

Marcus Whitfield

Senior Data Engineer

Austin, TX • (512) 555-0187 • marcus.whitfield@email.com • [linkedin.com/in/marcuswhitfield](https://www.linkedin.com/in/marcuswhitfield) • github.com/mwhitfield

SUMMARY

Senior data engineer with 10 years of experience designing large-scale data platforms and leading engineering teams through migrations from legacy warehouses to modern cloud architectures. Deep background in distributed processing, pipeline orchestration, and data governance. Focused on building systems that stay reliable as data volume and team size grow.

CORE SKILLS

Languages Python, SQL, Scala, Java	Processing Apache Spark, Apache Flink, Apache Kafka
Orchestration Airflow, Dagster, dbt	Warehouses Snowflake, BigQuery, Redshift
Cloud & Infrastructure AWS, GCP, Terraform, Kubernetes	Leadership Technical mentoring, architecture reviews, roadmap planning

EXPERIENCE

Senior Data Engineer, Ridgeline Health AnalyticsSep 2020 – Present

Austin, TX

- Lead architect for a healthcare data platform processing over 500 million records monthly across claims, clinical, and billing systems, built on Snowflake and Airflow.
- Redesigned the ingestion layer using Kafka and Spark Structured Streaming, cutting data latency from 12 hours to under 15 minutes for critical reporting feeds.
- Manage technical direction for a team of five data engineers, setting architecture standards and running design reviews across two product lines.
- Led the migration from an on premises Hadoop cluster to a cloud-native stack on AWS, reducing infrastructure spend by roughly \$310,000 per year.
- Built a data governance framework covering lineage tracking and access controls, helping the company pass a HIPAA compliance audit with zero major findings.
- Mentored three engineers into senior roles through regular design reviews, code mentorship, and ownership of increasingly complex pipeline projects.

Data Engineer II, Prairie Point LogisticsJun 2017 – Sep 2020

Dallas, TX

- Built and scaled ETL pipelines in Python and Spark that processed shipment and routing data for a network of 400 distribution centers.
- Designed a dimensional data warehouse in Redshift that became the primary source for operations reporting, replacing six disconnected spreadsheets.
- Introduced automated testing and data validation for pipelines, reducing production data incidents by 65 percent over two years.

Data Engineer, Cobalt SystemsJul 2015 – Jun 2017

Dallas, TX

- Developed batch ETL jobs in Python and SQL to integrate data from five acquired companies into a unified reporting database.
- Optimized long-running SQL queries on a legacy Oracle system, cutting report generation time from 20 minutes to under 3 minutes.

EDUCATION

M.S. in Data Science, University of Texas at Austin

Completed 2015

B.S. in Computer Science, Texas A&M University

Completed 2013

CERTIFICATIONS & TALKS

AWS Certified Data Analytics – Specialty, 2022

Speaker, "Streaming Data at Scale in Healthcare," Regional Data Engineering Summit, 2023