

Entrepôts de données, Hébergement, Données massives en santé

Les Entrepôts de données cliniques = Plateformes utilisées pour l'intégration de plusieurs sources de données au travers d'outils d'analyses spécialisés afin de **faciliter le traitement et l'analyse de données massives (Big Data)**.

Les Données Massives (en santé) = Big Data = Les gros volumes de données qui alimentent l'activité quotidienne d'un hôpital.

3 caractéristiques :

- **Volume**
- **Vitesse**
- **Variété**

Problèmes des Big Data :

- **90% du volume total** des données ont été produits ces 2 dernières années
- **Plus de 80%** de ces données ne sont toujours pas exploitées
- **8,9 milliards** de feuilles de soins informatiques dans la base de la Sécurité Sociale

Données structurées = données représentées ou stockées avec un format prédéfini (≈ 20% de données).

Extract – Transform – Load (ETL) :

Extract :

- Connecter les différentes sources de données et extraire les données nécessaires
- **Problème** : hétérogénéité des sources de données

Transform :

- Données extraites transformées dans un format spécifique, défini à l'avance
- **Problème** : définition et reconnaissance des formats à appliquer, prise en charge des nouvelles données, **évolution des formats** de données en fonction du temps, **interopérabilité des formats**.

Load :

- Données transformées dans leurs formes/dimensions finales
- **Problème** : La gestion des "anciennes" données VS celles à mettre à jour

General architecture with optional CDSS :

- Data mart sont harmonisés et transférés dans un CDW
- Avantage : CDSS apporte une fonctionnalité d'aide à la décision
- Chercheurs
- Médecins

Bio-bank driven :

- Liste des échantillons biologiques disponibles
- Avantage : Accès direct aux données brutes des échantillons pour faire des contrôles qualités
- Chercheurs

User-controlled :

- Données pré-traitées et intégrées directement
- Avantage : rapidité
- Inconvénient : exhaustivité des données
- Chercheurs

4) Federated architecture :

- Différents établissements, données intégrées dans un entrepôt de données « virtuel »
- Avantage : Sécurité + flexibilité
- Inconvénient : A la fin de la session, les données sont supprimées
- Chercheurs

- **Project management** = sécurité, identification des utilisateurs/trices, rôles
- **Ontology management** = gère la terminologie
- **File repository** = stocke les « gros » fichiers (images, puces).
- **Workflow Framework** = gère les interactions entre les différentes « hives ».
- **Identity management** = anonymisation des patients.
- **Web client application** = permet aux utilisateurs/trices d'interroger le CDW.
- **Workbench** = application permettant d'analyser les données de façon plus précise.