

ANKIT SAXENA

ankit282k@gmail.com | +91 7500335301
[linkedin.com/in/ankit282k](https://www.linkedin.com/in/ankit282k) | github.com/ankit282k

PROFESSIONAL SUMMARY

Data Engineer with 2+ years of experience designing, building, and optimizing scalable data pipelines and ML-ready platforms. Strong hands-on expertise in PySpark, Databricks, Snowflake, and Delta Lake for high-volume data processing. AWS Certified Cloud Practitioner with practical experience using AWS services for cloud-based analytics and AI workflows. Proven ability to deliver production-grade ETL systems, optimize performance and cost, and support end-to-end machine learning and GenAI use cases.

CORE SKILLS

Data Engineering

PySpark, Databricks, Delta Lake, Snowflake, SQL, ETL Pipelines, Data Modeling, Medallion Architecture, Kafka (Streaming)

Cloud & DevOps

AWS (Certified Cloud Practitioner – S3, EC2, IAM, CloudWatch, Redshift), Docker, GitHub Actions, CI/CD

Programming & Workflow

Python, FastAPI, Airflow/Prefect, Linux, Git

Analytics & ML (Supporting)

MLflow, Time Series Forecasting, Feature Engineering

PROFESSIONAL EXPERIENCE

AI/ML Engineer I — Simplifix Systems (Client : Vijay Sales)

Jan 2026 – Present

- Built end-to-end **Databricks pipelines** using **PySpark, AWS, and MongoDB** to process large-scale retail sales and inventory data for analytics and forecasting.
- Implemented **21-day time-series demand forecasting** at item–store level to support **inventory planning and stock replenishment**, improving product availability.
- Leveraged **Databricks AutoML** to train, evaluate, and select optimal forecasting models, reducing manual experimentation time and accelerating model development.
- Used **MLflow** to track experiments, manage model versions, and deploy validated forecasting models into production workflows for ongoing monitoring.

Associate Software Developer — Bosch Global Software Technologies

Feb 2025 – Jan 2026

- Designed and optimized PySpark-based ETL pipelines for internal AI platforms, reducing processing latency by **28%** and storage costs by **22%**.
- Built cloud-ready data workflows supporting ML and GenAI use cases using Databricks, Delta Lake, and distributed Spark jobs.
- Engineered LLM-powered microservices using **FastAPI and HuggingFace**, integrating secure data pipelines that improved team productivity by **32%**.
- Implemented CI/CD pipelines with **GitHub Actions and Docker**, automating testing and deployment and reducing production issues by **80%**.

Data Engineer Trainee — FranConnect Pvt. Ltd.

Sep 2024 – Feb 2025

- Built and maintained large-scale ETL pipelines using **PySpark, Hive, and SQL**, processing **30M+ records daily** with a **45% performance improvement**.
- Implemented incremental ingestion, schema evolution, partition tuning, and caching strategies, reducing Spark job runtime by **22%**.

- Automated data pipeline deployments using **Docker and GitHub Actions**, cutting integration failures by **80%**.

Data Scientist Trainee — Dyno India Pvt. Ltd.

Feb 2023 – Aug 2023

- Built deep learning pipelines using **CNN and LSTM**, improving validation metrics by **30%**.
- Implemented recommender systems using **Collaborative Filtering and FISM**, improving recommendation precision by **25%**.
- Executed complete ML workflows from EDA and feature engineering to model deployment and documentation.

PROJECTS

End-to-End Databricks ETL Pipeline (PySpark, Delta Lake, AWS)

- Designed a **Bronze → Silver → Gold** Medallion Architecture using Databricks and Delta Lake.
- Implemented ingestion, incremental loads, validation, and transformation jobs for large datasets.
- Used **AWS S3** as the storage layer with IAM-based access control.
- Improved pipeline runtime performance by **42%**.

Snowflake Automated Data Warehouse (Snowpipe, Streams, Tasks)

- Built automated ingestion pipelines using **Snowpipe** from cloud object storage.
- Designed **fact and dimension tables** with SCD logic for analytics use cases.
- Implemented **Streams and Tasks** for continuous incremental data processing.

Time Series Forecasting with MLflow

- Developed forecasting pipelines using **XGBoost and ARIMA**, achieving **21% MAPE improvement**.
- Tracked experiments, parameters, and models using **MLflow**.

CERTIFICATIONS

- **AWS Certified Cloud Practitioner (CLF-C02)**

EDUCATION

Bachelor of Technology – Electronics & Communication Engineering

Chandigarh Engineering College

CGPA: **7.93** | Year: **2023**