

# MOUNIKA KATTA

+1 (860) 984-3693 | [mounikakatta070699@gmail.com](mailto:mounikakatta070699@gmail.com)

<https://www.linkedin.com/in/mounika-katta-7387a7135> | Github: <https://github.com/MounikaKatta30>

---

## Professional Summary

Results-oriented Data Engineer with 5+ years of experience transforming raw data into reliable, scalable, and analytics-ready systems across finance, healthcare, and wealth management domains. Began career as a Data Analyst, delivering business insights through Power BI, Tableau, and SQL, and transitioned into data engineering by designing and automating cloud-native data pipelines using Azure Data Factory, Databricks, and Snowflake. Skilled in Python, dbt, Airflow, and real-time streaming (Kafka, Event Hubs), with proven success in implementing data governance, security, and compliance frameworks (HIPAA, GDPR, FINRA, SEC). Strong collaborator experienced in modernizing legacy systems, optimizing data workflows, and enabling self-service analytics to drive data-driven decision-making and digital transformation.

---

## Core Skills

**Data Engineering & ETL:** Azure Data Factory, Databricks, Apache Airflow, dbt, Synapse Analytics, SSIS, PySpark, Spark Streaming, AWS Glue, and large-scale ETL/ELT pipelines.

**Cloud Platforms:** Azure (Synapse, Data Lake, Event Hubs, Key Vault), AWS (EC2, S3, Lambda, Redshift, Glue, KMS), Snowflake, and Google BigQuery.

**Programming & Scripting:** SQL (advanced joins, CTEs, window functions), Python (Pandas, NumPy), R, Java, and Terraform for automation and CI/CD.

**Data Modeling & Warehousing:** Dimensional modeling, Star/Snowflake schema design, data marts, metadata management, and lineage in Snowflake, Azure Synapse, and Redshift.

**Streaming & Real-Time Processing:** Kafka, Azure Event Hubs, Spark Streaming, and event-driven data pipelines.

**Databases:** MySQL, Oracle, SQL Server, Postgres, and MongoDB with query optimization and schema design.

**API & Integration:** FHIR, HL7, REST, and SOAP for healthcare and financial system interoperability.

**Visualization & Analytics:** Power BI, Tableau, Domo, Looker, Qlik Sense, Metabase, Apache Superset, Salesforce Analytics, and Excel.

**Machine Learning & NLP:** scikit-learn, TensorFlow, Keras, classification models, regression, NLP, and time-series forecasting.

**Data Governance & Quality:** Informatica DQ, Ataccama, Collibra DQ, RBAC, encryption, data masking, BCBS 239, GDPR, CCPA, HIPAA compliance.

**Tools & DevOps:** Git, GitHub, Bitbucket, Azure DevOps, Jenkins, Jira, Confluence, Jupyter Notebook, Google Colab, and VS Code.

**Domain Knowledge:** B2B SaaS, PaaS developer platforms, financial services, healthcare, Oracle HCM, and infrastructure tooling.

**Other Skills:** Root-cause analysis, data validation, troubleshooting, Agile/SDLC, KPI definition, and cross-functional collaboration.

---

## Professional Experience

**Client:** PayPal | New York, NY

**Role:** Data Engineer | **Duration:** Jan 2024 - Present

**Project:** Finance & Operations Analytics Modernization

- Design, build, and maintain scalable ETL/ELT pipelines using Azure Data Factory, dbt, or Apache Airflow to automate data ingestion from ERP, CRM, and financial systems.
- Modernize legacy data flows into cloud-native data warehouses like Snowflake or Azure Synapse, ensuring high performance and fault tolerance.
- Implement incremental load strategies, schema evolution, and data partitioning to optimize storage and query efficiency.
- Develop data validation and anomaly-detection scripts in Python or SQL to maintain data accuracy across finance and operations domains.
- Establish data lineage and cataloging with tools like Purview or Collibra, ensuring compliance with organizational governance standards.
- Implement role-based access controls (RBAC) and encryption policies to secure sensitive financial data and meet audit requirements.

