

# SOHAM HARIPRASAD KATKAR

+1-201-284-1498  [sohamkatkar2024@gmail.com](mailto:sohamkatkar2024@gmail.com)  <https://www.linkedin.com/in/soham-katkar-35742515a/>

## EDUCATION

**University of Maryland, Baltimore County (UMBC)**

**Dec 2025**

*Master's, Information Systems*

**GPA:** 3.62/4.0

**Savitribai Phule Pune University (SPPU)**

**Jun 2021**

*Bachelor of Engineering, Electronics and Computer Engineering*

**GPA:** 3.30/4.0

## WORK EXPERIENCE

**HSBC Technology India**

**Dec 2023**

*Data Engineer*

*Pune*

- Optimized system efficiency by 15% and reduced processing errors by 20% by developing scalable, multi-threaded C++ and Java-based data models, applying advanced analysis techniques, and enabling distributed system design to support high-volume global transactions.
- Resolved over 200 real-time issues collaboratively with a team of 10 engineers in an agile work flow using scrum meetings, and effective communication ensuring seamless operation of the Global Payment System via Kubernetes, Jenkins, Splunk, and GitHub with a focus on robust data processing.
- Collaborated with developers in cross functional teams in debugging and code optimization, improving code quality and reducing defects by 15% through rigorous peer reviews and collaborative root cause analysis.
- Developed shell scripts to automate the server installation process and streamlined Jenkins jobs cutting setup time by 40%.
- Improved service deployment reliability by 30% by designing Docker images and Internal Kubernetes Platform (IKP) using Docker and Kubernetes cluster monitoring.
- Enhanced internal monitoring tools using JavaScript and Node.js, resulting in a 25% reduction in average incident resolution time by enabling real-time log visualization and quicker diagnostics for production-level issues.
- Integrated Apache Airflow to orchestrate and monitor complex data pipelines across multiple environments, improving job scheduling reliability and reducing manual intervention by 40%.
- Developed and integrated RESTful APIs to enable secure, scalable internal data access and supported domain-driven design principles in collaboration with product teams.

**Incentius Solution Ltd**

**Jul 2021**

*Data Analyst Intern*

*Pune*

- Collaborated with a regional analytics and marketing team to streamline the data processing of sales records across multiple European countries and increase business performance.
- Modeled complex sales data using **star schema** designs and SQL partitioning techniques, improving query performance and ensuring high data quality and consistency across dashboards.
- Led the development of automated data pipelines to manage and process raw sales data used in calculating performance-based incentives for over 100,000 sales representatives. Built a scalable, rule-based system that cleaned, matched, and categorized input data efficiently using Python for data transformation and SQL for aggregations and validations resulting in 30% improvement in speed and accuracy of the incentive disbursement.
- Increased stakeholders report engagement by 50% by developing interactive dashboards using Google Analytics 4, Tableau and Microsoft Excel, which boosted decision-making transparency.

## ACADEMIC PROJECT EXPERIENCE

**IPL Data Analysis using Apache SPARK (Apache Spark, Databricks, AWS Lambda, Power BI)**

**Jan 2025**

- Built an end-to-end data pipeline using Apache Spark on Databricks with data from Amazon S3, EC2 and AWS Lambda, enabling ETL, SQL analytics, and Power BI data visualizations.
- Developed a React-based dashboard that visualized match performance trends using REST APIs and Spark-transformed datasets, improving user interactivity and data accessibility.

**YouTube Data Analysis using Amazon Web Services (AWS Glue, Athena, IAM, SQL, Lambda)**

**Jul 2024**

- Implemented a scalable Data Lake on Amazon S3, orchestrated ETL workflows using AWS Glue and Lambda, and reduced processing time by 30%.
- Enabled secure, insight-driven analytics by executing 500+ SQL queries/month via Athena and improving access control through IAM and CloudWatch monitoring.

## TECHNICAL SKILLS & CERTIFICATIONS

- Programming Skills:** Java, Springboot, Unix Shell Script, C++, SPARK, Python, PL/SQL, MySQL, Kafka, Airflow, React.js, Node.js
- Data Visualization:** MS Excel, PowerBI, Tableau
- Tools & Workflow:** Git, GitHub, Agile methodology, CI/CD pipelines, JIRA, SDLC
- Statistical Modeling:** Numpy, Pandas, Keras, TensorFlow, Pytorch, Feature Engineering, scikit learn
- Machine Learning/AI Tools:** Scikit-Learn, TensorFlow, Keras, Google Analytics 4
- Cloud & DevOps:** AWS, Azure, Lambda, Docker, Jenkins, Kubernetes, Terraform, CloudWatch, CloudFormation
- Certification:** Oracle Database SQL Certified Associate