



datalex



Homework Lab

DP-203 Microsoft Azure Data Engineering



Alexandre BERGERE

Head of Data & AI Engineer & Partners at
DataGalaxy, Data Architect independent 🐦
Delta & openLineage lover

Data Expert

(Audit, Modeling, Urbanization, BI's project –
Azure, Big Data, Spark, openLineage & Delta
lover)

Trainer

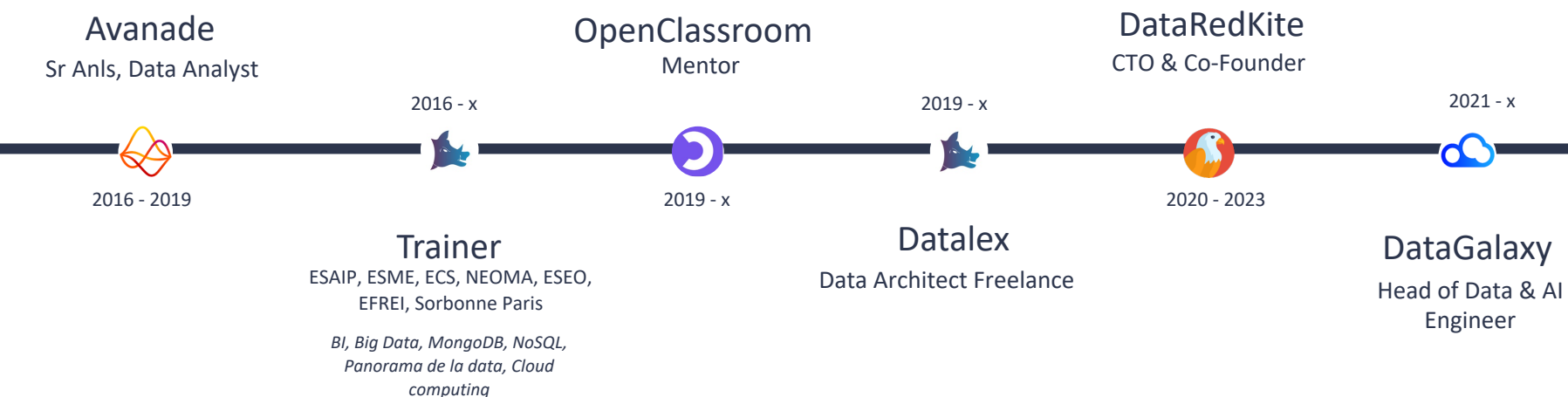
(BI, Big Data, MongoDB, NoSQL, SQL, Cloud
Computing)

Speaker

(Pass Summit 2018, Spark AI Tour 2020, ...)

Cloud & data tech articles

(medium, LinkedIn, slideshare)



Requirements

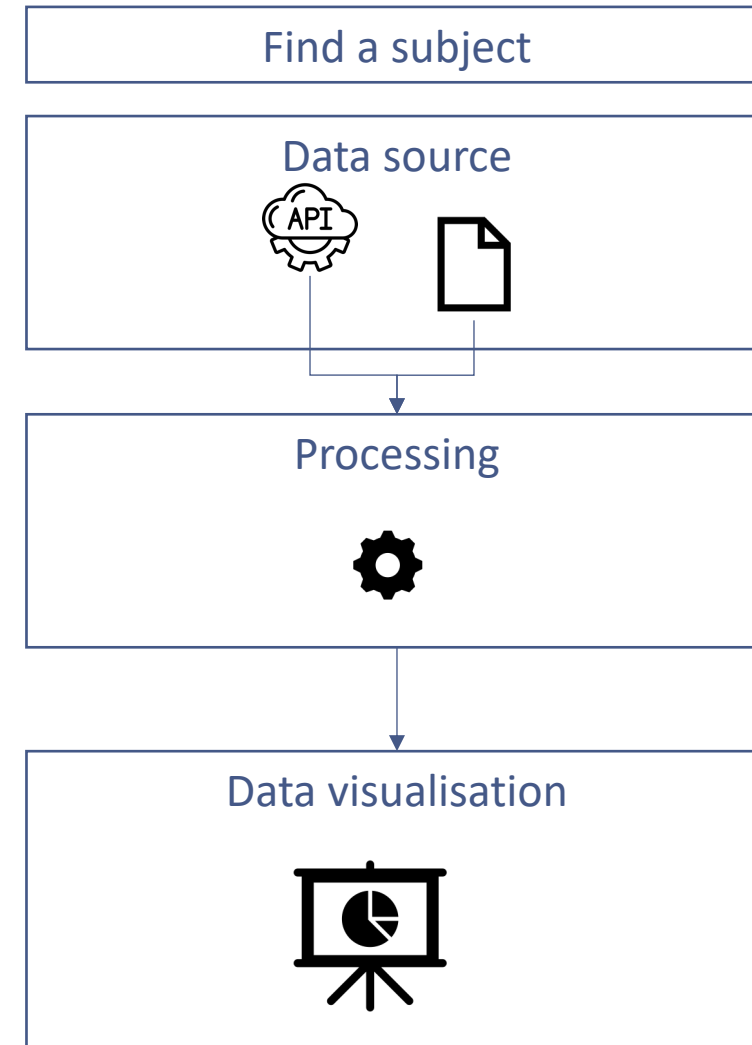
- Azure fundamentals
 - [Microsoft : Course AZ-900T00-A: Microsoft Azure Fundamentals](#)
 - [Summary Course AZ-900](#)
- Understand Azure Synapse Analytics, SQL Pool, Spark, Azure Data Factory
- SQL
- Python



Labs

Par groupe de 2:

- Trouver un sujet et les données associées
- Minimum 3 sources différentes
 - si vous avez la source parfaite avec une seule donnée, faite un split de votre fichier en amont en plusieurs fichiers pour les récupérer plus tard
- Mise en place d'une zone de staging + une zone datawarehouse
- L'architecture peut être soit de type Lakehouse, soit de type DataWarehouse



Labs

- Attendu:
 - Un rapport au format PDF qui explique:
 - Le projet, les données, l'objectif
 - Les traitements effectués (code, capture des traitements ...)
 - Le MCD (Modèle Conceptuels des Données = schéma) de votre DataWarehouse
 - Les scripts utilisés pour toutes transformation (repository Github ou GitLab)
 - Un rapport PowerBI ou Apache Superset

