

BIG BANG
BIG DATA

Survivre dans l'univers de l'Infinite Data

Les 6 bonnes pratiques
sur la protection des données

VERITAS[™]
The truth in information.

ARROW

La datasphere mondiale d'aujourd'hui représente **40 Zo de données** créées, dupliquées et enregistrées chaque année, aussi bien dans les data centers, le cloud, les serveurs que dans les smartphones et autres IoT.

D'ici 2025, ce chiffre fera plus que quadrupler : il atteindra les 175 Zo. **Plus de doute, la data est partout et ce n'est pas prêt de s'arrêter !**

Petit rappel :
Zo (zettaoctet)
= 10^{21} octets

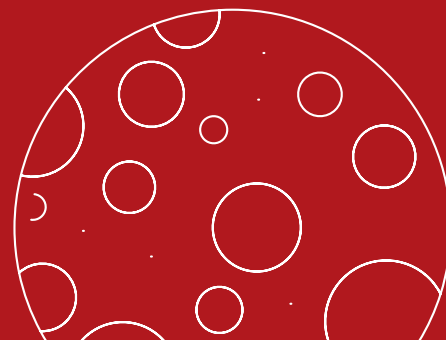
Si l'on devait stocker toute la donnée mondiale sur des DVD et ensuite les empiler les uns aux autres, cette jolie pile pourrait largement faire 23 voyages sur la Lune...

Faut-il pour autant céder à la panique ? Non, pas encore.

Toutefois, réveiller les consciences ne serait pas inutile ! Si les pratiques face à cette surabondance de données commencent à changer, qu'en est-il du monde de l'entreprise ?


Le secteur professionnel parle le plus souvent en termes de « stockage » et pas encore assez en termes de « données ». Derrière cette nuance sémantique se cache une réalité : les utilisateurs stockent encore trop et cumulent sans limite les données. Quand il n'y a plus d'espace de stockage, on se contente de racheter du disque et des octets, sans se dire qu'en multipliant ainsi la donnée, on multiplie les risques de cyberattaques et la perte de données vraiment importantes.

Si l'on comprend qu'il faut faire mieux mais avec moins, si des décisions éclairées sont prises et si l'on commence à s'intéresser vraiment et jusqu'au bout à sa donnée, alors, on pourra gouverner sa data, en toute tranquillité !





Les 6 commandements pour survivre à l'univers de l'Infinite Data

- 
- 01 Ta data tu connaîtras
 - 02 Ta data tu cartographieras
 - 03 Des règles tu créeras
 - 04 Les utilisateurs tu sensibiliseras
 - 05 Des formations tu dispenseras
 - 06 Au moins un CDO tu engageras



Connaître sa donnée : le fil conducteur des bonnes pratiques

Avant d'aborder la notion de gestion de données, il est indispensable que toute entreprise puisse connaître et surtout comprendre sa donnée. La connaître, c'est se poser les questions suivantes : où est rangée la donnée ? Quelle est sa taille ? Quel est son usage ? Les réponses à ces questions de base permettent de considérer la donnée dans son cycle de vie.

La phase de naissance

Lorsqu'une donnée se crée, elle évolue et grandit très rapidement. Elle nécessite ainsi une sauvegarde à très haute fréquence.

La phase de vie

La donnée est active, constamment sollicitée et enrichie par les utilisateurs. Elle nécessite donc une sauvegarde régulière. Il s'agit du cœur de la donnée d'une entreprise, c'est la donnée qui a la plus haute valeur stratégique.

La phase informative

Moins souvent consultée, la donnée garde toutefois une valeur consultative. Sa fréquence de sauvegarde est donc moins importante.

La phase d'archivage puis de destruction

La donnée est détruite si elle n'est pas soumise à une obligation légale ou administrative de conservation et d'archivage, comme les factures ou autres documents à valeur légale.

Si l'on ne connaît pas la finalité de sa donnée, comment appliquer jusqu'au bout son cycle de vie ?

Les informations et les données sont les actifs les plus précieux d'une entreprise. Pourtant, le cycle de vie des données est encore trop peu mis en œuvre. La cause ? Une méconnaissance de son usage. Les phases d'archivage puis de destruction sont à peine envisagées et absolument toutes les données sont logées à la même enseigne et au même endroit, peu importe leur valeur stratégique ! Par exemple, si des données sensibles doivent, par mesure de sécurité, s'inscrire à deux emplacements différents et si on applique ce mirroring par défaut à toutes les données, cela signifie qu'on duplique aussi des données que l'on ne consulte plus depuis des années, au même titre que la donnée sensible.

