Sécurité

Conseils pour sécuriser le Big Data

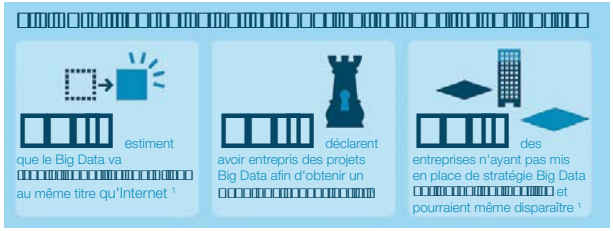
Pourquoi le Big Data n'est pas forcément un casse-tête

IBM

Le Big Data ou comment l'analytique en temps réel peut aider les entreprises à améliorer leurs ventes

Véritables moteurs d'innovation, les technologies de l'information ont totalement révolutionné la manière dont les entreprises, les gouvernements et les particuliers travaillent et interagissent. Les volumes massifs de données, baptisées Big Data, ont ouvert la porte à une collaboration universelle en temps réel et ont changé la donne de manière incontestable. Le Big Data a alimenté de nombreux débats et discussions dans la presse, sur les blogues, mais également entre les analystes et les grandes entreprises de technologie. Il n'en reste pas moins qu'il demeure difficile de discerner les avantages, les possibilités, les usages pratiques et les défis que présentent réellement les technologies de Big Data.

Aujourd'hui, les entreprises se servent des environnements Big Data pour traiter, analyser et extraire la valeur ajoutée des formats de données modernes, mais également des formats structurés traditionnels, en temps réel ou à des fins ultérieures afin de les aider à prendre des décisions à la fois plus informées et rentables. Le célèbre institut de recherche américain Forrester Research définit le Big Data comme ceci : « *Le Big Data se compose de trois dimensions : volume, vélocité et variété. Il représente un tournant majeur dans les domaines de la business intelligence et de la gestion des données, bouleversant les principes de base de la gouvernance et des services informatiques. Alors que les solutions traditionnelles sont devenues trop onéreuses à étendre ou adapter pour faire face à l'évolution rapide des technologies, les entreprises redoublent d'efforts pour trouver des technologies abordables pour stocker, traiter et analyser l'intégralité de leurs données. Seules des solutions novatrices peuvent aider les entreprises à extraire la valeur ajoutée des données de leur Big Data, et à créer des expériences client individuelles et plus personnalisées.* »



Du fait de l'expansion des volumes de données gérés dans les environnements Big Data, les entreprises assistent à l'émergence de risques et menaces considérables qui compromettent les référentiels des données stockées. Incapables de trouver le juste équilibre entre la sécurité et la qualité des données, les décisions deviennent de moins en moins certaines. En fait, des études montrent que les décideurs qui ne font pas confiance aux résultats analytiques trouveront toujours des raisons pour les rejeter à moins qu'ils aient développé une très grande confiance dans les données et qu'ils soient persuadés de leur fiabilité.

Nous sommes face à un paradoxe. En effet, les entreprises n'ont jamais généré autant de données qu'aujourd'hui. Pourtant, elles n'ont toujours pas compris leur pertinence, leur contexte ni comment les protéger.

Le Big Data se compose de trois dimensions : volume, vélocité et variété.

Volume

Chaque jour, nous produisons à travers le monde 2,5 quintillions d'octets de données provenant de sources traditionnelles et modernes (capteurs climatiques, messages échangés sur les réseaux sociaux, photos et vidéos numériques, enregistrements de transactions sur des sites marchands, signaux GPS de téléphones cellulaires, etc.).

Vélocité

Parfois, deux minutes c'est déjà trop tard ! Pour les processus chronosensibles tels que la détection de fraudes, une réponse en temps réel est nécessaire.

Variété

Le Big Data gère des données de tous types — structurées et non structurées — telles que le texte, les données mesurées par les capteurs, les données audio et vidéo, les parcours de navigation, les fichiers journaux, etc.

