

## Fiche métier big data

# Data Scientist

*Être Data Scientist, c'est être au cœur de la valorisation des données, c'est comprendre les enjeux et les problématiques de l'entreprise autour de la donnée et proposer des méthodes et des algorithmes issues du machine learning et de la Datascience.*

*Les data scientists vont à la rencontre des métiers pour en recueillir les besoins, identifient les données pertinentes permettant d'extraire les signaux faibles caractérisant un phénomène et son évolution. Ils interviennent à toutes les étapes du cycle de valorisation de la donnée : définition de la problématique, collecte des données, nettoyage des données, choix des algorithmes (modèles de classification ou de prédiction) et validation à partir des données.*

## 1 METIER, 5 MISSIONS

### 1\_RECUPERER LES DONNEES

- Identifier les données à collecter et les phases de nettoyage à effectuer pour assurer la cohérence globale avant traitement des données (pour éviter les fausses informations, les informations en double, les formats différents, ...).
- Développer des algorithmes de collecte et de nettoyage des données ;
- Etudier et mettre en place les meilleures solutions techniques pour gérer les grands volumes de données ;

### 2\_DECOUVRIR LES DONNEES STATISTIQUES

- Effectuer une 1<sup>ère</sup> caractérisation des données : identifier différentes segmentations, catégories de pré-classement de donnée ;
- Qualifier la donnée sans développer de modèles prédictifs ;
- Mettre en œuvre et développer la modélisation statistique des données ;

### 3\_DEVELOPPER LES MODELES DE PREDICTION

- Créer des algorithmes permettant d'identifier des informations prévisionnelles à partir d'une grande quantité de données, en fonction d'un objectif défini ;
- Développer des algorithmes d'apprentissage et des scénarios prédictifs ;

### 4\_COMMUNIQUER

Concevoir des outils de datavisualisation, en interaction forte avec le data analyst, qui rendent visibles et lisibles les informations issues du traitement des données auprès d'interlocuteurs métier.

### 5\_ASSURER LA VEILLE TECHNOLOGIQUE SUR LES ALGORITHMES ET A MOINDRE MESURE SUR LES OUTILS DE DATASCIENCE

- Effectuer une veille sur les nouvelles technologies et solutions logicielles d'analyse des données ;
- Rechercher et expérimenter de nouvelles méthodes de modélisation et d'analyse des données. (math) ;
- Contribuer à l'évolution des pratiques dans son domaine de compétences ;

## UN METIER DE DIVERSITE

Salarié(e) d'une entreprise industrielle ou de service, il/elle travaille dans un environnement très technique. Les entreprises requièrent généralement pour des postes de data scientist des personnes ayant plusieurs années d'expérience dans le domaine du big data.

### MOBILITE ET DISPONIBILITE

Le/la Data scientist est plutôt sédentaire mais il/elle peut aussi être amené(e) à se déplacer chez ses clients.

## UN METIER, DES COMPETENCES

### SAVOIR

- A de solides connaissances mathématiques : Arithmétique, Algèbre linéaire, Calcul différentiel et intégral, Probabilités et Statistiques, Analyse multivariée ;
- A de solides connaissances algorithmiques : Algorithmique et Optimisation ;
- A de solides connaissances informatiques ;
- A de solides connaissances en machine learning\* : régression linéaire, régression logistique, méthode des "k plus proches voisins", arbres de décision, forêts aléatoires, machines à vecteur de support, boosting, réseaux de neurones, algorithmes génétiques ;

\*machine learning = une manière de construire un algorithme qui fait de l'auto apprentissage sur la base de l'observation du passé et qui permet de résoudre des problèmes complexes

