

GOUVERNANCE DE LA DONNEE

Qu'est-ce que le data gouvernance : DIAPO 2 :

La gouvernance de la donnée est un modèle de cycle de vie de la donnée qui va décider des étapes à franchir par la donnée de sa collecte à sa restitution dans un data center ou bien un data lake. Avant même de collecter de la donnée on va s'interroger de comment on va le faire, avec quel outil et dans quel but. Généralement au sein d'une même entreprise un seul modèle va être instauré de sorte à uniformiser l'output de ses applications qui à un moment ou un autre va se retrouver dans un data center.

Au sein d'une entreprise il peut y avoir une entité destinée à la gouvernance de la donnée mais ce n'est pas toujours le cas. Donc sur quel axe pouvons-nous nous appuyer pour savoir si cela est nécessaire.

Que prendre en compte DIAPO 3 :

- Premièrement la taille de l'entreprise
- Deuxièmement temps ou l'argent
- Et enfin à qui est destiné l'application ? Sera-elle publique ou bien privé/interne ?

Alternative DIAPO 4 :

Avant même de mettre en place un MDM (Master data management), la gouvernance de la donnée c'est avant tout beaucoup de bon sens et donc une réflexion dès la collecte, par exemple : quel format monétaire allons-nous choisir : des euros ou bien des livres, le séparateur une virgule ou bien un point, un format de date international donc les mois avant les jours ou encore certains caractères spéciaux et accents. Ça commence donc par de petit geste au cours d'une éventuelle formation que l'on commence à instaurer une gouvernance.

Les Avantages DIAPO 5 :

Un modèle de gouvernance bien désigné en fonction des besoins, évite d'allouer plus d'argent et de temps à de service de maintenance comme le data cleaning et améliore donc la fluidité du cycle de vie ainsi que sa compréhension.

La qualité et/ou l'intégrité d'une donnée DIAPO 6 :

Le terme de data qualité, fait référence à sa qualité ainsi qu'à son intégrité on se base sur quel critère ? La bonne qualité d'une donnée ce n'est pas une donnée qui n'a pas d'accent, ce n'est pas une donnée en capital case, c'est tout simplement une donnée qui respecte les points de contrôle établis par un modèle de gouvernance.

Ces données vont être utilisées dans des algorithmes procéduraux chargés à prendre des décisions il est donc essentiel que ses données soient pertinentes. Le plus anodin des données est sensible et ce pour n'importe quel opérateur data. Notamment en milieu hospitalier où il est fondamental d'avoir des données pertinentes comme pour un dossier qui va être attribué à l'ID d'un patient ou bien une température à ne pas dépasser à un générateur d'énergie ou bien un moteur ect.

Comment améliorer la qualité d'une donnée ? DIAPO 7 :

Dans un premier temps nous sommes tous acteurs du domaine de la data, quand on crée une adresse mail ou quand on contracte un abonnement c'est déjà une étape de référencement de la donnée. L'utilisateur doit être conscient et vigilant des données qu'il référence, cela va du prénom à son social ID. Mais on doit aussi mettre à jour ces données, changement de numéro de téléphone, changement d'adresse ect.

A ce jour être sensibilisé et sensibiliser son entourage à une meilleure compréhension de la data et son évolution est un atout majeur au sein de notre société, car c'est aussi par des gestes simples que l'on peut contribuer à sa qualité.

Qu'est ce qui peut nuire à la qualité d'une donnée ? DIAPO 8 :

Désynchronisation d'une base de données, donc une perte lors d'une transaction. Une donnée tronquée, ou une mauvaise gestion dite "manuelle" par un opérateur data en charge du traitement de la donnée. Un piratage un bug/virus et du matériel défaillant.

Voici quelques exemples de données qui peuvent nuire à des procédures ou à sa qualité directement :

Une donnée remplie par défaut, une donnée absente, une donnée incohérente ou tout simplement erronée, une même donnée calculée différemment ou sujette à plusieurs interprétations donc imprécise ou tout simplement une donnée obsolète.

Dans la majorité des cas cela va résulter à une défaillance qui sera traduite par un temps de traitement plus long ou bien des opérations en suspens, des coûts plus importants dans un système financier, des mauvais conseils à un client ou encore un ciblage marketing imprécis ect...

Cycle de vie DIAPO 10 :

Pour revenir à son procédé, comment se déroule ce cycle de la vie d'une donnée ?

La première étape étant de mettre en place des contrôles à différent point du cycle de la fluctuation de la donnée à commencer au moment de sa collecte auprès d'une source qui pourra certifier de la qualité de cette donnée que l'on appelle "GOLDEN SOURCE". Ensuite en se référant à son modèle on va vérifier à chaque point de contrôle si les KPI sont respectés à savoir si la donnée est constante et qu'elle n'a pas été altérée.

La deuxième étape étant la gouvernance de la data qui a pour but de définir clairement la responsabilité de chaque opérateur de la data chargé d'une procédure comme un data owner qui va réfléchir sur l'usage de cette donnée, un data designer qui va se poser sur l'architecture et enfin un data officer qui va certifier de la qualité de cette donnée. C'est donc par l'officier de la data que les décisions seront prises.

One more thing

Pour instaurer ou respecter une gouvernance nous devons nous soucier également de l'intégrité d'une donnée qui représente la légalité de l'utilisation de la donnée, de la divulgué ou même de la stocker.

Une évolution notable à prendre en compte pour le traitement de la donnée et la réforme RGPD mise en vigueur en 2018 qui se chargera de vérifier vis à vis des entreprises la conformité des données des utilisateurs de son respect, sa circulation et sa divulgation. Cette réforme nous a généralement impacté depuis, vous l'avez sans doute constaté mais sur différents services vous avez dû réaccepter des conditions d'utilisation sur certains services.

Il y a d'autres évolutions plus spécifiques comme la bcbs 239 regroupant 14 principes qui ont pour but de participer à la stabilisation du système financier mondial contraignant les banques à identifier les risques auxquelles elles sont exposées. En 2018 une deuxième réforme a été mise en vigueur, la réforme BCE anacredit c'est un data lake européen qui contraint les banques à un reporting conséquent à un niveau national pour chaque prêt contracté qui sera enfin transmis à la BCE.

CONCLUSION :

UNIFORMITE UTILITE LEGALITE CONSTANCE