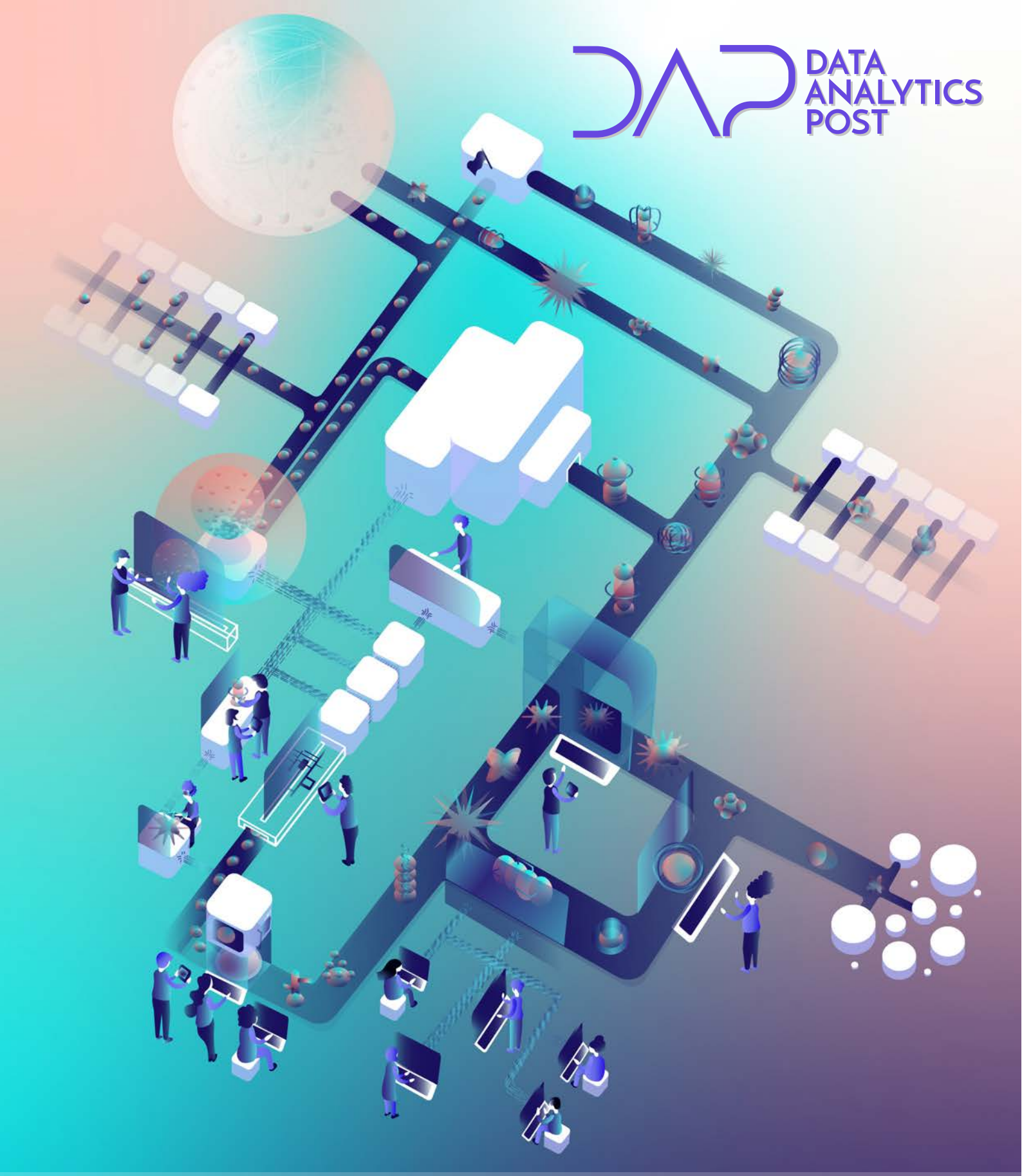
DATA ANALYTICS POST



# GUIDE DES MÉTIERS DE LA DATA

Édition 2019

**ISABELLE BELLIN**
Rédactrice en chef

# Qui fait quoi ?

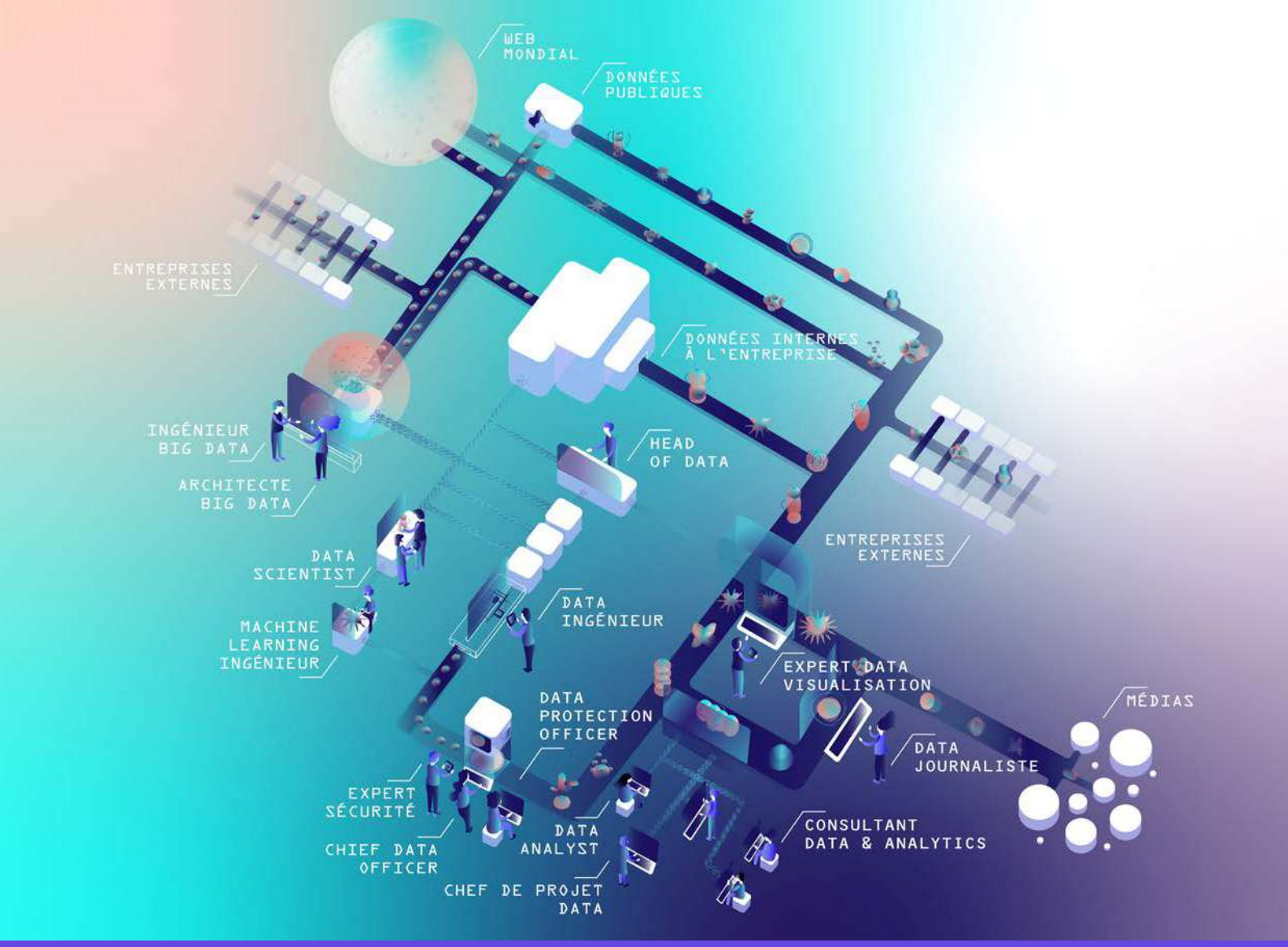
**Découvrez le métier d'un acteur de la data, ses missions dans le long cheminement de la valorisation des données, son travail au quotidien, les compétences et le type de formation requises.**

Les contours des principaux métiers de la data commencent à se préciser de plus en plus clairement, et les entreprises semblent enfin s'accorder sur les intitulés de postes. En dépit du flou qui reste de mise autour de nombre de ces métiers, cela nous a semblé opportun de réunir et publier des informations autour des principaux profils data. De quoi aider tant les entreprises que les candidats à s'y retrouver. Deux ouvrages peuvent également vous être utiles autour de ces sujets : Le Guide des entreprises qui recrutent en data, d'Upward data – ce cabinet de recrutement spécialisé data nous a aidé à préparer ces articles ; le Guide de Telecom ParisTech : Data scientist, les métiers qui façonnent les transitions vers demain, dont une nouvelle édition devrait être en ligne en mars prochain.

