# Abhishek Teli

Data Engineer

Atlanta, GA — Open to Relocate

+1 315-450-5629 | teli.abhi01@gmail.com | LinkedIn: abhishekteli | GitHub: abhishekteli


## Professional Experience

**EXL** June 2023 - Present

*Data Engineer*

- Engineered a **Multi-Account AWS Data Pipeline** enhancing **data ingestion** and **transformation** efficiency by **40%** and bolstering data security in the **data warehouse** and **data lake**.
- Augmented data ingestion throughput by **30%** into the **S3 data lake** using **AWS Data Sync** and **Kinesis**, utilizing **AWS Lambda** and **EventBridge** for efficient pipeline automation and orchestration.
- Enhanced **data validation** and transformation accuracy by **35%** with **AWS Lambda** and **Glue PySpark scripts**, ensuring data integrity within the **data lake** and **data warehouse**.
- Employed **Glue DataBrew** for PII data management in the **S3 data lake**, improving compliance and data preparation efficiency by **25%**, and optimized data querying capabilities using **Redshift Spectrum** and **Athena**.
- Advanced **CI/CD** integration with **GitHub** and **CloudFormation**, boosting deployment efficiency and infrastructure reliability by **20%** for the **data warehouse** and **data lake** environments.

**Barclays** August 2018 - June 2021

*Data Engineer*

- Successfully created a **Data pipeline** that transformed a large dataset of **150 million records** allocating **1 billion Euros** and Business rules parsing using **PySpark** and **Airflow DAGs**, reducing manual parsing work by **2 days**.
- Implemented a **Hadoop Data Warehousing** solution using Hadoop echo system by creating **PySpark** ETL scripts and integrating data from various sources through effective **Data modeling**.
- Developed a **AWS data pipeline** which leverages **S3**, **AWS Glue** for migration of on-prem database to **S3 Datalake** in AWS cloud and orchestrated pipeline using **AWS Step Functions**.
- Developed **PySpark scripts** and **AWS Lambda functions** to develop unit tests ensuring source data authenticity.
- Collaborated in developing AWS Migration plan adhering to **design patterns** with **Business stakeholders and Scrum Master**, reducing storage costs by **25%** and improving runtime by **40%** and improving project **scalability**.

## Education

- **Masters in Computer Science, Syracuse University** Aug 2021 - May 2023
  **Coursework** : Data Structures, Machine Learning, Alogrithms, Database Managements (Data Modeling, Data Warehousing) ,Cloud Computing
- **Bachelors in Computer Science, Ramaiah Institute of Technology** Aug 2014 - Jun 2018
  **Coursework** : Data Structures, Alogrithms, Database Managements (Data Modeling, Data Warehousing) ,Cloud Computing, Data Mining, Python

## Skills

- **ETL/ELT tools** : Spark, Kafka, HiveQL, HDFS, Hadoop, AWS Glue, AWS Lambda, Google Data Proc, Spark Streaming, Kinesis, EMR
- **Data Warehouse/ Data Lake**: S3 , Cloud Storage, Redshift, Big Query, Snowflake
- **Programming Languages/Scripting**: Python, Java, Scala, Shell, Unix commands, SQL
- **Relational Databases**: MySQL, Oracle, Teradata, PostgreSQL, DB2, SQL Server, RDS, Athena
- **Cloud** : AWS, GCP
- **NoSQL Databases**: DynamoDB, Cassandra, MongoDB
- **Orchestration Tools and CI/CD**: Airflow, Step Functions, Composer,GitHub, Jenkins, CloudFormation

## Certifications

- Spark Developer Associate - Databricks(Python) February 2024
- Google Cloud Professional Data Engineer November 2022
- AWS Associate Solution Architect August 2022
- SQL Programming - Advanced March 2023

## Projects

**Real Estate Chat Bot**

- Engineered an advanced real-time and batch data processing pipeline using **LangChain**, **ChatGPT**, **Kafka**, and **Spark Streaming**, facilitating dynamic user interactions with a chatbot for real estate inquiries and streamlined data retrieval and storage in **CSV** and **PostgreSQL**.
- Leveraged **Lang Chain** for natural language understanding, transforming real estate listings into embedded vectors to enhance the chatbot's query accuracy, integrating these vectors with the ChatGPT model via **ChromaDB**.
- Orchestrated periodic data updates and statistical analytical processes using **PySpark** and **Airflow**, employing a **Gradient Boosting Regressor** model to analyze and predict property investment values based on market trends.