

DIVYA SIRUGUDI

Mobile: +1 469-596-0539 **Email:** divyastorm28@gmail.com

IT professional with 6+ years of hands-on experience specializing in Big data, Data Governance, Data Warehousing Big Data environment management, and end-to-end data lifecycle management. Proficient in data acquisition, ingestion, modeling, storage analysis, integration, and data processing. Demonstrated expertise in API integration, data extraction, analysis, and metadata management. Adept at spearheading strategic initiatives to optimize data workflows and ensure compliance while driving actionable insights. Excels in collaborating with cross-functional teams to deliver solutions aligned with business objectives.

Education:

January 2021-December 2022

]University of North Texas++

M.S, Data Science

Potti Sriramulu College of Engineering and Technology

August 2011 - April 2015

B.Tech, Electronics and Communications Engineering

Technical Skills:

Data Engineering: ETL Operations, Data Pipeline Development, Data Integration, Big Data

API Integration: API Solution Design, RESTful APIs, Connectivity Solutions

Data Analysis: Data Extraction, Data Cleansing, Data Validation

Metadata Management: Metadata Extraction, Metadata Cleansing

Programming Languages: Python, SQL

Tools & Technologies: AWS (Glue, EMR, S3, Redshift, Lambda), GCP, Terraform, Spark, Flask, Django, Tableau

ML and NLP Algorithms: Linear Regression, Logistic Regression, Naïve Bayes, Decision Tree, Image Processing, Text Processing, Text Classification & Clustering, Sentiment Analysis.

Professional Experience:

Eficens Systems LLC, Atlanta, GA

(client- McKinsey and Company)

December 2022- March 2024

Senior Data Engineer

- Spearheaded the development and implementation of strategic data initiatives for Controllers, leveraging expertise in data governance, data quality, data management, and data warehousing.
- Documented end-to-end Data Lineage from source to reporting, ensuring transparency, compliance, and accuracy in data processes within the data warehousing environment.
- Initiated and modified systems to meet evolving user requirements, while ensuring alignment

with strategic business objectives and data warehousing best practices.

- Conducted in-depth analysis of Controllers' business processes and system issues to propose and implement strategic technological and procedural workflow improvements within the data warehousing framework, driving efficiency and effectiveness.
- Partnered with technology teams to ensure user tools for analyzing data within the data warehousing environment meet requirements, facilitating seamless integration and utilization of data analytics tools.
- Proficient in AWS cloud services such as EC2, S3, RDS, ELB, EBS, VPC, Route53, auto-scaling groups, CloudWatch, CloudFront, and IAM for server migration, configuration management, and data warehousing solutions.
- Spearheaded the design and implementation of MLOps (Machine Learning Operations) pipeline within the data warehousing environment to automate Data Quality and Data Masking processes.
- Conducted comprehensive evaluations of imputation and outlier detection methods within the data warehousing framework, fine-tuning parameters, and optimizing performance to meet specific business objectives and domain requirements.
- Proficient in data extraction and enrichment utilizing Spark SQL within the data warehousing environment, including Cassandra tables joins and conversion of Hive queries into Spark transformations.
- Worked closely with Data Science teams to run Machine Learning models on Spark EMR clusters within the data warehousing framework, delivering data-driven insights aligned with business requirements.
- Involved in writing unit tests within the data warehousing environment and worked along with DevOps team in Installing libraries, productionized ETL jobs.
- Used Ansible to provision the environment and deployed applications in a CI/CD process within the data warehousing framework using a Jenkins pipeline. Also managed and deployed configurations using Terraform.
- Wrote Kafka producers to stream the data from external rest API to Kafka topics, facilitating data ingestion and processing within the data warehousing environment.

