Surya Sarkar

Data & AI Engineer at Google

suryasarkar2@gmail.com | +917407218691 | Hyderabad, India | suryasarkar2 | in/surya-sarkar

Profile

Data & AI Engineer with 5+ years of experience building large-scale data pipelines, designing cloud-native architectures, and implementing AI/ML-driven analytics. Proven track record at Google, Mercedes-Benz, and Infosys in optimizing costs, enhancing operational efficiency, and improving data quality at scale. Known for delivering production-grade, impact-focused solutions across supply chain, partner finance, and connected vehicle analytics.

Work Experience

Data & AI Engineer
Google
Sep 2024 - Present | Hyderabad

- **Architected** data warehouse in BigQuery and Google internal engine **(10+ TB/day)** for reporting, forecasting and analytics, **reduced** query latency by **38%** via **clustering** and **partitioning** with no storage growth.
- **Unified** partner-payment analytics by building ETL pipelines integrating **SAP, Salesforce**, and other sources, cutting payment-claim errors from **40% to 10%**, saving **~2.5 hours/week** for Partners.
- **Launched** device-quality early-warning system that correlates factory shipment logs with tracking data to quarantine faulty hardware pre-launch, **preventing expensive recalls** and **improving brand reputation**.
- **Created** an AIOps alerting system to detect unusual resource usage patterns, the system can **flag issues** and **suggest targeted fixes** with **AI insights** based on RCA, preventing sudden incidents.
- **Built** LLM-based SQL Optimizer using Google ADK, reducing runtime by **45%** and cut BigQuery slot-cost by **30%.**
- **Automated** leadership reporting via HTML-based executive dashboards from raw BigQuery data using **multiple AI agents**, **reducing** turnaround time by **4×**.

Data/ML Engineer
Mercedes-Benz Research and Development India
Mar 2023 - Aug 2024 | Bangalore

- **Designed and developed** scalable data pipelines using **Spark**, **Hive**, and **Python** on Azure cloud, improving system throughput and reducing latency.
- **Led** system design for ML and analytics use cases for 15M+ connected vehicles, enabling **early fault detection** and **risk mitigation**.
- **Drove** stakeholder engagement for data governance and monitoring strategy, ensuring **100%** compliance with data quality SLAs.
- **Built** reusable data validation modules that **cut QA efforts** by **40%** across multiple ingestion layers.

Technologies involved: PySpark Dataframe API, Spark SQL, SQL Server, Hive, Azure Databricks, Azure Data Factory, Azure Data Lake, Power BI, AWS, LLM, Generative AI

Data Engineer (Digital Specialist Engineer)
Infosys
Jan 2021 - Mar 2023 | Bangalore, India
Client: Modern Analytics and Insights team for Microsoft

- **Accomplished cost-saving** by implementing cube processing from Azure Analysis Server to VM which led to saving **$2-3 Million** dollars quarterly.
- **Developed** metadata-driven framework to onboard **30+ data sources**, **cutting** development effort by **40%**.
- **Enhanced** ETL systems with **real-time logging** and **alerting**, **improving** incident response by **50%**.
- **Automated** data quality alerts to catch inconsistencies early, improving reliability and trust in reporting.
- **Delivered** robust Python/SQL code and led code reviews to ensure high-performance data solutions.
- **Led** a team of 3-4 engineers on multiple important deliverables and gained client appreciation.

Technologies involved:
PySpark Dataframe API, Spark SQL, SQL Server, Hive, Azure Databricks, Azure Data Factory, Power BI

Skills

AI & ML Technologies – Gemini and OpenAI APIs | Multi-agent orchestration (Google ADK) | LLM prompt engineering | Fine tuning LLM models

Cloud Platforms – - Google Cloud (GCP) | - Azure Synapse | - Azure DataBricks | - Azure Data Factory

Big Data Frameworks – - Apache Spark (Spark-SQL, PySpark) | - BigQuery | - Hadoop | - Hive

Languages – - Python | - Java

Database – - SQL (MySql, SQL Server) | - Relational Databases

Core CS – - Data Structures | - Distributed Systems | - Networking | - Design Patterns

Web Technologies – - HTML | - CSS

Version Control – - Azure Repos | - Github

Education

M.S
Georgia Institute of Technology
Jan 2026 - Present | Remote
Major: Computer Science
Specialization: AI

B.Tech
Maulana Abul Kalam Azad University of Technology
Aug 2016 - Jul 2020 | Kolkata, India
Major: Computer Science and Engineering
CGPA: 8.12/10

Certificates

Introduction to Generative AI with Google Cloud
Google

AZ-900: Microsoft Azure Fundamentals
Microsoft

Programming in Java
NPTEL, IIT KGP