



Cataloguer son modern data stack et en faire profiter ses équipes de Data Science

DRESSE LAURENT - laurent.dresse@datagalaxy.com

- DataGalaxy - 69003 Lyon

Thématique : Structuration des Données

Résumé : *Au sein de la complexité des systèmes transactionnels et de stockage, comment un data catalogue va vous permettre d'inventorier, classer et exposer vos actifs de la donnée et aider les équipes de data science à identifier les bons jeux de données et capitaliser ou documenter les données.*

Mots clés : *Dictionnaire de données, Data Catalogue, Glossaire Métier, Data management, Data Gouvernance, Data Management*

1. Introduction

De plus en plus d'organisations et d'entreprises sont confrontées à un nombre croissant du volume des données. Il devient difficile de garder une inventaire exhaustif des actifs de données et surtout de les exposer de manière cohérente à vos équipes de data science; Quelle colonne de quel système utiliser dans mon algorithme? Est-ce que cet élément de donnée correspond bien à la description faite par les métiers?

Nous verrons comment réconcilier ces deux mondes et capitaliser sur la connaissance de chacun à travers un Data catalogue.



2. Méthodologie

La méthodologie appliqué est d'inventorier de manière automatisée dans une solution de Data Catalogue les éléments suivants:

- Environnements Data Lake
- La Data Pipeline
- Les Data Wharehouses
- La BI/DataViz

Ensuite de restituer l'information au sein d'une plateforme collaborative ou directement dans l'environnement digital/web de l'utilisateur.

3. Originalité / perspective

L'approche permet la création d'une Data Knowledge Workplace où tous les acteurs data de l'entreprise pourront soit obtenir, soit documenter l'utilisation que est faite de la donnée.

De plus, cette démarche s'inscrit clairement dans la perspective d'une entreprise centrée sur la donnée.

Références

Profil LinkedIn: <https://www.linkedin.com/in/ldresse/>

Biographie (En Anglais): voir PDF joint

Interventions:

https://www.youtube.com/watch?v=a_2ilY2pfRE&t=24s

<https://www.youtube.com/watch?v=tloDcWD3e6Q&t=367s>



DataGalaxy