personne en combinant les informations personnelles avec les commentaires postés sur des réseaux sociaux et les *« Likes »*, les critiques publiées sur les produits et les services et l'historique de navigation Internet capturé par les cookies Web. Les organisations sont confrontées à des difficultés lorsqu'il s'agit de garantir que toutes les données collectées sont utilisées à des fins légitimes et que l'organisation respecte les lois et règlements. Les clients peuvent se sentir mal à l'aise face aux campagnes de marketing personnalisées, basées sur l'analyse des données personnelles, même lorsque les données sont utilisées pour des raisons valables d'un point de vue opérationnel.

En outre, les violations d'accès aux données du secteur public et privé aboutissant aux vols de celles-ci sont devenues trop fréquentes et les coûts associés à l'incapacité d'une organisation à protéger les informations à caractère personnel de ses employés, clients et fournisseurs augmentent constamment. Les sanctions réglementaires et les amendes, qui varient en fonction de l'emplacement et de la compétence, peuvent avoir pour effet d'engager la responsabilité légale et financière de l'organisation de manière considérable. En outre, une violation de données constitue un important préjudice pour l'image de la marque et la réputation de l'organisation, qui se solde par une baisse du chiffre d'affaires et une augmentation des coûts.

Les menaces et les vulnérabilités associées à un accès interne inapproprié (par exemple, par des collaborateurs, des consultants et des fournisseurs de *Big Data)* sont souvent aussi importantes que les menaces externes, compte tenu des connaissances et de la position privilégiée inhérentes des acteurs concernés. Ces actions d'un agent interne concernent notamment le vol de données confidentielles et sensibles, l'obtention de secrets d'affaires ou des activités malveillantes utilisant les connaissances de l'agent en question. Le vol des connaissances et des informations acquises grâce aux systèmes de *Big Data* à des fins personnelles passe souvent inaperçu parce que les entreprises concentrent les efforts de cybersécurité sur les menaces externes et ne mettent pas forcément en place des contrôles inadéquats lorsqu'il s'agit de prévenir et de détecter l'activité interne. Les privilèges liés au compte devraient être strictement limités à l'accès nécessaire pour les responsabilités professionnelles de l'utilisateur, et des contrôles supplémentaires devraient être mis en place pour surveiller et détecter les activités suspectes.

S'assurer que tous les systèmes sont sécurisés de manière appropriée et cohérente devient plus difficile car les systèmes de *Big Data* deviennent plus complexes et puissants et abritent des volumes plus importants de données disparates. Un correctif ou des configurations de sécurité inadéquats peuvent ouvrir la voie aux vulnérabilités qui peuvent à leur tour être exploitées pour visualiser ou modifier des données sensibles. Des perturbations du système peuvent également se produire, ce qui entraîne l'indisponibilité des services et une perte de productivité.

Veuillez vous reporter au Guide pratique de l'IIA intitulé « L'audit des risques d'atteinte à la vie privée », 2ème édition pour des informations supplémentaires concernant les risques et les défis liés à protection des données personnelles. En effet, les éléments proposés dans ce guide restent tout à fait pertinents pour des programmes et des environnements de *Big Data*. De plus, reportez-vous au GTAG « Evaluer le risque de cybersécurité - Le rôle des trois lignes de maîtrise défense » pour en savoir davantage sur les risques et les considérations liés à la sécurité.

