

Hemavarthini Balaji

Chennai, India | +919840820120 | hemavarthini3004@gmail.com | www.linkedin.com/in/hemavarthini

PROFESSIONAL SUMMARY

Senior Software Engineer building **scalable ETL pipelines and cloud data platforms** using Python, SQL, Databricks, and Azure. Strong foundation in **AI/ML and Data Structures & Algorithms**, demonstrated by solving **600+ LeetCode problems**.

SKILLS

Programming: Python, SQL, PL/SQL, PySpark, HTML, CSS, JavaScript

Cloud: Databricks, Snowflake, Azure Data Factory, Microsoft Azure, AWS S3, Google Cloud Platform (GCP)

Streaming: Apache Kafka, Kafka Producers/Consumers, Event-Driven Systems, Streaming Pipelines

Data: ETL Pipelines, Data Modeling, Incremental Loads, CDC, SCD (Type 1/2), Data Transformation, Talend

APIs: REST API, HTTP, OAuth2, API Authentication, Flask

Databases: SQL Server, Snowflake, SQLite, ChromaDB, Pinecone, OLAP, OLTP

Libraries: PyTorch, Keras, OpenCV, YOLO, Pandas, NumPy, LangChain

Analytics: Power BI, Tableau, Exploratory Data Analysis (EDA), Statistical Analysis, Query Optimization

DevOps: GitHub Actions, Azure Logic Apps, Workflow Automation (n8n), Monitoring & Alerts

Core: Data Structures & Algorithms, System Design, Problem Solving, Clean Code, SOLID Principles

ML: Machine Learning, Deep Learning, CNN, SVM, Transformers, Feature Engineering, Hyperparameter Tuning, PCA

GenAI: LLM, RAG, Recommendation Systems, Embeddings, Vector Search

CERTIFICATIONS

- AI-900: Microsoft Azure AI Fundamentals - Microsoft
- AZ-900: Microsoft Azure Fundamentals - Microsoft
- Databricks Generative AI Fundamentals - Databricks

ACHIEVEMENTS

- **8x Bravo Award (Optum):** Recognized for **data engineering impact, innovation, mentoring, and teamwork**.
- **Outstanding Leadership Skill (SRM University):** Earned for **initiatives and impactful team contributions**.

WORK EXPERIENCE

Optum

Chennai, India

Senior Software Engineer

Oct 2025 - Present

- Modernized **Talend-based ETL workflows to PySpark on Azure Databricks and AKS**, improving pipeline scalability by **40%** and enabling cloud-native data processing.
- Designed and implemented a **policy age-limiting validation module** for **3M+ subscribers**, automating eligibility checks and reducing manual approval efforts by **70%** while accelerating policy processing cycles.
- Enhanced **Talend XML processing pipelines and business logic**, introducing a policy duration attribute and resolving void-date inconsistencies, improving downstream data accuracy by **35%** and reducing manual reprocessing by **30%**.
- Resolved critical **Vision Policy multiline parsing failures** by conducting root cause analysis and implementing upstream **data validation checks**, improving pipeline reliability and preventing recurring **ETL processing failures**.
- Strengthened **application security and data consistency** by remediating **100+ GitHub security vulnerabilities** and updating **BRC literals and SQL transformation logic** to align policy datasets with evolving product requirements.

Optum

Hyderabad, India

Data Engineering Analyst

Sep 2024 - Oct 2025

- Orchestrated and architected **scalable Databricks-based data ingestion pipelines for 7+ API sources**, implementing validation, retry, and fault-tolerant processing frameworks to ensure **99.9% pipeline uptime**.

- Optimized **end-to-end ETL pipelines in Databricks** processing data from **5+ API endpoints**, handling edge cases such as invalid credentials, token failures, malformed payloads, and authentication changes, reducing production data processing errors by **30%**.
- Migrated enterprise datasets from **SQL Server to Snowflake using Azure Data Factory**, improving query performance by **60%** and enabling scalable **OLAP-based analytics** across **10+ business teams** with a throughput increase of **50%**.
- Modernized legacy workflows by **migrating pipelines to Databricks and implementing automated monitoring**, reducing runtime by **99.3%**, lowering compute/storage costs by **80%**, and improving pipeline reliability by **50%**.
- Streamlined a **Python-based Git automation script for GitHub asset migration**, eliminating manual repository management processes and reducing operational effort by **90%**.

Optum

Hyderabad, India

Associate Data Engineering Analyst (TDP)

Sep 2023 - Sep 2024

- Established and Deployed **scalable ETL pipelines using Azure Data Factory**, processing **10,000+ records daily** to support enterprise data integration and improve data throughput by **45%** for analytics workflows.
- Developed and optimized **data ingestion pipelines integrating CSV, JSON, and SharePoint sources**, enabling structured data transformation and reducing reporting errors by **30%** while accelerating report generation by **20%**.
- Enhanced **Power BI dashboards and automated reporting workflows**, improving reporting accuracy by **80%** and reducing manual operational effort by **60%** through automation and notification systems.
- Led migration from **SQL Server to Snowflake** platform, boosting data retrieval speeds by **60%**, which enabled analysts to perform complex queries **3x faster** and supported scaling of analytics capabilities across multiple departments.

PROJECT EXPERIENCE

Research & Engineering Projects

Chennai, India

Machine Learning & Data Engineering Projects

Jun 2019 - Present

- **Real-Time Recommendation System:** Orchestrated a real-time product recommendation system using **Kafka** streams and **ChromaDB** embeddings, enriching user experience by increasing **click-through rate 3%** for recommended items.
- **Clinical Intelligence System:** Architected an **LLM-driven clinical intelligence** platform leveraging Retrieval-Augmented Generation (**RAG**) to extract insights from healthcare datasets, accelerating clinical data analysis by **35%**.
- **Multimodal Emotion Recognition:** Engineered an **audio-visual emotion detection system** using STFT/MFCC features and OpenCV with decision-level fusion, achieving **92% accuracy across 5+ emotional states**.
- **Edge AI Emotion Detection (Jetson Nano):** Applied a **real-time emotion recognition pipeline using YOLO and OpenCV on NVIDIA Jetson Nano**, achieving **90% accuracy** while reducing inference latency by **60%**.
- **Breast Cancer Prediction using SVM:** Built a **machine learning classifier using Support Vector Machines (SVM)** to detect malignant tumors with **90% accuracy**, reducing misclassification by **15%**.
- **AI-Powered Requisition Assistant:** Programmed an **LLM-driven recruitment analytics assistant** enabling natural language queries on hiring datasets, reducing manual analysis errors by **30%**.
- **Scrape-to-Insight Data Platform:** Crafted a **data pipeline ingesting NASDAQ market data into Snowflake with dbt modeling and Streamlit dashboards**, improving reporting efficiency by **40%**.

EDUCATION

SRM Institute Of Science And Technology

Chennai, India

*Bachelor of Technology - Computer Science Engineering
CGPA - 9.57*

Graduation Date: May 2023

Woolf University, Learnbay

Online (London)

Masters of Science in Artificial Intelligence and Machine Learning

Graduation Date: Sep 2026