- Create semantic data models and materialized views for finance, budgeting, and operational KPIs.
- Collaborate with analysts to deliver self-service BI datasets for Power BI or Tableau, reducing report turnaround time.
- Build automated data refresh and monitoring workflows, integrating with CI/CD pipelines to support agile analytics releases.

---

**Implementation Partner: Tata Consultancy Services (TCS) | Hyderabad, India**

**Client: AdventHealth**

**Role: Data Analyst | Duration: Jun 2021 - Dec 2023**

**Project: Patient Experience Analytics Modernization**

- Partnered with client business teams to develop a knowledge-based Q&A platform using LlamaIndex and RAG models, applying advanced data analytics methods to address critical business challenges.
- Designed custom solutions based on client feedback for a job fraud detection classifier leveraging Hugging Face Transformers, ensuring robust data preprocessing and model accuracy.
- Provided predictive analytics and customer churn insights for clients in finance, healthcare, and retail sectors, utilizing statistical analysis and data visualization techniques.
- Built customized Power BI dashboards tailored to healthcare program analysis, demonstrating expertise in spreadsheet analysis and advanced Excel functions such as pivots and intricate formulas.
- Collaborated with client IT teams to integrate real-time fraud detection pipelines using Azure Databricks and Synapse Analytics, supporting efficient ETL processes and data-driven decision-making.
- Conducted client workshops and training sessions to facilitate the adoption of new dashboards and reporting tools, enhancing overall data literacy and cross-functional collaboration.

---

**Implementation Partner: Axis Max Life Insurance Corporation | Hyderabad, India**

**Client: Morgan Stanley**

**Role: Junior Data Analyst | Duration: Jan 2020 - May 2021**

**Project: Wealth Management Digital Transformation**

- Partnered with operations and finance teams to document leasing and property workflows.
- Designed Tableau dashboards for occupancy, revenue performance, and lease renewal analytics.
- Conducted business capability mapping, aligning process improvements with Microsoft Dynamics 365 CRM.
- Automated monthly reporting workflows with Power Automate and SQL, reducing turnaround time by 20%.
- Performed root cause analysis on revenue leakage and process gaps in tenant billing.
- Collaborated with IT to integrate tenant data into ERP and BI systems.
- Conducted UAT and provided training to property managers on new reporting tools.

---

## Education

**University of Bridgeport | Bridgeport, CT, USA | Jan 2024 - May 2025**

Master of Science, Computer Science | 3.7/4.0 GPA

- **Coursework:** Gained hands-on experience through real-time projects in insurance, finance, healthcare, and retail. Built Q&A systems using LlamaIndex and RAG models, developed fraud detection models using NLP and Hugging Face Transformers, created predictive churn models and Power BI dashboards for healthcare analytics. Designed real-time fraud detection pipelines using Azure Databricks and Synapse Analytics
- **Teaching Assistant:** Web App Development Course

**MVSR Engineering College | Hyderabad, India | Jun 2018 – Jul 2021**

Bachelor of Engineering, Electronics and Communication Engineering | 3.8/4.0 GPA

- **Coursework:** Studied core subjects in electronics, communication systems, and programming

---

## Projects:

- **Violence Detection System: Revolutionizing Safety through Machine Learning:** Built an end-to-end ML pipeline and secure web interface to detect violent content in uploaded videos using CNN models, real-time alerts, and encrypted data processing.
- **Stock Market Kafka Real Time Data Engineering Project:** Designed a Kafka-based real-time data pipeline using Python, AWS S3, Glue, Athena, and EC2 to ingest, process, and query live stock market data.
- **Text summarization pipeline with a pretrained LLM/SLM such as Llama-3, or Phi-3, or H2O:** Developed an LLM-based summarization workflow using Llama-3/Phi-3/H2O models with preprocessing, inference, and performance optimization.

---

## Certifications

- Certified in Python Programming
- Certified in Django Web Framework
- Windows Server Administration Certificate
- Create Machine Learning Models in AWS Course Certification
- AI and Machine Learning Cloud Fundamentals Course Certification
- Microsoft Office Specialist Excel Associate (Office 2019)
- Oracle SQL Certified Associate
- Microsoft Certified: Power BI Data Analyst Associate
- AWS Certified Data Analytics – Specialty