Comme vous le verrez, certains acteurs de la data ont plutôt des rôles d'animateur à l'échelle de l'entreprise comme le Chief data officer ou le Head of data, d'autres ont des rôles très techniques comme l'architecte big data, l'ingénieur big data ou encore le data ingénieur – deux profils particulièrement recherchés –, d'autres encore sont très orientés business, comme le chef de projet data ou le data analyst, ou encore très mathématique ou informatique, comme le data scientist ou le machine learning ingénieur, une fonction de développeur apparue très récemment pour industrialiser les algorithmes de machine learning développés par le data scientist. Nous verrons également que de nouvelles disciplines creusent leur sillon, notamment la data visualisation, qui compte désormais ses propres experts, y compris pour les médias, avec les data journalistes.

Ces intitulés de poste ne sont toutefois pas figés. Ils ont toutes les chances d'évoluer pendant encore quelques années. Rappelons que le terme big data est l'étendard de la transition numérique depuis environ cinq ans seulement, et que la plupart de ces métiers ont été inventés ou réinventés ces dernières années. Les missions des uns et des autres ne sont d'ailleurs pas forcément simples à décrire, car elles restent variables selon l'orientation plus ou moins technique, business, mathématique ou informatique du candidat. Elles varient également selon la maturité des entreprises en termes de culture de la donnée, mais aussi de type de données à valoriser.

Quoi qu'il en soit, le marché de l'emploi autour de la data est dynamique, et constitue de belles perspectives d'avenir pour de nombreux étudiants en mathématiques, en sciences, en écoles d'ingénieurs, d'informatique ou même de commerce. Les passerelles entre les métiers ou vers ces métiers via une reconversion existent aussi. Le panorama des possibilités est particulièrement large tant la valorisation des données est au cœur de l'activité des entreprises de toute taille et de tout domaine d'activité, tout comme elle est à la croisée de nombreuses thématiques de recherche tant dans le public que dans le privé.




# SOMMAIRE


## GOUVERNANCE DES DONNÉES
Chief Data Officer p.4
Expert Sécurité p.5
Data Protection Officer p.6

## LES MÉTIERS BIG DATA
Architecte Big Data p.7
Ingénieur Big Data p.8

## HEAD OF DATA
Head of Data p.9

## DATA CONSULTANT
Chef de Projet Data p.10
Consultant Data & Analytics p.11

## DATA ANALYST
Data Analyst p.12

## DATA VISUALISATION
Expert Data Visualisation p.13
Data Jounaliste p.14

## DATA SCIENTIST
Data Scientist p.15

## DATA INGÉNIEUR
Data Ingénieur p.16

## MACHINE LEARNING INGÉNIEUR
Machine Learning Ingénieur p.17

## INFOGRAPHIE
La donnée dans tous ses états p.18

GOUVERNANCE DES DONNÉES

# CHIEF DATA OFFICER

C'est le « **directeur des données** » de l'entreprise. Ce cadre dirigeant participe au pilotage de la stratégie globale de l'entreprise en plaçant les données au cœur de la stratégie et de la culture de l'entreprise, pour que chaque collaborateur devienne « *data fluent* ».

Ce poste, présent principalement dans les grandes structures manipulant une grande variété de données, prend une importance croissante. Le Chief Data Officer (CDO) a un rôle stratégique : il met en place les moyens et les équipes – notamment les datalabs – pour que l'entreprise développe l'usage des données de tous ses métiers et services dans le respect des principes de sécurité et de protection des données. Il s'assure de la collecte des données et de la transmission des informations les plus pertinentes, ainsi que de la gouvernance des données (contrôle, sécurité, niveau de confidentialité, définition de leurs propriétaires). Proche du comité exécutif comme des différentes directions métiers, le CDO insuffle la culture des données à l'entreprise. Il assure le partage des données entre les différentes équipes, qui n'ont pas toujours conscience de la disponibilité de telles informations. Il est parfois rattaché au Chief Digital Officer, dont la mission plus globale est de piloter la stratégie de transformation numérique de l'entreprise.

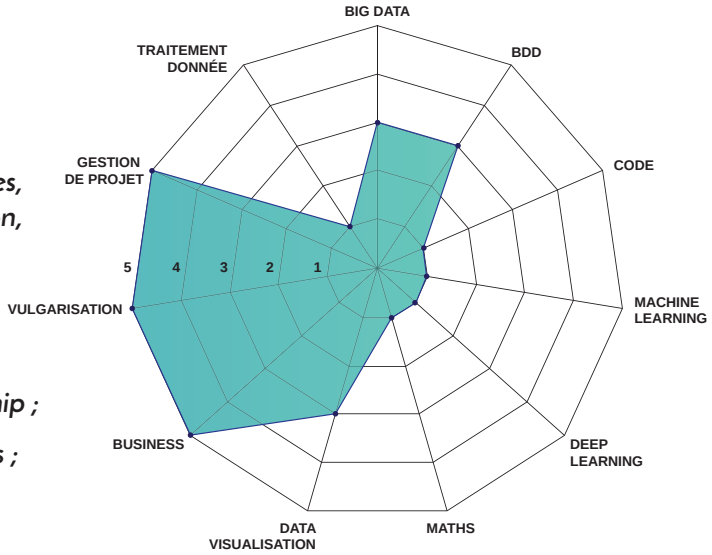
## LE PILOTE STRATÉGIQUE

### Au quotidien :

- *Veiller au partage des données entre les équipes ;*
- *Définir avec les métiers les actions sur les données, leur cohérence et mise en qualité, leur protection, la promotion de leur usage.*

### Compétences :

- *Connaissance du secteur d'activité de l'entreprise ;*
- *Communication, pédagogie, diplomatie, leadership ;*
- *Connaissance des réglementations internationales ;*
- *Communication, bon relationnel, vulgarisation.*




### Formation :

- *Ecole d'ingénieur ou de commerce puis évolution de carrière ;*
- *Bac + 4 ou 5 en informatique, management, statistiques ou marketing ;*
- *Ce métier étant assez récent, il est également possible d'y parvenir lors d'une évolution de carrière.*

GOUVERNANCE DES DONNÉES

# EXPERT SECURITÉ

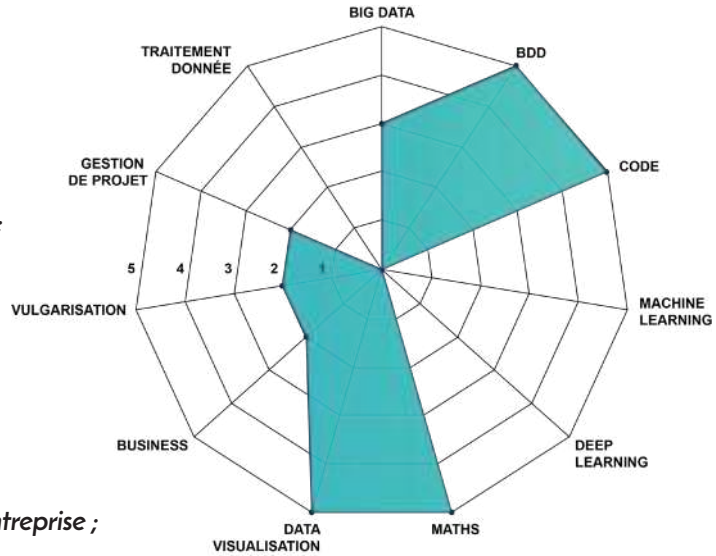
Ce spécialiste de la protection des données assure la politique de sécurité de l'entreprise. Il intervient principalement dans les couches informatiques et télécoms de l'entreprise. Il doit anticiper les risques augmentés ou créés par l'utilisation des big data.

