



DataScientest

Trouvez gratuitement le profil idéal

Focus sur la POEI Cloud Data Engineer



DATASCIENTEST

C'EST :

UNE ÉQUIPE
POLYVALENTE

4K

De personnes formées

35

Professeurs

95

Collaborateurs

UN PARCOURS
HYBRIDE

Enseignement
asynchrone



Enseignement
synchrone

2000

Heures de contenu certifié
par l'École des Mines ParisTech

L'HISTOIRE DE NOTRE ORGANISME DE FORMATION

- Fondé in 2016, **DataScientest** est un acteur **spécialisé** de la formation B2B et B2C dans les Data sciences
- Nos utilisateurs bénéficient d'une plateforme de formation entièrement pensée et conçue par nos experts. Notre communauté compte près de 4000 formés issus de plus de 70 grandes entreprises partenaires
- Nos principaux modules de **formation** permettent de dispenser des formations aux métiers de **Data Analyst, Data Scientist, Data Engineer, et Data Management Officer.**

LE LEADER FRANÇAIS DE LA FORMATION EN DATASCIENCE



PARTENARIATS EDITEURS



aws

→ DataScientest est depuis peu **parmi les 4 organismes de formations partenaires d'Amazon web services en France**. AWS est spécialisé dans les services de cloud computing à la demande pour les entreprises et particuliers.

→ Notre équipe de spécialistes Cloud computing forme sur 6 certifications allant d'un niveau débutant à professionnel.



→DataScientest bénéficie du statut de **Microsoft Learning Partner**, nous sommes ainsi habilités par Microsoft à former des équipes sur les produits et services de l'entreprise américaine.

→Nous avons créé plusieurs **cursus de formation** qui attestent d'un niveau d'expertise sur Azure, un ensemble de produits et services de **cloud computing** ; et sur **Power BI**, l'outil de **Business Intelligence de Microsoft**.

→Nous sommes dans **le club très fermé de seulement une dizaine d'organismes de formation** pouvant faire passer les certifications Microsoft.

PARTENARIAT ACADEMIQUE



À l'issue de la formation, chaque apprenant se voit à la fois opérationnel et bénéficiant d'un diplôme délivré par une école d'ingénieur de rang mondial



LE DISPOSITIF DE LA POEI

La POEI est un **dispositif de Pôle Emploi** permettant de **recruter totalement gratuitement** des profils répondant exactement à vos attentes.

Dans le cadre d'une telle action, **vous définissez vous-mêmes l'expérience professionnelle et académique que vous attendez** des candidats.

Ces derniers sont ensuite **filtrés selon vos critères par nos soins** avant de vous être présenté pour un **job dating** au cours duquel vous validerez ou non chaque profil.

En parallèle du processus de recrutement, **nos équipes pédagogiques co-construisent avec vos équipes techniques le contenu de la formation**. A l'issue de celle-ci, les personnes formées répondent donc exactement à vos attentes.



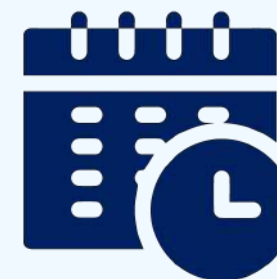
100% gratuite



Profils opérationnels
immédiatement



Sourcing et job dating



Calibrage de la formation



POURQUOI CHOISIR DATASCIENTEST ?

Pour certains métiers, comme celui de Cloud Data Engineer, **il n'existe pas de formation initiale**. Ce profil nécessite donc **une formation** précédée d'**une certaine expérience**, souvent en dev (Java, DBA...), avant de pouvoir prétendre à une reconversion vers ce métier. Le marché fait donc face à **une véritable pénurie de Cloud Data Engineer** que nous vous proposons de **recruter sans aucun frais**.

Notre expertise dans le domaine de la POEI repose sur 3 piliers :

UN SOURCING MASSIF ET UNE SELECTION MINUTIEUSE

- Plus d'**1 million de visiteurs annuels** sur notre site et de **nombreux médias partenaires**.
- Plus de **10 000 abonnés sur LinkedIn** et près de **3 000 inscrits à nos newsletters**.
- Une équipe dédiée au **screening des CV**.
- Un **test d'éligibilité corrigé à la main** par nos Data Engineers
- Plus de **100 entretiens de motivation réalisés chaque jour** par notre équipe admission (15 personnes à temps plein).

