

# PRATHAMESH WALI

+1(857) 313-2184 | [wali.prathamesh@gmail.com](mailto:wali.prathamesh@gmail.com) | [LinkedIn/prathameshwali](https://www.linkedin.com/in/prathameshwali) | [GitHub/prathameshwali](https://github.com/prathameshwali)

## PROFESSIONAL EXPERIENCE

**Data Engineer — Spare-it Inc** **Boston, USA** — Jan 2024 - Present  
**Technology Stack:** Spark, ETL, AWS (Glue, DynamoDB, Lambda, EMR, Kinesis, CloudFormation(IaaC), S3, QuickSight)

- Designed and built robust and scalable ETL pipelines processing over 5TB of big data from unstructured non-relational databases per day from IoT sensor devices, achieving a 40% reduction in processing time and enabling real-time analysis of streaming data, resulting in actionable insights delivered to stakeholders within seconds.
- Demonstrated expertise in optimizing resource consumption, notably reducing DPU (Data Processing Unit) hours from 0.9 to a highly efficient 0.3. Established advanced strategies to boost glue jobs efficiency, including the addition of pushdown predicates for time series big data, resulting in substantial performance enhancements and automated testing
- Pioneered the adoption of an incremental approach by implementing Lambda functions for live streaming big data from DynamoDB, leveraged innovative techniques to ensure real-time data integration which led to a 30% reduction in data latency, enabling more timely insights and decision-making for customers

**Associate Software Engineer — LogicMonitor Inc** **Pune, India** — Oct 2020 - Sep 2022  
**Technology Stack:** Python, Redshift, Java, SOA, Flask, Spring Boot, RestAPI, Docker, bash, Kafka, Looker

- Fine-tuned python and SQL queries to transform raw monitoring data into actionable insights for the product, supporting the enhancement of the RCA feature and contributing to the overall increase in customer satisfaction by 30%
- Launched a cloud monitoring feature for AWS, resulting in an 18% increase in monitoring efficiency, optimizing SaaS resource utilization and enhancing overall system performance in a agile scrum setting
- Enhanced the AI Ops topology mapping feature and enabled the RCA feature for Virtual Machine devices in distributed system and security which lead to an increase in sales by around 29%.

**Data Engineer — Ameya Industries** **Pune, India** — Jul 2018 - Sep 2020  
**Technology Stack:** SQL, Azure, Databricks, DataLake, Synapse, Snowflake, Power BI

- Constructed and integrated Azure Data Factory and Databricks pipelines for multiple APIs, increasing efficiency by 15% and facilitating faster decision-making processes
- Developed Snowpark stored procedures for MDM, reducing data duplication by 40% and ensuring seamless data integration across systems using SQL and Azure DataLake.
- Optimized Snowflake solutions, enhancing data processing efficiency by 35% through scalable warehouses and optimized SQL queries.

## LEADERSHIP EXPERIENCE

- Mentored 150+ students, as a Lead Teaching Assistant and improved comprehension, assignment scores, and code quality for Data Engineering, Data Science and Machine Learning projects through tailored support, assignments, and code reviews.

## SKILLS

Programming Languages	: Spark, SQL, Python, R, Java, DAX, JavaScript
Databases	: MySQL, PostgreSQL, MS SQL Server, NoSQL, Oracle
Machine Learning	: NumPy, Pandas, Matplotlib, SciKit-Learn, Tensorflow, Pytorch
Data Management Tools	: Airflow,Snowflake, dbt, Talend, Alteryx, Power BI, Tableau, Excel
Cloud Technologies	: AWS, Azure, GCP
DevOps & Other	: Docker, Kubernetes, Terraform, Kafka, REST APIs, Hive, Streamlit, FastAPI, Flask

## EDUCATION

Northeastern University, Master's of Information Systems   <b>Boston, USA</b>	<b>GPA:</b> 4.00/4.00
Pune University, Bachelor's of Engineering in Computer Science   <b>Pune, India</b>	<b>GPA:</b> 3.75/4.00

## PROJECT EXPERIENCE

**Flight Fare Predictor — Data Science Hackathon** **Boston, USA** — Jul 2023 - Sep 2023  
**Technology Stack:** AWS (SageMaker, EC2, IAM, IaC), Machine Learning, Python, Heroku, Flask

- Forecasted flight fares by conducting EDA, data cleaning, feature engineering, and selection, developed a 92% accurate predictive model with Random Forest and deployed a Flask app on Heroku and integrated with GitHub for CI/CD

**IMDb ETL Automation — Northeastern University** **Boston, USA** — Oct 2023 - Dec 2023  
**Technology Stack:** Airflow, dbt, Snowflake, GCP(Big Query, DataFlow), Tableau

- Automated scalable Data Warehouse for IMDB data using Kimball's methodology, orchestrated ETL pipelines across multiple databases and file formats, and created BI Dashboards in Tableau for top 1000 movies, TV shows, and production entities