

HITESH DAGER

Data Engineer · Snowflake DWH

Gurugram, India | +91-7011482715 | hiteshdagar123@gmail.com | linkedin.com/in/hiteshdagar

PROFESSIONAL SUMMARY

Data Engineer with 3+ years building Snowflake DWHs, Python ETL pipelines, complex SQL transformations, and Power BI dashboards across fintech and IoT. Skilled in Python - **PySpark, SQL, Airflow, Snowflake**, and PostgreSQL with strong AWS CloudOps expertise using AWS Glue, Docker, and Kubernetes.

TECHNICAL SKILLS

Data Warehousing: Snowflake (DWH design, multi-cluster WH, Zero-Copy Cloning, Time Travel, Streams & Tasks, Data Sharing, Snowpipe)

Data Engineering: Python · SQL · PySpark · Apache Airflow · Databricks · PostgreSQL · AWS Glue · ETL/ELT · Dimensional Modeling · pandas · NumPy · SQLAlchemy · psycopg2 · BeautifulSoup · Selenium

Cloud & DevOps: AWS (S3, EC2, Redshift, Glue) · Kubernetes · Docker · Terraform ·

Analytics & BI: Power BI (DAX, semantic models, paginated reports, DirectQuery, row-level security, Power Query, scheduled refresh) · Advanced Excel ·

PROFESSIONAL EXPERIENCE

Software Engineer | Vector Space Technologies Ltd *Dec 2025 – March 2026*

- Designed and managed production-grade Kubernetes clusters (on-prem and cloud), ensuring high availability, auto-scaling, and fault tolerance for business-critical workloads.
- Deployed and maintained multi-namespace microservices architectures using Helm charts, ConfigMaps, Secrets, and RBAC policies for secure, environment-isolated deployments.
- Managed AWS infra (EC2, S3, RDS) via Terraform; deployed containerised services with Docker and Kubernetes for high-availability data environments.
- Authored and maintained Docker Compose configurations for local development environments, ensuring dev-prod parity across engineering teams

Retainer – BI Engineer | Piramal Finance *Oct 2024 – Aug 2025*

- Designed Snowflake enterprise DWH from scratch – SQL DDL/DML for fact & dimension tables (loan origination, disbursement, collections) enabling single-source-of-truth reporting.
- Boosted Snowflake query performance by 50%+ via clustering keys, materialised views, result caching, and SQL profiling – cutting credit costs and dashboard load times.
- Built PySpark ELT pipelines on Databricks to ingest raw data into Snowflake staging – significantly faster than the legacy SQL-only approach.
- Automated Snowflake incremental loads via Snowpipe (S3 event-driven) and Streams + Merge for CDC updates across 20+ source tables.
- Built Power BI dashboards (DAX) for seller KPIs, portfolio health & collections – collaborated with **McKinsey** to drive ₹8,000+ crore disbursement decisions.
- Automated pre-ingestion cleansing via Python (pandas, NumPy) + SQL stored procedures – zero bad records entering the Snowflake staging layer.

Executive Engineer (Data Engineer) | Nippon Audiotronix *Feb 2022 – Aug 2023*

- Built Kafka real-time pipelines ingesting CAN bus + GPS data from 500+ IoT devices into Snowflake via SQL-modelled staging layers for live fleet reporting.
- Improved telemetry data accuracy by 30% via stream processing optimisation, deduplication filters, and schema enforcement at the Kafka consumer layer.
- Delivered Power BI dashboards (GPS KPIs, AIS-140 compliance, route deviation alerts) for 500+ vehicles – replaced Excel reports with real-time regulatory visibility.
- Parsed raw CAN bus frames with Python; applied SQL aggregation logic in Snowflake by route, time & device – boosting pipeline throughput and accuracy.

Python Developer (SWE) | Searchosis Marketing *Jul 2021 – Feb 2022*

- Built Python (pandas, Requests) + SQL ETL utilities for web-scale crawling, transformation, and loading – cutting manual data prep time.
- Deployed Power BI dashboards (DAX, Power Query) for supply chain and project analytics across multiple clients – drill-through reports with automated refresh.
- Implemented Snowflake ETL + SQL transforms + AWS Snowpipe for a scaffolding client – automated end-to-end ingestion enabling near-real-time inventory visibility.
- Automated client data extraction using Python (BeautifulSoup, Selenium, Requests) – cut data collection time by 60% across multiple client verticals.

KEY PROJECTS

Sphurti DWH Analytics Platform — *Piramal Finance*

- Snowflake DWH with star schema: 6 fact tables (disbursements, collections, loan pipeline, repayments, agent activity, write-offs) + 15+ dimension tables — zero reconciliation gaps across all business units.
- PySpark (Databricks) ELT → Snowflake raw/staging → dbt-modelled presentation layer → consumed by Power BI semantic model.
- Full Airflow DAG orchestration: S3 → Snowpipe → SQL staging → dbt → Power BI refresh – fully automated, zero manual intervention.
- Power BI dashboards (DAX KPI cards, drill-through) for 1,000+ field agents – C-suite visibility over ₹8,000+ crore disbursement portfolio in real time.
- Python (pandas, SQLAlchemy) + SQL stored procedures for data validation and reconciliation – reduced manual analyst effort.

Mercury Bank Cloud Data Platform — *Vector Space Technologies Ltd*

- Python (pandas, SQLAlchemy) ELT pipelines: PostgreSQL extract → SQL transforms (CTEs, window functions) → Snowflake load – 100K+ daily records with full audit trail.
- Configured Snowflake data sharing to expose curated SQL views to compliance team – eliminated manual weekly extract process entirely.
- Airflow DAGs with parameterised SQL tasks, SLA callbacks, and XCom-based metadata propagation for full downstream dependency tracking.
- Power BI paginated reports (DAX time-intelligence, Snowflake DirectQuery) – marketing team slices campaign performance by channel, region & cohort in real time.
- Python data quality framework (Great Expectations) enforcing SQL schema contracts and row-count SLAs – caught anomalies upstream before reaching Snowflake.

AIS-140 Fleet Telematics Analytics Platform — *Nippon Audiotronix*

- Kafka real-time ingestion for 500+ IoT fleet vehicles → SQL-modelled Snowflake staging & aggregation layers → live AIS-140 compliance reporting.
- Python-parsed CAN bus frames with deduplication + schema enforcement – improved telemetry data accuracy and reduced quality incidents.
- Power BI dashboards: speed, route deviation & AIS-140 KPIs for 500+ vehicles – fleet managers and R&D generate compliance reports on demand.

Client Data Automation & Analytics Platform — *Searchosis Marketing*

- End-to-end Python platform (BeautifulSoup, Selenium, pandas) + SQL ETL for multi-client web crawling – cut manual data collection time.
- Snowflake ETL + AWS Snowpipe for scaffolding client – automated ingestion enabling near-real-time inventory and supply chain visibility.
- Power BI dashboards (DAX, Power Query) for supply chain & project analytics across multiple clients – drill-through self-service reports reduced turnaround from days to minutes.

EDUCATION

B.Tech – Computer Science Engineering 2017 – 2021

Manav Rachna International Institute

Data Science Certification 2023 – 2024

Great Learning