Quelques cas d'utilisation des projets de Big Data :

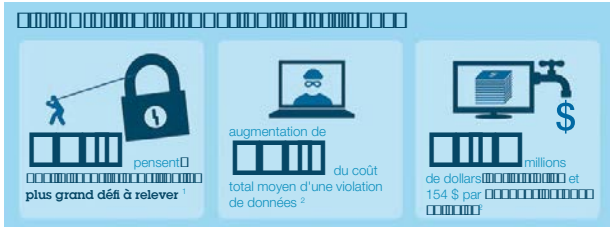
- Transformer 12 téraoctets de tweets en analyse poussée des opinions sur un produit
- Scruter 5 millions d'événements commerciaux par jour afin d'identifier les fraudes potentielles
- Utiliser les centaines de flux vidéo des caméras de surveillance pour identifier les menaces de sécurité

de défis uniques

Certes, l'analyse du Big Data offre d'importantes opportunités. Les entreprises ne doivent néanmoins pas perdre de vue les défis de sécurité sous-jacents, par exemple :

- L'analyse du Big Data offre aux entreprises un formidable levier de croissance en leur permettent d'accéder aux informations clés dissimulées dans leurs données. Or, ces informations sont généralement issues de données sensibles qui ont été agrégées dans le référentiel de données.
- La nature du Big Data (volume, vélocité et variété importants) permet difficilement de garantir l'intégrité des données.
- Les environnements répartis sans schéma prédéfini, dans lesquels peuvent être regroupées et agrégées les données de sources hétérogènes, compliquent la mise en place de contrôles d'accès.
- Les référentiels Big Data constituent une autre source de données à sécuriser. Or, la majorité des approches actuelles en matière de sécurité et conformité des données ne sont pas adaptées.

Les environnements Big Data permettent aux entreprises d'agréger de plus en plus de données — essentiellement des données financières ou privées, des données de propriété intellectuelle, des données propriétaires, ainsi que toute autre information sensible. La plupart des données sont soumises à des réglementations de conformité, telles que la loi Sarbanes-Oxley (SOX), la



loi HIPAA (Health Insurance Portability and Accountability Act), la loi PCI DSS (Payment Card Industry Data Security Standard), la loi FISMA (Federal Information Security Management Act) ou encore la Directive européenne sur la protection des données. Les données sensibles constituent la cible privilégiée des pirates informatiques.

C'est pourquoi les professionnels de la sécurité des données doivent engager des plans d'action incisifs sans plus attendre. Cela étant, les entreprises sont soumises à une telle pression pour prendre des décisions rapides que les experts en sécurité de données risquent d'être tenus à l'écart des décisions majeures ou d'être considérés comme des freins à la croissance. Fort heureusement, le risque de négligence de la sécurité des données est bien connu et documenté, et la sécurité n'est plus forcément un obstacle.

En cas de mauvais usage des données, les entreprises contrevenantes et leurs responsables s'exposent à de lourdes amendes allant de 5 000 à 1 000 000 de dollars par jour, voire des peines de prison. Selon l'étude 2015 du Ponemon Institute consacrée à l'analyse du coût de la violation des données (parue en juin 2015), on apprend que le coût moyen d'une violation de données s'élèverait à 3,8 millions de dollars. L'étude révèle que le coût moyen par personne induit pour chaque violation de données est passé de 145 à 154 dollars.

Mais les lourdes peines requises ne constituent pas le seul moyen de dissuader les entreprises de détourner vos données, car la violation de données peut également occasionner d'autres retombées particulièrement négatives (chute du cours de l'action, mauvaise publicité) et porter un préjudice irréversible à la marque.

Pour l'heure, les entreprises sont de plus en plus nombreuses à se doter d'environnements Big Data. Le temps est venu d'inviter les équipes sécurité au cœur de ces décisions et déploiements, d'autant plus que les environnements Big Data — dépourvus de fonctions complètes de sécurité des données — et les masses de données sensibles qui y sont stockées constituent une cible facile pour les pirates informatiques.

La sécurité des données est un processus continuel et détaillé qui doit devenir systématique pour les environnements Big Data. La sécurisation des données implique une approche globale pour protéger les entreprises de l'ensemble de menaces complexes provenant de systèmes hétérogènes. La sécurité des données doit venir en complément d'autres mesures de sécurité, telles que la sécurité au niveau des nœuds finaux, du réseau, des applications et des sites physiques, afin de permettre une protection maximale.

« La sécurité des données doit évoluer afin de traiter les données comme un actif qui est en train de se répandre au-delà des limites des silos de données traditionnels sur site, sur les plateformes de Big Data et dans les clouds publics. Le piratage, la résidence des données et les obligations de conformité sont autant de questions qui menacent la mise en œuvre des stratégies en matière de sécurité des données. »

— Gartner : Hype Cycle for Data Security, 2015, juillet 2015

[illegible]

La sécurité des données peut être gérée de manière efficiente et efficace pour satisfaire toutes les parties intéressées. Ne tardez pas à planifier la sécurité de votre Big Data : la sécurisation des environnements Big Data est la clé pour diminuer les risques, les coûts et les efforts de déploiement.

