

# Harsh Kr Gupta

Delhi | hgupta7629@gmail.com | +91-9871822508 | linkedin.com/in/hgupta7629 | github.com/harsh1330

## SUMMARY

---

**Software Engineer** with **3+ years** building ETL and reverse-ETL data infrastructure at scale: 50+ production connectors across platforms like HubSpot, Salesforce, Amazon SP-API, and Facebook Ads, including one integration that carries roughly half of the company's client traffic. Comfortable owning a pipeline end-to-end, from API integration through cloud deployment, monitoring, and on-call reliability.

## SKILLS

---

- **Languages** Python, TypeScript/JavaScript, SQL, Bash
- **Data & Storage** MySQL, PostgreSQL, MSSQL, BigQuery, Redshift, Snowflake, MongoDB
- **Cloud & Platforms** AWS (Lambda, S3, SQS, EventBridge), GCP (Cloud Functions, BigQuery), Linux, Django/Flask
- **Streaming & CDC** Debezium, Kafka Connect, Confluent Platform, Change Data Capture (CDC)
- **Tooling** Docker, Git, GitHub Actions, Selenium, Vitest
- **Focus Areas** ETL/Reverse-ETL Pipeline Design, Real-Time CDC Replication, Backend API Development, Event-Driven Architecture

## EXPERIENCE

---

**Software Engineer**, DataChannel – Gurugram June 2023 – Present

- Co-built a real-time Change Data Capture (CDC) platform as one of a 3-engineer team, streaming continuous replication from MySQL, PostgreSQL, and MSSQL into BigQuery, Redshift, and Snowflake (and across those systems in reverse).
- Built the capture layer on Debezium and Kafka Connect deployed on Confluent Platform, streaming row-level inserts, updates, and deletes with minimal replication lag.
- Designed dead-letter-queue (DLQ) handling for schema drift and malformed events, quarantining bad records for review instead of breaking the pipeline.
- Designed and built 50+ custom ETL and reverse-ETL connectors (HubSpot, Salesforce, Amazon SP-API, Facebook Ads, and others) using Python, Flask, and AWS Lambda, cutting new-client onboarding time by 30%.
- Architected a modular ingestion framework spanning APIs, databases, and flat files with schema auto-detection, giving plug-and-play compatibility across BigQuery, Redshift, and MongoDB destinations.
- Built serverless extraction pipelines (AWS Lambda, GCP Cloud Functions, headless Selenium, API mocking) for platforms with no public API, replacing manual report downloads with fully automated ingestion into S3 and client warehouses, with reverse-proxy routing to work around geo-restrictions where needed.
- Set up CloudWatch monitoring and alerting across the pipeline fleet to catch failures early and keep data delivery on schedule.
- Managed frequent Amazon SP-API version upgrades and schema changes without breaking existing client integrations, through a modular, backward-compatible connector design.

## PROJECTS

---

**Git Differ** – Personal Project [github.com/harsh1330/gitdiffer](https://github.com/harsh1330/gitdiffer)  
Live Site

- Built and shipped an internal tool for sharing a `git diff` with a teammate via a short-lived private URL, skipping throwaway branches and PRs for quick review; now used org-wide.
- Built the full-stack app with Astro (hybrid SSR) and Netlify Functions for the API, using Netlify Blobs for storage – no database or auth provider to run.
- Published a companion CLI (`npx gitdiffer`) in TypeScript/Node to npm, letting teammates share and apply diffs from the terminal with one command.
- Designed for safety by default: unguessable 131-bit share URLs, configurable expiry with automatic cleanup, and per-IP rate limiting.

## EDUCATION

---

**SRM University**, Bachelor of Technology in Computer Science

June 2019 – May 2023