

Rushikesh S. Ware

(571) 435-1707 | rushikeshsware@gmail.com | linkedin.com/in/rushikesh-ware10 | github.com/Rushikesh-S-Ware | Fairfax, VA

SUMMARY

Database Analyst with 2+ years of experience in **ETL pipeline development**, **data mapping**, **data modeling**, and **data quality management**. Proficient in **SQL**, **Python**, and **shell scripting** with hands-on experience profiling complex data sources, enforcing **schema validation**, and building source-to-target mappings across structured and semi-structured datasets. Strong attention to data integrity, metadata documentation, and communicating technical findings to cross-functional teams.

EDUCATION

George Mason University Fairfax, VA
Master's in Data Analytics Engineering (M.S.) — GPA: 3.90 Dec 2025
• Coursework: **Big Data**, **Machine Learning**, **Predictive Analytics**, NLP with Deep Learning, Data Engineering, Database Design

Savitribai Phule Pune University Pune, India
B.E. in Mechanical Engineering — GPA: 3.52 May 2021

EXPERIENCE

Programmer Analyst / Data Analyst Pune, India
Cognizant Technology Solutions Feb 2022 – Nov 2023

- Designed and optimized **SQL queries**, stored procedures, and **ETL workflows** in BigQuery on **GCP**, improving data delivery time by **25%** and reducing compute cost by **18%**.
- Performed **data profiling**, source-to-target **data mapping**, and pre/post-load **data quality checks** to validate accuracy and completeness across reporting pipelines.
- Implemented **data validation rules** including constraint enforcement, null handling, and referential integrity checks to maintain data quality across systems.
- Documented **data definitions**, transformation logic, and **metadata** standards to support governance, auditing, and team knowledge sharing.
- Automated deployment and testing workflows using **Docker**, **Airflow**, and **CI/CD** pipelines, reducing deployment errors by **20%**.

Data Engineer Fairfax, VA
Climate Data Integration Platform (Academic + Deployed) Aug 2025 – Dec 2025

- Profiled and mapped multi-source datasets (APIs, relational databases, CSV files) into a unified **relational schema** supporting 500K+ records across 165 countries.
- Performed **source-to-target data mapping** and built **fact and dimension data models** reducing query latency by **70%**.
- Enforced **data quality** rules across 4 validation dimensions — schema match, value ranges, unit consistency, and null handling — improving pipeline reliability by **85%**.
- Authored **data dictionaries**, pipeline documentation, and **data flow diagrams** to support governance and maintainability.
- Containerized pipelines using **Docker** and deployed via **CI/CD** with **GitHub Actions** for consistent, reproducible execution.

PROJECTS

Public Health Analytics | *Python, SQL, AWS Glue, PostgreSQL, Tableau* Jan 2024 – May 2024

- Profiled and cleaned **20 years** of public health records using **AWS Glue DataBrew**; performed **data mapping** and transformation across multiple source schemas into a unified analytical model.
- Built **SQL-driven** analytical queries and interactive **Tableau** dashboards surfacing demographic trends for research reporting.

Cybersecurity Risk Modeling | *Python, SQL, XGBoost, REST API, PostgreSQL* Jan 2025 – May 2025

- Ingested and mapped heterogeneous **CVE/CVSS/EPSS/CPE** data from multiple sources into a normalized **relational schema**; applied **data harmonization** to resolve inconsistencies across datasets.
- Built automated **test scripts** to validate pre- and post-transformation data integrity; deployed results via **REST API**.

Infrastructure Analytics & Optimization | *Python, SQL, XGBoost, Random Forest* Aug 2024 – Dec 2024

- Modeled structural condition data for **7,500+ assets**; applied Chi-Square, ANOVA, and NPV analysis to identify data relationships and support maintenance decision-making.

TECHNICAL SKILLS

- Database & SQL:** PostgreSQL, MySQL, BigQuery, T-SQL, PL/SQL, stored procedures, triggers, views, indexing, query optimization, schema design, data modeling (star, snowflake, normalized)
- Data Engineering & ETL:** Data profiling, source-to-target mapping, ETL/ELT, batch and incremental loads, CDC, data quality validation, reconciliation, Airflow, Spark, dbt
- Programming & Scripting:** Python (pandas, NumPy, SQLAlchemy, PySpark), Bash, R, Java, regex, REST APIs, pytest
- Cloud Platforms:** GCP (BigQuery, GCS, Dataflow), AWS (S3, EC2, Glue, Redshift, Athena), Snowflake, Databricks, Azure
- DevOps & Orchestration:** Docker, Kubernetes, Git, GitHub Actions, CI/CD, Terraform, Jenkins, UNIX/Linux, shell scripting
- Data Formats & Storage:** JSON, CSV, Parquet, Avro, XML, YAML, columnar vs row stores, partitioning, compression
- Visualization & BI:** Power BI, Tableau, Looker, Streamlit, matplotlib, seaborn, Plotly
- Governance & Workflow:** Data dictionaries, data flow diagrams, metadata standards, lineage, runbooks, Agile, Scrum, Jira, Confluence