

# Rishabh Jain

[www.rishabh.info](http://www.rishabh.info)



📍 Vancouver | 📞 236-866-6315 | ✉ [rishabhjain2017@gmail.com](mailto:rishabhjain2017@gmail.com) | [in linkedin.com/in/risj/](https://www.linkedin.com/in/risj/) | [github.com/rja40](https://github.com/rja40)

## SUMMARY

**4+ years** industry experienced as a **Data Engineer & ETL Developer**. Proven track record of driving user growth, boosting user retention & enhancing market share by deploying ETL components, techniques & best practices. Technical leadership skills for designing and developing Data Warehousing(DW)/Business Intelligence(BI) jobs, having **1 year onsite(UK)** experience. Played a key role in generating GBP 2 million in personalized marketing and achieved **Star Performer Award** by Tesco Bank for Inspiring Performance. Proficient in **Big Data** and **Cloud computing** to deliver compelling business value to clients & successfully execute projects.

## TECHNICAL SKILLS

Languages: **Python**, Scala, **SQL**, R, Shell Script

Big Data Tools: **Spark**, **AWS**, **Terraform**, **Docker**, **Kubernetes**, Hadoop, Mapreduce, Packer, Kafka

Database and ETL: **Airflow**, **Snowflake**, **DBT**, **Athena**, **Redshift**, Oracle 11g, MS-SQL Server, Postgres

Data Science Frameworks: Tensorflow, PyPy, **Scikit-learn**, **Pandas**, **Numpy**

CI/CD and Version Control Tool: Github, Gitlab, Github actions, Jenkins, SVN, Autosys

Visualization Tools: Tableau, Excel, Visio

## PROFESSIONAL EXPERIENCE

**Data Engineer | Skillz | Vancouver, BC**

Jun '21 – Present

- Building systems to provide real-time streaming analytics and event processing pipeline based on fast data architecture.
- Building enterprise-grade data warehouse to support both analytical needs and next-generation data infrastructure.
- Building data integration toolkit for backend services.
- Improve monitoring and alarms that impact data integrity replication lag.

Tech Stack – **Python**, **DBT**, **Streamsets**, **Snowflake**, **AWS**, **Terraform**, **Kubernetes**

**Big Data Engineer Intern | Samsung Electronics | Vancouver, BC**

May '20 – Dec '20

- Migration of AWS EMR to EKS to run Spark jobs using Spot instances to save cost.
- Created Airflow/AWS data pipelines to run ETL batches running 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.

Tech Stack – **Python**, **Scala**, **SQL**, **AWS**, **Terraform**, **Kubernetes**, **Docker**

**ETL Developer | ITC Infotech | Bangalore, IN**

Jan '19 – Aug '19

Client : Danske Bank

- Reporting - Boosted business via data manipulation and analysis to analyse customer habits & create user-friendly reports Identified and analysed the impact of marketing and product changes on customer behaviour, acquisition and retention.
- Data Processing & Management - Analyzed source systems to identify business rules for data profiling, data flows, data transformations & cleansing Directed the design, execution & monitoring of key metrics, reports, dashboards, etc., to drive business transformation. Enhanced & optimized query performance by deploying profiling tools and SQL

## Data Engineer | Tesco | Bangalore, IN

Aug '15 – Jan '19

ETL Developer/Designer

- Worked on multiple ETL projects and created end to end ETL pipelines.
- Experience in IBM Datastage 8.1 and 11.5. Worked in GRID architecture.
- Exposure in Batch scheduling process using Autosys. Deployed DS jobs using Jenkins.
- 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.

## Data Analyst | Tesco Bank | Edinburgh, UK

- 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.
- Worked on Requirement gathering, BSD, ETL Technical designs, Business analysis, Data modelling and build phase for different life cycle of Tesco Bank –Mortgages, Insurance, Offers, Loans, Savings, Current Accounts and Credit Card.

Tech Stack – Python, SQL, AWS, IBM Datastage

## PROJECTS

### Find Your Home ([Link](#))

Jan '21 – April '21

- Online platform that enables landlords and their tenants to gain a common consensus and enhance transparency. A Landlord would want to list and publicize his properties, manage his tenants and attend to service requests to ensure meeting his side of the deal.
- The services for the following entities are running on a Kubernetes cluster hosted on **Azure** and the database used is DynamoDB, a NoSQL DB running on AWS Cloud.

Tech Stack – Python, Azure, AWS, Rest API

### Strategic Asset Manager ([Link](#))

Jan '20 – Apr '20

- System that evaluates an investment decision taking into account the stock's historical performance, global news sentiment and company's Edgar reports.
- Used AWS Comprehend to find insights and relationships in text using machine learning. Created chatbot using AWS Lex. BERT was hosted on EC2 and files were stored on S3 which were used to answer user questions on Edgar reports. Lambda was the central co-ordinator between all components to work with Lex and deliver output.

## EDUCATION

### Master's in Computer Science, Big Data | Burnaby, CA

Sept '19 – April '21

Simon Fraser University

- CGPA: 4.0 / 4.3 Relevant courses taken: Machine Learning, Statistics, Big Data, Algorithms

### B.Tech - Computer Science | Vellore, IN

May '11 – Jun '15

Vellore Institute Of Technology

- CGPA: 8.27 / 10 Relevant courses taken: Data Mining, Database, Cloud computing

## CERTIFICATIONS

- Certified Kubernetes Application developer ([CKAD](#) LF-i19jtupyst)
- AWS Certified Data Analytics Speciality ([DAS C01](#))
- Tableau Desktop Specialist [Certification](#)