

(205) 222-6249
San Francisco, Bay Area
Anirudhashokpatil54@gmail.com

Anirudh Patil

Data Engineer

Portfolio: anirudhpatil367.github.io/portfolio
Github: github.com/Anirudhpatil367
LinkedIn: linkedin.com/in/anirudhpatil367

PROFESSIONAL SUMMARY

Data Engineer with 2+ years of experience in building ETL pipelines, cloud-based data workflows, and transforming raw data into actionable insights using Python, SQL, and PySpark. Experienced in the healthcare and banking domains through roles at UAB and Mashreq Global Services. Seeking full-time Data Engineering roles to contribute to scalable, high-impact data solutions in fast-paced environments.

SKILLS

Programming Languages: Python, SQL, SparkSQL, Bash, HiveQL, Git, Scripting Language.
Cloud and DevOps: AWS (S3, Glue, Athena), Azure (Data Factory, Synapse), GCP (BigQuery, Dataflow).
Data Engineering: Apache Spark, Apache Kafka, Airflow, AWS Glue, Databricks, ETL Pipelines, Terraform, Apache NiFi.
Databases: PostgreSQL, PL/SQL, MongoDB, Snowflake, object storage, key-value and document db.
Data Warehousing: Amazon Redshift, Google BigQuery, Data modeling, Data Warehousing, Workflow Ochestrations.
DevOps/Tools: Docker, GitHub, Jupyter Notebook, VS Code, Statistical Tools, CI/CD, Informatica, SSIS, Datastage.
Data Visualization: Power BI, Tableau, Grafana.

TECHNICAL EXPERIENCE

Research Assistant | Research Technician | Data Engineering **Oct 2024 – Present**
University of Alabama at Birmingham
• Promoted to Research Technician in July 2025 for contributions to ETL automation and data engineering.
• Built scalable ETL pipelines in Databricks using PySpark and AWS Glue, reducing gene-protein data processing time by 40%.
• Conducted statistical analysis using Pearson's correlation and p-values, improving accuracy by 35% across gene samples.
• Designed modular storage pipelines for 10K+ new records/day, improving data retrieval performance by 30%.
Quality and Performance - Senior Associate | Data Engineer **Feb 2022 – Aug 2023**
NTT Data
• Streamlined data workflows (client data ingestion Minitab/Prosim-ready analysis SLA reporting), cutting latency by 30%.
• Ensured 99% data accuracy for analytics SLA monitoring by integrating diverse client data sources with validation rules.
• Designed monitoring dashboards and automated alerts for SLA pipelines using Power BI and Python, achieving 95% uptime.
• Improved data integrity 20% by implementing data quality checks performing RCA on anomalies SLA deviations.
Apprenticeship | Internship | Data Engineering **Apr 2021 – Feb 2022**
Mashreq global services
• Converted from Intern to Apprentice based on performance.
• connected with senior engineers to design Databricks pipelines, reducing manual effort by 25% enhancing workflow efficiency.
• Refactored db queries and transformations, boosting processing speed by 20% and improving overall system performance.
• Monitored pipeline performance, identifying optimization opportunities and reducing troubleshooting time by 15%.

PROJECTS

Spotify Data Pipeline – AWS, PySpark, Athena, Databricks ([GitHub Link](#)) **Aug 2023 – Mar 2024**
• Built a serverless ETL pipeline on AWS (Lambda, S3, Glue, Athena), reducing data latency by 60%.
• Enhanced PySpark jobs on Databricks, decreasing transformation time by 35%.
• Standardized raw Spotify data using AWS Glue , reducing data discrepancies by 40% .
• Created SQL queries in Athena to analyze 1M+ Spotify tracks, driving content strategies and improving engagement by 20%.
YouTube Comment Tracker – Kafka, Spark Streaming, AWS([GitHub Link](#)) **Mar 2023 – Aug 2023**
• Engineered a real-time pipeline using Kafka and Spark Streaming to process 1M+ YouTube interactions, reducing feedback analysis time from hours to minutes.
• Developed custom PySpark UDFs to process 12,000+ comments, boosting sentiment analysis accuracy by 30%.
• Deployed sentiment analysis models and optimized AWS S3 storage, reducing report generation time by 70%.

EDUCATION

M.S in Computer Science, *University of Alabama at Birmingham* Aug 2023 - May 2025
Minor: Data Science GPA: 3.7 / 4.0
Bachelor of Engineering in CSE, *KLE Institute of Technology, Hubli* Aug 2017 - May 2021

CERTIFICATIONS AND ACTIVITIES

AWS Certified Data Engineer - Associate July 2025 – July 2028
Google Cloud Platform Fundamentals: Core Infrastructure May 2020
Data Engineer Associate July 2025