AMITHA REDDY PEDDIREDDY

(940)-274-1476| amithapeddireddy@gmail.com| [LinkedIn](#)|

SUMMARY

Data Engineer with expertise in designing and optimizing scalable **data pipelines**, managing **ETL workflows**, and ensuring high-performance data processing. I specialize in **cloud platforms (AWS, Azure, GCP)**, **Big Data technologies (Spark, Databricks, Kafka)**, and **SQL/NoSQL databases**. In addition to data integration, I leverage **automation** to streamline data workflows and ensure system reliability. With knowledge of **machine learning**, I support data-driven decision-making by building robust data infrastructures that enable analytics and model deployment at scale. I am committed to creating efficient, scalable data solutions that drive business growth.

SKILLS

- **Programming Languages:** Python, SQL, T-SQL, R, Scala, Linux, Unix
- **Big Data Processing:** Apache Hadoop, Spark, Flink, Hive, HBase
- **Data Streaming:** Apache Kafka, Storm, Flink, Beam
- **ETL & Data Integration:** Apache NiFi, Talend, Informatica, AWS Glue, Azure Data Factory
- **Cloud Platforms:** AWS (S3, EMR, Redshift, DynamoDB, Lambda, EC2), Google Cloud (Big Query, Dataflow), Azure (Data Lake, Databricks)
- **Databases:** MySQL, PostgreSQL, Oracle, SQL Server, Cassandra, MongoDB, Teradata, IBM DB2
- **Data Warehousing:** Redshift, BigQuery, Snowflake, Databricks
- **Containerization & Orchestration:** Docker, Kubernetes, Apache Airflow
- **DevOps & CI/CD:** Jenkins, GitLab CI, GitHub, AWS CloudFormation, Terraform
- **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
- **Data Analysis & Statistical Tools:** Pandas, NumPy, SPSS, MATLAB
- **Data Cleansing & Preparation:** Data wrangling, normalization, transformation
- **Soft Skills:** Problem-Solving, Troubleshooting, Collaboration

WORK EXPERIENCE

Data Engineer | American Airlines | Dallas, TX   *Feb 2024 - Current*

- Led the migration of large-scale data from **Snowflake to Azure Databricks**, managing the smooth transition of raw data from **Azure Blob Storage** and building robust **ETL pipelines** in Databricks with **PySpark**.
- Streamlined **ETL processes** by automating workflows with **Azure Data Factory**, ensuring reliability and quickly addressing issues through detailed log analysis.
- Refined **SQL-based transformations** with **DBT** and **Python**, converting legacy Snowflake scripts into modular, high-performance SQL code within Databricks using **Delta Lake** features.
- Created **Python scripts** with **Pandas** and **NumPy** for thorough data validation and reconciliation, guaranteeing accuracy across both Azure Databricks and Snowflake.
- Spearheaded seamless **data migration** across development, staging, and production environments using **Azure Data Factory's dynamic pipeline generation**, automating synchronization and data processes.
- Optimized storage and query performance by processing diverse **data formats (Parquet, Avro, JSON, CSV, XML)** with **PySpark** within **Azure Databricks**.
- Enabled real-time data streaming by integrating **Azure Event Hubs**, automating workflows through **Azure Functions** and **Azure Logic Apps** for improved data accessibility and integration.
- Automated infrastructure setup using **Azure Resource Manager (ARM) templates** and **Terraform**, ensuring consistent deployment across all environments.
- Oversaw system administration tasks and orchestrated **data pipelines** using **PowerShell** (Windows) and **Bash** (Unix/Linux), automating server management, log monitoring, and **ETL scheduling** for seamless data operations.

Data Engineer |Optum Health| Hyderabad, TN   *Jan 2021- Dec 2022*

- Designed and implemented **scalable, cloud-based data pipelines** using **Azure Data Factory and Azure Stream Analytics** to collect, transform, and load large-scale clinical trial audit data, ensuring real-time data processing and compl**iance**.
- Developed and managed **OLAP and OLTP solutions** within Azure Synapse and Azure SQL Database, optimizing data accessibility, reporting efficiency, and real-time decision-making for healthcare analytics.
- Engineered **end-to-end ETL workflows** using SSIS and Azure Data Factory, integrating Azure Blob Storage and Azure Functions to handle millions of data points per second, ensuring high data integrity and availability.
- Leveraged **Azure Databricks with PySpark** for distributed processing of terabytes of healthcare data, enhancing data accuracy, transformation pipelines, and advanced analytics capabilities.

- Designed **predictive analytics models** utilizing **TensorFlow, PyTorch, and Scikit-learn** for clustering and decision trees, optimizing clinical trial scheduling and operational efficiencies.
- Developed and maintained **CI/CD pipelines** with **Azure DevOps and Jenkins**, automating data pipeline deployments, testing, and version control, leading to faster and more reliable data integrations.
- Created interactive, KPI-driven dashboards in **Power BI** to provide data-driven insights for clinical researchers, compliance teams, and operational leaders, enhancing decision-making and performance monitoring.
- Implemented containerized data processing solutions using **Docker** and **Azure Kubernetes Service (AKS)**, enabling scalability and automation of data workloads.
- Monitored pipeline performance and system health using **Azure Monitor and Grafana**, setting up automated alerts for error detection and SLA compliance, reducing data latency and downtime.

Data Analyst| Capital One| Bangalore, KN   *May 2019 - Jan 2021*

- Designed and optimized complex SQL queries to extract, transform, and analyze large datasets, utilizing **window functions, stored procedures, CTEs, and advanced joins** to improve performance and efficiency.
- Developed and maintained **data pipelines** to ensure seamless data flow across business units, supporting key decision-making processes.
- Created and automated interactive **Power BI and Tableau dashboards**, leveraging **DAX functions** for dynamic calculations and performance optimization.
- Leveraged **Excel functionalities**, including **pivot tables, VLOOKUP, macros, and data validation**, to streamline data processing and reporting.
- Automated recurring data extraction and transformation processes using **Python and SQL**, reducing manual effort and improving accuracy.
- Collaborated closely with **business users and senior stakeholders** to gather requirements, define KPIs, and deliver data-driven solutions that aligned with business objectives.
- Conducted **A/B testing and statistical analysis** to evaluate product and marketing strategies, helping optimize customer engagement.
- Ensured data integrity, security, and compliance with internal policies and industry regulations, supporting risk management initiatives.
- Documented analytical processes, workflows, and methodologies to enhance knowledge sharing and maintain data governance best practices.

EDUCATION

Master's in Advance data Analytics   **Jan 2023 - Dec 2024**
The University of North Texas, Denton, TX   GPA: 3.9

LICENSES AND CERTIFICATE

- **Google Data Analytics** Certification Link
- **Google Analytics** Certification Skill shop Link
- **Microsoft Certified** Power BI Data Analyst Associate Link
- Certifications in **MySQL**, **Data Visualization using Tableau and Power BI**, **Python**, and **SAS**.