# SHOURYA PARANJAPE

shauryaparanjape@gmail.com | 647-717-5041 | Ontario, Canada

## CORE COMPETENCIES:

Python, Google Cloud, Tableau, SQL, Snowflake, GCP BigQuery, Airflow, AWS, EMR, Lambda, Redshift, S3, RDS, Git, Docker, DBT, Spark, FiveTran, Github, Gitlab, JIRA, Looker

## RELEVANT EXPERIENCE:

**Senior Data Engineer | Evercommerce**                                            Sep 2023 –
- Used data mesh architecture to build a reporting warehouse framework in DBT on top of Azure databricks. Created raw, staging and mart layers as per best practices. Utilized Jinja to reduce SQL code redundancy and macros for reusability. Created pre-hooks and post hooks for procedural programming. Utilized DBT utils package and DBT Audit helper package for enhanced code modularity and testing. Incorporated CI principles in the form of unit tests and custom tests to reduce data quality errors in production environment. This increased the efficiency of BI team to build reports by reducing development times by 25%.
- Created a data ingestion pipeline using Fivetran to load data from Postgres database into Azure Synapse. This helped increase efficiency by removing the legacy file ingestion process and reducing operational downtime by 30%.
- Used performance tuning technique of creating table partitions and clustering to reduce table scan times.
- Participated in code reviews and provided constructive feedback to colleagues, fostering a culture of collaboration and continuous improvement within the data modeling team.
- Collaborated with the data governance team to enforce data quality standards and best practices, validating data accuracy and consistency across all DBT models.
- Mentored junior members of the data modeling team, fostering their skills data modeling best practices, leading to enhanced team productivity and cohesion.

**Senior Data Engineer Architect | Infostrux**                              Feb 2023 – July 2023
- Built a data pipeline in Snowflake to ingest IoT data based on event notifications to create and analyze reporting on failure metrics for IoT devices.
- Collaborated and executed an ELT framework which utilized Snowflake warehouse structure through DBT. This framework utilized ingest, clean, normalize & analyze layers through the use of DBT models and macros.
- Developed and deployed stored procedures in Snowflake to automate testing processes and reduce testing efforts by 50%. Stored procedures in Snowflake were triggered every time a development model ran. As a result code quality improved and time to deploy code to production environment reduced drastically.

**Senior Analyst | Home Depot**                                            Oct 2021 – Dec 2022
- Built a data pipeline for anomaly detection using Python. This resulted in more actionable reports on data failures and improved data integrity standards. Used Git & Github for version control and pull requests through CI/CD development. Used Docker containers for local code testing and deployment.
- Developed ETL pipeline using SQL, Airflow, Python - to run reports on online clickstream big data to understand the effects of sample purchases from Google Cloud Big Query database service.
- Built data warehouse by means of data modeling techniques in GCP Big Query to collect post purchase data to analyze customer pain points when an order is returned or canceled. Successfully conducted performance tuning on queries for Google Cloud tables for performance enhancements and faster query retrieval. This reduced query costs by 60%
- Created a star schema data model in Big Query to model customer behaviors post purchase. This model reduced query costs by 35%

**Data Analyst | Libreum International**                                     Jan 2019 – April 2021

- Developed data pipeline to deploy Reinforcement-learning ML model into production which adapted marketing outreach communications in response to customer engagement metrics. The machine learning algorithm optimized for customer click-through rate. Pipeline was developed in Python, Airflow and Snowflake data warehouse.
- External vendor data sharing: Setup Python Spark in AWS EMR for data extraction from S3 files containing customer loan payment data. Data was then random sampled and transferred securely over SFTP protocol to share PII customer-level data with external vendors. Used Lambda functions, Airflow to trigger batch process jobs based on vendor criteria.
- Deployed a data engineering ETL pipeline to convert audio files in S3 storage to transcribed text using AWS cloud speech to text service using Python & AWS Lambda function. The transcribed text was then updated on AWS Redshift data warehouse. This reduced customer issue resolution time by 25%.
- Business Intelligence Reporting: Was responsible for developing the pipeline of integration between Tableau extract data sources and AWS Redshift data warehouse. As a result, the stakeholders from compliance and capital markets teams were able to focus on KPIs and led to 35% increase in raising capital for deployment. Data analysis from dashboards created for underwriting KPI measurement led straight to a 3% decrease in first payment default rates.
- Responsible for Tableau cloud administration in order to serve reports to stakeholders from different areas of business including finance, operations, marketing, compliance & analytics.
- Changed SQL schema to create partitions, indexes on existing tables for better query performance.
- Created federated queries in AWS Redshift to stitch together customer payment information and customer bank account transaction data to be used for analysis of first payment defaults in various lending portfolios

**Analyst | Accenture**                                                     Aug 2015 - Jul 2017

- ETL pipeline development. Quickly gained an understanding and platform development goals for ETL pipeline. Rapidly prototyped a solution using Informatica, Alteryx which met client requirements for ingesting large amounts of product retail data. Produced automation tools that cut the time required for testing constraints by 40%.
- Rapidly prototyped and deployed a data engineering solution in PySpark SQL for ingesting big data related to credit cards and transactions. Data from main-frames was ingested into AWS RDS using business logic.

**EDUCATION:**

Georgia State University,                                                    Aug 2017 - Dec 2018
 Master of Science, Major: Analytics, 4.0 GPA

Savitribai Phule Pune University                                            Aug 2011 - May 2015
 Bachelor of Engineering, Major: Electronics, 3.75 GPA

**CERTIFICATIONS:**

   Snowflake SnowPro Core Certification
   Data Engineering on Google Cloud Platform Specialization. (Coursera)