Dans un contexte global où les cyberattaques se déploient à grande échelle, l'expert sécurité (ou cybersécurité) évalue le niveau de vulnérabilité des systèmes d'information et des systèmes de gestion de données, que ce soit lors de l'acquisition des données, leur transport, leur traitement ou leur stockage. Il prépare les solutions de sécurisation des données, met en échec les tentatives d'intrusion ou de déni d'accès aux systèmes d'information et de gestion de données de l'entreprise. Son rôle est en amont (prévention, analyse de risques, détection, lutte) et en aval (explication et réparation). Il doit également effectuer une veille technologique poussée pour anticiper les futures défaillances ou attaques.

## L'ANTI CYBERATTAQUES

### Au quotidien :

- *Analyser les risques ;*
- *Chercher et repérer les failles des systèmes de l'entreprise ;*
- *Rédiger les politiques et les standards de sécurité ;*
- *Administrer les droits d'accès au réseau et aux données.*




### Compétences :

- *Très bonne connaissance informatique (théorie, architecture, protocoles) ;*
- *Connaissance de l'organisation et des besoins de l'entreprise ;*
- *Organisation, rigueur, communication et pédagogie.*

### Formation :

- *Ingénieur ou master en informatique, en télécommunications & réseaux, en systèmes d'information, avec une spécialité ou une dominante sécurité ;*
- *Cryptologue (si possible doctorat dans le domaine).*

GOUVERNANCE DES DONNÉES

# DATA PROTECTION OFFICER

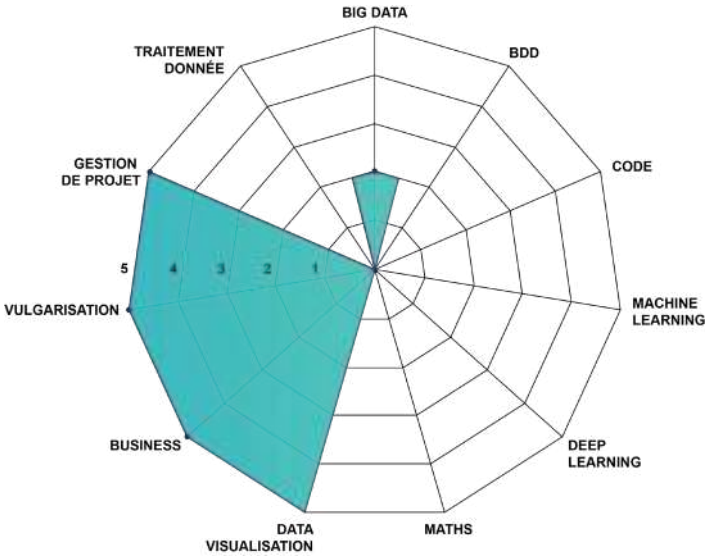
Ce métier récent prend en compte l'évolution des contraintes réglementaires en matière de gouvernance des données personnelles pour y sensibiliser les collaborateurs de l'entreprise.

Le Data Protection Officer (DPO) ou délégué à la protection des données a une mission d'information, de conseil et de contrôle en interne de la gouvernance des données personnelles. C'est un métier au carrefour du droit, de la sécurité informatique, de la conformité et de l'éthique. Depuis la publication, en avril 2016, du Règlement européen sur la protection des données (RGPD ou GDPR en anglais), ce poste est obligatoire en Europe depuis mai 2018 dans les entreprises et administrations qui traitent des données sensibles à grande échelle. Dans ce cadre, le DPO est chargé d'assurer la sécurité et la conformité des données personnelles, de définir les rôles et responsabilités de chacun, d'établir une cartographie des traitements et flux de données. Au quotidien, il dialogue avec les parties prenantes internes et externes afin de les sensibiliser à la culture de la protection des données. En France, ce rôle prend la suite de celui du Correspondant informatique et libertés (CIL). Son défi est de se tenir au courant de tous les projets de l'entreprise lancés autour des données, pour pouvoir y apporter des préconisations suffisamment en amont.

## LE PROTECTEUR DES DONNÉES PERSONNELLES

### Au quotidien :

- *Gérer et améliorer les processus de conformité notamment via des audits ;*
- *Garantir la conformité de l'organisme ;*
- *Sensibiliser les collaborateurs aux implications juridiques de l'utilisation des données ;*
- *Mettre en place des formations, des actions de sensibilisation ;*
- *Assurer une veille juridique.*




### Compétences :

- *Connaissances juridiques (droit NTIC), des règlements et des normes ;*
- *Bonne maîtrise de l'informatique ;*
- *Communication et pédagogie.*

### Formation :

- *Diplôme CIL / DPO (comme à Paris Nanterre) pour juristes ou informaticiens ;*
- *Evolution de carrières après écoles de commerce, d'informatique ou juristes ;*
- *Évolution du Correspondant informatique et libertés.*

LES MÉTIERS DU BIG DATA

# ARCHITECTE BIG DATA

À partir d'un besoin, cet expert du big data conçoit l'infrastructure et les combinaisons technologiques permettant de collecter et de traiter de gros volumes de données de toute nature. Il apporte une vision d'ensemble des technologies big data.

L'architecte big data intervient en amont du traitement de la donnée pour organiser la récupération et la gestion des données brutes en lien avec les équipes informatiques et les managers de la donnée. C'est l'acteur le plus expérimenté pour conseiller, planifier et mettre en place l'infrastructure de gestion des données. Grâce à sa vision d'ensemble des technologies big data, il s'assure de la cohérence de la structure des bases de données et des logiciels pour qu'ils soient toujours en phase avec les besoins et les enjeux de l'entreprise. Il crée l'architecture du système de gestion de données de telle sorte que l'acquisition, la manipulation, la circulation et la restitution des données soient facilitées : il prévoit comment et quand les différentes sources de données et les différents outils d'analyse seront mis en ligne et comment ils s'intégreront. Il affine et surveille en permanence cette infrastructure. C'est également un interlocuteur important du data scientist, à qui il fournit les données brutes à traiter.

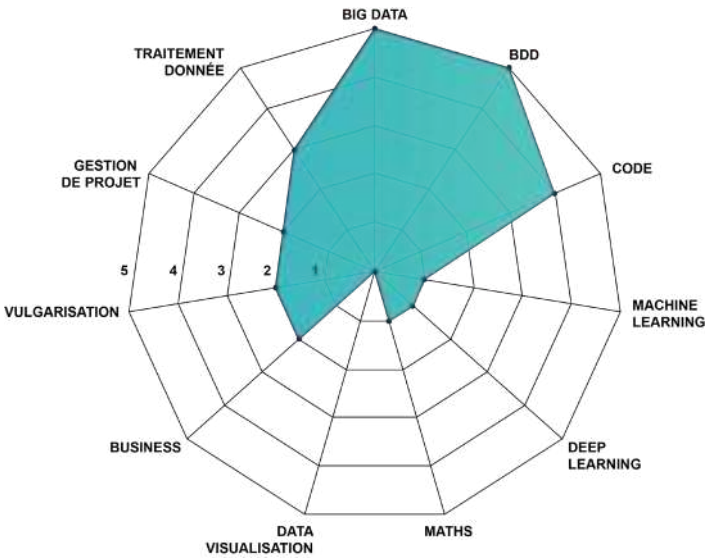
## LE VISIONNAIRE TECHNIQUE

### Au quotidien :

- *Concevoir l'architecture de l'infrastructure de gestion des données ;*
- *Garantir son fonctionnement ;*
- *La faire évoluer selon les besoins.*

### Compétences :

- *Maîtrise de toutes les technologies big data : bases de données NoSQL (MongoDB, Cassandra ou Redis), infrastructures serveurs (Hadoop, Spark) et stockage de données en mémoire (Memtables) ;*
- *Esprit d'équipe ;*
- *Curiosité, autonomie, communication.*




### Formation :

- *Bac + 4 ou 5 en informatique, management, statistiques ou marketing ;*
- *Formation big data ;*
- *École d'ingénieur spécialisée ;*
- *Expérience professionnelle Business Intelligence.*

LES MÉTIERS DU BIG DATA

# INGÉNIEUR BIG DATA

Il développe l'infrastructure mise en place par l'**architecte big data**. Il gère sa maintenance au quotidien et la fait évoluer selon les solutions les plus récentes.

Responsable du maintien de la solution big data développée, ce spécialiste logiciel maîtrise les outils de big data et met « les mains dans le cambouis ». Il écrit des logiciels qui tournent dans ces environnements. Il réalise des tests et des évaluations sur la structure pour être sûr qu'elle résiste à la masse de données exploitée par les Data Scientists. Il construit également les solutions techniques qui permettent aux acteurs de la data de les exploiter.

