Thomas K. John

thomaskjohn25@gmail.com | +91-7760988918

thomaskjohn.com | linkedin.com/in/thomaskjohn | github.com/Thomas-K-John

Summary

Experienced ML Engineer with 7+ years of delivering AI/ML solutions across on-premise and cloud environments, spanning e-commerce, manufacturing, and enterprise IT. Proficient in building scalable data pipelines and machine learning models using Python, PySpark, and SQL on platforms like Databricks, AWS Bedrock, and Dataiku DSS. Proven ability to drive business impact through:

- Achieving a 12% profit increase through AI-driven traffic blacklisting
- Delivering cost and performance gains by migrating workloads from SAS to Apache Spark
- Implementing computer vision-based defect detection with 95% accuracy
- Reducing inventory redundancy by 15% using AI-driven MRO optimization

Proficient in designing and deploying ML workloads using a broad range of AWS services, including compute, storage, networking, and ML tools—combined with infrastructure-as-code automation using Terraform.

Holds an AWS-hosted portfolio demonstrating real-world architecture and deployment expertise. Open to global remote roles focused on cloud solution design, migration, and architecture with a strong foundation in ML integration.

Data Science Experience

DATA SCIENCE LEAD - BRILLIO TECHNOLOGIES PVT. LTD., BANGALORE, INDIA
FEB 2025 - MAY 2025 | HYBRID

Designed a feature-rich RAG chatbot leveraging Databricks Mosaic AI, with multilingual support and image input handling to enhance user experience and precision in knowledge retrieval.

Key Skills: Python, Databricks Mosaic AI, Generative AI, Large Language Models (LLMs)

SENIOR DATA SCIENTIST - HEIDELSOFT TECHNOLOGIES PVT. LTD., KOCHI, INDIA
JAN 2022 - DEC 2024 | REMOTE

Client: Solute GMBH - German tech company behind the price comparison platform billiger.de

Project 1. Traffic Partner Blacklisting Project
- Spearheaded a data-driven initiative to enhance online shop performance by improving traffic quality and profitability.
- Designed and implemented an analytics pipeline using Python and PySpark for large-scale traffic and profit analysis.
- Defined performance-based blacklisting criteria, significantly improving return on ad spend (ROAS)
- Ensured quality and transparency through peer reviews, Confluence documentation, and Jira-based workflow.

Project 2. Reporting Tool Evaluation Project
- Led a 3-month strategic initiative to evaluate the optimal reporting and BI tool for future cloud-based integration and analytics modernization.
- Conducted an in-depth analysis of Looker, Power BI, and the existing on-premise cube-based reporting tool.
- Evaluated tools based on criteria such as web-based development, version control, collaboration features, self-service capabilities for business users, and performance.
- Collaborated with cross-functional teams to gather requirements and understand current and future reporting needs.
- Developed and executed test cases to compare performance, usability, and integration capabilities of each tool.

- Presented findings and recommendations to senior management, highlighting the pros and cons of each solution and aligning them with our strategic goals.

Key Skills:
Programming Languages: Python and PySpark (Apache Spark for Python)
Data Analysis: SQL, Hive and Spark SQL
Data Science Platform: Databricks
BI Reporting Tools: Google Looker, Microsoft Power BI, icCube
Optimization: Mixed Integer Linear Programming

DATA SCIENCE MANAGER - CAPGEMINI TECHNOLOGY SERVICES INDIA PVT. LTD., BANGALORE, INDIA
APR 2021 - JAN 2022 | REMOTE

Client: Morgan Stanley (A US based global investment bank and financial services company)
- Led financial data analysis using Python, Hadoop, Hive, and SQL to generate accurate forecasts.
- Utilized Apache Spark for large-scale data processing, delivering timely and actionable insights.
- Designed and maintained data pipelines using Dataiku DSS for efficient data preparation and transformation.
- Migrated legacy data science workflows from SAS to Python, improving scalability and flexibility.
- Trained team members on Apache Spark and Dataiku DSS, fostering knowledge sharing and consistency.
- Applied agile methodologies using Jira to manage project timelines and cross-functional collaboration.
- Presented key analytical findings and financial forecasts via Tableau dashboards to business stakeholders.

Key Skills:
Programming Languages: Python and PySpark (Apache Spark for Python)
Data Analysis: SQL, Hive and Spark SQL
Data Science Platform: Dataiku DSS
Machine Learning and Deep Learning: Supervised and Unsupervised Learning, Financial Forecasting, Model Training & Evaluation, Feature Engineering, Hyperparameter Tuning
Reporting Tool: Tableau

DATA SCIENTIST - INSOFE EDUCATION PVT. LTD., BANGALORE, INDIA
MAY 2017 - APR 2021 | INITIALLY ONSITE, LATER REMOTE

Project 1. Semiconductor Wafer Defect Signature Analysis

- Developed a computer vision model to analyze semiconductor wafer images, accurately detecting defect patterns such as circles, lines, and blobs to assist in quality control.

