



CERTIFICAT  
BIG DATA ENGINEER

Diplôme d'Université

# Big Data Engineer

#certificatBDE

Ouverture septembre 2019



CCM Benchmark  
INSTITUT

utt  
UNIVERSITÉ DE TECHNOLOGIE  
TROYES



Cours à  
Paris

Partenariat  
CCM  
Benchmark  
Institut - UTT

Programme  
certifiant

Formation  
dispensée  
en temps  
partagé

Blended  
Learning

A mesure que l'exploitation des grandes masses de données devient un outil incontournable dans la stratégie des entreprises, le technicien d'exploitation Big Data (Big Data Engineer) est un métier en plein développement et très demandé. Il accompagne les Architectes Big Data dans le dimensionnement et le déploiement des infrastructures et environnements applicatifs nécessaires aux traitements, et met en œuvre les algorithmes définis par les Data Scientists. Il administre les systèmes en production en veillant à leur approvisionnement en données, leur optimisation et leur évolution (scalabilité, passage à l'échelle, maintenance) afin de garantir la qualité de la chaîne de valorisation mise en place pour les analyses et la décision. Il veille à faire évoluer les systèmes afin qu'ils soient en adéquation avec la croissance de l'entreprise et sa stratégie.

Le Certificat « Big Data Engineer » délivré sous la forme d'un Diplôme d'Université, proposé par l'UTT et CCM Benchmark Institut, répond aux besoins des entreprises et des administrations publiques face aux défis qu'induisent les grandes masses de données et la gestion de projets complexes de collecte, d'entreposage et de valorisation de ces données.

## OBJECTIFS PÉDAGOGIQUES ET PROFESSIONNELS

Les participants développeront au travers de cette formation les compétences qui leur permettront de :

- Comprendre les enjeux et les règles de gouvernance des projets Big Data.
- Identifier et dimensionner les outils pour la conception d'architecture Big Data.
- Maîtriser les outils de l'écosystème des architectures distribuées en batch et en temps réel.
- Développer des applications de collecte, déployer les algorithmes et les solutions de visualisation des données.

## TECHNOLOGIES UTILISÉES

Apache Hadoop • Spark • Kafka • PostgreSQL • MongoDB • BigQuery • Elasticsearch • Kibana

## PROGRAMME DE FORMATION

Le Certificat Big Data Engineer compte 154 heures d'enseignement en présentiel, soit 22 jours de formation. Des activités pédagogiques en ligne correspondant à 26 heures d'enseignement seront organisées entre chaque séquence de cours en présentiel (*blended learning*).

Un projet « fil rouge » vient valider les compétences mises à l'œuvre, en prenant appui sur une problématique ou un projet de l'entreprise dans laquelle le participant exerce ses activités, depuis la phase de compréhension du problème jusqu'au déploiement en passant par la préparation des données et leur traitement.

La formation est dispensée à Paris dans les locaux de CCM Benchmark Institut (94 rue de Provence, Paris, 9<sup>e</sup>), à raison de 2 jours par mois.

