

Anchit JV Rao

Tracy, CA | 408-313-0873 | anchit.jv.rao@gmail.com | linkedin.com/in/anchitrao | github.com/anchitr | anchitjvrao.com

SUMMARY

Data engineer with 5+ years across enterprise hardware and consulting, building ETL pipelines and cloud data infrastructure in Python, SQL, and Airflow. Has put LLMs into production data pipelines at both employers, for translation, classification, and structured extraction from unstructured documents. Owns pipelines through production, including sole on-call for a 32-DAG fleet and data quality monitoring over a 1B+ row database.

EXPERIENCE

Data & Analytics Consultant

Uday Shukla Corp.

Sep. 2024 - Present

Lansing, MI (Remote)

- Owned end-to-end delivery of 8 custom data engineering and AI solutions across 6 client teams, translating ambiguous business requirements into production pipelines and integrations
- Designed and deployed batch data pipelines in Python, SQL, and Kestra, consolidating multiple source systems and processing 3-5K records weekly for downstream analytics and AI applications
- Integrated LLM APIs (Claude, OpenAI) into data pipelines to extract and classify request intent from unstructured email, PDF, and free-text form content that intake forms left uncategorized; replaced a manual review process consuming 2-3 hours daily across a 4-person team
- Built n8n automation agents integrating CRM data with LLMs for a real estate client, classifying inbound requests as new leads or existing-client follow-ups, routing each to the right team, and auto-generating templated responses that fed client requirements into their tracking database; cut follow-up response time from 2+ hours to under 40 minutes
- Architected automated Power BI dashboards with DAX and Power Query, consolidating multi-source raw data into executive KPIs; eliminated 15+ hours/week of manual reporting and improved data accuracy 25%

Software Engineer, Data

Western Digital Inc.

Jun. 2020 - Jan. 2024

Milpitas, CA

- Engineered a multilingual ingestion pipeline in Python and LLMs to process 10,000+ customer forum posts every 48 hours across 5 languages, automating translation, sentiment analysis, and issue classification; routed cases to the correct regional support teams and enabled RMA logistics to track defect frequency, cutting manual triage time 60%
- Designed and owned 32 Airflow DAGs consolidating sales, inventory, and executive reporting across 4 source systems and 100K+ daily rows; served as sole on-call for the full pipeline fleet, replacing 30+ hours/week of manual ad-hoc requests with scheduled delivery
- Built an automated validation pipeline with Great Expectations over a 1B+ row production database, monitoring regional SKU mapping against the master list and quantity drift between physical and reported inventory; traced recurring mismatches to faulty upload processes and drove permanent fixes with APAC and EMEA manufacturing teams
- Led the audit and refactor of an entire production application database into a single source of truth, defining schema and naming conventions adopted by 4 teams; reduced data integrity errors 25% and supported SOX compliance reporting
- Built a self-service Python and pandas CLI suite enabling a 34-person logistics support team to query production data without engineering involvement; drove 80% adoption before departure and improved issue resolution turnaround 35%

TECHNICAL SKILLS

Languages: Python, SQL (Postgres, Oracle)

Data Processing: PySpark, pandas, Polars, Great Expectations

Orchestration & ETL: Apache Airflow, Kestra, n8n, ETL/ELT Data Pipelines, Workflow Automation

AI/LLM: Claude Integration, LLM APIs, LLM Orchestration, Prompt Engineering, Structured Output

Data Analysis & Visualization: Power BI (DAX, Power Query), Exploratory Data Analysis, Statistical Modeling

Cloud & Infrastructure: AWS (S3/Glue), Databricks, Delta Lake

Tools: Git/GitHub/GitLab, CLI Development, Data Quality & Validation

EDUCATION

Boston University

Master of Science in Computer Information Systems

Boston, MA

Nov. 2022 - May 2024

Indiana University

Bachelor of Science in Informatics

Bloomington, IN

Aug. 2016 - May 2019

CERTIFICATIONS

Databricks Certified Data Engineer Associate

Apr. 2026

DeepLearning.AI Data Engineering Professional

Sep. 2025

Google Cloud Professional Data Engineer

In Progress