

SKILLS

- **Programming Languages:** Python, SQL, Java, R, Bash/Linux/Shell, HTML, CSS, JavaScript
- **Databases & Data Integration:** Snowflake, Teradata, MS SQL Server, MySQL, Oracle, PostgreSQL, SSIS, Talend, Alteryx
- **Cloud & Big Data Technologies:** Microsoft Azure, ADLS, Data Factory, Synapse, Event Hubs, GCP, BigQuery, Dataflow, Dataproc, Cloud Composer, AWS, Redshift, Glue, Lambda, Kubernetes, Docker, Kubeflow, Airflow, Databricks, Spark, Kafka, Hadoop, Hive, Pig, HBase, MongoDB
- **Business Intelligence & Analysis:** Tableau, PowerBI, QlikSense, QlikView, SSRS, Excel, Gretl
- **Other Tools & Technologies:** Git, GitLab, GitHub, JIRA, Rally, SharePoint, Confluence

WORK EXPERIENCE

Data Engineer - Senior Consultant at CVS Health, Wellesley, MA Mar 2021 - Current

- Conducted walk-through, requirements gathering meetings and worked with the stakeholders for gathering specifications.
- Implemented a large **cloud data warehouse** platform to host Clinical Trials Services business on **Microsoft Azure** and **Snowflake**.
- Implemented strategy, system architecture design and road map to establish & maintain a comprehensive patient database across multiple lines of business.
- Assisted in building **data model**, design process and governance for mapping patients to diseases (Clinical Knowledge Library).
- Developed **data pipelines** to extract and transform different data types (including structured/semi-structured) from multiple sources and scale/containerize using orchestration tool built using **Python** and **Kubeflow** on **Azure Kubernetes Service (AKS)**.
- Designed, developed, tested, deployed, and continually improved existing Targeting data framework using **Python** and **SQL**.
- Imported and exported data to and from multiple data sources using API calls, RDBMS, flat files, parquet files etc.
- Worked on multiple Feasibility study **data analyses** that uncovered business worth of **\$123 Million** in total sales.
- Worked with cross-functional teams to design, develop product, implement features, and deploy them by coordinating with teams like IT project team, Data Science team, Release Management DevOps (CI/CD) team and other partners.
- Implemented **automaton** using **Python** to run standard Feasibility **SQL** queries which led to **70%** reduction in manual effort.
- Identified, assessed, and resolved complex analytical problems including data sourcing, data ingestion, data transformation issues, potential risks, facilitated issue resolution, risk mitigation, and implemented data standardization.
- **Mentored** new team members to bring them up to speed on the product implementation and assisted with technical difficulties.
- Conducted **code review** and suggested improvements in existing workflow/data processes.
- **Led** and **managed** offshore resources to support operations on a day-to-day basis, standardized escalation hierarchy and established SOP/guidelines to support SLA per customer needs.

Implementation Consultant - Data Management at NeoXam, Boston, MA Nov 2017 – Feb 2021

- Designed and updated existing **data models** to build a large-scale **data warehouse** based on the requirements after interacting with non-technical business users.
- Developed end to end **data pipeline** from multiple data sources like RDBMS, XML, Parquet, csv, web services like REST & SOAP API, and **stream data** using JMS messages by performing extraction, transformation, and load (**ETL**). Implemented data wrangling, cleaning, transforming, merging, and reshaping data frames using **Spark** on **Databricks**.
- Collaborated with data governance, IT, and business teams to check for data quality and develop a microservice architecture.
- Wrote high quality complex **SQL** queries using joins, nested subqueries, and temporary tables to generate reports.
- Developed **Tableau** visualizations and dashboards integrating data from multiple data sources using Data Blending.
- Automated multiple routine tasks for end users using **OOPs** concepts in **Python** and **shell scripting**, which reduced the manual work by **95%**. Hosted and managed DataHub application on the cloud using **AWS**.
- Provided **project management leadership** following **Kanban** methodologies, to have streamlined execution of the projects.

Reporting and Business Intelligence Developer at GMO LLC, Boston, MA May 2016 – Dec 2016

- Applied **Agile & Scrum** methodologies, following full software development life cycle (**SDLC**) gathering requirements, designing, developing, documenting, testing, and deployment in **Business Intelligence**.
- Wrote optimized **SQL** queries, implemented tables, and views to generate verification reports.
- Designed **ETL** data pipeline for refreshing data in data warehouse using direct load, SCD 1 and SCD 2 **SSIS** mappings.
- Migrated **QlikView** dashboards to **Tableau** for finding **Key Performance Indicators (KPI)**.

EDUCATION

Master of Science, Information Systems at Northeastern University, Boston, MA	Sept 2015 – Aug 2017
Bachelor of Engineering, Computer Engineering at University of Mumbai, Mumbai, India	Aug 2010 – July 2014