### SAVOIR-FAIRE

- A de fortes aptitudes en développement / programmation : Bases de données (SQL), programmation (Python, R, Scala), Frameworks de calcul distribué (Hadoop, Spark, CUDA, ...) ;
- A une excellente maîtrise du machine learning ;
- Sait utiliser des outils de *Web analyse* (Omniure, Google analytics etc.) ;
- Fait preuve d'intelligence économique en écoutant ses intuitions stratégiques (esprit d'entrepreneur) ;
- A une bonne maîtrise de l'anglais (*converser en anglais en contexte pro.; utiliser un vocabulaire technique; comprendre de la documentation technique; écrire en anglais les livrables, notes, e-mails... nécessaires à la réalisation des activités...*) ;
- Collabore fortement avec les autres membres de l'équipe (*coopération et transparence permanente sur l'avancement du travail, ...*) ;
- *compétence permettant de « travailler en équipe AGILE »*
- Travaille par incrémentation et itération de 1 à 4 semaines (*le processus de développement est découpé en plusieurs phases appelées des "sprint", et chaque itération répond à des objectifs courts termes*) ;
- *compétence permettant de « travailler en équipe AGILE »*

## UN METIER, DES COMPETENCES

### SAVOIR-ETRE

- A un bon esprit d'analyse ;
- Fait preuve d'une grande honnêteté intellectuelle et de prise de recul par rapport aux méthodes utilisées pour garantir qu'elles soient adaptées au contexte des données traitées ;
- A une bonne capacité d'organisation pour structurer ses méthodes de travail et son plan d'intervention ;
- Est à l'écoute afin de recueillir avec précision les informations et les besoins des clients internes et externes de façon à mieux les satisfaire ;
- Est assez curieux pour suivre les nouvelles tendances et découvrir de nouveaux outils ;
- Est force de proposition auprès des différents acteurs de l'entreprise ;
- Sait demander de l'aide à bon escient ;
- *compétence permettant de « travailler en équipe AGILE »*
- Sait se contenter d'un travail individuel satisfaisant (pour faire le maximum avec les ressources allouées, sans recherche de perfection) ;
- *compétence permettant de « travailler en équipe AGILE »*
- A une grande capacité d'adaptation (sait rester réactif face au changement) ;
- *compétence permettant de « travailler en équipe AGILE »*
- Est capable de travailler dans un domaine non connu (en décalage avec son domaine d'expertise, mais toujours en gardant un lien) ;...
- *compétence permettant de « travailler en équipe AGILE »*

### SPECIFICITES

Dans une très petite, petite et entreprise de taille moyenne (TPE / PME) : le/la data scientist sera un rôle qui pourra être cumulé avec d'autres rôles comme le/la data miner et le/la data analyst.

Dans une Entreprise de Taille Intermédiaire (ETI) ou un grand groupe : le/la data scientist est un métier à part entière qui travaillera en équipe agile avec un/e data miner et/ou un/une data analyst et plus largement l'équipe big data. Dans cette configuration, il/elle pourra aussi en plus avoir un rôle de « scrum master » (chef de projet dans le big data).

Les compétences associées au travail en équipe agile, sont plus recherchées dans les entreprises d'une certaine taille qui a une équipe big data.

## ACCES AU METIER

Il n'existe pas véritablement de voie royale menant au métier de data scientist. Il existe presque autant de parcours professionnels que de data scientists sur le marché. Le point commun est qu'ils/elles possèdent tous/toutes un bagage technique conséquent.

Les connaissances acquises sont à compléter par des expériences sur le terrain, au sein des entreprises, et par l'auto-formation, grâce à des Moocs par exemple.

