Eegapuri Lokeshwar Reddy

Data Engineer

9959242262 ◇ elokesh4292@gmail.com ◇ Telangana, India ◇ LinkedIn ◇ GitHub ◇ Website

SUMMARY

Data Engineer with 3 years of experience in ETL operations, data pipeline automation, and enterprise data warehousing. Proven ability to monitor, troubleshoot, and optimize large-scale data workflows in production and disaster recovery environments. Hands-on expertise in dimensional modeling, OLAP, datamarts, and workflow orchestration, with skills honed through EDW experience and a strong foundation in scalable data infrastructure. Passionate about development and automation, leveraging modern Big Data tools to build efficient data pipelines. Combines operational discipline with a problem-solving mindset to design robust, high-performance data solutions. Eager to apply deep operational knowledge and engineering skills to develop innovative, data-driven systems.

WORK EXPERIENCE

System Engineer   Apr '19 — Feb '22
Tata Consultacy Service (TCS)   Hyderabad, India

Project 1: STC (Saudi Telecom Company) – Enterprise Data Warehouse (EDW) Operations
Environment: Production & Disaster Recovery(DR)
Tools: Ab Initio (ETL), Teradata (EDW), Red Hat Linux, SQL, Shell Scripting, Jira, Remedy

- Monitored **2000+** enterprise-scale **ETL jobs** across OLAP and data mart environments; responded to daily email requests and escalations from source and reporting teams to ensure uninterrupted data flow.
- Oversaw data ingestion from multiple upstream systems (including CRM, CDR, and IRB platforms) using both pull (SFTP, SQL extraction) and push (file transfers) methods; implemented automation to reduce manual interventions.
- Handled **200+ Change Requests (CRs)** via Jira, including **45+ new ETL Jobs** deployment, optimizations, data reloads, and job decommissions; gained strong understanding of pipeline and SQL-level logic within **distributed, MPP-based systems**, maintained a **flawless deployment track record**, and helped prevent multiple potential SLA breaches through careful planning and execution.
- Developed and optimized SQL queries to monitor daily business-critical KPIs, including postpaid, prepaid, landline, and sales metrics; automated discrepancy reporting and resolution tracking.
- Investigated and resolved critical system issues such as database locks, CPU spikes, schema mismatches, connectivity errors, and data rejections; coordinated with developers for patching and long-term fixes.
- Developed **shell scripts** to monitor system resources, detect transformation failures, and alert teams. **Proactively** identified and cleaned unwanted files, recovering **30% storage space** and improving operational efficiency by **35%.**
- Led real-time communication during escalations, tracked incidents via Jira and Remedy, and ensured internal and client teams received timely updates.
- Reprocessed rejected data records and validated fixes to ensure **100% data completeness** post-resolution.
- Participated in **4–5 disaster recovery (DR) drills** as part of a cross-functional team, replicating live production environments to validate failover processes.
- Tested and verified critical components — including job workflows, source connections, scheduling configurations, and stakeholder access mappings — to ensure seamless business continuity during disruptions.
- Created and maintained documentation of all ETL workflows, data dependencies, run schedules, and recovery steps to support smooth DR execution.
- Managed pre- and post-DR communications, coordinated system validations, and confirmed data consistency and readiness across environments.

Project 2: STC (Saudi Telecom Company) –Kafka-Based Event Streaming Operations
Environment: Production
Tools: Confluent Kafka, Python, Kafka Client Libraries, ksqlDB, Avro, JSON.

- Managed operations and maintenance of **250+ Kafka topics**, supporting real-time event streaming alongside EDW batch workflows.
- Handled **Avro and JSON** message formats to ensure schema compatibility between producers and consumers, maintaining downstream data integrity.
- Developed Python scripts using Kafka client libraries to automate health checks, monitor consumer lag, and send real-time alerts — significantly reducing manual monitoring.
- Resolved connector issues related to **JDBC/ODBC** and **HTTP** integrations, including connectivity failures, SSL/certificate errors, schema mismatches, and data type casting challenges that impacted reliable data delivery.
- Addressed core Kafka issues including RBAC permission errors, client-side driver mismatches, producer/consumer configuration tuning for performance, and **schema evolution conflicts** on topic fields.

