

Naveen Kumar Reddy

Data Engineer

Tampa, Florida | [LinkedIn](#)

Summary

- Accomplished Data Engineer with 5+ years of experience in building, scaling, and optimizing large-scale data pipelines, ETL processes, and cloud-based data solutions to support enterprise data needs.
- Expertise in managing data lakes, warehouses, and databases, ensuring data integrity, security, and high performance for data-driven decision-making.
- Strong proficiency in cloud infrastructure services (AWS, Azure, GCP), leveraging resources like S3, Redshift, BigQuery, and Dataflow to build secure and cost-effective data platforms.
- Strong knowledge of SQL and NoSQL databases, including PostgreSQL, MongoDB, and DynamoDB, optimizing queries, and managing complex database architectures to support high-traffic data applications.
- Proficient in leveraging big data technologies like Hadoop, Spark, and Kafka to handle massive datasets and perform real-time data processing, resulting in a 40% reduction in processing time.
- Led initiatives to improve CI/CD workflows using Jenkins and GitLab CI, automating the deployment of data engineering solutions, reducing release times by 40%, and improving overall system reliability.
- Proficient in metadata management and data lineage tracking using tools like Apache Atlas, ensuring compliance with regulatory requirements and improving data transparency across the organization.

Key Skills

Languages:	Python, R, C, C++, MATLAB, Linux (Bash), SQL, NoSQL, Java
Libraries:	Pandas, NumPy, Keras, Tensorflow, PyTorch, Scikit-Learn, ggplot, and Matplotlib
Databases:	MySQL, SQL Server, Oracle, and Snowflake
Machine Learning:	Linear & Logistic Regression, Decision Tree, Random Forest, Hierarchical Clustering, Ad-Hoc Analysis, A/B testing, XGboost, NLP, SciPy, Opencv, AWS Sagemaker
Tools:	Power BI, MS Office Suite, Jupyter Notebook, ETL, and IDEs
EDA:	Data Modelling Cleaning of Data, Data Manipulation, Data Wrangling, Data Warehousing, and Visualization
Methodologies:	SDLC, Agile, Scrum, and Waterfall
Operating Systems:	Mac OS, Windows, and Linux
Data Engineering:	JIRA, Docker, CI/CD Jenkins, Git, Flask API
Deep Learning:	Keras, TensorFlow, Transformers, Diffusion models
Cloud/Big Data:	AWS, (S3, Lambda, Kinesis, Glue, EC2, ECR, Redshift), Azure (Azure Blob Storage, Azure Functions, Azure Event Hubs, Azure Data Factory, Azure Virtual Machines, Azure Container Registry, Azure Synapse Analytics), GCP, PySpark, Salesforce, MongoDB, HDFS, Hadoop, Map-Reduce, HBase, Apache Spark, Kafka, Hive & Pig
Data Visualization:	Tableau, Matplotlib, ggplot2, R-shiny, Plotly, ArcGIS, Seaborn

Professional Experience

ACL Digital, USA | DATA ENGINEER

Feb 2023 – Present

- Designed, developed, and maintained scalable and efficient data pipelines using Apache Airflow and AWS Step Functions, automating data ingestion processes from diverse sources, resulting in a 25% reduction in processing time.
- Migrated on-premise data warehouses to cloud platforms such as AWS Redshift and Google BigQuery, utilizing Azure Data Factory for orchestrating complex data workflows.
- Designed and implemented robust data models and schemas in Snowflake and Amazon Redshift, improving query performance and data accessibility.
- Developed and managed ETL processes to integrate data from various sources, using tools like Informatica and Talend to ensure data consistency and integrity.
- Conducted performance tuning and optimization for large datasets in Oracle and SQL Server, reducing query run times by 40%, leading to improved system efficiency.
- Designed and implemented large-scale data pipelines using Apache Spark and Hadoop to process terabytes of data, reducing data processing times by 35%.
- Developed real-time data processing applications using Apache Kafka and Spark Streaming, enabling low-latency data

ingestion and processing for time-sensitive analytics.

- Created comprehensive data visualizations and dashboards in Tableau and Power BI, providing actionable insights and enabling data-driven decision-making for stakeholders.
- Led the migration of data from an on-premise Teradata system to Google Cloud Platform (GCP), achieving a 50% reduction in infrastructure costs and enhanced scalability.
- Managed and optimized databases like PostgreSQL, MongoDB, and DynamoDB, ensuring high performance and minimizing query execution times by 30%.
- Integrated CI/CD pipelines with Jenkins and GitLab CI for automated testing and deployment of data engineering solutions, improving code quality and reducing deployment time by 40%.

Avenir Technology, India | DATA ENGINEER

Jun 2018 – Aug 2021

- Designed and deployed data lake house solutions integrating Delta Lake with Apache Spark, providing a unified data management platform combining data lakes and data warehouses.
- Established a comprehensive data catalog using tools like Alation and AWS Glue Data Catalog, facilitating easy data discovery and metadata management across the organization.
- Processed and analyzed geospatial data using tools like PostGIS and GeoPandas, enabling advanced spatial queries and location-based insights.
- Implemented real-time data streaming applications with Spark Streaming integrated with Kafka, improving event processing latency by 30%.
- Set up and maintain CI/CD pipelines using Jenkins and GitLab CI, enabling automated testing and deployment of data applications.
- Implemented metadata management solutions using tools like Apache Atlas, ensuring comprehensive documentation and easy data lineage tracking.
- Developed and executed data archiving and retention policies to manage data lifecycle and storage costs effectively, utilizing cloud solutions like AWS S3 and Google Cloud Storage.
- Designed and implemented data warehousing solutions on AWS, Azure, and GCP platforms, using services like AWS Redshift, Azure Synapse, and Google BigQuery, and developed robust ETL pipelines for data integration.
- Maintained MySQL, SQL Server, Oracle, Teradata, and PostgreSQL databases with 80% uptime.

Education

Master's in Computer Science | University of South Florida

(08/23/2021-12/09/2022)

Bachelor in Computer Science | SIR C R Reddy College of Engineering, India

(09/18/2014 - 05/30/2018)

Certifications

- AWS Certified Solutions Architect – Associate