

Lundi 29/03

Professeur : Hervé TONDEUR, herve.tondeur@yahoo.fr

La donnée... L'or noir du 21e siècle

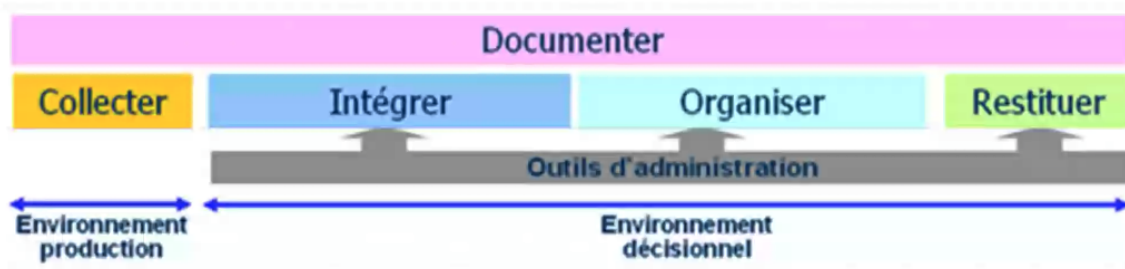
- Que veut réellement dire cette expression?
- Comment une (E) peut réellement transformer ses données en richesse?
- Quelles sont les données dont on parle?
- Où se trouvent les gisements dont l'(E) peut prétendre explorer?
- Quels sont les coûts pour extraire la valeur des données?
- Comment, avec quels moyens et en combien de temps les (E) pourront profiter de cette richesse?

Introduction

Nous allons essayer de comprendre :

- Ce qu'est une donnée et toutes les idées qui tournent autour d'une donnée
- Comment stocker, manipuler, interpréter ces données
- Comment construire des outils de stockage et restitution des données
- Comment comprendre le fonctionnement des différents types d'entrepôts
- Quelle est la \neq entre les entrepôts de pilotage
- Les entrepôts décisionnels
- Les entrepôts d'études et recherches
- Comment peut-on exploiter ces entrepôts, les principes

Place d'un entrepôt de données dans un SI, notion d'urbanisation



SI : "Véhicule" des entités de l'organisation. Sa structure est constituée de l'ensemble des ressources (le personel, le matériel, les logiciels, les procédures) **organisées pour** :

- **Collecter / La collecte**

L'origine de l'info peut être interne (*comptes, stocks*) ou externe (*infos sur le concurrent, disposition nouvelle d'ordre fiscale ou sociale, météo, etc.*)

Pour les infos d'origine externe, il est nécessaire d'y être **tout particulièrement attentif**, il convient donc d'**organiser des veilles techno**

- **Stocker ou Intégrer**

- **Traiter ou Organiser**

La phase de traitement commence avec le choix du support utilisé car il faut trouver une construction formalisée pour traiter l'info:

- soit la centralisation (réalisée à un seul endroit dans l'(E)
- soit la décentralisation (permet à chaque poste de travail d'échanger des infos et travailler en autonomie)
- soit la distribution (permet un traitement au niveau d'un site unique ; la saisie et la diffusion s'effectuent grâce à des terminaux)

- **Communiquer ou Restituer** la donnée aux utilisateurs, aux managers et autres chefs de décision qui vont comprendre et appréhender l'info pour agir sur la collecte des données

Elle doit répondre à 4 critères :

- Origine et destination
- Forme ?
- Délai

→ Diffusion large ou restreinte ?

⇒ il s'agit (les 4 étapes précédentes) de la documentation du SI

→ La qualité coûte cher mais apporte ..

Le SI coordonne les activités de l'organisation et lui permet ainsi, d'atteindre ses objectifs, grâce à **la structuration des échanges**

Un SI se construit à partir de l'analyse des processus métier de l'organisation et de leurs interactions/interrelations, et non seulement autour de solutions informatiques plus ou moins standardisées par le marché

⇒ Le SI est un **système sociotechnique** composé de 2 sous-systèmes : l'un social (constitué de personnes et de relations entre ces personnes) et l'autre technique (Hardware et Software)

