+91 7678636923,9667082711
Rishabh16rawat@gmail.com
New Delhi
GitHub: https://github.com/Rishabh1627rawat
LinkedIn: https://www.linkedin.com/in/rishabh-rawat-6921b017a/

Rishabh Rawat

SUMMARY

Dedicated Aspiring Data Engineer with hands-on experience in developing scalable big data applications using Python, Apache Spark, Hive, and Power BI. Proficient in building and optimizing ETL pipelines, ensuring data quality, and leveraging tools like Airflow, Hadoop, and Kafka for seamless data integration. Skilled in data modeling, performance tuning, and creating interactive dashboards to support data-driven decision-making. Collaborative team player with strong problem-solving and communication skills. , I am eager to expand my expertise into the Experience with Apache Spark; currently exploring Databricks for cloud-based transformations." Proficient in AWS data tools (S3, EC2, EMR), with foundational understanding of Azure cloud services and Databricks (currently upskilling in ADF & Synapse)."

TECHNICAL SKILLS

Amazon Aws S3, EC2, EMR | Advanced SQL (PostgreSQL/SQL) | RDBMS Basics, Data Modeling (Star Schema, Data Vault)

Python (NumPy, Pandas), OOPS | Apache Spark | Big Data & ETL Pipelines, | Data Cleaning | Power BI | Data-Lake

Git/Jupiter | Snowflake | dbt (Test,Macros,Documentation) | Linux | kafka | Airflow | CI/CD pipelines

KEY SKILLS

Process Optimization | **Multi-tasking** | **Problem Solving** | **Interpersonal Communication**

Reporting & Documentation | **Issue Resolution & Troubleshooting**

EDUCATION

Amity University Haryana, MCA. Aug '23 - Jun '25

Amity University Haryana, BCA Aug '20 - Jun'23

PROJECTS

Big Data Pipeline for Global Retail POS Analytics [https://github.com/Rishabh1627rawat/retail-sales-pipeline]

- Designed and implemented a scalable data engineering pipeline to process daily **POS transactions from 4000+ global retail locations**.
- Utilized **Apache Spark and Hadoop HDP for distributed data processing** and storage, enabling efficient handling of high-volume data.
- Integrated product and vendor metadata to enable supply chain visibility
- Gathered **reporting and visibility** needs from business users and converted them **into Power BI dashboards** in the Retail POS project.

Stock Price & News Sentiment Correlation Analysis [https://github.com/Rishabh1627rawat/Data-Engineering-Project]

- Analyzed 5 tech stocks and 500+ news articles using Spark & Python (92% correlation accuracy)
- Processed **500+ news articles & stock data** to identify patterns in price fluctuations based on sentiment analysis.
- Built a dimensional data model (Star Schema) to structure sentiment analysis results.
- Built **Power BI dashboard** to visualize sentiment impact, boosting insight access by 60%
- Documented data mappings and pipeline architecture for team and stakeholder reference.

Real Estate Data Pipeline Project [https://github.com/Rishabh1627rawat/-Real-Estate-Data-Engineering-Analytics-Pipeline]

- Developed a streaming pipeline to migrate real estate data from JSON to PostgreSQL via AWS S3, enabling smooth dashboard integration and analysis.
- Used Spark for data cleaning and transformation, storing processed data in S3 (Parquet format) as part of a structured “silver” layer.
- Designed a normalized PostgreSQL schema and built interactive dashboards using PowerBi to visualize trends like price per sq.ft and BHK distribution.
- Containerized services with Docker and implemented modular, scalable ETL workflows.
- Ensured data integrity and quality at each stage of the ETL process