

SUMER SHINALKAR

+91-8390302819 | shinalkarsumer@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

M.Tech Data Science student at COEP, AWS-trained Data Engineer and Analyst with expertise in Real Time pipelines (Kafka, Spark). Proven ability to architect scalable ETL systems and optimize query performance. Seeking to leverage skills in Big Data Engineering.

EDUCATION

- **M.Tech, Data Science**
COEP Technological University (CGPA: 8.0/10)
- **B.E, Computer Science**
Savitribai Phule Pune University (CGPA: 8.5/10)

TECHNICAL SKILLS

- **Languages & Compute:** SQL (Window Functions, CTEs), Python (Pandas, PySpark), Apache Spark (Structured Streaming).
- **Data Engineering & Cloud:** Apache Kafka, Delta Lake, dbt, DuckDB, Airflow, AWS (S3, Lambda, EventBridge, IAM, Glue), Docker.
- **Storage & Visualization:** PostgreSQL, Redis, DynamoDB, Tableau, Power BI.
- **Soft Skills:** Agile Methodology, Cross-functional Collaboration, Technical Documentation, Problem Solving.

KEY PROJECTS

BidLock: Distributed Real Time Bidding Engine | Kafka, Spark, Redis, Docker

- Video Link :- [Demo of BidLock](#)
- Architected a negotiation platform handling concurrent user bids with sub-millisecond latency using Redis for state.
- Engineered a Spark Structured Streaming pipeline to process Real Time events, enforcing business logic validation.
- Solved concurrency issues by implementing Optimistic Locking, eliminating race conditions during simultaneous bidding.
- Deployed microservices architecture using Docker Compose, orchestrating Kafka and Spark containers.

ShopPulse: Real Time E-Commerce Data Lakehouse | Kafka, Spark, Delta Lake

- Video Link :- [Demo of ShopPulse](#)
- Built a scalable ETL pipeline to ingest high-velocity clickstream logs via Kafka, processing data using Spark Streaming.
- Implemented ACID transactions on the data lake using Delta Lake, ensuring strict data consistency for analytics.
- Optimized query performance by 40% through storage techniques like Partitioning and Z-Ordering.

Modern Data Warehouse Pipeline | dbt, DuckDB, SQL

- Video Link :- [Demo of Modern Data Warehouse Pipeline](#)
- Developed a modular ELT pipeline using DuckDB for local compute and dbt for SQL-based transformation.
- Automated rigorous Data Quality Tests (Schema validation, Null checks) within dbt, reducing bad data ingestion by 90%.

EXPERIENCE

Data Analytics Intern | Ceres Canopus Pvt Ltd | July 2025 – January 2026

- Optimized complex SQL queries for client reporting, reducing data retrieval time by 40% to accelerate decisions.
- Automated manual daily workflows using Python scripts (Pandas), saving the team 5+ hours weekly.
- Collaborated with engineering teams to validate data integrity across source systems, ensuring 100% accuracy.

PUBLICATIONS

- FinSight: A Hybrid Real Time Multi-Source Framework for Stock Price Prediction (IEEE ESCI 2025(Accepted))
- Stock Price Prediction Using LSTM (IJRPR - International Journal of Research Publication & Review)

CERTIFICATIONS

- AWS Cloud Practitioner (Udemy, 2025)
- AWS Data Engineering Essentials (Udemy, 2025)
- Data Analytics Professional Certificate (Google, 2025)
- Responsible & Safe AI (NPTEL – IIT Hyderabad)