Objectif 5 : Comprendre la sécurité et la protection des données personnelles du *Big Data*	
Objectif de contrôle	**Description**
5.1 La gestion de la sécurité de l'information devrait être intégrée dans la stratégie de *Big Data*.	Un programme de cybersécurité existe au sein de l'organisation pour lutter contre les menaces internes et externes.
	Une configuration générique de sécurité renforcée est mise en place pour assurer un environnement d'exploitation cohérent et sécurisé adapté aux systèmes de *Big Data* et à leur infrastructure.
	Les utilitaires de système sont capables de contourner le système d'exploitation, le réseau et les contrôles d'application sont interdits ou contrôlés de manière appropriée.
	L'accès et l'utilisation des outils d'audit sont segmentés et limités pour éviter tout compromis, mauvaise utilisation et / ou destruction des données du journal. Les données du journal sont examinées pour identifier les activités suspectes.
	Tous les services fondés sur le *Cloud* utilisés par l'organisation sont approuvés pour l'utilisation et le stockage des données de l'organisation.
	La DSI évalue la sécurité des prestataires concernés pour répondre aux préoccupations concernant l'infrastructure partagée, les systèmes hébergés à l'extérieur et l'accès des fournisseurs aux données avant la mise en place d'une solution informatique fondée sur le *cloud* ou autre.
	Les processus de gestion des patchs sont documentés et mis en œuvre pour s'assurer que les systèmes sont corrigés en temps opportun selon les correctifs *(patch)* approuvés les plus récents (cf. GTAG « Contrôles de la gestion du changement et des patchs : un facteur clé de la réussite pour toute organisation », 2ème édition).
	Veuillez vous reporter au GTAG « Evaluer le risque de cybersécurité - Le rôle des trois lignes de maîtrise » pour de plus amples informations concernant les risques de cybersécurité et les contrôles connexes.

Objectif 5 : Comprendre la sécurité et la protection des données personnelles du *Big Data*	
Objectif de contrôle	**Description**
5.2 La gestion de la sécurité de données devrait être intégrée dans la stratégie de *Big Data*.	Seuls les utilisateurs autorisés de l'organisation ont accès aux données et aux rapports générés par des systèmes de *Big Data*. L'accès est conditionné par les responsabilités inhérentes au poste et basé sur le concept du moindre privilège.
	Seul un petit groupe d'utilisateurs techniques autorisés bénéficie d'un accès privilégié aux systèmes de *Big Data*, y compris les systèmes d'exploitation, les bases de données et les applications.
	Les outils de *reporting* des utilisateurs finaux sont configurés de manière appropriée pour s'assurer que les personnes autorisées sont en mesure de consulter les données sensibles.
	Les droits d'accès aux systèmes de *Big Data* sont revus périodiquement pour s'assurer de leur pertinence.
5.3 L'accès de tiers doit être géré de manière adéquate.	Les exigences en matière de sécurité, contractuelles et réglementaires concernant les fournisseurs sont traitées avant d'accorder l'accès aux systèmes de données et d'information. La direction évalue le respect de ces dispositions dans le cadre des systèmes de gouvernance des fournisseurs.
	Veuillez vous reporter au GTAG « Gestion des identités et des accès » pour de plus amples informations concernant les contrôles des accès, le provisionnement, l'administration de la sécurité et la mise en place.
5.4 La confidentialité des données devrait être intégrée dans la stratégie de *Big Data*.	Les données sont répertoriées et classées pour s'assurer que les données critiques de l'organisation, y compris les informations à caractère personnel nécessitant une protection, sont protégées de manière appropriée.
	Les informations à caractère personnel identifiables et d'autres données sensibles sont assainies ou brouillées avant la réplication entre la production et le développement ou les environnements de test.
	Un processus de réponse aux incidents a été documenté, approuvé et mis en œuvre pour s'assurer que les violations des données sont traitées de manière appropriée.
	Veuillez vous reporter au Guide pratique de l'IIA intitulé « L'audit des risques d'atteinte à la vie privée, 2ème édition » pour obtenir des renseignements sur les cadres de confidentialité et les principes d'audit relatifs à la protection des données personnelles.

