

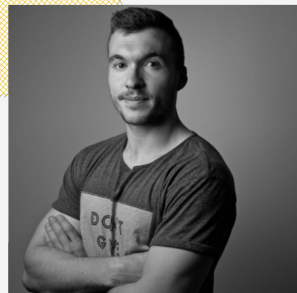
DATA GURU

EXPERTISE EN DATA



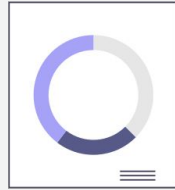
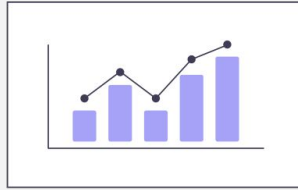
NOTRE HISTOIRE

DataGuru est le fruit de l'association de deux passionnés. Soucieux des bonnes pratiques de développements, ils ont accompagné plusieurs grands comptes sur la conception de plateformes de données et sur l'utilisation des méthodes de collaboration en agile.



NOTRE VOLONTÉ

La volonté de **DataGuru** est le partage des connaissances et l'expertise sur les sujets data. Équiper les futurs Data Engineer de demain avec des connaissances solides en programmation et en Big Data.





RÔLE DU DATA ENGINEER

Les **Data Engineer** travaillent dans divers contextes pour construire des systèmes qui collectent, transforment les données brutes en informations exploitables que les Data Scientists et les Data Analysts peuvent interpréter. Leur objectif ultime est de rendre les données accessibles afin que les organisations puissent les utiliser pour évaluer et optimiser leurs performances.

Voici quelques tâches courantes que vous pouvez effectuer lorsque vous travaillez avec des données :

- Acquérir des ensembles de données qui correspondent aux besoins de l'entreprise.
- Développer des algorithmes pour transformer les données en informations utiles et exploitables.
- Construire, tester et maintenir des architectures de pipeline de base de données.
- Collaborer avec la direction pour comprendre les objectifs de l'entreprise
- Créer de nouvelles méthodes de validation des données et des outils d'analyse des données
- Assurer la conformité avec les politiques de gouvernance et de sécurité des données



CARRIÈRE DANS LE DATA ENGINEERING ?

Une carrière dans ce domaine est à la fois gratifiante et stimulante. Vous jouerez un rôle important dans la réussite d'une organisation, en facilitant l'accès aux données pour les **DataScientists**, les analystes et les décideurs. Vous ferez appel à vos compétences en programmation et en résolution de problèmes pour créer des solutions évolutives.

Tant qu'il y aura des données à traiter, les Data Engineer seront en demande. En fait, Dice Insights a indiqué en 2019 que l'ingénierie des données est un des emplois les plus tendance dans l'industrie technologique, devançant les informaticiens, les concepteurs web et les architectes de bases de données [\[1\]](#). LinkedIn l'a répertorié comme l'un de ses emplois en hausse en 2021 [\[2\]](#).

44.515 €

Le salaire moyen d'un Data Engineer en France en 2021 selon d'après Glassdoor



NOTRE PÉDAGOGIE

Pas de profil Atypique

Que vous soyez manager, entrepreneur, salarié, demandeur d'emploi, freelance, nos portes accueillent tous types de profils, et c'est ce qui nous rend fiers ! Reconversion, création d'entreprise montée en compétences, vos objectifs peuvent être divers : c'est notre mission de prendre en compte votre niveau de compétences initial pour réaliser ces objectifs.

MasterClass

Vous recevrez 2 à 3 Masterclass par semaine, la durée d'une masterClass est de 2H. Les objectifs et les formats pédagogiques de chaque Masterclass changent en fonction du sprint abordé : corrections de use cases concrets, cours magistraux sur des thématiques spéciales, travaux-pratiques, ateliers d'architectures...

Learning By Doing

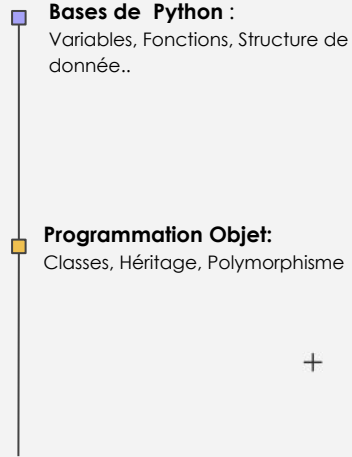
En dehors des masterClass vous allez avoir des exercices à faire pour appliquer la théorie vue en cours. Ainsi que des projets plus ou moins long

Une formation 100% en distanciel

30 % du temps, vous êtes accompagnés par vos professeurs et votre groupe pour des sessions de visioconférence. Pour les **70 %** restants, vous travaillez sur notre plateforme d'enseignement ou nous suivons votre progression et vous accompagnons pour assurer la bonne complétion de votre formation. Ce format vous donne la flexibilité nécessaire pour vous organiser à votre manière.

