

## IBM Certified Data Engineer - Big Data

CATEGORIE : B

### Vue d'ensemble

Domaine(s) d'activité professionnel dans lequel(s) est utilisé la certification :

Transverse : ■ **Transverse et tout secteur d'activités**

Tout type d'entreprise ou association ayant plusieurs postes informatisés, et entreprises du secteur informatique à partir d'une personne

Code(s) NAF : **62.02A**

Code(s) NSF : **326**

Code(s) ROME : **M1802**, **M1805**, **M1806**, **M1810**

Formacode : **31054**

Date de création de la certification : **11/05/2015**

Mots clés : **Ingénieur Big Data**, **Traitement de données**, **Analyse de Données**, **Big Data**

### Identification

Identifiant : **3127**

Version du : **06/11/2018**

### Références

Consensus, reconnaissance ou recommandation :

Formalisé :

- [Détails sur les sites :](#)
- [\\*](#)
- [, rubrique certification IBM](#)
- [La recherche s'effectuera ensuite par mot-clé ou n° de test ou de certification sur le site en anglais des certifications IBM.](#)

Non formalisé :

- **N/A**

Norme(s) associée(s) :

—

### Descriptif

#### *Objectifs de l'habilitation/certification*

Nous évoluons actuellement dans un monde connaissant une croissance exponentielle des données et, parallèlement, des flux de données.

Ces données proviennent de partout : de messages sur les sites de médias sociaux, d'images numériques et de vidéos publiées en ligne, d'enregistrements transactionnels d'achats en ligne, de signaux GPS de téléphones mobiles, de capteurs utilisés pour collecter les informations climatiques, pour ne citer que quelques sources.

Le concept de **Big Data** représente un ensemble de techniques concernant des données qui correspondent à plusieurs des caractéristiques suivantes : volume massif, grande variété ainsi que grande vitesse.

Les données, structurées (ex. : fichiers d'un logiciel), ou non structurées (ex. : galeries de photos), jouent un rôle prédominant auprès des entreprises qui les capturent, les gèrent et les analysent. Ces connaissances leur permettent de prendre rapidement des décisions métier bien informées.

La certification **IBM Certified Data Engineer - Big Data** s'adresse aux ingénieurs Big Data. L'Ingénieur Big Data travaille en étroite collaboration avec l'Architecte Data et les développeurs qui ont une expérience pratique. Grâce à lui, les données définies pour le système-cible et modélisées par l'Architecte Big Data se

concrétisent. L'Ingénieur Data maîtrise parfaitement au niveau technique, sur le plan théorique et pratique, une large gamme de produits et de technologies.

Il résout, en outre, les problèmes Big Data en appliquant les technologies adéquates et détient l'aptitude requise pour construire des systèmes de traitement des données à grande échelle pour les entreprises.

Les ingénieurs Data développent, testent et qualifient des solutions Big Data au sein des organisations, tout en s'assurant que l'infrastructure reste fluide et opérationnelle.

Ils fournissent des indications aux architectes sur les besoins en matériels et logiciels.

## *Lien avec les certifications professionnelles ou les CQP enregistrés au RNCP*

- 15719
- 13082
- 12051
- 12112

## *Descriptif général des compétences constituant la certification*

A l'issue de sa certification **IBM Certified Data Engineer - Big Data**, le spécialiste sera en mesure de démontrer les compétences suivantes :

Réunir, analyser, administrer de vastes ensembles de données, afin de fournir aux Data Scientists les données adéquates et les outils visuels qui leur serviront pour les analyses.

Saisir la complexité des données et être en mesure de traiter différentes sortes de données (structurées, semi-structurées, non structurées), ainsi que différents volumes, vitesses (y compris traitement des flux) et véracités de données.

Répondre aux problèmes posés par la gouvernance de l'information et aux défis associés aux données dans le domaine de la sécurité.

Posséder de solides connaissances en génie logiciel et une grande expérience de la programmation et du cryptage des données.

La journée type de l'Ingénieur Big Data est constituée de 50% de développement, 20% d'étude et spécification, 20% d'accompagnement et résolution de bugs et 10% de veille.

## *Modalités générales*

Le candidat peut préparer cette certification grâce à des lectures web gratuites disponibles sur internet, ainsi qu'à des ressources en ligne.

