BHARADWAJ BAIRI

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

SUMMARY

AWS Certified Data Engineer with 4+ years of experience designing and implementing cloud-native data pipelines on AWS and GCP. Delivered scalable ETL workflows that improved ingestion speeds by up to 67% and reduced compute costs by 25% through performance tuning and warehouse optimization. Skilled in Python, SQL, Spark, Airflow, and streaming technologies like Kinesis and Kafka. Committed to delivering high-availability, cost-efficient data solutions that drive business insights and support data-driven decision-making.

WORK EXPERIENCE

Data Engineer *Jan 2024 - Present*

SAAS Tech | NJ, USA

- Designed and optimized ETL pipelines using AWS Glue, Apache Spark, and Hive; improved processing efficiency by 30% through Glue-Spark integration and partition pruning on large datasets.
- Automated MySQL-to-S3 and S3-to-Snowflake ingestion workflows using Airflow and Lambda, reducing sync time by 2x and ensuring fault-tolerant scheduling across batch pipelines.
- Built real-time streaming pipelines using AWS Kinesis, Kafka, and Lambda, reducing data lag by 40%; integrated S3 lifecycle policies to automate cost-efficient cold storage management.
- Deployed containerized ETL applications using Docker on AWS ECS with Fargate for scalability and resilience; configured real-time monitoring using CloudWatch and CloudTrail across 15+ pipelines.
- Developed analytics-ready data models in Snowflake using SQL, Python, and Airflow; tuned warehouse sizing and query performance to reduce compute costs by 25%.

Data Engineer *May 2022 - Dec 2023*

Mayo Clinic | Hyderabad, India

- Built 20+ ETL pipelines using Python and Airflow to process data from 5–6 sources; improved data integrity by 30% and increased conversion rates by 16% through efficient scheduling and transformation.
- Migrated legacy systems from Oracle and SQL to BigQuery and Amazon Redshift, saving $678K annually; designed fault-tolerant ingestion workflows using Lambda, S3, and Glue triggers.
- Processed 10 TB of data daily using Python and AWS Glue, improving ingestion speed by 67%; integrated AWS Data Pipeline and GCP Data Proc for Spark job orchestration.
- Developed and maintained scalable data pipelines with BigQuery, GCS, and AWS S3; analyzed large datasets using SQL, boosting operational efficiency by 29%.
- Automated reporting dashboards in BigQuery and QuickSight; implemented version-controlled CI/CD pipeline framework using Git, Jenkins, and Terraform for reproducibility in AWS.

Junior Data Analyst Intern *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%.
- Performed data validation and duplicate removal using Excel, enhancing dataset reliability by 20%.
- Formulated SQL queries that streamlined the extraction process, reducing data retrieval time by 30%.

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
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*

Certified in designing scalable, secure data pipelines on AWS using Glue, S3, Redshift, Kinesis, and Lake Formation to support real-time analytics and cost-optimized data lake architectures.