

Rishabh Jain

Data Engineer | Data Analytics | Cloud Solutions

SUMMARY

With 7+ years of experience designing and implementing scalable data pipelines using Python, SQL, and AWS tools, I bring extensive expertise in optimizing ETL processes, following best practices and improving data quality for large-scale datasets using functional data engineering practices.

WORK EXPERIENCE

WineDirect, Vancouver, Canada

Data Engineer 06/2023 – Present

- Designed and implemented idempotent ETL pipelines using Glue and PySpark, reducing data processing time and improving data workflow efficiency.
- Introduced and productionized DBT within the data platform, improving data quality and governance using various custom and singular tests.
- Optimized Snowflake performance by analyzing query profiles, fine-tuning warehouse configuration, built-in optimization tools, reducing costs by 25%.
- Extensively worked on data modeling, leveraging Slowly Changing Dimensions and daily partitioned snapshots to optimize historical data tracking.
- Architected a centralized Data Lake to store and manage Petabytes of historical data, enabling seamless access for clients through a client-facing API.
- Streamlined database schema migrations using Flyway, ensuring version control and minimizing downtime during deployments.

Skillz, Vancouver, Canada

Data Engineer 06/2021 – 05/2023

- Worked on complex SQL queries, contributed to database schema design, and performed query optimization and tuning.
- Migrated data from Redshift to Snowflake while maintaining data integrity and consistency while tackling schema evolution and enforcement.
- Spearheaded the integration of Snowflake cloud data warehouse which enhanced data analytics and reporting capabilities for cross-functional teams.
- Implemented DBT custom tests, snapshots, incremental models and macros using DBT Core.

Samsung Electronics, Vancouver, Canada

Big Data Engineer Intern 05/2020 – 12/2020

- Migrated Spark jobs from AWS EMR to EKS, leveraging Spot instances to reduce infrastructure costs by 15% while processing Petabytes of data efficiently.
- Developed data pipelines using Apache Airflow and AWS, orchestrating ETL batches that executed SQL queries and Python code.
- Provisioned infrastructure using Terraform to create AWS ASG for running Spark Scala code on Kubernetes.
- Implemented data quality checks, validation rules, and governance processes to ensure the accuracy and reliability of IOT based events data.

Tesco, Bengaluru, India

Software Engineer, Data 08/2015 – 08/2019

- Experience in IBM Datastage 11.5. Effectively processed huge data sets by integrating different sources like Parquet, CSV, Json. Successfully processed structured, semi-structured data sets to identify & assess key takeaways.
- Deployed solution to handle schema evolution and scalability challenges.
- Traveled to the UK for one year to understand the business requirements and build a robust application for sending personalised offers across different banking products like Savings, Mortgages, Loans etc.

CONTACT

- Vancouver, Canada
- +1-236-866-6315
- rishabhjain2017@gmail.com
- [Rishabh's Portfolio](#)
- [Linkedin](#)

SKILLS

Data Engineering & ETL:

- DBT
- Airflow
- Glue
- Step Functions
- AWS DMS
- Terraform
- Data Warehousing

Big Data & Cloud:

- Pyspark
- Snowflake
- Redshift
- EMR / EKS
- Kubernetes

Programming Languages:

- Python
- SQL
- Scala

Version Control & CI/CD:

- Github
- Gitlab
- Circle CI
- Github Actions

EDUCATION

Master of Science in Big Data
Simon Fraser University
Burnaby, Canada
09/2019 - 05/2021 | GPA - 4.0/4.3

Bachelor of Science in
Computer Science
Vellore Institute Of Technology
Vellore, India
05/2011 - 06/2015

CERTIFICATIONS

- AWS Certified Data Analytics Speciality (DAS C01)
- Certified Kubernetes Application Developer (CKAD)
- Tableau Desktop Specialist Certification