SRILEKHA TUMMALAPALLY

+1 (775) 300-8759 - reddysrilekha0301@gmail.com - Charlotte, NC, 28217.

OBJECTIVE

Creative-minded and detail-oriented individual Data Engineer proficient in Pyspark and Scala, with a passion for optimizing cloud-based data solutions. Experienced in leading and mentoring teams, leveraging advanced cloud technologies, including Azure and AWS, to design and implement fault-tolerant data pipelines.

SUMMARY

Accomplished and seasoned Data Engineer with over 4 years of proven expertise in architecting impactful solutions. Proficient in leveraging Python, Apache Airflow, Kafka, Spark, Scala, Java, Snowflake, Hadoop, and SQL/NoSQL databases including MySQL, PostgreSQL, MongoDB, and DynamoDB to design fault-tolerant data solutions on Azure and AWS. Led the design and implementation of ADF pipelines, optimizing data workflows and reducing processing times by 30%. Managed Azure subscriptions and leveraged Azure Blob Storage for efficient data storage solutions. Pioneered query optimization strategies, resulting in a 20% improvement in database performance. Skilled in creating reports and dashboards using Tableau and PowerBI for data visualization and analysis. Experienced in implementing CI/CD pipelines using Jenkins for automated testing and deployment. Proficient in Linux for operating system management and scripting. Demonstrated expertise in integrating and building REST APIs for seamless data communication and integration. Committed to delivering solutions while maintaining security and compliance standards.

EXPERIENCE

Senior Data Engineer (Client: Vanguard, Charlotte, NC) July 2023 - Present

- Led end-to-end Python development projects, integrating Apache Airflow for workflow orchestration and REST APIs for seamless data integration, ensuring scalability and interoperability.
- Responsible for building highly Reliable and Scalable ETL Pipelines using highly specialized knowledge of Data Collection, Data integration, and Data Loading that facilitate deeper analysis and reporting by the Data and Analytics department.
- Developed and optimized data ingestion pipelines using Apache Iceberg to support incremental data processing and time travel capabilities.
- Implemented data validation and quality checks within Pandas workflows to ensure accuracy and consistency of data for downstream tasks.
- Implemented end-to-end data processing solutions on AWS, leveraging AWS Glue and AWS Sagemaker for automated ETL operations, AWS Lambda to implement governance solutions and Athena for Querying.
- Built and maintained scalable and efficient data pipelines from ingestion to consumption using Java, Python, Scala and developed API endpoints for serving and consuming data in JSON and other formats.
- Implemented containerized microservices architecture with Docker, Docker Compose for multi-container deployments.
- Utilized DynamoDB's partition keys, sort keys, and secondary indexes, alongside PostgreSQL and MySQL optimization techniques, and complex SQL queries, enhancing data access and analysis capabilities.
- Designed and implemented scalable solutions using Spark, Scala, spark-streaming, and Kafka to process and analyze large datasets for server analytics and machine learning applications.

Data Engineer (PepsiCo, Plano, TX) April 2021 - July 2023

- Deployed VM instances, provisioned Azure Storage, configured security groups. Experience includes working with Azure products like Azure VMs, Azure Data Lake Storage, and Azure Blob Storage for raw data migration and processing.
- Collaborated with cross-functional teams to design, develop, and implement ETL architectures using Azure Databricks with Pyspark and SparkSQL, enabling efficient data integration for both batch and real-time processes and transformed RDDs into DataFrames to take advantage of DataFrame's optimized execution plans and built-in optimizations for structured data processing.
- Designed and implemented data pipelines using Teradata and Snowflake's data warehouse architecture to efficiently handle large-scale data processing tasks.
- Utilized Azure Data Factory and Azure Databricks (Delta tables, Delta lake, Delta live tables) for data processing.
- Gathered, analyzed, and drew conclusions from large, diverse data sets to identify problems and contribute to decision-making in service of secure, stable application development.
- Created and maintained stored procedures, triggers, and views to ensure data integrity and consistency and

developed and maintained reports and dashboards using Tableau and PowerBI.

