SAHIL DILIP GHULE

Python Backend Engineer | Distributed Systems · Data-Intensive Platforms · Applied GenAI

work.sahilghule@gmail.com | +91-8830729336 | linkedin.com/in/sahil-ghule | Bengaluru, India

PROFESSIONAL SUMMARY

Backend engineer with 3+ years designing and owning Python services and data-intensive platforms in production. Architected a distributed analytics backend on FastAPI, ClickHouse and asynchronous workers that ingests ~1M events per day; re-architected async job execution for 3x worker concurrency and 42% faster processing. Shipped applied AI to production, including a Retrieval-Augmented Generation (RAG) pipeline and deployed machine-learning pricing models. B.Tech in Computer Science with an AI/ML specialization.

TECHNICAL SKILLS

Languages: Python, SQL, C++

Backend & APIs: FastAPI, GraphQL, gRPC, Flask, REST API design, asynchronous Python (asyncio), service integration

Distributed & Async Systems: Async worker architectures, job queues and orchestration (ARQ, RQ), scheduling pipelines, high-throughput event ingestion, concurrency and throughput optimization

Data Systems: ClickHouse, PostgreSQL, MongoDB; analytical schema design, query optimization, ETL and backfill pipelines, large-scale event data modeling

AI / ML / GenAI: Production RAG (Retrieval-Augmented Generation) pipeline integration, model serving and deployment, automated training-data pipelines, predictive modeling (Random Forest), deep learning (DNN, project work)

Tools & Practices: Git, Playwright, cloud deployment, performance profiling and system reliability

PROFESSIONAL EXPERIENCE

Software Engineer II | UrbanPiper Aug 2025 – Present

Bengaluru, India

- Own the architecture and development of Periscope, a distributed backend platform (Python, FastAPI, ClickHouse, asynchronous workers) that serves availability and downtime analytics to enterprise customers.
- Design high-throughput ingestion and analytics pipelines processing ~1M events per day across 10,000+ stores (15-minute polling, 96 checks per store daily).
- Re-architect asynchronous job execution by migrating from RQ to ARQ, increasing worker concurrency 3x and reducing end-to-end job processing time by 42%.
- Build ETL and backfill systems that process millions of records into query-ready analytical datasets without disrupting live workloads.
- Build large-scale scheduling and automation workflows integrating multiple external systems, reducing manual operational effort by 70%.
- Develop distributed crawling and indexing pipelines with Playwright for automated real-world data acquisition.
- Model and optimize ClickHouse schemas and SQL queries over large distributed datasets to keep analytical reads fast as data volume grows.

Software Engineer | Apperture AI May 2024 – Aug 2025

Remote

- Integrated parameter-extraction models into a production Retrieval-Augmented Generation (RAG) pipeline, shipping an AI-powered capability inside a live product.
- Automated AI training-data generation pipelines with reproducible Python workflows, reducing dataset preparation time by 70%.
- Built backend services powering analytical dashboards that monitor store- and item-level availability across food-delivery ecosystems.
- Designed scalable analytical storage and query systems on ClickHouse (schema design, aggregation strategy, query tuning) to sustain large analytical workloads.
- Built reusable GraphQL and gRPC services for efficient large-scale data access, and optimized backend services for performance and reliability under heavy load.

Software Development Engineer I | Cogoport Feb 2023 – Oct 2023

Mumbai, India

- Developed and deployed machine-learning dynamic pricing models (Random Forest, Python) to production for real-time pricing recommendations on large-scale logistics data.
- Built backend services powering logistics analytics dashboards, improving pricing prediction accuracy by 15% through data-quality improvements and feature engineering.

EDUCATION

B.Tech, Computer Science Engineering — Specialization in Artificial Intelligence & Machine Learning 2019 – 2023

Vellore Institute of Technology, Bhopal, India

PROJECTS & PUBLICATIONS

- AI-Powered Prostate Cancer Detection — deep-learning application predicting cancer probability; Python, Flask, DNN, deployed to cloud infrastructure.
- Power Plant Output Prediction — machine-learning models predicting full-load electrical power generation for operational efficiency decisions.
- Publication: 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Dec 2020.