Qualité, gestion et reporting des données

L'expression « *Garbage in, garbage out* » est courante dans le domaine informatique. Elle signifie que les données de mauvaise qualité entrées dans le système entrainent des résultats d'aussi mauvaise qualité. Les problèmes de qualité des données peuvent entraîner l'inexactitude des rapports de gestion et fausser le processus de prise de décision. Les bases de données qui ne sont pas conçues pour assurer l'intégrité des données peuvent comporter des données incomplètes ou invalides. Les analyses qui en découlent sont susceptibles d'aboutir à des résultats erronés. Par conséquent, les analyses de *Big Data* doivent tenir compte de ces risques liés à la qualité des données.

En outre, les données qui ne sont pas obtenues et analysées en temps opportun peuvent entraîner des résultats d'analyse incorrects, des décisions de gestion erronées ou une perte de revenus. Les données émanant des tiers devraient être obtenues en temps utile, exactes, complètes et extraites d'une source de bonne réputation. Les données émanant des tiers sous un format inapproprié présentent le risque d'être inadaptées à l'analyse et de retarder la prise de décision.

Une fois les données réceptionnées et analysées, il peut être difficile de s'assurer que les utilisateurs finaux assurent la gestion et la protection des données. L'absence de contrôles informatiques au niveau des utilisateurs finaux peut entraîner l'inexactitude des rapports et des fuites de données. Les rapports de production des utilisateurs finaux, les rapports ad hoc et les résultats d'analyses prédictives doivent tous être examinés et approuvés pour limiter les décisions de gestion erronées. Les rapports de *Big Data* devraient également respecter les politiques de classification des données d'une organisation afin de s'assurer que seules les données appropriées sont partagées, tant à l'interne qu'à l'externe. Les options de rapport et les canaux de diffusion peuvent être appropriés uniquement pour les données de tailles et de formats spécifiques. Les organisations peuvent faire face à des obstacles lors de la détermination des options et des canaux de rapport appropriés pour chaque résultat de l'analyse.

Objectif 6 : Comprendre la qualité, la gestion et le reporting du *Big Data*	
Objectif de contrôle	**Description**
6.1 Il convient d'établir des politiques et des procédures visant à garantir la qualité des données.	Les procédures sont établies et documentées pour s'assurer que la qualité des données est conforme aux normes de l'organisation en matière de données.
	Des processus doivent être mis en place pour valider formellement les systèmes d'enregistrement des éléments de données critiques afin de garantir une solution cohérente aux différences de données entre les systèmes.
	Les bases de données du *Big Data* sont conçues et mises en œuvre pour assurer l'intégrité des données, y compris un rapprochement et un suivi appropriés des principaux éléments de données jusqu'aux systèmes sources ou d'enregistrement.
	Des processus sont en place pour assurer la réception en temps opportun des données dans les systèmes de *Big Data* et la fourniture en temps voulu des analyses aux consommateurs de *Big Data*.
	La direction doit mettre en œuvre des mesures visant à préserver et à protéger l'intégrité des données des applications critiques de l'utilisateur final et des calculs qui en découlent. Les exemples incluent des contrôles d'accès, l'identification de la formule, les totaux de contrôle, etc.
	Les exigences pour la préparation des entrées de données pour les applications de *Big Data* et les analyses sont définies et documentées.
	Les données sont disponibles en temps opportun, ce qui permet une prise de décision rapide en fonction des exigences et des besoins de l'organisation.
6.2 Des politiques et des procédures devraient être établies pour s'assurer que les données obtenues auprès de tiers respectent les normes de qualité des données.	Les procédures permettant de s'assurer que les données achetées auprès de tiers sont exactes, complètes et proviennent d'une source de bonne réputation sont établies et documentées.
6.3 Il convient d'établir des politiques et des procédures visant à garantir l'exactitude des rapports.	Les procédures d'examen et d'approbation des résultats des rapports analytiques sont établies et documentées. Les analyses prédictives qui motivent des décisions bénéficient d'une attention toute particulière.
6.4 L'accès aux rapports devrait être accordé en fonction des besoins de l'organisation.	Les éléments de rapport appropriés pour la consommation interne par rapport à la consommation externe ont été définis et documentés. Les propriétaires de données ont approuvé la consommation de ces rapports et données par des personnes autorisées.

