

Syllabus

Formation

Data Engineering

**Développer une architecture
Framework Big Data**



Devenez **l'expert en Data** que toutes les entreprises s'arrachent

Propulsez votre carrière en évoluant vers les **métiers de la Data les plus demandés**

Accédez à notre **réseau d'entreprises partenaires**

Bénéficiez d'un **suivi personnalisé** avec des mentors expérimentés

Formez-vous **où que vous soyez** grâce à notre plateforme

Découvrir





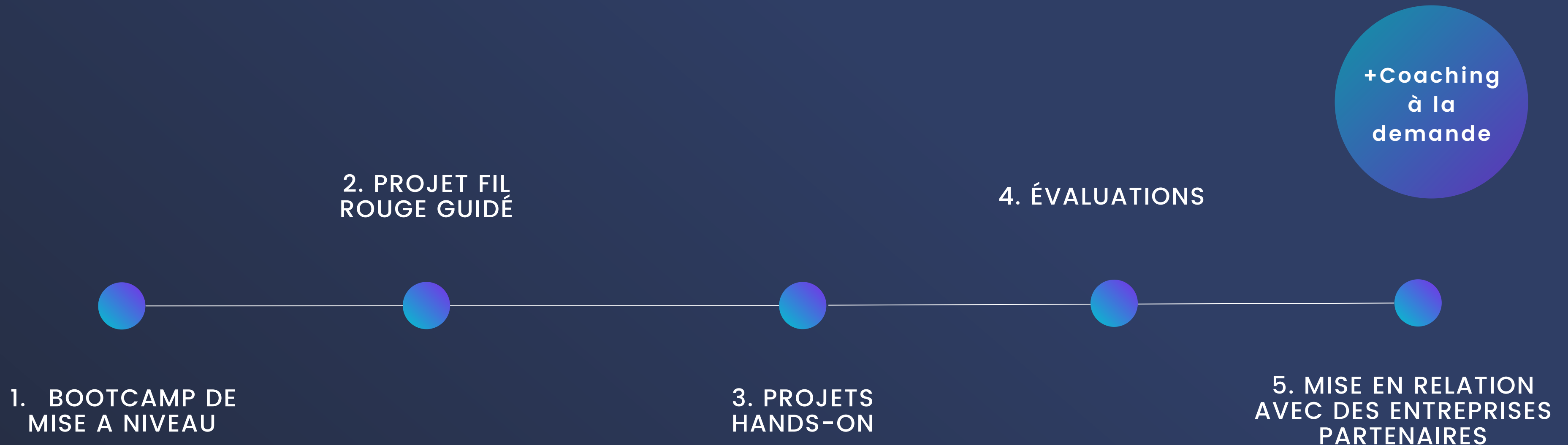
Une formation opérationnelle pour maîtriser toutes les compétences **Data Engineering** du calcul parallèle à l'automatisation en passant par le processing temps réel.

La pédagogie Blent



Un parcours pensé pour vous rendre opérationnel.le

La clé pour devenir réellement opérationnel.le est de pratiquer sur des projets qui correspondent à la réalité des équipes Data. Au sein de Blent, nous proposons un dispositif de formation totalement pensé par nos experts Data contenant :



