



## Fiche métier big data

# Data Engineer

*Les data engineers sont les professionnels de la donnée qui préparent l'infrastructure big data pour l'exécution de traitement, notamment ceux conçus par les data scientists.*

*Ce sont des ingénieurs en développement informatique qui préparent et mettent en service le socle matériel et logiciel qui permet d'exécuter différents traitements informatiques optimisés :*

- l'intégration de grands volumes de données à partir de diverses sources,*
- la transformation de ces données au moyen de différents traitements de nettoyage, normalisation, et enrichissement,*
- les traitements analytiques conçus par les data scientists,*
- et enfin des traitements décisionnels permettant d'établir des rapports périodiques automatisés, ou des visualisations à la demande.*

## 1 METIER, 3 MISSIONS

### 1\_ METTRE EN PLACE OU SUPERVISER LA MISE EN PLACE D'UNE ARCHITECTURE BIG DATA

- Coordonner, en lien avec la Direction des Systèmes d'Information, la mise en place de l'architecture big data, conçue par le data architect : infrastructures (*machines physiques, cloud, canaux de communications*), logiciels (*middleware et systèmes d'exploitation*) ;
- Assurer la maintenance de l'architecture big data en lien avec la Direction des Systèmes d'Informations (DSI) ;

### 2\_ INDUSTRIALISER / AUTOMATISER

Industrialiser / automatiser dans un environnement de développement :

- les algorithmes de récupération des données ;
- les algorithmes de découverte des données ;
- les modèles de prédiction des données ;

### 3\_ ASSURER LA VEILLE TECHNOLOGIQUE SUR LES ARCHITECTURES BIG DATA

- Effectuer une veille sur les nouvelles technologies et solutions matérielles de transport et de stockage des données ;
- Effectuer une veille sur les nouvelles technologies et solutions logicielles de structuration et d'infrastructures des données (*solutions middleware*) ;
- Contribuer à l'évolution des pratiques dans son domaine de compétences, en étant actif dans des réseaux de veille ;

## 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 engineer des personnes ayant plusieurs années d'expérience dans le domaine du big data.

### MOBILITE ET DISPONIBILITE

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

## UN METIER, DES COMPETENCES

### SAVOIR

- A des connaissances en mathématiques : en statistique de bases, en modélisation, en analyse de données (Math appliqués) ;
  - A de solides connaissances des principales architectures applicatives : modèles de conception logicielle (ex : MVC, ...), Framework (ex : java,...) ;...
  - A de fortes connaissances en algorithmique : informatique théorique (théorie de la complexité,...) (base de la recherche opérationnelle) et des bases en machine learning ;
  - Connaît bien les technologies d'infrastructures informatiques et leurs performances :
    - Cloud : support de l'externalisation de services, facturés à l'usage
    - non Cloud : technologies de calcul intensif (machines virtuelles, CPU\*, carte graphique, ...) n'ayant pas recours au cloud
- \*CPU = Control Processing Unit, API = Applicable Programming Interface

### SAVOIR-FAIRE

- A de fortes aptitudes en développement / programmation :
  - A une excellente maîtrise des langages de programmation suivant : Python, Scala, Java, Shell Unix
  - Sait programmer en langage C/C++, javascript,
  - A une excellente maîtrise des technologies Big data (HADOOP, SPARK)
  - A une excellente maîtrise du langage SQL
  - Avoir des bases dans la programmation de quelques base de données NoSQL (not only SQL) est un plus (Neo4j, MongoDB, Cassandra, 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 professionnel; 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'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 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 »*
- Se contente 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 (rester réactif 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, entreprise de taille moyenne (TPE / PME) : le/la data engineer sera un rôle qui pourra être cumulé avec d'autres rôles comme le data architect.

Dans une Entreprise de Taille Intermédiaire (ETI) ou un Grand groupe : le/la data engineer sera un métier à part entière qui travaillera en équipe agile avec un data architect. 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

Pour devenir Big data engineer, il faudra envisager des études relativement longues puisqu'un niveau bac +4/5 est généralement requis pour exercer cette activité.

