Tejeswar Raju Vempalli

Data Engineer

Email: tejeswarvempalli3@gmail.com   LinkedIn: linkedin.com/in/Tejeswar Raju   Mobile: +1 (216) 966-5599   Place: United States

SUMMARY

Experienced Data Engineer with 5+ years of expertise in designing, developing, and optimizing end-to-end data pipelines across AWS, Azure, and GCP platforms. Built scalable ETL frameworks and automated data quality checks using Python and PySpark, improving data availability by 40% for analytics teams. Engineered high performance Redshift, Synapse Analytics, and BigQuery solutions for efficient data storage and retrieval. Created interactive dashboards in Tableau, Power BI, and QuickSight to deliver actionable insights to stakeholders. Demonstrates expertise in transforming complex datasets into streamlined, business-ready analytics solutions.

TECHNICAL SKILLS

- **Programming & Scripting Languages:** Python, C++, R
- **Cloud Platforms & Services:** AWS, Glue, EC2, S3, Lambda, DynamoDB, Redshift, Kinesis, Azure, Synapse Analytics, Data Factory, Azure MySQL, Azure Data Lake, EventHub, Databricks, GCP, BigQuery, Pub/Sub, Dataproc, DataFlow, Snowflake, Fivetran, DBT(Data Build Tool)
- **Databases & Data Warehousing:** SQL, SQL Server, MySQL, PostgreSQL, T-SQL, Oracle, MongoDB
- **Big Data & ETL Tools:** Pyspark, Data Engineering, Data Analytics, MapReduce, Kafka, SSIS, SSMS, Talend, Airflow, DBT, Informatica, Apache Flink
- **Machine Learning & Analytics:** Supervised & Unsupervised Learning, Neural Networks, NLP, Time-series Analysis, XGBoost, TensorFlow,PyTorch, Inferential Statistics, Descriptive Statistics
- **DevOps & CI/CD Tools:** GitHub, Terraform, Ansible, Jenkins, Docker, Kubernetes
- **Business Intelligence & Visualization Tools:** Tableau, SAS, Power BI, MS Excel
- **Methodologies:** Agile, Scrum, Waterfall
- **Operating Systems**: Windows, Mac OS, Linux
- Software Development Life Cycle

PROFESSIONAL EXPERIENCE

CVS Health   October 2023 – Present
Data Engineer
Technologies: AWS (Glue, Lambda, S3, Redshift, Data Pipeline, Kinesis, Elastic Beanstalk, Airflow), Python, PySpark, Apache Spark, Delta Lake, Talend, Azure Databricks, ETL/ELT, Tableau, Power BI, QuickSight, SFTP, APIs, Databases, Snowflake, DBT,

- Developed AWS-based ETL pipelines using Glue, Lambda, and S3, streamlining healthcare claims data processing and enabling faster data accessibility for analytics teams.
- Implemented scalable data integration pipelines from many sources into Redshift, enhancing accessibility and reliability of data for analytics teams and improving data readiness by 40%.
- Constructed automated data quality checks using Python and PySpark, maintaining data integrity across all ETL processes.
- Created scalable AWS Lambda functions and Glue Jobs to process daily transactions, ensuring pipeline uptime.
- Established interactive dashboards in QuickSight/Tableau integrating Redshift and S3, enabling leadership to monitor key healthcare KPIs with 30% faster insights.
- Collaborated with teams to implement data governance and security policies on AWS, ensuring HIPAA compliance across all pipelines.
- Optimized healthcare claims data processing by implementing Delta Lake with AWS Glue and Databricks, reducing data duplication by 35% and improving query performance for downstream analytics teams.

Tech Mahindra   November 2018 – August 2021
Junior Data Engineer
Technologies: Azure Data Factory, Azure Synapse Analytics, Azure Data Lake Storage, Azure Monitor, Log Analytics, Azure Databricks, Apache Airflow, PySpark, Python, SQL (JOIN, CASE), SSIS, AWS RDS, Spark (DataFrames & Spark SQL), Tableau, Power BI, ANOVA, ETL, Cloud-native architecture, Snowflake, DBT

- Engineered end-to-end ETL pipelines on Azure Data Factory, ensuring seamless data flow and integration across many systems.
- Developed data ingestion frameworks from many sources into Azure Synapse Analytics, improving analytics data availability by 40%.
- Applied data validation checks using PySpark and Python, identifying and resolving inconsistencies before reporting.
- Optimized Azure Databricks notebooks for large-scale data transformations, enhancing pipeline performance by 25%.
- Configured and managed Azure Data Lake Storage and Blob Storage, ensuring highly secure and scalable data storage solutions.
- Executed monitoring and alerting using Azure Monitor and Log Analytics, reducing downtime and operational issues by 20%.
- Collaborated with stakeholders to define data models and KPIs, leveraging Power BI and SQL to deliver actionable business insights on time.

EDUCATION

Master of Science in Computer Science
Rivier University, Nashua, New Hampshire   January 2022 – August 2023

CERTIFICATIONS

- Achieved Databricks Certified Data Engineer Associate.
- Achieved AWS Certified Solutions Architect Associate.