Un parcours centré sur des projets

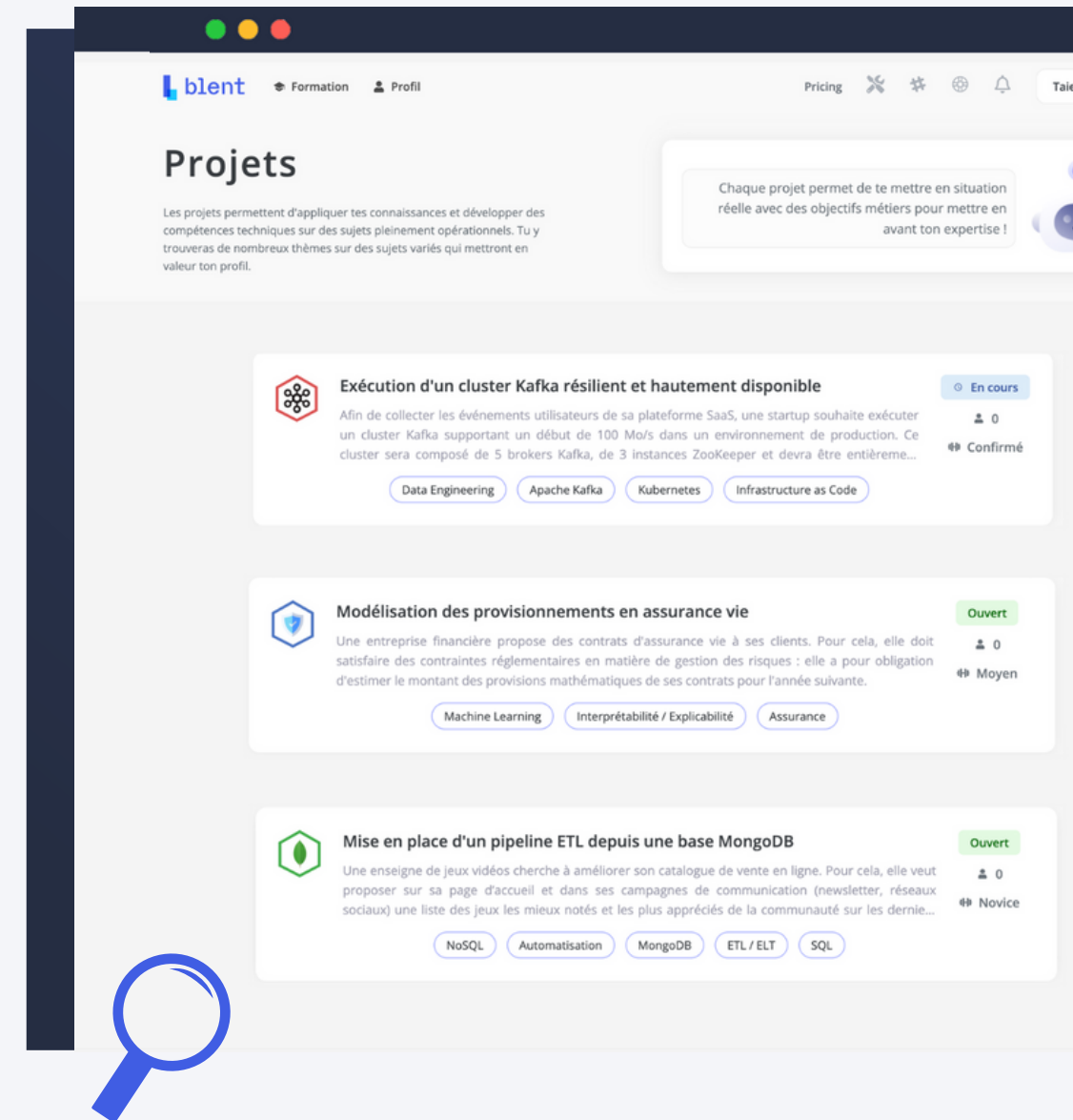
Notre pédagogie est basée sur des projets avec des données réelles d'entreprises. Dans le cadre de votre parcours, vous pourrez apprendre et pratiquer grâce à deux types de projets :

1 Un projet fil rouge guidé

Tout au long de la formation, vous allez dérouler les différentes parties de la formation avec du contenu varié et interactif (vidéos, cours, Notebooks interactifs, exercices, quiz). Lors de cette phase, vous apprenez à réaliser pas à pas un premier projet de Data Engineer en menant à bien les différentes phases d'un projet avec des données de Uber.

2 Les projets Hands-On

Des projets sur lesquels vous continuez à pratiquer avec plus d'autonomie et gagner en expérience. L'objectif est d'appliquer les connaissances acquises lors du projet fil rouge guidé en menant à bien un projet de A à Z et en faisant face à différentes contraintes métiers.



Exemple d'un projet Hands-On

Automatisation d'un pipeline ETL pour la consommation électrique

Une entreprise fournisseuse d'énergie souhaite acquérir de nouveaux clients depuis son site Internet.