Mais quelles données faut-il protéger ? De nombreuses entreprises se tournent vers les plateformes Hadoop pour réaliser des analyses massives de manière fiable et rapide. En déployant Hadoop avec leurs systèmes de bases de données actuels, elles peuvent ainsi combiner des données structurées classiques et des données non structurées modernes de manière optimale. D'autres entreprises préfèrent les bases de données NoSQL qui fournissent le même niveau d'évolutivité exceptionnelle tout en répondant aux besoins de débit élevé et de faible latence.

Quelle que soit la base de données choisie, les environnements Big Data nécessitent une protection capitalisant les mêmes stratégies de sécurité rigoureuses que celles actuellement appliquées aux environnements classiques pour garantir le respect des exigences de conformité et l'immunité des données stockées.

Quelques exemples de stratégies de sécurité à mettre en œuvre
pour les environnements Big Data :

- Détection et classification des données sensibles : à effectuer avant de migrer les données vers la base de données Hadoop ou NoSQL afin de mettre en place les mécanismes de sécurité adéquats en aval.
- Accès aux données et contrôle des changements : établir des règles afin que seuls les utilisateurs et applications autorisés puissent accéder aux données et y apporter des modifications.
- Audit et surveillance en temps réel de l'activité des données : déterminer les conditions (qui, quoi, où, quand, comment) d'accès aux données et veiller à la conformité.
- Protection des données : recourir à des techniques de masquage, de chiffrement ou d'occultation selon le cas.
- Prévention contre la perte de données : créer un journal d'audit détaillé des activités d'accès aux données et de leur utilisation afin de s'assurer qu'aucune donnée ne soit perdue.
- Gestion et évaluation des vulnérabilités : comprendre les points faibles et mettre en place des plans d'action adéquats pour les éliminer.
- Gestion de la conformité : intégrer un processus de production de rapports de conformité à la base de données Hadoop ou NoSQL afin de gérer la production, la distribution et la validation de rapports.

En outre, les entreprises doivent être capables de répondre aux questions telles que :

1. Qui exécute des requêtes particulières sur votre Big Data ?
2. Les utilisateurs sont-ils autorisés à exécuter des requêtes ?
3. Quelles opérations Map/Reduce les utilisateurs emploient-ils ?
4. Les utilisateurs tentent-ils de télécharger des données sensibles ou la requête est-elle associée à une requête plus vaste, comme une analyse marketing ?

Les environnements Big Data doivent également se soumettre à un cadre réglementaire. La ruée vers le nouvel « or noir » ne dispense pas les entreprises de respecter le cadre et leurs obligations réglementaires. Les entreprises doivent être prêtes à respecter les exigences d'audit (voir Figure 3).

Cadre réglementaire					
Exigences d'audit	**COBIT (SOX)**	**PCI-DSS**	**ISO 27002**	**Lois sur la protection et la confidentialité des données**	**NIST SP 800-53 (FISMA)**
1. Accès aux données sensibles (réussite/échec d'opérations SELECT)		✓	✓	✓	✓
2. Changements de schémas (DDL) (Create/Drop/Alter Tables, etc.)	✓	✓	✓	✓	✓
3. Changements de données (DML) (Insert, Update, Delete)	✓		✓		
4. Exceptions de sécurité (échecs de connexion, erreurs SQL, etc.)	✓	✓	✓	✓	✓
5. Comptes, rôles et autorisations (DCL) (Grant, Revoke)	✓	✓	✓	✓	✓

DDL - Data Definition Language (pour les changements de schémas)
DML - Data Manipulation Language (pour les changements de valeurs des données)
DCL - Data Control Language

[illegible]

Planifier à l'avance

Beaucoup d'entreprises commencent à peine à explorer la mise en œuvre d'un environnement Big Data et n'ont aucune idée des types de données (structurées ou non) qu'elles souhaitent inclure dans leur référentiel.

L'étape de planification est le moment idéal pour engager une discussion entre les différents experts (sécurité des données, juridique, métier et informatiques) et s'intéresser à la compréhension, la détection et la classification des données sensibles. Une équipe interfonctionnelle se chargera de localiser les données, de déterminer les définitions communes à appliquer aux données sensibles et enfin d'identifier les types de données qui seront migrées dans l'environnement Hadoop ou NoSQL. Par ailleurs, il faudra établir une approche de cycle de vie afin d'extraire en continu les données à l'échelle de l'entreprise — tout en identifiant et en gérant les problèmes tels que les droits « dormants » et les données « dormantes ».

