

Sanjeevi M

Chennai, Tamil Nadu, India – 600019. | Mob: +91 8610889030 | sanjayvm66@gmail.com

LinkedIn: www.linkedin.com/in/sanjeevi-m | GitHub: <https://github.com/askmrsanjay>

Tableau Public: <https://public.tableau.com/app/profile/sanjeevi.m/vizzes>

Website: [https://askmrsanjay.github.io/sanjeevi data engineer online resume/](https://askmrsanjay.github.io/sanjeevi%20data%20engineer%20online%20resume/)

Professional Summary

Azure & Databricks-certified Data Engineer with 4 years of experience designing, automating, and optimizing scalable ETL/ELT pipelines using PySpark, Snowflake, Delta Lake, and Azure cloud services. Proven track record of delivering enterprise-grade data solutions for Fortune 500 clients including Humana, Nike, and Apple. Skilled in building metadata-driven frameworks and improving pipeline performance.

Technical Skills

- **Languages:** Python (Advanced), SQL, (**Vibe Coding Languages :** PowerShell, Bash, Java Script, HTML/CSS)
- **Big Data & Frameworks:** PySpark, Apache Spark, Apache Kafka, Delta Lake, Delta Live Tables, Apache Iceberg, Apache Airflow, Snowpark
- **ETL & Orchestration:** Azure Data Factory, Informatica, Databricks Workflows, Pandas, NumPy, Great Expectations
- **Databases & Storage:** Snowflake, Teradata, Oracle, Elasticsearch, ADLS, Unity Catalog, Parquet
- **Cloud & DevOps:** Azure (ADLS, Key Vault, ADF), Docker, Git, GitHub, Azure DevOps
- **Visualization & Monitoring:** Tableau Desktop/Server, Prometheus, Grafana, Kibana, Streamlit

Professional Experience

COGNIZANT TECHNOLOGY SOLUTIONS | ASSOCIATE DATA ENGINEER | JUL 2025 – PRESENT

- Build and optimize **cloud-native ETL pipelines** using Azure Databricks, PySpark, and Delta Lake for enterprise-scale workloads.
- Design **metadata-driven workflows** to streamline data ingestion, improve reusability, and reduce development effort.
- Implement **Delta Lake ACID pipelines**, supporting incremental loads, schema evolution, and time-travel.
- Monitor and tune Spark jobs to **improve performance**, reduce compute costs, and ensure reliability.

INFOSYS LTD | TECHNOLOGY ANALYST | MAY 2022 – JUL 2025

- Designed and implemented **end-to-end ETL pipelines in Azure Databricks**, improving performance through PySpark optimizations and Delta Lake tuning for datasets exceeding 20M records.
- Developed ingestion workflows for **structured (Teradata)** and **unstructured (JSON, Parquet)** data with automated data quality checks.
- Engineered **CDC (Change Data Capture)** solutions for customer billing and sales order pipelines achieving <1% latency and 99.9% uptime.
- Built automated reporting pipelines with Slack integration, eliminating manual Tableau refreshes and reducing reporting time.
- Collaborated with Apple's Data Operations team to design dashboards for anomaly detection, pipeline health monitoring, and KPI tracking.

Professional Projects

Snowpark-Based Dynamic Rule Processing Framework

Data Engineer | Humana Inc. | Technologies: Python (Advanced), SQL, PySpark, Databricks, Snowflake, ADLS, Delta Lake

- Designed and developed a **config-driven Snowpark framework** that dynamically reads transformation and validation rules from JSON, enabling flexible, code-light processing.
- Implemented **dynamic rule parsing** for dataset mappings, join logic, transformation sequences, and parameters.
- Engineered Snowpark pipelines to **load, combine, and transform** data using joins, unions, and rule-based logic.
- Built a **sequential rule execution engine** ensuring deterministic and fully auditable workflows.
- Developed an **audit logging module** capturing rule name, datasets, parameters, timestamps, initiator, and execution status.
- Implemented robust **error logging for exceptions, validation failures**, and diagnostic information.
- Enabled full **extensibility** by making rule updates driven through JSON without code changes or redeployment.
- Ensured operational transparency and governance compliance through end-to-end traceability.

Databricks – Snowflake Dynamic Extract Framework

Data Engineer | Humana Inc.

- Built a **parameter-driven, config-based PySpark framework** to dynamically process any system + file ID combination without code changes.
- Automated metadata retrieval from **Snowflake control tables** (ADLS paths, target tables, transformation SQL).
- Developed ingestion pipelines to extract Snowflake data into **Delta format on ADLS**.
- Automated **Unity Catalog External Table creation** for seamless SQL access.
- Implemented ACID-compliant **INSERT...SELECT** pipelines loading into managed Delta tables.
- Developed a centralized **Audit & Error Logging module**.
- Secured pipelines using **Azure Key Vault / Secret Scope**.
- Deployed using **Databricks Jobs & Workflows** ensuring reliability.

Heritage EMEA Billing & Sales Order Pipelines

Data Engineer | Nike Inc. | Technologies: Python (Advanced), SQL, PySpark, Databricks,

- Gathered, cleansed, and processed customer billing and sales order data from **Teradata** for EMEA analytics.
- Designed **Informatica workflows** for EMEA pipeline objects.
- Created and tested **CDC SQL queries enabling** accurate incremental data updates.
- Implemented multilayer validation with automated email-based verification reports.
- **Achieved <1% latency and 99.9% EMEA pipeline uptime.**

Additional Personal Projects

Real-Time E-Commerce Data Lakehouse:

Github: https://github.com/askmrsanjay/data_engineering_projects.git

Built medallion architecture (Bronze/Silver/Gold) data lakehouse using Apache Iceberg, Kafka, Spark Streaming, MinIO, and Airflow with Great Expectations for data quality validation

Real-Time Analytics Pipeline:

Github : <https://github.com/askmrsanjay/Real-Time-E-Commerce-Analytics-Pipeline>

- Developed end-to-end streaming analytics with Kafka, PySpark, Elasticsearch, and Kibana for processing high-throughput e-commerce transactions

Career Flow Engine:

Github : https://github.com/askmrsanjay/Product_company_job_alert.git

- Engineered automated web scraping pipeline using Selenium, BeautifulSoup, and Pandas to extract and analyze LinkedIn job market data across 15+ global markets

Education

POST GRADUATE STATISTICS | MAY 2019 | PRESIDENCY COLLEGE, CHENNAI

Certifications

- Databricks Certified Data Engineer Associate
- Microsoft Azure Data Engineer Associate
- Infosys Certified PySpark Developer
- Infosys Certified Python Programmer
- Infosys Certified Tableau Developer