UN CURSUS CERTIFIE ET PERSONNALISABLE SELON VOS BESOINS

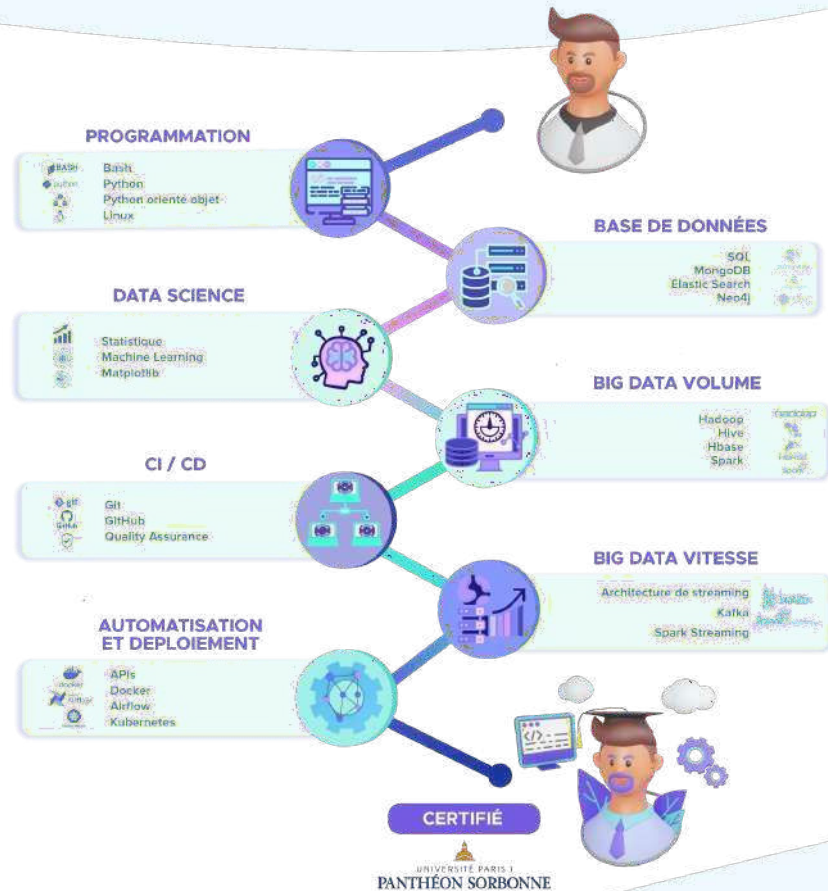
- Seul organisme à être **partenaire officiel à la fois de Microsoft et AWS**, nous vous proposons d'inclure des **certifications adaptées à votre environnement** (y compris GCP).
- Nos Data Engineers coconstruisent un **parcours de formation sur-mesure** avec vos équipes techniques
- Une **certification de l'Ecole des Mines ParisTech** délivrée l'issue de la formation

UNE EXPERIENCE INEGALEE AUPRES DES GRANDS GROUPES

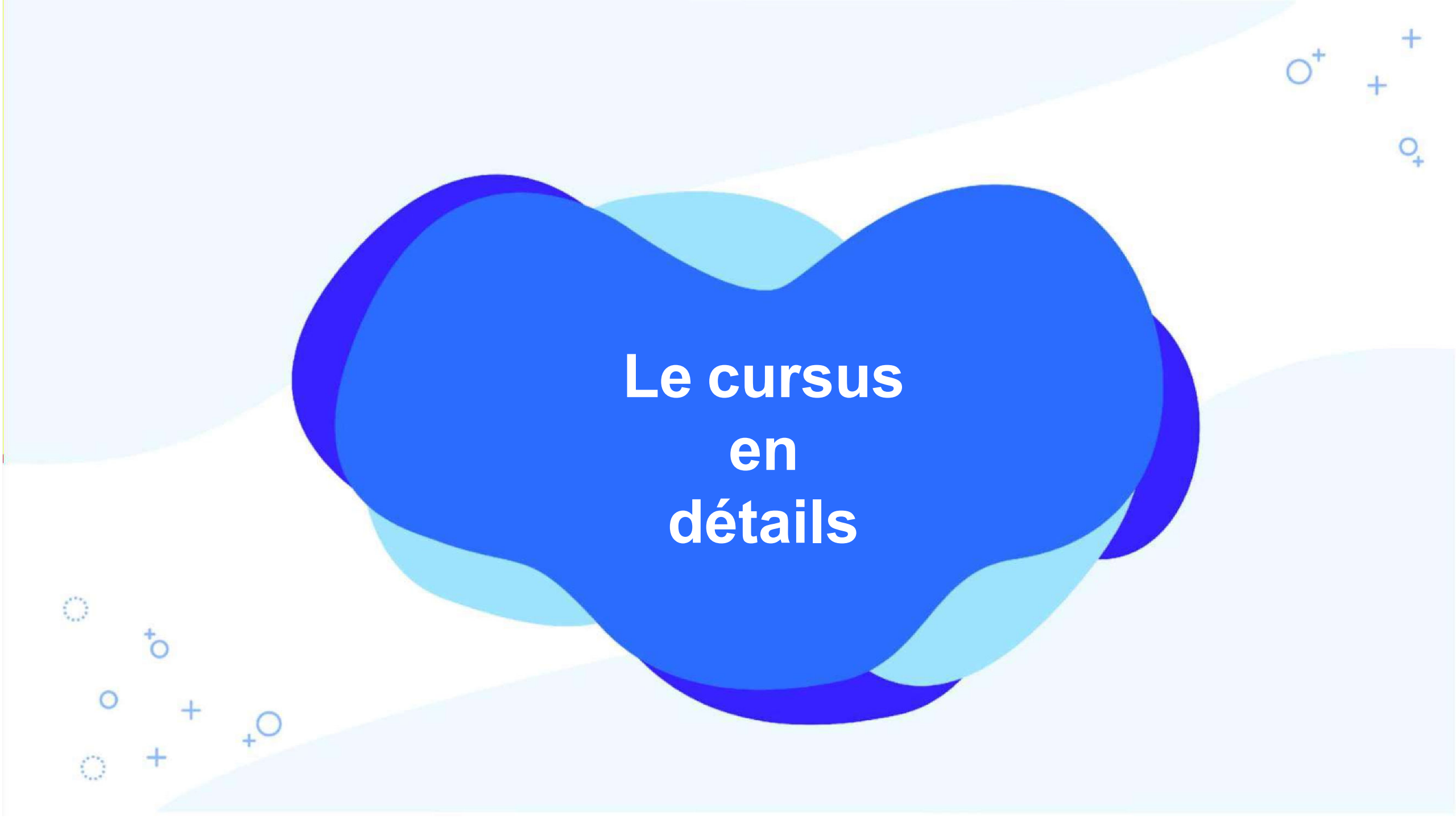
- Plus de **70 grands groupes partenaires** dont plus de 30 du CAC 40.
- Relation privilégiée avec **Pôle Emploi**.
- Des **Data Engineers employés à temps pleins** et anciens responsables **Data de groupes** comme Capgemini ou Devoteam.
- Des **mini-projets pratiques** tout au long de la formation construits des **use-cases issus du monde de l'entreprise**.



