Ali Haidar

ahaidar97@outlook.com | **in** | ⦿

PROFESSIONAL EXPERIENCE

Oproma Inc.   Jan 2021 – Jan 2025
Data Engineer   Ottawa, ON

AEM Migration - Correctional Services Canada
- Engineered **metadata and asset collection workflows for 10,000+ web pages**, leveraging **Groovy scripts** and **MiCore tools**, eliminating manual errors and ensuring scalability for larger datasets.
- Generated and validated **Content and Asset Inventories**, enriching **20+ metadata fields** and **resolving 95% of flagged errors** through automated validation scripts.
- Transformed **15+ legacy components** and mapped content to **20+ AEM templates**, deploying all content consistently across **1,000+ target paths**.

Archive Digitization - Library and Archives Canada
- Engineered a **scalable ETL pipeline** for Library and Archives Canada (LAC) to **digitize 1,001 PDFs** into structured CSV datasets, **scaling for 80,000+ records**.
- Streamlined content from **36.5 GB to 3 GB** through optimized page filtering, then transformed data via Tesseract OCR, further **reducing the data size by ~90%**, preparing it for metadata extraction.
- Processed **95% of metadata extraction** via regex, controlled vocabulary matching, and MiCore tools, **validating data** through **ISBNdb and Google Books APIs**.

Mailbox Migration
- Migrated **500+ mailboxes** (~5 TB of data) from Exchange Server 2019 to Exchange Online using **C#** and **.NET Core**, achieving **95% time savings** compared to manual workflows.
- Developed **end-to-end mailbox extraction** and transformation workflows using **Microsoft Graph API** and **EWS API**, processing mailboxes at a rate **of 80–100/hour** with **99% data accuracy**.
- **Implemented** Role-Based Access Control (RBAC) and OAuth 2.0, **securing 1,000+ API interactions** while maintaining **100% security compliance**.

Canadian Radio-Television Committee
- Structured a **20+ table SQL Server schema**, powering an ETL pipeline that **processed 100K+ records** for regulatory Canadian content compliance tracking.
- Improved **query execution times by 50%** using Visual Studio profiling tools and **SQL batch processing**, **optimizing overall pipeline performance by 35%**.
- Architected a three-tiered web application for **1,000+ users**, integrating **RBAC** and secure CRUD operations for **100+ Protected B data submissions** monthly.

Department of National Defense   Jan 2020 – Jan 2021
Junior Data Engineer   Ottawa, ON
- Built an **MVC** application utilizing **.NET Core and C#** to automate Security Technical Implementation Guides (STIG) compliance checks, **cutting manual review times for security audits by 75%.**
- Structured an efficient **ETL pipeline** to ingest DISA STIG data directly from Excel spreadsheets, **reducing manual preparation time by over 15 hours monthly**.

PERSONAL PROJECTS

Fantasy PL Stats – www.fantasyplstats.com
- Founded and deployed a fantasy league management web app via **Azure** using **TypeScript** and **Tailwind CSS**, integrating **Azure ML** and **React-Table** to predict the results across **760 fixtures.**
- Built a **Python ETL pipeline** to ingest historical football data from GitHub into **Azure Data Lake**, cleaning **160,000+ records** across **7 seasons** before loading into **Azure SQL**.
- Performed correlation analysis on cleaned entries, identifying **20+ key predictive features** for player performance and training an AI model in **Azure ML**, achieving **R^2 scores of 0.94 to 0.96**.

Restaurant Data Dive – results
- **Analyzed 3,000+ transactions** from **DoorDash**, **Uber Eats**, and **SkipTheDishes** to identify peak order times and customer preferences, driving data-backed marketing strategies.
- **Ingested raw sales data** into **AWS S3**, then designed and implemented **ETL pipelines in Databricks** using **Apache Spark**, streamlining the data cleaning and standardization process.
- **Evaluated sales trends** and visualized customer behavior using **Matplotlib,** uncovering actionable insights that **drove a 15% increase in off-peak sales** through targeted promotions.

SKILLS

Programming Languages: C#, Groovy, Java, Python, SQL.
Data Platforms & Tools: Apache Spark, AWS S3, Azure DevOps, Azure Functions, Azure Monitor, Databricks, Docker, GitLab CI/CD, Pandas, PostgreSQL, SQL Server.
AI & Machine Learning Tools: Azure AI, Azure Machine Learning, GPT.
Data Engineering Workflows: Data Cleaning & Transformation, ETL Pipelines, Large-Scale Data Processing, Metadata Automation, Schema Design, Data Warehousing.
Data Governance & Security: OAuth, Protected Data Standards, RBAC, Secure API Development.

CERTIFICATIONS

Microsoft Certified	**Dec 2024**
Azure AI Engineer Associate	Microsoft
Design Gurus	**Dec 2024**
Grokking the Modern System Design Interview	Educative
IBM	**Apr 2023**
Introduction to Data Engineering	Coursera
Government of Canada	**Feb 2021**
SECRET Clearance	Ottawa, ON

EDUCATION

University of Ottawa	**Aug 2020**
BASc. Computer Engineering	Ottawa, ON

- Awards: Dean's Honor List