Vipin Kumar Karthikeyan

Boston, MA - (413) 510 7910 - vipinkumarka@umass.edu - LinkedIn

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 *April 2017 - May 2022*

Skills

Industry Knowledge

- Data Engineering, Extract, Transform, Load (ETL), Data Transformation, Data Cleansing, Data Warehousing, Data Modeling, Data Visualization, Cloud Architecture, Big Data, Machine Learning, Artificial Intelligence (AI), Business Analytics, Data Processing, Project Management, Agile Environment, Data Governance, Cloud Security, AI Model Deployment, DevOps, Data Security, Client Interaction, Stakeholder Engagement, Leadership & Communication.

Tools & Technologies

- **Cloud Platforms:** Amazon Web Services (AWS), Microsoft Azure (Azure), Google Cloud Platform (GCP)
- **Big Data & Data Engineering:** MS SQL Server, Databricks, Apache Kinesis, Hadoop, Kafka, Snowflake, Data Warehouse, Airflow
- **Programming Languages:** Python, SQL, PySpark, C, C++, Scala, Java, NoSQL
- **Visualization & BI:** Microsoft Power BI (DAX, M Query), Tableau, Excel
- **Version Control & Other Tools:** GitHub, GitLab, Docker, Azure DevOps

Experience

Incoming Cloud Engineer Intern
Waters Corporation – Milford, MA *May 2025 – August 2025*

- I will be joining Waters as a Cloud Engineer Intern, where I will collaborate with the IT and Engineering teams to develop and implement cloud-driven solutions.

Lead Data & Analytics
Thorogood Associates - Bangalore, India *July 2022 - July 2024*

- Architected and managed data pipelines using Azure Data Factory and Databricks, processing 500K invoices and 100K suppliers/factories, and deployed via CI/CD pipelines using Azure DevOps for automated delivery.
- Designed an Azure Data Lake to integrate SAP data and sustainability metrics, enabling seamless reporting through Microsoft SQL Server and Power BI.
- Automated data validation and quality assurance with Databricks Delta Lake, Power Automate, and Power Apps, achieving 99.9% accuracy and improving efficiency by 30%.
- Developed scalable data models for channels, categories, and SKUs, providing real-time analytics on inventory performance with a 98.4% fill rate.
- Built pipelines with Azure Data Factory, Databricks Delta Lake, and Azure Analysis Services, processing billions of records to support dashboards with $3.06B in sales and $65.2M in lost sales.
- Improved Power BI refresh times by 40% using Delta Tables and CSV-based storage, enhancing large-scale data analysis across 15+ categories and 3 sales channels.

Data Engineer
EXL Service - Chennai, India *July 2021 - June 2022*

- Migrated on-premises systems to cloud-based architectures using AWS and Azure, enhancing scalability, reducing latency, and cutting costs by 35% while boosting performance by 60%.
- Designed and implemented ETL workflows using AWS Glue and Azure Data Factory, processing datasets with up to 500K claims per file and achieving speeds of over 350K edits per second in high-load scenarios.
- Built data pipelines integrating S3, Lambda, Kinesis, Aurora DB, Azure Blob Storage, Databricks, and Service Bus, ensuring reliable data flow and compliance for pre-pay and post-pay audits.

Data Scientist
Neewee Analytics - Bangalore, India *May 2020 - June 2021*

- Deployed an early fault detection system in steam turbines using Multivariate State Estimation Technique and Statistical Process Control, enabling predictive maintenance and reducing downtime by 25%.
- Built a data pipeline to process high-frequency sensor data in JSON format from MQTT brokers, leveraging Python, Scala, and PostgreSQL for real-time anomaly detection with low latency.
- Automated the fault detection workflow, including preprocessing, imputation, and feature engineering, and visualized operational metrics through dashboards, reducing maintenance costs by 20%.

Academic Projects

Voice-to-Text Q&A Chatbot *September 2024 - Present*

- Built an AWS based chatbot system using Lambda, Transcribe, and S3 for voice-to-text processing, and deployed RAG models via Bedrock and SageMaker for real-time user interaction with GPT-powered responses.

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