



cegefos  
numérique  
académie

# BIG DATA

---

Outre la maîtrise des enjeux et des perspectives, le cursus Big Data se distingue par la maîtrise de multi-compétences en Gestion de données massives. Cette formation va vous permettre d'acquérir les compétences pour devenir Full Stack Big Data, à savoir : l'architecture et l'engineering Big Data, le stockage NoSQL, l'analyse et la science de données, la visualisation de données, le déploiement en Cloud d'une solution Big Data et l'administration d'une plateforme Big Data.

Maîtriser le concept Big Data pour la recherche, l'analyse, le partage et le stockage de données massives.

Explorer le potentiel BIG DATA au service de tous les métiers (Marketing, Contrôle de Gestion, Ressources humaines, Renseignement, Santé, Banque, Assurance...)

# Big data



## PRÉ-REQUIS

Pour le bon déroulement de la formation certifiante BIG DATA, les participants doivent :

- Avoir suivi la formation Développeur Java J2EE ou Microsoft.NET serait un plus
- Avoir une expérience des bases de données relationnelles
- Posséder des connaissances approfondies en statistique est un plus
- Etre capable de travailler en équipe avec un sens d'écoute et d'analyse

Les inscriptions se font sur dossier et après étude par l'équipe pédagogique CEGEFOS



## PUBLIC

Le parcours certifiant Big Data : Récolte et analyse des données volumineuse, s'adresse au :

- Directeurs/Chefs de projet ou Responsable métier
- Responsable système d'informations
- Développeurs informatiques
- Consultants en business intelligence
- Ingénieurs d'étude, de recherche et développement, Architecte système et logiciel
- Consultants techniques, Consultants business
- Statisticiens et Mathématiciens
- Tout Développeur ou informaticien

## En chiffres



### 214

Vidéos dédiées



### 198

Documents écrits



### 6

Formateurs



### 48

Tests et quiz



### 4

Projets fil rouge



### 4

Etudes de cas



## DÉBOUCHÉS

Le certificat « Big Data » va vous ouvrir un large panel de possibilités et vous permettre d'exercer les métiers suivants :

- Consultant Big Data
- Data Architect
- Data engineering
- Data Analyst
- Data Scientist avec Python
- Data Scientist avec R
- Data Scientist et Intelligence artificielle
- Architect cloud AWS, Azure ou GCP pour Big Data
- Développeur Full Stack Big Data

53  
JOURSRS  
2234

93835

certif  
info

## SON RÔLE

Outre la récolte et la Gestion des données volumineuses et les nouvelles technologies de Cloud, la formation certifiante BIG DATA se distingue par l'acquisition et la parfaite maîtrise des fonctionnalités stratégiques du traitement pointu des données afin de les transformer en valeur ajoutée au service d'une entreprise, une institution ou une administration.

Un Consultant Full Stack Big data doit être en mesure de :

- Définir la source de données et préparer l'infrastructure de stockage
- Collecter les données structurées et non structurées, nettoyer ces dernières et les stocker dans des bases de données NoSQL
- Créer des requêtes d'interrogation et d'analyse de données
- Créer des algorithmes d'analyse et de science des données et maîtriser les technologies de cette dernière : Machine learning, Deep learning, DataViz...
- Maîtriser les technologies spécifiques au Big Data comme Hadoop, Pig, Hive, Spark, Yarn, Cloudera