### QUELQUES EXEMPLES DES DIPLOMES LES PLUS CONNUS PERMETTANT L'ACCES AU METIER :

#### BAC+5 :

- Formation pilotée par Grenoble INP - Ensimag et l'EMSI Grenoble (GEM) ;
- Ecole d'ingénieurs, Master en mathématiques, physique, Informatique, actuariat ou statistiques qui aspirent à une insertion professionnelle de haut niveau ;
- Ecole d'ingénieurs ou docteur spécialisé en analyse statistique et programmation informatique ;
- Ex : Diplôme d'ingénieur de l'École nationale de la statistique et de l'analyse de l'information de l'ENSAI
- Filière **IAMD** (Ingénierie et Applications des Masses de Données) à Télécom Nancy ;
- Master 2 mathématiques pour la science des masses de données de l'école Polytechnique ;
- Master 2 « Data Sciences », Telecom ParisTech ;
- Master 2 Informatique : parcours « Data & Knowledge » et « Data Scale », Telecom ParisTech ;
- Master 2 Big Data , Polytechnique (Université Paris-Saclay) ;
- Mastère spécialisé en data sciences et business analytics de l'Essec-Centrale Paris ;
- Mastère spécialisé Big Data de Télécom Paris Tech ;
- Mastère Spécialisé « Big Data : gestion et analyse des données massives », Telecom ParisTech  
(→ *destiné aux diplômés en poursuite d'études ou en reconversion*) ;
- Mastère Spécialisé Assurance, Actuariat et Big Data, ESILV (Ecole Supérieure d'Ingénieurs Léonard de Vinci) ;
- Mastère spécialisé *Big data analyse*, management et valorisation responsable, ENSIMAG et EMSI Grenoble (GEM) ;

**FORMATION CONTINUE OU VAE :** CES « Data Scientist » proposé par Télécom ParisTech

**TITRE PROFESSIONNEL : 22490 :** Master traitement de données volumineuses

- MOOC :**
- Sur le site du MIT, de Coursera ou de Big Data University ;
  - Fondamentaux pour le Big Data (IMT Institut Mines Telecom) sur FUN-MOOC ;
  - The Data Scientist's toolbox, COURSERA ;

#### PLATEFORME DE COMPETITIONS EN LIGNE :

(pour comparer les modèles de prédiction les plus performants)

- [Kaggle](https://www.kaggle.com/)
- [Datascience.net](https://www.datascience.net/)
- ...

## CE QU'IL FAUT RETENIR

### CENTRE D'INTERET

- Apprécier le travail en équipe ;
- Aimer informer, communiquer ;
- Avoir un goût pour la veille technique (technologique et mathématique) ;
- Avoir un goût pour les défis mathématiques ;
- Avoir un goût pour la recherche de solutions ;

### FONCTIONS

- Production ; Maintenance ;
- Support ;
- Marketing & Commercial ;

### SECTEURS D'ACTIVITES

La métallurgie, la chimie, l'agroalimentaire, ... et plus généralement tous les secteurs d'activités de l'industrie et des services. A noter également que les métiers du big data, bien au-delà de l'industrie sont présents et très recherchés dans de nombreux autres secteurs d'activités : immobilier, santé, ...

### STATUT

Salarié(e) d'une entreprise (plutôt Entreprise de Taille Intermédiaire – ETI - et grande entreprise) ou d'une société de services intervenant pour toute taille d'entreprises et notamment pour les PME. Il/elle peut aussi être consultant/e indépendant/e.

### CODE ROME\*

- M1403 - Études et perspectives socio-économiques
- M1803 - Direction des systèmes d'information
- M1805 - Études et développement informatique
- M1806 - Conseil et maîtrise d'ouvrage en systèmes d'information
- M1810 - Production et exploitation de systèmes d'information

\* Le Répertoire opérationnel des métiers et des emplois (ROME) est un répertoire créé en 1989 par l'ANPE (Agence nationale pour l'emploi), aujourd'hui Pôle emploi en France. Il sert à identifier aussi précisément que possible chaque métier. Ce répertoire comprend plus de 10 000 appellations de métiers et d'emplois. Cette nomenclature et la codification du ROME sont utilisées par d'autres organismes publics ou privés traitant de l'emploi. Le ROME est utilisé pour l'élaboration des familles professionnelles (FAP) en France<sup>1</sup>.