## Modules :

Introduction au Big Data	<ul style="list-style-type: none"> <li>• L'écosystème du Big Data : technologie, solutions, entreprises</li> <li>• Méthodologies et process de mise en œuvre : dimensions, scalabilité</li> <li>• Le Big data en entreprise : interactions entre défis technologiques, dimensions et métiers [ EN DISTANCIEL ]</li> <li>• Enjeux et gouvernance de la donnée en entreprise : acteurs, sécurité, éthique</li> </ul>	2 JOURS 24 et 25 septembre 2019
Bases de données avancées	<ul style="list-style-type: none"> <li>• Les différents formats de données : comparatif des différents types de bases de données (structurées, semi-structurées, et non structurées)</li> <li>• Architecture des différentes bases NoSQL et technologies de données distribuées [ EN DISTANCIEL ]</li> <li>• Identification et déploiement des formats de bases de données les plus pertinents pour son projet et ses traitements (NoSQL, graphes, etc)</li> </ul>	3 JOURS 14, 15 et 16 octobre 2019
Ecosystème Hadoop	<ul style="list-style-type: none"> <li>• L'écosystème Hadoop (langage, stockage, modèle de calcul, administration, streaming)</li> <li>• Architecture de projets Big Data</li> <li>• Conception de projets Big Data hautement parallélisés [ EN DISTANCIEL ]</li> <li>• Manipulation de données avec Hive &amp; Flume</li> </ul>	4 JOURS 21 et 22 novembre 2019 18 et 19 décembre 2019
Intelligence artificielle et Machine Learning	<ul style="list-style-type: none"> <li>• Introduction au Data Mining</li> <li>• Mise en forme des données : Feature Engineering</li> <li>• Mise en œuvre des principaux algorithmes de machine learning dans un contexte de projet Big Data</li> <li>• Industrialisation des modèles et passage à l'échelle (scalabilité)</li> </ul>	3 JOURS 22, 23 et 24 janvier 2020
Applications Big Data en mode batch	<ul style="list-style-type: none"> <li>• Développement et optimisation d'applications avec Apache Spark</li> <li>• Création d'applications distribuées avec Java en se basant sur le paradigme MapReduce [ EN DISTANCIEL ]</li> <li>• Automatisation et orchestration des traitements avec Oozie</li> </ul>	4 JOURS 24 et 25 mars 2020 23 et 24 avril 2020
Applications Big Data en mode streaming	<ul style="list-style-type: none"> <li>• Déploiement et gestion d'un cluster pour traiter les données en streaming</li> <li>• Gestion des applications en temps réel avec Kafka [ EN DISTANCIEL ]</li> <li>• Développement d'applications avec Apache Spark Streaming</li> </ul>	4 JOURS 26 et 27 mai 2020 15 et 16 juin 2020
Conception et gestion d'un projet Big Data	<ul style="list-style-type: none"> <li>• Projet Big Data : comprendre ses spécificités technologiques, organisationnelles et juridiques</li> <li>• Mise en place et pilotage d'une équipe projet Big Data [ EN DISTANCIEL ]</li> <li>• Problématiques de protection et de sécurité des projets Big Data</li> </ul>	2 JOURS 2 et 3 juillet 2020
Projet fil rouge (projet de fin d'études)	• Dépôt et validation des sujets de projets	Du 25 nov. au 8 déc. 2019
	• Dépôt des rapports de projets	Du 20 juillet au 10 août 2020
	• Soutenances	Sept. 2020 Date à définir

Le projet « fil rouge » donne lieu à soutenance devant un jury constitué des responsables pédagogiques et des intervenants du programme, à l'issue de la formation.



## Babiga BIRREGAH

*Enseignant-chercheur, Co-responsable du Certificat « Big Data Engineer »  
Département Recherche Opérationnelle, Statistiques Appliquées et Simulation*

« Le Diplôme d'Université Big Data Engineer vous offre l'opportunité d'approfondir vos connaissances en informatique décisionnelle et dans le domaine du Big Data. Il est spécialement conçu pour répondre aux besoins des entreprises et des administrations publiques face aux défis qu'induisent les grandes masses de données. Au travers de cette formation, nous vous équipons pour répondre aux besoins du marché pour adresser des projets complexes de collecte, d'entreposage et de valorisation de données. »



## Damien MANGIN

*CTO, CCM Benchmark & Chief Data Officer, Groupe Figaro  
Co-responsable du Certificat « Big Data Engineer »*

« La technologie nous permet depuis quelques années de mieux exploiter, plus rapidement et de manière plus performante, la data. Jusqu'à présent, les entreprises mesuraient l'importance de la data mais n'avaient pas les moyens de l'exploiter. Le certificat Big Data Engineer permet de se former aux dernières technologies et méthodes innovantes. »

<https://certificat-big-data-engineer.ccmbenchmark.com>

## Cours à Paris

94 rue de Provence  
75 009 Paris



## PUBLIC CONCERNÉ

La formation est destinée aux ingénieurs d'étude, de recherche et développement, consultants en business intelligence & SI, développeurs ou encore statisticiens, data scientists et data analysts souhaitant acquérir des compétences dans le Big Data. Certaines compétences techniques sont pré-requises (pratique d'un langage de programmation objet, bases en statistiques et probabilités, connaissances en réseaux informatiques et en base de données).

Les candidats devront être titulaires au moins d'un diplôme Bac+3 avec expérience professionnelle ou équivalent.

ADMISSION SUR DOSSIER  
ET LETTRE DE MOTIVATION

## PRIX

11 900 € HT (14 280 € TTC)

pour les 22 jours de formation en présentiel et les activités pédagogiques en ligne correspondant à 26h d'enseignements qui auront lieu entre chaque cours présentiel.

## INFORMATIONS & CANDIDATURE

Pour obtenir un conseil, un renseignement sur cette formation ou un devis, vous pouvez nous contacter par téléphone au 01 47 79 51 08 ou par e-mail à l'adresse [formation@ccmbenchmark.com](mailto:formation@ccmbenchmark.com)