## OBJECTIFS A l'issue du cursus certifiant BIG DATA, le certifié sera en mesure de :

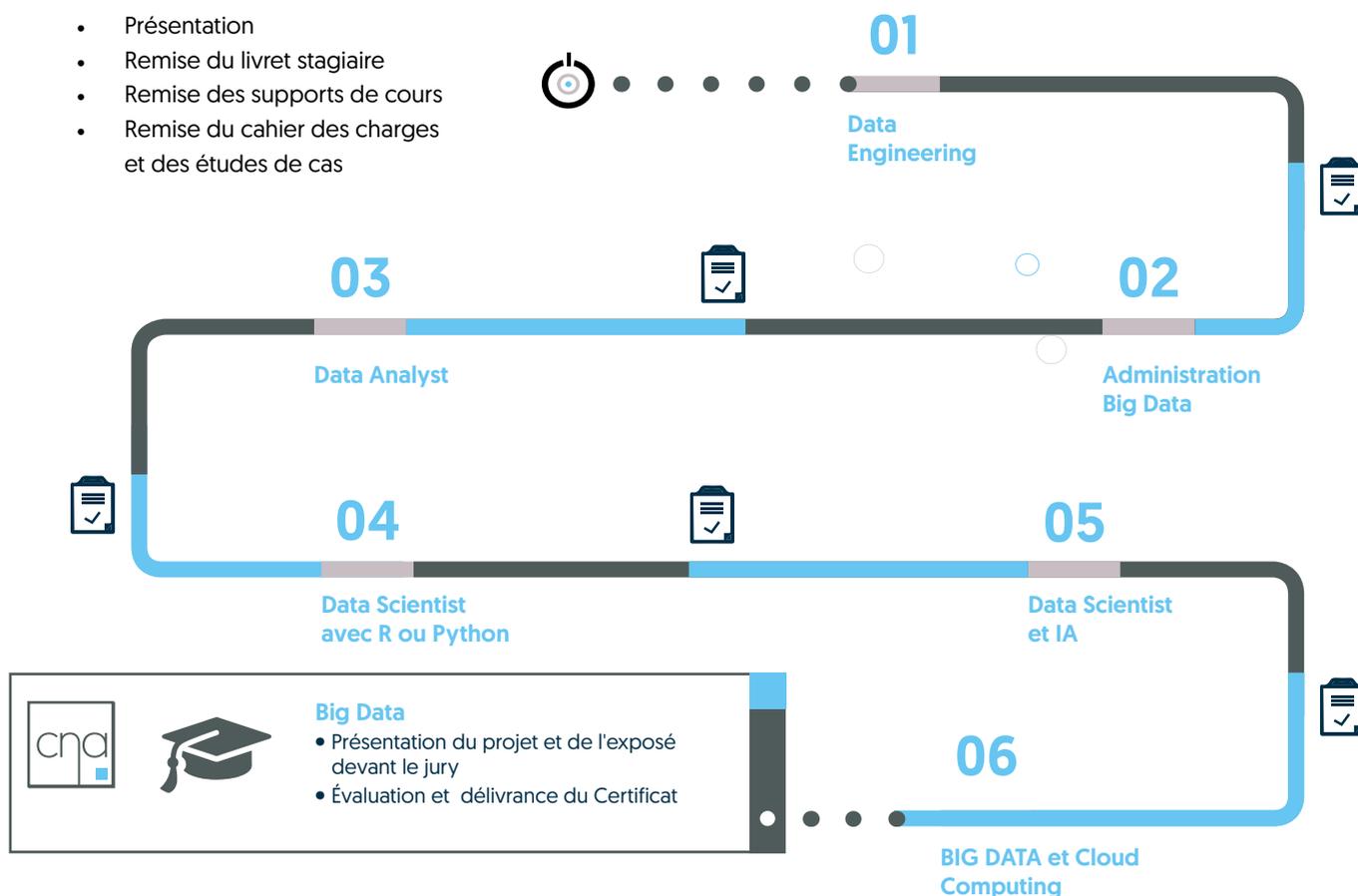
- Identifier les besoins et la problématique des directions métiers
- Maîtriser les technologies spécifiques au Big Data comme Hadoop, Pig, Hive, Spark, Yarn, Kafka, ELK, ou Cloudera
- Mettre en place un Data Lake
- Maîtriser les bases de données NoSQL : MongoDB, Cassandra, Neo4j, hBase, Redis...
- Maîtriser les technologies spécifiques à la science, à l'analyse ainsi qu'à la visualisation de données : Statistiques, Machine Learning, Deep Learning, langage R, Python, Tableau, PowerBI, R Shiny...
- Maîtriser le déploiement des solutions Big Data sur le Cloud : AWS, GCP, Microsoft Azure...
- Maîtriser la science de données et l'IA « Intelligence Artificielle » : Sprak Mlib, Scala, PySpark, RPA « Robotic Process Automation »
- Construire des modèles prédictifs pour répondre à la problématique
- Construire des algorithmes pour améliorer les résultats de recherche et de ciblage
- Trouver et rassembler l'ensemble des sources de données structurées ou non structurées nécessaires à l'analyse pertinente
- Identifier les opportunités à travers l'open data et les cas d'usage métiers
- Concevoir un projet Big Data (acquisition et stockage des données, traitement distribué, analyse de données à large échelle)
- Maîtriser les technologies par des études de cas concrètes
- Maîtriser les enjeux juridiques et liés à la protection des données

