# Shrey Niraula

Data Engineer | ETL & Pipelines | Lakehouse Architecture | AWS
Huntsville, AL • +1 (475) 251-5056 • nshrey53@gmail.com • LinkedIn • GitHub • Portfolio

## SUMMARY

Data engineer with 4+ years building and scaling ETL pipelines, OLAP reporting, and lakehouse platforms across digital marketing and satellite-data domains. Proven record migrating legacy pipelines to modern lakehouse architectures, automating manual workflows, and partnering with US stakeholders to turn business requirements into reliable, high-performance data infrastructure.

## TECHNICAL SKILLS

**Languages:** Python, SQL, Scala, C++
**Cloud & Big Data:** AWS (S3, Glue, Athena, Redshift, EMR, Lambda, DynamoDB), Apache Spark (PySpark), Apache Airflow, PySTAC
**Data & Reporting:** Data Lakehouse, Data Warehousing, Dimensional Modeling, SCD, OLAP, ETL/ELT, Data Acquisition, Data Governance, Materialized Views, pandas
**Databases & Caching:** MySQL, Postgres, Redis, Dremio, Iceberg
**Engineering:** Docker, Terraform, CI/CD, RabbitMQ, Flask, REST APIs, Multithreading, Distributed Systems, OOP & Design Patterns (Factory, Decorator), Pytest, Git
**AI/ML:** Local LLMs (Ollama), Multi-Agent Systems, Pydantic, Prompt Engineering, LLM Evaluation

## PROFESSIONAL EXPERIENCE

**Data Engineer, NASA ODSI (CSDA Program)** *Aug 2024 – Present*
Huntsville, AL, USA

- Built and evaluated a **multi-agent system on a local LLM** that converts Amazon S3 inventory reports into auditable, cost-saving storage-migration plans, using verifier-driven retries for numerical accuracy; validated on petabyte-scale CSDA production buckets against a status-quo baseline.
- Onboarded a new satellite vendor (Airbus DEM) end-to-end, building its ingestion and PySTAC-based STAC transformation from scratch.
- Rebuilt legacy Airbus SAR and BlackSky ingestion into a centralized, partitioned Airflow pipeline, with Pytest suites validating ingestion logic.
- Built an Airflow solution to automate the manual MAXAR vendor checksum-verification process, consolidating DynamoDB export and Athena checks into a single DAG.

**Mid-Level Data Engineer, GrowByData** *May 2022 – Jul 2024*
Kathmandu, Nepal · Hybrid (US-based stakeholders)

- Owned the **LeadGen data source end-to-end** (POC, acquisition, ETL, and reporting backend).
- Re-architected the **Facebook Ads pipeline end-to-end** (acquisition, ETL, reporting) into a modular, multithreaded design supporting easy extension of pulls and filters.
- Migrated the reporting warehouse from Redshift (ra3.4xlarge) to a **Dremio lakehouse (S3 + Glue Catalog)**, decoupling storage and compute to scale compute independently for heavy ad-hoc/aggregation queries.
- Refactored demo-report anonymization from existing duplicated query classes into a reusable Python decorator with pandas post-processing, cutting 100+ report development **from ~3 months to ~2 weeks**.
- Sped up acquisition **2–5×** across Pinterest, TikTok, Google/Bing Ads, and Shopify via multithreading and memory tuning.
- Integrated Redis caching into the reporting layer, cutting query latency **~70%**.
- Mentored junior engineers and interns across data acquisition, ETL, reporting logic, Spark, and data modeling.

**Associate Data Engineer, GrowByData** *May 2021 – May 2022*
Kathmandu, Nepal · Hybrid (US-based stakeholders)

- Built the first production acquisition pipeline for Pinterest Ads with OAuth2 authentication.
- Refactored the Facebook Ads acquisition to handle API throttle limits, stabilizing ingestion.
- Migrated Talend ETLs to distributed Spark (Scala + SparkSQL) using star-schema modeling and SCD, **reducing Redshift CPU load and ETL latency**.

## EDUCATION

**M.S., Computer Science (Data Science Concentration), University of Alabama in Huntsville** *Aug 2024 – Aug 2026*
GPA: 4.0 • Coursework: Advanced Database, Big Data Computing, Cloud Computing, Machine Learning

**B.E., Electronics & Communication Engineering, Tribhuvan University (Pulchowk Campus)** *2016 – 2021*
Top 1% (Rank 4 of 700+), 81.70% • Ncell Excellence Award (consecutive semester topper)

## CERTIFICATIONS

- AWS Certified Data Engineer - Associate (DEA-C01) *(In Progress)*
- Deep Learning Specialization; Stanford Machine Learning (Andrew Ng); JLPT N2 (Japanese)