Une formation sur-mesure assurée par un formateur pourra également être proposée.

## *Liens avec le développement durable*

Aucun

## Public visé par la certification

- multiples : salariés, demandeurs d'emploi

## Valeur ajoutée pour la mobilité professionnelle et l'emploi

### *Pour l'individu*

Le professionnel certifié **IBM Certified Data Engineer - Big Data** développe sa crédibilité en mettant en place et en améliorant les outils permettant le stockage, le traitement et l'analyse de données volumineuses, que ce soit en batch ou en streaming.

### *Pour l'entité utilisatrice*

Les entreprises sont plus nombreuses à stocker, traiter et exploiter leurs données, quelles qu'en soient la forme et le volume. Les systèmes qui prennent en charge de grandes quantités de données, structurées et non structurées, continuent à se développer. Les plates-formes doivent

Il valide des compétences techniques en développement et en déploiement de solutions d'envoi et de récupération de données. Il maintient et assure le suivi de l'évolution de l'architecture Big Data en vue de traiter des volumes de données toujours plus importants. Il a bien assimilé par ailleurs les enjeux business de l'entreprise.

Le Big Data et les métiers liés à ce domaine sont particulièrement émergents en France actuellement. Certains géants sont à la pointe et à l'affût de talents. La montée des technologies numériques dans le monde et la quantité massive de données qu'elles génèrent, crée également des emplois liés au Big Data.

Qu'ils s'agissent de start-ups, ou d'entreprises de toutes tailles, elles manifestent déjà un vif intérêt pour ces nouveaux profils. De ce fait, le professionnel certifié **IBM Certified Data Engineer - Big Data** acquiert une expertise transposable dans de multiples environnements de travail et dans de très nombreuses entreprises.

permettre aux personnes qui s'occupent des données d'assurer la gouvernance et la sécurité du Big Data et en même temps donner aux utilisateurs finaux les moyens d'analyser ces données. Une fois arrivées à maturité, ces plate-formes s'intègrent sans difficulté aux systèmes et aux normes IT des entreprises.

Dans une même entreprise, de plus en plus de départements sont concernés par la mise en place et l'utilisation du Big Data : commercial, marketing, tous les domaines de production selon les secteurs (banques, télécoms, énergie, santé, ...) Le Big Data est notamment mis à profit par les professionnels du marketing pour affiner leur ciblage et analyser toutes les facettes du comportement du consommateur.

Le professionnel certifié **IBM Certified Data Engineer - Big Data** mettra son expertise au service de son entreprise en développant et en entretenant les systèmes de collecte, stockage et mise à disposition de données sécurisées, les aidant ainsi à concevoir, développer et mettre en place une stratégie couronnée de succès tournée vers la performance.

## Evaluation / certification

### Pré-requis

Préalablement à sa préparation à la certification, le professionnel devra détenir les connaissances et aptitudes suivantes :

Couche de données et ses zones d'enjeux/de risques potentiels  
Capacité à transformer les exigences fonctionnelles en spécifications techniques  
Capacité à aboutir à une architecture physique en partant d'une solution globale ou d'une architecture logique  
Gestion des clusters  
Exigences réseau  
Principales interfaces  
Modélisation des données  
Capacité à identifier pour la solution les exigences non fonctionnelles et à apporter un soutien technique  
Latence, flexibilité, haute disponibilité, réplication et synchronisation des données, restauration après sinistre  
Performance globale (performance des requêtes, gestion de la charge de travail, optimisation de la base de données)  
Proposer les meilleures pratiques ou les plus recommandées en matière de déplacement, manipulation et stockage des données dans une solution Big Data  
Options techniques d'ingestion des données  
Options et ramifications du stockage de données, dont exigences supplémentaires et enjeux imposés par les données dans le Cloud

### Certificateur(s)

- IBM

### Centre(s) de passage/certification

- De nombreux centres Pearson Vue France entière.

Techniques de requêtes de données et leur disponibilité en matière de soutien à l'analyse.