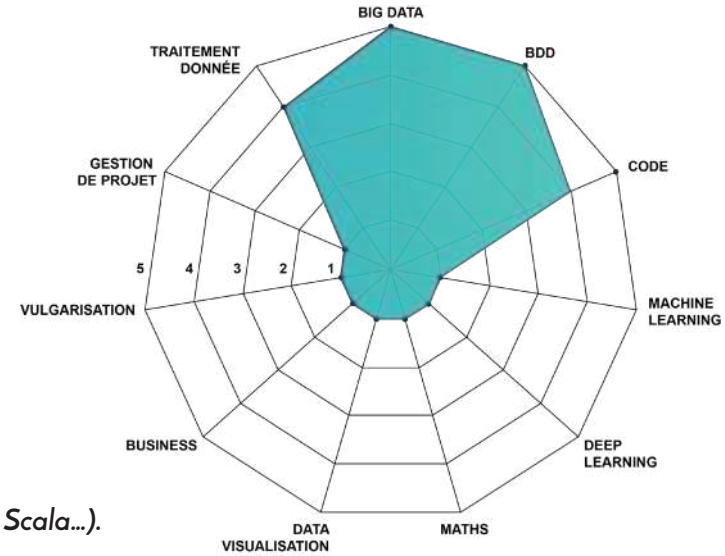
## L'ACTEUR DU BIG DATA

### Au quotidien :

- *Gérer l'architecture big data de l'entreprise ;*
- *Assurer une veille technologique.*

### Compétences :

- *Expertise d'internet et des bases de données (SQL, NoSQL) ;*
- *Maîtrise des technologies big data (Hadoop, Spark, Hive, Storm, Pig) et des infrastructures serveurs ;*
- *Maîtrise des langages de programmation (Java, Scala...).*




### Formation :

- *Bac+5 big data ;*
- *Bac+8 Statistiques ;*
- *Formation big data.*

# HEAD OF DATA

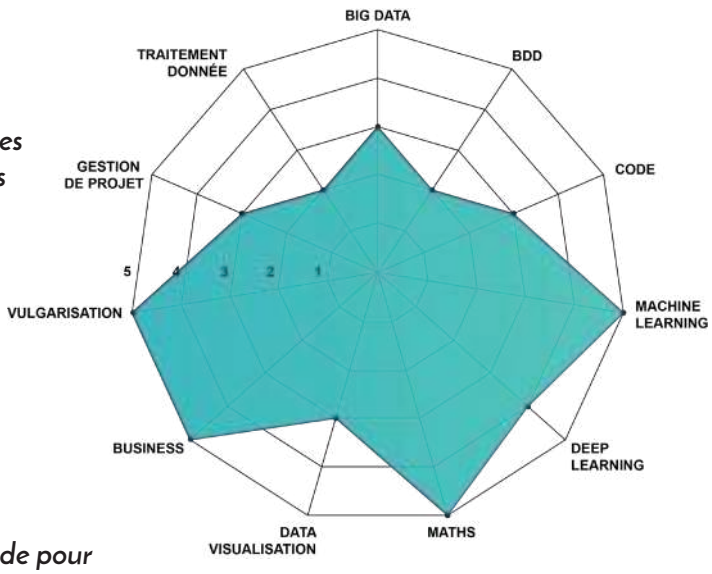
Le *Head of Data* (ou *Lead Data Scientist*) est responsable de l'équipe data science d'une organisation, en charge de créer un datalab et de le diriger. Il explore les valorisations de données de l'entreprise.

Le rôle de ce *data scientist* d'origine est de piloter et mettre en œuvre des projets exploratoires avec l'objectif de créer de la valeur à partir de la donnée. Il doit faire émerger de nouvelles idées, repérer les données pertinentes à collecter ou à traiter pour alimenter les algorithmes d'apprentissage machine, des data visualisations... Selon les cas, il encadre l'équipe de *data scientists* pour identifier ce qu'une meilleure utilisation des données pourrait apporter aux métiers de l'entreprise, en mode exploratoire ; dans certains cas, il doit également avoir une légitimité technique avérée pour animer la communauté scientifique de l'entreprise. Il met en place des processus d'idéation pour identifier des valorisations possibles de données en travaillant en lien avec les directions métiers et leurs équipes de *data scientists*. Il est également chargé de trouver les sponsors en interne pour financer ses travaux exploratoires, organiser des hackathons internes et mettre en place des datalabs.

## LE CRÉATEUR DE VALEUR

### Au quotidien :

- *Encadrer les data scientists et gérer les projets ;*
- *Animer la communauté scientifique ;*
- *Faire émerger de nouvelles idées de valorisation des données avec les spécialistes des différents métiers de l'entreprise ;*
- *Explorer de nouvelles solutions.*




### Compétences :

- *En sciences de données (analyse de données, modélisation, machine learning) ;*
- *Maîtrise de Python, R, Java, Matlab, Spark, écosystème Hadoop ;*
- *Bon relationnel, capacités de vulgarisation, aptitude pour le travail en équipe, rigueur, aptitudes managériales.*

### Formation :

- *Mathématiques, statistiques ou informatique (niveau > bac + 5) selon les enjeux de l'entreprise, de ses problématiques données et de la taille des équipes.*

DATA CONSULTANTS

# CHEF DE PROJET DATA

Il définit les cas d'usage et suit les projets data de l'entreprise grâce à ses connaissances des enjeux business de l'entreprise ou d'un de ses secteurs d'activités, ainsi que des enjeux et des problématiques liés au big data.

Rattaché au responsable de l'équipe données (Head of Data), le chef de projet data déploie la stratégie de l'entreprise à plus ou moins long terme, par exemple dans le cadre du datalab de l'entreprise. Pour cela, il gère certains projets fondés sur l'analyse des données stratégiques de l'entreprise. Il fait la liaison entre les profils des technologies de l'information (IT), les profils métiers, plus techniques et les profils business. Selon les entreprises, le profil recherché sera plus ou moins technique, incluant éventuellement la data visualisation. Dans sa version « *data manager* », le chef de projet data est également garant de la conformité et de la bonne organisation des données dans les systèmes d'information (données référentielles liées aux catalogues fournisseurs, clients, articles de l'entreprise et métadonnées liées aux normes et aux règlements).

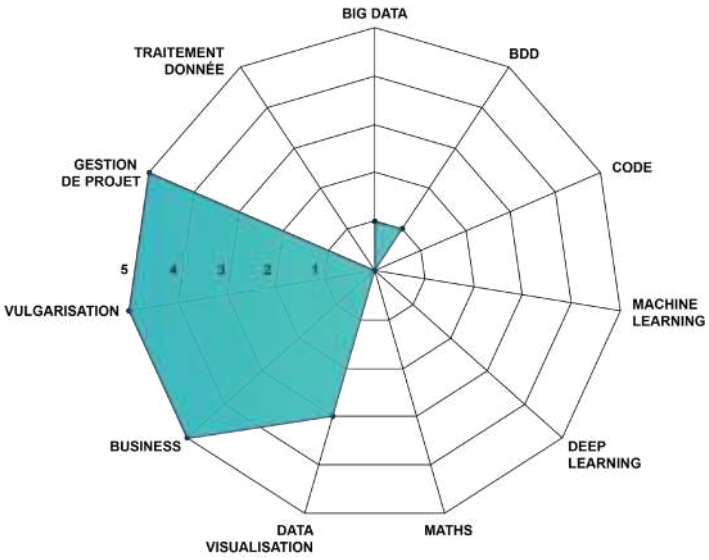
## L'EXPERT BUSINESS DE L'ENTREPRISE

### Au quotidien :

- *Animer les équipes projets ;*
- *Faire le pont entre les équipes techniques et business, en interne comme en externe, éventuellement à l'aide de data visualisation ;*
- *Assurer une veille technologique.*

### Compétences :

- *Très bonne connaissance des enjeux data ;*
- *Maîtrise de Python, R, Java, Matlab, Spark, écosystème Hadoop ;*
- *Très bonne connaissance métier ;*
- *Qualités d'écoute, d'analyse et de synthèse ;*
- *Communication, esprit d'équipe, leadership, autonomie.*




### Formation :

- *Ecole d'ingénieur ou de commerce.*

DATA CONSULTANTS

# CONSULTANT DATA & ANALYTICS

Le consultant data intervient auprès de diverses organisations. Selon les cas, il peut les aider à définir leur stratégie data sur le long terme, mettre en place des business units spécifiques, ou encore développer des outils pour des clients.

