

# Rahul Middha

Toronto, ON, CA | (647) 383 - 9553 | rahulmiddha@gmail.com | www.linkedin.com/in/rahulmiddha

## SUMMARY

---

Results-oriented ML/Data Engineer with over 6 years of experience developing machine learning models and scalable data pipelines. Proven expertise in leveraging cloud services for optimized data processing and creating impactful dashboards. Known for collaborating with cross-functional teams, mentoring peers, and driving data strategy. Dedicated to solving real-world problems using ML, focusing on search relevance, recommendations, fraud detection, and time-series forecasting. Experienced in microservice-based architecture and scalable data pipelines.

### TECHNICAL SKILLS:

- **Programming:** Python, NumPy, Pandas, Scikit, SQL
- **ETL Tools:** Informatica, dbt, Airflow
- **Visualization:** Looker, Power BI
- **Big Data Technologies:** Hadoop, Spark, Kafka, Hive, Flink, Beam
- **Cloud Platforms:** Google Cloud Platform, Azure, AWS, Snowflake
- **Database & Storage:** NoSQL databases, Time-series databases, Data Lake solutions, Distributed storage, MySQL, MongoDB
- **Other Tools & Technologies:** Databricks, Git, Docker, Kubernetes

## EDUCATION

---

**MASTER OF SCIENCE – COMPUTER SCIENCE** **2017**  
The University of Texas at Arlington, Texas

**BACHELOR OF ENGINEERING – INFORMATION SCIENCE** **2014**  
Visvesvaraya Technological University, India

## PROFESSIONAL EXPERIENCE

---

**CMiC** **CMiC**  
*Data Engineer L3*

TORONTO, CANADA  
*July 2023 – June 2024*

Tech Stack: Python, Scikit, NumPy, Pandas, TensorFlow, Oracle, AWS

- Designed efficient data solutions in AWS Redshift which incorporated real-time customer Kafka data achieving a 25% increase in data processing efficiency.
- Pioneered a predictive analytics system utilizing Python and Spark that accurately forecasted sales trends with a 95% confidence interval for retail clients.
- Used Pinecone as a Vector Database to store Hugging Face embeddings for LLM chatbot which improved accuracy and faster responses.
- Enhanced team productivity by implementing CI/CD pipelines, which facilitated a smoother, more reliable code deployment process that decreased deployment failures by 60%.
- Transformed and cleansed data from existing CI/CD pipelines into BI reports which helped stakeholders make critical decisions.
- Spearheaded the migration of legacy systems to an AWS Glue and Kafka-based scalable streaming ETL infrastructure, reducing processing time by 40%.
- Introduced robust data governance frameworks that improved data quality by 35%, enabling more accurate machine learning models for customer behavior prediction.
- Directed the adoption of containerization with Docker and Kubernetes, reducing development-to-production lead times by 50% and operational costs by 20%.

 **AFFINE**  
*Consultant Data Engineer*

BENGALURU, INDIA  
*January 2022 – July 2023*

Tech Stack: AWS Glue, AWS S3, BigQuery, RedShift, Python, SQL, Power BI

### Social Media Datawarehouse Project

- Designed and implemented real-time data pipelines using AWS Glue, accelerating processing by 30% and reducing latency by 25%.

- Created ETL scripts in Databricks to handle sensitive customer data for complex data transformations, enhancing retrieval speed for data scientists.
- Led ML Solution Pods to develop production-ready AI applications, focusing on fraud detection and anomaly detection, achieving a 40% reduction in false positives.
- Collaborated with data analysts and data architects, to design and feed proper data models for downstream usage.
- Identified new opportunities and insights from the data to improve models and projected ROI of proposed modifications.

### **Cosmos DB Data Views and Pipeline Creation Project**

- Developed Cosmos DB data views for popular metrics, improving data accessibility and accuracy.
- Reduced processing time by 30% through efficient data pipelines using Azure Data Factory, Azure Databricks, and Synapse Analytics.
- Increased stakeholder satisfaction by 25% by creating informative and visually appealing dashboards in Power BI.
- Improved team skills and productivity by 40% through mentoring and training sessions.
- Collaborated closely with engineering and ML platform teams to ensure adoption of best practices in building and deploying scalable ML services.



BENGALURU, INDIA  
*August 2018 – January 2022*

Tech Stack: AWS Glue, AWS RedShift, Snowflake, Informatica, Python, SQL

- Developed ETL pipelines in Informatica, ensuring data integrity and security, reducing data discrepancies by 30%.
- Automated batch ETL processes in Airflow, reducing manual workload by 40% monthly and increasing operational efficiency.
- Utilized full stack AWS development for scalable AI solutions which improved daily operations.
- Modeled and shifted custom SQL queries into dbt for creating materialized views.
- Implemented large-scale data warehousing programs on Snowflake Cloud and AWS Redshift, reducing storage costs by 20%.
- Enhanced application deployment efficiency by 80% using Docker.
- Designed and developed complex aggregate, join, lookup transformation rules using Informatica PowerCenter tools.

### **PROFESSIONAL CERTIFICATES**

---



**Associate Cloud Engineer - Google Cloud**  
**Issued Jun 2023 – Expires Jun 2026**  
**Credential ID [75891934](#)**

---