## EN SAVOIR +

### PUBLICATIONS

KEEPING UP WITH THE QUANTS, Thomas H. Davenport et Jinho Kim

<http://www.amazon.com/Keeping-Up-Quants-Understanding-Analytics/dp/142218725X>

DATA SCIENCE IS A TEAM SPORT, Michael Walker

<http://www.datasciencecentral.com/profiles/blogs/data-science-is-a-team-sport>



### Les métiers de la data

Tour d'horizon d'un secteur d'avenir à la recherche de nouveaux talents

Auteur(s) : [Camille Regache](#)

Editeur(s) : [L'Étudiant](#)

Collection : [Les guides de l'Étudiant - Métiers et formations](#)

Nombre de pages : 190 pages

Date de parution : 13/04/2017

EAN13 : 9782817605746



### Big data, penser l'homme et le monde autrement

Auteur(s) : [Gilles Babinet](#)

Editeur(s) : [Le Passeur](#)

Nombre de pages : 248 pages

Date de parution : 24/11/2016

EAN13 : 9782368904923



### Data Scientist et langage R

Guide d'autoformation à l'exploitation des Big Data

Auteur(s) : [Henri Laude](#)

Editeur(s) : [Eni](#)

Nombre de pages : 663 pages

Date de parution : 12/03/2016

EAN13 : 9782409000430

**EN SAVOIR +**



## Big data, open data et valorisation des données

Auteur(s) : [Jean-Louis Monino](#), [Soraya Sedkaoui](#)

Editeur(s) : [Iste](#)

Collection : [Innovation, entrepreneuriat et gestion - Smart innovation](#)

Nombre de pages : 158 pages

Date de parution : 26/02/2016

EAN13 : 9781784051228



## Le big data

Ses origines, ses outils et méthodes, ses effets

Auteur(s) : [Pierre Delort](#)

Editeur(s) : [PUF](#)

Collection : [Que sais-je ?](#)

Nombre de pages : 128 pages

Date de parution : 22/04/2015

EAN13 : 9782130652113



## Du big data au smart data

Au service d'un monde connecté

Auteur(s) : [Fernando lafrate](#)

Editeur(s) : [Iste](#)

Collection : [Systèmes d'information, web et informatique ubiquitaire](#)

Nombre de pages : 86 pages

Date de parution : 02/03/2015

EAN13 : 9781784050689



## Apprentissage machine

De la théorie à la pratique - Concepts fondamentaux en Machine Learning

Auteur(s) : [Massih-Reza Amini](#)

Editeur(s) : [Eyrolles](#)

Collection : [Algorithmes](#)

Nombre de pages : 272 pages

Date de parution : 05/02/2015

EAN13 : 9782212138009 

**EN SAVOIR +**

## **AUTRES SOURCES**

[http://www.lemonde.fr/campus/article/2017/05/12/le-data-scientist-nouvelle-coqueluche-des-recruteurs\\_5126695\\_4401467.html](http://www.lemonde.fr/campus/article/2017/05/12/le-data-scientist-nouvelle-coqueluche-des-recruteurs_5126695_4401467.html)

[http://www.lemonde.fr/campus/article/2017/05/16/serge-abiteboul-data-scientist-est-un-metier-qui-demande-enormement-de-neurones\\_5128539\\_4401467.html](http://www.lemonde.fr/campus/article/2017/05/16/serge-abiteboul-data-scientist-est-un-metier-qui-demande-enormement-de-neurones_5128539_4401467.html)

<https://www.fedbusiness.fr/actualites-fed-business/metier-de-data-scientist-cest-qui-cest-quoi>

<https://www.usine-digitale.fr/article/data-scientist-un-metier-d-elite-en-voie-de-democratisation.N444472>

## **LES SITES WEB**

[www.syntec-numerique.fr](http://www.syntec-numerique.fr)

<http://www.assopascaline.fr/>

## **LES VIDÉOS LES PLUS PERTINENTES SUR LE MÉTIER**

Vis mon job de data scientist (2'44)

<https://www.youtube.com/watch?v=KedSyJzIJo&feature=youtu.be>

Dessine-moi un data scientist (3'11)

<https://www.youtube.com/watch?v=iydqa8NuzE>

**EN SAVOIR +**

**EXEMPLE D'OFFRES D'EMPLOI :**

**DATA SCIENTIST JUNIOR H/F**

**CGI - experience the commitment**

Catégorie : Renseignements numériques/Décisionnel (Gestion des données/intelligence d'affaires)

Ville : Paris, Paris, France

Type d'emploi : Temps plein

**Description du poste :** Intégré(e) à notre équipe « Valorisation de la donnée », vous participerez en qualité de **Data Scientist Junior** aux différentes phases de nos projets : conseil, avant-vente, cadrage des besoins, choix et optimisation des algorithmes et mise en place technique de la solution retenue. Encadré(e) par une équipe d'experts, vous serez amené(e) à intervenir sur des problématiques pointues de Data Science en utilisant, selon les choix technologiques de nos clients, des infrastructures, des outils et des langages variés : Python, Spark, R, Hadoop, MongoDB, ElasticSearch, Cassandra, Teradata,...

