

## ATUL YADAV

+91 7747945847 | Indore (MP), India  
[LinkedIn](#) | [atulyadav.er@gmail.com](mailto:atulyadav.er@gmail.com) | [GitHub](#)

---

### Education

**Master of Technology in Computer Science and Engineering (AI & ML)** 2023 - 2025  
*Technocrats Institute of Technology, Bhopal, MP*

**Bachelor of Technology in Computer Science** 2019 - 2023  
*Swami Vivekananda College of Engineering, Indore, MP*

---

### Skills

Python, SQL, Apache Spark, Hadoop, Hive, Airflow, Kafka, Fivetran, AWS Glue, MySQL, PostgreSQL, MongoDB, AWS (S3, Redshift, EC2), Git, GitHub, Docker, Kubernetes, Jenkins, Terraform, Ansible, Snowflake, Shell Scripting, Scrapy, Linux, Windows

---

### Professional Experience

#### Data Engineering Intern

*Gumanco Software Solution* Nov 2021 – Oct 2022

- Engineered and optimized ETL pipelines, processing over **10TB of raw data**, enhancing data processing efficiency by 20%.
- Automated data collection workflows using **Python** and **Apache Airflow**, reducing manual efforts by 30%.
- Collaborated with data scientists to preprocess datasets for machine learning models, boosting model accuracy by 15%.
- Acquired **AWS Cloud** expertise, scoring 95% in internal training, and applied solutions to optimize data storage and workflows.

#### Data Engineer Intern

*Zintellix* Sep 2024 – Present

- Designed and optimized **ETL pipelines** for an advanced analytical tool, ensuring efficient data ingestion, transformation, and reporting.
- Performed extensive **data cleaning** and preprocessing, improving data quality by 25% and ensuring seamless downstream analytics.
- Developed and maintained **real-time data streaming pipelines** using **Kafka**, enabling dynamic and responsive analytics dashboards.
- Architected data storage solutions using **Snowflake** and **AWS S3**, reducing query latency by 30%.

---

### Projects

#### 1. Automated Web Scraping and ETL Pipeline for Financial Data

- Built an auto scraping system to collect financial data from multiple sources.
- Python, Scrapy, Apache Airflow, AWS S3, AWS Redshift, SQL*

#### 2. Competitor Analysis Tool for Retail Industry

- Developed a competitor analysis tool to scrape product details, prices, and promotions.
- Python, Scrapy, Pandas, AWS S3, Matplotlib*

#### 3. Data Insight Tool

- Build a tool to analyze uploaded datasets (CSV, Excel, JSON) for cleaning and insights.
- Python, Pandas, SQL, Apache Superset*

#### 4. Stock Market Real-Time Data Analysis

- Engineered a real-time stock market analysis platform to ingest data streams using **Apache Kafka**.
- Python, Pandas, Plotly, Kafka, AWS*

---

### Certifications

Apache Airflow Fundamentals, AWS Cloud Computing, Data Science, Python nanodegree