La plupart du temps, les entreprises privilégient les candidats ayant suivi des cursus dans les domaines de l'ingénierie et de l'informatique.

Il n'existe pas véritablement de voie royale menant au métier de data engineer.

Il existe presque autant de parcours professionnels que de data engineer sur le marché d'autant que ces profils "technologiques" sont souvent des techniciens confirmés.

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

### BAC+5 : DIPLOMES D'ECOLES

- Expert en ingénierie informatique (CRESPA école du Campus Sciences-U Lyon) avec le Mastère Professionnel Expert en Ingénierie Informatique option Business Intelligence et Big Data ;
- Filière *IAMD* (Ingénierie et Applications des Masses de Données) à Télécom Nancy ;
- Expert en ingénierie informatique (EFFICOM de Lille avec son Mastère Professionnel ingénierie informatique ;
- Expert en ingénierie informatique (EPITA), avec une Majeure MTI (Multimédia et Technologies de l'Information) ;
- Expert en ingénierie informatique (ISTM Institut Supérieur de Technologie et Management) de Paris avec un MBA Management & Big Data ;
- Expert en ingénierie informatique (ESIMED) ;

### MOOCs

- Sur le site du MIT, de Coursera ou de Big Data University ;
- Fondamentaux pour le Big Data (IMT\_Institut Mines Telecom) sur FUN-MOOC ;

### CERTIFICAT PROFESSIONNEL

Google Certified Professional - Data Engineer

<https://cloud.google.com/certification/data-engineer?hl=fr>



## CE QU'IL FAUT RETENIR

### CENTRE D'INTERET

- Avoir le sens du contact ;
- Apprécier le travail en équipe ;
- Aimer assumer des responsabilités ;
- Aimer manipuler les chiffres et la logique mathématique ;

### 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 prospectives 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



### Big data, open data et valorisation des données

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

Editeur(s) : [Lste](#)

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

Nombre de pages : 158 pages

Date de parution : 26/02/2016

EAN13 : 9781784051228

**EN SAVOIR +**



## 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 : Systemes d'information, web et informatique ubiquitaire

Nombre de pages : 86 pages

Date de parution : 02/03/2015

EAN13 : 9781784050689

## LES SITES WEB

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

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

**EN SAVOIR +**

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

**DATA ENGINEER**

2017-09-07/dans [Data Engineering CDI D3.js](#), [Linux](#), [Management](#), [Python](#), [SQL](#) /par [jharousseau](#)

*A propos de Datawords*

Datawords est le pionnier d'une nouvelle discipline, l'e-Multiculturalisme, qui permet à ses clients, grandes marques et acteurs du e-commerce, d'accélérer leur déploiement international.

Datawords a en effet développé un savoir-faire et une expertise uniques dans l'adaptation, la localisation et l'actualisation digitale multilingue. Au croisement de la sémantique et de la technique, l'agence intervient à tous les niveaux de la production digitale multilingue pour s'assurer que le marketing digital d'une marque soit déployé à l'international, de façon pertinente et synchronisée, pour des performances optimales.

Datawords compte aujourd'hui 500 collaborateurs, représentant une quarantaine de nationalités et plus d'une cinquantaine de langues. La société est implantée en France, en Italie, en Allemagne, en Angleterre, en Corée du Sud, à Hong Kong, au Japon, à New-York et, depuis peu, à Barcelone et à Milan. Datawords est une entreprise en pleine croissance et à l'identité forte qui offre aux candidats passionnés par les nouvelles technologies et le multiculturalisme une opportunité unique de participer à l'aventure du web et d'appréhender ses multiples enjeux à l'international.

Afin de répondre aux évolutions technologiques et aux défis croissants du secteur en termes d'innovation, Datawords recherche un/une Data Engineer (H/F).

**Description du poste**

Equipe

La croissance continue de Datawords se traduit par des projets d'adaptation de contenu digital multilingue toujours plus nombreux et d'envergure grandissante.

