

Sourab Saklecha

(408) 581-4470 | saklechasourab@gmail.com | linkedin.com/in/sourabsaklecha | ssaklecha.github.io

TECHNICAL SKILLS

SQL | Python | Scala | PySpark | Snowflake | PostgreSQL | AWS EC2 | AWS S3 | AWS RDS | AWS Glue | AWS Redshift | AWS Lambda | Databricks | DBT | GCP | BigQuery | Azure Data Factory | Azure Data Lake Storage | Azure Synapse Analytics | Microsoft Power BI | Tableau | Apache Spark | Apache Airflow | Apache Kafka | Apache Iceberg | Git | API Design

EXPERIENCE

WinWinLabs

Plymouth, MI

Data Engineer

Feb 2025 – Present

- Designed and deployed a GCP BigQuery-based customer analytics warehouse in just 8 weeks, enabling real-time user behavior analytics across product engagement and conversion funnels.
- Modeled multi-source event data from GA4, Tag Manager, and custom trackers into a star schema in BigQuery, optimizing it with partitioning, clustering, and materialized views—boosting dashboard performance by 45%.

Core Molding Technologies Inc

Columbus, OH

Data Engineer

Jul 2022 – Jan 2023

- Led the design and implementation of an enterprise data warehouse on Azure Synapse using a snowflake schema, integrating ERP, on-prem, and ADLS data—centralizing analytics across sales, supply chain, and finance teams.
- Developed 15+ modular ETL pipelines in Azure Data Factory (ADF) and dbt-based transformations—accelerating delivery of high-trust data to business teams and reducing ad-hoc data escalations by 41%.
- Reduced storage costs by 16% using columnstore indexing, type optimization, and partitioning.
- Improved Power BI dashboard load times by 35% through performance tuning and optimized query logic.
- Migrated legacy reporting infra from Cognos to Power BI, reducing licensing and maintenance costs by ~ 30%.

Santander Bank

Boston, MA

Data Engineer

Oct 2020 – Apr 2022

- Migrated 10+ TB of operational data from on-prem MS SQL Servers to Azure using 40+ ETL pipelines on Databricks, validating downstream integrations to ensure compatibility across reports, apps, and services.
- Designed and deployed CI/CD pipelines via Azure DevOps to automate ADF and Databricks notebook deployments across environments, reducing release time by 40% and improving deployment consistency.
- Achieved 99.9%+ post-migration data accuracy using automated hash checks in Python.
- Leveraged Azure Monitor for alerts and Power BI to deliver near real-time dashboards on pipeline health and migration KPIs, enabling early bottleneck detection and improving SLA compliance.

Game On Fitness

Bangalore, India

Data Engineer

Oct 2019 – Sep 2020

- Migrated on-premises BI system to AWS—enabling horizontal scaling of data pipelines during peak traffic periods.
- Reduced batch processing time by 35% by implementing dynamic resource allocation.
- Built scalable ETL/ELT pipelines using AWS Glue and PySpark to move data from multiple sources to Redshift.
- Engineered automated data validation checks across the DE lifecycle using AWS Lambda and dbt to detect schema mismatches—saving 9 man-hours of daily manual effort for 6 teams.
- Led cross-team efforts to implement RBAC and row-level security in AWS Redshift and Power BI—cutting downstream unauthorized access incidents by ~ 75% and aligning with audit controls.

PROJECTS

Real-Time Distributed Data Processing using Spark, Kafka, and Airflow

Stock Market Data Analytics

- A real-time stock market analytics pipeline using Kafka, Spark, and Airflow to ingest data via external market data APIs—enabling dynamic price analytics and real-time dashboards for monitoring trends and volatility.

EDUCATION

San Jose State University

San Jose, CA

MS Data Analytics

Dec 2024