Abhishek Teli

Data Engineer

Atlanta, GA — Open to Relocate

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

Summary:

Data Engineer with 3 years of experience in designing and deploying scalable data pipelines using AWS, GCP, and Big Data Hadoop ecosystems. Proficient in Python, Spark, SQL, and data visualization with Tableau, delivering comprehensive data solutions from ingestion to insightful reporting.

Professional Experience

Data Engineer   June 2023 - Present

EXL

- Orchestrated end-to-end **data migration** pipelines on AWS, utilizing **Glue**, **Lambda**, **Data Sync**, **Kinesis**, and **Step Functions** to construct robust **data pipelines** for seamless **ingestion**, **transformation**, and **loading** of data from diverse sources such as **database files**, **SMB files**, **API data**, and **streaming data**.
- Implemented automated infrastructure provisioning with **CloudFormation** templates across **data lake**, **data warehouse**, and **data lake house** architectures, optimizing resource management.
- Executed intermediate data **validations** and **transformations** using **PySpark** scripts within **AWS Glue**, ensuring data integrity and accuracy throughout the **data pipeline**.
- Leveraged **AWS Glue** and **DataBrew** for advanced data transformations, automating the identification and processing of **PII data** to ensure compliance with privacy regulations.
- Utilized **AWS Redshift** and **Athena** for SQL-based data querying and analysis, providing valuable insights for stakeholders.
- Contributed to **CI/CD** practices with **GitHub Actions**, enabling automated testing, code quality checks, and deployment workflows for continuous integration and deployment objectives.

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