

# SHRADDHA WAPHARE | DATA ENGINEER

Redmond, WA, USA 98052 • (716) 325-2163 • [shraddha.waphare30@gmail.com](mailto:shraddha.waphare30@gmail.com) • [LinkedIn](#) • [Tableau](#)

## SUMMARY

Experienced Data Engineer with over 4 years in designing and deploying large-scale data processing systems, handling over a million records daily on AWS cloud.

Proficient in Agile/Scrum methodologies & database/SQL for efficient project management.

Expertise in Python and PySpark for data processing, analysis, and visualization. Skilled in ETL development, data modeling, and data warehousing with PySpark, PySpark SQL, and PySpark Data Frame to architect scalable data solutions. Experienced in cloud storage using Amazon S3 and Azure Blob Storage, as well as in-memory data processing with Apache Spark and Apache Ignite. Deep understanding of Hadoop ecosystem tools such as HDFS, MapReduce, Spark, Yarn, and Sqoop, used to design and develop Big Data Analytics solutions. Strong skills in SQL programming, including developing Stored Procedures, Triggers, Functions, and Packages using SQL and PL/SQL.

A keen eye for detail and exceptional communication skills. Quick learner who thrives in fast-paced environments, leveraging technical expertise to deliver high-quality results under tight deadlines.

## SKILLS

Programming languages:	Python, R, Shell Scripting, SQL, and PL/SQL
Data processing tools:	Hadoop, HDFS, MapReduce, Hive, Spark, Oozie, Sqoop, Flume, Pig
Libraries:	Pandas, Matplotlib, Seaborn, SciKit-Learn, and NumPy
Data & Analytics Techniques:	Document analysis, Data Profiling, Data Integration, Data Modeling, Data Mining, Data Visualization, Market analysis, GAP analysis, SWOT analysis, MOST analysis, Cost-benefit analysis, Feasibility Analysis, Big Data Analytics, Data Analysis, Data Migration
Databases & Cloud:	HBase, MySQL, SQL Server, PostgreSQL, Teradata, AWS (S3, Ec2, lambda, Apache airflow, Cloud Watch, IAM, Cloud Formation), and Azure
Data visualization and reporting:	Tableau, PowerBI, SSRS, SSMS, SSIS, Spark, Informatica, and MS Office suite
Other Tools:	Jenkins, Maven, Docker, CI/CD, Git/Bitbucket, Unix/Linux, MapR, EMR, and ETL
Environment:	Agile, Scrum, Waterfall, Windows, MacOS, and Linux/Unix

## PROFESSIONAL EXPERIENCE

**DELL, USA; Data Engineer** Sep 2023 – Present

- Enhanced model evaluation efficiency and accuracy using Python and R, leading to more precise data analysis.
- Conducted statistical data analysis and visualization using Python and Pandas, providing actionable insights for business decisions.
- Developed containment scripts for data reconciliation using SQL and Python, improving data accuracy by 30%.
- Presented dashboards to higher management using Tableau, leading to data-driven decision-making.
- Utilized MySQL and MS Excel to identify trends, resulting in more informed business strategies.
- Prepared and presented comprehensive BRD, SRS, and FRD, ensuring clear communication of technical requirements.
- Converted charts into Crosstabs in MS Excel for deeper data analysis.
- Developed and executed action plans, implemented control panel structures, conducted audits, and provided technical solutions, enhancing operational efficiency by 25%.

**Bitwise Solutions Pvt. Ltd., India; Programmer/Data Engineer** Jun 2019 – Oct 2021

- Managed data across 9 environments, each containing over 100TB of financial data, ensuring data integrity using Informatica.
- Achieved 100GB reduction of table space by compressing tables with no data loss using multivalue compression in shell scripting.
- Demonstrated expertise in SQL performance tuning to improve DB efficiency, significantly reducing execution time by 20%.
- Played pivotal report in migration project from Teradata to Snowflake and achieved 100% uptime during the migration period.
- Implemented an automated Spark ETL process for processing 1M of data, resulting in a 29% reduction in manual workload.
- Extracted bureau, loan performance, and behavioral attributes, streamlining the data preprocessing pipeline and achieving a 20% reduction in processing time and conducted comprehensive data cleaning and variable reduction using SQL and Python.
- Scalability achieved through Apache Spark led to a 50% increase in data processing capacity, accommodating larger datasets.
- Implemented Apache Airflow to automate & streamline complex data workflows, resulting in 40% improvement in operational efficiency through enhanced scalability.
- Automated 15% of total reports using Tableau and SQL which are crucial for decision making in higher management.
- Worked closely with senior manager for migration report, ensuring that all reporting was relevant, accurate, and actionable.
- Conducted data analysis to identify trends, patterns, and insights, providing valuable information to support decision-making.
- Developed Python-based anomaly detection algorithms, reducing false positives by 15% and improving system health.
- Elevated Tableau dashboards with interactive filters, parameters, and dynamic visual elements, resulting in a 30% increase in user engagement emphasizing specific metrics of interest, and contributing to a 25% improvement in decision-making

**CueTech Systems, India; Data Engineer** May 2018 – Jun 2019

- Improved model evaluation efficiency using Python and R, resulting in more accurate analyses.
- Worked with AWS cloud (S3, EC2, Lambda, Apache Airflow, CloudWatch, IAM, FSX, CloudFormation) to implement scalable data solutions.
- Analyzed business requirements and composed functional technical data solutions.
- Identified integration impacts, data flows, and data stewardship responsibilities.
- Generated and executed SQL scripts for DB changes, including table updates and index creation.
- Reviewed and revised data models for structural soundness and adherence to standards.
- Performed ETL using Informatica Power Center, defining ETL mapping specifications and designing processes.
- Created complex SQL queries to validate data accuracy and translated business requirements into clear specifications.
- Conducted data analysis to identify patterns, enhancing data preparation processes.
- Developed MySQL queries and views for Tableau, facilitating comprehensive data visualization.

## EDUCATION

**Master of Science in Data Science** | Jan 2022 - Jun 2023  
University at Buffalo, SUNY, USA

**Bachelor of Engineering in Computer Engineer** | Jul 2015 - May 2019  
PCCOER; Pune University, Pune, India