MANISH REDDY GOLI

+1 (601) 699 2338 | manishgolid@gmail.com | Piscataway, NJ | LinkedIn

SUMMARY

- Data Engineer with over 4 years of experience in building scalable data pipelines and AI-powered systems for companies like Microsoft and Amazon. Key achievements include:
- Architected end-to-end data pipelines using Spark, Kafka, and Hadoop, processing 5TB of data daily and reducing data latency by 40% for real-time analytics dashboards.
- Deployed machine learning models leveraging Python, R, TensorFlow, and PyTorch that streamlined data-driven decisions, delivering actionable insights to the executive team within 24 hours.
- Designed and deployed cloud-native applications on AWS, utilizing EC2, Lambda, S3, and API Gateway to streamline software delivery and improve operational efficiency.
- Optimized data pipelines through Python and Apache Airflow, reducing throughput time by 30% and ensuring effortless synchronization of real-time applications.
- Led the development of over 15 interactive dashboards in Tableau and Power BI, providing stakeholders with actionable insights and improving key business metrics by 15%.
- Orchestrated efforts across engineering, product, and cross-disciplinary data experts to establish data engineering standards, accelerating ETL workflow deployment speed by 40% and directly driving business expansion.
- Established a fully automated CI/CD pipeline with Jenkins and GitHub Actions, slashing the mean time to resolution (MTTR) for critical bugs by 20% and boosting overall system reliability.
- Engineered and maintained the front-end architecture using React.js, JavaScript, HTML, and CSS, ensuring a 99.9% uptime and optimizing web application performance metrics by 40%.

TECHNICAL SKILLS

- **Languages:** Python (Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch), C/C++, SQL, R, Java(Basics).
- **Front-End:** HTML5, CSS, Bootstrap, JavaScript, React.js, TypeScript.
- **Databases:** MySQL, PostgreSQL, MongoDB, Microsoft SQL Server, Oracle.
- **Big Data Technologies:** Data Modeling, Data governance, Cloud-based data storage, Data Visualization and reporting Data Profiling, Validation and reconciliation.
- **Cloud:** AWS (EC2, S3, Lambda, ECS, SQS, DynamoDB, API Gateway), Azure (App Services, SQL Server, Functions), Docker, CI/CD (Jenkins, GitHub Actions)
- **Data Engineering Tools:** Apache Airflow, Talend, ETL processes, Data Warehousing, Data Modeling, Tableau, Power BI
- **Version Control:** Git, GitHub, GitLab.
- **Tools & Methodologies**: JIRA, Agile (Scrum & Kanban), System Design, Redis, Hadoop, Spark, Kafka.
- **Machine Learning:** Supervised & Unsupervised Learning, Classification, Regression, Clustering, Deep Learning (CNN), Natural Language Processing (NLP)
- **Software Engineering & System Design:** Object-Oriented Programming (OOP), Microservices Architecture, Design Patterns, REST & GraphQL APIs, Redis, Multithreading, Event-Driven System

WORK EXPERIENCE

Client: University of Southern Mississippi   **Aug 2024 – Present**
Role: Graduate Assistant
Key Responsibilities:

- Facilitated 20+ computer hardware and software malfunctions weekly, decreasing student downtime by 30% and enabling continuous access to essential academic resources and a smooth learning environment.
- Instituted Python-based data validation checks, reducing data errors by 25% and improving the reliability of research findings to Communicate the primary root causes of crashes in the lab.
- Pioneered ETL pipeline optimization with Python, cutting data preprocessing time by 30% and resolving 15+ technical issues weekly, while guiding 10+ students in Python fundamentals.
- Enhanced data processing efficiency by automating ETL workflows with Python and Apache Airflow, reducing pipeline execution time by 25% and guaranteeing flawless real-time data analysis.

Client: ToursLog   **Jun 2022 – May 2023**
Role: Data Engineer Intern
Key Responsibilities:
- Migrated and Streamlined large-scale data pipelines from Cloudera to AWS (EMR, S3, Redshift), improving data accessibility, scalability, and throughput for real-time decision-making.
- Streamlined data pipelines by optimizing Spark scripts through caching and serialization techniques, decreasing processing time by 30% and improving workflow reliability across critical data transformation jobs.
- Led the automation of ETL workflows with Apache Airflow, with resulting in a 40% reduction in data processing time and improved system uptime to 99.9% while standardizing data engineering principles.
- Architected resilient data pipelines utilizing PySpark, SQL, HBase, and Kafka, partnering with business intelligence and data science teams that reduced data processing latency by 30%, which enabled real-time analytics dashboards for stakeholders, enhancing decision-making.
- Deployed real-time monitoring dashboards utilizing Grafana and Prometheus, catching and fixing 15+ pipeline inefficiencies that improved data pipeline stability and slashed system failures by 25%.
- Visualized key performance indicators using Tableau, creating 15+ dashboards that tracked user engagement and conversion rates, informing marketing strategies and improving customer acquisition cost by 15%.

Client: Suntek Corp Pvt Ltd   **Jun 2021 – May 2022**
Role: Software Engineer Intern
Key Responsibilities
- Constructed and Revamped scalable web applications using React.js, Django, and Flask, improving page load speed by 40% and enhancing user experience.
- Conceptualized and implemented RESTful APIs and microservices using Spring Boot and FastAPI, reducing API response time by 30% and supporting high concurrency.
- Built and managed cloud-native applications on AWS (EC2, Lambda, S3, API Gateway) and implemented CI/CD pipelines with GitHub Actions and Docker, automating deployments and reducing release time by 50%.
- Optimized backend databases (PostgreSQL, MySQL, MongoDB), leading to 25% faster query performance and ensuring efficient data processing and retrieval.
- Constructed robust test suites with JUnit, PyTest, and Selenium, catching 95% of potential bugs before production and reducing post-release hotfixes by 60% in the first quarter.

PROJECTS
Scalable ETL Pipeline for Retail Sales Data Processing   **Aug 2024 – Dec 2024**
- Executed five distinct ETL workflows to consolidate sales data from CSV files and APIs into a unified MySQL, PostgreSQL database, delivering comprehensive sales performance dashboards using Tableau and Power BI.

Stock Price Prediction System with Machine Learning   **Jan 2024 – May 2024**
- Engineered a predictive modeling system for stock prices by extracting, cleaning, and transforming financial data from Yahoo Finance APIs, achieving a data processing speed of 5,000 data points per minute.
- Designed a stock price prediction system leveraging Linear Regression, LSTM, and XGBoost, achieving 15% more accurate predictions than previous models, and presenting through Flask and Docker.

Advanced Sentiment Analysis of Product Reviews   **Jul 2022 – May 2023**
- Automated sentiment scoring of product reviews with NLP, achieving 95% accuracy, enabling product managers to address the root causes of negative reviews and improve overall product satisfaction ratings.
- Unified the model with real-time data feeds, streamlining data workflows by 40% using AWS cloud services and SQL; democratized data access by creating interactive visualized key metrics with Tableau and Power BI

CERTIFICATIONS
- Python Proficiency Certification.
- AWS Certified Data Engineer - Associate.
- Database Programming with SQL Certification.
- Certified in Java Fundamentals – Oracle Academy.

EDUCATION
University of Southern Mississippi | Hattiesburg, MS   **Aug 2023 - May 2025**
Masters in Computer Science | Grade: 3.925/4
Osmania University | Hyderabad, India   **Aug 2019 – Jun 2023**
Bachelors of Engineering in Computer Science and Engineering |Grade: 3.8/4