Neha Yadav

Data Engineer

neha.yadav@example.com +91 87XXX XXX93 Gurugram, Haryana linkedin.com/in/neha-yadav-example

PERSONAL STATEMENT

Data engineer with 5 years of experience building batch and streaming pipelines for e-commerce and insurance analytics. Works with PySpark, SQL, Apache Airflow, Kafka, Databricks and AWS (S3, Glue, Redshift), and has delivered platforms processing over 2 TB of data a day.

WORK EXPERIENCE

Senior Data Engineer · **Aravalli Data Systems Pvt Ltd, Gurugram** Sep 2023 – Present

- Built a lakehouse on Databricks and Delta Lake for an e-commerce client, ingesting 2 TB of order, clickstream and inventory data a day.
- Rewrote 60+ legacy Hive jobs in PySpark with better partitioning and broadcast joins, reducing nightly batch time from 7 hours to 2.5 hours.
- Orchestrate 140 Airflow DAGs with Great Expectations data-quality checks, cutting data issues reported by business teams by 65%.
- Set up Kafka streaming for real-time order status, giving the operations team updates within 2 minutes instead of the next day.

Data Engineer · **Kurukshetra Insurtech Analytics, Noida** Jun 2021 – Aug 2023

- Designed star-schema data marts in Amazon Redshift for policy, claims and agent-performance reporting.
- Built AWS Glue and Lambda ingestion jobs for 25 source systems, including SFTP files from partner banks.
- Wrote SQL reconciliation queries with window functions that flagged ₹3.2 crore of mismatched premium records in the first audit.

EDUCATION

B.Tech in Computer Science and Engineering · **Faridabad Institute of Engineering, Faridabad** 2017 – 2021

Grade: 7.5 CGPA

SKILLS

Python / PySpark · SQL · Apache Airflow · Apache Kafka · Databricks / Delta Lake · AWS (S3, Glue, Redshift) · dbt · Data modelling

CERTIFICATIONS

Databricks Certified Data Engineer Associate · **Databricks**
Jun 2024

LANGUAGES

English · Professional

Hindi · Fluent

CORE SKILLS

- Pipeline design
- Data quality
- Performance tuning
- Dimensional modelling
- Stakeholder collaboration

MyPerfectResume Free