Pour cela, l'entreprise cherche à proposer un formulaire gratuit en ligne permettant à chaque utilisateur visitant le site Web d'évaluer rapidement et simplement sa consommation.

En tant que Data Engineer, votre objectif est de :

- Automatiser le processus de récolte et de transformation de données journalières de la consommation électrique moyenne
- Déployer le service de calcul sur un serveur ou dans un service Serverless
- Réaliser des tests unitaires et fournir un benchmark sur l'efficacité technique du service de tarification

Contraintes :

- Le temps de calcul de chaque estimation doit être inférieur à 6 secondes.
- Les coûts générés par l'utilisation du formulaire doivent être inférieurs à 20€ / 1000 estimations.
- Les coûts de stockage et d'exécution sur le Cloud doivent être inférieurs à 100€ HT par mois.

Un format flexible adapté aux professionnels

Notre format Blended Learning intégrant du E-learning, du mentorat à la demande et un suivi par nos équipes tout au long de votre parcours vous permet d'avancer à votre rythme tout en vous donnant la possibilité d'exploiter au maximum les ressources que vous aurez à disposition.

- 1 Adaptez la formation à **votre emploi du temps** et suivez la formation où que vous soyez
- 2 **Notre équipe vous suit** pour s'assurer que vous ne perdez pas le rythme
- 3 Posez vos questions à n'importe quel moment grâce à notre **module de questions-réponses interactif** et via le **Slack de la communauté**
- 4 Bénéficiez de **séances de mentorat individuelles** à la demande tout au long de votre parcours

Plateforme 100% en ligne

Une **plateforme unique** dans la formation en Data

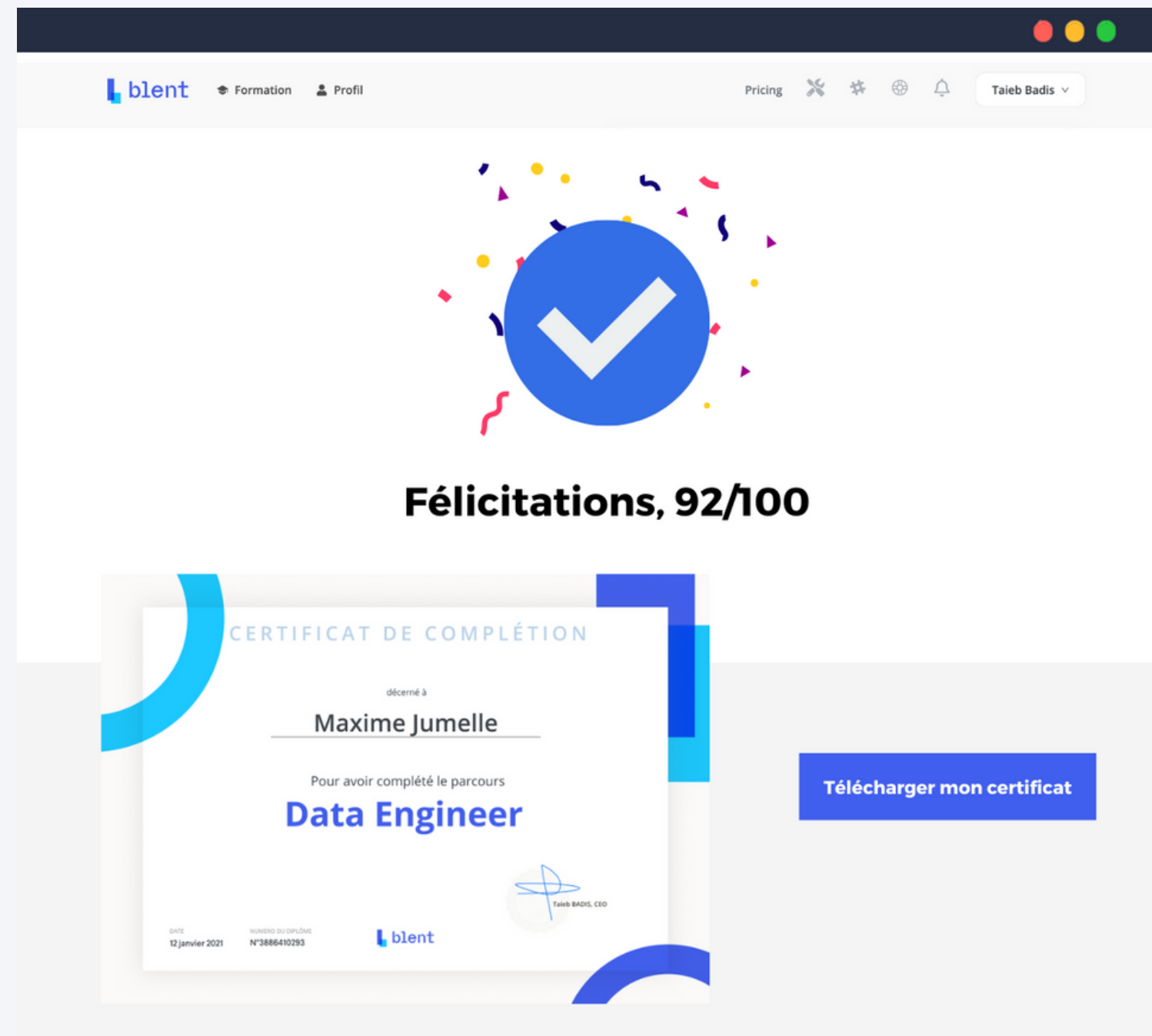
Notre plateforme intègre un environnement complet pour vous permettre de coder sans rien installer.