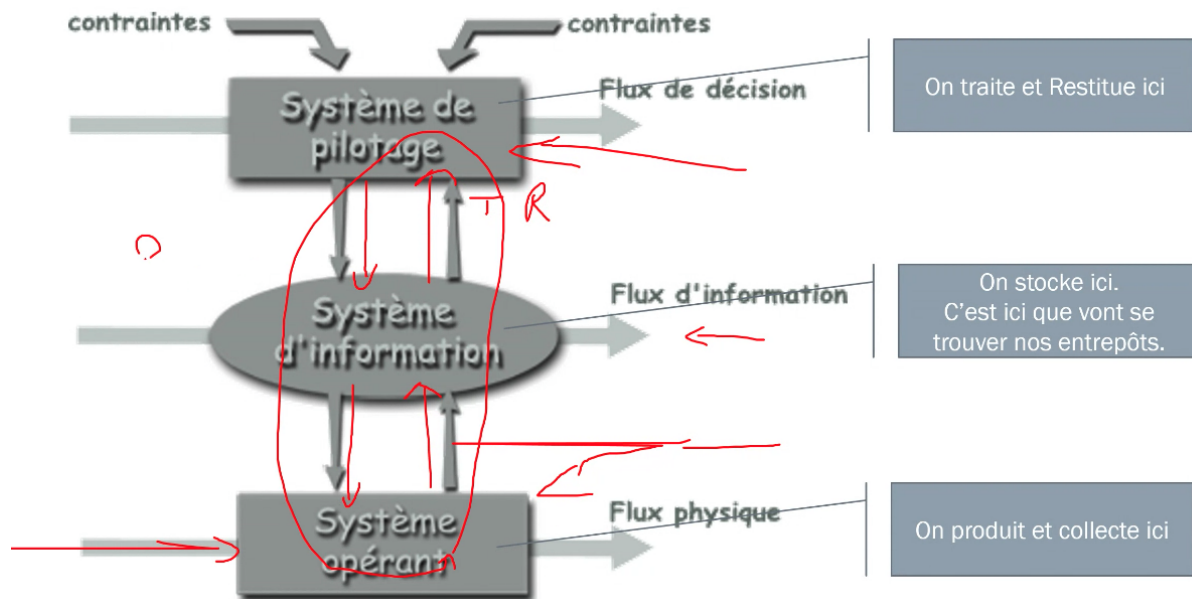
2 grandes catégories de systèmes selon les types d'application informatique :

- **les systèmes de conception** : fonctionnent selon des techniques temps réel ; ici il faut restituer et traiter au plus vite.
- **les SI de gestion** : qui emploient des techniques de gestions (majoritaire)

Ce qui ≠ ces 2 catégories c'est la notion du temps

Organisation : On peut tj décomposer une organisation en 3 sous-systèmes :

- le système opérant : on **produit** et **collecte** l'info
- le système d'information : on **stocke** l'info ici ; **les entrepôts se trouvent ici**
- système de pilotage : on **traite** et **restitue** l'info



Vision des entrepôts de données selon la perspective utilisateurs / DSI / Managériale /technicien informatique

Du point de vue utilisateurs, un entrepôt est un système qui permet d'avoir a disposition pour son quotidien d'informations pour des prises de décisions métiers rapides, de lui permettre de compléter son environnement informatif dans sont domaine opérationnel.

Du point de vue du Directeur du système d'information, un entrepôt est un système qui va permettre de rationaliser son SI et de permettra de pérenniser les informations et consolider le SI.

Du point de vue Managériale, les entrepôts de données sont des outils de pilotages, qui permettent de piloter l'entreprise, d'améliorer la qualité, d'ouvrir des marchés de donner un instantané de l'état de l'entreprise (du point de vue financier (dépenses/recettes, marketing, qualité, ressources humaines).

Du point de vue du technicien est un système informatique composé de SGBD, d'outils d'ingestions de données, d'outils de restitutions des données, de traitement et affichage des données pour les utilisateurs et manageurs.

Difficile de discuter d'entrepôt de données avec \neq personnes car \neq points de vue

Quelques définitions

Entrepôt : toute solution technique permettant de conserver des données de toutes provenance selon certaines règles:

les données doivent être :

→ ordonnées : l'entrepôt est organisé est structuré

- intégrées : les données proviennent de sources hétérogènes utilisant chacune un type de format, elles sont intégrées avant utilisation
- orientées sujet : dans un entrepôt, les données sont organisées par thème
- historisées : les données non volatiles sont horodatées
- non volatiles :

Données : toute info simple ou complexe qui permet de qualifier un info du monde réel ou imaginaire. Elle doit absolument représenter une information, attribut d'un objet ou d'une entité

Sources de données : les sources de données ont de multiples domaines de productions. Tous les domaines métiers produisent de la donnée de tout type

- par la saisie d'opérateurs
- par la capture de mesures par des capteurs
- par le travail et mesure de robots de productions
- par le calcul agrégé de données

Les sources de données internes :

- les données internes sont l'ensemble des elt d'information produites et stockées au sein d'une organisation

Les sources de données externes :

- les données externes sont toutes les données non générées par l'organisation et accessible via des sources extérieures

Données non structurées :

Données semi-structurées : On ajoute aux données non structurées des métadonnées

Métadonnées : ce sont des données qui décrivent des données

Données structurées :

Typage des données et représentation des données métiers et métadonnées
: