Naga Mutya Kumar Kumtsam
469-927-2610 | nkumtsam@gmail.com
LinkedIn: https://www.linkedin.com/in/nagamutyakumarkumtsam

SUMMARY

- Data Engineer with 6+ years of experience designing, building, and supporting large-scale ETL pipelines, data migrations, and cloud-based data platforms.Strong background in Python-driven data processing, ETL testing, and production support for enterprise analytics systems.Experienced in delivering cloud data solutions, supporting migration initiatives, and ensuring data reliability across complex workflows.Comfortable working in hybrid environments and collaborating closely with cross-functional teams to deliver high-impact data solutions.

PROFESSIONAL EXPERIENCE

***Data Engineer*,Applied Materials** Aug 2024 – Present

- Designed and supported cloud-based data pipelines on Azure, processing high-volume operational and manufacturing datasets for analytics and reporting
- Developed Python and SQL-based ETL/ELT workflows to ingest, transform, and validate data across batch and near-real-time pipelines
- Built and optimized scalable data processing solutions using Azure Databricks and Spark, improving pipeline performance and reliability
- Supported data lake and warehouse integrations, enabling analytics-ready datasets for downstream reporting and business intelligence use cases
- Implemented data quality checks, reconciliation logic, and schema validation to ensure accuracy and consistency across enterprise datasets
- Monitored and troubleshot production data pipelines, resolving issues related to schema evolution, job failures, and performance bottlenecks
- Collaborated closely with analytics, data science, and business teams to translate requirements into reliable data solutions
- Assisted with cloud migration initiatives, refactoring legacy pipelines to align with Azure-native architecture and best practices
- Integrated event-driven and streaming datasets into Databricks-based workflows to support near-real-time analytics
- Documented data architecture, pipelines, and operational runbooks to support maintainability and knowledge sharing

***Data Engineer*,Rapid Canvas** Sept 2022 – July 2024

- Built and maintained scalable ETL pipelines using Python and SQL to support analytics and reporting use cases across multiple business domains
- Automated data ingestion and transformation workflows, reducing manual data preparation effort by 40%
- Designed analytical datasets and data models supporting dashboards and downstream analytics workloads
- Performed ETL testing to validate data integrity, accuracy, and performance across batch pipelines
- Supported cloud migration efforts by refactoring existing pipelines to align with cloud-native design patterns
- Implemented validation and anomaly-detection logic to reduce recurring data quality issues in production datasets
- Enhanced pipeline reliability by introducing structured logging and alert mechanisms, improving issue detection and recovery time
- Collaborated with cross-functional teams to gather requirements and deliver reliable, reusable data assets
- Worked with orchestration tools such as Airflow to schedule and monitor ETL workflows
- Gained basic exposure to Kafka through ingestion of event-based datasets as part of broader batch processing workflows

***Software Engineer*,ITC Infotech, Bengaluru, India** Jul 2018 – Jul 2021

- Designed, developed, and maintained 30+ ETL pipeline**s** using SSIS, Talend, and SQL to process structured and semi-structured enterprise data

- Built data integration workflows combining procurement, vendor, purchase order, and approval workflow data for reporting and analytics
- Developed Python and SQL-based data validation routines, reducing recurring data inconsistencies and improving trust in reporting outputs
- Created dimensional and normalized data models to support analytics, improving query performance and simplifying downstream reporting
- Automated finance and operational reporting pipelines, reducing manual data preparation effort and improving data refresh reliability
- Developed Power BI dashboards to provide insights into procurement spend, operational performance, and approval timelines
- Supported data reconciliation and audit processes by implementing checks for completeness, accuracy, and consistency across datasets
- Collaborated with business stakeholders to translate functional requirements into scalable ETL and reporting solutions
- Assisted with performance tuning of SQL queries and ETL jobs to support growing data volumes
- Documented data workflows, transformations, and business rules to support long-term maintenance and knowledge transfer

PROJECTS

Predicting Customer Bookings (Python, Machine Learning, Pandas, Scikit-learn)

- Developed predictive models using Python, Pandas, and Scikit-learn, achieving ~90% accuracy on the Kaggle Hotel Booking Demand dataset
- Applied probability, statistics, feature engineering, and algorithm optimization to improve model performance and reliability
- Designed data preprocessing pipelines and evaluation metrics to support data-driven decision making

EDUCATION

M.S. in Computer Engineering – The University of Texas at Dallas, May 2023
B.Tech in Electronics and Communication Engineering – GITAM University, May 2018

SKILLS

- **Languages:** SQL, Python, C#, Java
- **Data Engineering & Warehousing:** Databricks, SSIS, Talend, Azure Data Factory
- **Cloud & ETL Platforms:** AWS (S3, Glue, Redshift, Lambda),Azure Databricks
- **Orchestration & Streaming:** Airflow, Kafka
- **Analytics & BI:** Power BI, Data Modeling (Star/Snowflake Schema)
- **Tools & Practices:** ETL Testing, CI/CD, Git, Linux, Agile