Nous recueillons ainsi de plus en plus de données sémantiques multilingues et de données liées à l'utilisation de nos plateformes.

Partant de ce constat et en anticipant la situation dans quelques années, nous sommes convaincus que les dernières technologies web et l'intelligence artificielle seront au cœur de notre stratégie pour atteindre notre objectif : devenir un leader mondial de notre secteur. C'est pourquoi nous avons récemment créé l'équipe Machine Learning et NLP composée de Data Engineers.

Elle est encadrée par notre CTO et un Consultant Digital et collabore étroitement avec l'équipe de Semantic Solutions Developers.

Son objectif premier est la construction d'une offre pérenne et complète de services "intelligents" en lien avec nos activités. Ces services doivent pouvoir interagir avec nos plateformes logicielles qui leur mettent des données à disposition et qui reçoivent en échange des feedbacks data-driven.

*Quelques aspects de l'ADN de l'équipe :*

- Approche systémique
- Travaux organisés autour de sprints de 2 semaines
- Benchmarking des solutions déjà existantes sur le marché
- Développement et déploiement en MVP (Minimum Viable Product)
- Amélioration et innovation continues
- Coordination et communication auprès d'autres équipes



**EN SAVOIR +**Missions :

En rejoignant l'équipe Machine Learning et NLP en tant que Data Engineer, vos défis seront nombreux. Ils seront tout d'abord d'ordre technique et technologique avec la prise en main de notre environnement data, la compréhension de nos problématiques et le développement de services à haute valeur ajoutée. Il sera également attendu de vous que vous développiez l'équipe et pérennisiez son offre en interne dans un premier temps et auprès de nos clients à plus long terme. De manière plus précise, vous pourrez être amené à travailler sur les missions suivantes :

- Contribution au monitoring et à la maintenance des services en production
- Conception et déploiement de nouveaux services "intelligents" (NB : les services sémantiques devront être développés en ayant en permanence le souci de la perspective multilingue)
- Si les résultats liés à un service innovant que vous avez contribué à développer et déployer sont concluants, la rédaction d'un papier dans l'optique d'une soumission pour publication dans une revue scientifique ou pour une conférence pourra être envisagée
- La veille technologique et concurrentielle/ Veille sur les différentes librairies open source intéressantes pour nos services
  - > Veille sur les services cognitifs (IBM Watson, Google, Azure...)
  - > Participation aux principaux salons d'intelligence artificielle de la région
  - > Suivi de l'évolution de l'offre de nos concurrents
- Le développement de l'équipe
  - > Pédagogie auprès des autres équipes sur la proposition de valeurs de l'équipe
  - > Participation à la rédaction de supports de communication interne et externe de l'équipe
  - > Représentation de l'équipe lors des actions école
  - > Participation au recrutement de stagiaires et de profils juniors

**Profil recherché**Parcours

- Diplômé(e) d'une Grande Ecole d'Ingénieurs
- Expériences dans la conception et le déploiement de solutions d'intelligence artificielle

Qualités recherchées

- Compétences en programmation et en déploiement de services (Python, Django REST, SQL, Docker, Redis...)
- Connaissances des principales librairies de Machine Learning et de NLP (Scikit Learn, Tensor Flow, NLTK, Gensim, OpenNMT...)
- Grande autonomie, rigueur
- Esprit d'initiative et de créativité
- Capacités de synthèse et de restitution des informations

Compétences et expériences appréciées

- Connaissances des services cognitifs IBM Watson, Google, Azure
- Expérience à l'international

## EN SAVOIR + EXEMPLE D'OFFRES D'EMPLOI :

### **DATA ENGINEER EXPERIMENTE H/F**

#### *Informations générales*

Date de parution : 06/10/2017

#### *Description du poste*

Métier : Application - Analyse, Conception, Développement

Intitulé du poste : Data Engineer Expérimenté H/F

Contrat : CDI classique (00)

#### *Description de la mission*

Intégré(e) au sein de notre Pôle d'Excellence Insights & Data, déployé au niveau mondial et qui représente aujourd'hui 10 000 consultants dans le monde, vous participerez à aider nos clients à construire ou à transformer leur stratégie de gestion de l'information et évoluerez au sein d'un marché passionnant, innovant et en pleine expansion.