NOTRE PROPOSITION DE FORMATION



- Concevoir et développer des logiciels et des applications
- Maîtriser les **données du SI**
- Analyser et représenter **des données**
- Assurer la **qualité de la donnée**
- Mettre en œuvre la gestion et la gouvernance des données
- Être sensibilisé à la veille technique et aux innovations et mises à jour des outils utilisés
- Maîtriser les bases de données



Le cursus en détails

PROGRAMMATION

Durée : 40h

LES COMPÉTENCES ACQUISES À L'ISSUE

SYSTÈME LINUX & SCRIPT BASH

- Présentation des Systèmes Linux
- Prise en main et utilisation d'un terminal
- Mise en place de scripts Bash

PYTHON & PYTHON ORIENTÉ OBJET

- Maîtrise des variables et des types
- Présentation des divers opérateurs et de leurs applications
- Introduction au concept de boucles et aux structures de contrôle
- Définition d'une fonction sur Python et ses applications
- Initiation aux classes et modules
- Préparation de la mise en place, du paramétrage et de l'enchaînement de Décorateurs
- Différenciation et implémentation du multithreading et du multiprocessing sur Python
- Application d'une fonction asynchrone sur Python
- Introduction aux annotations et utilisation de la bibliothèque MyPy

- Maîtriser le système d'exploitation Linux
- Apprendre à utiliser un Terminal
- Créer et gérer des exécutable Bash
- Maîtriser le langage Python et toutes ses applications
- Comprendre et utiliser la programmation orientée objet
- Créer des scripts complexes avec Python



Bases de données

Durée : 50h

SQL

- Introduction aux bases de données relationnelles
- Initiation aux bases du langage SQL
- Présentation de SQL Alchemy et applications
- Approfondissement de SQL et de ses applications

MONGODB

- Introduction aux bases de données NoSQL (base de données orientée document, colonne, graphe)
- Présentation de MongoDB
- Familiarisation avec la syntaxe des requêtes MongoDB

NEO4J

- Introduction aux bases de données orientées graphe
- Mise en place d'un premier graphe
- Initiation au langage de requête Cypher
- Chargement de données dans Neo4J
- Utilisation d'un client Python pour Neo4J

ELASTICSEARCH

- Description d'un moteur de recherche
- Présentation d'un index et mode d'emploi
- Mise au point d'un Mapping
- Découverte des différentes opérations
- Prétraitement des données avec Ingest Node
- Extraction des données avec les Texts Analyzer

Déploiement de bases de données sur Azure



- Présentation des bases de données propriétaires Microsoft Azure
- Déploiement d'une base de données relationnelle Azure SQL Database
- Déploiement d'une base de données Azure PostgreSQL managée
- Introduction à Azure Cosmos DB, base de données NoSQL multi modèle

