

Harsh Singh

+ (91) 6359120976 | hsingh050803@gmail.com | linkedin.com/in/harsh | github.com/harsh

EDUCATION

Birla Institute of Technology and Science Pilani
Bachelor of Engineering in Electrical and Electronics Engineering

Pilani, RJ
Oct 2021 – June 2025

EXPERIENCE

- SDE-1** Jul 2025 – Present
Vegapay (Credit Fintech Startup) Bengaluru, KA
- Developed **Analytica**, a **high-throughput CDP** using **FastAPI** and **Kafka** that processes **300+ events/sec** with **<100ms p95 latency**, routing schema-validated events to **ClickHouse** for OLAP analytics, **Dittofeed** for triggered campaigns, and **Grafana** for real-time metrics with pluggable consumer architecture and async I/O
 - Owned a **Spring Boot** reporting platform on **ClickHouse** and **Postgres** EDW, generating and **validating reports** over **10M+ row datasets** and **exporting 400MB+** files with secure **S3/SFTP delivery** and full run auditability.
 - Designed and developed **Athena**, a multimodal **AI assistant** with **real-time voice/text interaction**, leveraging **LangGraph** agents, **RAG (vector + graph DB)**, **Whisper**, and **Kokoro-TTS** to enable intelligent onboarding and contextual guidance.
 - Optimized PostgreSQL to ClickHouse data sync pipeline (Temporal + Kubernetes) by implementing batched **COPY**-based ingestion, **reducing latency from 2.5 s to 0.6 s** per 10 k-row batch and **improving overall ETL throughput by 4x**.
- SDE Intern** Jan 2025 – June 2025
Vegapay (Credit Fintech Startup) Bengaluru, KA
- Built WaveCtrl, a **full-stack automation tool** using **Streamlit**, **FastAPI**, and **Celery**, enabling teams to configure and trigger data ingestion pipelines through a self-serve dashboard; **reduced manual onboarding time by automating backend workflows** and job scheduling via **RESTful APIs** and async task queues.
 - Migrated financial reporting queries from Amazon Athena to **ClickHouse**, **reducing query execution time by 70%** and **cutting operational costs by 50%**, enhancing fraud detection and transaction analytics.
- SDE Intern** Jun 2024 – Aug 2024
Samsung R&D Institute Bangalore Bengaluru, KA
- Collaborated with the Language AI Team to refine Speech Keyword Detection and Verification modules, **boosting recognition accuracy by 35%** through advanced signal processing and model optimization
 - Used Tensorflow lite models in C++ to **reduce memory footprint and power usage the model footprint by 40%**

PROJECTS

- Distributed In-Memory Cache** | *Go, Distributed Systems* Aug 2024
- Created a **Golang**-based distributed in-memory cache focused on **high availability and partitioning**
 - Used **leader-write, follower-read architecture** to dynamically elect leaders for replication
 - Utilized **Consistent Hashing** technique for sharding and seamlessly handling node failures
- BITS Fest Backend and Infrastructure** | *Python, Django, Go, Docker, Redis* Apr 2023
- Created a **Golang** based wallet system for BITS Pilani Fest apps, enabling transactions, ticketing, T-OTP and an ordering system seamlessly handling around **2.49Cr in app transactions**
 - Created internal tools for Registrations, Event, and Inventory management using **Django** and **Django REST Framework** accumulating **revenue of 48 Lacs total**
 - Managed 5000+ users** with a **peak of 1000+ concurrent users**, ensuring minimal downtime by implementing a microservice structure with **Docker**, **Redis** with scalable **SSE** servers for real-time app updates

TECHNICAL SKILLS

Languages: JavaScript, Python, Java, Go, C++

Frameworks: Django, Django-REST, Express.js, Springboot, FastAPI, Gin

Developer Tools: Git, Docker, Kubernetes, Redis, PostgreSQL, MongoDB, Redis, Kafka, Linux