

Data Analytics

Infos pratiques

- > ECTS : 1,5
- > Nombre d'heures : 18.0
- > Niveau d'étude : BAC +5
- > Période de l'année : Enseignement neuvième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral et Travaux dirigés
- > Composante : Sciences économiques, gestion, mathématiques et informatique

Objectifs

Un Data Lake est un repository de stockage contenant une quantité massive de données originelles stockées dans leur format natif, dans un but d'exploitation ultérieure. Un Data Lake a une architecture plate pour le stockage de données rendant les données accessibles, flexibles et scalables. L'objectif du cours est de motiver le recours aux lacs de données en exposant le passage de la business intelligence vers le business analytics (exploitation directe des données), l'objectif étant de comprendre la différence entre un datawarehouse et un data lake tout en exposant la méthodologie, la stratégie et l'implémentation.

Approche pédagogique et plan de cours.

- Les lacs de données : introduction et définition
- Les lacs de données : méthodologie
- Data lake vs Data warehouse
- Data lake: stratégie et implémentation
- Acquisition de données en batch ou en streaming

Évaluation

Session 1 : Contrôle Continu Intégral (cf. règle par défaut de la section « Modalités spécifiques » des M3C spécifiques)

Session 2 : Règle par défaut décrite dans la section « Modalités de contrôle et examens / Modalités spécifiques »

Pré-requis nécessaires

- Les concepts de BI datawarehouse
- Les ETL

Compétences visées

- Comprendre le concept de lac de données
- Maîtriser la méthodologie de la mise en place d'un lac de données
- Positionner les lacs de données par rapport aux entrepôts de données
- Un tour d'horizon des outils

Bibliographie

- [Saurabh Gupta, Venkata Giri](#): Practical Enterprise Data Lake Insights: Handle Data-Driven Challenges in an Enterprise Big Data Lake, Apress, 2018
- [Alex Gorelik, The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science](#), O'Reilly Media , 2019
- [Tomcy John et Pankaj Misra : Data Lake for Enterprises: Lambda Architecture for building enterprise data systems](#), Packt Publishing Limited, 2017