**Fonctions et responsabilités :**

*Quelques exemples de missions :*

- Énergie : maintenance prédictive, prévision de consommations, détection de fraudes, ...
- Télécom : segmentation clients, analyse des logs, ...
- Média : système de recommandations, analyse de sentiments,...
- Data Lab' interne : veille technologique, concours de Data Science, benchmark,...

Enfin, nous attendons de vous que vous soyez moteur dans le développement de notre offre en réalisant de la veille technologique, en proposant et en développant des démonstrateurs sur notre Data Lab' interne et que vous évangélisiez vos connaissances techniques et/ou fonctionnelles au sein de CGI.

**Qualités requises pour réussir dans ce rôle :**

De formation Bac+5, vous justifiez d'une première expérience dans un environnement similaire et avez acquis des compétences dans les domaines suivants :

- Manipulation de données: SQL, R, Python, Java, Scala,...
- Data Science : analyses statistiques avancées, machine learning, deep learning,...
- Une première expérience en Big Data est un plus : Hadoop, Spark, MongoDB, Cassandra,...
- Une première expérience en Data Visualisation est un plus : Qlik, Tableau, PowerBI, D3.js,...

Votre curiosité et votre rigueur vous permettent d'évoluer dans un contexte exigeant, nécessitant des qualités de communication, souvent directement avec les équipes métier du client.

Vous êtes capable de communiquer les résultats d'une analyse de manière claire, précise et efficace.

Vous savez collaborer en équipe et faites preuve de capacité d'adaptation.

Ce poste est ouvert aux personnes en situation de handicap.



**EN SAVOIR +**

**Ce que vous pouvez attendre de nous :**

Chez CGI, nous sommes une équipe de bâtisseurs. Chacun de nos membres (ou employés) concourt à bâtir sa propre entreprise - une firme qui regroupe 65,000 professionnels présents dans 40 pays. Fondée en 1976, CGI est un chef de file des services en TI et en gestion des processus d'affaires qui contribue au succès de ses clients. Nos clients - tout comme nos membres - peuvent compter sur nos ressources mondiales, sur notre savoir-faire, sur notre stabilité et sur nos professionnels chevronnés pour obtenir les résultats escomptés. Joignez-vous à nous. Pour en savoir davantage : [www.cgi.com](http://www.cgi.com).

Cette offre est une excellente occasion de vous joindre à une équipe gagnante. CGI offre des conditions de rémunération avantageuses et des possibilités de croissance et de développement professionnel. Les avantages des membres permanents à temps plein entrent en vigueur dès le premier jour et comprennent des vacances payées, un régime de participation aux bénéfices ainsi qu'un régime de participation à l'actionnariat.

Nous aimerions remercier tous les candidats de leur intérêt pour notre organisation. Cependant, nous communiquerons seulement avec les candidats retenus.

Les candidatures non sollicitées provenant d'agences de placement ne seront pas retenues.

CGI favorise l'équité en matière d'emploi. De plus, CGI s'engage à offrir des accommodements aux personnes handicapées, conformément à la législation provinciale. Veuillez nous indiquer si vous vivez avec un handicap nécessitant la mise en place d'accommodements raisonnables dans le cadre de notre processus de recrutement; nous collaborerons avec vous pour répondre à vos besoins.

---

**Compétences** : Hadoop Hive, Java, Python, Scala, SQL



**Data Scientist est un métier qui vous intéresse.**

**N'hésitez pas à demander encore plus d'informations :**

**[osonslindustrie@industrie-dufutur.org](mailto:osonslindustrie@industrie-dufutur.org)**