

La gestion du patrimoine digitale



Table des matières

Objectifs	3
Introduction	4
I - La gestion des bases de données	5
1. La centralisation des bases de données	5
1.1. L'importance de la centralisation des données	5
1.2. Le processus de centralisation	6
2. Les données distribuées ou décentralisées.....	7
3. L'équipe de gestion de la qualité des données.....	7
3.1. Le directeur général des données.....	7
3.2. Le sous directeur des données	8
3.3. L'assistant chargé des données	8
4. Les outils de données	8
II - Le classement et l'archivage de données	9
1. Le classement des données	9
1.1. La catégorisation des données	9
1.2. Les méthodes de classification des données.....	9
2. L'archivage des données.....	9
2.1. Définition	9
2.2. La réglementation de l'archivage des données	10
2.3. Les outils de l'archivage des données	10

Objectifs



- Présenter la gestion des bases de données ;
- Identifier les types de classement et d'archivages des données numériques.

Introduction

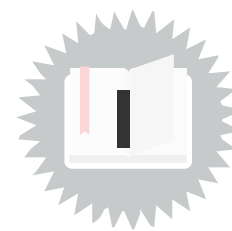


L'idée est de savoir si les données doivent être conservées dans un endroit unique pour faciliter l'opérationnalité ou la disponibilité des données ou si on doit disperser ces données pour garantir la spécification géographique.

Une étude conduite en 2020 a montré que sur un échantillon de 1000 entreprises, seules 39% d'entre elles sont capables d'exploiter la puissance des données (<https://www.capgemini.com/fr-fr/perspectives/publications/data-powered-enterprise>). En plus, seuls 20% des 15 dirigeants d'entreprise interviewés ont confiance dans leurs données.

A contrario, les entreprises qui comprennent l'importance des données voient une amélioration des ventes de nouveaux produits de 19% (https://www.capgemini.com/fr-fr/wp-content/uploads/sites/2/2021/02/Rapport_data-powered-enterprise_FR.pdf).

La gestion des bases de données



1. La centralisation des bases de données

Il s'agit de la manière de collecter les données numériques provenant de divers supports ou de diverses activités dans un espace donné. Dans entreprise privée telle que DILITRUST par exemple fournissent des accès sur le Cloud pour les données d'une entreprise cliente.

1.1. L'importance de la centralisation des données

a) Une fiabilité élevée

Quand les informations numériques sont centralisées, cela permet de les corriger et de les compléter au fur et à mesure des évolutions de l'entreprise ou des changements qui l'affectent. Il est aussi facile de les standardiser, donc de les exploiter directement. En effet, stockées, les données en un même lieu permettent de construire un schéma (système de classement) ou un chemin (organisation des bureaux des différents des centres de données internes) d'accès rapide.

Les données fiables facilitent le reporting périodique.

b) Une vision globale

Avec les données centralisées, il y a une autorité unique qui planifie la gestion des données, donnant une certaine cohérence dans la gestion des informations. Les acteurs des différents services traduisent en données opérationnelles liées à leur service, car comprenant la logique globale d'évolution de l'entreprise.

c) Une réduction des coûts et la sécurité

Quand les données sont stockées dans un même parc informatique, on note une réduction de déplacements et de la maintenance. Par ailleurs, il est facile de sécuriser des objets qui se situent sur un même espace que sur une diversité d'endroits.

d) Le problème des données cloisonnées

Les données cloisonnées sont des informations dispersées dans les départements de l'entreprise ou ses différents services, soit du fait des formats différents (clés, textes papiers, feuilles Excel, Cloud via Drive, etc.), soit des informations contradictoires.

Ces données nécessitent une consolidation avant d'être exploiter efficacement. Cette consolidation est plus facile à réaliser avec des données centralisées.

e) La conformité avec les lois en vigueur

Les données centralisées facilitent la conformité à la réglementation. Etant dans un endroit unique, donc accessibles facilement, il est facile de les adapter au changement dans la législation.

1.2. Le processus de centralisation

<https://www.astera.com/fr/type/blog/managing-data-quality-in-healthcare/>

L'idée est d'obtenir un système de données structurées. Une donnée structurée est quantitative et se présente sous forme de tableau par exemple ou de schéma d'arbre (ordinogramme). Une donnée non structurée est un texte. Il convient de le structurer pour une exploitation optimale des bases de données. Il y a aussi des données semi-structurées. Mais, avec l'avènement des méga-données et des centres de données, on note de plus en plus de tentative de gestion des informations apparaissant non structurées.

On parle de la qualité des données pour désigner le fait que les informations soient conformes à un certain nombre de critères. Ces derniers sont : l'exactitude, la cohérence, la validité, l'exhaustivité, l'actualité et l'unicité.