MODULE PROGRAMMATION

01 PYTHON



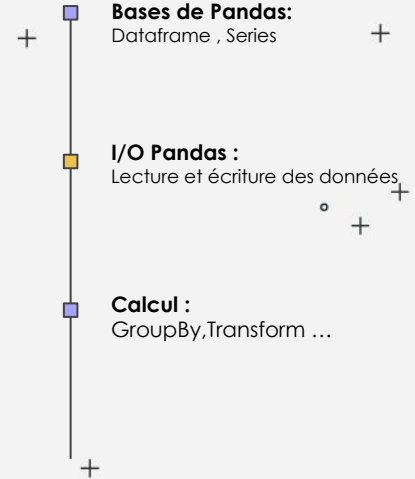
02 LINUX



03 GIT



04 Pandas



MODULE CRAFT



01 CLEAN CODE

- Code Propre :**
Maintenabilité, Intention
- SOLID :**
Les 5 principes
- Architecture Code :**
Hexagonale

02 TESTS

- Bases des Tests :**
Types, Couverture
- Stratégie de tests :**
Test first, Early Test
- TDD :**
Définition et exemple(s)

03 AGILE

- Introduction:**
Emergence de l'agile
- Framework:**
Scrum

MODULE BIG DATA

- **Introduction:**
SGBD, Concepts
 - **SQL :**
SELECT, JOIN , GROUPBY
 - **Transactions:**
Définition et exemples
- 01 RELATIONNEL**

- **Introduction:**
4 types
 - **Garanties:**
CAP
 - **Moteur de recherche:**
Elasticsearch
- 02 NOSQL**

- **Introduction:**
Scalabilité
 - **Réplication:**
Types et exemples
 - **Sharding:**
Types et exemples
- 03 BASES DISTRIBUÉES**

MODULE BATCH

01 HADOOP

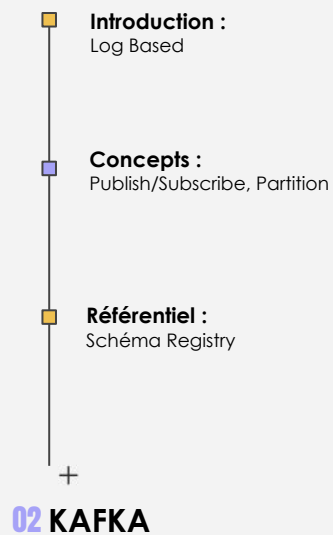
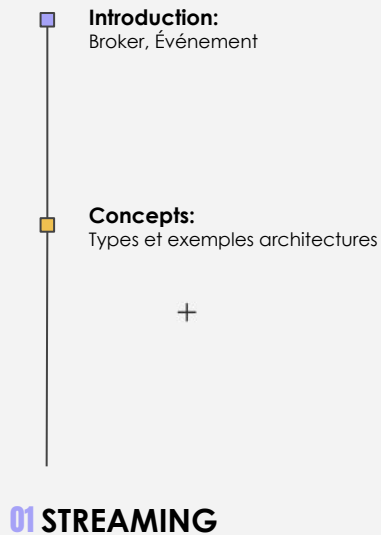
- Ecosystème:**
Distributions
- HDFS:**
Définition
- Calcul Distribué :**
Map Reduce

02 SPARK

- Introduction:**
Data Flow Engines
- Architecture Spark:**
Driver, Worker
- STRUCTURES :**
RDD , Dataframes

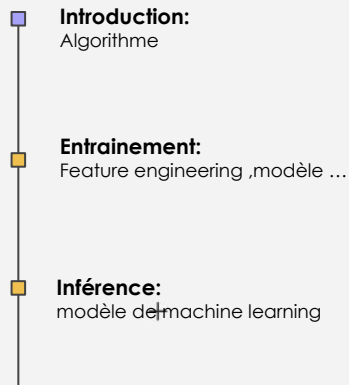


MODULE STREAMING





MODULE DATA SCIENCE

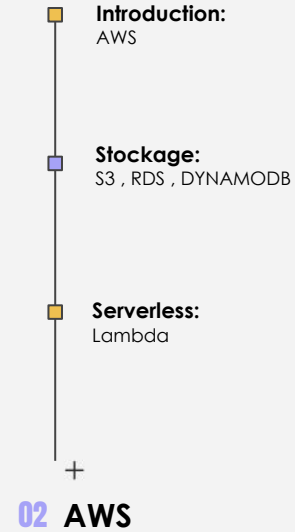
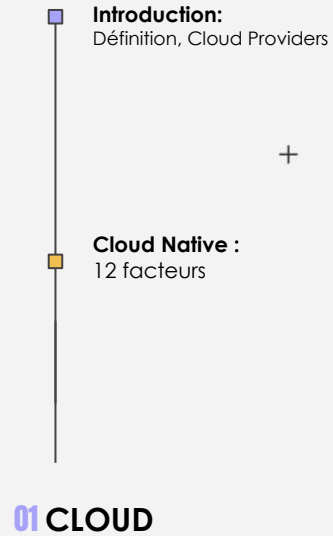


01 DATA SCIENCE



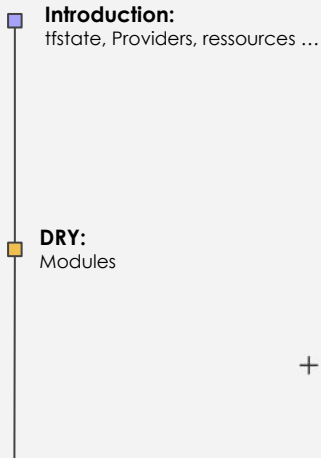
02 INDUSTRIALISATION

MODULE CLOUD

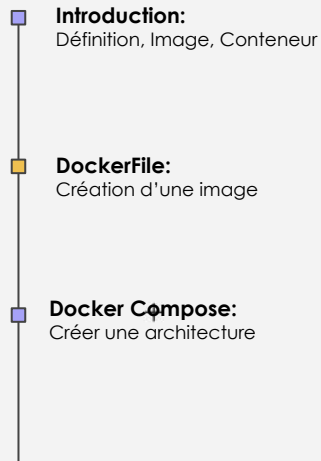


MODULE OPS

01 TERRAFORM



02 DOCKER



03 CI/CD



04 Ansible

