Vipin Kumar Karthikeyan

(413) 510 7910 – vipin.k.karthikeyan@gmail.com - [LinkedIn](#) - [Portfolio](#) - [GitHub](#)

Education

University of Massachusetts Amherst - Isenberg School of Management
Master of Science in Business Analytics   *September 2024 - December 2025*
Anna University - Coimbatore Institute of Technology
Master of Science in Data Science   *July 2017 - May 2022*

Skills

- **Programming Languages:** Python, SQL, PySpark, Scala, Java, C, C++, NoSQL
- **Cloud Platforms:** Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP)
- **Data Engineering & Big Data:** Databricks, Airflow, Kafka, Kinesis, Snowflake, MS SQL Server, Data Warehousing, ETL/ELT
- **Data & Analytics:** Data Modeling, Data Transformation, Machine Learning, AI, Data Architecture, Cloud Security
- **Visualization & BI Tools:** Microsoft Power BI, Tableau, Excel, DAX, M Query
- **DevOps & Tools:** Docker, Azure DevOps, GitHub, GitLab, CI/CD Pipelines, Terraform
- **Core Competencies:** Project Management, Agile Methodologies, Stakeholder Engagement, Cross-functional Collaboration

Experience

Cloud Analytics & Automation Intern
Waters Corporation – Milford, MA   *June 2025 – Present*
- Engineered a real-time, cross-cloud observability platform using Azure, AWS, ServiceNow, Jira, and Prisma Cloud, improving cost visibility, usage tracking, and security alerting by 60%.
- Automated end-to-end data pipelines with Power Automate, Azure Functions, and REST APIs, cutting manual data handling by 70% and accelerating cloud analytics workflows across business units.
- Integrated Azure OpenAI for LLM-driven cost optimization insights and remediation triggers, reducing cloud spend by 15% and elevating operational efficiency through AI-powered automation.

Senior Data & Analytics Engineer
Thorogood Associates - Bangalore, India   *July 2022 - July 2024*
- Led end-to-end data pipeline architecture using Azure Data Factory, Databricks, and SQL Server, processing over 500K invoices and 100K suppliers, with CI/CD automation via Azure DevOps.
- Designed and implemented a cloud-based sustainability analytics platform, integrating SAP and factory-level data into Azure Data Lake, boosting sustainable procurement visibility by 34%.
- Built and optimized enterprise BI dashboards using Power BI and Azure Analysis Services, supporting $3.06B in sales analysis and identifying $65.2M in lost sales across 15+ categories.
- Automated data validation workflows with Delta Lake, Power Automate, and Power Apps, achieving 99.9% data accuracy and a 30% improvement in processing speed.
- Streamlined data processing and reporting by implementing partitioned delta tables, improving Power BI refresh performance by 40% and enhancing time-to-insight for critical business users.

Data Engineer
EXL Service - Chennai, India   *July 2021 - June 2022*
- Migrated on-premise SQL Server infrastructure to cloud-native architectures using AWS (S3, Lambda, Kinesis, Aurora) and Azure, reducing compute costs by 35% and improving query performance by 60%.
- Developed scalable ETL pipelines using AWS Glue and Azure Data Factory, processing 500K+ medical claims per file, enabling real-time ingestion and 350K+ edits/sec in fraud detection scenarios.
- Designed end-to-end data orchestration with Databricks, Service Bus, and Azure Blob Storage, ensuring secure audit trails and SLA-compliant pre-pay and post-pay claim analytics.

Data Scientist
Neewee Analytics - Bangalore, India   *May 2020 - June 2021*
- Deployed a predictive maintenance solution for steam turbines using Multivariate State Estimation Technique (MSET) and Statistical Process Control, reducing unplanned downtime by 25%.
- Engineered real-time sensor data pipelines from MQTT brokers, processing high-frequency JSON streams with Python, Scala, and PostgreSQL to enable low-latency anomaly detection.
- Automated fault detection workflows with data imputation, feature engineering, and dashboarding, cutting manual monitoring efforts and lowering maintenance costs by 20%.

Academic Projects

Voice-to-Text Q&A Chatbot   *September 2024 – Present*
- Architected an Agentic AI chatbot on AWS, integrating Lambda, Transcribe, and S3 for speech-to-text processing and dynamic task delegation based on intent classification and user prompts.
- Deployed Retrieval-Augmented Generation (RAG) pipelines with Amazon Bedrock and SageMaker, enabling real-time interactions with LLM-based responses fine-tuned for domain-specific Q&A tasks.

Know Your Coin   **[GitHub](#)**
- Built a deep learning-based coin classification system using InceptionV3 for numismatics; implemented in Google Colab with image augmentation, resulting in 99% training and 84% test accuracy.

Certifications

- Azure Certified Data Engineer Associate - DP-203: Data Engineering on Microsoft Azure
- Databricks Certified Data Engineer Associate