

# Sourab Saklecha

San Jose, CA (Open to Relocation) | +1 408 581 4470 | saklechasourab@gmail.com | LinkedIn: [/sourabsaklecha](#) | [GitHub](#) | [Data Portfolio](#)

## WORK EXPERIENCE

### Data Engineer | WinWin Labs– Plymouth, MI

Feb 2025-Present

- **Data Warehousing:** Building and managing data pipelines to integrate client databases into central data warehouse.
- **SQL Querying:** Ran ad-hoc SQL queries on BigQuery to analyze Google Analytics data.

### Data Engineer | Core Molding Technologies Inc – Columbus, OH, USA (Remote)

Jul 2022-Jan 2023

- **Data Warehousing:** Engineered an end-to-end data warehouse with a snowflake schema on SQL Server utilizing SSIS and SSMS to integrate data from 7 different sources.
- **Data Transformations:** Developed 15+ T-SQL scripts to perform advanced data transformations.
- **Data Optimization:** Reduced warehouse size by 20% using normalized data models and data type optimization.
- **Query Performance Enhancement:** Improved query performance by ~30% using partitioning and indexing strategies.
- **ETL Automation:** Automated ETL jobs using SQL Server Agent to load over 1 GB of data daily into the warehouse.
- **Job Failure Reduction:** Improved ETL job reliability by reducing the job failure rate by 95%, ensuring consistent data processing.

### Data Engineer | JoulesToWatts Pvt Ltd – Bangalore, India

Apr 2022-Jul 2022

- **ETL Pipeline for OLAP:** Migrated on-premises business intelligence system to AWS. Designed ETL pipelines using AWS Glue to transform and ingest S3 data into AWS Redshift OLAP, achieving a 40% reduction in analytics dashboard generation time.
- **Data Validation Automation:** Automated validation scripts (AWS Lambda) to detect discrepancies, reducing data cleaning time by 16%.
- **Data Governance:** Established frameworks, including role-based access control (RBAC) at the data and reporting level using AWS IAM group roles and policies, to reduce unauthorized data access incidents by ~75%.
- **Cross-functional Collaboration:** Coordinated efforts across six teams to optimize data consumption and reporting workflows.

### Data Engineer | Infosys Limited – Bangalore, India

Oct 2020-Apr 2022

- **Data Migration Optimization:** Optimized ETL pipelines using Azure Data Factory and Azure Databricks notebooks to migrate over 10 TB of on-premises data to Azure, ensuring zero critical operations downtime.
- **CI/CD Workflows:** Implemented incremental build processes using Azure DevOps CI/CD pipelines, ensuring seamless deployment.
- **Post-migration Data Validation:** Collaborated on validation post-migration using Azure Data Factory, ensuring 99.9% data accuracy.
- **Performance Monitoring:** Leveraged Azure Monitor and Log Analytics to track migration performance and ensure system stability.
- **Power BI Dashboard Development:** Leveraged DAX measures to build dashboards and visualize migration progress and KPIs.
- **Stakeholder Reporting:** Delivered near real-time reports with 2-hour refresh rate, helping in detecting bottlenecks early.

### Data Analyst Intern | Infosys Limited – Mysore, India

Sep 2019-Mar 2020

- **Data Collection and Cleaning:** Collected, cleaned, and validated over 100K records from flat files (JSON, CSVs) using Power Query.
- **SQL Querying:** Worked on SQL query tasks that included window functions, subqueries, CTEs, stored procedures, and views.
- **Ad-hoc Reporting:** Used advanced Microsoft Excel functions such as VLOOKUPS and Pivot Tables to deliver reports.

## PROJECTS

### Real-Time Data Streaming using Apache Kafka – Data Engineering (Apache Kafka, Python, AWS) ([Git](#))

Aug 2023-Dec 2023

- Built a real-time pipeline that ingests high volumes of stock market data (handling 100,000+ records per minute) using Kafka.
- Employed AWS S3 for scalable storage, Glue for automated schema inference, and Athena for interactive SQL-based analytics.

### Urban Taxi Analytics – Big Data Analytics (GCP, BigQuery, Microsoft PowerBI) ([Git](#))

Jan 2024-May 2024

- Developed an ETL pipeline to process 84.5M records from NYC taxi rides using GCP (Storage), Mage (ETL), and BigQuery (Warehousing).
- Achieved high throughput (200k+ records per min) through efficient orchestration and scalable architecture for Parquet files.

## TECHNICAL SKILLS

- **Programming Languages:** SQL (T-SQL), Python, Scala.
- **Databases/Storage:** MySQL, PostgreSQL, MongoDB, Cassandra.
- **Big Data Processing/ETL:** Apache (Spark, Airflow, Kafka), Hadoop (HDFS, Hive, MapReduce), Power Query, SSIS, PySpark.
- **Data Warehousing:** AWS Redshift, Google BigQuery, Snowflake.
- **Cloud Platforms:** AWS (S3, Glue, RDS, Athena, Lambda), Microsoft Azure (Blob, Data Factory, Databricks, Synapse Analytics, Fabric).

## LICENSES, CERTIFICATIONS AND AWARDS

- Microsoft Certified: Power BI Data Analyst ([PL-300](#)) and Azure Data Fundamentals ([AZ-900](#)) among [others](#).
- AWS Certified DevOps Engineer – Professional.
- NPTEL Certified: Programming, Data Structures and Algorithms using Python (*From IIT, Madras*).

## EDUCATION

- **M.S., Data Analytics**, San Jose State University, San Jose, CA, USA
- **B.E., Computer Science**, Visvesvaraya Technological University, India

Jan 2023-Dec 2024

Aug 2016-Sep 2020