Pranav Kapale

pranavkapale11@gmail.com | 9130173253 | LinkedIn | GitHub | Leetcode

SUMMARY

Software Engineer with 3+ years of experience in **Databricks**, **Python**, **Apache Spark**, **Scala** and **CICD**, specializing in processing geospatial data, optimizing data pipelines, streamlining ingestion, and automating CICD.

TECHNOLOGIES

Languages : Python, SQL, Apache Spark, Scala, C++

Libraries : PySpark, Mosaic, GeoPandas, Pandas, Numpy, Shapely

DevOps & Cloud : Azure :- DevOps, Pipelines(YAML); AWS :- S3, EMR, EC2, Lambda, Step Function

Databases & Platforms : SQL Server, Postgres, Databricks

Familiar : Git, REST, JSON, Splunk, Maven, Delta Lake, Geospatial Data, MCP, AI Agents, Docker, Confluence, Agile

EXPERIENCE

Software Developer 2

S&P Global, Ahmedabad Oct 2025 – Present

- Built **H3-indexed geospatial data layers** by selecting optimal resolutions and migrating datasets to Iceberg format, **boosting geometry precision and storage efficiency**
- Enhanced AI Agent capabilities by implementing robust session management in **MCP** based architecture

Software Engineer

S&P Global, Remote Aug 2022 – Sept 2025

- Engineered end to end ingestion pipelines onboarding multiple geospatial datasets (nearly 200 GB dataset each) fetched via **REST APIs**, transformed in Databricks, and loaded into Postgres using JDBC with high reliability and throughput
- **Achieved 60% faster execution** by optimizing Databricks workflows using **Apache Spark**, Mosaic and **Multithreading**
- **Reduced cloud compute costs by 40%** through workflow refactoring and **modular orchestration** strategies
- **Lowered cluster usage costs by 30%** via Spark UI based performance tuning and migration to Graviton instances
- Performed **data ingestion into Production SQL Server** using stored procedures and controlled batch processing while enforcing strict backfill logic and data quality governance
- Crafted **Azure YAML pipelines** for Python and Scala projects, including respective Wheel and Maven deployments. Integrated DR Deployments, AI-Review Pipelines and various automated security scans such as SonarQube, Fortify and Mend
- Addressed critical/high ArmorCode vulnerabilities achieving in **60% reduction on overall security findings**
- Built a **scalable ETL framework** delivering **curated data subsets** to multiple downstream channels with minimal rework, partnering with cross-functional teams to align delivery with stakeholder requirements
- **Automated file transfer workflows** across AWS environments using SharePoint as a data exchange intermediary
- Converted manual data generation scripts into **CICD driven and job-orchestrated pipelines**, reducing manual effort by 90%

PROJECTS

Covid Outbreak Analyzer | *Python, HTML, Flask* GitHub

- Created site which tracks daily statistics of corona cases of only in Indian states, based on number of confirmed cases including foreign tourist corona cases alongside discharged and death case

Online Electronic Component Selection Tool | *Python, HTML, Django* GitHub

- Built a web app comparing product prices across electronics distributors to help users identify the best pricing options

EDUCATION

Walchand College of Engineering, BTECH in Electronics Aug 2018 – Aug 2022

- CGPA: 7.41/10

ACHIEVEMENTS

- **Pubication:-** IoT Based Automated Paralysis Healthcare System Link
- **Certifications:-** AI for Everyone, EssentialTECH Foundations