

## EDUCATION

### Northeastern University, Boston

Expected May 2024

Master of Science in Information Systems

**Relevant courses:** Data Architecture and Business Intelligence, Advance Data Science/Architecture, Big Data Architecture and Governance, Data Management and Database Design

### Savitribai Phule Pune University, India

Jun 2016 – May 2020

Bachelor of Engineering in Information Technology

## SKILLS

- **Programming Languages:** Python, R, SQL, PySpark, Pandas, NumPy
- **Data Visualization Tools:** Tableau, Microsoft Power BI, Matplotlib, SAP Business Objects
- **Cloud Services:** Google Cloud Platform: GCP Compute Engine, GCP BigQuery  
Azure: ADLS, Databricks, Data Directory, Data Factory, Kafka, Synapse, SSIS  
AWS: AWS EC2, S3, SageMaker, AWS Redshift, AWS Glue, AWS Lambda

## WORK EXPERIENCE

### Praxis Info Solutions, India, Pune

#### Data Engineer- II

Jun 2021 – Aug 2022

- Developed a comprehensive batch of **15+ ETL pipelines to Azure Databricks (Lakehouse)** for real-time data ingestion from 7 primary data sources (**APIs and ERP databases**), significantly enhancing system throughput to manage an **10,000+ transactions per hour**
- Implemented SQL queries to **extract and analyze data** from multiple databases, enhancing **the accuracy of incident management insights** by 15%

#### Data Engineer- I

Jun 2020 – May 2021

- Created a **relational database system** to store and analyze company-wide data from **multiple sources**, increasing **data accuracy by 20%**
- Led the successful implementation of a real-time dashboard visualization project, utilizing **Tableau**, to **track KPI and monitor data accuracy**, resulting in a 15% increase in operational efficiency and data-driven decision making
- Collaborated with cross-functional teams to develop and maintain **data governance policies and data security protocols**, ensuring compliance with industry regulations and protecting sensitive data

#### Data Analyst - Intern

Jan 2020 – May 2020

- Collaborated with clients and technical teams in assessing existing data sources and data quality, and also worked on **cleansing and standardizing data**.
- Created a comprehensive mapping model using **SAP Business One** to define the relationships between data in the legacy systems and the corresponding data fields in SAP Business One.
- Worked on a project for leading **FMCG company for implementation of SAP Business One**

## ACADEMIC PROJECTS

### IMDB integration for Advanced Analytics (Talend | Alteryx | Tableau | SQL)

Dec 2023

- Executed end-to-end **ETL processes** in Talend, seamlessly integrating data from **MySQL, TSV, and JSON sources**
- Designed and implemented a Type 2 Movie Titles Dimension table, incorporating **SCD2** methodology for historical tracking
- Conducted **SQL scripting** to extract relevant information, ensuring alignment with **Power BI/Tableau visual values**

### Stock Market Real-Time Data Analysis (Apache Kafka | AWS | Athena | Python | SQL)

Sep 2023

- Implemented a comprehensive ETL pipeline leveraging Python, **AWS services (Kafka, Glue, Athena)**, and SQL for efficient processing of real-time stock market data, ensuring seamless execution of the project.
- Orchestrated **real-time data flow** through **Apache Kafka**, performed transformations with Glue, and **optimized storage on AWS S3** for a robust end-to-end solution. and **accurate financial analysis**.

### Airflow Twitter Integration (Python | Apache Airflow | EC2 | S3)

July 2023

- Orchestrated an **end-to-end ETL pipeline** employing Apache Airflow, utilizing **custom DAGs and Operators** to automate Twitter **data extraction** via the Twitter API.
- Ensured **data integrity** through Python-based transformations, deploying the code seamlessly on Airflow/EC2, and storing the finalized results on Amazon S3 for comprehensive and accessible **data analysis**.

## CERTIFICATION

- DP - 900 Azure Data Fundamentals, Microsoft Certified | Google Data Analytics