Ankit Singh

Data Engineer

📍 Sector 15, Noida, UP   📞 +91 82996 57259   ✉ ankitsingh231225@gmail.com   ⌨ github.com/Ankit-singh05   🔗
linkedin.com/in/ankit-singh-2b2a71295

Professional Summary

Data Engineer with hands-on expertise in **Python, SQL, and PySpark**, specializing in building robust **ETL pipelines** and data transformation workflows. Experienced with **Azure Databricks** for large-scale data processing, with a strong focus on data quality, pipeline scalability, and delivering analytics-ready datasets.

Experience

Data Engineer Trainee   **Feb 2025 – Feb 2026**
Highendservices (Payroll: CETPA InfoTech)   *Delhi*

- Worked on **SQL** for data extraction and transformation; performed data cleaning using **Python** and **Pandas**.
- Assisted in **ETL pipeline development** using **PySpark** on **Azure Databricks** for large-scale distributed data processing.
- Implemented **SCD Type 2** for historical data tracking; ensured **data quality and validation** across pipelines.

Technical Skills

Programming	Python, SQL	**Big Data**	PySpark
Cloud	Azure Databricks	**Libraries**	Pandas, NumPy
Databases	MySQL, SQLite	**Data Formats**	CSV, JSON, Excel
Tools	Git, GitHub, VS Code, Jupyter Notebook		

Projects

⚡ **Smart Meter Data Cleaning Pipeline** — *Python, PySpark, SQL, Azure Databricks*

- Engineered an end-to-end pipeline for Smart Electric Meter data including **duplicate removal**, **missing value handling**, and **timestamp standardization** using PySpark on Azure Databricks.
- Delivered clean, analytics-ready datasets enabling downstream reporting on household energy consumption.

🔻 **Data Cleaning & Transformation Pipeline** — *Python, Pandas, NumPy, SQL* | ⌨ GitHub

- Built a full ETL pipeline for large-scale CSV datasets with automated **duplicate detection**, **null value imputation**, and **data standardization** enabling reliable downstream analytics.

☁ **API-to-Database Integration Pipeline** — *Python, MySQL, REST APIs* | ⌨ GitHub

- Developed an automated ingestion pipeline to extract **JSON data from REST APIs**, apply transformation logic, and load structured records into a **MySQL** database, eliminating manual data entry.

Education

B.Tech in Computer Science & Engineering   **2021 – 2025**
Dr. A.P.J. Abdul Kalam Technical University (AKTU)   *Lucknow, Uttar Pradesh*

Higher Secondary Education (12th Grade) – CBSE   **2021**
St. Francis Public School (CBSE)

Training & Certifications

Azure Databricks Training – CETPA Infotech & Infogain Collaboration

- Industry-focused training on Azure Databricks covering distributed data processing, PySpark workflows, and cloud-based pipeline development.

Languages

English (Proficient) — **Hindi** (Native)