Koushik Reddy

McKinney, TX | (316)519-0532 | [Koushik.sama9@gmail.com](mailto:Koushik.sama9@gmail.com)

**LinkedIn:** https://[www.linkedin.com/in/koushik-s-0b4048289/](http://www.linkedin.com/in/koushik-s-0b4048289/)

**PROFESSIONAL SUMMARY:** Data Engineer with 5+ years of experience building scalable ETL and real-time data solutions on GCP, with expertise in BigQuery, Dataflow, Airflow, and Spark. Proven track record in designing secure, production-grade pipelines for analytics and ML workloads, and leading automation and optimization initiatives in Agile/DevSecOps environments.

# CORE SKILLS:

* Experience working with large-scale data integration architectures, focusing on scalable, highly available ETL and real-time streaming solutions.
* Strong inclination towards learning new technologies and delivering robust cloud-native data solutions within defined timeframes.
* Ability to develop efficient, maintainable, and high-performance data engineering code using Python, SQL, Spark, and Airflow.
* Proficient in data engineering best practices, including ETL framework design, workflow orchestration, source control management, and deployment automation.
* Recognitions: Known for self-discipline, proactive problem-solving, and driving data solution standardization and automation initiatives.

# TECHNICAL SKILLS:

* Technologies**:** Python, SQL, Java, Apache Spark, Apache Airflow, Kafka, REST APIs, Terraform, CI/CD
* Cloud Platforms**:** Google Cloud (BigQuery, GCS, Pub/Sub, Dataflow, Dataproc, cloud functions), AWS, Snowflake, Hadoop
* Databases**:** MySQL, PostgreSQL, Oracle, Netezza
* Certifications: AWS cloud solutions architect – Associate, Snowflake SnowPro Core Certified.

# WORK EXPERIENCE:

**Data Engineer:** CareFirst BCBS – Baltimore, Maryland 09/13/2023 – Present

* Gathering data integration requirements from stakeholders and designing scalable ETL solutions aligned with business needs.
* Developing and maintaining Python, SQL, and Spark based ETL pipelines, implementing SCD Type 2 data models in SQL Server and Snowflake.
* Collaborated with data governance teams to validate data quality, perform unit/integration testing of data pipelines in dev/QA env, ensure compliance, and align with enterprise metadata standards.
* Building secure API integrations and automating file-based data transfers using containerized GCP applications.
* Deployed CI/CD pipelines using Terraform and Jenkins to automate GCP ETL workloads with HIPAA-compliant security.
* Performing risk and fraud analysis by validating ETL outputs, creating mock datasets, and contributing to feature engineering and model dataset preparation for ML workflows.
* Participating actively in Agile ceremonies, driving on-time delivery and continuous improvement across the data engineering lifecycle.

**Data Engineer:** Planet Fitness – Hampton, NH 12/19/2022 – 09/08/2023

* Built and modernized scalable GCP-based data lakes and data warehouses, enabling structured data access and analytics across the organization.
* Actively contributed to technical design discussions, leading improvements in data architecture, storage optimization, and query performance tuning in BigQuery.
* Built event-driven pipelines using GCP Dataflow and Pub/Sub for real-time ingestion into BigQuery, reducing Storage and cost.
* Automated application API-based data retrieval and loading processes using Cloud Functions, Python, and SQL, supporting real-time business operations and analytics.
* Developed orchestration workflows in Cloud Composer (GCP Airflow) using Python and PySpark for scalable data pipelines.
* Integrated reporting platforms like Looker with BigQuery to enable faster business insights and data visualization across multiple departments.

**Data Engineer:** USAA – San Antonio, TX 05/17/2021 – 12/16/2022

* Led the end-to-end migration of structured and semi-structured data from Oracle and Netezza to GCP Cloud, modernizing

legacy systems into a scalable cloud-based architecture.

* Developed reusable ETL frameworks for batch data processing using Python, SQL, Shell scripting, and PySpark, standardizing ingestion and transformation workflows across multiple domains.
* Automated batch and real-time pipelines using Cloud Composer and Dataflow, loading data from MySQL staging to GCP and Snowflake.
* Designed and implemented real-time streaming data pipelines using Kafka and Java, enabling continuous ingestion of critical data into GCP Cloud and Snowflake for near real-time analytics.
* Leveraged Dataproc Spark clusters for distributed batch processing of large datasets, enhancing data throughput, transformation efficiency, and supporting advanced reporting needs.
* Implemented CI/CD pipelines using Concourse CI and Terraform for GCP-based data engineering workflows.

**Data Engineer:** Techborn Software Systems Pvt. Ltd, India 05/01/2018 – 12/13/2019

* Contributed to the development of cloud-based data integration solutions supporting business reporting, analytics, and decision- making using Python, SQL, Spark, and AWS technologies.
* Built and maintained real-time data ingestion frameworks using Kafka to enable timely and actionable insights across critical operational and financial business processes.

# EDUCATION:

**Master of Science in Computer Science,** University of Missouri, Kansas city. January 2020 – May 2021