LES COMPÉTENCES ACQUISES À L'ISSUE

- Savoir choisir un système de gestion de base de données en fonction du cas d'utilisation
- Comprendre la notion de schéma et leur implémentation dans une base de données relationnelle
- Comprendre comment requêter un SGBDR (système de gestion de base de données relationnelle) avec le langage SQL
- Manipuler une base de données orientée document comme MongoDB
- Améliorer la recherche de vos données textuelles à l'aide d'Elasticsearch
- Gérer une base de données orientée graphe
- Utiliser Cypher pour requêter des bases de données orientées graphe
- Connaître les principales bases de données déployables sur Azure



Data Science

Durée : 50h

LES COMPÉTENCES ACQUISES À L'ISSUE

STATISTIQUES

- Exploration des variables numériques
- Exploration des variables catégorielles
- Étude des relations entre les variables

MACHINE LEARNING

- Prétraitement de données.
 - Sélection et optimisation d'un algorithme de Machine Learning
 - Définition et application d'un algorithme de régression
 - Définition et application d'un algorithme de classification
- Développement d'algorithmes de Clustering
Introduction au PCA

DATA VIZ AVEC MATPLOTLIB

- Présentation de différents types de graphes:
 - Courbes
 - Graphiques
 - Nuage de points
 - Histogrammes
 - Boîtes à Moustaches (Box Plots)
 - Camemberts (Pie Plots)

- Comprendre les bases des principaux algorithmes de Machine Learning
- Être directement opérationnel en machine learning
- Entraîner des modèles de machine learning avec la bibliothèque ScikitLearn
- Manipuler vos données avec des Dataframes Pandas
- Maîtriser Numpy
- Visualiser vos données dans des graphiques variés grâce à Matplotlib



Big data: volume

Durée : 50h

LES COMPÉTENCES ACQUISES À L'ISSUE

HADOOP ET HIVE

- Fonctionnement de Hadoop
- Installation et configuration de Hadoop
- Traitement et stockage des données avec HDFS
- Présentation de MapReduce
- Utilisation de Hadoop Streaming pour exécuter un fichier Map/Reduce
- Mise en place d'entrepôts de données
- Présentation du fonctionnement de Hive

HBASE

- Présentation des bases de données orientée colonne
- Association de Hadoop (HDFS) et de Hbase
- Requêtes de données
- Modification des données par Python et happybase

SPARK

- Distinction entre Spark et Hadoop
- Introduction au calcul distribué avec Spark
- Présentation des APIs RDD et Dataframe de Spark
- Pipeline de processing de donnée distribuée avec Pyspark
- Machine Learning distribué avec Spark MLlib

- Appréhender les concepts fondamentaux du Big Data Comprendre la théorie sur les architectures des systèmes distribués
- Stocker et traiter de manière distribuée des données avec Hadoop distributed file system (HDFS)
- Maîtriser les principaux outils pour la gestion des Big Data:
 - Hadoop
 - Hive
 - Hbase
 - Spark



CI / CD

Durée : 25h

LES COMPÉTENCES ACQUISES À L'ISSUE

GITHUB

- Découverte de la plateforme Github pour le travail collaboratif sur Git
- Présentation des fonctionnalités majeures de GitHub :
 - Fork
 - Pull Request
 - Issues
- Partager ses modifications avec pull et push
- Participation à l'amélioration de projet public (open source)
- Présentation des principaux workflows git

GIT

- Introduction au système de gestion de version Git
- Initialisation d'un dépôt Git
- Présentation et approfondissement des concepts git :
 - Branches
 - Tag
 - Merge

QUALITY ASSURANCE

- Mise en place de tests unitaires avec Pytest
- Introduction aux Tests d'intégration et leurs fonctions
- Présentation des avantages des tests : gain de temps, lisibilité, qualité et amélioration de code

- Maîtriser les outils de versionning
- Travailler de manière collaborative et versionner les projets avec Git et GitHub
- Être capable de mettre en place des tests unitaires
- Appliquer des méthodes adaptées en fonction des différentes problématiques
- Vérifier le fonctionnement d'unités de code indépendantes lors du développement



Big Data Vitesse

Durée : 25h

LES COMPÉTENCES ACQUISES À L'ISSUE

KAFKA

- Présentation de la plateforme de streaming distribuée Kafka :
 - Architecture
 - Avantages
- Gestion des paramétrages de Producers :
 - Clef de partitionnement
- Maîtrise des paramétrages de Consumers :
 - Consumer group

SPARK STREAMING

- Prise en main de Spark Streaming pour le traitement de données temps réel
- Présentation du minibatch streaming nécessaire pour le fonctionnement de Spark Streaming

ARCHITECTURE DE STREAMING

- Gestion de flux de données en temps réel
- Conception d'une architecture Big Data hybride (batch et temps réel)
- Mise en place d'une architecture Lambda

