**A close up of a logo

Description automatically generated**A logo for a company

Description automatically generatedA blue hexagon with white text

Description automatically generated

AMITHA PEDDIREDDY

(940)252-2233 | [amithapeddireddy@gmail.com](mailto:amithapeddireddy@gmail.com) | [LinkedIn](https://www.linkedin.com/in/amitha-reddy-peddireddy-652132155/) | [Portfolio](file:///Users/amithareddy/Downloads/amitha%20copy/index.html)

**PROFESSIONAL SUMMARY**

**Results-driven Data Engineer** with **5+ years** of expertise in designing and implementing data solutions across **major cloud platforms (AWS, Azure, GCP)**. Proven track record of developing **scalable ETL pipelines and real-time processing systems**, consistently achieving significant performance improvements. Advanced **analytics professional** with a Master's degree and multiple cloud certifications, combining technical mastery **in SQL, Python, Spark, and Kafka** with strong business acumen. Experienced in translating complex data requirements into efficient, production-ready solutions for enterprise organizations.

# TECHNICAL SKILLS

**Programming Languages**: Python, SQL, T-SQL, JavaScript, R, Scala

**Big Data & Processing**: Apache Spark, Hadoop, Hive, Flink, Kafka, Beam, AWS Kinesis, Azure Event Hubs

**Cloud Platforms**: AWS (S3, EMR, Redshift, DynamoDB, SQS, Lambda, EC2, IAM, Key Management Service), GCP (BigQuery, Dataflow, Pub/Sub, Google Cloud Storage), Microsoft Azure (Data Lake, Cosmos DB, Azure Databricks, Azure Data Factory, Azure Functions, Logic Apps), IAAS

**ETL & Data Integration**: Apache NiFi, Talend, Informatica, AWS Glue, Azure Data Factory

**Databases**: MySQL, PostgreSQL, Oracle, SQL Server, Cassandra, MongoDB, HBase, NoSQL, Teradata, IBM DB2

**Data Warehousing**: Amazon Redshift, Google BigQuery, Snowflake, Databricks, Azure Synapse Analytics

**Containerization & Orchestration**: Docker, Kubernetes, Apache Airflow

**DevOps & CI/CD**: Jenkins, GitLab CI, DevOps, GitHub, AWS CloudFormation, Terraform

**Scripting Languages & Shell Environments**: Bash, Unix Shell, PowerShell

**Data Governance & Compliance**: Apache Atlas, Collibra, GDPR, HIPAA

**Data Visualization**: Tableau, Power BI, Looker, Excel

**Data Modeling Tools**: Erwin R2, ER Studio, Power Designer

**Power BI External Tools**: DAX Studio, Tabular Editor, ALM Toolkit

**Version Control**: Git, SVN

**Monitoring Tools**: Splunk, Grafana, Azure Log Analytics

**Agile & Project Management**: Agile, Scrum, JIRA, Kanban

**Soft Skills**: Problem-Solving, Troubleshoot, Collaboration, Critical Thinking, Code Review, Time Management, Debugging

# WORK EXPERIENCE

**American Airlines| Data Engineer** Feb 2024 – Current

* Contributed to data migration using Azure Databricks, Snowflake, and Data Lake Storage, implementing PySpark ETL pipelines with Delta Lake optimizations, improving processing speed by 40% and enabling modular DBT and Spark SQL transformations for consistency and lineage tracking.
* Built scalable ETL workflows using Apache Airflow and Azure Data Factory, ensuring automated orchestration, data validation, and monitoring with Python (pandas, NumPy) and Great Expectations.
* Designed dimensional and relational data models (Star & Snowflake schemas) to improve query efficiency and report generation, raising performance by 20% via indexing, partitioning, and denormalization.
* Built real-time streaming pipelines with Azure Event Hubs, Kafka, and Snowpipe, achieving 99.9% data accuracy by integrating Great Expectations for data validation.
* Architected a Lambda-based data processing framework to handle 10+ TB of daily data, integrating batch and stream processing with Apache Spark & Python, enhancing scalability and performance.
* Automated infrastructure provisioning with Terraform, Docker, and Kubernetes, integrating CI/CD pipelines in Azure DevOps & GitHub Actions and leveraging Git-based version control, reducing deployment time by 65%.
* Implemented automated monitoring and alerting using Python scripts to detect pipeline failures, performance issues, and data quality anomalies.
* Deployed data governance frameworks with Unity Catalog in Databricks, enforcing row-level security and compliance while using Python, PowerShell, and Bash for automated validation and log analysis.
* Integrated Power BI and Tableau with data platforms, creating optimized SQL views and Python-based custom functions to deliver actionable business insights.
* Actively contributed in Agile-based sprint planning, backlog grooming, and stand-ups, ensuring iterative delivery of Python-powered data pipelines and platform enhancements.

