## KOUSHIK REDDY SAMA

Mobile no: 316-519-0532

LinkedIn: https://[www.linkedin.com/in/koushik-sama-0b4048289/](http://www.linkedin.com/in/koushik-sama-0b4048289/) Email: [**Koushik.sama9@gmail.com**](mailto:Koushik.sama9@gmail.com)

# EDUCATION

**Master of Science in Computer Science Jan 2020 – May 2021**

University of Missouri, Kansas City

# PROFESSIONAL SUMMARY

Data Engineer with around 5+ years of expertise in data solutions, specializing in ETL optimization and infrastructure design. Proficient in Python and SQL for data analysis and manipulation. Skilled in managing data warehouses, including Snowflake, and adept at ensuring data integrity and security. Experienced in migration and integration projects and proficient in cloud technologies, with a focus on AWS and Azure. Eager to leverage technical skills to drive data excellence in diverse projects.

# TECHNICAL SKILLS/CERTIFICATION:

* **Technologies**: Python, SQL, ETL, SQL/NoSQL DBs, CICD, GIT, Restful APIs, Control-M, Airflow, AWS Services, Hadoop Map Reduce, Spark, Pyspark, Databricks, Snowflake, Informatica, PL/SQL, Power BI, Tableau, business intelligence.
* **Version Control Tools:** GIT, SVN, TFS, Docker, Jenkins.
* **SCM:** Github, Bitbucket, GitLab, Azure Devops
* **Mics:** Data Structures, Algorithms, problem solving, Design Patterns and Data Modelling.
* **Databases:** MySQL, SQL Server, Oracle, DB2, PostgreSQL, Netezza, **Cloud**: AWS, Azure
* **Operating Systems:** Windows, Linux, Unix
* **Certifications: AWS Certified Solutions Architect – Associate, Snowflake - Snowpro core Certification**

# PROFESSIONAL EXPERIENCE

## CAREFIRST BCBS

**Data Engineer Sep 2023 –Present**

* + Worked on Multiple multi-phase data integration projects, delivering **risk** **analysis** **data** for the business **LexisNexis** **Risk** **Solutions** and **Medicare**, **Membership**, **Health** **insurance** and **claims** data for the **Optum** client.
  + Designed data models for SQL Server and Snowflake tables to implement SCD Type 2, for membership data integration.
  + Developed and optimized ETL workflows using **SSIS**, designing and implementing complex **Data flow tasks, components** and

**ETL jobs** to efficiently integrate Membership and claims data from source systems into **Snowflake** and **SQL Server**.

* + Developed complex SQL queries to source data from Snowflake, utilizing advanced techniques Partitioning, window functions, **CTEs**, subqueries, and joins to optimize data retrieval and ensure accuracy in ETL processes.
  + Contributed to **GenAI** initiatives by configuring Snowflake Semantic Model Generator and LLM YAML files for prompt-driven query generation.
  + Utilized **Tableau** to connect to **Snowflake**, creating interactive dashboards to visualize **claims** and **member enrollment**

data, providing real-time insights that supported data-driven decision-making and optimized healthcare operations.

* + Developed **Python RESTful APIs** for secure data integration, automating data retrieval and seamless transfer of healthcare and claims data into **Snowflake** for ETL processing.
  + Engineered DevOps Ingestor tools for creating tables and loading data in Snowflake and for Azure storage integration.
  + Completed **HIPAA** compliance training, ensuring adherence to healthcare privacy and security standards.

**Technologies Used**: Snowflake, Python, Azure Devops, SQL, Informatica IICS, SSIS, SQL Server Management Studio, Datawarehouse solutions, Data Modelling, SCD, CDC, CI/CD, GenAI, LLMs, Data Visualization, Tableau, SDLC, Agile.

