Darpan Jiyani

+1 408-590-8885 | darpanjiyani78@gmail.com | linkedin.com/in/darpanjiyani | github.com/darpanjiyani | San Jose, CA, USA

SUMMARY

Data Engineer building batch and streaming pipelines with PySpark, Databricks, Delta Lake, Airflow, and AWS. Unified RF test data from 60+ stations into a single Delta Lake table and cut a Power BI refresh from 15 minutes to under 15 seconds on 2M+ records through incremental ETL. Also builds LLM data tools, including a multi-agent text-to-SQL system at 99.3% execution accuracy.

SKILLS

Languages: Python, SQL
Data Engineering: ETL/ELT, PySpark, Spark SQL, Spark Structured Streaming, Kafka, Airflow, Databricks, Delta Lake
Cloud & Databases: AWS (S3, EMR, Redshift, Lambda, EventBridge), Azure, SQL Server, PostgreSQL
Modeling & Optimization: Star schema, SCD Type 2, CDC, schema evolution, partitioning, query tuning
AI & LLM: LangGraph, LangChain, multi-agent systems, Claude API, OpenAI API, RAG, ChromaDB, LLM evaluation
BI & DevOps: Power BI (DAX), Tableau, Streamlit, Power Automate, Git, Docker, GitHub Actions

WORK EXPERIENCE

Data Engineer - UL Solutions | Fremont, CA | May 2026 – Present

- Automated RF test-data collection and visualization with Python across 60+ test stations, reducing end-to-end test-cycle analysis from 2 days of Excel-based processing to 30 minutes
- Delivered a Python and React validation app that analyzes 1M+ RF test records and 1,000+ plots per test, cutting manual review from ~30 minutes to ~5 seconds
- Handled schema drift across 12+ RF test-case formats with differing column sets via Delta Lake mergeSchema, unifying them into a single queryable Delta table without breaking pipelines on new test types
- Wrote PySpark transformations and Spark SQL with joins and window functions; date partitioning cut RF query time 20%
- Developed 5+ Tableau dashboards on Delta tables surfacing RF test results to 3 client teams and 20+ internal engineers

Data Science Intern - Keysight Technologies | Santa Rosa, CA | May 2025 – December 2025

- Reduced Power BI dashboard refresh time from 15 min to <15 sec on 2M+ records via incremental ETL with timestamp-based change detection, enabling near real-time wafer monitoring for a 30-person team
- Deployed Dockerized Airflow DAGs on 15-min schedules with task retries and email failure alerts
- Designed a Power BI monitoring dashboard showing tool status (up/down/at-limit) across three production lines, helping operators identify at-limit equipment before downtime
- Implemented SQL status-change detection with Python alerts, cutting operator notification from 2–3 hours to under 5 minutes
- Streamlined shift reporting for 50+ supervisors with Power Automate, saving 30+ min/shift and eliminating formatting errors
- Shipped a Power BI dashboard (DAX) surfacing underperforming shifts against rolling baselines, replacing manual review with targeted root-cause investigation

Data Engineer - Cerebulb | Gujarat, India | January 2022 – December 2023

- Engineered an EventBridge-scheduled Lambda pipeline to fetch telemetry from 100+ boiler sensor APIs every minute and store it in S3 for near real-time monitoring
- Transformed 500GB+ of raw sensor data from S3 to Redshift via daily PySpark jobs on EMR for analyst dashboards
- Modeled boiler KPIs into a Redshift star schema with fact and dimension tables for Tableau reporting
- Partnered with plant managers on Tableau monitoring dashboards that flagged failures earlier and cut unplanned downtime

PROJECTS

AI Data Agent | Python, LangGraph, Streamlit, Claude, PostgreSQL, Pandas, sqlglot | Code | Live Demo

- Built a LangGraph multi-agent SQL/ETL app with memory (follow-ups right in 32/32 chats) and prompt caching (-59% tokens)
- Raised text-to-SQL accuracy from 89.4% to 99.3% on 47 questions (3-run avg) with schema notes, SQL rules, and retry loops
- Built Open-Meteo to PostgreSQL ETL agent with human-approved, hash-checked loads; weather-query accuracy 0/12 to 11/12
- Blocked 22/22 SQL writes (sqlglot, read-only role); sandboxed AI pandas via AST allow-list and audit hooks

Real-Time Ride Analytics Pipeline | FastAPI, Azure Event Hubs, Kafka, Spark Structured Streaming, Delta Lake | Code

- Streamed booking events from a FastAPI service to Azure Event Hubs, consumed via Kafka with Spark Structured Streaming
- Merged bulk backfill and live Kafka streams into one Delta table using Lakeflow Declarative Pipelines append flows
- Joined live rides to six lookup tables with a watermark, and built a star schema via Auto CDC with SCD Type 2 city history

EDUCATION

Master of Science in Data Science | San Jose State University, San Jose, CA | December 2025
Coursework**:** Big Data Technology, Machine Learning, Deep Learning, Generative AI, Data Visualization

Bachelor of Engineering in Computer Engineering | LDRP-ITR, Gujarat, India | May 2023
Coursework**:** Data Structures & Algorithms, Python Programming, Object-Oriented Programming (Java)