MANIVARDHAN REDDY

Saint Louis, Mo | +1 (618) 517-5957 | manivardhanreddyreddem93@gmail.com | [LinkedIn](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

- Data Engineer with 4 years of experience designing, building, and operating scalable cloud data platforms, batch and streaming data pipelines, and analytics-ready data warehouses across AWS, Azure, and GCP environments.
- Strong expertise in SQL performance optimization, dimensional data modeling, Apache Spark, Hadoop, Kafka, Airflow, and cloud-native ETL/ELT frameworks, supporting high-volume analytical and operational workloads.
- Proven ability to deliver reliable, production-grade data systems through DataOps practices, CI/CD automation, and infrastructure-as-code, while enabling BI and analytics consumption through governed datasets.
- Experienced in building and maintaining batch and real-time data processing systems that support enterprise reporting, operational analytics, and downstream machine learning use cases.
- Recognized for translating complex business and platform requirements into cost-efficient, high-availability data architectures aligned with enterprise-scale data engineering standards.

TECHNICAL SKILLS

Programming & Querying: Python, SQL, PySpark, Scala, Bash

Data Engineering & Streaming: Apache Spark (SQL, Structured Streaming), Hadoop, Hive, Apache Kafka, AWS Kinesis

Databases & Warehousing: MySQL, PostgreSQL, SQL Server, Oracle, MongoDB, Amazon Redshift, Snowflake, Google BigQuery

ETL / ELT & Orchestration: Apache Airflow, AWS Glue, Databricks, dbt

Cloud Data Platforms: AWS (S3, Redshift, Glue, EMR, Lambda, EC2, Kinesis), GCP (BigQuery, Cloud Storage, Dataflow), Azure (Databricks, Data Lake, Synapse)

Data Modeling & Analytics Enablement: Dimensional Modeling, Star & Snowflake Schemas, Analytical SQL, Data Quality Frameworks

BI & Consumption: Tableau, Power BI, Looker

DataOps & Automation: Docker, CI/CD, GitHub, Jenkins, Terraform, Monitoring & Alerting

Delivery & Collaboration: Agile, Scrum, Jira, Confluence

PROFESSIONAL EXPERIENCE

Data Platform Engineer   **Feb 2025 - Present**
Juniper Networks

- Redesigned the MySQL analytics layer by introducing table partitioning, optimized indexing strategies, and SQL query refactoring, which improved query response times by 35% while supporting 500K+ daily analytical and operational queries.
- Implemented scalable batch ingestion pipelines using Python, PySpark, Apache Spark, and Hadoop to process 2M+ daily network telemetry records, ensuring reliable data availability and consistent schema enforcement across analytics workloads.
- Established analytics-ready dimensional models using star and snowflake schemas in SQL, enabling faster BI queries and significantly reducing repeated ad-hoc transformations across Tableau, Power BI, and Looker datasets.
- Introduced data quality validation and reconciliation logic within transformation workflows using Python and SQL, leading to a ~30% reduction in reporting inconsistencies and improved confidence in network performance metrics.
- Delivered near-real-time KPI datasets by implementing Spark Structured Streaming for selected telemetry feeds, improving metric freshness by 25% for operations teams responsible for network capacity and reliability planning.
- Strengthened DataOps reliability by managing SQL and Python pipeline code through GitHub-backed CI/CD workflows, resulting in more predictable releases and fewer manual intervention issues during deployments.
- Converted network engineering and operations requirements into production-grade data platform components through Agile collaboration in Jira and Confluence, accelerating delivery of analytics features tied to network health monitoring.
- Improved downstream analytics efficiency by publishing governed, performance-optimized datasets to BI tools, reducing dashboard refresh latency and lowering analyst dependence on custom extracts by over 20%.

Data Engineer   **Jul 2021 - Nov 2022**
Datamatics

- Built a scalable AWS data lake with Amazon S3, AWS Glue, and Amazon Redshift, centralizing 50TB+ of monthly data and enabling consistent, enterprise-wide analytics and reporting.
- Processing bottlenecks were eliminated by redesigning Apache Spark and PySpark batch pipelines, which cut overall transformation runtimes by 40% and improved data availability for downstream consumers.
- Reliable data delivery was achieved through Apache Airflow-orchestrated ETL and ELT workflows, where dependency management and retry controls sustained 99% pipeline success rates.
- Near-real-time analytics became possible by integrating Apache Kafka with Spark Streaming, increasing data freshness by 25% for operational dashboards and analytics use cases.
- Legacy warehouse workloads were modernized through migration to Amazon Redshift, where schema redesign and optimized sort and distribution keys delivered a 60% improvement in query performance.
- High availability across data stores was maintained by tuning PostgreSQL, MongoDB, and Redshift systems, strengthening performance stability and safeguarding data integrity.
- Deployment cycles accelerated by 18% after introducing GitHub-based version control and Jenkins CI/CD pipelines, reducing manual deployment errors across data engineering workflows.
- Analytics and data science teams received curated, analytics-ready datasets built with SQL and Spark, improving consistency and reliability for reporting, forecasting, and model development.