L'exactitude se rapporte à l'idée que les données sont censées représenter ce pourquoi elles ont été générées. Il peut s'agir d'événements ou d'entités particulières. La cohérence, elle, désigne le caractère uniforme des informations d'un système à un autre. La validité indique que les données sont conformes aux règles et paramètres définis par l'entreprise, qui garantissent que les données sont correctement structurées et contiennent les valeurs qu'elles devraient contenir.

On parle de données exhaustives quand celle-ci comprennent toutes les valeurs et tous les types de données qu'elles sont censées contenir, y compris toutes les métadonnées qui devraient accompagner les ensembles de données. Dans ce cas, les données se doivent d'être mises à jour régulièrement pour être utilisables sur le moment. C'est l'actualité des données. Enfin, l'unicité des données indique qu'il n'y a pas de doublon dans l'emplacement des données par exemple.

a) L'analyse des données existant

On parle aussi d'évaluation de l'infrastructure. L'idée est de rechercher les différentes sources de données, les redondances, les lacunes, la qualité des données, etc.

b) La détermination des objectifs de la centralisation

On cherche ici à savoir comment la centralisation des données s'adapte à la politique globale de l'organisation (Etat, entreprise, etc.). L'objectif de la centralisation peut être l'amélioration de la qualité des données, la simplification ou la conformité.

Dans l'ensemble, les avantages ci-dessus constituent aussi des objectifs de la centralisation.

c) La précision du cadre de gouvernance

Il s'agit de déterminer la propriété des données, les niveaux de responsabilité sur les formats choisis. Il s'agit aussi de préciser les niveaux d'accès et la politique de sécurité.

d) La sélection de l'architecture des données

C'est le choix entre le Cloud, les lacs de données, les systèmes de gestion référence, les entrepôts de données et les magasins de données (datamarts).

Les lacs de données sont des techniques de stockage rapide de données volumineux sans modifications préalables (données originales ou dans le format original).

Les systèmes de gestion des données de référence sont des systèmes de gestion des bases de données (SGBD), c'est-à-dire des logiciels permettant de stocker, de partager des données d'une base de données dans une autre base de données de façon à garantir la confidentialité des informations en masquant la complexité des opérations.

Les entrepôts de données sont des systèmes de stockage de données (textes, fichier Excel, vidéo, XML, etc.) qui implique un traitement préalable des données. Ce qui donne une architecture cohérente des données, passant par exemple de données hiérarchiques à des données transversales. Gérer un entrepôt de données implique de définir des normes de traitement, la gestion des références et un système pour les métadonnées.

Quand les données sont mises dans un entrepôt, il est difficile de les modifier, car elles apparaissent lecture seule.

Les data marts ou magasins de données sont une composante des entrepôts. Leur particularité est qu'un magasin de données porte sur un sujet précis et n'implique de nombreuses actions ou recherches pour trouver les données (<https://www.youtube.com/watch?v=HvvNEYABKDo>).

e) L'intégration des données

Il convient de chercher une façon de rassembler les données de diverses sources (migration) en vue de l'analyse des données par exemple.

NB : Il existe plusieurs autres étapes dont la sécurité, l'accès aux données, les itérations régulières pour actualiser le système des données aux évolutions de l'organisation ou pour gérer les redondances.

2. Les données distribuées ou décentralisées

Son fondement est d'éviter les pannes qui pourraient paralyser l'accès aux données. Toutefois, dans le cadre de la gestion des données en informatique, les spécialistes préfèrent les données décloisonnées ou centralisées aux données décentralisées ou cloisonnées.

On parle de silos d'information lorsque les données sont cloisonnées par département ou service par exemple. Cela signifie que l'accès à ces données est limité pour les personnes ou structures qui ne sont pas la structure en question. Ce qui rend difficile la coordination générale des tâches et l'analyse globale des statistiques de l'organisation. On voit que les silos de données empêchent la création d'un entrepôt de données.

3. L'équipe de gestion de la qualité des données

3.1. Le directeur général des données

Il définit les règles de gouvernance, suit la mise en œuvre et coordonne l'ensemble des activités de Data Management.

Le rôle du CDO :

- Développer et maintenir une stratégie data,
- Organiser la gouvernance des données,
- Définir et promouvoir le modèle de données d'entreprise,
- Sponsoriser les projets et les services de gestion des données,
- Communiquer et promouvoir les politiques, les standards et les procédures de gestion des données,
- Superviser et coordonner les activités de gestion des données,
- Suivre les coûts liés à la gestion des données et proposer des optimisations.

3.2. Le sous directeur des données

Le Data Management Executive (DME) organise les services de gestion des données, en cherchant à développer le partage et la réutilisation pour plus d'efficacité et de cohérence.

