

GOKUL SUBRAMANIAN | DATA ENGINEER

+971 54 302 4015 | gokul12cool@gmail.com | Dubai, UAE (*Open to relocation to Poland or Spain*)

<https://www.linkedin.com/in/gokul-subramanian-068381191>

Data Engineer with over 4 years of experience designing, building and maintaining reliable data platforms supporting batch and streaming pipelines to enable insight generation and AI initiatives. Proven ability to collaborate with cross-functional teams to transform abstract concepts into fully functional, data driven solutions that deliver measurable business value.

SKILLS:

Tools: Databricks, Snowflake, BigQuery, Fivetran, Terraform, AWS Glue, Azure Data Factory, Apache Spark, Airflow, Hadoop, Hive, Impala, PostgreSQL

Programming Languages: Python, PySpark, Scala, SQL, Unix Shell Scripting

EXPERIENCE:

Deriv | Data Engineer | Dubai, UAE

Apr 2025 – Present

- Developed **streaming pipelines** to ingest data from a **Redis Queue** to **BigQuery** using Python, streaming millions of trading events daily
- Led the refactoring and migration of several event-driven pipelines from **GCP Cloud Functions Gen1 to Gen2** by conducting impact analysis and testing (TDD) to ensure seamless migration
- Collaborated with cross-functional teams to enable **AI/ML** initiatives by building curated datasets and low latency REST APIs to provide real time data for client facing applications and AI agents
- Deployed several containerized Python ETL jobs (**Docker & GCP Artifact Registry**) to build a scalable batch data ingestion pipeline to BigQuery, orchestrated with Airflow
- Co-designed a high-availability architecture for production **PostgreSQL** databases, implementing asynchronous **logical replication** using WAL files to support real-time data services with zero impact on primary database performance
- Automated infrastructure management with **Terraform** reducing manual cloud configuration efforts and enforcing platform standards

Solution BI | Data Engineer | Dubai, UAE

Sep 2022 – Mar 2025

- Developed a metadata-driven ETL framework to migrate over 30,000 tables from multiple source systems to **Snowflake** using **PySpark** on **AWS Glue**, eliminating manual effort and achieving 65% increase in data processing speed
- Applied data consolidation strategies using dynamic SQL scripts on Snowflake to automate data standardization, which eliminated manual effort and increased data quality
- Designed an automated testing process to identify and flag data anomalies after migration which led to 100% data accuracy
- Migrated 100 TB+ data from a legacy SQL Server Parallel Data Warehouse (PDW) to **HDFS** using **Apache Sqoop** and built a data warehouse on **Apache Hive**
- Unified siloed data systems by implementing a data lakehouse with medallion architecture using **Azure Data Factory** and **Databricks** to deliver advanced analytics solutions
- Improved Spark job performance on Databricks for large datasets (10 TB+) by salting the key column to improve parallelism and reducing job execution time by 30%
- Automated a scorecard reporting system to provide an actionable list of KPIs to monitor business growth and operations for the C-Suite

Brighter Prep | Data Analyst | Dubai, UAE

Feb 2021 – Sep 2022

- Developed **Python** script to web scrape information about universities and the courses provided eliminating manual data entry and saving time

- Supported ad-hoc requirements by creating dashboards on excel to visualize popular courses, top universities based on ranking and country

CERTIFICATIONS:

[Azure Data Engineer Associate - 2024](#), [Azure Administrator Associate -2024](#), [Azure Solutions Architect Expert - 2024](#), [Databricks Certified Data Engineer Associate - 2024](#), [IBM Data Analyst Professional Certificate \(V2\) - 2022](#)

EDUCATION:

Bachelor of Engineering (B.E.) – Aeronautical Engineering
Madras Institute of Technology - Chennai, India

Aug 2016 - Apr 2020