La maturité des entreprises en matière de données aboutira lorsqu'elles sauront garder la bonne donnée, au bon endroit, au bon moment et surtout quand elles auront conscience qu'une donnée a toujours une finalité : l'archivage puis la destruction.

—○— Amaury COUFFIGNAL
Ingénieur Avant-Vente, Veritas Technologies LLC

Cartographier et appliquer le cycle de vie de la donnée

L'afflux massif et souvent non structuré de la donnée nécessite qu'elle soit analysée, triée, segmentée puis qualifiée au sein des 4 étapes du cycle de vie. À chacun de ces axes, la donnée devrait se voir attribuer un repository et des règles au regard de son modèle de protection, d'usage et de stockage ainsi qu'un budget spécifique.

Qu'est-ce ça implique ?

Une cartographie induit de prendre des mesures distinctes au regard de chacune des étapes du cycle de vie. On ne stocke pas la donnée active et celle qui n'a pas été utilisée depuis des années au même endroit. Une fine cartographie de la donnée permet donc une optimisation plus judicieuse du stockage, notamment en termes de coût. On optera ainsi pour des environnements de stockage et d'architecture (des serveurs aux clouds) les plus adaptés.

Il en va de même pour les mesures de protection : on ne protège pas la donnée qu'on utilise chaque jour de la même façon que la donnée que l'on consulte une fois par an. La donnée en phase de création et celle qui a une valeur informative, ne nécessitent pas non plus la même fréquence de sauvegarde.

Une donnée en phase informative n'a pas vocation à être sauvegardée aussi souvent qu'une donnée en phase active et encore moins qu'une donnée en phase de création.

Une fois encore, les mauvaises habitudes perdurent. Trop d'entreprises sauvegardent toute la donnée à la même fréquence. Les conséquences ? Des sauvegardes qui durent des heures et dont la longueur nuit au bon déroulement de l'activité. Il suffirait simplement de déterminer la donnée stratégique, moins dense, des données informatives voire inutiles, pour faire des sauvegardes plus régulières et plus rapides. Segmenter ainsi les sauvegardes selon le degré de valeur de la donnée permet non seulement de mieux la contrôler, mais aussi de mieux la protéger !

La cartographie de la donnée permet à une entreprise de mesurer si à terme, elle applique des pratiques efficaces ou si elle se dirige vers un modèle déviant.

—⊙— Amaury COUFFIGNAL
Ingénieur Avant-Vente, Veritas Technologies LLC

Créer et appliquer des règles normatives

Créer des règles pour chacune des étapes du cycle de vie est un bon début : à chaque phase, un budget, un modèle de sauvegarde et de protection. Les inscrire dans une charte informatique leur confèrera une forte valeur normative que l'on respectera plus facilement. Par ailleurs, la charte informatique rappelle les droits et les devoirs de chaque utilisateur au regard de l'outil informatique professionnel.

La charte s'avère être le bon endroit pour rappeler que les données personnelles, comme les photos de vacances, ne sont pas vraiment à leur place sur le serveur de l'entreprise.

Quelles normes pour les données non structurées ?

Il existe pour un certain nombre de données, comme des factures ou des documents à valeur légale et administrative, une obligation d'être conservées un certain nombre d'années. Une fois ce délai passé, on peut les détruire en toute certitude, puisqu'une norme l'y autorise.

Mais qu'en est-il des données comme les documents textes ou multimédias, ou encore les emails ? Ces données sont les plus difficiles à référencer et à quantifier tant elles sont nombreuses et sans format

défini sur les périphériques, les serveurs et autres environnements de stockage. Pourquoi ne pas créer des règles spécifiques à ces données ? On pourrait imaginer que les emails de plus de 2 ans doivent être supprimés ou que certains documents de travail puissent être détruits au bout de 5 ans, et même moins si son auteur a quitté l'entreprise ! Pourquoi pas s'inspirer des normes du RGPD pour établir sa charte ?

Si tous les utilisateurs d'une société prenaient l'habitude d'effacer leurs emails de plus de 2 ans et que l'on étendait cette pratique à l'échelle mondiale, l'économie de stockage serait phénoménale... et gratuite !

