Rohit Ashva

📞 +919828799900   ✉ rohit.ashva900@gmail.com   in linkedin.com/in/rohit-ashva/   💻 GitHub

Career Objective

Detail-oriented Data Engineer with a Master's degree and a Data Science certification from IIT Roorkee. Currently gaining hands-on experience through an internship, specializing in data engineering and analytics.

SKILLS

Programming Languages: Python, SQL, **PySpark**, Pandas

Cloud & ETL Tools: Azure Data Factory, **Apache Airflow, Databricks,** Microsoft Azure Data Lake Gen2, Amazon S3

Databases & Data Warehousing: PostgreSQL, MySQL, Azure Synapse Analytics

Data Processing & Automation: Docker, Git, Jupyter Notebooks, Linux, APIs, Power BI, Beautiful Soup, Selenium, Requests

Data Modeling & Quality: dbt (transformations & testing)

Soft Skills: Communication, **problem-solving,** attention to detail

Professional Experience

Data Engineer | LegisIQ | April 2025 – Present

- Designed and shipped **30+ production ETL pipelines** in Python (Selenium, Playwright, BeautifulSoup, Requests) to extract, clean, and normalize multi-site web data.
- **Packaged pipelines as Python wheels** and **deployed on Azure Databricks** as scheduled jobs for dependable daily refreshes.
- Wrote curated datasets to **Azure SQL Database** using idempotent upserts, schema validation, and basic quality checks.
- Added pragmatic monitoring (logging/retries/alerts) and GitHub-based versioning to keep runs stable and easy to maintain.

Data Engineering Intern | Mactores Cognition Pvt. Ltd. | April 2025

- Training in **Databricks** with company-sponsored certification.
- Developing skills in **big data processing & ETL pipelines**.

Data Engineering Intern | Success Ladder Technologies Private Limited. | (Jan 2025 - March 2025)

- Built and optimized **ETL pipelines** using **Azure Data Factory** and **PySpark** for efficient data ingestion and transformation.
- Implemented **Delta Lake storage** in Azure Data Lake for **scalability and versioning**.
- Developed **SQL queries and Power BI dashboards** to generate actionable business insights

PROJECTS

Dynamic ETL Pipeline Project with Azure - GitHub

- Developed an **end-to-end ETL pipeline** using **Azure Data Factory** to process **800K+ rows** from the **New York City Taxi API**.
- Utilized **Databricks** for transformation, implementing **Bronze, Silver, and Gold** data architecture.
- Optimized storage with **Parquet and Delta Lake**, reducing query execution time by **30%**.
- Implemented **recursive file lookup and DDL schema management** for efficient workflow execution.

Python Data Pipelines on AWS using Airflow and S3 - GitHub

- Designed **Apache Airflow DAGs** to automate **ETL pipelines** on **AWS EC2**, extracting and processing raw data.
- Orchestrated data workflows and automated **data transformations**, reducing execution time by **40%**.
- Stored **processed data in Amazon S3** for scalable and efficient access.

EDUCATION

Indian Institute of Technology, Roorkee, India (E-Learning)	**Jul 2023 - Aug 2024**
Advanced Certificate Program in Full Stack Software Development (**Specialization: Data Science**)	3.8 GPA
Jaipur National University (Completed a degree through Distance Learning)	**Jul 2019 - Aug 2021**
Master of Computer Applications (MCA)	7.6 CGPA
Maharshi Dayanand Saraswati University, Ajmer, India	**Jun 2014 - Jul 2017**
Bachelor of Computer Applications (BCA)	6.7 CGPA

CERTIFICATES

Microsoft Certified:Databricks Certified Data Engineer Associate *(In Progress)*

The Complete SQL Bootcamp: Master SQL & PostgreSQL (HackerRank)

Achievements & Extra-Curricular

- Solved 100 + DSA questions on various platforms like Leetcode, GfG, HackerRank, and more