

Ashutosh Sharma

+91 97940 98116 | sharma.aashu1208@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Software Engineer and Data Engineer with almost 4 years' experience building ETL/ELT pipelines using AWS, Databricks/PySpark, Python, and SQL, along with REST API integration (FastAPI, client-side API consumption). Domain experience in healthcare (HIPAA/PHI), FMCG e-commerce, and gaming data, with Medallion Architecture (Bronze/Silver/Gold) pipelines, data governance, and Git-based CI/CD workflows.

TECHNICAL SKILLS

Data Engineering: PySpark, Databricks, Delta Lake, Unity Catalog, Medallion Architecture, ETL/ELT Pipelines, Data Quality, Workflow Orchestration, Spark Performance Tuning, Data Modeling (Star Schema)

Languages: SQL (Window Functions, CTEs, Joins, Query Optimization), Python, C#

Cloud & Tools: AWS (S3, Redshift, Lambda), Azure DevOps (CI/CD), Git, GitHub, Bitbucket, Docker, Postman

Databases & Backend: MySQL, SQLite, Pandas, FastAPI, REST APIs

PROFESSIONAL EXPERIENCE

Senior Associate

March 2026 - July 2026

ABHARTECH

On-site

- Designed scalable ETL/ELT pipelines using Python and AWS (S3, Lambda) to ingest player telemetry, game session logs, and in-game transaction data from card game applications.
- Optimized **SQL** queries for extraction, aggregation, and performance tuning on large datasets; maintained code quality via Git/Bitbucket workflows and code reviews.

Software Developer

Oct 2024 - Feb 2026

Sequent Inc.

On-site

- Built **Lakehouse pipelines** on **Databricks** using **PySpark** and **Delta Lake** following **Medallion Architecture** (Bronze/Silver/Gold).
- Engineered end-to-end pipelines with Python, PySpark, and AWS (S3/Lambda) for health-tech and dental care product telemetry, applying **HIPAA** and **PHI** masking/privacy standards.
- Processed 10-20+ GB structured and unstructured datasets in cloud storage; configured access controls and lineage for sensitive healthcare datasets using **Unity Catalog**.
- Implemented **data quality** checks (validation, completeness, accuracy, consistency), reconciliation, and exception handling across Bronze/Silver/Gold layers; built reusable PySpark transformation modules.
- Tuned **Spark** workloads (partitioning, joins, caching) for performance and cost; orchestrated pipelines with **Databricks Workflows**, including monitoring and production troubleshooting.

Software Developer

Mar 2022 - Aug 2023

Hashtag Bazaar

On-site

- Built ETL pipelines for an **FMCG** e-commerce platform to ingest product catalogs, user activity logs, and transaction data into cloud storage; used **AWS S3** for large datasets and assets.
- Performed **data cleaning and filtering** on high-volume FMCG sales and product data using Python and staged datasets in **Databricks** for analytics.

DATA PROJECTS

Walmart Sales Analytics & SQL Dashboard

[Repository](#)

- Analyzed retail sales data with advanced **SQL** (window functions, CTEs, aggregations, JOINS) for revenue trends and product performance.

Amazon Prime Customer Churn Analysis

[Repository](#)

- Built a reproducible **SQL** workflow for retention analysis: churn rate and high-risk customer segments.

Document Reader Chatbot (LLM-Powered)

[Repository](#)

- Developed a PDF-querying chatbot using **Python** and **FastAPI** with system prompting and conversational memory.

EDUCATION

Data Engineering Program

2026

Bosscoder Academy

Remote

MCA (2019-2021) & BCA (2016-2019)

Integral University

Lucknow, India