Snehil Upadhyay

+91-9717001225 | snehilupadhyay007@gmail.com

LinkedIn | **GitHub**

PROFILE

Data Engineer with 3+ years of experience in building large-scale data pipelines, ETL and ELT processes, and data warehouse solutions. Utilized technologies like **Python, SQL, Spark, BigQuery, DataProc, and Airflow** to develop scalable big data solutions for Accenture's **Marketing and Communication** domain.

TECHNICAL SKILLS

Programming Languages:	Python, SQL, Scala
Big Data Technologies:	Spark, PySpark, Spark SQL, Polars
Cloud computing:	BigQuery, Composer, DataProc, Cloud Functions, Pub/Sub, Cloud Storage(GCS)
Data Engineering Tools:	Azure DevOps, Data Modelling, ETL/ELT Data Pipeline
Orchestration:	Airflow
Version Control:	Git, GitHub
Familiar:	LangChain, Vertex AI, LLM Models, Prompt Engineering, RAG, Streamlit

EXPERIENCE

Accenture — August 2021 - Present
Data Engineer Analyst — *Noida, India*

- Designed and implemented a **Polars-based pipeline**, achieving **20x** faster performance compared to Pandas, reducing processing time and costs.
- Developed Dynamic Priority Mapping using **Apache Airflow Task Groups**, resolving job dependencies in concurrent pipelines, achieving **99%** reliability for critical jobs, and improving system efficiency.
- Transitioned from annual full loads to daily delta loads, implementing SCD1 to refresh data **12x** faster, enabling near-real-time insights for key business processes.
- Created an asynchronous Google Cloud Function using asyncio to fetch data from an API, processing **800+ pages in 6 minutes** and successfully storing (**38 million records, 15GB**) in Google Cloud Storage for under **$10**.
- Streamlined data transformation and retrieval in BigQuery, implementing an optimized append-load strategy and creating structured pivot tables for advanced analytics.
- Collaborated with CIO M+C stakeholders to normalize complex nested JSON data in BigQuery, addressing ambiguities and PII challenges.
- Developed aggregated BigQuery tables by converting complex SQL queries into Spark Transformations, supporting business intelligence and reporting.
- Worked on a GENAI Use Case in GCP: Text to SQL:
  * Designed and implemented NLP2SQL systems to automate SQL query generation from natural language, streamlining data query processes and ensuring accuracy.
  * Optimized LLM performance in Vertex AI Studio by fine-tuning parameters and leveraging prompt engineering techniques, including few-shot examples.
  * Engineered an HTTP-triggered Semantic Search Cloud Function using RAG (Retrieval-Augmented Generation) and cosine similarity to retrieve precise results from BigQuery embeddings.

AWARDS & HONORS

- Received **25+** recognitions and awards, including Client Value Creation, Best People, and Exemplify Client-Centricity from Accenture Leadership, Business Stakeholders, and Tech Architects.
- Awarded the **Pinnacle Award** as a high performer across the Delivery Unit (DU) for FY23.
- Recognized as **Star of the Month** across the Delivery Unit (DU) in July FY22.

EDUCATION

Galgotias University — Greater Noida, Uttar Pradesh
Bachelor of Technology - Computer Science & Engineering — *August 2017 - May 2021*

JBS Children Paradise — Kota, Rajasthan
Senior Secondary School — *April 2015 - March 2016*