

RNCP36739 – Expert en Ingénierie de Données

Session : 2023-2024

DOSSIER FINAL

Organisme formateur : Efrei Paris

Formation : Mastère Data Engineering & AI

Candidats :

Salaheddine HOUKMI - 20221124

Clément ROUANET - 20220481

RNCP36739BC01 : Concevoir et développer une architecture de stockage de données

RNCP36739BC02 : Concevoir, développer et déployer une solution de traitement des données massives

RNCP36739BC03 : Implémenter et optimiser des solutions de stockage et de traitement de données sur le cloud

RNCP36739BC04 : Implémenter des méthodes d'intelligence artificielle pour modéliser et prédire de nouveaux comportements et usages

RNCP36739BC05 : Concevoir une stratégie de management et de gouvernance de données pour transformer les données en informations créatrices de valeur

EMPLACEMENT DES PROJETS

Tous les projets cités ci-dessous, validant les blocs de compétences, sont disponibles sur le lien Git suivant :

[Dossier_RNCP36739-HOUKMI_ROUANET](#)

BLOCS 1 & 2

Concevoir et développer une architecture de stockage de données & Concevoir, développer et déployer une solution de traitement des données massives

| | |
|----|--|
| B1 | Concevoir et développer une base de données relationnelle en réponse aux besoins d'un client en vue de la mise à disposition de ses données structurées pour un tout en utilisant les technologies et les langages de requêtes adaptés aux développements envisagés |
| | Concevoir et développer une base de données non-relationnelle en vue de la mise à disposition des données semis-structurées et non- structurées pour un traitement analytique ou d'intelligence artificielle tout en utilisant les technologies et les langages de requêtes adaptés |
| | Concevoir et construire un lac de données « Datalake » en choisissant les architectures, les indicateurs de performance et les solutions de stockage appropriées afin d'intégrer les données provenant des systèmes d'information multivariés : Données non structurées, semi-structurées, structurées et des solutions de stockage de données massives ; redondantes. |
| | Créer une API en utilisant les technologies qui permettent de rendre accessible les données en augmentant l'efficacité et la praticité des applications et des services |
| B2 | Concevoir en s'appuyant sur une veille technologique et mettre en œuvre une architecture distribuée répondant au besoin du client pour traiter les données massives en entreprise en utilisant les technologies de traitement |
| | Implémenter un système distribué en utilisant des technologies de streaming identifiées à partir d'une veille pour traiter des données sur une période précise ou en temps quasi réel |
| | Transformer les données provenant de différentes sources en prenant en compte la variété de données pour faire de l'analytique à échelle (intégration, formatage, manipulation, stockage données multidimensionnelles) |
| | Optimiser la performance des pipelines en utilisant les techniques d'intégration et de mise en scène adéquates pour le traitement des données massives |
| | Automatiser la création, les tests, l'intégration et le déploiement des pipelines de données en s'appuyant sur une veille technologique qui permet d'identifier et de mobiliser les solutions pour maximiser l'efficacité et réduire le 'time to market' tout en utilisant les technologies de containerisation et d'ordonnancement. |

Projets :

- API Web Services
- Database NoSQL
- Projet de fin d'études

BLOC 3

Implémenter et optimiser des solutions de stockage et de traitement de données sur le cloud

Mettre en œuvre des solutions de stockage de données dans le cloud pour permettre aux entreprises d'explorer leurs données en implémentant des techniques et stratégies adaptées à leur utilisation et en choisissant les types de stockage et d'archivages appropriés

Concevoir et développer des pipelines et des solutions de traitement de données dans le cloud en chargeant les données en les transformant et les mettant à disposition des utilisateurs pour permettre leur exploitation

Mettre en place une politique de sécurité des données dans le cloud en développant une stratégie de chiffrement et de gestion des identités et des accès et protéger les données en étant et être conforme aux règlements en vigueur

Optimiser les solutions de stockage et de traitement des données dans le cloud en définissant des indicateurs de performance pour assurer la disponibilité des services et optimiser les coûts

Projets :

- Data Engineering on Cloud
- Projet de fin d'études

Certification :

- AZ-900 : Microsoft Certified – Azure Fundamentals

BLOC 4

Implémenter des méthodes d'intelligence artificielle pour modéliser et prédire de nouveaux comportements et usages

Extraire des données en provenance de systèmes d'information multivariés pour les exploiter et les analyser en utilisant les outils professionnels courants

Préparer les données en les transformant et les nettoyant pour faire l'analyse et le reporting selon les besoins des différents métiers

Élaborer une communication infographique visuelle en construisant des tableaux de bord interactifs afin de communiquer les résultats d'analyses et d'assurer l'extraction de connaissances en temps réel.

Développer un modèle prédictif pour identifier de nouveaux comportements et usages en implémentant des algorithmes d'apprentissage automatiques supervisés

Évaluer la performance d'un modèle de Machine Learning en analysant ses résultats et en comparant avec d'autres modèles afin d'implémenter la solution la plus convenable à un cas d'usage

Projets :

- Hackathon
- Data Visualisation
- Projet de fin d'études

BLOC 5

Concevoir une stratégie de management et de gouvernance de données pour transformer les données en informations créatrices de valeur

Définir une gouvernance de données en mettant en place les politiques et les standards afin d'établir les rôles, les responsabilités et la propriété des données

Mettre en place les bonnes pratiques de gestion de données en appliquant les réglementations en vigueur pour protéger les données et respecter la vie privée dans une logique de transparence.

S'assurer que la qualité de la donnée permet d'atteindre les objectifs business en faisant un audit de qualité et en proposant des plans de prévention et de remédiation en cas de non-qualité

Identifier l'approche stratégique pour faire de la donnée un actif central ancré dans la culture de l'entreprise et intégrer l'approche dans un plan de communication diffusé aux concernés

Élaborer une stratégie de sécurité afin de protéger les données en utilisant les technologies adéquates permettant de parer aux menaces et en mettant en place une politique d'accès aux données

Projets :

- Data Gouvernance
- Projet de fin d'études