Penser de manière stratégique

Les environnements Big Data constituent un outil particulièrement précieux pour les entreprises. Les experts en sécurité des données pourront également en tirer profit, car les référentiels de données massives comme Hadoop permettent de stocker des informations de sécurité vitales.

En effet, ils peuvent se servir des environnements Big Data pour prioriser leurs analyses de sécurité et mettre en place les mécanismes de sécurité adéquats de manière nettement plus efficace.

En apprenant à connaître davantage les particularités des attaquants potentiels (qui a accédé aux données dans Hadoop, comment les données ont-elles été consultées et à quel moment), les experts peuvent alors réajuster leurs stratégies de sécurité en envoyant par exemple une alerte en temps réel qui demandera à l'équipe en charge de la sécurité des informations d'intervenir.

La protection des environnements Big Data nécessite des fonctionnalités critiques, notamment :

- **Fonctions d'audit sans diminution notable des performances**
- **Alertes en temps réel en cas de comportement suspect afin d'intervenir de manière proactive**
- **Séparation des responsabilités afin que l'expert sécurité et l'administrateur Hadoop ne soient pas la même personne, par exemple.**
- **Analyse avancée de l'utilisation des ressources et du comportement des utilisateurs**
- **Évolutivité de la plateforme, performance et capacité à intégrer des environnements classiques et de Big Data**

Sécuriser et protéger
Le regroupement des données augmente le risque que des informations sensibles puissent être interceptées par un cybercriminel ou un initié (malveillant ou non). C'est pourquoi les entreprises doivent limiter le nombre de personnes pouvant accéder aux référentiels de données massives.

Les environnements Big Data doivent être dotés d'une protection des données sensibles. Par exemple, le contrôle d'accès permet de s'assurer que seuls les utilisateurs autorisés accèdent à certaines données et la surveillance continue de l'accès des utilisateurs et des applications présente un intérêt capital, en particulier lorsque les utilisateurs modifient des rôles. La surveillance permet de s'assurer que les stratégies de sécurité sont appliquées et efficaces, d'analyser les modèles d'accès et d'utilisation des données, mais également d'alerter les équipes sécurité d'un éventuel abus (par exemple, lorsque le système détecte que des fichiers journaux sont modifiés par un ID administrateur). Les cybercriminels ou attaquants internes laissent généralement des traces de leur passage ou des artefacts qui peuvent être détectés par une analyse approfondie. En vous dotant de fonctions de blocage, de mise en quarantaine et d'alerte, vous pouvez lutter activement contre la perte de données.

Pour sécuriser leurs données, les entreprises peuvent également employer des techniques d'abstraction des données, telles que le chiffrement ou le masquage. Une fois chiffrées ou masquées, les cybercriminels ne peuvent généralement pas décoder ni récupérer les données.

[illegible]

IBM® Security Guardium® sécurise les environnements Big Data en :

- Surveillant étroitement l'activité des bases de données Hadoop et NoSQL par les applications et les utilisateurs en temps réel ; déclenchant des alertes en cas de violation de règles ; en faisant le suivi des tentatives d'accès et d'utilisation des données afin de déceler tout comportement inhabituel parmi les utilisateurs privilégiés et externes ; et en notifiant les tableaux de bord SIEM pour engager les mesures correctives adéquates (alerte, blocage, résiliation de connexion).
- Soutenant les activités d'audit et de production de rapports afin satisfaire aux exigences de conformité et contribuer aux enquêtes : les données d'activité peuvent être regroupées au sein de pistes d'audit non répudiables et dans des rapports au format adéquat ; la séparation des responsabilités permet d'observer les observateurs également ; des règles et des rapports préconfigurés prêts à l'emploi sont également disponibles.
- Mettant en place un contrôle des changements pour contrôler le volume, la vélocité et la variété du Big Data.
- Implémentant des contrôles automatisés et centralisés au sein à l'échelle de l'entreprise (bases de données, applications, fichiers, Big Data, etc.).
- Protégeant les données sensibles au moyen de techniques de chiffrement, de masquage et d'occultation.
- Évaluant et résolvant les faiblesses de l'environnement de manière à sécuriser l'ensemble du Big Data.

Conclusion : renforcer la sécurité à [illegible]

Les entreprises ne doivent pas s'affoler lorsqu'on leur parle de sécurisation des environnements Big Data. En effet, la sécurisation des mises en œuvre Hadoop et NoSQL est gouvernée par les mêmes principes que ceux qui s'appliquent aux bases de données, aux entrepôts de données et aux systèmes de partage de fichiers. La plateforme de sécurité de données Guardium est évolutive, ce qui vous permet de protéger à la fois vos architectures classiques de gestion de données et vos environnements Big Data, mais également de protéger votre entreprise de l'ensemble de menaces complexes (notamment les menaces internes, les modifications non autorisées et les attaques externes) tout en restant concentré sur vos objectifs et en automatisant les contrôles de conformité.