Ses missions consistent à :

- Définir et mettre en œuvre des architectures de données répondant aux besoins métiers,
- Définir et maintenir un cadre de mutualisation des services transverses liés aux données,
- Coordonner et piloter les projets d'évolution de l'architecture des données,
- Suivre les coûts liés à la gestion des données et proposer des optimisations,
- Garantir la mise en œuvre dans le SI des contrôles permettant le respect des exigences de qualité et des contraintes réglementaires.

3.3. L'assistant chargé des données

Le Data Steward (par famille de données) décrit les besoins de gestion des données dont il a la charge en fonction des usages. Il suit la mise en œuvre et veille à délivrer des données de juste qualité.

Son rôle consiste à :

- Comprendre les processus liés à la collecte de la donnée et à sa consommation,
- Recenser les besoins et les exigences relatifs aux données, les hiérarchiser, et proposer des solutions / services en coordination avec le DME,
- Mettre en œuvre les bonnes pratiques de gestion des données en coordination avec le CDO,
- Gérer et résoudre les problèmes liés aux données,
- Suivre les coûts liés au Data Management sur son périmètre et proposer des optimisations.

4. Les outils de données

Talend

(<https://www.youtube.com/watch?v=UrUiKhJ2EOo>)

<https://www.youtube.com/watch?v=HlAXBcEKZ0s>

Collibra

Data Galaxy,

Ab Initio,

Informatica, etc.

Le classement et l'archivage de données



1. Le classement des données

Il s'agit de catégoriser les fichiers ou données de sorte à déterminer les accès et les politiques de stockage des données et de diffusion.

1.1. La catégorisation des données

On peut catégoriser les données de plusieurs manières. En fonction des objectifs, on peut avoir des données très sensibles, sensibles et des données peu sensibles.

Les données très sensibles ont des accès limités en termes de nombre et de qualité des personnes. Les données sensibles sont des fichiers ou des données qui ne peuvent pas être divulgués au public, mais pour lesquels une violation des données ne constitue pas un risque important. Ils nécessitent des contrôles d'accès comme les données à haute sensibilité, mais un plus grand nombre d'utilisateurs peuvent y accéder. Enfin, les données peu sensibles ont un accès plus ou moins libre.

1.2. Les méthodes de classification des données

Il s'agit d'un ensemble de technologies de gouvernance et de protection des données.

a) La gestion de l'accès aux identités

On parle aussi d'outils IAM (Identity Access Management). Ils permettent aux administrateurs de déterminer les personnes susceptibles d'accéder aux données. Les groupes reçoivent des niveaux d'autorisation et sont gérés comme une seule unité. Lorsqu'un utilisateur quitte l'entreprise, il peut être retiré du groupe. Ce qui élimine toutes les autorisations pour cet utilisateur. Ce type de regroupement et d'organisation rationalise la gestion des autorisations sur le réseau.

b) Le chiffrement des données

Il s'agit d'un ensemble de techniques permettant de chiffrer certaines parties des données pour éviter les attaques sur les données.

c) L'automatisation

Les données sont étiquetées et classées de façon automatique grâce à l'intelligence artificielle.

2. L'archivage des données

2.1. Définition

L'archivage des données se rapporte au fait de pouvoir déplacer une donnée ancienne d'un espace de stockage à un autre. L'idée est de pouvoir récupérer cette donnée pour des besoins de réutilisation ou d'enquête par exemple. L'archivage est donc différent de la sauvegarde (mise de la donnée sur un support numérique).

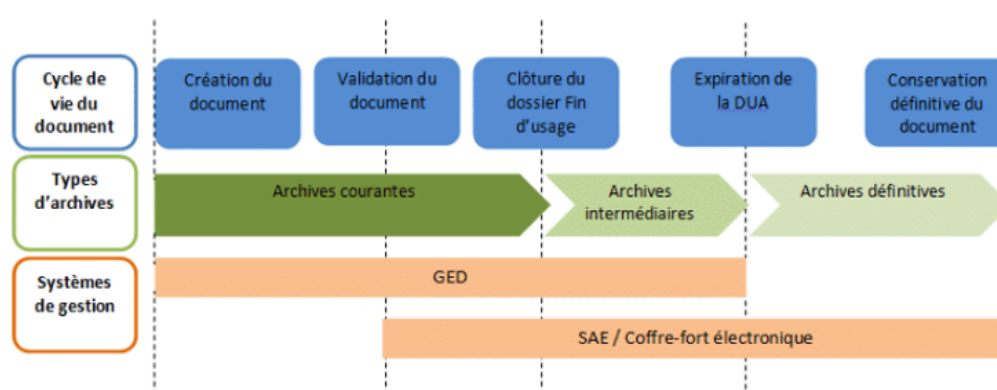
2.2. La réglementation de l'archivage des données

