

# Sneha Roseline Augustine

Email: [roseline.dataengineer@gmail.com](mailto:roseline.dataengineer@gmail.com)

<https://github.com/sneha-roseline> | <https://www.linkedin.com/in/sneha-roseline-a-48238b115/>

A Data geek with a flair for innovation and attention to detail. Effective collaborator with strong communication skills, decision-making and problem-solving abilities towards complex business challenges. Known for seamlessly blending teamwork and leadership with diverse cross-functional teams to achieve collective objectives. Self-starter with the ability to work independently.

## TECHNICAL SKILLS

**Programming:** Python, SQL, PySpark, SparkSQL, Java, HTML, CSS, shell scripting

**Database:** SQL Server, AWS Redshift, Snowflake, BigQuery, Oracle, MySQL, HBase, Teradata

**Big Data:** Apache Spark, Hadoop, HDFS, Map Reduce, Hive, Sqoop, Kafka

**Cloud:** AWS (S3, RDS, EMR, EC2, Lambda, Glue, Athena), Azure (Data Lake, Data Factory, Synapse Analytics, Blob storage, Databricks), GCP Compute Engine

**Methodologies:** Agile, Scrum, SDLC, JIRA, GIT, A/B testing, CI/CD

**Tools:** Salesforce, Airflow, Power BI, Tableau, DBT, Docker, Kubernetes, Terraform, Fivetran, SSIS, SSRS, Looker, Github, Visual Studio Code, PyCharm, Jupyter Notebook, Jenkins

## PROFESSIONAL EXPERIENCE

### Senior Data Engineer

May 2023 - Present

Programsoft LLC – Philadelphia, Pennsylvania

- Created spark clusters and configured high concurrency cluster using Azure Databricks to speed up the preparation of high-quality data.
- Proficient in writing pyspark and spark SQL transformation in Azure Databricks to perform complex transformations for business rule implementation.
- Built automated data pipeline to load data from Amazon S3 to snowflake using Airflow

### Senior Data Engineer

Feb 2022 - Mar 2023

Connect America – Philadelphia, Pennsylvania

- Built and maintained data pipeline for ingesting tables from on premise SQL Server Database using Azure services like Data Factory, Data Lake, Databricks, Synapse Analytics and created interactive dashboards using Power BI.
- Utilized Azure Active Directory and Azure Key Vault for data governance and monitoring
- Performed relational and dimensional data modeling to create logical and physical designs of databases and ER Diagrams with all related entities and relationships using ERWIN.
- Developed python data pipelines to extract data from on-premises source, using AWS S3 as external stage and loaded the data into Snowflake using snowpipe, stream and tasks.
- Performed all necessary day-to-day GIT support for different projects, responsible for maintenance of the GIT repositories, and access control strategies.
- Implemented, designed and managed CI/CD pipelines using Jenkins for enabling automated testing, validation and deployment of ETL pipelines.

### Data Analytics Engineer

Aug 2018 – Jan 2022

Connect America – Philadelphia, Pennsylvania

- Extracted and transformed data performing analysis on healthcare dataset to create automated reports using SQL, Power BI, Excel and Python libraries like NumPy and Pandas.
- Proficient in data modeling and schema design, creating effective and efficient data models to support business objectives and data-driven decision-making

- Maintained strict regulatory compliance through data handling, quality assurance and adherence to HIPPA regulations.
- Collaborated and coordinated with stakeholders from cross-functional teams to understand, define data requirements and develop data-driven solutions to support strategic initiatives.
- Designed Power BI data models with multiple fact tables and dimensions.
- Implemented ETL processes with SSIS for diverse data sources and created reports in SSRS.
- Coordinated with the marketing team on data ingestion from disparate sources using SQL, Google Analytics API, Salesforce API and python to support vendor solutions.

**STEM Coding Instructor & Trainer**

Feb 2017 – July 2017

**Community Partnership Charter School** – New York, New York

- Mentored HTML, JavaScript, SQL in program design, data and curriculum.

**Graduate Research Assistant**

Dec 2016 – June 2017

**New York Institute of Technology** – New York, New York

- Developed an employee management android application using Java and MySQL to effectively track activities and manage employee data.

**Data Analyst**

May 2014 – Jan 2016

**CITI** – Chennai, India

- Synthesized transaction data for statistical metrics using SQL and VBA.
- Achieved 35% improvement in account management with ad-hoc reports using MySQL, Excel and CSV files.
- Created KPI reports and dashboards from over 100,000 banking data using Tableau for customer churn analysis, sales and performance.

**EDUCATION**

**Master of Science in Computer Science**

Aug 2017

New York Institute of Technology, New York

**Bachelor of Technology in Information Technology**

Apr 2014

Anna University, India

**CERTIFICATION**

Databricks Accredited Lakehouse fundamentals

Taming Big Data with Spark and Python

Analyzing and Visualizing data with Microsoft Power BI

Data Warehouse and Business Intelligence (Informatica & Cognos)

**PERSONAL PROJECTS**

- Built a data pipeline in Azure environment with Data Factory, Data Lake for storage, Databricks for Spark transformations and Synapse Analytics for analysis with Tokyo Olympic 2020 dataset.
- Scraped data with News API building an end-to-end ETL pipeline on AWS Cloud using Python, EC2, S3, Glue Crawler, Data Catalogue and used Apache Airflow to orchestrate the process.
- Deployed Google Cloud Compute Engine for Jupyter notebooks, facilitating data-driven decision-making with advanced data analysis and visualization expertise in real-world aviation datasets.