## PLANET FITNESS

**AWS Python Developer Dec 2022 –Sep 2023**

* + Worked on Data Modernization project, standardizing data transfers from **DynamoDB** to **AWS S3** and **Snowflake** Data Warehouse.
  + Re-engineered Snowflake Stored Procedures, optimizing query performance and indexing.
  + Automated data exports using AWS Lambda and Firehose Kinesis for real-time data transfer.
  + Developed AWS **Glue** ETL jobs for data processing and orchestrated jobs with Step Functions.
  + Utilized **Databricks** with **Apache Spark** for real-time data ingestion and processing, integrating seamlessly with Kinesis and S3 to enable high-speed data streaming and scalable processing.
  + All the scripts are developed in **Python**, **PySpark**, and shell scripting, leveraging AWS EMR clusters for large dataset processing and connected **Tableau** servers to AWS **Athena** for querying.

**Technologies Used**: Hadoop, HDFS, SQL, Python, Pyspark, Apache Spark, Snowflake, Databricks, AWS Dynamo DB, Lambda, Step Functions, Firehouse Kinesis, Glue jobs, APIs, Athena, Tableau, S3, CI/CD, SDLC, Agile.

## USAA

**Data Engineer May 2021 – Dec 2022**

* + Worked on end-to-end data migration from **Oracle** and **Netezza** to AWS **Snowflake** andintegrationproject.
  + Automated ETL processes using **Python, PySpark, Spark, SQL,** and **Unix** **Shell scripting**, developing reusable and robust migration scripts to efficiently process and transform large-scale datasets and utilized IICS tool for ETL integration work.
  + Utilized Control-M for developing batch services/user interface for single and bulk asset registrations and Airflow for job scheduling, creating dags and ensuring seamless migration execution.
  + Created complex queries and stored procedures in Oracle PL/SQL for efficient data extraction and transformation.
  + Designed and developed **Python RESTful APIs** for data integration between Oracle and AWS Snowflake.
  + Leveraged **Kafka producers, consumers, and topics** to design real-time data streaming pipelines, efficiently ingesting and transforming data into **Snowflake**, ensuring high performance and adherence to best data warehouse practices.
  + Utilized **Snowpipe** for real-time data ingestion into AWS Snowflake during the migration from Oracle and Netezza, enhancing data availability and processing efficiency.
  + Employed **dbt (Data Build Tool)** to optimize data models and transformations within **Snowflake** post-migration, ensuring efficient data structuring and performance for analytics by automating and modularizing the transformation process.
  + **Utilized Power BI** to create real-time dashboards and reports from **Snowflake**, enabling business stakeholders to track migration progress and monitor data quality metrics during the Oracle and Netezza migration.

**Technologies Used:** Hadoop, HDFS, Hive, SQL, Snowflake, Python, Pyspark, Spark, Shell, AWS S3, IAM, Snowflake, Unix scripts, Kafka, Control-M, DBT, Oracle, PL/SQL, Netezza, IICS, Power BI, CSV Files, XML files, CI/CD, SDLC, Agile.

**Techbion Software Systems Private Limited**

**Data Engineer May 2018 – Dec 2019**

* + Migrated enterprise data from legacy **DB2** systems to Azure Data Lake Storage (**ADLS**), enabling scalable and centralized storage for downstream analytics.
  + Built and maintained **ADF** and **SSIS** pipelines to move data across Teradata, DB2, SQL Server, and cloud platforms like Azure Synapse and Blob Storage using PolyBase and Delta Lake formats.
  + Developed Databricks notebooks for data transformation and enrichment, supporting complex ingestion from structured and semi-structured sources.
  + Utilized Python, PySpark, and SQL within Databricks notebooks to build robust data transformation logic, including handling large-scale joins, aggregations, and data validations before loading into Azure Data Lake and Synapse.
  + Integrated Azure Functions and custom Unix scripts for file validation, JSON file handling, and event-triggered processing via webhooks and API endpoints.
  + Underwent rigorous internal training and evaluation tests on Azure Data Engineering and passed performance benchmarks required to continue on the project.

**Technologies Used:** Azure Data Factory (ADF), Azure Data Lake Storage (ADLS), Azure Synapse, Databricks, Python, PySpark, SQL, DB2, Teradata, SQL Server, PolyBase, Unix Shell Scripting, Azure Functions, SSIS, JSON, Delta Lake.