

BHARADWAJ BAIRI

Denton, Texas, USA | bharadwajbairi3@gmail.com | +1-660-528-8192
linkedin.com/in/bharadwajbairi | github.com/bharadwajbairi3

SUMMARY

AWS Certified Data Engineer with 3+ years of experience designing and implementing scalable cloud-native data pipelines on AWS and GCP. Proven ability to deliver ETL workflows that improved ingestion speeds by up to 67% and reduced compute costs by 25%. Proficient in Python, SQL, Spark, Airflow, and streaming technologies like Kinesis and Kafka, committed to building cost-efficient data solutions that drive business insights.

WORK EXPERIENCE

Data Engineer | Cloud-Native ETL & Streaming Pipelines

Jan 2025 - Present

SAAS Tech | NJ, USA

- Designed and optimized ETL pipelines using AWS Glue, Apache Spark, and Hive, improving processing efficiency by 30% through Glue-Spark integration and partition pruning.
- Automated MySQL-to-S3 and S3-to-Snowflake ingestion workflows using Airflow and Lambda, reducing sync time by 2x and ensuring fault-tolerant scheduling.
- Built real-time streaming pipelines with AWS Kinesis, Kafka, and Lambda, reducing data lag by 40%, and optimized cold storage using S3 lifecycle policies.

Data Engineer | Enterprise Data Migration & Automation

May 2022 - Dec 2023

Mayo Clinic | Hyderabad, India

- Built 20+ ETL pipelines using Python and Airflow, processing data from 5-6 sources to improve data integrity by 30% and increase conversion rates by 16%.
- Migrated legacy systems from Oracle and SQL to BigQuery and Amazon Redshift, resulting in annual savings of \$678K, utilizing Lambda, S3, and Glue triggers.
- Processed 10 TB of daily data with Python and AWS Glue, improving ingestion speed by 67%, and integrated AWS Data Pipeline and GCP Data Proc for Spark job orchestration.

Junior Data Analyst Intern | SQL-Driven Reporting & Cleansing

Mar 2021 - Mar 2022

Vak Tech INC | Hyderabad, India

- Facilitated data cleaning and analysis using SQL and Excel to ensure data accuracy and integrity.
- Delivered comprehensive reports and summaries which improved team task efficiency by 15%.
- Formulated SQL queries that streamlined the extraction process, reducing data retrieval time by 30%.

PROJECTS

AWS Serverless Student Data Pipeline

Jul 2025

- Designed and implemented a scalable AWS serverless data pipeline for real-time and batch processing of simulated student academic data (up to 100,000 records daily) using DynamoDB, Lambda, Glue, and Athena.

Spotify ETL: AWS Music Metadata Pipeline

Jun 2025

- Developed a serverless AWS ETL pipeline to integrate 3 distinct music metadata datasets, leveraging Glue for Parquet conversion (reducing storage by 70%) and QuickSight for 5 interactive dashboards.

SKILLS

Programming: Python, SQL, PySpark, NoSQL

Databases & Warehousing: Snowflake, Amazon Redshift, BigQuery, PostgreSQL, Relational Databases

Cloud Platforms: AWS (S3, Glue, Lambda, Kinesis, EMR, Athena, Lakeformation, CloudWatch), GCP (BigQuery, Cloud Composer, GCS)

Data Orchestration & Streaming: Apache Kafka, Apache Airflow, Spark Streaming, AWS Stepfunctions

Containerization: Kubernetes, Docker

Reporting & Visualization: Amazon QuickSight, Tableau, SQL-Based Dashboards, BigQuery Materialized Views

Version Control & CI/CD: Git, Jenkins

Development Methodologies: Agile, Scrum

EDUCATION

Master of Science: Information Systems

Jan 2024 - May 2025

Northwest Missouri State University, Missouri

Bachelor of Engineering: Civil Engineering

Jun 2018 - Mar 2022

Jawaharlal Nehru Technological University Hyderabad, Telangana

CERTIFICATIONS

AWS Certified Data Engineer Associate

Jul 2025

Snowflake SnowPro Core Certification

Jul 2025