

# CHARAN TATINENI

LinkedIn://charantatineni • Github://charantatineni • tatineni.dev • charantatineni11@gmail.com

## EXPERIENCE

### GOODRX

#### *Data Engineer*

*Santa Clara (Remote), CA*

(Aug 2025 -- Present)

#### **Data Infra and Platform Team**

- Own design and delivery of **production pipelines** using **PySpark, Airflow, and Databricks**, producing analytics-ready datasets that powered product, engineering, and finance decisions across pricing and operational domains.
- Design ingestion and normalization pipelines to normalize and redact raw vendor feeds from vendor S3 buckets, establishing a stable foundation for downstream **DAG orchestration** and business-critical reporting.
- Implement low-latency ingestion patterns using **zero-ETL and Kafka**, processing over 1B+ records weekly to continuously sync operational datasets and reduce data availability from hours to minutes for product analytics.
- Provision AWS and Databricks infrastructure via **Terraform**, enforcing IAM-based access control, CI/CD pipelines and environment isolation while enabling reproducible deployments across dev and prod.

#### **Vendor Data and Data Marts**

- Architected Medallion data models** and dataset contracts (schema guarantees, freshness SLAs, completeness checks), transforming fragmented raw data into trusted **domain data marts** and accelerating ad-hoc analysis for stakeholders.
- Led cross-functional design reviews to validate business logic and orchestrated the onboarding of **Databricks** into the AWS ecosystem, establishing the infrastructure standards and metric definitions required for production-ready DAGs.

### HYPERWATER.AI

#### *Data Engineer*

*Dallas, TX*

(May 2024 -- Aug 2025)

- Built Ingestion services and REST **APIs** to onboard vendor data into AWS S3, enforcing schema contracts, idempotency, and validation gates to deliver reliable, analytics-ready datasets.
- Implemented centralized **logging, monitoring**, and alerting across pipelines using CloudWatch and custom metrics, reducing incident detection time by 40% and preventing downstream data corruption through automated quality gates.
- Developed modular, configuration-driven **ETL, ELT** pipelines and automated Codefresh **CI/CD** workflows, maintaining 99.99% data availability and meeting strict 15-minute freshness SLAs for business-critical dashboards.

### GEORGE MASON UNIVERSITY

#### *Database Administrator*

*Fairfax, VA*

(Sep 2023 -- May 2024)

- Administered **hybrid cloud infrastructure** and engineered Python ETL pipelines to automate the ingestion and normalization of healthcare research data, optimizing **database performance** and accelerating data availability for clinical teams.

## SKILLS

### DATA ENGINEERING & CLOUD

- Python, SQL, Java, PySpark (Spark SQL), Databricks (Delta Lake, Iceberg), Medallion Architecture, AWS (S3, Redshift, EMR, Kafka(Confluent), DynamoDB)

### ORCHESTRATION, DEVOPS AND TOOLS

- Apache Airflow (Astronomer), DBT, Terraform, GitHub Actions, Docker, Monte Carlo, Fivetran, Looker

## EDUCATION

### GEORGE MASON UNIVERSITY

#### *M.S. in Computer Science*

*Fairfax, VA*

GPA: 3.73 / 4.0

Graduate coursework in Databases, Distributed Systems, Big Data Processing, Machine Learning, and Advanced Algorithms

### JNTU-K INDIA

#### *B.S. in Computer Science*

*India*

GPA: 3.5 / 4.0

Coursework includes Software Engineering, Object-Oriented Programming, Systems Programming, API Design, Algorithms