Shreeyash Shrestha

Old Baneshwor Kathmandu, Nepal | shreeyashofficialpc1@gmail.com | 9767658897 |

PROFILE

Data Engineer and BCS (Hons) Data Science student with hands-on experience building batch and streaming ETL pipelines in Python, SQL, PostgreSQL, Apache Airflow, Apache Spark (PySpark), Apache Kafka, dbt, and Docker. Skilled in medallion (Bronze/Silver/Gold) architecture, Kimball dimensional modeling, incremental loads, watermarking, and automated data quality checks. Seeking data engineering opportunities to build reliable, scalable, production-ready analytics platforms.

EDUCATION

IIMS College, Kathmandu, Nepal, Kathmandu

Bachelor of Computer Science (Hons) — Data Science, 2027

- Specialization in Data Science with coursework covering Machine Learning, Database Management, Big Data Analytics, Statistics, Data Structures, and Software Engineering
- Engaged in hands-on lab work involving real-world datasets and applied analytics techniques

TRAINING & CERTIFICATIONS

Data Engineering Training, Kathmandu, Nepal

- Designed Python-based ETL pipelines with integrated data quality checks and watermarking.
- Orchestrated multi-step workflows using Apache Airflow DAGs with branching, retries, and monitoring.
- Applied dbt for SQL-based transformations and testing across layered data architectures.
- Built batch and streaming pipelines using Apache Kafka and Apache Spark for scalable data processing.
- Modeled analytics-ready data in PostgreSQL using dimensional modeling principles.

Professional Ethical Hacking Training, Kathmandu, Nepal

- Completed 60 hours of structured training on ethical hacking methodologies, tools, and cybersecurity fundamentals.
- Built foundational knowledge of network security, vulnerability assessment, and penetration testing concepts.

Workshop: Future Forward — Communication Skills in Action

- Participated in an intensive communication and professional skills development workshop.

PROJECTS

Airflow + dbt Containerized Data Pipeline

- Built a fully containerized 3-layer medallion data pipeline using Docker to ingest public API data and process it through Bronze, Silver, and Gold layers.
- Orchestrated ingestion and transformation tasks using Apache Airflow DAGs for automated scheduling and dependency management.
- Implemented dbt models to build Kimball-style dimensional analytics tables in the Gold layer, optimized for analytical queries.

End-to-End ETL Pipeline with Data Lake Integration

- Developed a production-ready ETL pipeline integrating a Rustfs/S3 data lake with PostgreSQL as the final analytics destination.
- Applied branching practices within Airflow DAGs for conditional pipeline execution and error handling.
- Showcased scalable storage and workflow orchestration patterns for enterprise-style ingestion.

dbt Medallion Architecture & Kimball Dimensional Modeling

- Implemented a banking analytics data model using dbt and a Bronze/Silver/Gold medallion architecture for structured transformation.
- Designed Kimball-style fact and dimension tables to support analytics use cases, with synthetic data generation scripts included.
- Automated data enrichment and loading with Python scripts and containerized the environment using Docker Compose.

Python ETL Pipeline with Airflow & Data Quality Checks

- Built Python-based ETL pipelines covering extraction, transformation, and loading stages with built-in data quality validation.

- Implemented watermarking for incremental loads and configured Airflow for automated scheduling and monitoring.

TECHNICAL SKILLS

Python, Apache Spark (PySpark), ETL Pipelines, Kimball Dimensional Modeling, NumPy, Docker Compose, Windows, SQL, Apache Kafka, ELT, Data Warehousing, Matplotlib, Git, PostgreSQL, dbt, Data Modeling, Workflow Orchestration, Seaborn, Version Control (GitHub) , Apache Airflow, Rustfs/S3, Medallion Architecture (Bronze/Silver/Gold), Pandas, Docker, VS Code, Cloud Platforms (AWS, GCP, Azure), Python Libraries (Pandas, NumPy), Business Intelligence (Power BI, Tableau), SQL Optimization.

LANGUAGES

Nepali (Native), English (Professional Proficiency), Hindi (Conversational)

LINKS

Github: https://github.com/ShreeyashShrestha

Linkedin: https://www.linkedin.com/in/shreeyash-shrestha-169a67367/

ACCOMPLISHMENTS

- Built and deployed 4 end-to-end batch and streaming data pipelines using Python, Apache Airflow, Apache Spark (PySpark), Apache Kafka, and dbt to produce analytics-ready Bronze/Silver/Gold datasets.
- Designed Kimball-style dimensional models and implemented ELT transformations in dbt to deliver query-ready analytics tables.
- Implemented incremental loads, watermarking, and automated data quality checks to improve data freshness, consistency, and reliability.
- Containerized pipeline environments with Docker and Docker Compose to support reproducible, production-like deployments.
- Integrated Rustfs/S3 data lake storage with PostgreSQL as the analytics destination to support scalable storage and reporting.
- Orchestrated Airflow DAGs with branching, retry policies, and monitoring to automate scheduling, error handling, and observability.