

Léa Tremblay

Data Engineer Junior

Data engineer diplômée d'un Master Data Science de l'Université de Montréal avec **6 mois de stage** chez Ubisoft Montréal. Construction de pipelines traitant **50 millions d'événements par jour** sur Apache Spark et Airflow. Réduction du temps d'exécution des pipelines de **40 %** et optimisation des coûts Snowflake de **15 000 CAD par mois**. Maîtrise de Python, Spark, Airflow, dbt et Snowflake.

COMPÉTENCES

Python (pandas, PySpark) • Apache Spark • Apache Airflow • dbt (data build tool) • Snowflake • SQL avancé • AWS (S3, Glue, R-edshift) • Great Expectations (qualité de données) • Git et CI/CD • Modélisation dimensionnelle

LANGUES

Français (langue maternelle) • Anglais (C1)

EXPÉRIENCE

Stagiaire Data Engineer, Ubisoft Montréal, Montréal

Janv. 2025 – Juin 2025

Stage de 6 mois au sein de l'équipe Data Platform (8 data engineers, périmètre gaming analytics).

- Construction de **5 pipelines ETL** sur Apache Spark et Airflow, traitant **50 millions d'événements de jeu par jour**
- Modélisation de **12 tables** dans le data warehouse Snowflake avec dbt, suivant le modèle dimensionnel
- Réduction du temps d'exécution des pipelines batch de **40 %** par l'optimisation des partitions Spark et du clustering Snowflake
- Mise en place de **45 tests de qualité** avec Great Expectations, détectant **8 anomalies de données** par semaine en moyenne

Assistante de recherche en traitement de données, MILA (Institut québécois d'intelligence artificielle), Montréal

Mai 2024 – Déc. 2024

Préparation et nettoyage de datasets pour des projets de recherche en NLP.

- Traitement de **3 datasets** totalisant **8 millions de documents texte** en français et anglais
- Construction d'un pipeline de déduplication réduisant la taille des données de **22 %**
- Développement de scripts Python d'extraction de features utilisés par **4 chercheurs**

FORMATION

Diplôme et spécialité, Université de Montréal, Montréal

Sept. 2023 – Août 2025

Diplômée avec mention d'excellence (**3,8/4,0 GPA**).

- Cours principaux : Big Data Analytics, Data Warehousing, Machine Learning, Systèmes distribués
- Mémoire : « Optimisation des pipelines de données en temps réel pour les applications de jeux vidéo », noté A+

CERTIFICATIONS

Snowflake SnowPro Core Certification, Snowflake

Avr. 2025 – Avr. 2025

AWS Certified Cloud Practitioner, Amazon Web Services

Nov. 2024 – Nov. 2024

PROJETS

Pipeline de données sportives en temps réel

Sept. 2024 – Févr. 2025

Projet personnel de collecte et analyse de données sportives en temps réel (hockey sur glace).

- Pipeline Kafka vers Snowflake traitant **10 000 événements par match** en temps réel
- Dashboard Streamlit consultable par **25 utilisateurs** avec des statistiques avancées
- Open source sur GitHub avec **85 étoiles** et **12 forks**

RÉFÉRENCES

Dr. Marie-Claire Bouchard, Lead Data Engineer, Ubisoft Montréal, mc.bouchard@ubisoft.com, +1 514 490 2345

ACTIVITÉ EXTRA-SCOLAIRE

Organisatrice – Meetup Data Montréal

Janv. 2024 – Juin 2025

Co-organisation du meetup mensuel de la communauté data de Montréal.

- Organisation de **8 sessions** avec une moyenne de **45 participants**
- Présentation d'un talk « dbt en pratique : retour d'expérience chez Ubisoft » devant **60 personnes**