- 1 Accès à **tout le contenu de formation** pendant 12 mois
- 2 Des **accès illimités** aux principaux services Cloud
- 3 Des **serveurs dédiés pré-configurés** à la demande
- 4 Des environnements de code avec les **Workspaces**



Un Certificat reconnu sur le marché

À la fin de la formation, vous passerez une évaluation. Si vous la validez, nous vous délivrons un certificat reconnu sur le marché à faire valoir auprès de votre réseau.



Un portfolio de projets à valoriser

En tant que futur Data Engineer, il est extrêmement important de valoriser sur le marché les projets basés sur des données réelles sur lesquels vous avez travaillé, car cela démontre votre capacité à appliquer vos compétences en Data Engineering dans des situations réelles et à résoudre des problèmes concrets.



Mise en place d'un Data Warehouse sur Apache Hive

Création d'une base d'apprentissage ML à partir de données brutes

Exécution d'un cluster Kafka résilient et hautement disponible

Les entreprises qui recrutent nos alumni

Si vous le souhaitez, vous serez également référencé sur notre plateforme de mise en relation avec nos entreprises et recruteurs partenaires.



Programme de formation



Bootcamp de préparation



Lorsque vous êtes officiellement inscrit, nous vous donnons accès à 30h de e-learning pour que vous puissiez apprendre les bases en Python, aux commandes systèmes Bash et à la manipulation de données.

Cette phase de préparation est disponible dès votre inscription à la formation, et vous pourrez poser toutes vos questions à notre équipe de professeurs sur Slack.

Ce que vous allez apprendre dans cette phase

- Maîtriser le langage Python et son utilisation avancée pour les systèmes en production.
- Utiliser les commandes Bash et les principaux éléments utilisés dans les environnements UNIX.
- Utiliser Visual Studio Code en tant qu'environnement de développement intégré.
- Les bases du Cloud Computing et les concepts essentiels de virtualisation.

Ce que vous saurez faire à l'issue de cette phase

- Vous saurez écrire des programmes Python robustes pour les systèmes en production.
- Vous saurez utiliser les principaux outils opérationnels tels que Visual Studio Code, Bash ou encore Git.



Stockage de données



OpenRefine



Lors de cette première phase, nous aborderons le stockage des données, et notamment les différentes solutions qui existent. Nous étudierons ces solutions, tant sur le plan économique que technique, sur comment nous allons stocker des centaines de Go de données.

