SACHIN KHAJURIA

Sr. Analyst - Data Engineering

sachinkace@gmail.com | +917889848384 | Banglore, karnataka | github.com/Sachinkayy

SUMMARY

Senior Data Engineer with 4+ years of experience building and optimizing large-scale data pipelines across AWS, Azure, and Snowflake. Proficient in SQL, Spark/PySpark, Kafka, DBT, Airflow, and modern ELT tooling (Fivetran, HVR, Matillion), with hands-on focus on data modeling, data quality, and data governance. Experienced across Marketing Analytics, Data Migration, and Healthcare domains, with a track record of building CI/CD-driven, production-grade pipelines that uphold data accuracy, integrity, and accessibility.

PROFESSIONAL EXPERIENCE

LatentView Analytics, *Sr. Analyst - Data engineering - Present* 05/2024 | Bangalore, Karnataka

Tools Used: DBT, Snowflake, AWS, Fivetran, Airflow, Python, Fivetran

- Designed and orchestrated scalable, production-grade data pipelines using **Airflow, Matillion, and Fivetran,** enabling reliable ingestion and processing of high-volume data across multiple sources and destinations.
- Developed and optimized complex SQL transformations on large-scale datasets, improving **query performance**, **reducing execution time**, and ensuring efficient data processing for downstream analytics.
- Led the migration of **60+** legacy Matillion ETL workflows to a modern **DBT + Astronomer (Airflow) architecture**, enhancing pipeline modularity, maintainability, and scalability while significantly reducing operational overhead.
- Implemented robust **pipeline monitoring, alerting, and failure-handling mechanisms**, proactively identifying issues and ensuring high data reliability, availability, and SLA adherence.
- Built and maintained curated, analytics-ready datasets to support **marketing use cases, enabling self-service analytics and powering dashboards** in **Sigma** and other **BI** tools for business stakeholders.

Futurense Technologies, *Data Engineer* 04/2023 – 05/2024 | Banglore, Karnataka

Tools: SQL, Python, Power BI, Pandas, Snowflake, PySpark, Spark-SQL, Azure Data Factory, Azure Synapse, Azure Databricks, FiveTran

- Used the existing logics from **HMCL** and migrated all the pre-developed reports on the **Oracle OBIEE** in **PowerBI** after the successful migration and validation of the data using **SnowSql.**
- Worked in the migration of **Hero Motto Corp Limited's** data and their **Informatica jobs** with their data existing on **Oracle DB** to **Snowflake** using **Fivetran.**
- Led a comprehensive Uber Data Analytics project, harnessing tools like **MySQL**, **Hadoop**, **Hive**, **Sqoop**, **PySpark**, and a suite of **cloud services** from **Azure** storage to **Snowflake**. Designed an end-to-end data pipeline with **Azure DataFactory**, **SnowPipe**, and Spark connector, ensuring efficient data processing and prompt solution deployment.
- Through deep analysis with **Hadoop** and **Spark**, unveiled critical demand patterns and driver behaviors, subsequently presenting these actionable insights via **PowerBI** dashboards leading to not only steered enhanced profitability strategies but also promoted informed decision-making and a culture of continuous learning.

Tech Mahindra, *Associate Software Engineer* 03/2022 – 04/2023 | Pune, Maharashtra

- Managed and optimized **SQL Server** and **SAP ABAP** databases, enhancing **query performance** by 20% and reducing storage requirements by **15%**. Developed efficient database solutions using **open SQL** for seamless data retrieval and manipulation.
- Led and participated in **Java** and **Python programming** initiatives, showcasing strong **problem-solving skills**. Completed challenging coursework and personal projects, **developing 3 functional programs**.

TECHNOLOGY SKILL SETS

ETL
- Fivetran/HVR, Matillion
- AWS Glue
- Astronomer Airflow
- Kafka, Docker
- Jenkins, CI/CD Pipelines

Core Competencies
- Data Quality, Data Modeling, Data Governance
- Root Cause Analysis
- REST/SOAP API

Data Analysis & Visualization
- Numpy, Pandas
- MS Excel
- PowerBi, Tableau, Sigma Reporting

Tools and others
- VsCode, PyCharm
- JupyterNotebook
- JIRA/Confluence
- LLMs, Copilot, Coco Desktop, Claude Cowork

SQL/Databases
- SQL, PL/SQL, No SQL
- MySQL Workbench
- Sql Server
- PostgreSQL

Cloud & Tools
- Snowflake, RBAC
- Azure Data Factory, Azure Data bricks, Azure SQL, Azure Synapse, Blob Storage, ADLS
- Glue, Redshift, S3, EC2, RDS, Lambda, EMR

Big Data Ecosystem
- PySpark, DSL, SparkSQL, Spark Streaming, Databricks, Delta Lake, Delta tables
- Hadoop, Hive, Hbase, HDFS, Data Lake

VCS
- Git, Github

EDUCATION

B.Tech. - ECE, *SRM Institute Of Science & Technology, Delhi Ncr* 06/2017 – 05/2021

Higher Secondary (XII) - PCM, *Air Force School, Jammu J&K* 03/2016 – 05/2017

CERTIFICATION

- SQL FOR BEGINNERS - UDEMY (2022-05)
- Core Java Made Easy - UDEMY (2022-05)
- Excel Skills for Business - COURSERA (2020-06)
- ChatGPT Prompt Engineering for Developers - Futurense (2023-07)
- Introduction to Git and GitHub - COURSERA (2021-02)
- Python 3 Specialization - Coursera (2021-07)
- Agile Fundamentals - UDEMY (2022-05)
- Learn Linux - UDEMY (2022-05)
- Data Bricks Lake-house Fundamentals - Academy Accreditation (2023-08)

PROJECTS

Marketing Analytics and Research System (MARS) - Autodesk, *LatentView, Bangalore* 06/2024 – Present

- Architected scalable marketing analytics solutions, enabling data-driven insights across campaigns and customer behavior.
- Engineered and orchestrated robust data pipelines using Apache Airflow (Astronomer) and Matillion for efficient data ingestion and transformation.
- Designed high-performance data models and implemented complex SQL transformations in Snowflake to support large-scale analytics workloads.
- Modernized legacy workflows and strengthened pipeline reliability through monitoring, optimization, and seamless integration with Sigma for business reporting.

Hero Moto Corp Ltd. Migration, *BluePi, Gurgaon* 06/2023 – 03/2024

- Developed a wrapper function for the team to convert the Informatica expressions to their equivalent Snowflake queries, singly handled 40+ Informatica functions resulting into a 95+% accuracy in the conversion.
- Worked with the team in the onboarding process of the migration of the Informatica jobs into the snowflake by parsing the XML data generated from Informatica and using them to make working Snowflake procedures with an acceptance rate of 99+%.
- Spearheaded the Validation process of the onboarded data by developing a python script which checked the tables on 6+ different metrics and finally generated accuracy table for the particular tables.

Healthcare Analysis Using SQL, *Futurense, Banglore* 03/2023 – 05/2023

- Analyzed healthcare sector, addressing data from multiple sectors. Processed and validated 10,000+ records from medicine industry, hospitals, insurance, pharmacy etc
- Utilized MySQL, Python's Pandas, NumPy, and Seaborn libraries for advanced data cleaning, merging, and visualization. Ensured data accuracy and consistency across 4+ different sectors.
- Conducted sector-wise analysis to evaluate industry conditions. Identified areas for improvement and provided actionable insights based on data analysis. Solved 50+ current market problems.

PROGRAMMING LANGUAGES

- SQL
- Python
- JAVA (CORE)

INTERESTS

- Sketching
- Watching Movies
- Sports and Athletics