Objectif 6 : Comprendre la qualité, la gestion et le reporting du *Big Data*	
Objectif de contrôle	**Description**
6.5 Les outils et procédures de *reporting* devraient être flexibles et permettre l'élaboration de rapports *ad hoc*.	Des procédures doivent être mises en place dans le cas où un *reporting ad hoc* est autorisé au lieu de rapports génériques qui ne peuvent être modifiés sans une gestion des changements et des développements au niveau des systèmes.
6.6 Les utilisateurs devraient être périodiquement formés pour optimiser l'utilité du rapport.	Au fur et à mesure des modifications des données sous-jacentes et des méthodes de *reporting*, les utilisateurs finaux reçoivent une formation sur les changements qui en résultent et sur les impacts en aval de leurs rapports, systèmes d'utilisateurs finaux, etc.
6.7 La sélection des fournisseurs de rapports devrait être conforme aux besoins de l'organisation.	Un processus est en place pour sélectionner les options / canaux de notification appropriés et privilégiés (par exemple, *Tableau®, Splunk®, Oracle® Business Intelligence Enterprise [OBIEE])* pour les résultats analytiques.

Contributeurs

Brian Allen, CISA, CISSP

Stephen Coates, CIA, CISA, CGAP

Brian Karp, CIA, CISA, CRISC

Hans-Peter Lerchner, CIA, CISA

Jacques Lourens, CIA, CISA, CGEIT, CRISC

Tim Penrose, CIA, CISA, CIPP

Sajay Rai, CISM, CISSP

L'IIA remercie l'*Information Technology Guidance Committee (ITGC)* pour sa contribution à la rédaction de cette ligne directrice.

Réviseurs de la traduction :

Marie-Elisabeth Albert, CIA

Béatrice Ki-Zerbo, CIA

Sébastien Lepers, CIA, CGAP, CRMA, CFE

À propos de l'Institute of Internal Auditors

Porte-parole mondial de la profession d'audit interne, l'Institute of Internal Auditors (IIA) est une autorité reconnue et un leader incontesté dans la formation et la formulation de normes, lignes directrices et certifications. Fondé en 1941, l'IIA compte actuellement quelque 185 000 membres dans plus de 170 pays et territoires. Son siège se situe à Lake Mary (Floride) aux États-unis. Plus d'informations sont disponibles sur le site www.globaliia.org ou www.theiia.org.

À propos des lignes directrices complémentaires

Les lignes directrices complémentaires font partie intégrante du Cadre de référence international des pratiques professionnelles (CRIPP) et proposent des recommandations supplémentaires (facultatives) pour la réalisation des activités d'audit interne. Si elles reflètent les *Normes*, ces lignes directrices complémentaires ne visent pas directement la mise en conformité aux dites *Normes*. Elles couvrent au contraire des domaines particuliers, ainsi que des questions sectorielles, et contiennent des procédures et processus détaillés. Ces lignes directrices ont été officiellement révisées et approuvées par l'IIA.

Avertissement

L'IIA publie ce document à titre informatif et pédagogique. Cette ligne directrice n'a pas vocation à apporter de réponses définitives à des cas précis, et doit uniquement servir de guide. À ce titre, il doit être utilisé uniquement à des fins d'orientation. L'IIA vous recommande de toujours solliciter un expert indépendant pour avoir un avis dans chaque situation. L'Institute dégage sa responsabilité pour les cas où des lecteurs se fieraient exclusivement à ce guide.

Copyright

Le copyright de ce guide pratique est détenu par l'Institute of Internal Auditors et par l'IFACI pour sa version française.
Pour l'autorisation de reproduction, veuillez contacter l'Institute of Internal Auditors à l'adresse guidance@theiia.org ou l'IFACI à l'adresse recherche@ifaci.com.

Juin 2017 - ISBN 978-2-915042-81-8