

HEMANYA ARORA

North York, Ontario | Ph: 647-807-5541 | Email: hemanya56@gmail.com

Portfolio: harora.ca | LinkedIn: linkedin.com/in/hemanya-arora | GitHub: github.com/hemanyaarora

PROFESSIONAL SUMMARY

Senior Data Engineer with proven expertise architecting enterprise-scale data platforms that drive strategic business decisions. Specialized in cloud-native solutions (Azure, Databricks), data modeling, and team leadership. Track record of delivering high-impact data products—from modernizing legacy systems to building ML-powered analytics—that generate measurable business value, cost savings, and operational efficiency across Fortune 500 organizations.

RELEVANT EXPERIENCE

Senior Data Engineer: Cineplex Entertainment

April 2024 – Present

- **Platform Migration & Modernization:** Spearheaded the migration of 40+ legacy data pipelines from on-premises infrastructure to Databricks, establishing a scalable cloud-native data platform. Orchestrated end-to-end migration strategy, ensuring zero data loss and minimal business disruption while modernizing critical data workflows.
- **High-Impact Performance Optimization:** Re-architected 25-year-old legacy OLAP cubes into modern star schema data models (Media360, Film360) in Databricks, powering executive-level financial reporting in Power BI. Achieved 87% reduction in processing time (4.5 hours → 36 minutes), enabling real-time decision-making for C-suite stakeholders on film performance and revenue distribution across all business verticals.
- **Technical Leadership & Team Management:** Led cross-functional teams of 6-8 engineers on enterprise-scale data initiatives. Owned complete project lifecycle for Media360 and Film360 data platforms—strategic solutions that drive advertising pricing decisions and power the company's flagship financial reporting system, directly informing multi-million dollar revenue decisions.
- **Strategic Data Product Development:** Architected and deployed production-grade data products integrating multi-source film metadata (YouTube API, Wikipedia, MovieXchange). These pipelines serve as foundational datasets for advanced analytics including long-term attendance forecasting, film recommendation engines, and international market models—enabling data-driven strategic planning across the organization.
- **Production Operations & Reliability Engineering:** Established robust monitoring frameworks and BAU operational excellence practices, ensuring 99.9%+ pipeline reliability. Implemented proactive alerting, automated recovery mechanisms, and comprehensive data quality validations across mission-critical data flows serving enterprise reporting and ML workloads.
- **Cost Optimization & Vendor Management:** Identified and negotiated alternative data sourcing strategy, preventing \$150K+ in annual licensing costs while maintaining comprehensive data coverage. Demonstrated business acumen in balancing technical requirements with financial constraints.
- **DevOps & Platform Engineering:** Championed infrastructure-as-code practices, implementing CI/CD pipelines with Azure DevOps, service principal authentication, and automated deployment workflows. Enhanced data platform security posture and accelerated development velocity through standardized engineering practices.

Data Science Intern: Kinaxis

September 2022 – December 2023

- **Data Modeling & ETL:** Prototyped data-driven models and optimized feature engineering processes, improving AI frameworks for supply chain forecasting.
- **ML Workflow Optimization:** Analyzed feature importance and implemented automation algorithms to streamline data segmentation, yielding more accurate forecasting outcomes.

- **Large-Scale Data Handling:** Conducted data transformations and exploratory data analysis using Python and PySpark, with strong focus on data security, compliance, and performance.
- **Agile Collaboration:** Engaged in daily scrums with cross-functional teams, ensuring alignment of project deliverables with business goals.

Business Intelligence Analyst (Co-op): Co-operators

August 2021 – October 2021

- Developed BI solutions (MicroStrategy, IBM Netezza) to improve reporting and analytics.
- Optimized and validated SQL-based data extractions, ensuring accuracy and consistency.
- Analyzed relational and non-relational data structures to contribute to more efficient data pipelines.

TECHNICAL SKILLS

- **Programming:** Python, SQL, T-SQL, PySpark
- **Big Data & Cloud:** Databricks, Azure Data Factory, dbt, Apache Spark, Hadoop (HDFS), Azure DevOps
- **Data Engineering:** ETL, Data Modeling, Star/Snowflake Schema, Medallion Architecture, CI/CD, Data Orchestration
- **Visualization:** Power BI, Tableau, Plotly
- **Workflow/Collaboration:** Jira, Confluence, Git (GitHub), Agile/Scrum

EDUCATION

Bachelor of Science, Computer Science – York University, Toronto, ON

January 2019 – September 2023

CERTIFICATIONS

- **Databricks Certified Data Engineer Associate** – Issued March 2025
- **Machine Learning (Offered by Stanford University), Coursera** – Issued September 2022

RELEVANT PROJECTS

NFT Market Analysis using Twitter Data ([GitHub](#)):

- End-to-end ML pipeline: web scraping with TWINT, feature engineering, predictive modeling (XGBoost, Random Forest), sentiment analysis achieving 87.3% accuracy, and interactive Plotly visualizations.

Credit Card Fraud Detection ([GitHub](#)):

- Anomaly detection using ensemble methods, handling imbalanced datasets with SMOTE, achieving 97.5% accuracy with Logistic Regression and precision-recall optimization.

VOLUNTEERING EXPERIENCE

Moderator/Co-facilitator: Code Club Canada

September 2020 – September 2022

Educated students aged 8-12 in programming basics and Raspberry Pi projects, fostering an early interest in technology.