— Amaury COUFFIGNAL
Ingénieur Avant-Vente, Veritas Technologies LLC

Le rôle moteur du RGPD

Depuis le 25 mai 2018, le Règlement Général sur la Protection de la Donnée (RGPD) sur la confidentialité des données, au champ d'application le plus vaste au monde, occupe l'esprit de tous les aficionados de la data. Et pour cause, le RGPD a introduit des notions de pénalités et d'amendes fortes en cas de non-conformité. Les directions et les entités juridiques sont en droit de vérifier si une direction des systèmes d'information sait identifier et localiser rapidement la donnée.

Cela va même plus loin, puisque désormais la conformité au RGPD est de plus en plus une condition pour faire du business. Si une société ne se plie pas aux exigences du RGPD, elle prend le risque de se voir fermer la porte à des opportunités commerciales.

04

Sensibiliser tous les utilisateurs

Faire prendre conscience du coût réel de l'IT

Les métiers et les utilisateurs n'ont pas forcément connaissance des coûts réels de stockage. Ils se tournent vers les DSI pour demander une certaine quantité de stockage pour tel budget sans vraiment savoir si le stockage souhaité répond à leurs besoins réels. Bien souvent, les demandes excèdent ces besoins.

La DSI doit être en mesure de lui dire si ce stockage est nécessaire pour un an, 5 ans ou plus. Elle doit aussi pouvoir l'informer que ce stockage va lui coûter tant en termes d'administration et de protection. Fort de ces informations précises, les utilisateurs peuvent réaliser les coûts relatifs à la demande de stockage et se rendre compte qu'ils peuvent se contenter de moins, pour moins cher !

Héberger la donnée sur le cloud, c'est souvent faire appel à des sociétés externes.

Elles présentent le volume de données hébergées et sont par exemple en mesure de dire combien de données inactives sont présentes et combien elles coûtent. Par cette verbalisation, la prise de conscience est plus efficace du côté des utilisateurs.

Trop de données, plus de risques

L'accumulation de données inutiles et dupliquées à l'infini augmente le risque de vols ou de pertes de la donnée. L'équation est logique, plus on a de biens, plus on a de chance de les perdre ou de se les faire voler. Pour la donnée, c'est la même chose : trop d'échanges et de mauvais comportements permettent aux ransomware de s'infiltrer parmi l'amas de données. Aujourd'hui, toutes les entreprises, quelle que soit leur taille, sont sujettes à de tels risques.

Les risques sont bien réels et les pertes financières... astronomiques.

Une entreprise moyenne immobilisée à cause de la perte ou du vol de données **perd 4900 € par minute soit 260 000 € par heure**. Pour les grands groupes, ces chiffres s'envolent pour atteindre des millions d'Euros à la journée !

Le ransomware a encore de beaux jours devant lui !

Une entreprise est victime d'une attaque de ransomware toutes les **14 secondes en 2019, et d'ici 2021, ça sera toutes les 11 secondes...**

La data, ça consomme !

Il s'avère très compliqué de saisir l'ampleur de l'impact du numérique sur la planète. Pourtant, bien que silencieuse et furtive, elle est bien réelle. Pour que les consciences s'éveillent sur le sujet, il suffit simplement de rappeler qu'être éco-responsable, c'est consommer moins. Il en va de même pour la data, si on consomme moins de données, on est en mesure de mieux la gérer et donc, d'être plus efficace !

Qui n'a jamais perdu de longues minutes à chercher un document sur son ordinateur ?

Les DSI peuvent engager la démarche de sensibilisation en mettant en place par exemple des actions RSE ou encore en s'engageant à n'utiliser que du matériel « green certified ». Tous les scénarios sont possibles !

