Rishabh Jain

Data Engineer | Data Analytics | Cloud Solutions

SUMMARY

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

WORK EXPERIENCE

WineDirect, Vancouver, Canada

Data Engineer   06/2023 – Present

- Designed and implemented ETL pipelines using AWS Glue and PySpark, reducing AWS data processing time by 25% and improving data workflow efficiency.
- Migrated data from Redshift to Snowflake while maintaining data integrity and consistency. Conducted comprehensive data mapping, data validation, and testing.
- Developed and deployed 20+ APIs using Ruby on Rails within an MVC architecture to enhance functionality and scalability for an e-commerce platform.
- Architected a centralized Data Lake to store and manage Terabytes 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.
- Introduced and productionized DBT (Data Build Tool) within the data platform, integrating it with Snowflake and orchestrating pipelines using Apache Airflow.
- Spearheaded the integration of Snowflake cloud data warehouse which enhanced data analytics and reporting capabilities for cross-functional teams.
- Optimized alert thresholds for data integrity metrics, contributing to a 30% reduction in unnecessary notifications, enabling the team to concentrate on genuine data integrity challenges and enhancing overall productivity.

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 complex SQL queries and Python code.
- Provisioned infrastructure using Terraform to create AWS ASG for running Spark Scala code on Kubernetes.
- Worked on Kubernetes, Docker, CI/CD, Distributed systems and Functional Data Engineering paradigm.

Tesco, Bengaluru, India

Software Engineer, Data   08/2015 – 08/2019

- Experience in IBM Datastage 8.1 and 11.5. Worked in GRID architecture.
- Effectively processed huge data sets by deploying advanced querying, visualization and analytics tools. Successfully processed structured, semi-structured and unstructured data sets to identify & assess key takeaways.
- 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
- https://rishabh.info/
- linkedin.com/in/risj/

SKILLS

Data Engineering & ETL:
- DBT
- Airflow
- Glue
- Step Functions
- AWS DMS
- Terraform
- Data Warehousing

Big Data & Cloud Technologies:
- Pyspark
- Snowflake
- Redshift
- EMR / EKS
- Kubernetes

Programming Languages:
- Python
- SQL
- Ruby
- 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 | GPA - 9.1/10

CERTIFICATIONS

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