Sa vision très large des écosystèmes de la donnée permet au consultant data d'apporter son assistance et son savoir-faire, aussi bien aux équipes de *data scientists* qu'aux entreprises n'ayant pas ces compétences en interne. Il est chargé de préciser à une organisation ce qu'elle peut attendre de ses données (y compris hors de ses domaines les plus courants) et d'apporter une réponse opérationnelle. Tour à tour *data scientist* et expert métier, il est en contact avec les opérationnels métiers et avec l'ensemble des personnes agissant sur les données, des ingénieurs big data aux *chief data officers*. Il a un rôle clé dans la préparation des données en amont et dans le dépouillement en aval.

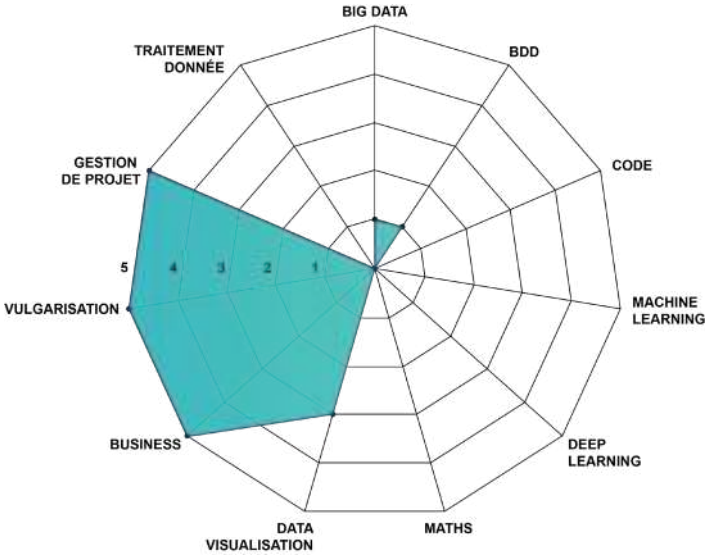
## L'ARTISAN DE LA DONNÉE

### Au quotidien :

- *Animer des comités de direction, des ateliers d'émergence de besoin ;*
- *Découvrir et comprendre un métier ;*
- *Identifier les problèmes sur lesquels la donnée peut ou non aider ;*
- *Manipuler les données, coder, tester de nouvelles approches techniques ;*
- *Assurer une veille technologique.*

### Compétences :

- *Bonne connaissance de l'écosystème data ;*
- *Compréhension des enjeux métier ;*
- *Capacité à structurer des projets (organisation, suivi, rétroplanning) ;*
- *Créativité, curiosité.*




### Formation :

- *École d'ingénieur ou de commerce ;*
- *les data scientists, les data analysts, les ingénieurs big data et les architectes big data peuvent devenir des consultants s'ils ont des qualités de communication, un bon relationnel, des capacités de synthèse et l'appétence pour la diversité de métiers de leurs clients.*

# DATA ANALYST

Le *data analyst* travaille en général sur un type spécifique de données issues d'une source unique et connue, qu'il analyse avec un regard marketing ou finance selon le secteur afin d'orienter les prises de décisions stratégiques de l'entreprise.

Il utilise les techniques statistiques et informatiques afin d'explorer, d'organiser, de synthétiser et de traduire les données brutes de l'entreprise, qu'elles soient issues des réseaux sociaux, du e-commerce, de l'internet des objets ou bien liées à ses clients, à ses produits, à ses propres performances, à son marché ou à sa concurrence. À partir de cette analyse menée à l'aval de la chaîne de traitement de la donnée, en collaboration avec le *data scientist* sur les aspects technico-scientifiques, le but du *data analyst* est de faciliter les prises de décision de l'entreprise lui permettant de prendre un avantage compétitif. Il définit notamment des indicateurs clés de performance (KPI) pour vulgariser et restituer ses résultats aux décideurs sous un format exploitable, notamment au travers de data visualisations.

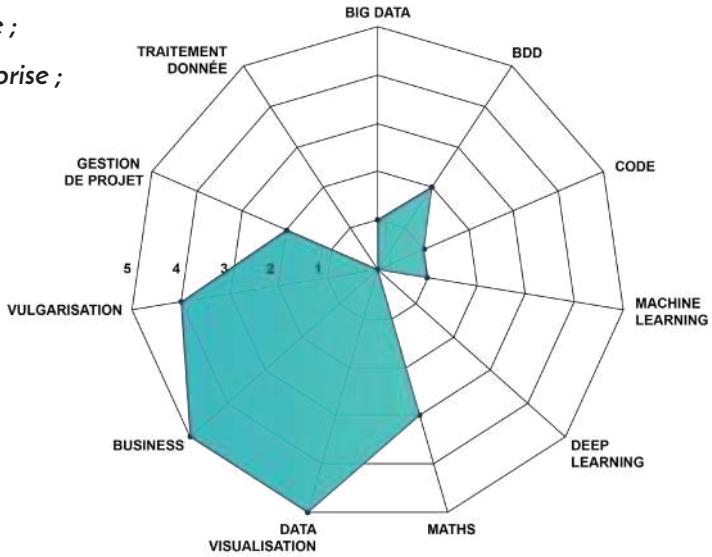
## LE SPÉCIALISTE BUSINESS D'UNE SOURCE DE DONNÉES

### Au quotidien :

- *Analyser les données pour les transformer en informations exploitables ;*
- *Participe à la stratégie Data-Driven de l'entreprise ;*
- *Créer et maintenir les bases de données de l'entreprise ;*
- *Élaborer les critères de segmentation.*

### Compétences :

- *Une certaine expertise en mathématiques et en statistiques ;*
- *Maîtrise des bases de données (SQL) ;*
- *Maîtrise des outils de visualisation (type Tableau, QlikView) ;*
- *Sensibilité aux enjeux business ;*
- *Esprit d'équipe, Communication, curiosité, rigueur, aisance rédactionnelle.*




### Formation :

*Ces profils, type écoles de commerce ou école d'ingénieur, peuvent également être des consultants free-lance.*

- *Bac + 4 ou 5 en statistiques, informatique ou marketing ;*
- *Formation big data.*

DATA VISUALISATION

# EXPERT DATA VISUALISATION

Ce spécialiste data, attaché à une entreprise, est capable d'exploiter les données de l'entreprise et de les mettre dans un contexte simplifié de visualisation pour en explorer le sens et les impacts.

Grâce à un choix judicieux d'organisation spatiale, de liaisons entre les données, de couleurs, de formes, l'expert en data visualisation met en scène des données complexes, les rend intelligibles, plus accessibles dans le but de les présenter à des acteurs sans expertise technique. Ce profil a deux facettes : celui d'un expert d'outils de data visualisation faisant du *reporting* et du *storytelling* sur les données, ou bien celui d'un développeur qui crée des applications de data visualisation que ce soit en intranet, sur le web, sur des applications mobiles ou encore sur papier. Grâce à son travail sur les interfaces, cet expert permet également aux équipes opérationnelles d'y voir plus clair dans leurs données en posant les bonnes questions, et d'identifier de nouvelles pistes d'analyse en explorant les données sous un nouveau jour. Il doit être capable de choisir les visualisations les plus pertinentes et susceptibles d'apporter le moins de biais. Ces profils peuvent évoluer à terme vers le métier de *data scientist* au sens large.

