Amyuktha Gummadi

New jersey | 603-202-2426 | amyuktha.gummadi.20@gmail.com | **linkedin** | **Github** | **Portfolio**

PROFESSIONAL SUMMARY

Databricks Certified Data Engineer Associate with hands-on experience building **end-to-end, metadata-driven ELT pipelines** using **AWS S3, Snowflake, dbt, and Azure data services**. Strong foundation in **SQL, Python, data modeling, incremental processing, and Medallion architecture**. Experienced in working with **Azure Data Factory and Azure Data Lake** for data ingestion and orchestration. Actively building practical expertise in **Generative AI for data applications**, including **LLMs, transformer architectures, embeddings, vector databases, and Retrieval-Augmented Generation (RAG)**. Skilled in developing scalable, analytics-ready datasets and modern **cloud-native data and GenAI pipelines**.

TECHNICAL SKILLS

Programming & Querying - Python, SQL
Data Engineering & Pipelines - ETL / ELT pipelines, metadata-driven pipelines, incremental loads, data ingestion, transformation, analytics-ready datasets
Big Data & Processing - Apache Spark (PySpark), Databricks
Data Modeling & Architecture - Medallion Architecture (Bronze / Silver / Gold), Star Schema, Fact & Dimension Tables, One Big Table (OBT), Slowly Changing Dimensions (SCD Type 1 & Type 2)
Cloud Platforms - **AWS:** S3 (data lake fundamentals). **Azure:** Azure Data Factory, Azure Data Lake. **Snowflake:** Cloud data warehouse
Generative AI & LLMs (Foundational, Hands-on) - LLM fundamentals, Transformer architecture (encoder–decoder, attention), embeddings, vector databases (ChromaDB), Retrieval-Augmented Generation (RAG), LangChain (chains, agents), vector store persistence, schema validation (Pydantic, TypedDict)
Analytics & Engineering Tools - dbt (models, snapshots, tests, Jinja templating), Power BI, Git, CI/CD fundamentals
Data Quality & Governance - Data validation, data testing, metadata documentation

PROFESSIONAL EXPERIENCE

Data Engineer - ARETE IT Services Pvt. Ltd. — India - *Jul 2022 – Jul 2023*
- Designed and supported **production ELT pipelines**, integrating multiple source systems into a centralized MySQL data warehouse.
- Built **Spark-based transformations** in Databricks, implementing **Silver and Gold layers** aligned with Medallion and lakehouse architecture patterns.
- Orchestrated ingestion workflows using **Azure Data Factory** and managed data storage in **Azure Data Lake**.
- Developed **star schema data models**, implemented **SCD Type 1 & Type 2**, and optimized SQL queries to deliver analytics-ready datasets.

Data Engineer - Avishkar Tech Solutions — Remote - *Jun 2021 – Jun 2022*
- Developed **Python-based ETL pipelines** for data cleaning, transformation, and validation.
- Used **PySpark and Pandas** for scalable data processing and performance optimization.
- Prepared curated datasets for analytics and reporting use cases.

Full-Stack Developer - Saayam For All — Remote, USA - *Aug 2025 – Present*
- Built backend data workflows using **Python and PostgreSQL** to support analytics and operational reporting.
- Automated data refresh processes and implemented basic monitoring to improve pipeline reliability.
- Collaborated with cross-functional teams to support data-driven application features.

PROJECTS

Metadata-Driven Data Engineering Pipeline (AWS S3 | Snowflake | dbt)
- Built an **end-to-end metadata-driven ELT pipeline** using **AWS S3** as the data lake, **Snowflake** as the cloud data warehouse, and **dbt** for transformations.
- Organized transformations across **Bronze, Silver, and Gold layers**, producing analytics-ready outputs including **fact tables, star schema models, and One Big Table (OBT)**.
- Used **dbt Jinja templating** to build reusable SQL models and **incremental models** to process only new or updated data.
- Implemented **dbt snapshots** to track historical changes using **SCD Type 2** logic.
- Applied production-style design principles to improve pipeline **scalability, flexibility, and maintainability**.

Mood Song Predictor (SQL | Power BI)
- Designed a relational data model and SQL transformations to categorize songs by mood.
- Built **Power BI dashboards** enabling interactive filtering and insight generation.

Business Continuity Planning Project
- Designed **data backup, recovery, and incident response strategies** to ensure data availability and operational resilience.

EDUCATION

Rowan University, Glassboro, NJ - **Master of Science — Computer and Information Systems Security** | May 2025
Koneru Lakshmaiah Education Foundation (KL University) - **Bachelor of Technology — Computer Science** | May 2023

CERTIFICATIONS

- Databricks Certified Data Engineer Associate
- Microsoft Azure Fundamentals (AZ-900)
- Oracle Cloud Infrastructure Foundations Associate