05

+

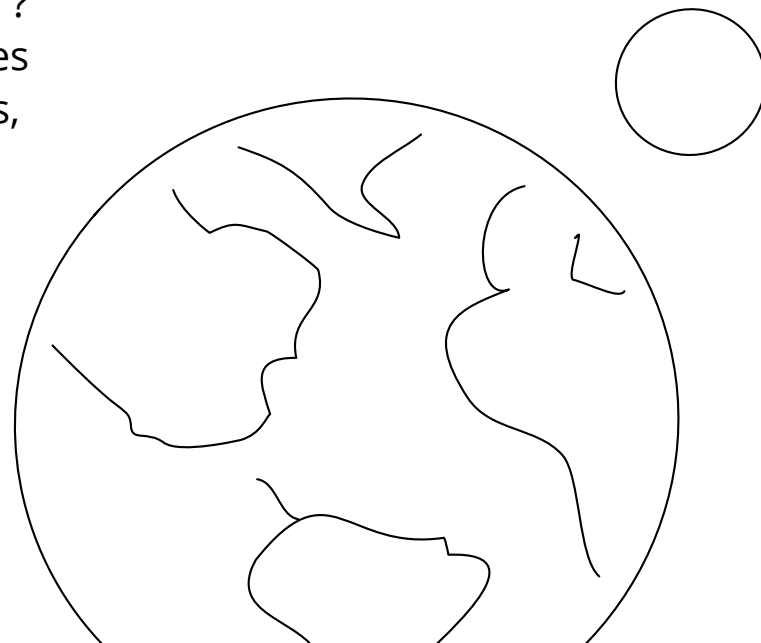
Dispenser des formations

Il revient à la DSI et à toutes les instances en charge du numérique de veiller à ce que les bons gestes perdurent. Pour défendre leurs données, beaucoup d'entreprises se sont lancées dans la sensibilisation de leurs utilisateurs par des simulations d'attaques par phishing ou par hameçonnage. Les résultats sont souvent édifiants et bien que cela génère chez certains une forte prise de conscience, beaucoup trop de collaborateurs tombent encore dans le piège, même lorsque l'exercice est renouvelé sous différentes formes.

Pourquoi ne pas aller plus loin, en établissant par exemple des programmes de formations ? Celles-ci seraient obligatoires pour tous les collaborateurs,

quel que soit leur niveau hiérarchique ou leur niveau supposé de connaissances de l'outil informatique. Protéger sa donnée passe par des comportements éclairés face aux ransomware. Mais l'erreur humaine fait aussi beaucoup de dégâts ! Combien de fichiers ont été perdus ou volés à cause de mauvais comportements ou d'oublis de la part des collaborateurs ?

+



Et si la meilleure défense contre la perte et le vol de données était d'avoir des utilisateurs aguerris aux bonnes pratiques ?

A l'issue de ces formations, il convient aussi de mesurer si les connaissances ont bien été assimilées. Les utilisateurs avec par exemple moins de 80% de bonnes réponses à des quiz ou à des tests devraient suivre à nouveau les formations.

Des milliards de fichiers à la dérive chaque année !

18,5 millions de fichiers et de données ont été volés ou perdus chaque jour durant le premier semestre de 2018, dans le secteur professionnel mondial !
33,6 % sont liés à des pertes accidentelles.

06

Développer le rôle du Chief Data Officer

C'est un statut qui en est encore à ses balbutiements et qui reste encore aujourd'hui l'apanage des grands groupes. Mais cet expert de la donnée est pourtant un élément central de toute entreprise avec la croissance infinie de la data et des objets connectés. Ce rôle a tout à gagner d'intégrer massivement les PME, les agences et autres start-up.

Le CDO, véritable gourou de la data

Le Chief Data Officer joue un rôle essentiel dans la stratégie d'entreprise parce que l'une de ses actions principales consiste à sélectionner parmi l'amas de données, celles qui seront les plus pertinentes pour la bonne marche de l'activité. Le CDO doit également insuffler une véritable culture de la data auprès des acteurs, mais aussi auprès des clients d'une entreprise : sensibiliser sur son importance tout comme avertir sur les dérives possibles, de concert avec les DSI.

Pour que cela fonctionne réellement, le CDO et les autres experts de la data doivent, soit dépendre de la direction générale, soit dépendre de l'entité juridique et administrative d'une entreprise. De cette manière, il leur est plus facile d'entamer des actions, donner des directives et pour tout simplement être mieux écoutés. Pour qu'un travail commun autour de l'enjeu de la data soit réellement et efficacement entrepris, il faut pouvoir s'appuyer sur des services qui sont les voix de la notion de risques.

Sensibiliser les utilisateurs à la donnée, la priorité des CDO

Pour **56% des CDO français**, l'objectif ultime est de démocratiser la donnée à travers les différents services d'une entreprise. Mais seulement **30% d'entre eux sont rattachés à sa direction générale.**

bigbang-et-bigdata.fr

BIG BANG & BIG DATA

VERITAS™

The truth in information.

ARROW

Sources :

IDC et Seagate : Data Age 2025 The Digitization of the World From Edge to Core

Gartner 2018 Ensure Cost Balances Out With Risk in High-Availability Data Centers by Dave Cappuccio

Cybersecurity Venture : Annual Cybercrime Report

Data Breach Level & Gemalto : Data Privacy and New Regulations Take Center Stage - POWERED BY 2018 FIRST HALF REVIEW

Informatica : Un Chief Data Officer - Pour quoi faire ? 50 Données clés sur les CDO français

Réalisation : swpl.fr