SHIVA SAI AMARAVADI

Lewisville, TX |+1 (940)-843-8137 | amaravadishivasai@gmail.com | LinkedIn

---

PROFESSIONAL SUMMARY

Results-driven Data Engineer with 3+ years of experience designing and deploying cloud-native data platforms, specializing in AWS-based solutions. Expertise in building scalable ETL/ELT pipelines, data lakehouse architectures, and real-time streaming systems using AWS Glue, Redshift, S3, Lambda, Kinesis, Step Functions, and DynamoDB. Skilled in Databricks for big data analytics and machine learning workflows. Proficient in SQL, Python, PySpark, Terraform, and CI/CD automation, enabling robust, secure, and high-performing data delivery. Adept at implementing data governance, HIPAA/GDPR compliance, and security best practices to support business-critical analytics. Additionally experienced in Azure Data Factory and Synapse for integration and reporting when required.

TECHNICAL SKILLS

Data Analysis & Visualization: SQL, Tableau, Power BI, Looker, Amazon QuickSight, Microsoft Excel (Power Query, Pivot Tables), Data Storytelling, KPI Reporting, Dashboard Development.
Programming & Analytics: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn), R (tidyverse, ggplot2), Exploratory Data Analysis (EDA), Statistical Modeling, Hypothesis Testing, A/B Testing.
Data Engineering & Pipelines: Apache Spark, PySpark, Airflow, dbt, AWS Glue, AWS Lambda, Amazon S3, Redshift, DynamoDB, Kafka, Kinesis, Azure Data Factory, Google Cloud Dataflow, ETL/ELT Development.
Cloud & Big Data Platforms: Snowflake, Databricks, AWS (S3, Glue, Redshift, EMR, Kinesis), Azure Synapse, Azure Data Lake, Google BigQuery, Google Cloud Storage.
Databases & Storage: PostgreSQL, MySQL, SQL Server, Oracle, MongoDB, Cassandra, Data Lakehouse Architectures (Delta Lake, Iceberg, Hudi).
Data Governance & Security: Data Quality, Data Lineage, Data Catalogs, PHI/PII Handling, GDPR, HIPAA, Role-Based Access Control (RBAC), IAM, Encryption (AWS KMS).
Collaboration & DevOps: Git/GitHub, GitLab CI/CD, Jenkins, Docker, Kubernetes, JIRA, Confluence, Agile/Scrum.

PROFESSIONAL EXPERIENCE

Data Engineer   **Sep 2024 - Present**
Micron Technology | St. Louis, MO, USA
- Designed and deployed AWS Glue, Lambda, and Step Functions pipelines processing 25M+ weekly healthcare and financial records, improving ETL reliability and cutting data latency by 50%.
- Built and optimized Redshift and S3-based data lakehouses supporting both batch and real-time ingestion, enabling analysts to query cross-domain insights with 40% faster response times.
- Leveraged Kinesis Data Streams and Firehose for real-time ingestion of claims and operational data, ensuring low-latency delivery for downstream analytics.
- Automated infrastructure provisioning using Terraform and AWS CDK integrated with GitHub Actions, reducing deployment cycles by 35% and improving consistency across environments.
- Secured sensitive data with IAM roles, AWS KMS encryption, and VPC isolation, achieving 100% HIPAA/GDPR compliance in US-based workloads.
- Developed CloudWatch and CloudTrail monitoring dashboards to proactively detect pipeline failures and SLA risks, reducing downtime incidents significantly.
- Collaborated with analysts to implement dimensional data models and star schemas in Redshift, standardizing KPIs and reducing manual SQL queries by 30%.
- Implemented PySpark-based validation frameworks on Glue jobs with Great Expectations, reducing bad records by 25% and ensuring data quality for executive reporting.
- Additionally delivered limited Azure Data Factory and Synapse Analytics pipelines for specific cross-platform integration and reporting requirements.

Data Engineer   **Jul 2021 - Aug 2023**
We Agency | Hyderabad, India
- Built scalable ETL workflows on AWS Glue, Lambda, and S3 to integrate 5-10 GB of daily customer and product data, ensuring analytics-ready datasets for sales and product teams.
- Optimized Amazon Redshift schemas and SQL queries, reducing dashboard load times by 40% and improving executive access to near real-time insights.
- Developed streaming pipelines with Kafka and Kinesis to monitor 100K+ daily customer interactions, strengthening fraud detection and real-time analytics.
- Migrated 500+ GB of legacy data into Amazon S3 and DynamoDB using PySpark on Glue, reducing query latency by 30% and improving downstream analytics accuracy.
- Automated ETL triggers and monitoring with Lambda + CloudWatch, cutting pipeline failures by 30% and eliminating 15+ hours of weekly manual fixes.
- Designed Power BI and Tableau dashboards covering churn, adoption, and ROI, improving reporting efficiency by 35% and guiding data-driven decision-making.
- Applied AWS KMS encryption, IAM policies, and PII masking, ensuring GDPR/HIPAA compliance while securing customer data for analytics workloads.

PROJECTS

Real-Time Flight Operations Data Platform
- Built ingestion pipelines with Azure Data Factory, Kafka, and Spark Streaming to capture flight, baggage, and customer events in real time, maintaining 99.9% uptime across airline operations.
- Implemented a Databricks Lakehouse with Delta Tables and Synapse that supported both batch and streaming queries, improving analytics speed by 40% and enabling faster incident response.
- Developed Power BI dashboards with live KPI tracking and integrated TensorFlow ML models into MLflow, increasing flight delay prediction accuracy by 20% and reducing operational disruptions.

Healthcare Risk Adjustment Analytics
- Centralized clinical and claims data into Snowflake using Azure Data Factory and Talend, streamlining compliance workflows and reducing manual data prep time by 30%.
- Automated STAR ratings and risk adjustment models with PySpark, improving reporting cycles by 25% and ensuring timely delivery of regulatory metrics.
- Designed interactive Tableau dashboards for patient risk stratification, helping identify high-risk groups and increasing accuracy of compliance reporting by 25%.

EDUCATION

Master of Science in Data Engineering   **Aug 2023 - May 2025**
University of North Texas | Denton, TX

CERTIFICATIONS

- Google Data Analytics Professional Certificate **- Coursera**
- Microsoft Power BI Data Analyst Associate **- LinkedIn Learning**
- AWS Data Analytics Specialty **- Coursera**
- Data Engineering with Google Cloud Professional Certificate **- Coursera**
- Tableau Data Analyst Professional Certificate **- Coursera**