

Gayathri Kumar

Dallas, TX 75234 | Gayathrikumar2@my.unt.edu | 4698031270 | [Linkedin](#)

Summary

Results-driven Data Engineer with 7 years of experience in designing, developing, and optimizing data pipelines, ETL processes, and cloud-based solutions. Proficient in Azure Data Factory, AWS Lambda, Python, SQL, and data visualization tools like QuickSight and Power BI. Strong expertise in OCR, Machine Learning, and automating data workflows. Proven ability to enhance efficiency, optimize performance, and drive business insights through data-driven solutions. Currently pursuing a master's degree at the University of North Texas, with an expected graduation date of May 2026.

Education

University of North Texas

Expected Graduation: May 2026

Master of Science in Information Science, with a concentration in Information Systems (GPA: 4.0)

Technical Skills

- **Cloud & ETL:** Azure Data Factory (ADF), AWS Lambda, AWS Glue, S3, Databricks, Data Lake Storage
- **Programming & Automation:** Python (Pandas, NumPy, OpenCV), SQL, Java, C++ , MLOps
- **Data Science & ML:** Machine Learning, Deep Learning, OCR (Tesseract, PyMuPDF), NLP, XGBoost, TensorFlow, Docker, CI/CD pipelines, Git workflows, predictive modeling, LLMs, ETLs
- **Visualization & Analytics:** QuickSight, Power BI, Tableau
- **Big Data & Databases:** Azure SQL, DynamoDB, Athena, Delta Lake, SAP HANA
- **Databases:** SQL, NoSQL(MongoDB), AWS cloud database , GCP, ML Model Deployment, Pytest
- **Developer Tools:** Git/GitLab, Linux, QNX, RapidMiner, draw.io, Excel
- **Others:** Data Collection, Cleaning, Feature Engineering, Pipeline Development, Data Visualization, Mathematics, Statistics, Data Analytics, Communication, time-series data, version control, code review

Work Experience

Deloitte , Lead Solution Advisor , India

Jul 2022 – Jul 2024

Azure Data Factory ETL & Analytics

- Engineered and deployed ETL pipelines using **Azure Data Factory** (ADF) to transform data from **SAP HANA** and SAP ECC into Azure **SQL**, reducing processing time by 25%.
- Configured **Self-Hosted Integration Runtime** (SHIR) to securely extract on-premises SAP data, improving data accessibility and security.
- Accelerated structured and semi-structured data processing by 30% through **Databricks (PySpark)**.
- Implemented scalable data retention solutions using **Blob Storage** and **Data Lake Storage Gen2**.

- Automated data workflows using ADF **Data Flows, triggers**, and metadata-driven frameworks, reducing manual tasks by 30%.
- Enhanced system monitoring by integrating **Azure Monitor** and Log Analytics for proactive issue detection.
- Streamlined self-service analytics with **Alteryx** integration, increasing business user efficiency by 20%.

AWS Serverless Data Pipeline & Dashboard

- Developed a serverless data pipeline using **AWS Lambda, S3, DynamoDB** for automated data processing.
- Leveraged **AWS Glue** for ETL operations and **Amazon QuickSight** for real-time dashboards.
- Optimized query performance with **Athena**, security with **IAM policies**, and monitoring via **CloudWatch**, resulting in a 35% performance gain.

HCL Technologies (Ford Motors) , Lead Engineer , India

Jan 2021 – Jul 2022

OCR-Based Data Extraction & Validation

- Developed an **OCR-powered system** using **Tesseract, OpenCV, PyMuPDF** to extract car part numbers from PDFs.
- Preprocessed images using **grayscale conversion, thresholding, contour detection** increasing data extraction accuracy by 15%.
- Designed a **Flask-based web app** for users to compare extracted data with Excel records.
- Automated mismatch detection using **Pandas, OpenPyXL, NumPy, Regex** , achieving a 20% reduction in validation time.

Fedo.ai , Data Scientist , India

Oct 2020 – Jan 2021

AI-Powered Smoker Detection for Insurance

- Built a **machine learning model** using **Scikit-Learn, TensorFlow, XGBoost** to classify individuals as smoker/non-smoker, achieving 93% accuracy incorporating features from biometric, lifestyle, and medical data for improved accuracy.
- Deployed the model as an **API using Flask/Fast API** for faster insurance verification.

Bosch & Siemens Home Appliances , Executive-II ,India

Feb 2018 – May 2020

- Developed web-based applications for process automation, reducing manual work by 30%.
- Conducted code reviews and optimizations, improving application performance by 20%.

Cognizant Technology Solutions , Programmer Analyst , India

Apr 2016 – Jan 2018

- Led **functional, regression, and integration testing** initiatives.
- Automated testing processes and optimized **system requirements tracking**.

Relevant Coursework – Information Science with Information Systems concentration

Data Modeling, Data Analysis, Information Security & Cybersecurity, Project Management, Knowledge Management, Data Visualization