Shardul Chavan
AWS Certified - Data Engineer Associate

+18573135138 | shardulc36@gmail.com | LinkedIn | GitHub

EDUCATION

Northeastern University - *Master of Science, Information Systems, **GPA:** 3.76/4.00*   **Sep 2022 — Dec 2024**

University of Mumbai - *Bachelor of Engineering, Computer Science and Engineering*   **Aug 2019 — May 2022**

WORK EXPERIENCE

Crewasis, Remote - Data Scientist Intern   **Jan 2025 — Jun 2025**

- **ETL Workflow Automation:** Designed and deployed scalable ETL pipelines processing 2M+ unstructured text records from social/news feeds, enabling downstream NLP, sentiment analysis, and actionable product-market insights.
- **Cloud ML Orchestration:** Built EC2-hosted machine learning pipeline orchestrated with AWS Lambda, automating Latent Semantic Indexing (LSI) training and reducing manual retraining cycles by 70%.
- **Infrastructure-as-Code (IaC):** Utilized Terraform scripts to deploy repeatable AWS infrastructure, ensuring high availability and reproducibility.
- **Interactive Reporting & Analytics:** Developed interactive Tableau dashboards combining NLP insights with CRM segmentation, empowering marketing teams to make data-driven decisions and increasing campaign conversions by 15%.

Skyworks Solutions Inc., Boston, MA - Data Analytics Engineer Co-op   **Jan 2024 — Jun 2024**

- **Airflow Pipeline Development:** Engineered 25+ Airflow and dbt pipelines across Azure SQL and data lakes, reducing data refresh latency by 60% and boosting accessibility for analytics teams.
- **Data Product Optimization:** Standardized RF testing data into analytics-ready formats, saving 5+ engineering hours weekly and accelerating high-volume RF module validation.
- **System Resilience Engineering:** Containerized data services with Docker and set up Prometheus monitoring to track performance and failures, reducing downtime by 45% through retry logic and better visibility.
- **Ensured data governance** by integrating validation with Great Expectations and enforcing RBAC + row-level security policies, ensuring compliance and secure data access.
- **Data Modeling & Reporting:** Designed scalable dimensional data models and built interactive Power BI dashboards for 3 business units, accelerating monthly reporting cycles and enabling cross-team insights.

Northeastern University, Boston, MA - BI Analyst & AI Researcher   **Sep 2023 — Dec 2024**

- **Big Data Pipeline Engineering:** Developed scalable PySpark pipelines on Hadoop to process 1.2M+ insurance records, enabling efficient feature
- engineering for risk modeling and improving predictive accuracy by 20%.
- **AI Solution Development:** Mentored 50+ graduate students in building applied AI/BI applications, integrating LLMs (OpenAI, Anthropic) into enterprise analytics use cases and accelerating delivery of intelligent insights.
- **Product-facing BI automation:** Built SnowGPT a LangChain-powered tool on Snowflake, enabling natural language-to-SQL translation driving 40% increase in self-service analytics adoption among non-technical stakeholders.
- **CI/CD for ML Workflows:** Deployed containerized ML models with Docker and GitHub Actions CI/CD, reducing manual deployment time by 60% and ensuring reproducibility across staging and production

Accion Labs, Mumbai, India - Data Analytics Engineer   **Jan 2022 — Jul 2022**

- **LLM Integration & Analytics Automation:** Engineered REST-based integrations between LLM APIs and ServiceNow, enhancing conversational AI for support bots, resulting in a 40% reduction in ticket escalations and improved NLU accuracy.
- **SQL Automation & Cross-Team Collaboration:** Built parameterized SQL automation scripts to streamline reporting cycles, while partnering with QA and product teams to validate data accuracy against KPIs and SLA thresholds.

PROJECTS

NewsSphere: Personalized News Digest Platform | Google Cloud, Docker, Pinecone, AzureSQL, FastAPI

- Engineered FastAPI-based services with optimized RAG indexing, reducing content retrieval times by 35% and driving higher user engagement.
- Deployed containerized Streamlit app and services with Docker + Kubernetes, ensuring fast, reliable personalized news delivery every 8 minutes

AWS-Based YouTube Data Analytics | IAM, Redshift, S3, Lambda, Spark, Glue, Athena

- Built a serverless ELT pipeline using AWS Lambda, Glue, Redshift to ingest multi-regional YouTube data, and analyze 80K+ user interactions in real time, powering QuickSight dashboards that informed product engagement and growth strategies for marketing teams
- Automated deployment with Terraform IaC and scaled workloads via Kubernetes (EKS), improving pipeline resilience and cost efficiency.

Retail Sales Data Pipeline | Azure Blob Storage, Databricks, Spark, Azure SQL

- Architected a Spark pipeline processing 1M+ retail transactions, enabling trend analysis and accurate forecasting.
- Implemented Grafana dashboards and OpenTelemetry tracing to proactively detect errors and ensure pipeline observability.

Iowa Retail Sales Intelligence & Reporting | Alteryx, Talend, Azure Data Factory, Power BI, Tableau, Azure Data Studio

- Engineered ETL workflows using Azure Data Factory and Alteryx to process 25M+ sales records, built Kimball-based models to boost ingestion speed by 60%, and developed Power BI dashboards to improve sales forecasting and inventory planning.

SKILLS

Programming & Scripting	: Python, SQL, Scala, Bash, Java, C++
Data Engineering & Orchestration	: Airflow, dbt, Spark (PySpark), Databricks, AWS Glue, Azure Data Factory
Cloud & Infrastructure	: AWS (S3, EC2, Lambda, Redshift, Glue, EKS), Azure (ADF, Synapse), GCP (BigQuery, Cloud Run), Kubernetes, Helm, ArgoCD, Flux, Terraform
Data Modeling & Governance	: Snowflake, Redshift, PostgreSQL, MongoDB, Dimensional Modeling, RBAC, Row-Level Security, Metadata Lineage
Observability & Monitoring	: Prometheus, CloudWatch, OpenTelemetry, Splunk, Grafana, ELK stack
Business Intelligence	: Tableau, Power BI, QuickSight, Streamlit, Superset
APIs & Frameworks	: FastAPI, REST APIs, Flask
DevOps & CI/CD	: Docker, GitHub Actions, Jenkins, CircleCI
Analytics & Experimentation	: A/B Testing, Hypothesis Testing, Product KPIs, Event Logging