La solution Guardium prévient la fuite des bases de données, des entrepôts de données et des environnements Big Data, tout en préservant l'intégrité des informations et en automatisant les contrôles de conformité sur des environnements hétérogènes. Elle propose une plateforme évolutive qui permet d'effectuer une surveillance continue du trafic des données structurées et non structurées, tout en veillant à la mise en application des règles concernant l'accès aux données sensibles à l'échelle de l'entreprise. L'objectif ? Améliorer la sécurité des données et les décisions en matière de sécurité en se fondant sur des informations exploitables et priorisées, issues du contrôle et de la surveillance de l'ensemble de l'environnement.

Pour en savoir plus sur la sécurité du Big Data, contactez votre revendeur ou votre partenaire commercial IBM, ou rendez-vous sur le site Web : **ibm.com**/guardium

De plus, IBM Global Financing peut vous aider à financer l'acquisition des solutions informatiques nécessaires à votre entreprise, de la manière la plus rentable et stratégique possible. Pour les clients éligibles au crédit, nous pouvons proposer une solution de financement personnalisée pour répondre à vos exigences métier et économiques tout en permettant une gestion efficace de votre trésorerie et en améliorant votre coût total
de possession. IBM Global Financing est le choix le plus judicieux pour financer vos investissements informatiques stratégiques et faire progresser votre activité. Pour plus d'informations, consultez le site Web : **ibm.com**/financing

Déclaration de bonnes pratiques de sécurité : La sécurité des systèmes informatiques implique la protection des systèmes et des informations via la prévention, la détection et la réponse en cas d'accès incorrect au sein et à l'extérieur de votre entreprise. Un accès non autorisé peut se traduire par la modification, la destruction ou une utilisation inadéquate ou malveillante de vos systèmes, y compris l'utilisation de ces derniers pour attaquer d'autres systèmes. Aucun système ou produit informatique ne doit être considéré comme étant complètement sécurisé et aucun produit, service ou mesure de sécurité ne peut être entièrement efficace contre une utilisation ou un accès non autorisé. Les systèmes, les produits et les services IBM sont conçus pour s'intégrer à une approche de sécurité complète, qui implique nécessairement des procédures opérationnelles supplémentaires, et peuvent avoir besoin d'autres systèmes, produits ou services pour optimiser leur efficacité. IBM NE GARANTIT PAS QUE TOUS LES SYSTÈMES, PRODUITS OU SERVICES SONT À L'ABRI DES CONDUITES MALVEILLANTES OU ILLICITES DE TIERS OU QU'ILS PROTÈGERONT VOTRE ENTREPRISE CONTRE CELLES-CI.

1 « Big Business with Big Data », Accenture. Avril 2014 https://www.accenture.com/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_14/Accenture-Big-Data-POV.pdf

2 « 2015 Cost of Data Breach Study : Global Analysis », Ponemon Institute. Mai 2015. http://ibm.com/security/data-breach

Sécurité
Route 100
Somers, NY 10589

Produit aux États-Unis d'Amérique
Novembre 2015

Le présent document est à jour à la date initiale de publication et peut être modifié par IBM à tout moment. Les offres ne sont pas toutes distribuées dans tous les pays dans lesquels IBM exerce son activité.

Les données de performances présentées dans ce document ont été obtenues dans des conditions de fonctionnement spécifiques. Les résultats réels peuvent varier.

Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits ou logiciels non expressément référencés par IBM.

LE PRÉSENT DOCUMENT EST LIVRÉ « EN L'ÉTAT » SANS GARANTIE D'AUCUNE SORTE, EXPLICITE OU IMPLICITE, Y COMPRIS TOUTE GARANTIE D'APTITUDE À L'EXÉCUTION D'UN TRAVAIL DONNÉ ET TOUTE GARANTIE OU CONDITION DE NON-CONTREFAÇON. Les produits IBM sont garantis conformément aux conditions de leur contrat de vente.

Le client est seul responsable du respect des lois et réglementations en vigueur. IBM ne donne aucun avis juridique et ne garantit pas que ses produits ou services assurent au client qu'il se conforme aux lois ou réglementations applicables.

Les déclarations d'IBM portant sur son orientation ou ses intentions futures sont susceptibles d'être modifiées ou annulées sans préavis, et ne représentent que ses objectifs.

Recyclable

SEB03025-FRFR-00