

Harshit Sharma

📍 Bangalore ✉ harshittsharma21@gmail.com ☎ 6239887837 in Harshit Sharma 📄 harsh007fc

Experience

Data Engineer

Finacplus([Toorak Capital Partners](#) [🔗](#))

Bangalore, KA
Feb 2022 – Present

- **Migrated** a **GCP DataProc + PySpark**-based system to a **SQLMesh**-powered, **Snowflake**-style architecture on **GCP**, reducing infrastructure costs by approximately **73%** (from **\$15K** to **\$4K** per month).
- **Redesigned** legacy data architecture to a Snowflake schema on **AWS** to improve query performance and report refresh times, then migrated the solution to GCP using **Airbyte**, Cloud Functions, DMS, Cloud Storage, DataProc, and **BigQuery**, resulting in approx **20%** cost reduction compared to the original **AWS** setup.
- **Designed** and implemented a scalable **Data Warehouse** using **PySpark**, AWS DMS, Kinesis, **Lambda** Function, S3, EMR, and Redshift to ingest and denormalize data from **10+ structured and semi-structured** sources; integrated AWS **Secret Manager**, CloudWatch, SNS, and SQS for secure operations, **monitoring**, and **alerting**.
- **Developed** a Python script to compare large DataFrames and generate a column-wise **difference Excel reports**, streamlining data reconciliation for QA and reducing manual comparison time from **hours to minutes**. Partnered with **Stake Holders, PMs, BI, QA** and **SRE** teams to deliver scalable analytics solutions across **AWS & GCP**, aligning with evolving **business** needs.

Software Engineer Intern

[Newton School](#) [🔗](#)

Bangalore, KA
Oct 2021 – Jan 2022

- **Developed** and maintained reusable **React** components to enhance the user interface of their website.
- **Contributed** to **backend** development with **Node.js** by building and optimizing APIs, and supported full-stack **debugging** and feature enhancements.

Skills

Languages: Python, SQL, LookML, Bash/Zsh, Java, Javascript

Technologies and Frameworks: Pyspark, Apache Spark, Sqlmesh, Hadoop, Airbyte, ETL/ELT Pipelines

Cloud Platforms: GCP (Bigquery, Cloud Storage, Cloud Functions, Logs Explorer), AWS (Redshift, S3, Lambda, EMR, Cloudwatch, Airflow)

Familiar With: Hive, Parquet, Delta lake, GCP Data Streams, GCP Scheduler, GCP DataProc, GCE, GKE, API Gateway, AWS Data Migration System, EC2, AWS Kinesis, Pub/Sub, Webhooks, PDTs, Jenkins

Other Tools: Git, Github, Jira, Confluence, Looker

Education

Sept 2018 – May 2022 **Guru Nanak Dev University Amritsar**
BE in Computer Science

- GPA: 8/10
- **Coursework:** Computer Architecture, Database Management Systems, Operating Systems, Engineering Mathematics

Projects

Retail Data Warehouse

[harsh007fc/retail-dw](#) [🔗](#)

- **Built an end-to-end retail data warehouse** with star schema dimensional modeling, encompassing all key business entities (**customers, products, orders and reviews**) with **fact** and **dimension** tables.
- Tools Used: **Pyspark, DuckDb, Airflow, Parquet, Delta**