

# ASHISH ARUN

Dallas, TX • Email: ashish1995@gmail.com • [linkedin.com/in/ashish-a](https://www.linkedin.com/in/ashish-a) • [github.com/ash1sh95](https://github.com/ash1sh95)

## Summary

Innovative Data Engineer with 5+ years' experience building scalable cloud-native solutions and advanced AIML data pipelines. Proven in designing and optimizing data and AI infrastructure to reduce costs, accelerate model deployment, and drive value at Fortune 500 companies. Specializes in Azure, Databricks, LLM integration, and infrastructure as code for secure, automated, and high-performance data systems.

## Education

### Master of Computer Information Systems

Colorado State University, Fort Collins, Colorado, USA

May 2021

### Bachelor of Engineering in Electronics and Communication

Visvesvaraya Technological University (VTU), Karnataka, India

Jul 2017

## Technical Skills

**Cloud Platforms:** Azure (Databricks, Synapse, Machine Learning), AWS (S3, Redshift, EC2, EMR, Lambda)

**Data Technologies:** Snowflake, MS SQL, MySQL, PostgreSQL, MongoDB, BigQuery, Apache Cassandra, Kafka

**Programming:** Python, Java, C#, Shell Scripting, SQL, PySpark, SparkSQL, SparkML

**DevOps & Tools:** Terraform, Docker, Kubernetes, Git, Airflow, CI/CD, Power BI, Tableau

**AI/ML:** MLflow, Azure Machine Learning, Large Language Models (LLMs)

## Work Experience

### General Motors Financial Company Inc

*Data Engineer III* ( Azure Cloud | Terraform | Databricks)

Feb 2025 - Present

- Led integration and optimization of generative AI/LLM pipelines (Azure Databricks, GPT-3, Llama 3, DBRX), cutting inference costs by 30% and boosting enterprise AI adoption.
- Designed and maintained real-time streaming data pipelines for near-real-time analytics, decreasing ingestion costs by 25%.
- Automated provisioning/configuration of Azure cloud resources with Terraform, reducing deployment time by 50% and enabling repeatable, scalable MLOps for data and AI/ML services.
- Developed and monitored CI/CD pipelines for Azure ML workloads, accelerating secure deployment and updates of AI services.
- Consulted on AI architecture advances, evaluating and integrating new LLM technologies for business applications
- Spearheaded the foundation of Center of Excellence for Databricks and GenAI

### General Motors Financial Company Inc

*Data Engineer II* ( Azure Cloud | Terraform | Databricks)

Aug 2023 - Feb 2025

- Spearheaded AI infrastructure implementation using Azure services, slashing model deployment time by 60% and enabling company-wide AI adoption
- Pioneered integration of Large Language Models (LLMs) including GPT-3 and DBRX in Azure Databricks, resulting in 30% reduction in inference costs
- Architected real-time streaming data pipelines, decreasing data ingestion costs by 25% and enabling near real-time analytics
- Implemented Infrastructure as Code (IaC) using Terraform, accelerating Azure resource deployment by 40%
- Optimized Delta Lake tables and implemented Delta Live Tables, boosting query performance by 20% and enhancing data reliability by 40%
- Leveraged MLflow for experiment tracking, improving model reproducibility by 50% and streamlining the ML lifecycle

### General Motors Financial Company Inc

*Data Engineer I* ( Azure Cloud | Terraform | Databricks)

Oct 2022 - August 2023

- Engineered 30+ Synapse data pipelines, boosting data ingestion efficiency by 40% for bronze layer sources
- Migrated Hive metastore to Unity Catalog in Databricks, enhancing data governance and reducing security incidents by 50%

- Established robust CI/CD pipelines in Azure DevOps, reducing deployment errors by 75% and doubling release frequency
- Automated Databricks cluster management with Terraform, cutting cluster setup time by 80% and yielding \$50,000 annual savings in compute costs

## **Snagajob**

**Associate Data Engineer** (Snowflake | Kafka | MongoDB | Docker | Argo)

Jul 2021 - Aug-2022

- Implemented 25+ SnowSQL scripts, functions and procedures for KPI metric calculations and automated them in Python
- Maintained and managed Kafka topics and kafka consumers with confluent kafka for data pipelines
- Developed 50+ kafka pipelines to snowflake and automated the process with docker containers and argo
- Designed and Architected automation to notify new kafka topics through Slack API webhooks which improved visibility and reduced tracking time by 80%
- Created and Maintained python code to pull data from RESTful APIs and mongoDB, transformed and ingested to Snowflake

## **JBS USA**

**Information Technology Intern** (MS SQL | PowerShell)

Jun 2020 - Aug 2020

- Automated database backup and restore processes using PowerShell, improving operational efficiency by 50%
- Developed GUI for system resource monitoring, enhancing visibility into disk space and memory utilization

## **NTT DATA Global Delivery Services**

**Testing Engineering Senior Associate** (JIRA | HP ALM | MS SQL | Python)

Feb 2018 - Jun 2019

- Executed comprehensive software testing strategies, achieving 90% defect detection accuracy
- Developed 50+ test scenarios daily, reducing defect slippage to under 5% and maintaining zero fatal defect slippage

## **Certifications**

Microsoft Certified: Azure AI Fundamentals (AI-900)

Databricks Accredited: Gen AI Fundamentals, Lakehouse Fundamentals, Platform Administrator

Business Intelligence and Application Development (Colorado State University)