

Mohamed AMTALSSI

Ingénieur BIG DATA

☎ 0778780036 — ✉ mohamed.amtalssi@hotmail.com — 📍 France

Éducation

ESIEA Laval

Ingénieur (Bac+5) en Big Data & IA

2017 - 2022

Expérience Professionnelle

Ingénieur Big Data à France Télévision Publicité

Depuis 08/2024

Contexte: Dans le cadre d'un projet de modernisation des infrastructures de données, j'ai été chargé de migrer et d'optimiser les processus de traitement de données vers la plateforme Microsoft Fabric.

Réalisations:

- PoC : Migration du Datawarehouse Oracle vers le Datawarehouse Microsoft Fabric.
- PoC : Conversion des pipelines Talend et ODI en Data Factory sur Fabric.
- Développement de nouveaux pipelines pour l'intégration des données AWS S3.
- Implémentation d'un système de chargement incrémentiel avec gestion des checkpoints(PySpark).
- Création de scripts PySpark pour le nettoyage et l'enrichissement des données.
- Réalisation de webscraping pour l'acquisition de données complémentaires.

Consultant Big Data à Orange Business

09/2022-07/2024

Contexte: Au sein d'Orange Business, la mission en tant que consultant Big Data impliquait des consultations techniques auprès de divers clients, visant à optimiser leur utilisation des données et technologies Big Data pour améliorer les performances et l'efficacité opérationnelle.

Client : Orange OINIS

10/2023-07/2024

Réalisations:

- Conception et déploiement d'une infrastructure de test pour évaluer les performances des serveurs de diffusion Gcore.
- Création et exécution de scripts Gatling scala pour simuler le trafic de streaming vidéo (VoD et live).
- Gestion d'un environnement Docker Swarm pour orchestrer les services de simulation de charge.
- Évaluation des résultats de test pour améliorer la capacité et l'efficacité du service CDN (Réseau de diffusion de contenu).
- Automatisation des tâches à l'aide de Bash et Python.

Client : Groupe Atlantic

09/2022-11/2023

Réalisations:

- Déploiement de pipelines avec Azure Data Factory pour orchestrer le flux de données.
- Capture des données IoT via Azure Event Hubs.
- Traitement et normalisation des données avec Azure Databricks en utilisant Scala.
- Stockage des données normalisées en format parquet dans Azure Data Lake Storage Gen2.
- Calcul de la consommation énergétique sous Databricks, ajusté pour heures creuses et pleines.
- Stockage des résultats dans Azure Cosmos DB pour accès rapide et intégration.
- Développement d'une API pour la BDD NoSQL CosmosDB permettant l'accès aux données aux outils internes.

Client : La Poste

03/2023-06/2023

Réalisations:

- Révision et optimisation de la chaîne de production NiFi existante par l'intégration de Kafka, remplacement des composants pour une meilleure fluidité des flux de données.
- Conception et déploiement d'une pipeline Kafka dédiée pour évaluer et comparer son efficacité par rapport à la solution NiFi existante.
- Rédaction de documentation technique détaillée pour chaque solution, facilitant la transition et la maintenance future.

Client : Orange INNOV

10/2022-03/2023

Réalisations:

- Conception et mise en œuvre d'un connecteur Kafka développé en Python pour une collecte de données efficace et en temps réel.
- Élaboration de bases de données anonymisées, conçues pour soutenir des initiatives de machine learning, tout en assurant la conformité avec les normes de protection des données.
- Optimisation des données pour l'analyse prédictive et l'apprentissage automatique.

Alternant Data Analyst à Orange OINIS

09/2020-08/2022

Réalisations:

- Collection, nettoyage et visualisation des données réseau sur PowerBI pour une analyse détaillée et l'optimisation des opérations réseau.
- Conception et implémentation d'une pipeline de données Kafka pour un modèle d'IA dédié à la détection efficace des alarmes, améliorant la précision et la réactivité de la surveillance réseau.
- Rédaction d'un manuel opérationnel pour l'utilisation d'AirFlow 2, optimisant les workflows d'orchestration de données au sein de l'équipe.
- Évaluation de la solution Feature Store et rédaction de documentation technique pour l'outil Feast, facilitant son intégration pour la gestion des features dans les projets d'IA.
- Modélisation et développement d'un modèle de clustering textuel des logs Kubernetes, améliorant la classification et la gestion des incidents dans l'environnement réseau.

Compétences

Techniques: Big Data, Azure Data Factory, Azure Databricks, Kafka, NiFi, Docker Swarm, Microsoft Fabric, AWS Athena

Programmation: Scala, Python, Bash

Outils: PowerBI, Gatling, CosmosDB, SQL Server, MongoDB, Oracle, ODI, Talend

Langues

Français: Langue maternelle

Anglais: Compétent

Intérêts

Développement personnel, sport, compétitions de programmation.