

Mit Dasondi

Senior Data Engineer

+91 95586 55755
mitdasondi777@gmail.com

SUMMARY

Senior Data Engineer with 6+ years building and owning cloud-native, fault-tolerant data platforms on GCP and AWS. I ship end-to-end data products every sprint — owning the full lifecycle from source ingestion and append-only bronze (Iceberg/BigLake) through dbt dimensional modeling (SCD2, fact/dim at multi-million-row scale) and a governed publish layer, across DEV/UAT/PROD. Strengths in root-causing data defects, source-to-target reconciliation, and BigQuery cost/performance tuning, with reliable daily orchestration on Airflow/Cloud Composer and delivery through GitLab CI/CD. Fluent in Python and SQL, and in AI-assisted engineering workflows.

PROFESSIONAL SKILLS

Cloud Platforms:

GCP (BigQuery, Dataproc, Cloud Composer/Airflow, Cloud Run, Scheduler, Monitoring, Cloud Logging), AWS (S3, Kinesis, Firehose, Lambda)

Data Engineering:

ETL/ELT, dbt, Dimensional Modeling (Fact/Dim, SCD2), Data Products, Medallion / Lakehouse (Bronze/Silver/Gold), Avro, Data Warehousing, Data Migration, Web Scraping

Databases:

BigQuery, PostgreSQL, MySQL, MongoDB

Languages:

Python, SQL, PySpark

Tools & Frameworks:

Git, GitLab CI/CD, Mage-AI, Airflow, Iceberg/BigLake, Artifact Registry, Fivetran, FastAPI, OpenAPI, Soda, Cosmos, Google Sheets

AI-Assisted Engineering:

Claude / Claude Code, LLM prompt engineering, agentic developer workflows

Visualization:

Looker Studio, Tableau, Power BI

WORK EXPERIENCE

Senior Data Engineer

Coforge (Client: QXO) · Apr 2026 – Present

- Own end-to-end data products shipped **every sprint** — driving the full lifecycle from source extraction and append-only bronze ingestion through dbt dimensional modeling and a governed publish layer — on a GCP lakehouse spanning DEV, UAT, and PROD.
- Designed, built, and own a **7M-row daily-snapshot last-mile delivery (OTIF) fact model** in dbt/BigQuery — the single source of truth powering dispatch %, confirmed-arrival %, on-time %, photo-completion %, and delivery-sales KPIs across the full order history (2020–present).
- Engineered the ingestion-to-bronze layer with Dataproc/Spark: extract from Mincron (Oracle) and other sources into a shared raw landing bucket, standardize to Avro with generated manifests, and load append-only **Iceberg bronze tables exposed to BigQuery as BigLake external tables**.
- Re-architected the fact model to be sales-order-driven (Mincron SNDO) with a UNION-ALL branch for same-day open/batched orders, so operational dashboards reflect current-day activity instead of lagging by days.
- Root-caused and eliminated a **~100K duplicate-key defect** — traced to an inclusive time-window join boundary and fixed with a half-open interval plus a merge→insert_overwrite strategy — removing ~95% of duplicates while preserving all business rows.
- Built a repeatable source-to-target reconciliation validating dbt output against authoritative Mincron/Tableau reports to within ~1pp, isolating variance root causes (grain mismatches, carrier filtering, line-type differences).
- Diagnosed and rebuilt a drifted customer-master (HMI) Iceberg/BigLake table and its schema config, then full-refreshed the downstream dbt SCD2 dimension chain to restore loyalty reporting.
- Cut incremental load cost and runtime via partitioning, clustering, and lookback tuning; orchestrated daily runs through fully managed Airflow/Cloud Composer DAGs and shipped every change via GitLab CI/CD with env-aware dbt sources (dev/uat/prod).

Data Engineer

Freedom Property Investors · 2024 – 2026

- Sole data engineer: designed, built, and operated the company's entire GCP data platform end to end — every dataset, pipeline, and streaming ingestion.
- Reduced pipeline failures and delivery delays by **80%** with resilient, modular-DAG ELT and retry logic, and cut incident response time by **60%** through real-time Cloud Monitoring alerts and dashboards.
- Developed dbt data marts for key business domains (CoreLogic, IPScape, KiXy) and built **10+ dynamic Looker Studio dashboards**, replacing Excel workflows and driving analytics adoption with consistent, validated metrics.
- Architected a full-stack pipeline for the AI Sales Scorecard, surfacing strategic performance insights for the business.
- Enabled document-availability tracking for **10K+ deals** by building a GCS metadata service and PDF validation system.
- Automated reporting across **50+ datasets** via a Google Sheets Apps Script (auto-fetching the latest BigQuery tables) and a reusable Gmail extraction pipeline, eliminating manual data handling.
- Hardened security by migrating all credentials to GCP Secret Manager, and pioneered Mage-AI to make transformations more modular and faster to ship.

Data Science Engineer

Meditab · Dec 2021 – May 2024

- Spearheaded data modeling and warehousing to support EHR transitions from legacy systems to IMS.
- Refactored database structures and optimized the event cache, reducing server load by **~26%**.
- Replaced a legacy ETL system with a robust SQL-based solution, improving reliability and performance.
- Automated data ingestion from spreadsheets and JSON using Python, significantly reducing manual processes.
- Designed dynamic, physician-centric reports from structured EMR datasets.
- Optimized AWS S3 storage with lifecycle policies, cutting cost and improving data version control.
- Built Python tooling to compress and export RTF documents, improving document-handling efficiency.

Junior Data Engineer

Beekay Infotech · Jun 2020 – Dec 2021

- Designed and implemented efficient SQL-based data pipelines for enterprise clients including Aditya Birla and Ambuja Cement.
- Achieved a **25% reduction** in operational cost by replacing slower Python pipelines.
- Built and maintained a centralized data warehouse, improving data accuracy by **15%**.
- Leveraged AWS S3, Kinesis, Firehose, and Lambda for scalable data ingestion and transformation.

EDUCATION

Master's in Computer Engineering (M.Tech) — 2021

AI-driven research on automated change detection using remote-sensing data for Indian defence (BISAG).

Bachelor's in Computer Engineering (B.E.) — 2019

Awarded Best Project of the Year; recognized for startup innovation and academic excellence.

CERTIFICATIONS

Google Cloud Certified — Associate Cloud Engineer

dbt Certified Developer