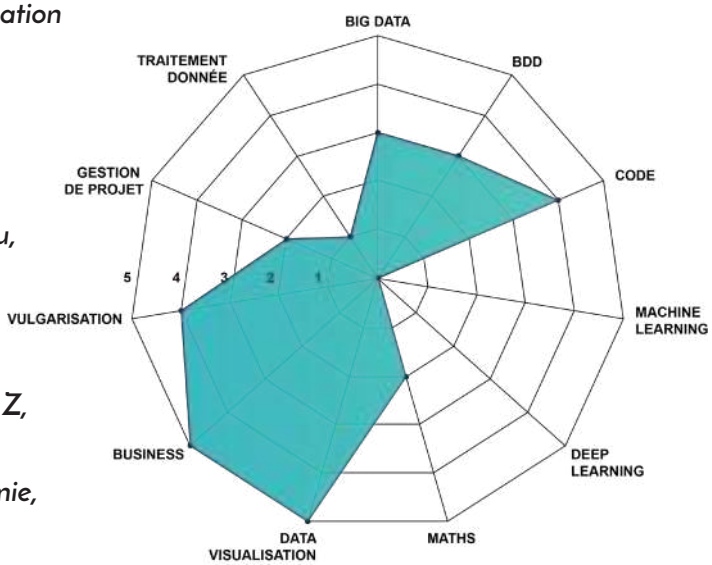
## LE DATA *STORYTELLER* DE L'ENTREPRISE

### Au quotidien :

- *Développer et concevoir des outils de data visualisation selon les enjeux business de l'entreprise ;*
- *Mettre en place des applications pour restituer de larges jeux de données.*

### Compétences :

- *Maîtrise des outils de visualisation comme Tableau, QlikView, Microsoft Power BI, Excel ;*
- *Pour les développeurs web : Javascript, D3.js, Angular.js, Django, html, CSS*
- *Pour ceux qui développent les applications de A à Z, technologies de type Node.js, Java ;*
- *Curiosité, créativité, sensibilité au design, autonomie, aptitude au travail en équipe, rigueur.*




### Formation :

- *Développeurs front end & back end.*

DATA ANALYTICS POST

DATA VISUALISATION

# DATA JOURNALISTE

Le data journaliste ou « journaliste de données » est un journaliste qui ne rédige pas nécessairement d'articles. Il propose une visualisation de l'information avec les mêmes outils graphiques que les experts en data visualisation.

Ce métier apparu il y a une dizaine d'années existe maintenant dans la plupart des rédactions, que ce soit pour faire du *fact-checking* – contrôler la véracité de propos et de chiffres – ou enquêter, trouver des informations nouvelles à partir de masses de données et les présenter de manière compréhensible, accessible et interactive. Pour cela, le data journaliste collecte, vérifie, recoupe les données – chiffres, données brutes, statistiques – les retranscrit et propose une visualisation de l'information sous un nouvel éclairage. Travaillant le plus souvent avec les développeurs web et graphistes de son média, il doit avoir des idées de mise en scène, de visualisation des informations, de storytelling.

## L'INVESTIGATEUR DES DONNÉES

### Au quotidien :

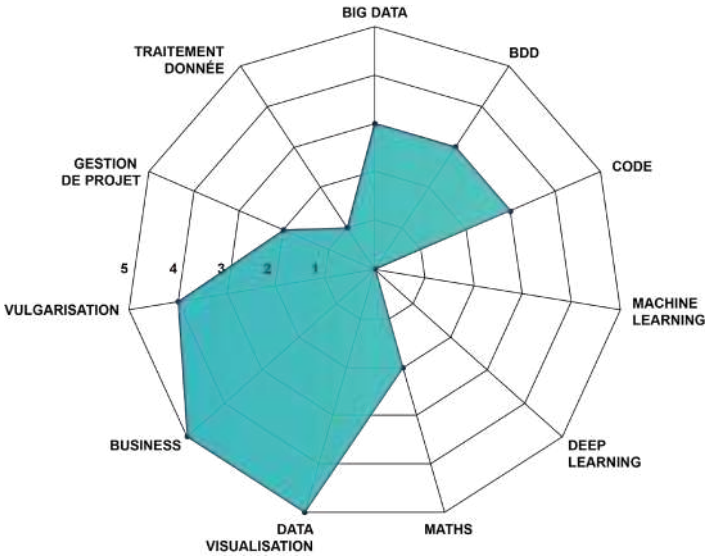
- *Collecter, nettoyer, vérifier, croiser et analyser des données massives ;*
- *Travailler en mode projet avec des développeurs, statisticiens et graphistes.*

### Compétences :

- *Esprit d'analyse, de synthèse, de hiérarchie des informations ;*
- *Goût pour les chiffres ;*
- *Maîtrise des outils numériques et logiciels spécifiques ;*
- *Curiosité, rigueur, grande culture générale, communication ;*
- *Graphisme.*

### Formation :

- *École de journalisme ;*
- *BTS communication.*



# DATA SCIENTIST

Le *data scientist* est expert en mathématiques, statistique et informatique. Ce métier d'explorateur de la donnée relève, selon les profils – plutôt orientés mathématique ou informatique –, de la recherche, du développement d'algorithmes, voire de l'industrialisation de solutions.

Selon les cas, le *data scientist* peut travailler sur des problèmes plus ou moins appliqués ou faire de la recherche pour une entité publique ou privée. Dans le premier cas, avec des compétences plus orientées informatique, il code et développe des algorithmes avancés pour répondre aux enjeux et aux problématiques stratégiques de l'entreprise. En allant à la rencontre des métiers pour en définir les besoins et en collaborant avec les équipes de R&D, il apporte son expertise en techniques et algorithmes d'apprentissage automatique et en systèmes d'intelligence artificielle. Il intervient à toutes les étapes de la chaîne de données : définition du problème, collecte des données, nettoyage, mise en place des modèles prédictifs et création des algorithmes. Il a d'excellentes capacités de communication et sait adapter son discours à chacun.

Concernant le *data scientist* orienté recherche, différents cas de figure existent. Certains chercheurs font progresser les techniques et technologies au cœur des différentes disciplines des data sciences (*machine learning*, apprentissage profond, etc.). D'autres utilisent les techniques de la data science, notamment l'intelligence artificielle, pour faire avancer les recherches dans leur propre discipline. D'autres encore mettent à profit les évolutions de la data science dans leur propre domaine d'activité : *designers*, juristes, sociologues, économistes…

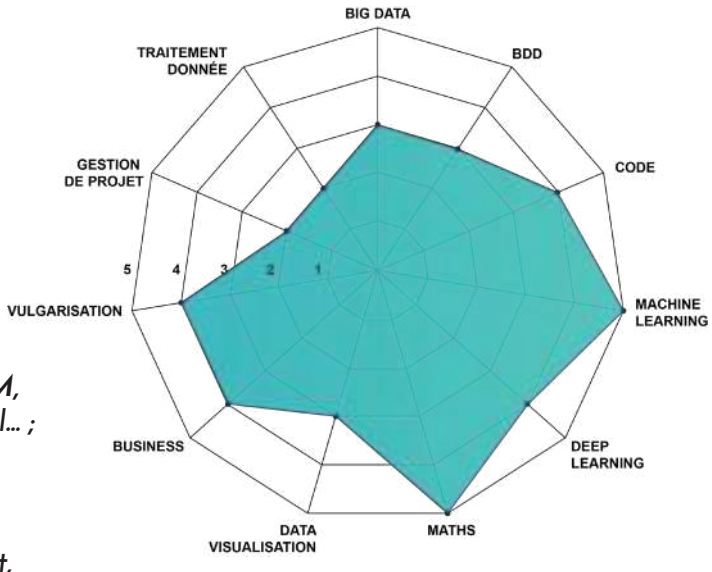
## L'EXPLORATEUR DE DONNÉES COMPLEXES

### Au quotidien :

- *Explorer de nouvelles pistes dans les ensembles de données et tester des hypothèses*
- *Créer de nouveaux modèles de données*
- *Concevoir des algorithmes à base de machine learning et les implémenter*

### Compétences :

- *Maîtrise du machine learning, deep Learning, Random Forests, modèles de Markov cachés, SVM, régression, séries temporelles, traitement du signal… ;*
- *Maîtrise des outils Big data ;*
- *Maîtrise de Python, R , Java, C, C++, Matlab, écosystème Hadoop, Spark… ;*
- *Esprit d'équipe, communication, ouverture d'esprit, créativité, curiosité, transdisciplinarité ;*
- *Sensibilité aux enjeux business, notamment dans les secteurs comme le marketing, le web, la publicité…*




### Formation :

- *Masters ou thèses en informatique, sciences cognitives, statistiques, mathématiques… ;*
- *Bac + 4-5 en mathématiques-statistiques ou informatique ;*
- *Formation big data ;*
- *Expérience professionnelle (Data Analyst).*

# DATA INGÉNIEUR

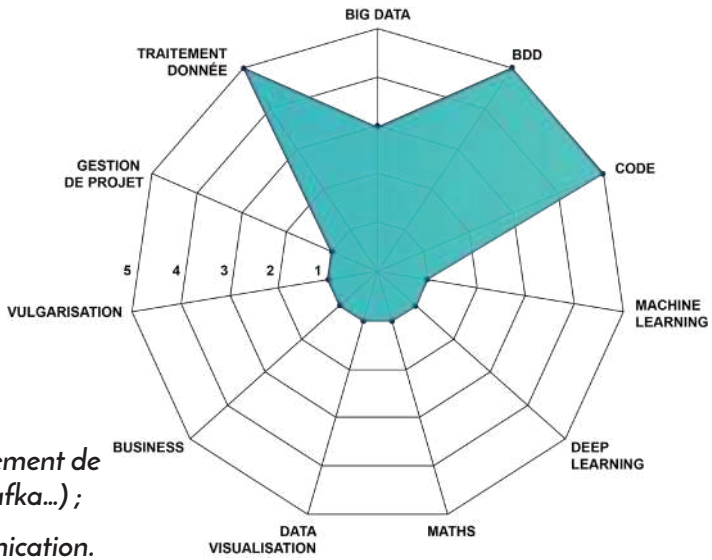
Ce métier d'expert de la gestion des données – de leur ingénierie au sens propre –, apparu très récemment, est vite devenu indispensable. Polyvalent, le data ingénieur un acteur clé des projets.

