Sourab Saklecha

saklechasourab@gmail.com | linkedin.com/in/sourabsaklecha | ssaklecha.github.io

TECHNICAL SKILLS

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

EXPERIENCE

WinWinLabs   **Plymouth, MI**
Data Engineer   **Feb 2025 – Present**
– Implemented end-to-end BigQuery data warehouse architecture for GA4 analytics within 8 weeks, establishing automated daily ingestion of user events and reducing reporting latency from 24 hours to near real-time.
– Reduced warehouse query costs by 20% and improved dashboard performance by implementing partitioned and clustered GA4 tables, and automating daily rollups for high-traffic analytics workloads.

Core Molding Technologies Inc   **Columbus, OH**
Data Engineer   **Jul 2022 – Jan 2023**
– Spearheaded the implementation of an enterprise data warehouse using a snowflake schema in Azure Synapse, integrating data sources on on-premises, ADLS, ADB, REST APIs, and ERP systems using Azure Data Factory.
– Built and orchestrated 15+ batch pipelines in ADF, using dbt for transformations, staging-to-prod workflows, and automated testing—enhancing pipeline reliability and deployment efficiency.
– Reduced storage costs by 20% with use of columnstore indexing, type optimization, and partitioning. Improved Power BI dashboard load times by 35% (9+ secs) by tuning query plans and performance tuning.
– Migrated legacy reports from Cognos to Power BI, enabling self-service analytics across Sales and Finance.

JoulesToWatts   **Bangalore, India**
Data Engineer   **Apr 2022 – Jul 2022**
– Migrated on-premises BI system to AWS. Built scalable ETL pipelines using AWS Glue and PySpark for HR data from S3 to AWS Redshift OLAP models—resulting in 40% faster dashboard load times.
– Engineered automated data validation across the DE lifecycle checks using AWS Lambda and dbt to detect schema mismatches— saving 9 hours of daily manual effort for 6 teams.
– Led cross-team efforts to implement Role Based Access Controls at the data/reporting level using AWS IAM roles and policies—cutting downstream unauthorized access incidents by 75% and aligning with audit controls.

Santander Bank   **Boston, MA**
Data Engineer   **Oct 2020 – Apr 2022**
– Migrated 10+ TB of operational data from on-prem SQL Servers to Azure using 40+ parameterized ETL pipelines (Azure Data Factory + Databricks), achieving zero downtime during business.
– 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.
– Developed automated validation workflows on ADF using Python with hash checks and sampling logic, achieving 99.9%+ data accuracy post-migration and streamlining compliance checks for internal audits.
– 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.

Infosys   **Mysore, India**
Data Engineer Intern   **Sep 2019 – Mar 2020**
– Cleaned and transformed marketing data from GA, Mailchimp, and Google Ads using Power Query; built automated Power BI dashboards and 20+ ad-hoc reports with DAX and Pivot Tables—cutting prep time by 80% and driving campaign strategy through insights on CTRs and conversions.

EDUCATION

San Jose State University   **San Jose, CA**
MS Data Analytics   **Dec 2024**