

NITHEESH CHANAMBATLA

DATA ENGINEER | GEOSPATIAL DATA ENGINEERING

Dayton, Ohio | (937) 907-2325 | nitheeshc2325@gmail.com

[LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Engineer with 4+ years of experience building scalable batch, streaming, and cloud data platforms across financial services and healthcare environments. Hands-on expertise in Python, SQL, Spark, Airflow, Kafka, PostgreSQL/PostGIS, and AWS, with emphasis on data quality, governance, observability, and production reliability. Independently built and verified a NASA FIRMS geospatial platform that demonstrates remote-sensing data engineering, deterministic lineage, replay-neutral analytics, and controlled cloud execution.

TECHNICAL SKILLS

Languages: Python, SQL, Java, Scala, Shell, JavaScript, HTML/CSS

Data Engineering: Apache Spark, Apache Kafka, Airflow, Hadoop, HDFS, YARN, Hive, Presto, AWS Glue, Informatica, ETL/ELT, batch and streaming pipelines, CDC

Cloud & Data Platforms: AWS EMR Serverless, S3, Glue Data Catalog, Athena, Kinesis, Data Firehose, Lambda, Redshift, IAM, KMS, CloudFormation, CloudWatch

Databases & Modeling: PostgreSQL, PostGIS, SQLite, dimensional modeling, star/snowflake schemas, SCD Type 2, partitioning, bucketing, indexing

Geospatial / Remote Sensing: NASA FIRMS, satellite-derived active-fire observations, geospatial event processing, EPSG:4326/WGS84, spatial indexing, bounding-box queries, acquisition-time semantics, replay-neutral spatial-temporal aggregation, lineage/provenance

Reliability & Delivery: Data validation, reconciliation, idempotency, retries, failure recovery, quarantine, CI/CD, Git, Jenkins, Docker, ELK, root-cause analysis

Web / Portfolio: Node.js, Vercel, Supabase, Zod, responsive design, accessibility, Content Security Policy

PROFESSIONAL EXPERIENCE

Data Engineer | Discover (Client Engagement) | Chicago, Illinois | 2023-Present

- Built and maintained Python, Airflow, and Spark ETL/ELT pipelines supporting enterprise analytics and machine-learning datasets.
- Designed an AWS data lake integrating S3, Glue Data Catalog, and Athena across structured and unstructured sources; implemented Kinesis, Data Firehose, and Lambda ingestion for real-time analytics and alerting.
- Migrated analytical workloads to Amazon Redshift and optimized dimensional structures and query patterns for BI consumption.
- Automated testing and delivery with Git, Jenkins, and Docker; monitored failures, latency, and data quality through CloudWatch and ELK.
- Applied IAM, KMS encryption, RBAC, lineage, and cataloging practices while supporting GDPR- and HIPAA-aligned data handling.

Data Engineer | Cigna | Bengaluru, India | 2021-2022

- Developed Spark, Hive, and Presto batch pipelines for large-scale financial transaction data in a healthcare client environment.
- Designed star and snowflake schemas and applied partitioning, bucketing, and indexing for analytical workloads.
- Built Python frameworks for ETL validation, error handling, reconciliation, and performance monitoring across HDFS and YARN workloads.
- Integrated CDC with Debezium and Kafka Connect; embedded cataloging, classification, and retention controls and supported production incident response.

SELECTED PROJECTS

NASA Earth Observation Event Intelligence Platform | GitHub

- Independently designed and built a governed NASA FIRMS geospatial platform with deterministic event, detection, and lineage identities, replay-neutral scientific aggregates, and reproducible batch and streaming contracts.
- Verified the complete local platform with 1,000,000 replay events from 10,000 real detections across Kafka, Spark, Parquet Silver/Gold, PostgreSQL/PostGIS, Airflow, read-only FastAPI, bounded caching, and Streamlit; measured Spark throughput at 6,688.88 events/second.
- Processed 5,000,000 replay events on AWS EMR Serverless at 8,734.80 events/second with 5,000,000 accepted, zero rejected, zero duplicates, and 10,000 replay-neutral detections; an independent verifier reconciled lineage/scientific aggregates and matched 82 Silver/Gold object sizes and SHA-256 checksums.
- Deployed a private CloudFormation-managed AWS foundation with S3/KMS, private subnets and VPC endpoints, least-privilege IAM, CloudWatch, and Budget/SNS controls; removed temporary execution authority and completed controlled teardown after evidence preservation.
- Verified deterministic local generation and independent read-back at 10,000,000 records; local Spark processing at that scale exceeded the measured JVM heap boundary, and AWS 10M was not executed.

Customer360 Analytics Warehouse

- Built a PostgreSQL dimensional warehouse loading 400,000 sales facts across five source domains with tested SCD Type 2 customer and product dimensions.
- Implemented surrogate-key resolution, validation, quarantine, constraints, repeat-safe loading, 20 analytical SQL queries, and 20 automated tests plus hosted CI.

EDUCATION & CERTIFICATIONS

Master's in Data Science and Artificial Intelligence | Campbellsville University | Dec 2024

- AWS Certified Data Analytics - Specialty
- AWS Certified Solutions Architect - Associate
- Google Cloud Professional Data Engineer