Ce profil technique est désormais parmi les plus recherchés, dès qu'une entreprise est assez mature pour industrialiser le traitement de ses données. Le data ingénieur rend la donnée exploitable en amont et en aval des projets. En amont, il est capable de comprendre l'environnement numérique du client, de collecter, préparer, nettoyer, normaliser, contextualiser, stocker ses données en s'assurant que les « tuyauteries » nécessaires (au sens des infrastructures) sont adaptées. C'est lui qui met en place l'ETL (Extraction, Transformation, *Loading*), le système de chargement de données depuis les différentes sources d'information de l'entreprise jusqu'à son entrepôt de données. En aval des projets, le data ingénieur accompagne les *data scientists* dans l'industrialisation de leurs modèles, relit leur code, l'optimise, assure l'exploitabilité des solutions, leur mise à l'échelle en suggérant des tests automatisés, en déployant et en ordonnant les traitements.

## L'ACCOMPAGNATEUR TECHNIQUE DE A À Z

### Au quotidien :

- *Collecter des données en explorant le système d'information d'un client et ses différentes sources de données, en alimentant une plateforme d'acquisition de données, etc. ;*
- *Nettoyer ces données en ajustant leur format, en normalisant les bases de données, en gérant les données manquantes ;*
- *Les stocker dans un format homogène.*




### Compétences :

- *Développeur compétent sur le stockage et le traitement de données volumineuses (Hadoop, Scala, Spark, Kafka...) ;*
- *Esprit d'équipe, curiosité, débrouillardise, communication.*

### Formation :

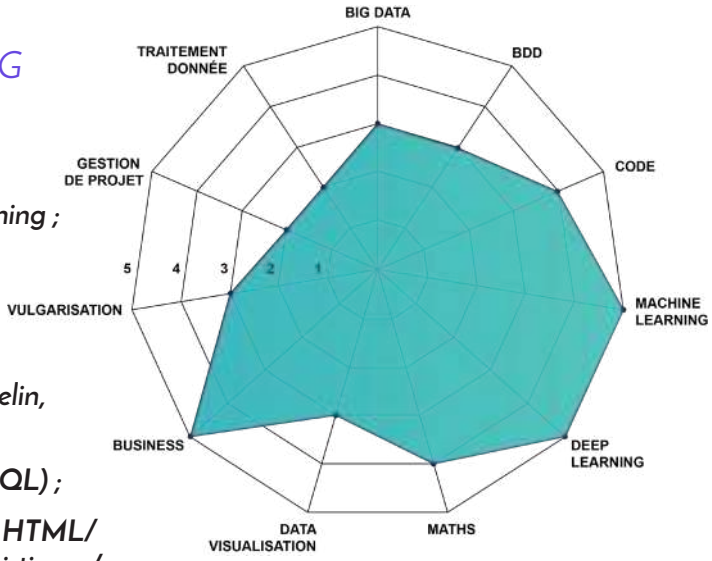
*Il n'en existe pas encore de spécifique. Les plus adaptées sont des écoles d'informatique (comme Epita, Epitech, Enseeiht) ou d'ingénieurs généralistes (comme UTC, UTT, Polytech). Quant aux data ingénieurs les plus expérimentés, ce sont en général d'anciens développeurs, des profils issus de l'informatique décisionnelle, voire d'autres sciences comme la physique théorique.*

# MACHINE LEARNING INGÉNIEUR

C'est l'un des métiers les plus récemment apparus en sciences de la donnée. Le machine learning ingénieur est au croisement de la data science et du data engineering

Le *machine learning* ingénieur a pour rôle d'optimiser et de mettre en production les algorithmes développés par le *data scientist* au sein de l'infrastructure préparée par le data ingénieur. Il travaille donc en collaboration avec ces deux acteurs : il conçoit, développe et teste des briques de *machine learning*, voire de data visualisation, par exemple pour outiller une plateforme « *datalab as a Service* », pour créer des modèles prédictifs à inclure dans un prototype « *Proof Of Concept* », ou encore pour comparer des méthodes d'apprentissage automatique à d'autres méthodes prédictives. Les contours exacts de ce poste restent cependant encore à définir, tant son émergence est récente.

## LE DÉVELOPPEUR EN *MACHINE LEARNING*




### Au quotidien :

- *Étudier la faisabilité de solutions de machine learning ;*
- *Coder pour mener des tests fonctionnels.*

### Compétences :

- *Environnements techniques : Spark, HDFS, Zeppelin, Jupyter, Stack ELK, Python, Java, Angularjs ;*
- *Connaissances en bases de données (SQL et NoSQL) ;*
- *Compétences en développement web (Javascript, HTML/CSS), analyse de données (Python, R, Scala), statistiques/probabilités ;*
- *Compétences en machine learning et data science ;*
- *Notions sur les technologies Hadoop, Spark et Kafka.*

### Formation :

*il n'en existe pas encore de spécifique. Les plus adaptées sont des écoles d'informatique (comme Epita, Epitech, Enseeiht) ou d'ingénieurs généralistes. On trouve à ces postes des data ingénieurs plutôt orientés machine learning et data science, ou des data scientists qui se sont mis au développement.*

# LA DONNÉE
## DANS TOUS SES ÉTATS

CETTE INFOGRAPHIE RECENSE LES PRINCIPALES PROBLÉMATIQUES QUE PEUVENT RÉSOUDRE LES NOMBREUSES MÉTHODES D'ANALYSE ET DE TRAITEMENT DE GRANDES MASSES DE DONNÉES. SUR LA BASE D'UN EXEMPLE (LA SURVEILLANCE D'UNE USINE OU D'UN MOTEUR D'AVION), VOYONS EN QUOI CONSISTENT LES DIFFÉRENTES PHASES POSSIBLES DE TRAITEMENT, DU PRÉTRAITEMENT DES DONNÉES À L'IMPLÉMENTATION FINALE D'UNE RÈGLE DE DÉCISION.

LES DONNÉES PEUVENT ÊTRE ISSUES DE CAPTEURS, TEXTES, IMAGES, VIDÉOS, SONS, NUAGES 3D, TABLEAUX, RÉSEAUX...

