# VENKATA PRANEETH THOTA

## **Data Engineer**

노 +1 (901) 799 - 0900 🛛 thotavpraneeth@gmail.com 🔗 LinkedIn 🔗 https://github.com/tpc0809

# EXPERIENCE

### Data Engineer

#### **Tenet Healthcare**

**m** 09/2022 - 01/1970

Healthcare organization

- Developed and optimized ETL pipelines using Python, Apache Spark, PySpark, SQL, and AWS (S3, Lambda, Glue), processing 1TB+ data weekly and reducing execution time by 30%, improving data reliability and scalability.
- Implemented scalable data architectures leveraging AWS Redshift and Azure Synapse Analytics, reducing cloud costs by 20%
- Designed real-time streaming solutions using AWS Kinesis and Lambda, achieving 99.9% system uptime and reducing latency by 35%
- · Automated complex ETL workflows using Apache Airflow, improving reliability and reducing manual workload by 40%
- · Optimized Redshift queries and indexing strategies, reducing SQL execution time by 25%
- · Designed a machine learning pipeline using AWS SageMaker to improve patient outcome forecasting by 20%
- · Implemented robust data governance and security measures ensuring HIPAA compliance
- Developed interactive dashboards using AWS QuickSight, Power BI, and Tableau for KPI monitoring, business analytics, and data visualization, integrating data from AWS Redshift, Snowflake, and SQL databases to enhance decision-making and performance tracking.
- · Designed and integrated graph-based analytics using Neo4j to enhance medical treatment pattern recognition
- Orchestrated multi-cloud data integration strategies across AWS and Azure to improve data interoperability
- Engineered cost-efficient storage management using AWS S3 lifecycle policies and automated access control with IAM
- Developed real-time data validation and monitoring tools using Great Expectations, improving data quality by 99.5%

### **Big Data Engineer**

#### Infosys

**i** 10/2018 - 12/2020

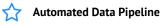
Multi-national corporation specializing in information technology services

- Designed and optimized Spark-based data pipelines using Python, Databricks, AWS Glue, and Snowflake, enhancing processing efficiency by 35%
- Automated ETL workflows using Apache Airflow and Informatica, reducing data transformation time by 30%
- · Implemented real-time data ingestion with Kafka, Flink, and Java, improving data availability and reducing latency
- Led the migration of legacy databases to Snowflake, optimizing data warehousing by leveraging Snowflake's scalability, partitioning, and compression techniques, reducing storage costs by 25% and query times by 20%, while enhancing performance, security, and data governance.
- Optimized AWS Athena queries and metadata management, enhancing serverless querying efficiency
- Developed CI/CD pipelines with Terraform, Kubernetes, and Jenkins, ensuring seamless data pipeline deployments
- Integrated predictive analytics models into ETL pipelines using AWS SageMaker and MLflow
- Designed and developed real-time, interactive analytics dashboards using AWS QuickSight and Tableau, transforming complex data into actionable business intelligence insights and enhancing strategic decision-making.
- Implemented advanced data lake architectures using AWS S3 and Snowflake, improving data accessibility and processing speeds
- · Enhanced security and compliance frameworks using IAM-based access control policies and encryption standards

# SUMMARY

Results-driven Data Engineer with 4+ years of experience in big data processing, ETL development, real-time streaming, and cloud-based data solutions. Skilled in designing and optimizing high-performance data pipelines, with deep expertise in AWS and Azure services. Passionate about building scalable data ecosystems that enable actionable insights, drive business growth, and support innovation...

# **KEY ACHIEVEMENTS**



Developed a fully automated data pipeline leveraging AWS Lambda, Kinesis, and Flink to support real-time processing



#### **Anomaly Detection System**

Built a machine learning-based anomaly detection system using AWS SageMaker, improving fraud detection rates

#### **ETL Cost Optimization**

Optimized ETL cost efficiencies by implementing AWS Glue and Redshift Spectrum, reducing cloud spend by 20%



#### **Automated Infrastructure Deployment**

Implemented CI/CD pipelines for data infrastructure using Terraform, Jenkins, and Git, reducing deployment time by 30%.



# **EDUCATION**

# Master of Science in Management Information Systems

## University of Memphis

**m** 01/2021 - 08/2022

## Bachelor of Technology in Information Technology

## V. R. Siddhartha Engineering College

**iii** 06/2016 - 05/2020

# SKILLS

Airflow Ansible Apache	
Apache Kafka Apache Spark Athena	_
Python SQL Scala Shell Scripting	]
Apache Spark PySpark Kafka Flir	ık
Airflow Informatica AWS Glue	
Databricks Azure Data Factory AWS	\$3
AWS Redshift AWS Glue AWS Lambd	a
AWS EC2 AWS Kinesis AWS QuickSig	ht
AWS IAM Azure Data Factory	
Azure Synapse Analytics Azure ML	
PostgreSQL MySQL MongoDB	
Cassandra DynamoDB Neo4j	
Snowflake Amazon Redshift Tableau	
Power BI AWS QuickSight Hadoop	
HDFS Docker Kubernetes Terrafo	orm
Jenkins GitHub AWS SageMaker	Jira
DBT Zookeeper FastAPI	