

# Introduction

Avec l'avènement du Big-Data, quels solutions pouvons-nous utiliser pour le stockage ?

Pour cela il existe deux méthodes bien connus le Data-Warehouse et le Data-Lake.

## Data-Warehouse

Le data Warehouse (entrepôt de donnée) consiste à stocker de manière structurée dans une base de données relationnelle une partie ou l'ensemble des données d'une entreprise pour les utiliser pour la prise de décisions dans l'entreprise par le biais de statistiques et d'effectuer des rapports grâce à des outils de reporting comme Power BI ou Cyfe. Grâce aux informations stockées dans un Warehouse il est possible d'avoir une vue d'ensemble des différentes échanges qui ont eu lieu au fil du temps. Le Warehouse est souvent utilisé pour permettre d'accéder plus facilement aux données stockées dedans.

## Data-Lakes

Un Data Lake (Lac de données) permet le stockage de données. C'est référentiel de données permettant de stocker une très large quantité de données brutes dans le format natif pour une durée indéterminée. Cette méthode de stockage permet de faciliter la cohabitation entre les différentes formes de données. Les données brutes y compris les copies des données système source, côtoient les données transformées d'une entreprise et sont stockées dans un seul Data Lake. Ces données sont ensuite utilisées pour établir des rapports, pour l'analyse de données ou pour le Machine Learning.

## Conclusion

En conclusion une méthode n'est pas meilleure que l'autre, cela dépend de la structure des données, les utilisateurs ciblés, les méthodes de traitement et l'objet global des données. Mais il existe quelques différences entre les deux : -le Data-Lake demande une plus grande capacité de stockage que le Data-Warehouse mais les données brutes contenues exclusivement dans les Data-Lakes sont facilement modifiables, elles sont plus rapides à être analysées et sont idéales pour le machine learning mais les données brutes présentes dans les Data-Lakes non classées peuvent porter à la confusion. Les Data-Lakes sont plus difficiles à explorer pour les utilisateurs qui n'ont pas l'expérience des données non transformées. Les données brutes exigent généralement les services d'un data scientist et des outils spécialisés pour les comprendre et les traduire. Les Data-Lakes n'ont pas de structure et sont donc faciles à consulter et modifier. Les modifications à apporter aux données peuvent être faites très rapidement.

-Le Data-Warehouse qui stocke uniquement des données transformées, permettent de réduire l'espace de stockage et donc de réduire le coût. Par ailleurs, les données transformées peuvent être facilement comprises par un éventail beaucoup plus large d'utilisateurs. Les Data-Warehouses sont plus structurés. Le traitement et la structure des données les rendent plus faciles à comprendre mais les restrictions imposées à leur structure rendent les Data-Warehouses difficiles à manipuler.

## Liens Sources

### Data-Warehouse

<https://lejournal.cnrs.fr/articles/le-difficile-stockage-des-masses-de-donnees>

[https://fr.wikipedia.org/wiki/Entrep%C3%B4t\\_de\\_donn%C3%A9es#D%C3%A9finition\\_et\\_construction](https://fr.wikipedia.org/wiki/Entrep%C3%B4t_de_donn%C3%A9es#D%C3%A9finition_et_construction)

<https://www.lebigdata.fr/data-warehouse-entrepot-donnees-definition>

### Data-lake

<https://www.lebigdata.fr/data-lake-definition>

### Conclusion

<https://fr.talend.com/resources/data-lake-vs-data-warehouse/>

From:

<https://wiki.sio.bts/> - **WIKI SIO : DEPUIS 2017**

Permanent link:

<https://wiki.sio.bts/doku.php?id=1b>

Last update: **2020/07/26 16:27**

