PURVA JAVALE

Software Engineer
United States | +1(502) 650-1425 | purvajavale@gmail.com | LinkedIn | Portfolio | GitHub
Work Authorization: H4 EAD – No sponsorship required.

EXPERIENCE

Hexanika | Pune, India *February 2015 – January 2022*
Software Engineer – Big Data
Key Projects and Contributions:

Data Ingestion & Processing Platform

- Designed and built a scalable data ingestion and transformation platform using Apache Spark, Hadoop, and AWS, enabling the automation of regulatory and operational data pipelines across multiple banking use cases.
- Developed data cleaning and mapping components that standardized inputs from heterogeneous sources, improving data accuracy and reducing pre-processing effort by 40%.
- Executed a Proof of Technology on IBM z/OS mainframe, enabling Spark-based data processing on legacy systems. This helped modernize data access for compliance reporting without disrupting the source systems.
- Incorporated data lineage and rule versioning features, ensuring that every transformation was traceable, auditable, and compliant with regulatory standards.

Regulatory Reporting Automation

- Co-developed rule-based engines for financial reports such as FRY9C, Call Report (FFIEC), HMDA-LAR, and AML-SAR, incorporating business and regulatory logic in Spark-based pipelines.
- Integrated machine learning models to predict rule applicability and detect inconsistencies in reporting data, thereby enhancing accuracy and reducing false positives.

Automation, Monitoring & Cloud Deployment

- Incorporated the automation for Hadoop-Spark-based ETL jobs and report generation workflows using custom schedulers, reducing data processing time by 60% while maintaining job reliability and failure recovery.
- Provisioned and maintained multi-node Hadoop clusters and AWS infrastructure (S3, EC2, EMR) for distributed data processing and reporting workloads across multiple environments (dev, UAT, prod).
- Designed and implemented SOAP APIs to expose metadata, lineage, and validation statuses for external and internal monitoring.
- Built real-time dashboards in Power BI, enabling teams to monitor pipeline, data, and Hadoop-Spark job metrics.
- Led Agile delivery cycles, participated in sprint planning, reviewed code, mentored junior engineers, and authored detailed runbooks and deployment documentation to support post-launch stability.

EDUCATION

Bachelor of Engineering in Information Technology *June 2010 - June 2014*
Pune University | Pune, India

SKILLS

Programming Languages: Java, Python, SQL, .NET Core
Big Data: Hadoop, Spark, Hive, HBase, Kafka, Airflow
Formats: TXT, CSV, XML, JSON, AVRO, ORC, PARQUET
Database: MySQL, SQL Server
Cloud & Infrastructure: AWS (S3, EMR, EC2), Databricks, Linux
Legacy: IBM Mainframe (z/OS), WebSphere
Machine Learning: Supervised, Unsupervised, Reinforcement Learning models, Model Evaluation and Tuning
Tools: Team Foundation Server, GitHub, Eclipse, Visual Studio, PyCharm, Power BI, Microsoft Project, Jira, Microsoft Office

CERTIFICATIONS

Big Data Masters Program – TrendyTech (In Progress)
Machine Learning Specialization – Coursera (2025)
Databricks Platform Architect (AWS) – Databricks (2025)
Databricks Fundamentals – Databricks (2025)
Agile Project Management – Google (2024)
AWS Certified Cloud Practitioner – Amazon Web Services (2021)
IBM Mainframe – IBM (2024)

LEADERSHIP & RECOGNITION

- Delivered guest lectures on Big Data at MIT Pune to mentor aspiring data engineers.
- Recognized as "Star Performer of the Year" and "Well-Done Achiever" for leading key innovations.
- Secured 2nd place in Hexanika's FinTech Hackathon for creative data engineering solutions.
- Organized Hexanika's annual offsite, fostering cross-functional collaboration.