

# PRASHANT KUMAR MISHRA

## SENIOR DATA ENGINEER



+91 8484093319



prashant\_mishra@outlook.com



Pune MH India

### PROFESSIONAL SUMMARY

- Results-driven Senior Data Engineer with over 13+ years of expertise in the Banking and Finance domain, proven track record in developing robust data pipelines for ML models developed for fraud detection and prevention.
- Recognized for optimizing existing data pipelines, achieving a remarkable 40% cost reduction and a 80% improvement in resource optimization.
- Implemented generic framework which automated data pipeline workflow for different ML models, leading to 70% reduction in manual efforts.

### EXPERIENCE

#### Assistant Vice President

February 2023 - Present | Barclays, Pune, Maharashtra

- Led the migration from legacy SAS application to PySpark application which reduced the execution time of jobs from 10Hrs+ to 2 Hours and enhanced the overall performance by 80%.
- Led a team of Data Engineers and designed complex data pipelines for ML Models designed to prevent credit card related frauds.
- Analyzed and remediated the DQ gaps present in existing data curation process which elevated the performance by 60% in order to get better ML scores.

#### Senior Developer

August 2021 - February 2023 | Barclays, Pune, Maharashtra

- Optimized the pipeline workflow for existing ML models to reduce the execution time by 40%.
- Developed the PySpark framework to automate the existing process which improved the performance in terms of execution and resource usage to 70% and data quality improved to 90%.

#### Manager

January 2019 - October 2021 | Citi, Pune, Maharashtra

- Designed and developed the Ingestion framework which is used to fetch data from multiple sources performs sanity checks along with basic transformations in order to store the data into staging layer and allows downstream applications to pull the data for analytics, it has reduced resource and infra cost by 20%.

### SKILLS

- Python
- SQL
- PySpark
- AWS
- Azure Databricks
- Data Warehouse
- Delta Lake
- Linux Shell Scripting
- HBase
- Hive
- Datalake
- Team Leading
- Presentation
- Stakeholder Management
- Jenkins
- Automation
- Performance Tuning

### LANGUAGES

- English**  
Fluent
- Hindi**  
Native

- Framework has reduced the turnaround time to ingest data to staging and serving layer to 80%.

### Assistant Manager

April 2016 - December 2018 | Citi, Pune, Maharashtra

- 
- Led the re-design and optimization of ETL framework which extracts data from files, oracle DB tables and perform the required transformations and load data to data warehouse for analytics and reduced the batch failures by 50%.
- Developed the shell scripts to automate the testing of Ab-Initio graphs using Jenkins CICD pipeline which reduced infra and storage costs by 40%.

### Associate - Projects

October 2014 - April 2016 | Cognizant, Pune, Maharashtra

- Developed ETL Ab-Initio workflows as per project requirements and supported end-to-end delivery of the tasks till prod deployment and L3 support post Go-Live which reduced the total batch execution time by 30%.

### Senior Systems Engineer

July 2011 - October 2014 | Infosys, Pune, Maharashtra

- Pioneered as ETL Developer (Ab-Initio) and L3 Support for ETL batches which loads data to Oracle DB. The data was used to generate reports for business use cases.
- Automated the process in order to send notification email in case of job or batch failures resulting to 70% improvement in manual process.

## EDUCATION

### Master Of Computer Applications

July 2011

Birla Institute Of Technology, Mesra, Ranchi

## CERTIFICATIONS

- AWS Certified AI practitioner
- AWS Certified Solutions Architect – Associate
- Databricks Certified Associate Developer for Apache Spark 3.0
- Databricks Lakehouse Fundamentals

## AWARDS

- Awarded for designing, resource optimization and benchmarking of generic framework for automating data curation and data quality report workflows. It has reduced 90% of manual efforts in curating datasets and preparing DQ reports on monthly basis.
- Awarded for designing and developing generic data ingestion framework which helped reducing 60% operational and 30% data storage costs across projects.

