

# Namita Chougule

Ph no: +1-571-564-7751 / namitac9817@gmail.com / https://www.linkedin.com/in/namitachougule

## SUMMARY

AWS Certified Data Engineer with 3+ years of experience designing and optimizing scalable data pipelines, warehouses, and real-time processing systems. Skilled in SQL, ETL, Python, Spark, Databricks, and AWS, with deep expertise in Snowflake, Talend, and relational databases (MS SQL Server, Oracle) as well as NoSQL systems. Proven track record of delivering end-to-end pipelines from ingestion to analytics while improving performance by 30%+ and enabling data-driven decision making across finance, insurance, and AI-driven applications.

## EDUCATION

|   |                     |
|---|---------------------|
| <b>Master of Science in Computer Science</b><br>George Mason University   GPA: 3.87/4.00  | Jan 2024 - May 2025 |
| <b>Bachelor of Engineering in Computer Science &amp; Engineering</b><br>Visvesvaraya Technological University   GPA: 8.18/10.00 | Aug 2016 - Aug 2020 |

## EXPERIENCE

|   |                     |
|---|---------------------|
| <b>Applied AI &amp; Data Developer</b> / onAir Tech Networks  | July 2025 - Present |
| - Engineered scalable <b>Python data pipelines</b> integrating <b>Hugging Face LLMs</b> (Pegasus, BART-MNLI) to ingest and transform <b>RSS feed</b> content in <b>MongoDB</b> .  |                     |
| - Automated <b>WordPress</b> content distribution through pipeline integration, reducing manual effort by <b>80%</b> .  |                     |
| - Developed <b>RESTful APIs</b> to integrate summarization and classification outputs into WordPress hubs, supporting live production usage with minimal downtime.  |                     |
| - Provisioned and optimized <b>AWS infrastructure</b> (EC2, S3, IAM, CloudWatch) to scale AI data pipelines, improving deployment reliability and efficiency by <b>30%</b> .  |                     |
| - Exploring <b>Airflow-based agentic AI</b> workflows to modularize data ingestion and transformation for scalable orchestration.   |                     |
| <b>Graduate Teaching Assistant / Grader – Database Systems</b>   George Mason University  | Jan 2025 - May 2025 |
| - Assisted instruction for a graduate-level course on Database Systems, covering relational schema design, dimensional modeling, normalization, indexing, DDL, DML, SQL, NoSQL, and transaction management.                       |                     |
| <b>Data Engineer: ETL Developer</b> / NTT DATA Services   | Jan 2021 - Dec 2023 |
| <b>Client: Tokio Marine Kiln</b>  |                     |
| - Led migration of <b>60+ TB</b> of insurance/finance data from on-prem <b>MS SQL Server</b> to <b>Snowflake</b> and shifted ETL from <b>SAP BODS</b> to <b>Talend</b> , improving scalability and reducing costs by <b>40%</b> . |                     |
| - Built and optimized <b>ETL/ELT</b> pipelines processing <b>1M+</b> daily records from AWS S3 (Parquet, Avro, CSV, JSON) and relational databases for curated downstream analytics.  |                     |
| - Built <b>event-driven</b> pipelines with <b>MSK</b> and <b>Spark Streaming</b> to process <b>15K+</b> transactions/min for near real-time analytics.  |                     |
| - Re-engineered workflows to reduce batch runtime by <b>50% (12 hrs to 7 hrs)</b> and enable faster reporting cycles.   |                     |
| - Monitored <b>1,000+</b> Snowflake and Talend workflows daily, resolving failures and bottlenecks within SLA timelines.  |                     |
| - Developed <b>DBT models</b> on Snowflake for transformations and automated quality checks.  |                     |
| - Integrated <b>Talend</b> with <b>AWS IAM, S3, and Snowflake</b> for secure and auditable data flows.  |                     |
| - Collaborated with stakeholders and analysts in an <b>Agile</b> environment to deliver actionable data solutions.  |                     |
| - Implemented <b>data governance</b> and <b>quality validation checks</b> , ensuring reliable and compliant reporting datasets.   |                     |
| - Created <b>Power BI dashboards</b> to visualize data insights, enabling senior management decisions.  |                     |
| - <b>Mentored</b> a team of 2 junior engineers by guiding best practices, conducting code reviews and troubleshooting issues.   |                     |

## CERTIFICATIONS

AWS Certified Data Engineer – Associate | Agile & Scrum (NLCI) | Oracle Data Engineer (NLCI) | Data Warehousing (NLCI)

## TECHNICAL SKILLS

**Programming & Development:** Python, SQL, Java, Javascript C/C++, Web Development, Spring Boot, React, Hugging Face.  
**Data Engineering:** Talend, ETL/ELT, SSIS, SSRS, DataBricks, SAP BODS, DBT, AWS Glue, Spark, Kafka, Athena, Airflow.  
**Databases & Warehousing:** Snowflake, MSSQL, Oracle, Postgres, MongoDB, Redshift, DynamoDB, AWS RDS.  
**DevOps & CI/CD:** Docker, Kubernetes, Rancher, Jenkins, GitHub, JIRA.  
**Cloud & Visualization:** AWS, GCP, Azure, Power BI, Quicksight.

## PROJECTS

|  |
|--|
| <b>Real-Time Data Pipeline for Smart City Data</b>   Kafka, Spark, AWS Redshift, Quicksight  |
| - Built an event-driven Kafka + Spark pipeline to process IoT sensor data, enabling real-time analytics in AWS Redshift and visualized insights via Quicksight dashboards. |
| <b>AI-Powered Hybrid Chatbot</b>   Python, React, GPT, Pinecone, RAG   |
| - Developed an LLM-based RAG chatbot with semantic memory and web search fallback, enhancing contextual awareness for user interactions.                                   |
| <b>Interactive Sales Performance Explorer</b>   MySQL, Power BI, DAX   |
| - Built a star-schema data model and Power BI dashboard analyzing 100K+ sales records to identify top SKUs, seasonal patterns, and stock recommendations.                  |
| <b>Cloud-Native Survey Application with CI/CD Deployment</b>   Spring Boot, Docker, Kubernetes, Jenkins, AWS RDS   |
| - Deployed a Spring Boot-based survey application with a CI/CD pipeline on AWS using Jenkins, Docker, and Kubernetes integrated with an Oracle RDS backend.                |