- Built, tested, and enhanced data curation pipelines integration data from a wide variety of sources like DBMS, File systems, and APIs for various KPIs and metrics development with high data quality and integrity.
- Handled data manipulation (extract, load, transform), data visualization, and administration of data and systems securely and in accordance with enterprise data governance standards.
- Developed automated testing (Unit Testing, Integration Testing, Regression Testing) frameworks and CI/CD pipelines using Python and Jenkins, reducing deployment time by 50%
- Utilized Azure key vault for security. Collaborated with data scientists and analysts to bring data into a centralized data repository and build additional data sets for analytics and reporting purposes.
- Used GIT for version control, Eclipse and IntelliJ IDEs for extensive development and JIRA to handle Software Development issues and involved SDLC life cycles and release management.

Data Engineer (Cloudhub IT Solutions, Plano, TX) January 2021 – April 2021

- Deployed Instances, provisioned EC2, S3 bucket, Configured Security groups and Hadoop eco system including HDFS, Hive, Sqoop, Impala for Cloudera in AWS. Experience in using distributed and cloud computing architectures like AWS products (e.g., AWSEC2, AWS Redshift, and AWS EMR, AWS S3, AWS Lambda) and working on raw data migration to Amazon cloud into S3 and performed refined data processing.
- Designed and implemented Excel-based ETL workflows to extract data from various sources, including databases, CSV files, and APIs, into Excel spreadsheets and utilized dbt for workflows.
- Designed, built and maintained tools and services that improved the internal data infrastructure platform using Python, AWS, CloudFormation, and Elastic Kubernetes Service (EKS)
- Utilized SQL like Oracle, Db2 databases and NoSQL DBMS for querying and maintaining data quality
- Conducted data analysis using SQL and Python to provide insights into customer behavior and preferences.
- Performed data profiling, data mapping, and data modeling using Kimball methodology to support data analysis and reporting

Data Analyst Intern (Advansoft International Inc, Arlington Heights, IL) August 2020 – January 2021

- Developed Datawarehouse process models, including sourcing, loading, transformation, and extraction using for streamlined data management and integration.
- Developed new programs or modify existing programs to meet customer requirements using SQL, PySpark, Scala.
- Designed and Implemented warehouse database using physical database design structures.

SKILLS

Programming Languages	:	Python, Spark, Scala, SQL, Snow SQL, Java, C, C++
Database Systems	:	Teradata, Snowflake, Cassandra, MySQL, Oracle, MongoDB, Redshift, MariaDB
Big data	:	Kafka, Spark streaming, Storm, Sqoop, MapReduce, Pig, Hadoop, HDFS, Hive
Azure Cloud Data Solutions	:	Azure Data Lake, Azure Data Factory, Azure Databricks, Azure IOT key vault
AWS Cloud Data Solutions	:	Redshift, S3, EMR, EC2, Glue, DynamoDB, Lambda, RDS, EKS, Athena
ETL/ELT Tools	:	AbInitio, Talend, DataStage, Informatica, SSIS, Alteryx, Unity catalog
DevOps	:	Docker, Streamlit, dbt
Data Visualization	:	Tableau, PowerBI
CI/CD	:	GIT, Jenkins
Operating Systems	:	Linux, Unix, Windows

EDUCATION

Master of Science in Computer Science August 2018 - May 2020

University of Nevada, Reno -Nevada – *CGPA - 3.8/4.0*

- Relevant Coursework: Data Structures, Analysis of Algorithms, Database Management systems, Machine Learning, Internet Security, Python Programming, Data Mining, Information Management and Visualization, Statistics in Decision Making

Bachelor of Technology in Electronics and Communications Engineering September 2014 - May 2018

Mallareddy Engineering College- Hyderabad, India – *CGPA – 8.5/10.0*

- Relevant Coursework: Machine learning, Operating systems, Computer architecture, Object oriented programming languages like python and java, Computer Networks, Data Structures, Algorithms