Eficens Systems LLC, Atlanta, GA
Data Engineer

February 2022-November 2022

- Involved in developing batch and stream processing applications that require functional pipelining using Spark APIs.
- Developed Spark jobs on Databricks to perform tasks like data cleansing, data validation, standardization, and then applied transformations as per the use cases.
- Worked with Data Science team running Machine Learning models on Spark EMR cluster and delivered the data needs as per business requirements.
- Experience with Apache bigdata Hadoop components like HDFS, MapReduce, YARN, Hive, HBase, Sqoop, Pig, Ambari and Nifi.
- Involved in building a data pipeline and performed analytics using AWS stack (EMR, EC2, S3, RDS, Lambda, Glue, SQS, Redshift).
- Developed Sqoop jobs for data ingestion, incremental data loads from RDBMS to HDFS.

- Utilized Spark's in memory capabilities to handle large datasets on S3 Data Lake. Loaded data into S3 buckets, then filtered and loaded into Hive external tables.
- Strong Hands-on experience in creating and modifying SQL stored procedures, functions, views, indexes, and triggers.
- Involved in developing a custom-built Rest API to support real time customer analytics for data scientists and applications.
- Involved in extracting and enriching multiple Cassandra tables using joins in Spark SQL. Also converted Hive queries into Spark transformations.
- Hands-on experience on API design and development using Spring Boot for Data movement across different systems.
- Fetched live data from Oracle database using Spark Streaming and Kafka using the feed from API Gateway REST service.
- Automated the process of transforming and ingesting terabytes of monthly data using Kafka, S3, Lambda and Oozie.
- Performed ETL operations using Python, Spark SQL, S3 and Redshift on terabytes of data to obtain customer insights.
- Developed Oozie workflows for scheduling and orchestrating the ETL process. Involved in writing Python scripts to automate the process of extracting weblogs using Airflow DAGs.
- Migrated Java analytical applications into Scala. Used Scala where performance and logic is critical.
- Involved in writing unit tests, worked along with DevOps team in Installing libraries, Jenkins agents and productionized ETL jobs and microservices.
- Used Ansible to provision the environment and deployed applications in a CI/CD process using a Jenkins pipeline. Also managed and deployed configurations using Terraform.
- Experience with analytical reporting and facilitating data for Tableau dashboards.

IMR Software PVT LTD, Hyderabad, India

January 2018- August 2020.

Data Engineer

- Using Hadoop/Big Data concepts, loaded and transformed large sets of structured, semi - structured and unstructured data.
- Developed Spark applications using **Pyspark** and **Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- Developed SQL queries to search the Repository DB for deviations from the Company ETL Standards for items such as Sources, Targets, Transformations, Log Files, Mappings, Sessions, and Workflows that were produced by users.
- Optimized Hive queries to extract the customer information from **HDFS**.
- Worked on a Scala log producer that monitors application logs, transforms incremental logs, and transmits them.
- Hands-on experience on developing **SQL** Scripts for automation purposes.
- Used Talend for Big data Integration using Spark and Hadoop.
- The **Spark-SQL** tables were read, processed, and generated using the Scala API.
- Using advanced approaches including bucketing, partitioning, and optimizing self joins, we used

structured data in Hive to run fast.

- Using the **Scala API**, we used **Spark-SQL** to read, analyze, and generate tables from parquet data.
- Exposure to Spark Architecture and how **RDDs** function by involving and processing data from local files.

S V Technologies, Vijayawada, India

February 2016- January 2018.

Data Analyst

- Played a key role in analytics, engineering, and modeling, contributing to the design and development of data solutions.
- Assisted in coaching and training business analytics solution users to enhance decision-making capabilities.
- Ensured and maintained high levels of data integrity through monitoring tools and mass update processes.
- Cultivated and managed relationships throughout the organization, facilitating effective collaboration and communication.
- Collaborated with Data Engineering Teams to develop and optimize pipelines, ensuring data quality and integrity throughout the process.
- Implemented and validated new data quality checks and measures, enhancing the accuracy and reliability of predictive modeling algorithms and applications.
- Streamlined data processes, leading to a 15% reduction in data redundancy and improving overall data quality.
- Conducted Exploratory Analysis of Internal and External data sources to identify data quality issues and opportunities for improvement.
- Contributed to CI/CD pipelines focused on data quality improvements, ensuring consistent and reliable data delivery throughout the software development lifecycle (SDLC).