**Optum| Data Engineer** Jan 2021 – Dec 2022

* Architected a unified patient data platform using AWS Lake Formation, Amazon EMR, Redshift, and GCP BigQuery, implementing OLAP-based dimensional modeling (star schema, partitioning, and aggregations) to consolidate 50M+ patient records, reducing data access time by 60%.
* Developed automated data quality frameworks using Python, SQL, PySpark, AWS Glue, and Google Dataflow, integrating dbt and Great Expectations, achieving 99.9% PHI detection accuracy.
* Engineered a real-time patient monitoring system processing 500K+ events per minute from medical devices & EHR systems using Amazon Kinesis, Kafka, AWS Lambda, and Google Pub/Sub.
* Built an ETL integration testing framework with pytest, Jenkins, and Cloud Build (GCP), ensuring 95% test coverage and enabling bi-weekly releases through Agile/Scrum.
* Designed a data reconciliation framework (Pandas, NumPy) for automated profiling & validation, reducing manual QA effort by 70%.
* Automated metadata management via AWS Glue Data Catalog, GCP Data Catalog, IAM, and Python Flask APIs, enabling data lineage tracking and HIPAA compliance reporting.
* Implemented CI/CD pipelines with GitHub Actions, Terraform, AWS CloudFormation, and GCP Deployment Manager, reducing deployment cycles from weeks to days.
* Optimized terabyte-scale data pipelines with Apache Spark, S3, Hive, and Google Cloud Storage, leveraging OLAP-based aggregations to improve query performance for analytics.
* Strengthened data security & compliance by implementing AWS IAM, GCP IAM, data encryption (AWS KMS & GCP KMS), and masking strategies, ensuring HIPAA compliance.

**Capital One| Data Analyst** May 2019 – Jan 2021

* Optimized SQL queries for large datasets in SQL Server and Oracle, improving performance by 30% using advanced techniques like window functions, stored procedures, and CTEs.
* Developed business intelligence solutions with Power BI, Tableau, and Excel, leveraging DAX, PivotTables, and charts to create dynamic reporting, KPI visualization, and executive dashboards.
* Built and automated ETL pipelines using SQL and Python, improving reporting accuracy and enabling real-time analytics, reducing report generation time by 20%.
* Led A/B testing and statistical analysis with R and Python, translating complex results into clear, actionable recommendations for marketing strategies, driving a 15% improvement in customer engagement.
* Implemented data governance practices, ensuring compliance with industry standards and reducing data discrepancies by 25%.
* Automated data extraction and transformation processes using Python and SQL, saving 40% in manual effort and boosting team productivity.
* Collaborated closely with cross-functional teams (Finance, Marketing, Customer Success) to translate technical insights into easy-to-understand recommendations, improving strategic decision-making.
* Worked in an Agile environment with JIRA for sprint planning and task tracking, driving analytics deliverables and refining workflows.

# EDUCATION

**Master’s in Advance data Analytics (GPA 3.9)**

The University of North Texas, Denton, TX

# PROJECTS

**Azure Movie Recommender:** Azure Databricks (with Spark ML), Azure Data Factory, Azure Storage, Azure Logic Apps, Azure Resource Group

**Collision Severity on Local Roadways:** Machine learning frameworks (for Logistic Regression and Decision Trees), statistical analysis tools, data visualization software.

**Layoff – Analysis:** SQL, Python, statistical libraries (for EDA and ANOVA), and modeling tools for Linear Regression.

**Car Sales:** PowerBI Desktop, DAX for calculations, and Excel, SQL for data preparation and management.

**Dallas Police Arrest Dataset:** Tableau and Tableau Prep for data preparation, analysis, and interactive dashboard creation.