

Shweta Ajay Shinde ✉ shwetashindeus@gmail.com | shweta.shinde@sjsu.edu | +16695775590 | Bay area, CA, 94086

 [LinkedIn - Shweta Shinde](#)  [GitHub – Shinde Shweta](#)  [Portfolio-Shweta Shinde](#)

ETL Data Analyst with 7 years of expertise in developing and optimizing ETL applications across all SDLC phases. Proven ability to resolve database issues, ensure client satisfaction, and implement efficient data strategies for timely solutions.

EDUCATION:

Master of Science in Data Analytics. (Current Student)

May 2026

San Jose State University, San Jose, CA.

Courses: Mathematical and Statistical Methods, Big Data Technologies, Machine Learning Technologies, Business Intelligence, Data Visualization, Deep Learning, Data Warehouse and Pipeline, Distributed Systems for Data Engineering, Generative Model Applications.

Master of Science in Computer Science, Pune University, India. GPA: 3.17

Jun 2018

Bachelor of Science in Computer Science, Pune University, India. GPA: 3.04

Jun 2016

PROFESSIONAL EXPERIENCE:

Cloud Engineer: NTT Data, Pune, India, Client: EXXON MOBIL (Oil and Fuel Domain)

Oct 2021-Apr 2023

- Gathered and documented cross-functional business requirements, creating Azure pipelines, SQL objects, and algorithms to manage data flow from Business Source through Azure Data Factory to Snowflake.
- Collaborated with global teams to extract and load operational data into RAW and Curated layers, resolving data issues and troubleshooting ETL pipelines to ensure 100% client satisfaction.
- Utilized Agile methodology for project management, Git for version control, and implemented end-to-end data strategies with computer vision to enhance data transformation and compliance.
- **Technologies:** Snowflake, Azure Data Factory, MySQL, Python (Advance), GIT, GitHub, Tableau, Visual Studio.

Programmer Analyst: Cognizant, Pune, India, Client: “DWBI ETL”- Merck (Life Science Domain)

May 2016-Oct 2021

- Developed ETL applications from requirement gathering through SDLC, improving data flow from source to target.
- Reviewed and analyzed functional requirements and mapping documents, assisting in impact analysis for new or changed ETL applications.
- Formulated and loaded client’s operational data into the Application Staging and ODS Layer while implementing specified business requirements, demonstrating a results-oriented approach to meet client needs effectively.
- Provided timely root cause resolutions to users via calls, discussions and effective communication.
- Processed COVID-19 vaccine trial data, delivering comprehensive reports used resulting in the production of resourceful reports used in COVID testing.
- Worked onsite in Argentina (2019-2021), demonstrating adaptability to diverse environments
- **Technologies:** Linux, Shell Scripting, Putty, Informatica Power Center (9.6/1Data 0.2), Informatica Intelligent Cloud Services (IICS), MySQL, Oracle(10g), Teradata SQL Assistant 15.1, PowerBI, Remedy, Autosys, Crontab Scheduling.

SKILLS:

Programming Skills:	C, C++, Linux, Shell Scripting, Python (Advanced), HTML
Data Processing:	Data Cleaning, Feature Engineering, ETL Pipeline, Data Wrangling
Cloud Tools:	Informatica Intelligent Cloud Services (IICS), Azure Data Factory, Apache Airflow, Docker
Database:	MySQL, Oracle(10g), Teradata SQL Assistant 15.1, Snowflake
Data Visualization:	Tableau, PowerBI, Excel, Matplotlib, Pandas, PandasAI, Seaborn
Tools:	Git, Remedy, Informatica, Putty, Microsoft Office 365, Visual Studio, Jupyter Notebook
Libraries:	NumPy, Turtle, SciPy Scikit-learn, SymPy, Pytorch etc.
Scheduling Tools:	Autosys, Crontab Scheduling

PROJECTS

Data Visualization Project: Paris Olympics Dataset

Aug 2024-Present

- Developing interactive visualizations using Tableau and Power BI to analyze a comprehensive dataset related to the Paris Olympics, uncovering trends to inform urban planning and community development while ensuring user-friendly narratives for stakeholders.

Data Warehousing Project: Hourly Weather Forecast

Aug 2024-Present

- Creating an ETL pipeline in Apache Airflow to transfer data from source using API keys into Snowflake, utilizing SQL and Python for data processing and transformation to ensure efficient and reliable data flow. The pipeline is containerized with Docker, ensuring consistent and scalable deployment across different environments.