- Comprendre comment gérer les flux de données en temps réel
- Mise en place et gestion d'architecture de streaming
- Traiter des données en temps réel
- Maîtriser le logiciel Kafka
- Traiter et transformer des données en temps réel, de manière distribuée avec Spark Streaming



Automatisation et Déploiement

Durée : 40h

APIs

- Introduction aux APIs et découverte des architectures micro services
- Présentation des différentes méthodes HTTP et de leurs fonctions
- Utilisation des bibliothèques FastAPI et Flask pour développer des API RESTful
- Documentation d'une API avec la spécification OpenAPI
- Gestion des erreurs et des performances d'une API

AIRFLOW

- Découverte des concepts d'Airflow:
 - Présentation des principes de l'orchestration et utilité
 - Graphe orienté acycliques ou DAG (Directed Acyclic Graphs)
 - Opérateurs
- Gestion de tâches par le biais d'Operators spécifiques
- Monitoring des DAGs via l'interface graphique d'Airflow

Docker

- Présentation de la conteneurisation et de son utilité par rapport à la virtualisation
- Initiation au fonctionnement de Docker
- Manipulation des images et des conteneurs
- Communication avec les conteneurs
- Persistance des données grâce aux Volumes
- Création d'une image Docker via un Dockerfile
- Partage des images sur le Dockerhub
- Utilisation de dockercompose

KUBERNETES

- Déployer et gérer des conteneurs

LES COMPÉTENCES ACQUISES À L'ISSUE

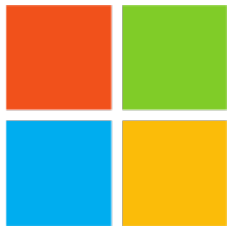
- Comprendre les APIs
- Apprendre à créer une Api avec Flask et FastApi
- Requête une API HTTP
- Automatiser vos tâches grâce à la maîtrise d'Apache Airflow
- Comprendre la virtualisation
- Maîtriser les techniques et outils de conteneurisation et d'orchestration des conteneurs





**Les certifications qui
peuvent être incluses
au cursus**

DP-203 Azure Data Engineer



Spécialisation faisant partie intégrante du **tronc commun**

Objectifs

- Implémenter des bases de données relationnelles et non relationnelles
- Fournir un accès sécurisé aux données (cryptage et data masking)
- Analyser des flux de données en temps réel
- Suivi et optimisation des solutions déployées
- Valider la certification officielle DP-203

MasterClass

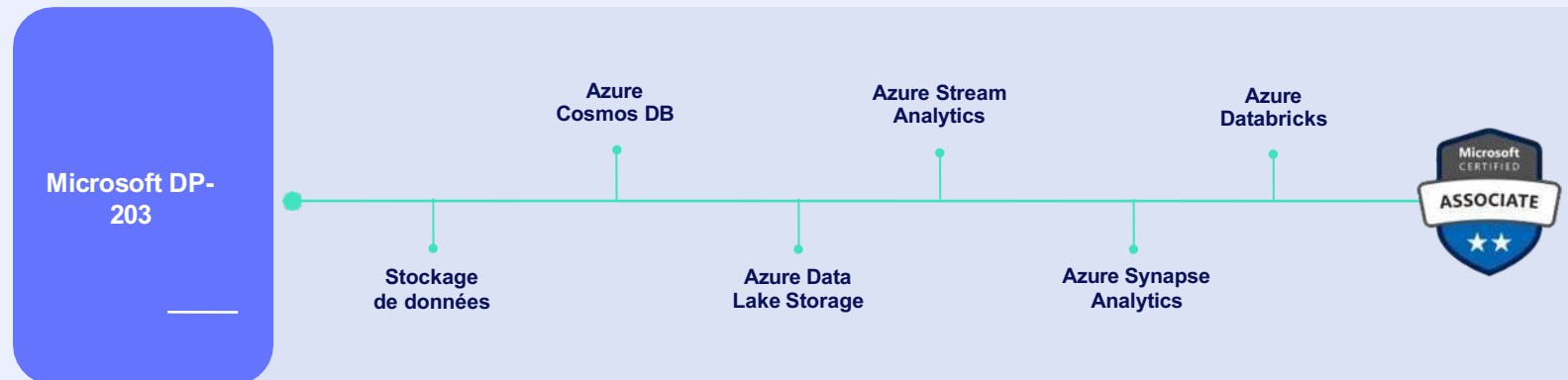
- Séance de Kick-off
- **14 heures** de MasterClass
- Séance de restitution
- Séance de préparation et examens blancs

28 h

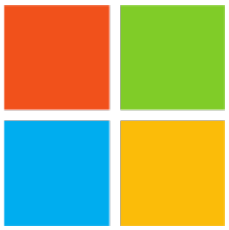
E-learning
asynchrone



50%



Le parcours formation DP-203



DP-203 atteste de votre expertise

La validation de DP-203 vous permet d'implémenter de manière éthique et responsable **des solutions Azure dans une entreprise**. Vous apprendrez à suivre toutes les solutions déployées afin d'évaluer leur performance, fournir une puissance de calcul et un stockage adapté. **Cette certification permet d'obtenir le titre de "Data Engineer Associate"**.