En plus des bases structurées classiques de la famille SQL ou du stockage persistant, nous découvrirons les bases de données orientées documents (MongoDB), les bases de données orientées colonnes distribuées (avec Cassandra) ainsi que les bases de données clé/valeur en mémoire dont Redis est le plus populaire. Enfin, une petite ouverture sur les bases NewSQL sera proposée pour être prêt face aux futures systèmes de gestion de bases de données.

- **Bases de données relationnelles**
 - Installation d'un serveur PostgreSQL et de pgAdmin 4
- **Bases orienté colonnes**
 - Création d'un ensemble de données sous BigQuery
 - Exécution de requêtes SQL-like
 - Insertion de données BigQuery via Pandas



Calcul parallèle



Une fois le stockage correctement établi, une des premières choses sur lequel le Data Engineer doit réfléchir est la bonne construction d'un processus ETL (pour Extract Transform Load) afin d'automatiser le processing de la donnée jusqu'à un Data Warehouse.

Du fait de la grande quantité de donnée à disposition, les outils de calcul parallèle, dont Hadoop et Spark sont les plus populaires, nous permettront d'effectuer un nombre impressionnant de calculs en un temps très court grâce aux clusters à disposition dans le Cloud. En complément des processus ETL, les applications dites serverless permettent aujourd'hui de lancer des calculs spécifiques sans se soucier de l'infrastructure cible.

- **Patron d'architecture MapReduce**
 - Étapes du MapReduce : de l'InputSplitter au Combiner
 - Implémentation d'algorithmes itératifs
- **Hadoop et Spark**
 - Gestion de ressources et YARN
 - Hadoop HDFS et Spark SQL
 - Cluster Spark sur DataProc : cycle de vie des clusters, exécution de tâches et monitoring
 - Drivers JDBC et BigQuery pour l'insertion de données



Data Streaming



CONFLUENT



PUBSUB



RabbitMQ

Aujourd'hui, beaucoup d'applications alimentent les bases avec des données en temps réel (ou à flot continu) : c'est le Data Streaming. Comment peut-on efficacement appliquer un processus ETL efficace ?

Pendant cette phase intermédiaire, nous mettrons en place plusieurs solutions pour le Data Streaming, avec notamment des architectures publish-subscribe avec Apache Kafka. Nous verrons également Confluent Platform, qui apporte de nouvelles fonctionnalités puissantes à Kafka.

- **Data Streaming avec Apache Kafka**

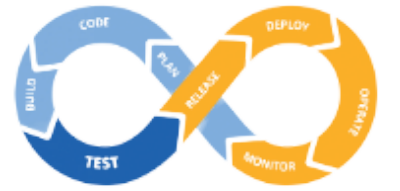
- Architecture publish-subscribe : consumers et producers, tolérance à la faute, topics et replicas.
- Installation d'un serveur Apache Kafka standalone sur une machine virtuelle.

- **Confluent Platform**

- Installation de Confluent Platform et utilisation de son API pour les appels à distance
- Registre de schéma et validation de données par sérialisation Avro.
- Connecteurs Kafka vers BigQuery



Automatisation et déploiement



Cette dernière étape est la plus importante : c'est elle qui définit comment les applications vont être intégrées dans une infrastructure Cloud.

Véritable institution dans le Cloud Computing, nous verrons la puissance de la conteneurisation d'applications avec Docker, et comment les solutions sur le Cloud permettent de répartir les charges de travail. De plus, les outils d'automatisation sont indispensables pour le Data Engineer : nous utiliserons Airflow, une référence en la matière.

- Automatisation avec Airflow
 - DAG, Executors, cycle de vie des tâches et Xcom
 - Pipelines de données automatisées
 - Entraînement continu de modèles et gestion de cycle de vie sous MLflow
- Déploiement sur le Cloud
 - Encapsulation de modèles de Machine Learning dans une API REST
 - Conteneurisation de l'API dans une image Docker
 - Exécution de conteneurs sur des orchestrateurs Cloud

La communauté Blent



Ce qu'en pensent nos Alumnis



Ismaila Djimera
Lead Data
Scientist @[Craft](#)

Le parcours Data Engineer est complet et axé sur la pratique ce qui permet une rapide montée en compétence. Il m'a permis de sortir de ma zone de confort de Data Scientist recevant des données presque propres.



