Pranav Kapale

pranavkapale11@gmail.com | +91 9130173253 | LinkedIn | Pune

SUMMARY

Software Engineer with 2.8+ years of experience in Python, Scala, Spark and Databricks, specializing in optimizing data pipelines, streamlining ingestion, and automating CICD. Eager to solve data challenges and grow with a dynamic team.

TECHNOLOGY STACK

- **Languages & Data Processing:** Python, SQL, Scala, Apache Spark, Delta Lake
- **DevOps: & Cloud Tools:** Azure DevOps, YAML Pipelines, Git, AWS S3, Docker, Maven, Splunk, JSON
- **Database & Platforms:** SQL Server, PostgreSQL, DBFS, Databricks, Confluence, REST APIs
- **Tooling & Workflow:** IntelliJ, PyCharm, VS Code, Postman, DBeaver, Agile

PROFESSIONAL EXPERIENCE

S&P Global

Software Engineer   ***Aug 2022 - Present***

- **Cost and Performance Optimization:** Improved performance of robust Databricks jobs built using PySpark and Scala by 60% through Delta Lake and Mosaic, leveraging parallelism & multithreading. Reduced cloud compute costs by 40% through workflow refactoring and modular orchestration, and further reduced costs by 30% by optimizing 15+ Spark jobs and migrating to Graviton instances.
- **Data Ingestion Expertise:** Participated in architectural design discussions to identify bottlenecks in ingesting large volumes of data. Collaborated on creating the data ingestion process and developed transformation logic in Databricks, ingesting data into different databases (SQL Server, PostgreSQL). Performed production data ingestion in SQL Server using stored procedures and batch processing, and automated PostgreSQL table transformations and ingestion via JDBC.
- **Mastered Data Delivery**: Partnered with stakeholders to automate the manual, script-based data generation process into a CICD and job-based delivery system. Designed scalable ETL pipelines in Databricks to process GIS spatial data using Mosaic, and identified data quality issues through an investigation-oriented approach using tools like GeoPandas, Mosaic and QGIS.
- **Engineered Azure Pipelines:** Collaborated with DevOps teams to use reusable components and crafted CICD pipelines of Python and Scala projects, including Wheel/Maven deployments. Integrated DR Deployments, AI-Review Pipelines and various automated security scans such as SonarQube, Fortify and Mend.
- **Security & System Enhancement:** Addressed vulnerabilities from security scans to enhance system security and compliance. Participated in technological discussions to identify and implement technical improvements, including integrating REST APIs for direct data downloads and Splunk logging for improved monitoring.
- **Cloud Integration:** Automated workflows for AWS file movement across environment via SharePoint intermediatory, ensuring cross-platform data consistency. Developed AWS Lambda/Step Functions for automated data workflows.
- **Agile Collaboration:** Actively participated in Agile ceremonies collaborated with cross-functional teams to prioritize features, deliver iteratively, and facilitate efficient communication for schema and data updates.

CERTIFICATION

- AI for Everyone
- EssentialTECH Foundations

EDUCATION

Walchand College of Engineering, Sangli	2018-2022	
BTECH, Electronics Engineering	(CGPA: 7.41)	
Dayanand College of Science, Latur	2016-2018	
Higher Secondary Certificate	(Percentage: 79.69)	

SKILLS

Problem Solving, Adaptability, General practices for Code Quality, Teamwork, Performance optimisation, Data quality and Validation

LANGUAGES

English (Professional), Marathi (Native), Hindi (Professional)