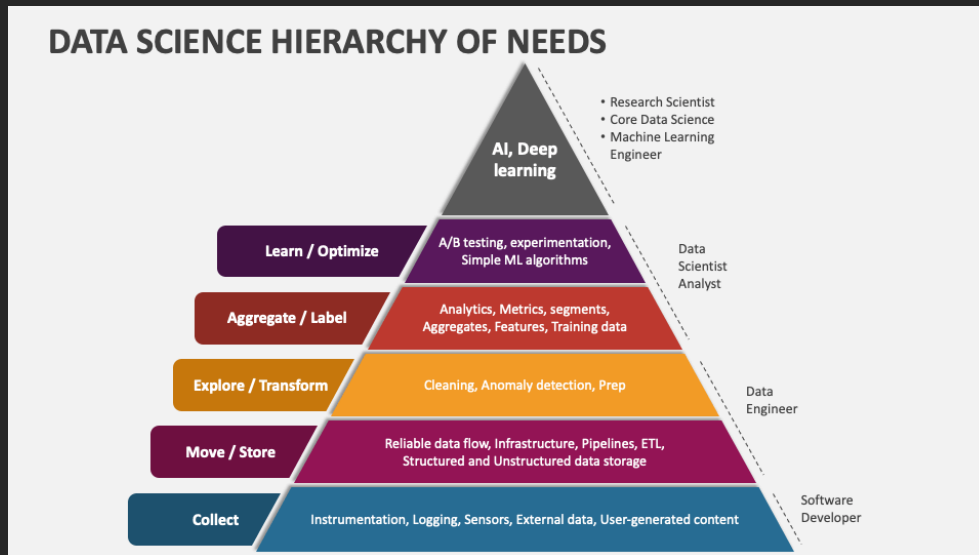


Dans le domaine des données, trois rôles clés se distinguent : **Data Scientist**, **Data Analyst**, et **Data Engineer**. Bien que chacun ait des responsabilités et des compétences spécifiques, ils partagent également des points communs et collaborent souvent pour transformer les données brutes en informations exploitables. Cette complémentarité permet aux entreprises de tirer pleinement parti de leurs ressources de données, de l'infrastructure à l'analyse avancée en passant par l'interprétation des résultats.



Un **Data Engineer** conçoit, construit et maintient l'infrastructure et les systèmes qui permettent de **collecter**, **stocker** et **accéder efficacement aux données**. Parmi ses missions, nous pouvons citer :

- La construction de pipelines de données
- Le développement et la maintenance d'architectures de données
- L'assurance de la qualité des données
- L'optimisation du stockage et de la récupération des données

Les compétences d'un **data Engineer** :

- Compétences en programmation (**Python**, **Scala**, **Java**)
- Connaissance des bases de données (**SQL** and **NoSQL**)
- Outils de big data (**Hadoop**, **Spark**)
- Le processus **ETL** (Extract, Transform, Load)
- Connaissance de plateformes cloud (**AWS**, **Azure**, **GCP**)

Un **Data Analyst** se concentre principalement sur l'**interprétation des données existantes pour fournir des connaissances exploitables**. Il travaille avec des données structurées pour générer des rapports et des tableaux de bord. Parmi ses missions, nous pouvons citer :

- Le nettoyage et l'organisation des données
- La génération de rapports
- La création de visualisations
- La réalisation d'analyses statistiques de base

Les compétences d'un **data Analyst** :

- Compétences de base en programmation (**Python, R**)
- Base de données et feuille de calcul : (**SQL, MS Excel**)
- Outil de visualisation (**Power BI, Tableau**)

Un **data Scientist** se concentre sur la création de modèles et d'algorithmes pour extraire des connaissances à partir des données, et en utilisant des techniques d'apprentissage automatique et statistiques. Parmi ses missions, nous pouvons citer :

- Le développement de modèles prédictifs
- La création de prototypes

Les compétences d'un **data Scientist** :

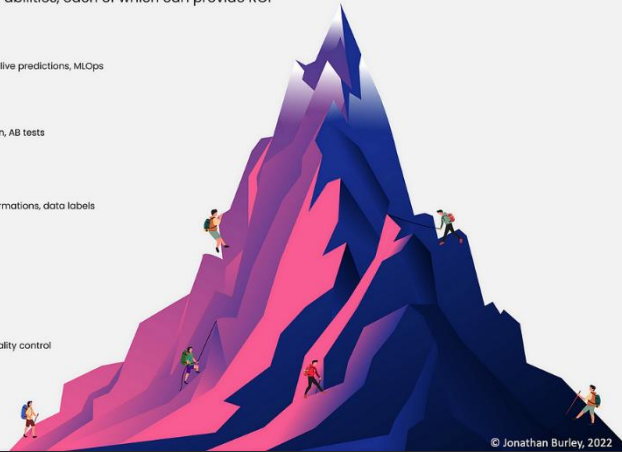
- Compétences en programmation (Python, R)
- Connaissances en statistiques et expertise dans le domaine
- Apprentissage automatique
- Visualisation de données

En somme, un data Engineer peut construire un pipeline de données pour collecter et traiter des données provenant de diverses sources. Un data Analyst utilise ensuite ces données pour générer des rapports et des connaissances. Un data Scientist pourrait aller plus loin en créant des modèles prédictifs pour prévoir les tendances futures basées sur les données.

the data science **hierarchy of needs**

Data science products are built on sequential abilities, each of which can provide ROI

- 6 **AI & continuous learning**
Prescriptive optimization, automated learning systems, live predictions, MLOps
- 5 **predict & prove**
Predictive models, experimentation, historical simulation, AB tests
- 4 **learn & label**
Diagnostic analytics, derived features, reusable transformations, data labels
- 3 **explore & exploit**
Descriptive analytics, anomaly detection, KPIs
- 2 **move & store**
Data storage (structured & unstructured), pipelines, quality control
- 1 **collect**
Raw data: internal, external, instrumentation, etc.



© Jonathan Burley, 2022