Siddharth Goradia

+1 6045180459 | siddharth.rajesh.goradia@gmail.com | LinkedIn

TECHNICAL SKILLS

- **Programming & Big Data Tools**: Python, SQL, PySpark, Apache Spark, Azure Databricks, Apache Airflow, Pandas, NumPy, Docker, Debezium, Apache Kafka, PostgreSQL, Scala, AirOps, Hadoop
- **Cloud & Data Platforms**: AWS (S3, EMR, Glue, Athena , Redshift), Azure (ADLS , ADF , Data Factory, Synapse), Snowflake, GCP (BigQuery, Dataflow), Hudi
- **Data Modeling & Visualization**: Tableau, Power BI, Looker

WORK EXPERIENCE

ZS Associates — **Toronto, Ontario**
Associate Data - Engineer — *Aug 2024 – December 2024*

- Developed and executed data ingestion scripts on Databricks by writing optimized SQL queries to transform financial data, ensuring data consistency and accuracy, which contributed to faster, more accurate financial reporting.
- Created and deployed scalable, fault-tolerant data pipelines using Azure Data Factory to automate the processing of financial datasets, improving data flow efficiency, which streamlined data ingestion and resulted in a 25% reduction in processing time.
- Implemented data validation checks within pipelines by embedding validation logic to detect and handle anomalies, ensuring data integrity and catching potential errors, which reduced data inconsistencies by 20% and quality of financial insights.

RajDeep Steel — **Mumbai, Maharashtra (Remote)**
Data Engineer — *Jan 2024 – July 2024*

- Designed and implemented ETL pipelines using Python and Apache Spark, enabling real-time data processing of sales trends and production schedules, reducing inventory holding costs by 15%.
- Built a data warehouse on AWS Redshift, integrating transactional data from ERP systems to generate actionable insights, improving procurement efficiency by 20%.
- Developed scalable data models with SQL and Tableau dashboards, enabling dynamic pricing strategies based on market fluctuations, resulting in a 10% increase in revenue margins.

SAP Inc. — **Vancouver, British Columbia**
HANA Support Intern — *May 2023 – December 2023*

- Streamlined ticket scheduling system through analysis and optimization of software performance and automation using Python and Tableau, improving case assignment efficiency by 20%.
- Spearheaded comprehensive analysis of sales data using SQL and Python in order to enhance marketing efficiency for client companies, resulting in segmentation based on customer purchase history, demographic insights and a 10% increase in sales.
- Developed an incident response model, helping with incident triage based on severity, potential impact, and relevance through employment of HANA in analyzing key attributes, historical data achieve an improvement in overall response time.

ZS Associates — **Pune, Maharashtra**
Business Technology Solutions Associate — *July 2021 – July 2022*

- Designed and maintained data pipelines and ETL processes using Databricks and Airflow that processed 5TB of pharmaceutical data daily, resulting in a 25% reduction in data processing time.
- Developed a medical text recommendation engine using unsupervised machine learning, saving 20+ work hours monthly by selectively compiling and recommending extracts based on key phrases from historical study documents and reports.
- Drove exploratory analysis to identify key gaps in sales and supply chain, employing SQL and PySpark to map correlations between product trends and location demographics resulting in cost-saving opportunities and a 10% reduction in expenses.

Publicis Sapient — **Gurgaon, Haryana**
Engineering Intern — *January 2021 – June 2021*

- Conducted complex data analysis into customer phone plan renewal turnover using SQL queries and Excel macros, identifying key trends that resulted in a 25% reduction in churn and a 15% increase in revenue.
- Developed and implemented a time series forecasting model using Python and Apache Spark to identify usage surges and accurately allocate resources, leading to a 30% increase in customer acquisition and a 20% improvement in retention.
- Collaborated with cross-functional teams to streamline data collection and reporting processes, resulting in a 30% reduction in data processing time and an improvement in data consistency.

PROJECT EXPERIENCE

Simon Fraser University — **Burnaby, British Columbia**
Real-Time Data Pipeline with Change Data Capture (CDC) — *January 2024 – April 2024*

- Designed and implemented a real-time data pipeline leveraging **Change Data Capture (CDC)** techniques with Debezium, Kafka, PostgreSQL, and Docker, enabling efficient and scalable data synchronization, real-time analytics, and event-driven architectures.

EDUCATION

Simon Fraser University — **Burnaby, British Columbia**
Master of Science in Professional Computer Science — *Graduation Date: 2024*

Vellore Institute of Technology — **Vellore, Tamil Nadu**
Bachelor of Technology Computer Science and Engineering — *Graduation Date: 2021*