Microsoft Azure étant utilisé par de nombreux groupes, **cette certification est un réel avantage pour trouver un emploi**.

Les profils de Data Scientist sont de plus en plus nombreux et cette certification peut vous différencier d'autres profils lors de vos candidatures.

Introduction à Azure

Présentation de Azure
Découverte de Stockage Azure
Stockage Blob Azure



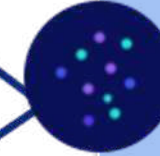
Gestion et traitement des données B.

Développement de l'engineering de donnée avec Azure Databricks
Développement de solutions de batch processing avec Data Factory et Azure Databricks
Traitement de données



Gestion et traitement des données A.

Appliquer l'Engineering données avec Azure Synapse Analytics et Apache Spark
Intégration de donnée avec Azure Data Factory ou Azure Synapse
Application de Azure Synapse Analytics
Ingérer des données et transformer des données avec Apache Spark



Suivi des solutions déployées pour le stockage et le traitement

Suivi des bases de données relationnelles et non-relationnelles
Suivi des pipelines Data Factory
Chargement et sécurisation des données dans Azure Data Lake Storage
Activer une messagerie fiable pour des applications Big Data avec Azure Event Hubs



DA-100 Analyzing Data with Microsoft Power BI



Spécialisation faisant partie intégrante du **tronc commun**

Objectifs

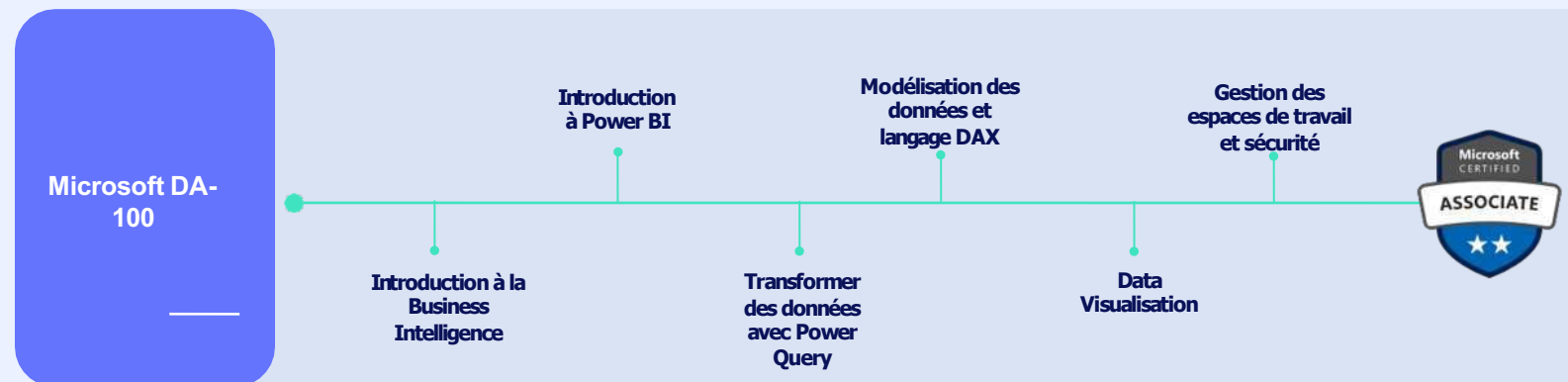
- Maîtriser les concepts fondamentaux de la **Business Intelligence**
- **Récupération** et **transformation** données
- Utilisation d'outils spécifiques : **Power BI, Power Query**
- **Modélisation des données** grâce au **Langage DAX**
- Comprendre le pipeline data jusqu'à visualisation
- Mise en place de **KPIs** pour le pilotage de projets
- Valider **la certification officielle DA-100** (frais d'inscription inclus)

MasterClass

- Séance de Kick-off
- **10 heures** de MasterClass
- Séance de restitution
- Séance de préparation et examens blancs

35 h

E-learning
asynchrone



Spécialisation faisant partie intégrante du **tronc commun**

Objectifs

- Configurer un environnement de solution de cloud computing à l'aide de GoogleCloud Platform
- Planification, installation et configuration d'une solution de cloud computing à l'aide de Google Cloud Platform
- Configurer l'accès et la sécurité à l'aide de Google Cloud Platform