Bangaly Camara
Senior Data Scientist & Analyst
Quantitatif @[Gradient](#)

Le Parcours Data Engineer de Blent est très complet. Il permet de comprendre toutes les composantes du métier à savoir : l'implémentation de Datalake et Data Warehouse, la collecte et le traitement de données, travailler avec des bases NoSQL, le traitement de flux continue de données avec Kafka et le déploiement d'algorithmes de Machine Learning.



Mehdi Olivier D.
Machine Learning
Engineer @[Soladis](#)

J'y ai réalisé 2 parcours de formation : Data Engineer et Machine Learning Engineer. Pour une personne qui a déjà les bases, c'est un régal de voir de vrais cas issus du monde réel avec les derniers outils. J'ai pu voir des choses que je n'avais jamais vu ailleurs.



Informations pratiques



Formation Data Engineering

Développer une architecture Framework Big Data

RÉSULTATS ATTENDUS :

Avoir la capacité d'obtenir le bloc 3 intitulé « Développement avancé : Comprendre et utiliser une architecture de type framework » de la certification intitulée « Développeur Web » délivrée par WEBECOM

OBJECTIFS PÉDAGOGIQUES :

- Confronter ses connaissances en développement avec la documentation d'un framework dans le but de s'approprier l'architecture et les fonctionnalités d'un nouveau cadre de travail.
- Configurer le framework en utilisant un gestionnaire de dépendances afin de préparer l'environnement de développement de l'application.
- Développer une application évolutive avec un framework afin de disposer d'un environnement de développement standardisé pour l'ensemble du projet.

PRÉREQUIS :

- Les critères d'acceptation ne prennent pas en compte le niveau scolaire mais, plutôt, la curiosité, la motivation, le projet professionnel, la rigueur et la logique ainsi que l'attrait auprès des nouvelles technologies. Les bases de l'anglais sont requises. Cette formation est destinée aux candidats les plus curieux et motivés.
- Un entretien individuel est effectué suite à chaque demande de candidature pour valider ces prérequis et évaluer le degré de motivation des candidats, si ce dernier est concluant, des tests d'aptitudes sont effectués basés sur de la logique, logique avancée, problèmes mathématiques et bases de l'Anglais.

Formation Data Engineering

Développer une architecture Framework Big Data

MODALITÉS D'ÉVALUATION :

- **Évaluation continue** : Réalisation d'une évaluation durant le parcours de formation sous forme d'un test en ligne sur notre plateforme d'évaluation.
- **Évaluation de fin de parcours (examens avec jurys externes)** : Mise en situation professionnelle sous forme de projet
 - Le candidat conceptualise une application dotée de fonctionnalités avancées à partir d'un cahier des charges et un cahier des fonctionnalités fournis ;
 - Lors de la soutenance auprès du jury de professionnels, le candidat présente son application fonctionnelle et déployée sur un serveur.

L'évaluation certificative a lieu la dernière semaine de la formation. L'obtention de la certification est conditionnée par le passage à un jury de certification intervenant dans les 3 mois suivants le jury d'évaluation. Pour l'année 2024, les prochains jury d'évaluation auront lieu le 21 mars 2024 ; le 6 juin 2024 et le 25 juillet 2024.

MODALITÉS DE SUIVI :

- Relevé de connexion à la plateforme et feuille d'émargement envoyée au stagiaire par mail
- Attestation de formation envoyée au stagiaire en fin de formation par mail
- Évaluation de la formation "à chaud" via des questionnaires de satisfaction sur la plateforme à la fin de chaque module
- Évaluation de la formation "à froid" via un questionnaire de satisfaction envoyé par mail à l'issue de la formation

Formation Data Engineering

Développer une architecture Framework Big Data

MOYENS D'ENCADREMENT :

