

Shraddha Mane

Data Engineer

+91 8788573233

shraddhamane34@gmail.com

Pune

[in](#) [LinkedIn](#)

SUMMARY

I am Data Engineer with 4.5 years of experience in Big Data and Analytics. Extensively worked on PySpark, Hive, Linux for optimized performance, also have good knowledge of implementing the scripts for the data ingestion. I am experienced in coordinating with client and on-site team. Looking forward to utilizing my skills as Data Engineer to develop and enhance my skill sets.

KEY SKILLS

Python | PySpark | Spark | Control- M tool | UNIX | LINUX | SPLUNK tool | Data Engineering | Big Data - Hadoop, Hive | Exploratory Data Analysis | Data Cleaning | Data pre-processing | Agile | V model | Waterfall model | Putty | WinSCP

TECHNICAL SKILLS

- **Languages:** Python, SQL, HQL,
- **Operating Systems:** Windows, Linux
- **Analytical skills:** Data Interpretation, Cleaning. Pre-processing, Visualization.
- **Tools:** Hive, Spark, Jupyter Notebook, Control-M, VS code, Git, Ms-Excel, Tableau, MySQL, Putty, WinSCP, JIRA, Confluence,

EDUCATION

B.E. in Information Technology

Aug '16 - Mar '20

D Y Patil collage of Engineering

Pune

- CGPA: 8.26/10.0

CERTIFICATIONS

- Google Associate Cloud Engineer Certification
- Automation Engineer Practitioner Certification

PROFESSIONAL EXPERIENCE

Data Engineer - Barclays

Jan '24 - Present

Project: People Analytics and Culture

Pune

Project Description:

- Empowered the global team with a remit to deliver HR reporting & analytics to HR & the business, driving **data-led decisions** impacting our 100k colleagues worldwide.
- Pioneered the development of a machine learning & predictive analytics capability to enhance an already mature HR reporting & dashboarding service, driving improved decision-making processes & elevating HR service quality leveraging **Python, Pyspark, & Hive**.

Responsibilities:

- Integrated and standardized diverse survey datasets, ensuring seamless onboarding into the data ecosystem and facilitating consistent data analysis across projects.
- Executed reusable & modular codebases utilizing Python, PySpark, & Hive to **optimize data processing tasks**, thereby enhancing efficiency & maintainability in data optimization projects.
- Enhanced text datasets through **advanced natural language processing techniques**, optimizing, tokenizing, & extracting key insights efficiently to reveal valuable patterns.
- Streamlined end-to-end data pipelines automation, ensuring reliable data flow from ingestion to storage, which **elevated data reliability for analytics**.
- Led project tasks & progress management through Jira, **oversaw version control using Bitbucket**, & curated comprehensive documentation in Confluence for team knowledge sharing support.
- Facilitated cluster migration activities, overseeing the **seamless transition of 10 TB of data & processes to new computing environments**, ensuring minimal disruption throughout migrations.
- **Participated in agile ceremonies**, including retrospectives and daily stand-ups, contributing to the development of team strategies and fostering continuous improvement.

Project Description:

- Delivered consistent & quality Global Compliance Data crucial for managing compliance risk at HSBC & enhancing the FCC function through a **sustainable & efficient operating model**.
- Implemented a scalable data storage solution for multiple source system data to enhance reporting analytics, facilitating **extraction, validation, & data formatting** for improved reporting efficiency.

Responsibilities:

- Designed, tested, & deployed scripts to Production as per client specifications, resulting in **boosted client satisfaction scores**, leveraging Python.
 - Conducted development & testing of HQL scripts, analyzed data using **Python & Hive** to drive increased efficiency & enhance decision-making.
 - Implemented the automation of production jobs with **Control-M Scheduling tool and monitor batches with help of Putty/WinSCP**, resulting in an 80% reduction in manual intervention.
 - Upgraded applications on the Cloudera environment to analyze outputs, **data insights & performance metric enhancements**.
 - Resolved incidents & user issues, managed Jiras, prevented Production errors, **data integrity** & maximizing resource efficiency.
 - Oversaw the deployment of applications onto production using **Jenkins deployment tool** to streamline production deployments, integrating technologies like Ansible for optimal efficiency.
 - Maintained configuration stats for each cluster on a daily basis & organized housekeeping activities as needed using **Python & Excel** for efficient data management.
 - Efficiently copied data from various production clusters based on requirements using **DISTCP & SCP tools** to enhance data accessibility.
 - Orchestrated the establishment of Ranger Policies for HDFS/Hive aligning with security requirements, fortifying data protection measures, & **optimizing data access control** using Apache Ranger.
 - Collaborated with development teams to conduct thorough testing of enhancements & fixes in test environments, ensuring **successful outcomes** before deploying to production upon approval & completion of signoff using GitLab & JIRA for a streamlined process.
-