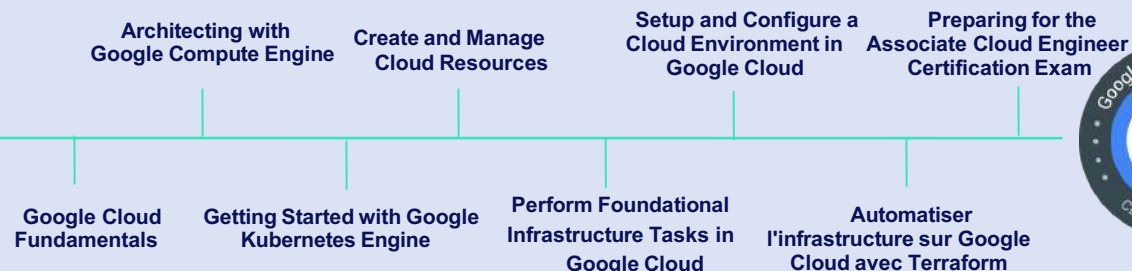
Prérequis

- Connaissance de base des produits et services de Google Cloud Platform
- Connaissance de base des concepts de cloud computing tels que les machines virtuelles, les conteneurs et les réseaux.

E-learning asynchrone  50%

21h

GCP Cloud Engineer



Certification AWS: Cloud Practitioner



Spécialisation faisant partie intégrante du **tronc commun**

Objectifs

- Définir ce qu'est le Cloud AWS et l'infrastructure globale de base
- Décrire les principes architecturaux du Cloud AWS
- Décrire les services clés de la plateforme AWS et leurs cas d'utilisation courants
- Aspect sécurité et conformité de base
- Modèle de facturation, de gestion de compte et de tarification

MasterClass

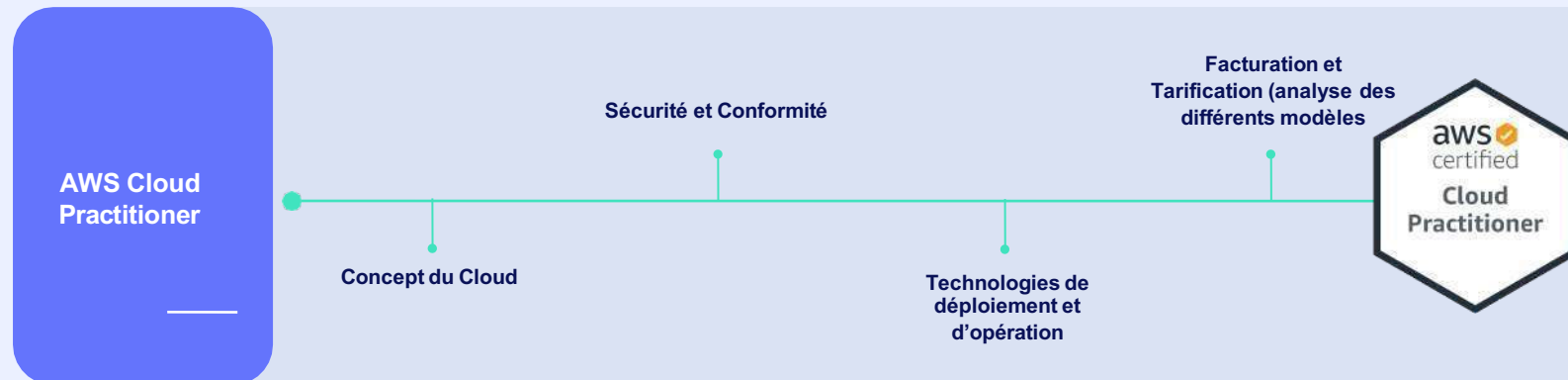
- Séance de Kick-off
- 100% présentiel
- Séance de restitution
- Séance de préparation et examens blancs

8 h

E-learning
asynchrone



70%



Spécialisation pour le métier de: **Concepteur Big Data**

Objectifs

- Définir les services AWS d'analyse de données et comprendre leur intégration mutuelle
- Développement évènementiel et modèle d'ingestion des données
- Obtenir une compétence dans la conception, le développement et le déploiement de solutions basées sur le cloud
- Mettre en œuvre un pipeline d'intégration et de déploiement en continu (CI/CD) pour déployer des applications sur AWS

MasterClass

- Séance de Kick-off
- 100% présentiel
- Séance de restitution
- Séance de préparation et examens blancs

