

# PURVA JAVALE

## SOFTWARE ENGINEER | BIG DATA | MACHINE LEARNING

United States | +1(502) 650-1425 | purvajavale@gmail.com | [LinkedIn](#) | [Portfolio](#) | [GitHub](#) | [Medium](#)

**Work Authorization:** H4 EAD – No sponsorship required.

### PROFESSIONAL SUMMARY

Software Engineer with 6+ years of experience building large-scale Big Data platforms, ETL pipelines, and audit-ready reporting solutions for U.S. banks and fintechs. Skilled in Java, Spark, Hadoop, AWS, Databricks, and Power BI, with exposure to Machine learning workflows. Specialized in automating regulatory reporting with a strong focus on data lineage, controls, data quality, and compliance accuracy, enabling reliable and scalable enterprise data workflows.

During 2022–2025, focused on upskilling in Big Data, Cloud, Databricks, and Machine Learning, completing certifications and applied projects to stay aligned with industry trends.

### CORE SKILLS

**Big Data:** Hadoop, Spark, Hive, HBase, Kafka, Airflow

**Database:** MySQL, SQL Server

**Formats:** TXT, CSV, XML, JSON, AVRO, ORC, PARQUET

**Machine Learning:** Data Preprocessing, Model-Ready Pipelines

**Visualization:** Power BI, Dashboard Design

**Tools:** Team Foundation Server, GitHub, Eclipse, Visual Studio, PyCharm, Microsoft Project, Jira, Microsoft Office

**Programming:** Java, Python, SQL, .NET Core

**Cloud & DevOps:** AWS (S3, IAM, EMR, EC2), Databricks

**Database:** MySQL, SQL Server

**Legacy:** IBM Mainframe (z/OS), WebSphere

**Methodologies:** Agile, SDLC, Data Governance, Compliance

### PROFESSIONAL EXPERIENCE

#### Software Engineer | Hexanika – Pune, India | Feb 2015 – Jan 2022

- Built and optimized end-to-end ETL pipelines using Java, Hadoop, Spark, and AWS, enabling automated regulatory reporting and reducing manual effort by 40%.
- Improved Spark distributed processing efficiency by 30% through query optimization, storage tuning, and execution-path analysis.
- Modernized legacy compliance workflows by deploying Apache Spark on IBM z/OS mainframe for parallel processing and automated rules execution.
- Built data lineage, rule versioning, and audit-trail capabilities, ensuring full traceability, regulatory compliance, and internal audit readiness.
- Automated ETL scheduling and recovery logic, improving pipeline reliability and reducing processing time by 60%.
- Led agile ceremonies, cross-functional delivery, and mentored junior engineers as the platform scaled from 1 to 50+ employees.

#### Platforms Delivered

- **Smart Join - Regulatory Data Preparation Platform** - Designed data cleaning framework and Spark-based transformation engine, boosting accuracy and reducing preprocessing workload by 40%.
- **Smart Reg - Regulatory Reporting Platform** - Developed Spark SQL rules engine for FRY-9C, Call Reports, HMDA-LAR, and AML-SAR, integrated predictive ML models, maintained Hadoop/Spark clusters and AWS infrastructure, built SOAP APIs and Power BI dashboards.

### EDUCATION

#### Bachelor of Engineering (B.E.) - Information Technology | Pune University, India | 2014

### PROFESSIONAL PROJECTS & CERTIFICATIONS

- **Machine Learning Specialization – Coursera (Stanford University, 2025):** Developed predictive models and ML pipelines for financial datasets using Python and Spark. Built a movie review sentiment classifier with Python, NLTK, and Scikit-learn.
- **Big Data Masters Program – TrendyTech (2025):** Built end-to-end ETL pipelines with Spark, AWS, and Hadoop, implementing data quality and governance frameworks for fintech datasets. Developed ETL pipelines in Databricks for HMDA datasets and created Power BI dashboards for compliance-ready insights.
- **Databricks Platform Architect – AWS (2025):** Designed Big Data workflows in Databricks integrated with AWS and automated batch processing pipelines. Created end-to-end pipelines integrating AWS S3, Databricks, and Power BI for audit tracking.
- **IBM Mainframe – IBM (2024):** Gained expertise in mainframe data handling for banking ETL and legacy integrations.
- **Agile Project Management – Google (2024):** Applied Agile principles to manage multi-phase ETL and data analytics projects.
- **AWS Certified Cloud Practitioner – AWS (2021):** Hands-on experience with AWS S3, Lambda, Glue, and IAM.

### LEADERSHIP & EXTRAS

- First employee at fintech startup, wore multiple hats across engineering, client delivery, and hiring.
- 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.
- Managed extracurricular activities at company offsite events.