

RISHABH JOHRI

Bengaluru, India | +91-6266046936 | [rishabh2000johri2@gmail.com](mailto:rishabh2000johri2@gmail.com)

LinkedIn: [linkedin.com/in/rishabh-johri-156a39174](https://www.linkedin.com/in/rishabh-johri-156a39174) | GitHub: [github.com/rishabhjohri](https://github.com/rishabhjohri)

Target Roles: Data Engineer | AI Data Engineer | PySpark Developer | AWS/GCP Data Engineer | MLOps Engineer

## PROFESSIONAL SUMMARY

AI & Data Engineering professional with hands-on experience in building cloud-based data pipelines, ETL/ELT workflows, large-scale data processing, and analytics platforms using AWS, GCP, Databricks, Snowflake and PySpark. Strong foundation in distributed computing, data lake architecture, data modelling and ML pipelines. Skilled in designing end-to-end data workflows across ingestion, transformation, orchestration and deployment using tools like Apache Spark, Airflow, AWS Glue, BigQuery and SageMaker. Strong problem-solving mindset and proven ability to adapt quickly to complex data engineering requirements.

## DATA ENGINEERING TOOLS & PLATFORM EXPERIENCE

Cloud Platforms: AWS (S3, Glue, Lambda, IAM, Redshift), GCP (BigQuery, Cloud Storage, Dataflow, Pub/Sub)

Big Data & ETL: PySpark, Spark SQL, Hadoop, Databricks, Delta Lake, AWS Glue, Dataflow

Analytics Warehouse: Snowflake, Redshift, BigQuery

Pipeline Orchestration: Apache Airflow, AWS Step Functions

Data Integration: Kafka, AWS DMS, REST APIs, JSON

ML & AI Platforms: AWS SageMaker, Vertex AI, MLflow, TensorFlow, PyTorch

Databases: PostgreSQL, MySQL, MongoDB

DevOps: Docker, Git, Linux, CI/CD (GitHub Actions)

## PROFESSIONAL EXPERIENCE

Data Scientist Intern – Evoastra Ventures Inc. (Remote)

Jun 2024 – Aug 2024

- Designed Python-based data ingestion and transformation workflows for structured/unstructured data.
- Developed automated ETL scripts to process raw business data into analytics-ready datasets on AWS S3.
- Built model training pipelines using AWS SageMaker and optimized inference performance by 30% through vectorization.
- Deployed ML-based microservices using Docker and AWS Lambda, reducing manual reporting effort.
- Built KPI dashboards in Tableau and Power BI for operational decision support.

## Data Analyst Intern – IOTAGI (Vellore)

Jul 2021 – Aug 2021

- Performed large-scale data analysis using SQL in Google BigQuery on high-volume transaction datasets.
- Developed reusable Python scripts to clean and enrich raw datasets to improve data quality.
- Created automated insights dashboards in Power BI that improved reporting efficiency by 40%.
- Wrote optimized SQL queries for downstream data analytics and data warehouse usage.

## PROJECT DELIVERY EXPERIENCE

### Cloud Data Lake with PySpark & Redshift – AWS + Databricks

- Built a scalable data lake on AWS using S3 for raw, staging and curated data zones.
- Orchestrated PySpark-based ETL pipelines on Databricks using Delta Lake for ACID transactions.
- Designed dimensional data models for Redshift to support analytical reporting.
- Implemented incremental loading using Change Data Capture logic.

### Real-Time ETL Pipeline – GCP + Dataflow

- Designed event-driven ETL architecture using Pub/Sub and Dataflow.
- Applied windowed aggregations and schema management using Apache Beam pipelines.
- Loaded curated data into BigQuery for analytics and dashboarding.

### ML Pipeline Automation with SageMaker

- Designed a feature engineering pipeline using Pandas and Spark.
- Trained and deployed classification models on AWS SageMaker.
- Implemented batch inference jobs triggered using event-based workflows.

## KEY ACADEMIC PROJECTS

### Autoencoder-Based Text Encryption – Python, PyTorch, Flask

- Built a secure encoder-decoder model for encrypted text communication.
- Deployed model APIs and served endpoints using Flask and Docker.

Big Data Behavioural Analysis – PySpark, Hadoop, Tableau

- Performed correlation and association analytics using Spark.
- Processed large datasets using HDFS and Hive.

#### EDUCATION

M.Tech – Artificial Intelligence, IIT Jodhpur (2023–2025) | CGPA: 7.69

B.Tech – CSE with Data Science, VIT Vellore (2019–2023) | CGPA: 8.68

#### CERTIFICATIONS

AWS Cloud Certification | Google Cloud Certification | IBM Applied AI | Kaggle ML | HackerRank (Python, SQL)

#### POSITIONS OF RESPONSIBILITY

Teaching Assistant – Introduction to Machine Learning, IIT Jodhpur (2023–2024)

#### ACHIEVEMENTS

GATE CS Qualified | Runner-up – Data Unchained Hackathon (IIIT Delhi)

#### KEY Skills:

AWS, GCP, Databricks, Snowflake, PySpark, ETL, Airflow, BigQuery, Data Pipelines, Data Lake, Delta Lake, Cloud Data Engineering, Machine Learning Pipelines, SageMaker, Dataflow, SQL, Python, Kafka