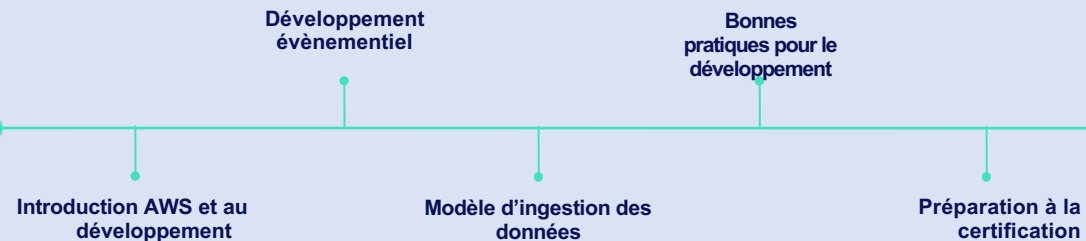
21 h

E-learning
asynchrone



70%

AWS Certified
Data Analytics



aws
certified

Developer

Associate



Spécialisation pour le métier de: **Architecte Cloud Devops**

Objectifs

- Capacité de démontrer de façon efficace que vous avez les connaissances pour réaliser l'architecture d'applications sécurisées et robustes
- Déployer les architectures sur les technologies AWS
- Définir une solution en utilisant des principes de conception architecturale basés sur les exigences client
- Fournir des conseils d'implémentation basés sur les bonnes pratiques à l'organisation

MasterClass

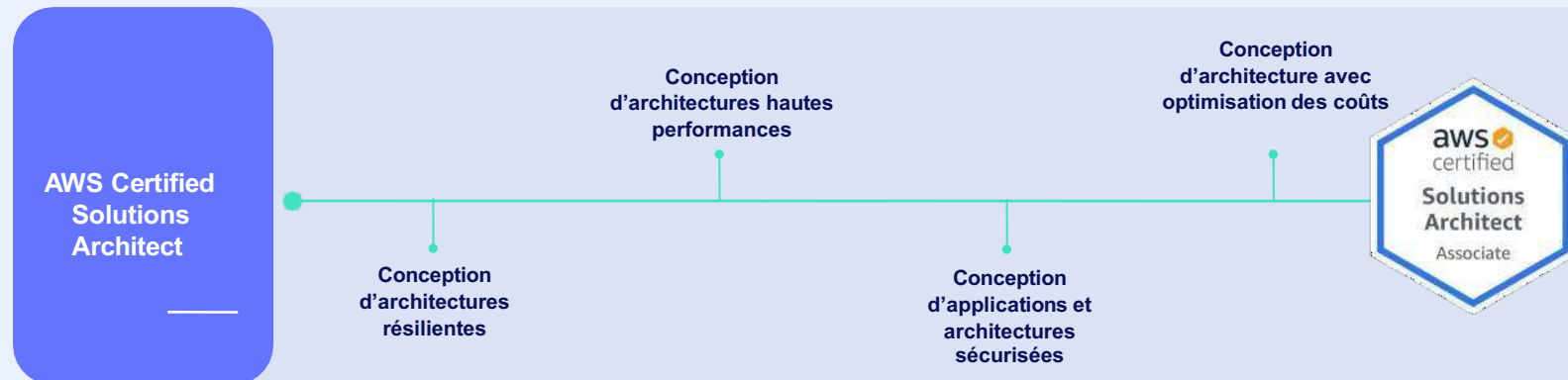
- Séance de Kick-off
- 100% présentiel
- Séance de restitution
- Séance de préparation et examens blancs

31 h

E-learning
asynchrone



70%



Spécialisation pour le métier de: **Administrateurs Cloud SysOps**

Objectifs

- Déployer, gérer, et exploiter des systèmes évolutifs
- Mettre en oeuvre et contrôler le flux de données depuis /vers AWS
- Sélectionner le service AWS approprié en fonction des exigences en termes de calcul, de donnée ou de sécurité
- Estimer le coût d'utilisation d'AWS et identifier les mécanismes de contrôle des coûts
- Migrer les charges de travail sur site vers AWS

MasterClass

- Séance de Kick-off
- 100% présentiel
- Séance de restitution
- Séance de préparation et examens blancs

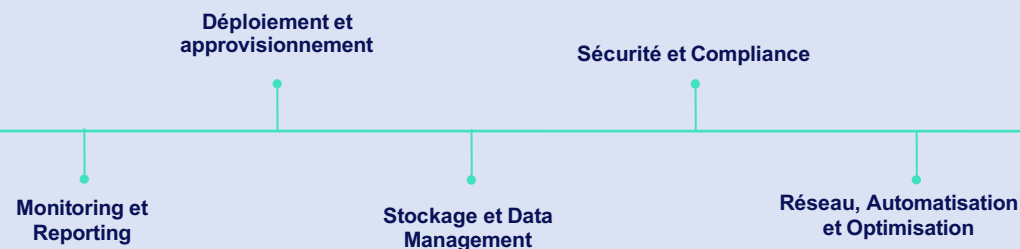
21 h

E-learning
asynchrone



70%

AWS Certified
SysOps
Administrator



Merci

*Des questions ?
N'hésitez pas à nous contacter*

