

# Yash Gupta

Data Engineer

+91 9990181300 yashgupta1470@gmail.com [linkedin/yashgupta7337](https://www.linkedin.com/in/yashgupta7337) Bangalore, India

## SUMMARY

Data Engineer experienced in independently owning and scaling data platforms in fast-paced environments. Led critical migrations and architectural shifts, reducing query latency by 90%+ and infrastructure costs by 35-40%. Specialized in building cost-efficient, reliable data pipelines and real-time systems, with strong ownership in lean engineering environments.

## EXPERIENCE

### Data Engineer I

Bangalore, India

#### [Connect and Heal](#)

06/2024 - Present

A healthcare technology company focused on digital solutions to improve patient engagement and management

- Migrated data platform from cloud-native Lakehouse to ClickHouse, optimizing warehouse design and reducing query SLA from 60+ seconds to 5 seconds on average.
- Consolidated and migrated databases, reducing AWS Aurora Postgres costs across production and lower environments by 40%.
- Designed and optimized scalable data workflows using AWS S3, MWAA, Glue, DMS, Hudi, Spark, and EMR, improving processing efficiency and system stability in a medallion architecture; drove a 35% cost reduction across the lakehouse platform.
- Led data migration of 300GB from AWS DocumentDB to AWS Aurora Postgres, transforming unstructured to structured data with zero downtime and no cost increase.
- Designed real-time monitoring dashboards with automated alerting, enabling proactive issue detection and improving data reliability.
- Led PoCs across emerging data platforms (RisingWave, Redshift, Dagster, dbt, Olake) to evaluate streaming, orchestration, and transformation trade-offs, influencing platform architecture decisions.

### Software Engineer Intern

Gurugram, India

#### [MagicPin](#)

05/2022 - 08/2022

A tech startup focused on enhancing retail and shopping experiences through data-driven insights

- Enhanced data crawling and parsing processes to efficiently extract product information from multiple sources, achieving a 10% improvement in speed and performance.
- Applied Web Crawling with Python to automate data collection, improving efficiency and scalability by reducing Docker image size by 20%.
- Assisted in deploying containerized data pipelines using Docker and Kubernetes, improving scalability of data ingestion workflows.

### Summer Research Intern

Singapore

#### [National University of Singapore](#)

06/2022 - 06/2022

A leading research university specializing in innovation and advanced technology.

- Improved AI model performance through optimization and experimental evaluation in collaboration with the National University of Singapore and Hewlett Packard Enterprise.
- Group Research Project: Designed an **AI-powered fashion design system** using **Neural Style Transfer (NST) and Generative Adversarial Networks (GANs)** to create unique and innovative designs using AI.

## SKILLS

**Data Processing & Platforms:** Spark, SQL, Python, ClickHouse, Hudi, Iceberg

**Cloud & Storage:** AWS (S3, EMR, Glue, Lambda, RDS, DynamoDB, Athena, DMS, IAM, EC2, ECS, ECR, DocumentDB, Aurora Postgres)

**Orchestration & Transformation:** Airflow (MWAA), Dagster, dbt

**DevOps & Tools:** Docker, Jenkins, Git

## PROJECTS

### Stock Market Data Pipeline with Airflow on Docker

Developed a data pipeline to efficiently collect, process, and analyze stock market data for improved insights. [\[Link\]](#)

- Built an end-to-end stock market data pipeline using Apache Airflow and Docker to ingest, process, and store daily data.
- Implemented data availability checks and processed data using Dockerized Spark for reliable and scalable transformations.
- Stored processed data in MinIO (S3-compatible) and PostgreSQL, and integrated with Metabase for analytics and visualization.

### AI-Generated Fashion Design with NST and GANs

Developed an AI-powered fashion design project leveraging Neural Style Transfer (NST) and Generative Adversarial Networks (GANs) to create innovative and trend-setting designs. [\[Link\]](#)

- Reduced design iteration time from hours to seconds by building an AI-driven fashion design system using NST and GANs.
- Developed algorithms to blend artistic styles into unique patterns using advanced image processing techniques.
- Secured 2nd place at NUS for innovation and practical application.

## EDUCATION

### B.Tech in Computer Science Engineering

Shiv Nadar University

India

08/2020 - 05/2024