Project 2. AI-based approach for MRO Optimization and Similar Parts Detection

- Designed an AI-based solution for MRO (Maintenance, Repair, and Operations) optimization, identifying and consolidating duplicate spare parts from an inventory of over one million items, significantly improving inventory efficiency and part approval workflows.

Project 3. Monitoring failures in a Hadoop cluster through log analysis

- Built a resource usage monitoring system for YARN-based applications and developed statistical analysis models using Apache Spark ML.

Key Skills:
Programming Languages: R, Python and PySpark (Apache Spark for Python)
Data Analysis: SQL, Hive and Spark SQL
Data Science Platforms: Databricks, Dataiku DSS
Machine Learning and Deep Learning: Supervised and Unsupervised Learning, Model Training & Evaluation, Feature Engineering, Hyperparameter Tuning, Deep Learning Architectures (CNNs, RNNs, Transformers)
Reporting Tool: Tableau
ML/AI Frameworks & Libraries: OpenCV, TensorFlow, Keras, Scikit-learn

Prior Experience (2009 - 2017)

SENIOR SOFTWARE ENGINEER - XORIANT SOLUTIONS PVT. LTD., BANGALORE, INDIA
SEP 2016 - MAY 2017 | ONSITE

- Developed a network controller to enhance traffic monitoring and analysis efficiency.
- Implemented a rule-based classification algorithm to automatically prioritize ACL rules for traffic redirection.
- Improved monitoring service accuracy and responsiveness, receiving excellent feedback from customers.

Key Skills: Java/J2EE, Linux, Database Replication, SDLC, SQL, Telecommunications NMS

SENIOR SOFTWARE ENGINEER - L&T TECHNOLOGY SERVICES LTD., BANGALORE, INDIA
JUL 2014 - SEP 2016 | ONSITE

Client: ADVA Optical Networking, Bangalore, India
- Developed an RPM-based ENC installer supporting fresh installations, high-availability (HA) setups, upgrades/downgrades, and patch management.
- Migrated build infrastructure from ANT to Maven.
- Automated CI/CD pipelines using Jenkins, streamlining build and deployment processes.
- Implemented key features including feature-based licensing, and auto-restoration of database replication.

Key Skills: Java, Maven, Jenkins, RPM, Linux, Database Replication, SDLC

IT ANALYST - TATA CONSULTANCY SERVICES, KOCHI, INDIA
JAN 2014 - MAY 2014 | ONSITE

- Worked on an application framework for developing hybrid mobile applications across platforms including Android, iOS, and Windows Phone.
- Contributed to cross-platform development efforts, ensuring code reusability and consistent UI/UX across devices.

Key Skills: Java, SDLC, SQL

SOFTWARE ENGINEER - AVAYA INDIA PVT. LTD., BANGALORE, INDIA
JUL 2012 - DEC 2013 | ONSITE

Contributed to the development and support of a Network Management System (NMS) for configuring and managing telecom devices. Worked on backend modules using Java, Hibernate, and JMS, and gained experience in asynchronous messaging, SNMP-based communication, and enterprise system integration.

Key Skills: Java, Hibernate, EJB, JMS, SNMP, SQL, VMware vSphere, SDLC, Telecommunications NMS

PROJECT ENGINEER - WIPRO TECHNOLOGIES PVT. LTD., BANGALORE, INDIA
AUG 2009 - JUL 2012 | ONSITE

Client: Alcatel-Lucent (France-based telecom company)
Developed and maintained large-scale Java-based OSS applications for telecom service provisioning. Gained hands-on experience with system design, API integrations (SOAP/XML), and test-driven development in an enterprise environment.

Key Skills: Core Java, SQL, JUnit, SDLC, Telecommunications NMS, Linux

Education

Bachelor of Technology (B. Tech), Electronics & Communication Engineering
Rajiv Gandhi Institute of Technology (RIT), Kottayam
Affiliated to Mahatma Gandhi University, Kerala
2004 – 2008

Certifications

- AWS Certified Machine Learning Specialty
- AWS Certified Solutions Architect Associate
- AWS Certified SysOps Administrator Associate
- AWS Certified Cloud Practitioner
- HashiCorp Certified: Terraform Associate (003)