# Composition et déroulement

## Quels sont les modules par BADGE et comment se déroule la formation ?

**BADGE** : Bilan d'Aptitude Délivré par les Grandes Ecoles. Un BADGE correspond à un bloc de compétences.

- Présentation
- Remise du livret stagiaire
- Remise des supports de cours
- Remise du cahier des charges et des études de cas



## Blocs de compétences

Chaque Bloc de Compétences [BADGE] est sanctionné par la délivrance d'un CCP "Certificat de Compétence Professionnelle"



Pour la validation du certificat Big Data il faut au moins valider les trois blocs de compétences suivants :

- Big data engineering
- Data scientist
- Cloud pour Big Data

### 01 Data Engineering

Nombre de jours

Big Data : Enjeux et perspectives	1
Introduction à l'Architecture Big Data	1
NoSQL	1
Hbase / MongoDB / Cassandra / Neo4j	4
Apache Hadoop	4
Apache Spark	3
Apache Kafka	2
Elastic Stack Elk	2
Talend Open Studio for Data Integration	2
Big Data et Objets connectés	2
Gouvernance et sécurité	1
Mise en place d'une Plateforme BIG DATA	1
Mise en place d'un Data Lake	1

## 02 Administration Big Data

Nombre  
de jours

Mise en place d'un cluster Cloudera	4
Mise en place d'un cluster Elastic	2
Mise en place d'un cluster KAFKA	2
Administration d'une plateforme BigData dans le Cloud	2
Fondamentaux Devops/FinOps/SysOps	1

RS 2234

## 03 Data Analyst

Nombre  
de jours

Fondamentaux des statistiques	1
Analyse de données et méthodes statistiques	1
Analyse de données avec Python	1
Statistiques avec R	2
Visualisation de donnée avec Power Bi	2
<b>Modules de visualisation optionnels</b>	
• Analyse et Visualisation de donnée avec Excel	2
• Visualisation de donnée avec Tableau	2
• Visualisation de données avec R-Shiny	2
• Power BI : Analyse de données avec R	1
• Power BI : Visualisations avancées et notions clés du DAX	1
• Power BI avancé : L'analyse de données avec DAX 1 jour	1

RS 2234

## 04 Data Scientist avec Python

Nombre  
de jours

Introduction à la data science	1
ElasticSearch	2
Spark avec Hadoop	3
Programmation en Python	3
Machine Learning avec Python	5
Deep Learning avec Python	2
Visualisation : Notebook Zipline-Jupyter	2

RS 2234

## 05 Data Scientist avec R

Nombre  
de jours

Introduction à la data science	3
ElasticSearch	4
Spark avec Hadoop	6
R : Initiation et Analyses statistiques	1
R : Programmation	2
Machine Learning avec R	1
Deep Learning avec R	1
DataViz avec R-Shiny	3

RS 2234

## 06 Data Scientist et IA

Nombre  
de jours

Introduction : IA et la data science Deep Learning vs Machine Learning	4
Data Science avec Python	2
Industrialisation des cas d'usages de Data Science et de Data prep:	3
Machine Learning : Concepts de base, modèles linéaires, perception, K nearest neighbors	3
Deep Learning: Introduction et Principaux algorithmes	2
Introduction à Spark avec Scala et PySpark	2
Spark MLlib : Machine Learning avec Spark	3
Deep Learning avec TensorFlow	1
RPA (Robotic Process Automation)	5
RPA avec UiPath	12

RS 2234

# Blocs de compétences

## 07 BIG DATA et Cloud Computing AWS

	Nombre de jours
Les Fondamentaux : AWS	1
Intégration d'une solution Big Data : AWS	1
Déploiement d'une solution Big Data : AWS	3
Sécurité, SysOps et DevOps : AWS	2
Atelier pratique [Projet file rouge]	-

RS 2234

## 08 BIG DATA et Cloud Computing AZURE

	Nombre de jours
Les Fondamentaux : AZURE	1
Intégration d'une solution Big Data : AZURE	1
Déploiement d'une solution Big Data : AZURE	3
Sécurité, SysOps et DevOps : AZURE	2
Atelier pratique [Projet file rouge]	-

RS 2234

## 09 BIG DATA et Cloud Computing GCP

	Nombre de jours
Les Fondamentaux : GCP	1
Intégration d'une solution Big Data : GCP	1
Déploiement d'une solution Big Data : GCP	3
Sécurité, SysOps et DevOps : GCP	2
Atelier pratique [Projet file rouge]	-