En tant que responsable technique Big Data au sein de la practice Insights & Data, vous serez amené à intervenir sur des projets d'implémentation de solutions Big Data pour :

- Encadrer techniquement l'équipe de Data Engineers
- Accompagner les développeurs junior s dans leur montée en compétences
- Développer des flux d'alimentation de Data Lake et de traitement de données
- Participer à la mise en place d'architectures Big Data
- Travailler en collaboration avec nos consultants fonctionnels, nos architectes Big Data et nos Data Scientists

#### *Intégré(e) au sein de l'une de nos équipes, vous pourrez :*

- Bénéficier du savoir faire de Capgemini (méthodologies projets, expertises sectorielles ...), de sa grande communauté Big Data réparti dans le monde, et des possibilités offertes à l'étranger.
- Mettre en pratique vos acquis sur des projets d'envergure, développer vos connaissances et acquérir de nouvelles compétences techniques, méthodologiques et fonctionnelles.
- Monter en compétences sur de nouvelles technologies
- Construire votre carrière grâce à notre programme de développement professionnel.

#### **Profil, compétences "Rôle" requis :**

Développement et bonnes pratiques associées.

Capacité à encadrer une équipe

Capacité à accompagner des profils junior

#### **Compétences techniques requises :**

Big Data: Hadoop (Cloudera/Hortonworks/MapR) • Spark • Compréhension des mécanismes de MapReduce et de HDFS • Script shell • Linux

#### **Les + :**

Développement Java • Moteur de recherche (ElasticSearch, SolR) • Complexe event processing : Kafka, Storm • sgbd: PostgreSQL, Oracle, MySQL • nosql : MongoDB, CouchBase, HBase, Cassandra, Neo4J • etl : Talend, Informatica, Pentaho • Analytics: R, Python, Dataliku, SAS, SPSS • Dataviz: Qlikview, Tableau, d3.js, openlayer

#### **Diplômes/Formation/Expérience :**

De formation Bac +5 (école d'ingénieur), vous avez 2 à 3 ans d'expérience en développement et implémentation de solutions Big Data.

Curieux et souhaitant progresser, vous êtes apprécié(e) pour votre rigueur et votre capacité de travail.

## **EN SAVOIR + EXEMPLE D'OFFRES D'EMPLOI :**

### **DATA ENGINEER - H/F**

Chez D2SI

#### **Le poste**

Vous rejoignez l'équipe de data engineers D2SI en prenant part à différents projets :  
Construction de pipelines de traitement de données sur l'un des 20 plus gros clusters Hadoop de France au sein d'une équipe de 12 data engineers. Cela passe par la mise en place de l'infrastructure, du traitement et du reporting de ces données.

Mise en place d'environnements Hadoop et Spark à la volée via le service EMR d'AWS avec des contraintes de chiffrement.

Construction d'une chaîne d'analytics streaming basée sur les services managés Google Cloud Platform.

#### **Les responsabilités**

- Vous mettez en place des solutions de data engineering performantes et résilientes avec Hadoop, Spark...
- Vous développerez en java et/ou scala des jobs permettant le traitement de données
- Vous mettez en places des outils d'automatisation tel que Terraform, Travis CI, Ansible...
- Vous conseillerez nos clients sur l'utilisation des BDD et les outils de l'éco-système big data

#### **Pourquoi D2SI ?**

- Une communauté de passionnés passionnants, fan d'événements, de conférences, de meetups et de débats !
- La possibilité d'avoir des jours dédiés pour faire de la veille, préparer des formations, travailler sur des projets internes
- Un suivi Commercial & RH humain & décontracté et la possibilité d'être parrainé par un expert technique
- Un entreprise aux valeurs humaines & éthiques fortes (labels RSE)
- Travailler dès aujourd'hui sur les sujets de demain :)



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

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

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