YERRAMSETTI DHARMA TEJA
Sr Data Engineer PH:+1 812-671-5428
tejayd01@gmail.com
Masters in Data Science at IUB (2022-2024)
LinkedIn: https://www.linkedin.com/in/teja-y-d-5a3871228

Professional Summary

Experienced Senior Data Engineer with 8+ years of experience designing, building, and optimizing scalable data pipelines and ETL solutions across healthcare, banking, and fintech domains. Proficient in modern cloud platforms including AWS and Azure, with hands-on expertise in orchestrating complex data workflows using services like AWS Glue, Lambda, EMR, and Azure Data Factory. Skilled in real-time and batch data processing using Spark (Scala/Python), Kafka, and SQL. Strong background in healthcare analytics with Epic Clarity and Caboodle, as well as regulatory compliance (HIPAA, PCI, GDPR). Proven ability to deliver high-impact data products, support analytics teams, and modernize legacy systems using cloud-native architectures and DevOps practices. Adept in tools like SSIS, Airflow, Informatica, and Tableau, driving data-driven decisions with high-quality, reliable data infrastructure.

Professional Experience

Senior Data Engineer
The Ohio State University Wexner Medical Center, Columbus, OH *October 2024 – Present*

- Supporting the reporting and analytics team by developing and maintaining SSIS packages within Visual Studio to automate ETL workflows.
- Designed and managed robust data pipelines to extract data from diverse clinical and administrative sources including **Epic Clarity**, flat files, Excel sheets, and API endpoints.
- Loaded and transformed data into on-premises SQL Server databases, enabling reporting for multiple business units.
- Created and deployed SSIS Catalog jobs with Tidal for automated execution and scheduling, improving data freshness and reliability.
- Optimized data transformations with SQL scripts and stored procedures to ensure accurate and efficient data loading.
- Collaborated with data analysts to gather requirements and deliver data marts, improving downstream reporting capabilities.
- Provided production support for ETL failures, resolved errors quickly, and implemented monitoring to minimize downtime.
- Participated in data governance discussions, ensuring adherence to data quality standards and compliance policies.
- Developed queries and data models from **Epic Clarity** and **Caboodle** for clinical and operational reporting.
- Utilized **Epic Reporting Workbench** and **Radar dashboards** to validate data consistency between source systems and SQL Server.
- Currently exploring cloud-native alternatives to modernize legacy workflows using AWS services like Glue, Lambda, and Redshift.

Environment: Epic Clarity, Epic Caboodle, Reporting Workbench, Radar Dashboards, SQL Server, SSIS, SSIS Catalog, Tidal, Visual Studio, SQL, Flat Files, Excel, APIs, Git, AWS

Data Engineer
Bank of America, Chicago, IL *May 2023 to May 2024*

- Designed and deployed AWS-based multi-tier applications using EC2, S3, RDS, DynamoDB, and CloudFormation to ensure high availability and scalability.

- Built scalable data pipelines integrating AWS Lambda, API Gateway, Glue, and Redshift/Snowflake for seamless data flow and transformation.
- Developed Spark applications in Scala and Spark SQL, migrating legacy MapReduce workflows and supporting real-time ingestion from S3 using Spark Streaming.
- Created Python-based Lambda functions for JSON parsing, validation, and transformation tasks.
- Managed Glue Catalogs with crawlers for automated metadata discovery and querying.
- Designed normalized data models and implemented AWS EventBridge for event-driven architecture across microservices.
- Automated and orchestrated ETL workflows using Apache Airflow and Jenkins, ensuring robust data dependencies and job scheduling.
- Optimized Hadoop clusters on AWS EMR for high-volume batch processing and data warehousing.
- Applied test-driven development using pytest, collaborated with DevSecOps to build secure CI/CD pipelines.
- Developed dashboards and visualizations in Tableau, enabling data-driven decisions for business stakeholders.