EDUCATION

B-Tech in Mechanical Engineering, CVR College of Engineering, Hyderabad (GPA: 9.56)   2015 — 2019
    Hyderabad, India

Intermediate in MPC, Narayana Junior College (GPA: 9.7)   2013 — 2015
    Hyderabad, India

CERTIFICATIONS

Confluent Fundamentals Accreditation, Confluent

Academy Accreditation - Databricks Fundamentals, Databricks

IBM Data Engineering Professional Certificate, IBM (In Progress)

ACHIEVEMENTS & AWARDS

- On the **Spot Award** for outstanding performance and exceptional contributions to **ETL operations** and **system optimization**.
- Received **client appreciation** for ensuring seamless system operations and maintaining **100% KPI availability** with **zero SLA breaches**.
- Cleared the first round of **TCS CodeVita Season 7 (2018)**, a global coding competition, demonstrating strong programming and problem-solving skills.

PERSONAL PROJECTS

RentLok: Real-Time Rental Management System | Link
- **Tech Stack:** Android, FastAPI, PostgreSQL, Apache Kafka, Kafka Connect, ksqlDB, SQLAlchemy, Pydantic, JSON
  - Built an end-to-end **OLTP** data pipeline for a rental management system with a native **Android app** and backend powered by **FastAPI**, **PostgreSQL**, and **Apache Kafka**.
  - Developed secure RESTful **APIs** over **HTTPS** to process **JSON** data, enforced schema validation with **Pydantic**, and usedKafka Connect to stream PostgreSQL data into Kafka topics.
  - Implemented **ksqlDB** for real-time transformations and aggregations, enabling live in-app metrics; optimized the **PostgreSQL** schema for low-latency access and modular microservice deployment.
- **Key Learning:**
  Successfully transitioned from OLAP to OLTP system design, demonstrating the ability to build reliable, API-driven, real-time data pipelines and translate business requirements into scalable, end-to-end data solutions.

Telecom Data Pipeline | Link
- **Tech Stack:** Apache Airflow, Apache Spark (PySpark), PostgreSQL, Shell Scripting, Python
  - Designed a batch processing **ETL pipeline** to process millions of CDR records using **Airflow** for orchestration and **PySpark** for scalable, distributed transformations.
  - Developed fraud detection logic to identify suspicious patterns like high-frequency calls and short-duration international interactions.
  - Implemented audit logging and error tracking, improving pipeline observability and operational reliability.
- **Key Learning:**
  Demonstrated the ability to design scalable, distributed data transformations using PySpark, while enhancing reliability through fraud detection logic and robust logging.

E-Commerce Data Pipeline | Link
- **Tech Stack:** Apache Airflow, MySQL, PostgreSQL, Foreign Data Wrapper (FDW), Python, Pandas, SQL
  - Built a modular batch ETL pipeline using Airflow to automate CSV ingestion and transformations across MySQL and PostgreSQL.
  - Enabled cross-database querying using PostgreSQL FDW, secured credentials with .my.cnf and .pgpass, and optimized performance through indexing.
  - Exported cleaned data for downstream analysis, showcasing a reusable and scalable data integration workflow.
- **Key Learning:**
  Strengthened skills in Airflow orchestration, relational database data ingestion, and secure, cross-database batch processing aligned with real-world data engineering practices.

Game Development & 3D Modeling | Link
- **Tech Stack:** Unity, C#, Blender
  - Developed a PC game (**"Cubon"**) using **Unity** and **C#**, implementing player controls, interactive mechanics, and UI features.
  - Created and animated 3D models in **Blender**, applied texturing and rendering, and integrated assets into the Unity engine.
- **Key Learning:**
  Gained hands-on experience in software development, event-driven programming, and 3D rendering pipelines — combining creative and technical skills foundational to full-cycle product development.

SKILLS

Tools/Platforms: Apache Airflow, Ab Initio, Confluent Kafka, Hadoop (HDFS, YARN), Apache Spark (SparkSQL, PySpark), Data Modeling, Dimensional Modeling, ERD, Data Warehousing, JIRA, Remedy, Red Hat Linux, Databricks, Git, CI/CD, Looker Studio, Teradata, PostgreSQL, AWS S3, AWS Glue

Programming Languages: Python, SQL, Shell Scripting, ksqlDB

Soft Skills: Problem Solving, Communication, Analytical Thinking, Adaptability, Attention to Detail, Risk Management, Decision-Making, Emotional discipline, Patience, Reliability, Knowledge Sharing