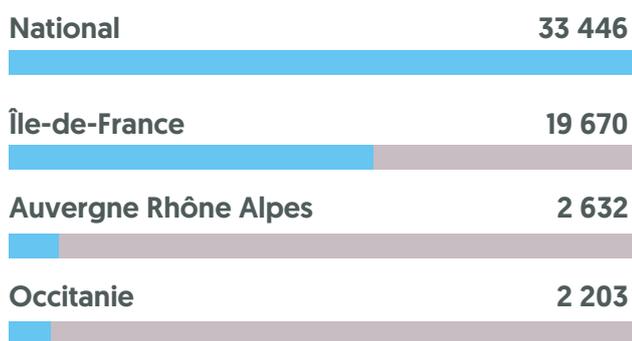
RS 2234

# Informations pratiques



Suivant le baromètre Besoin en Main d'Œuvre 2019 (BMO 2019) et la DARES, le nombre de postes à créer pour le consultant Big Data ainsi que les difficultés de recrutement en 2019 sont comme suit :

■ Nombre de postes



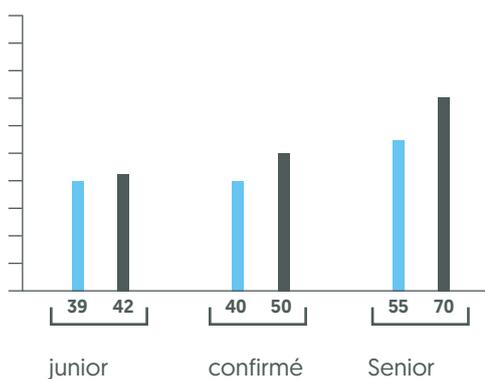
■ Difficulté de recrutement



K€



**SALAIRE EN K€ : Quels sont les chiffres associés aux postes et à la formation ?**



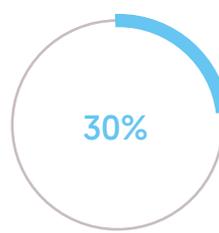
- Junior est déterminé entre 0 et 2 années d'expérience.
- Confirmé est déterminé entre 2 et 5 années d'expérience.
- Senior est déterminé au-delà de 5 années d'expérience.



**RÉPARTITION TEMPS DE FORMATION : Comment sont répartis les temps de formation ?**



Pratique



Théorique

# Et après où exercer ?

**Le certifié en Big Data travaille au sein des établissements aussi bien publics que privés, exemple :**

- A son compte comme indépendant ou Consultant en Big Data, Data engineering, Data Analyst, Data Scientist, Cloud computing pour Big Data
- Services informatiques d'une collectivité territoriale ou d'une administration
- Services informatiques d'une grande entreprise
- Entreprises de Services Numériques ESN
- PME – PMI
- Entreprises de vente en ligne e-commerce
- Sociétés de conseil en informatique et en système d'informations

## Les avantages



### POUR LES CERTIFIÉS

- Acquisition de nouvelles compétences en forte demande sur le marché de l'emploi, à savoir : Data Architect/Engineering, Analyse de données massives, Data Science, BIG DATA et cloud computing, Intelligence Artificielle
- Capacité de gérer un projet BIG DATA
- Augmentation de l'employabilité pour les personnes en insertion, en reconversion ou en recherche d'emploi, dans tous les secteurs d'activités aussi bien Publics que Privés
- Accès à de meilleures possibilités d'évolution de carrière pour un salarié
- Obtention d'un certificat de compétences Professionnel CCP Big data attestant l'acquisition du savoir et savoir-faire d'un niveau expert



### POUR LES ENTITÉS UTILISATRICES

- Garantie pour les entreprises d'avoir un personnel opérationnel et expert en Bid Data et Data Scientist ou Science des données
- Intégration de nouvelles compétences et technologies en pleines évolutions
- Autonomie et réactivité pour accomplir les tâches ainsi que l'indépendance vis-à-vis des prestataires externes
- Maîtrise du budget système d'informations et gestion des données massives et la garantie d'un meilleur retour sur investissement
- Capacité à anticiper les besoins et l'évolution de technologies en matière de récolte, de stockage et d'analyse des données volumineuses
- Motivation et fidélisation des salariés en sécurisant leurs parcours professionnels