En Côte d'Ivoire, une loi encadre l'accès et la diffusion de données personnelles (voir plus haut). Il y a en France le Règlement générale sur la protection des données (RGPD) qui donne des pratiques de gestion des risques en manières d'archivage des données numériques (https://www.cnil.fr/sites/cnil/files/atoms/files/cnil_guide_securite_des_donnees_personnelles-2023.pdf). On peut aussi citer dans le domaine de la santé la réglementation HIPAA, Health Insurance Portability Accountability Act, (<https://www.sergroup.com/fr/blog/article/reglementation-hipaa.html>), dans le domaine des cartes bancaires, il y a par exemple la certification PCI, Payment Card Industry Data Security Standard, (https://docs-prv.pcisecuritystandards.org/PCI%20DSS/Standard/PCI-DSS-v4_0_1.pdf).

Ces réglementations visent à protéger les données contre les fuites par vols d'identité, les erreurs humaines de manipulation et l'hameçonnage ou phishing. Ce dernier consiste à appâter ou à leurrer un individu pour obtenir de lui des données personnelles (cartes bancaires, mot de passe, codes Internet, etc.).

2.3. Les outils de l'archivage des données

Ces outils sont différents des serveurs. On peut positionner les outils de l'archivage de données numériques ou électroniques sur le cycle de vie d'un document comme on peut le voir ci-dessous.



Le parcours d'un document (source : Novarchive.fr)

a) Le GED

La GED (gestion électronique de documents) est un outil informatique dont l'objectif est de gérer les documents vivants, c'est-à-dire en cours d'utilisation, pour la conduite quotidienne de l'entreprise. La GED est l'espace centralisé dans lequel les documents peuvent être stockés.

Une GED n'a pas vocation à assurer la conservation des documents en apportant des garanties sur leur authenticité, ni sur leur pérennité à long terme. Elle contribue à améliorer l'échange et la manipulation des informations au sein d'une entreprise, par exemple via la gestion des différentes versions d'un même document numérique.

On peut voir une explication du fonctionnement des outils GED sur : <https://www.youtube.com/watch?v=SrdkyYHStbg> et https://www.youtube.com/watch?v=VIHc0a9S_Yk. On peut y voir par exemple qu'il y a plusieurs outils GED comme Zeendoc, Microsoft Sharepoint, Alfresco. Un exemple d'utilisation de l'outil de GED nommé Sarepoint est présenté sur : https://www.youtube.com/watch?v=fJPeQKYL_OU et <https://www.youtube.com/watch?v=BIkfNtTmI0M>.

Il y a aussi une présentation d'un logiciel open source (libre) de la GED. C'est le logiciel "LogicalDoc" : <https://www.youtube.com/watch?v=56hyZRXL9Q> ou sur : www.logicaldoc.fr¹.

1. <http://www.logicaldoc.fr/>

b) Le système électronique d'archivage

A la différence de la GED, les outils SAE permettent de conserver les données longtemps et ces dernières ne sont pas modifiables. Ce qui permet une signature électronique selon les réglementations usuelles. Toutefois, la signature électronique (horodatage par exemple) n'a pas la même valeur juridique (valeur probatoire ou probante) que la signature physique (sur papier). La priorité est donc accordée à la signature papier.

Par ailleurs, même si le SAE permet de conserver les données selon les règles légales (en 10 ans), il n'est pas sûr que les données numériques conservent une durée au-delà de 10 ans, voire 100 ans ou 200 ans par exemple.

Toutefois, le SAE reste le seul garant de la valeur juridique, culturelle et historique des données électroniques. Il a pour ce faire quatre principes : la pérennité des informations (le document doit être déchiffrable), l'authenticité (le document doit pouvoir être validé), l'intégrité (la garantie que ce document n'a pas été modifié) et l'accessibilité (le document numérique doit être communicable).

<https://www.youtube.com/watch?v=oGYdDs8Fn0I>

<https://www.youtube.com/watch?v=oGYdDs8Fn0I&t=728s>

https://www.youtube.com/watch?v=7pAucPsC_uY (pour un exemple de logiciel d'archivage numérique libre ou open source : VITAM). Voir aussi le site : www.programmevitam.fr

c) Le coffre-fort numérique

Un coffre-fort numérique permet une conservation longue des documents numériques. L'une des implications des coffres forts numériques est la définition de la politique d'archivage : types de documents, règles de conservation (la durée d'utilité administrative de conservation, droit à l'oubli, le format de stockage tel que jpeg, PDF, MPEG, etc.). Il faudrait donc définir un cahier de charge.

Les outils de coffre-fort sont par exemple Digiposte, Gerermesaffaires, Efallia, etc.