

Mise en place d'architectures et de traitements Big Data

RS5341 · RS · Publiée

L'essentiel

Code(s) NSF	326t, 326
Formacode(s)	31023, 31009, 31067
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	20/01/2021
Date d'échéance	19/08/2025

Certificateur(s)

DIGITAL SERVICES SCHOOL · SIRET : 38183776400095<https://ecoledeladataetia.com>**ORANGE · SIRET : 38012986648625****BUSINESS ET DECISION UNIVERSITY · SIRET : 38183776400087****ORANGE · SIRET : 38012986646850**

Résumé de la certification

OBJECTIFS ET CONTEXTE

Depuis les cinq dernières années, les projets Data Science et d'intelligence artificielle connaissent une forte croissance dans des domaines très variés, tels que l'industrie pétrolière, la téléphonie, la distribution, la banque, la finance et les services. Les récentes évolutions technologiques permettent de faire évoluer ces projets vers la mise en place de véritables solutions d'intelligence artificielle. La réalisation de ces projets nécessite de former des équipes de « Data-Engineers » et de « Data Scientists » pour lesquelles il est nécessaire de conjuguer l'acquisition des compétences techniques et méthodologiques, l'expérience des cas concrets et surtout, un parcours d'apprentissage jalonné.

Les principales compétences visées par cette certification sont celles nécessaires à la mise en œuvre des activités suivantes :

- Tester et corriger le fonctionnement des infrastructures big data : datalake et datahub
- Spécifier et programmer les traitements des données dans le datalake/datahub
- Optimiser les traitements des données dans le datalake/datahub

COMPÉTENCES ATTESTÉES

1. Savoir identifier la bonne infrastructure les différents types d'infrastructure Big Data et DataScience
2. Maîtriser le fonctionnement de l'écosystème Hadoop
3. Maîtriser le fonctionnement de l'écosystème Spark
4. Maîtriser le fonctionnement des bases de données NoSql
5. Maîtriser le fonctionnement d'une infrastructure cloud majeure
6. Savoir programmer en Scala
7. Savoir programmer avec Hive
8. Savoir utiliser le logiciel Kafka
9. Connaître les bases de programmation d'un ETL
10. Savoir utiliser le logiciel Nifi
11. Mettre en œuvre la mesure de la performance des traitements sur l'infrastructure cible
12. Savoir paralléliser les traitements sur infrastructure Spark
13. Pouvoir identifier les goulots d'étranglement et les traitements critiques
14. Savoir programmer en Scala avec Spark ML

Voies d'accès

MODALITÉS D'ACCÈS

Formation continue

COMPOSITION DU JURY — FORMATION CONTINUE

Le jury est présidé par un data-scientist confirmé. Il établit la composition du jury qui sera composée d'au minimum 4 consultants-experts des domaines couverts par le certificat et du directeur de Business & Decision University. Les membres du jury peuvent être des formateurs du certificat, mais dans tous les cas plus de 50% des membres du jury seront des personnes extérieures à l'organisation et si des formateurs participent au jury, leur voix sera strictement consultative

Base légale

Date de décision	20/01/2021
Durée de l'enregistrement	5 ans
Date d'échéance de l'enregistrement	19/08/2025

Pour plus d'informations

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rs/5341/>

Demande CPF — formations engagées

Dossiers CPF engagés	27
Montant total engagé	37 250 €
Coût moyen par dossier	1 380 €
Durée moyenne	113 h
Période	06/2021 – 12/2021
France compétences	92 %
État	6 %
Reste à charge titulaire	2 %

Formations engagées via le CPF — Caisse des Dépôts (open data), tous organismes confondus.