Afin de faciliter la montée en compétences de nos stagiaires et le bon déroulé des parcours de formation, les apprenants bénéficient d'une assistance conseil assurée par l'intervenant(e). Nos intervenant(e)s assurent l'assistance pédagogique en mode synchrone pendant les classes virtuelles et en mode asynchrone via les échanges de mails ou de messages sur notre plateforme. L'apprenant peut également contacter l'accueil de Blent via l'adresse suivante : contact@blent.ai

MOYENS PÉDAGOGIQUES & TECHNIQUES :

- L'assistance technique est assurée par notre équipe de gestion de la Plateforme de formation.
- Cette formation à distance se dispense de la manière suivante :
 - Le mode synchrone réalisé via des classes virtuelles en web-conférences ou visio-conférence téléphone, permettant l'échange d'informations avec les autres apprenants et l'intervenant ainsi que du partage de documents en direct et en temps réel.
 - Le mode asynchrone réalisé par l'intermédiaire d'un contenu de formation elearning (auto-formation) et d'exercices diffusés sur la plateforme et à pratiquer dans son quotidien. Les échanges avec le(s) intervenants(s) et les apprenants s'effectuent via des échanges directement sur la plateforme ou par envoi de mails.
 - Comment se connecter sur la plateforme : le stagiaire reçoit un mail d'invitation qui lui permet d'accéder à la plateforme où il pourra compléter et personnaliser son espace dédié.

Formation Data Engineering

Développer une architecture Framework Big Data

MODALITÉS PÉDAGOGIQUES ADAPTÉES AUX PERSONNES EN SITUATION DE HANDICAP :

Si un candidat signifie qu'il est en situation de handicap, nous adaptons son parcours pédagogique. Les adaptations peuvent être de plusieurs natures, notamment une adaptation du parcours d'apprentissage et du référentiel d'évaluation en fonction des besoins identifiés de chaque personne bénéficiant du statut de travailleur handicapé. En fonction des demandes spécifiques, nous mettrons tout en œuvre avec nos partenaires spécialisés pour réaliser la prestation.

ACCESSIBILITÉS DE L'ACTION AUX PERSONNES EN SITUATION DE HANDICAP :

Nos aménagements sont évolutifs et individualisés en fonction de la situation de chacun des candidats. L'ensemble de nos temps synchrones sont disponibles sous-titrés. Les apprenants peuvent demander un tiers-temps sur les examens.

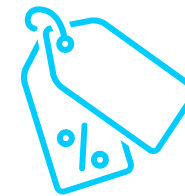
Pour plus d'informations à ce sujet vous pouvez consulter la page "[Accessibilité et handicap](#)" de notre site ou bien contacter directement notre référent handicap Charles Bourrat à l'adresse suivante : contact@blent.ai.

Frais de formation



Suivi mensuel

- Ressources accessibles pendant 12 mois
- Accès à un espace de Questions/Réponses dédié sur notre plateforme
- Workshops Live Q&A mensuels



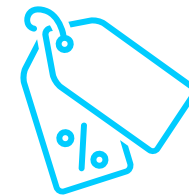
2500€

- ~~Séances de mentorat individuelles avec des Mentors expérimentés~~
- ~~Accompagnement pour appliquer directement sur vos propres cas d'usage~~

Frais de formation

Suivi hebdomadaire

- Ressources accessibles pendant 12 mois
- Accès à un espace de Questions/Réponses dédié sur notre plateforme
- Workshops Live Q&A mensuels



4600€

- Séances de mentorat individuelles avec des Mentors expérimentés
- Accompagnement pour appliquer directement sur vos propres cas d'usage

Financements

Vous disposez de plusieurs possibilités afin de financer votre formation avec Blent.

- Par un **financement public** via votre CPF ou par Pôle Emploi.
- Par un **financement de votre entreprise**, soit directement par celle-ci, soit en passant par un opérateur de formation OPCO, FIF PL ou AGEFICE.
- Par un **financement personnel**, où nous pouvons mettre en place un paiement en 3 fois sans frais.

Notre équipe Admissions est à disposition pour vous expliquer les procédures à suivre pour créer votre dossier de financement et commencer votre formation dans les plus brefs délais.



Prêt(e) à passer au **niveau
supérieur de la Data ?**



[NOUS CONTACTER](#)