### ANALYSE DESCRIPTIVE OU EXPLORATOIRE

RÉSUMÉ DE MASSES DE DONNÉES, TELLES QUELLES SI ELLES SONT QUANTITATIVES, SINON À PARTIR DE CARACTÉRISTIQUES QUANTITATIVES (FEATURES) JUGÉES PERTINENTES PAR MÉTIER OU EXTRAITES DES DONNÉES, NUMÉRISÉES

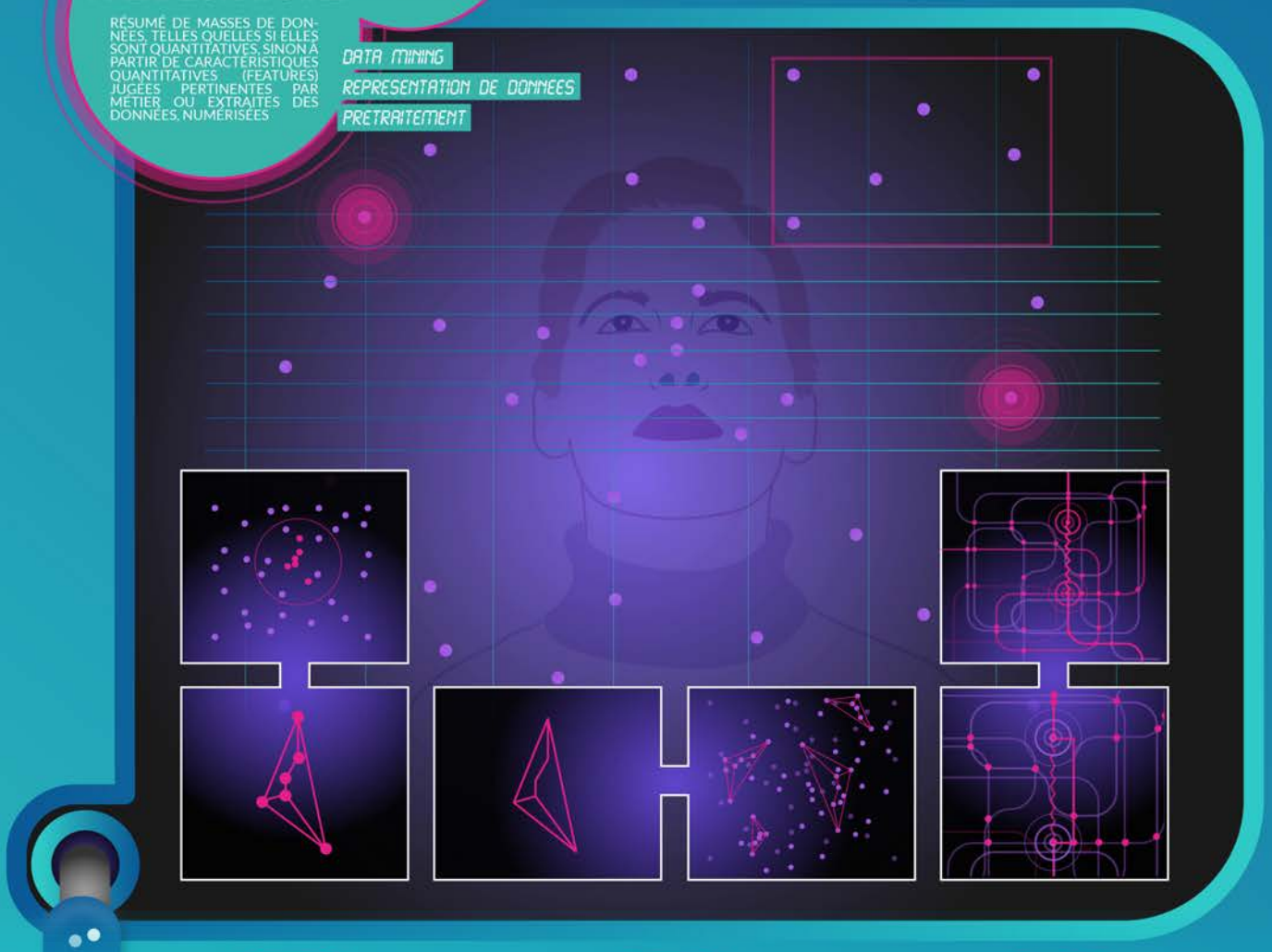
*EXEMPLE*

ALERTE GRÂCE À LA SURVEILLANCE EN TEMPS RÉEL DE GRANDEURS CLÉS (ÉNERGIE, PICS D'INTENSITÉ...) EN ANALYSANT LE SIGNAL ISSU DE CAPTEURS (QUANTITATIF)

*DATA MINING*

*REPRESENTATION DE DONNEES*

*PRETRAITEMENT*



DÉTECTION

ANALYSE

ANALYSE

## DÉTECTION OU RECHERCHE

RECHERCHE DANS UN ENSEMBLE DE DONNÉES D'UN ÉVÈNEMENT RARE (ATTAQUE SUR UN RÉSEAU INFORMATIQUE), D'UN MOTIF COMPLEXE À DÉTECTER (*PATTERN*, UN VISAGE DANS UNE IMAGE)

*EXEMPLE*

DÉTECTION D'ÉVÈNEMENT, D'ANOMALIE, DE NOUVEAUTÉ, DE SIGNAUX FAIBLES DANS LE COMPORTEMENT D'UNE INSTALLATION

RECONNAISSANCE DE FORME

TRAITEMENT D'IMAGES

MOTEUR DE RECHERCHE

## ANALYSE PRÉDICTIVE

CONSTRUCTION ET CALIBRATION, PARFOIS HORS LIGNE, D'UNE RÈGLE DE DÉCISION (UNE FONCTION, UN CALCUL) BASÉE :

- SOIT SUR L'EXPÉRIENCE, VIA DES SYSTÈMES DE RÈGLES NUMÉRISANT LE SAVOIR-FAIRE MÉTIER)
- SOIT SUR UN ALGORITHME D'APPRENTISSAGE.

*EXEMPLE*

PRÉDICTION SOUVENT EN TEMPS RÉEL DU COMPORTEMENT D'UN SYSTÈME, PRÉDICTION DE PANNE (MAINTENANCE PRÉDICTIVE)

MACHINE LEARNING

CLASSIFICATION

CATEGORISATION

SEGMENTATION

MOTEUR DE CIBLAGE/SCORING

MOTEUR DE RECOMMANDATION

REGRESSION/PREVISION

## ANALYSE PRESCRIPTIVE

PLANIFICATION DE LA MAINTENANCE D'INFRASTRUCTURES VOIRE AIDE AU DESIGN DE L'ARCHITECTURE DE SYSTÈMES OPTIMISANT PERFORMANCE ET RISQUE SUR LEUR CYCLE DE VIE

*EXEMPLE*

PRESCRIPTION DE MAINTENANCE, PLANIFICATION D'OPÉRATIONS DE MAINTENANCE (TIMING, ORDONNANCEMENT)

PROBLEME INVERSE

OPTIMISATION

MOTEUR DE RECHERCHE

## IMPLÉMENTATION (MISE EN PRODUCTION)

PROTOTYPAGE PUIS IMPLÉMENTATION DE LA RÈGLE DE DÉCISION ; L'APPRENTISSAGE DOIT PARFOIS ÊTRE FAIT EN ROUTINE COMME DANS LES SYSTÈME D'INFORMATIONS À GRANDE ÉCHELLE ; LES DONNÉES NE SONT PARFOIS VISIBLES UNE SEULE FOIS (LA THÉORIE DOIT ALORS S'ADAPTER)


## TEAM

RÉDACTRICE EN CHEF : ISABELLE BELLIN I EXPERTISE : NICOLAS VAYATIS

CHEF DE PROJET : ZEINA HADATI I DIRECTEUR ARTISTIQUE : VINCENT DEVILLARD


DAP DATA ANALYTICS POST

DAP

DATAANALYTICSPOST.COM

JE M'ABONNE !

TOUTES LES SEMAINES,
RECEVEZ GRATUITEMENT PAR EMAIL
UN ÉCLAIRAGE SCIENTIFIQUE, TECHNIQUE,
OPÉRATIONNEL, EMPLOIS...
SUR LES DATA SCIENCES

Le Data Analytics Post (DAP) est un média d'information et de réflexion autour des data sciences porté par le master MVA de l'ENS Paris-Saclay.

L'objectif du DAP est d'informer, décrypter et ouvrir le débat autour des problématiques de traitement et d'analyse des données, depuis leurs fondements jusqu'aux perspectives scientifiques et économiques prometteuses.

Notre ambition est d'animer l'écosystème que constituent les équipes de recherches académiques, les entreprises -des startups jusqu'aux grands groupes internationaux- et les filières de formations (enseignants et étudiants).

*www.dataanalyticspost.com*

***Inscrivez-vous à notre newsletter :*** *www.dataanalyticspost.com/sabonner-newsletter/*

## Remerciements :

**Upward Data** - cabinet de recrutement spécialisé dans les métiers du big data - pour son expertise et son aide pour la réalisation de ces fiches métiers.

**Isabelle Bellin**, pour la conception et la rédaction de ce guide.

**Vincent Devillard**, pour la création des illustrations.

**Thierry Cayatte**, pour la mise en page.

© Data Analytics Post 2019
Toute reproduction même partielle est strictement interdite sans l'autorisation du Data Analytics Post.