

**Certificat de compétence : Mise en place de traitements et d'architectures Big Data
Business & Decision University – Février 2020**

| REFERENTIEL DE COMPETENCES | |
|--|--|
| Activités et tâches | Compétences Associées |
| 1. TESTER ET CORRIGER LE FONCTIONNEMENT DES INFRASTRUCTURES BIG DATA : DATA LAKE ET DATA HUB | 1.1 Savoir identifier la bonne infrastructure les différents types d'infrastructure Big Data et DataScience 1.2 Maîtriser le fonctionnement de l'écosystème Hadoop 1.3 Maîtriser le fonctionnement de l'écosystème Spark 1.4 Maîtriser le fonctionnement des bases de données NoSql 1.5 Maitriser le fonctionnement d'une infrastructure cloud majeure |
| 2. SPECIFIER ET PROGRAMMER LES TRAITEMENTS DES DONNEES DANS LE DATA LAKE/DATA HUB | 2.1 Savoir programmer en Scala 2.2 Savoir programmer avec Hive 2.3 Savoir utiliser le logiciel Kafka 2.4 Connaître les bases de programmation d'un ETL 2.5 Savoir utiliser le logiciel Nifi |
| 3. OPTIMISER LES TRAITEMENTS DES DONNEES DANS LE DATA LAKE/DATA HUB | 3.1 Mettre en œuvre la mesure de la performance des traitements sur l'infrastructure cible 3.2 Savoir paralléliser les traitements sur infrastructure Spark 3.3 Pouvoir identifier les goulots d'étranglement et les traitements critiques 3.4 Savoir programmer en Scala avec Spark ML 3.5 Savoir programmer en PySpark avec Spark ML |

**Certificat de compétence : Mise en place de traitements et d'architectures Big Data
Business & Decision University – Février 2020**

| REFERENTIEL DE CERTIFICATION | |
|--|---|
| 1. Modalités et déroulement des sessions d'évaluation | <p>Les modalités d'évaluation de ce certificat s'articulent autour d'exercices pratiques, d'un questionnaire à choix multiples et d'un projet pédagogique :</p> <ul style="list-style-type: none">- Les compétences les plus conceptuelles et théoriques font l'objet uniquement de questionnaires à choix multiples ;- Les compétences les plus pratiques font l'objet :<ul style="list-style-type: none">o d'exercices qui permettent de valider que l'apprenant sait les mettre en application. Ainsi, ces compétences sont évaluées à la fois par des questionnaires à choix multiples et des travaux pratiques ;o d'un projet pédagogique permettant de mettre en pratique et d'articuler l'ensemble des compétences du certificat <p>- Exercices pratiques Les exercices pratiques sont réalisés pendant l'action de formation. Le résultat attendu est la validation de la réalisation intégrale et complète de chaque exercice par le formateur. Finir l'exercice pratique permet de valider que la compétence évaluée est acquise</p> <p>- Questionnaire à choix multiples Les questions permettent à l'apprenant de valider qu'il a bien assimilé les compétences techniques. Pour chaque question :<ul style="list-style-type: none">- 4 réponses possibles- 1 seule bonne réponseObtenir un score supérieur à la moyenne permet de valider que la compétence évaluée est acquise</p> <p>- Un projet opérationnel pédagogique Au début de la formation, les candidats recevront une information sur le projet à réaliser, la constitution des groupes et le planning de réalisation du projet pédagogique. Le thème du projet sera communiqué pendant la formation lorsque les candidats auront abordé les compétences nécessaires à la réalisation du projet.</p> |

**Certificat de compétence : Mise en place de traitements et d'architectures Big Data
Business & Decision University – Février 2020**

2. Modalités et critères d'évaluation

| Compétences évaluées | QCM | Travaux pratiques | Projet Pédagogique |
|---|-----|-------------------|--------------------|
| 1. TESTER ET CORRIGER LE FONCTIONNEMENT DES INFRASTRUCTURES BIG DATA : DATALAKE ET DATAHUB | | | OUI |
| 1.1 Savoir identifier la bonne infrastructure les différents types d'infrastructure Big Data et DataScience | OUI | | |
| 1.2 Maîtriser le fonctionnement de l'écosystème Hadoop | OUI | OUI | |
| 1.3 Maîtriser le fonctionnement de l'écosystème Spark | OUI | OUI | |
| 1.4 Maîtriser le fonctionnement des bases de données NoSql | OUI | OUI | |
| 1.5 Maitriser le fonctionnement d'une infrastructure cloud majeure | OUI | OUI | |
| 2. SPECIFIER ET PROGRAMMER LES TRAITEMENTS DES DONNEES DANS LE DATALAKE/DATAHUB | | | |
| 2.1 Savoir programmer en Scala | OUI | OUI | |
| 2.2 Savoir programmer avec Hive | OUI | OUI | |
| 2.3 Savoir utiliser le logiciel Kafka | OUI | OUI | |
| 2.4 Connaître les bases de programmation d'un ETL | OUI | OUI | |
| 2.5 Savoir utiliser le logiciel Nifi | OUI | OUI | |
| 3. OPTIMISER LES TRAITEMENTS DES DONNEES DANS LE DATALAKE/DATAHUB | | | |
| 3.1 Mettre en œuvre la mesure de la performance des traitements sur l'infrastructure cible | OUI | | |
| 3.2 Savoir paralléliser les traitements sur infrastructure Spark | OUI | | |
| 3.3 Pouvoir identifier les goulots d'étranglement et les traitements critiques | OUI | | |
| 3.4 Savoir programmer en Scala avec Spark ML | OUI | OUI | |
| 3.5 Savoir programmer en PySpark avec Spark ML | OUI | | |

La certification est délivrée si les deux conditions suivantes sont satisfaites :

- Une note globale égale à 70% de la note obtenue au QCM + 30% de la note obtenue au projet pédagogique supérieure à 10/20
- Accomplissement des exercices pratiques de chacun des modules