Data Analyst - Associate (Analytics Engineering Focus)   **Jul 2020 - Jun 2021**
3i Infotech

- Analytics adoption grew 28% after delivering SQL Server-backed datasets engineered with optimized SQL transformations that fed Tableau and Power BI reporting for sales, marketing, and operations.
- Performance bottlenecks were removed by indexing and refactoring complex SQL Server queries, cutting average query latency by 32% for recurring BI workloads.

- Business KPIs became consistent when 500K+ transaction records were analyzed with Python, Pandas, and NumPy, translating exploratory findings into standardized reporting logic.
- Reporting turnaround improved through SQL-based customer segmentation and behavior models, allowing reusable datasets to power multiple dashboards without repeated transformations.
- Manual effort declined after automating data preparation and transformations with SQL and Python, resulting in cleaner handoffs to downstream analytics consumers.
- Stakeholder questions were converted into scalable analytics datasets by aligning requirements with data platform standards, improving long-term maintainability of reporting logic.
- Onboarding time shortened once data models, SQL logic, and workflows were documented in Confluence, enabling faster ramp-up for analysts and junior engineers.
- Confidence in reports increased by embedding validation and reconciliation checks within SQL pipelines, supporting audit-ready outputs and reliable enterprise analytics.

Data Analyst Intern   **Jul 2018 - Dec 2018**
3i Infotech
- Manual processing effort dropped by 10+ hours per month after scripting recurring data tasks with Python and Java, allowing analytics teams to deliver reports faster and with fewer delays.
- Data accuracy increased by 20% when Oracle tables were validated and maintained using SQL checks, preventing inconsistent records from reaching reporting and analytics layers.
- Hands-on experience with large-scale data processing was gained by supporting Scala and Apache Spark transformations on Azure Databricks, enabling reliable batch ETL for enterprise datasets.
- Trend identification accelerated through exploratory analysis built with Python and Plotly, helping analysts surface anomalies and patterns earlier in the analysis lifecycle.
- Analytics workloads became more scalable by assisting with Azure-based data platform deployments, supporting storage and compute configurations used for reporting and transformation pipelines.
- Repeated data issues were reduced by collaborating with senior engineers to resolve SQL and Spark pipeline quality and performance problems, improving overall pipeline stability.

PROJECTS
Network Performance Analytics & Database Optimization Platform
- Built scalable analytics pipelines using Apache Spark, SQL, and Python to process high-volume network latency, throughput, and traffic logs, enabling reliable performance monitoring across large telemetry datasets.
- Query performance improved by 40% after redesigning MySQL schemas, indexing strategies, and execution paths with analytical SQL, supporting faster access to mission-critical network metrics.
- Operational visibility increased by publishing optimized datasets to Tableau and Power BI, enabling faster incident detection and root-cause analysis for network reliability teams.

Cloud Data Platform Migration & Real-Time Streaming Architecture
- Migrated enterprise ETL workflows to an AWS-native data platform using Amazon S3, AWS Glue, EMR, Lambda, and Redshift, improving scalability and standardizing cloud-based data processing.
- Enabled sub-minute analytics by implementing event-driven pipelines with AWS Kinesis and Spark Streaming, improving timeliness of operational metrics for downstream consumers.
- Deployment consistency improved through Terraform-based infrastructure provisioning and CI/CD pipelines, reducing manual configuration drift and supporting cost-optimized cloud operations.

Enterprise Customer Analytics & BI Enablement
- Unified CRM, marketing, and transaction data by designing analytics-ready dimensional models with SQL, creating a centralized warehouse optimized for analytical queries.
- Customer insights became more actionable after delivering curated datasets to Tableau, Power BI, and Looker, supporting churn, engagement, and performance analysis.
- Analytics turnaround time decreased after automating KPI refresh workflows using SQL and scheduled pipelines, enabling leadership teams to access up-to-date metrics consistently.

Data Automation & Enterprise Reporting Framework
- Reduced manual reporting effort by developing modular ETL and ELT pipelines with Python, SQL, and Apache Airflow, ensuring reliable and repeatable data transformations.
- Reporting consistency improved through standardized data models and reusable analytical layers, allowing multiple teams to consume governed datasets without duplication.
- Operational visibility expanded by implementing scheduled data pipelines that delivered daily, weekly, and monthly insights aligned with executive and operational reporting needs.

EDUCATION
Master of Science in Information Systems
Belhaven University
Bachelor of Science in Computer Science
Chaudhary Charan Singh University

CERTIFICATIONS
- Google Professional Data Engineer - **Coursera**
- IBM Data Engineering Professional Certificate - **Coursera**
- DeepLearning.AI Data Engineering Professional Certificate - **Coursera**
- Data Engineering Foundations - **LinkedIn Learning**
- Modern Data Engineering with Snowflake - **LinkedIn Learning**
- AI Fundamentals - **DataCamp**
- Certified ScrumMaster (CSM) - **Scrum Alliance**