Environment: AWS EMR, S3, EC2, Lambda, Redshift, Snowflake, DynamoDB, Glue, EventBridge, Spark, Airflow, Oozie, Hadoop, Kafka, PostgreSQL, MongoDB, Tableau, Jenkins, Grafana, Superset.

Data Engineer
Visa Inc., Bengaluru, India *April 2018 s to July 2022*

- Designed and implemented AWS-based data pipelines using Lambda, Glue, and Redshift to ingest, transform, and stage data from various sources including JSON and APIs.
- Built scalable Spark applications in Scala and Spark SQL to modernize legacy MapReduce workflows and process real-time data from S3 with Spark Streaming.
- Developed Python-based Lambda functions for processing complex JSON structures, and automated workflows using Apache Airflow.
- Managed Glue Catalogs with crawlers and applied data lake architecture principles for efficient staging and querying.
- Integrated Snowflake as a downstream warehouse and developed data pipelines using AWS Glue and Python.
- Administered AWS EMR clusters and scheduled Spark/Scala jobs in Oozie for structured data processing and lineage tracking.
- Designed Tableau dashboards to deliver actionable insights and supported regulatory compliance (PCI, GDPR) through secure data engineering practices.
- Collaborated with analytics and product teams to deliver consistent, governed data solutions.

Environment: AWS EMR, S3, EC2, Lambda, Redshift, Snowflake, Glue, Airflow, Spark, Oozie, Python, Scala, Tableau, Matillion, Git, HDFS.

ETL Developer
United Health Group, Hyderabad, India *January 2017 to March 2018*

- Built and maintained real-time ETL pipelines using **Informatica PowerCenter**, **Kafka**, and **Spark Streaming** for high-volume data ingestion.
- Integrated data from diverse sources (Teradata, Oracle, flat files) into Hadoop using Sqoop, Flume, and Informatica transformations.
- Developed Spark and MapReduce modules in Python for machine learning workflows, including a distributed random forest model.
- Designed data pipelines with **Azure Data Factory**, **Synapse**, and **Databricks**, implementing data lakes and warehousing solutions.
- Automated ETL operations using PL/SQL, Unix shell scripts, and Jenkins, ensuring reliable job execution and secure file transfers via SFTP.

- Ensured compliance with **PCI** and **GDPR** through secure pipeline development and data governance practices.

Environment: Informatica, Kafka, Spark, Python, Hadoop on Azure, ADF, Synapse, Sqoop, Flume, Teradata, Oracle, Unix, SFTP, QMC, SQL Loader.

Data Analyst- Python

Wells Fargo, Hyderabad, India *June 2015 to December 2016*

- Contributed to various projects utilizing machine learning and big data methodologies, with a focus on data visualization and exploratory data analysis using python libraries like NumPy, matplotlib, SciPy and pandas.
- Specialized in quantitative analysis and data mining, adept at interpreting complex datasets to uncover trends and provide actionable insights.
- Documented limitations in data quality that jeopardize the ability of internal and external data analysts' ability.
- Designed and executed advanced SQL queries and scripts for precise data extraction and aggregation, ensuring data integrity and compliance with business requirements. Skilled in translating complex business needs into clear, concise specifications and executable queries.
- Proficient in extracting and analyzing data from varied sources including CSV, Excel, HTML, and SQL, ensuring accurate data migration and reporting.
- Developed comprehensive analysis reports leveraging Excel and Tableau, highlighting data quality, billing patterns, and identifying outliers, which were instrumental in enhancing data integrity and accuracy.
- Utilized Tableau's robust sorting and filtering capabilities, such as basic, quick, context, condition, and top filters, to refine data presentation and filter operations.
- Designed and implemented backend business logic with Python, optimizing performance through effective use of Lambda functions and data frame operations.

Environment: Python, SQL, UNIX, Linux, Oracle, NoSQL, PostgreSQL, PySpark, NumPy, Pandas.