Généalogie des données et gouvernance des données

Diversité des données et volume des données

Mettre en place la sécurité des données et fournir des orientations dans ce domaine, conseiller en particulier sur la sécurité LDAP, les fonctions/la sécurité des utilisateurs, le monitoring des données, la sécurité des données à caractère personnel

Logiciels à connaître en priorité : BigInsights - BigSQL - Hadoop - Cloudera (NoSQL)

Logiciels à connaître accessoirement :

Information Server - Intégration à partir de BigInsights - Balanced Optimization pour Hadoop - Fonction Push Down de JAQL - Gouvernance des données - Spécificités de BigInsights liées à la sécurité - Metadata Workbench - Intégrer Optim avec BigInsights - DataClick pour BigInsights - BigMatch - Guardium (monitoring) - Outils d'analyse des données (SPSS) - BigSheets - Support technique de Hadoop/BigInsights - Disponibilité des données et support dans leur requête - Flux de données - Interface/intégration avec BigInsights - Données de streaming - Analyse de données intégrée - Netezza - DB2 BLU - Bases de données Graph - Machine Learning (système ML)

## Compétences évaluées

Les compétences particulières suivantes, spécifiques au développement (50% de son activité), seront évaluées :

### Architecture et Intégration

Implémenter MapReduce

Opter pour Hive, Big SQL ou HBase en fonction des scénarii

Créer un index Solr ou en faire la requête

Sélectionner les formats de fichiers appropriés (JSON, CSV, Parquet, Sequence)

Visualiser les recherches au travers d'Apache View

### Chargement des données

Charger des données non structurées dans InfoSphere BigInsights, importer des données de streaming dans Hadoop avec InfoSphere Streams

Créer un classeur BigSheets

Importer des données dans Hadoop et créer des définitions de tableaux Big SQL

Importer des données dans Hbase et dans Hive

Charger des données depuis des sources relationnelles dans InfoSphere BigInsights avec Data Click et en mode libre-service

Extraire des données d'une source relationnelle en utilisant Sqoop

Charger des données de fichiers log dans Hadoop avec Flume

Insérer des données via l'API du système de fichiers Posix IBM General Parallel File System (GPFS)

Charger des données avec l'utilitaire de ligne de commande de Hadoop

### Préparation, Transformation et Exportation des données

Transformer les données dans InfoSphere BigInsights en utilisant les techniques de requêtes Jaql

Capturer et préparer les données issues des médias sociaux en vue de leur analyse

Intégrer le scoring de modèle d'IBM SPSS dans InfoSphere Streams

Mettre en oeuvre une résolution d'entité dans une plate-forme Big Data, par ex. Big Match

Transformer et manipuler des données à partir de PIG

Transformer des données dans InfoSphere BigInsights à partir de Big SQL

Exporter de Hadoop les résultats du traitement des données (DataClick, DataStage)

Utiliser des zones homogènes dans InfoSphere Streams afin de permettre au moins un traitement

## **Sécurité des données**

Sécuriser les données dans les standards PCI

Protéger les données sensibles avec des masquages (Optim, Big SQL)

## **Performance et Evolutivité**

Améliorer les performances de MapReduce à l'aide de RDD (Resilient Distributed DataSet)

Sélectionner les formats de fichiers les plus adaptés pour optimiser les performances de Big SQL et de JAQL

Prendre des décisions spécifiques visant à optimiser les performances de Hive et de HBase

Analyser les facteurs de performance lors de l'utilisation d'Apache Spark

*Niveaux délivrés le cas échéant (hors nomenclature des niveaux de formation de 1969)*

N/A

La validité est Permanente

**Possibilité de certification partielle :** non

Matérialisation officielle de la certification :

Un Certificat électronique est envoyé par IBM aux candidats ayant réussi, ainsi qu'une carte professionnelle lors du 1er examen réussi. Le candidat accède à son transcript à tout moment sur un portail sécurisé dédié.

## Plus d'informations

### *Statistiques*

En France, depuis 2015, environ **800 personnes ont obtenu une certification IBM dans le domaine de l'Analyse.**

Le programme de certification professionnelle IBM existe depuis plus de 15 ans et compte entre 1000 et 2000 nouveaux certifiés par an.

### *Autres sources d'information*

<http://home.pearsonvue.com/IBM>

<http://www-03.ibm.com/certify/>

Lien de la certification :

<https://www.ibm.com/certify/cert?id=50001501>