

Soham Katkar

+12012841498 ◇ sohamkatkar2024@gmail.com ◇ Open to Relocate ◇ [LinkedIn](#)

SUMMARY

Expert in hardware and software integration with strong skills in digital and analog circuit design, embedded systems, computer architecture, and advanced C++, C, and Python programming. Proven track record in optimizing data structures, algorithms, and software-hardware solutions to enhance system performance. Skilled in microcontroller programming, signal processing, system simulation, and knowledgeable in networking protocols and operating systems. Dedicated to applying innovative engineering principles and analytical problem-solving to deliver impactful, high-quality results.

EDUCATION

Master in Information Systems, University of Maryland, Baltimore County (UMBC) (GPA: 3.62/4.0) Jan '24 — Dec '25

Bachelor in Electronics and Computer Engineering, Savitribai Phule Pune University (SPPU) (GPA: 3.30/4.0) Aug '17 — Jun '21

EXPERIENCE

Data Engineer

HSBC Technology India

Aug '17 — Dec '23

- Acted as data steward for global payment data pipelines, ensuring accuracy and lineage traceability for millions of transactions across highly regulated financial systems.
- Supported regulatory data fulfillment by developing scalable C++/Java data models and SQL-based transformations, reducing processing errors by 20% and improving compliance reporting accuracy.
- Partnered with cross-functional stakeholders (Risk, Compliance, Finance) to deliver audit-ready data solutions, contributing to smooth regulatory inspections and minimizing remediation efforts. Automated and monitored critical reporting processes with Airflow, Docker, and Kubernetes, improving reliability of regulatory submissions and reducing manual intervention by 40%.
- Conducted peer reviews and root cause analysis aligned with regulatory audit standards, improving defect detection by 15% and strengthening system integrity.

Data Analyst Intern

Incentius Solution Ltd.

Feb '21 — Jul '21

- Collaborated with regional analytics and marketing teams to manage KPIs across 5+ European markets to streamline sales data processing, reducing data consolidation time by 25% and enabling faster decision-making that supported double-digit revenue growth.
- Designed 3 enterprise-wide star schema models and applied SQL partitioning, improving query performance by 40%, ensuring 99.9% data accuracy, and standardizing reporting across 10+ executive dashboards.
- Led development of 6 automated Python and SQL data pipelines for incentive calculations impacting 100,000+ sales reps in 12 countries, increasing calculation accuracy and speed by 30%; boosted stakeholder engagement by 50% through deployment of interactive dashboards in Google Analytics 4, Tableau, and Excel, viewed by 200+ managers monthly.

CERTIFICATIONS

Oracle Database SQL Certified Associate

Feb '24

PROJECTS

IPL Data Analysis using Apache SPARK

Jan '25

- Built an end-to-end data pipeline using Apache Spark on Databricks, integrating data from Amazon S3, EC2, and AWS Lambda to enable ETL processing, SQL-based analytics, and interactive Power BI visualizations, reducing reporting latency by 40%.
- Developed a React-based dashboard powered by REST APIs and Spark-transformed datasets, improving match performance trend tracking, increasing user interactivity by 60%, and enhancing data accessibility for 300+ active users.

YouTube Data Analysis using Amazon Web Services

Jul '24

- Implemented a scalable Data Lake on Amazon S3, orchestrating ETL workflows with AWS Glue and Lambda to process 2TB+ of data daily, reducing overall processing time by 30%.
- Enabled secure, insight-driven analytics by executing 500+ SQL queries per month via Amazon Athena and enhancing security compliance by improving access control through 10+ IAM policy refinements and real-time CloudWatch monitoring.

SKILLS

Programming Java, Springboot, Unix Shell Script, C++, Python

Machine Learning Scikit-Learn, TensorFlow, Keras, SPARK, PL/SQL, MySQL, Kafka, Airflow

Data Visualization Excel, PowerBI, Tableau

DevOps AWS, Azure, Lambda, Docker, Jenkins, Kubernetes, Terraform, CloudWatch, CloudFormation

Statistical Analysis Numpy, Pandas, Pytorch, Feature Engineering