

# Karthick R

**Data Engineer | Snowflake | AWS | DBT | SQL**

+91 8660649915 | karthickgkmj@gmail.com | Bengaluru, India

---

## Professional Summary

Data Engineer with 3+ years of hands-on experience building and maintaining Snowflake data warehouses. Strong background in designing **dbt** transformation models, implementing change data capture **pipelines**, and optimizing warehouse queries. Comfortable working with **Python** for data validation and automation. Currently focused on migration projects and **CDC** implementations at Tech Mahindra.

---

## Technical Skills

**Cloud Platforms:** Snowflake, AWS (S3)

**Databases:** MySQL, Oracle, Snowflake

**Data Engineering:** dbt, SQL, SnowSQL, Python

**Snowflake Features:** Streams, Tasks, Snowpipe, Materialized Views, Zero Copy Cloning, Time Travel, Data Masking

**Formats & Tools:** CSV, JSON, PARQUET, Git, Jira, Mulesoft

**Visualization:** Power BI

**DevOps:** CI/CD pipelines, dbt Cloud, GitHub Actions

---

## Professional Experience

### Data Engineer - Tech Mahindra, Bangalore

July 2025 – Present

TMAP (Transportation Mobility Access Platform) - Walgreens

Stack: **Snowflake, AWS S3, dbt, Python, Snowflake Tasks**

- Migrated 15+ tables and 2TB+ of historical data from **Oracle RDBMS** to Snowflake using **COPY INTO**. Schema transformations applied during load to handle data type conversions from legacy systems.
- Built 80+ **dbt** models organized by staging, intermediate, and mart layers. Added 120+ dbt tests (generic and custom) to catch data quality issues before they reach reporting. CI/CD integration reduced code review cycles from 2 days to 4 hours.
- Created 5 CDC pipelines using Snowflake **Streams and Tasks** to capture insertions, updates, and deletes in real-time. Processes 50K+ daily incremental records with <5 minute latency. Implemented **SCD Type 2** logic in dbt for **slowly changing dimensions**.
- Optimized 8 critical queries in reporting layer. Used Query Profile to identify full table scans and missing indexes. Materialized 3 high-traffic **views**, cutting dashboard load time from 45 seconds to 8 seconds. Cluster keys added to **fact tables** to improve partitioning strategy.
- Wrote Python **stored procedures** in Snowpark to validate incoming data. Checks for null values, duplicates, and referential integrity. Flagged data anomalies are logged to an error table for investigation.

---

## Associate Consultant - MDI Insurance Pvt. Ltd., Bangalore

September 2022 – May 2025

### Allstate Insurance Mart - Reporting Infrastructure

Stack: **Snowflake, AWS S3, SQL, dbt, Power BI**

- Owned the migration of 40+ RDBMS tables (3.5TB) from on-premises to Snowflake. Used **COPY INTO** with transformation logic to handle character encoding issues and date format mismatches. Validated record counts post-load across all tables (100% match verified).
- Built 45 dbt models for insurance claims, premiums, and customer dimensions. Schema and data tests passed in all CI/CD deployments (zero production failures). Created reusable **macros** for common transformations (null handling, type conversions) used across 12+ downstream projects.
- Set up internal and external **Snowflake stages**. Internal stages store temporary files during load/unload operations. External stages point to S3 buckets with

lifecycle policies to auto-delete aged files. Reduced storage costs by 25% with this setup.

- Created 8 materialized views for daily reporting. Scheduled refresh **tasks** to run at 6 AM to populate dashboards. **Views** use secure headers to limit data by user roles, reducing exposure of sensitive information.
- Implemented CDC using Snowflake **Streams** on 6 source tables. Built Tasks to process stream rows incrementally. Idempotent design allows Tasks to re-run without duplicating data if a failure occurs mid-pipeline.
- Used **Zero Copy Clone** to spin up QA and Dev environments from production **snapshots**. Reduced environment setup time from 2 hours to 5 minutes. Team can now test schema changes safely without touching prod data.
- Wrote **dbt tests**: schema (column existence, data type), data (custom SQL to check for nulls and negative values), and relationship tests (foreign key validation). Tests run before deployment to catch issues early.
- **Cloned databases** for testing Snowflake account changes (warehouse sizing, retention policies). Allowed us to run performance experiments without impacting production reporting SLAs.

---

## Education

Bachelor of Computer Applications (BCA) - Rabindranath Tagore University

---

## Key Achievements

- **Data Integrity**: Reduced production data errors by 80% through comprehensive dbt testing framework and validation stored procedures.
- **Performance**: Cut query execution time from 3+ minutes to under 1 minute in 3 major reporting views through Query Profile analysis and indexing.
- **Cost Optimization**: **Materialized views** strategy saved \$8K/month in warehouse compute credits compared to previous approach.
- **Team Growth**: Onboarded 2 junior engineers in 1.5 weeks vs. typical 3-week ramp time through structured mentoring and documentation.
- **Migration Success**: Zero data loss in two major RDBMS-to-Snowflake migrations totaling 5.5TB across 55+ tables.

---

## Notable Projects

**Snowflake Performance Tuning for Claims Analytics** - Identified N-way joins causing full table scans. Refactored queries with proper filtering and cluster keys. Result: 65% faster query execution, immediate dashboard performance improvement.

**Real-time CDC Pipeline for Premium Updates** - Built Streams + Tasks solution to capture premium tier changes. Updates reflected in reporting layer within 3 minutes vs. 24-hour batch delays previously. Stakeholders could track promotions in near real-time.

**dbt Testing Framework** - Designed and implemented 120+ tests across 80 models. Found 18 duplicate records in historical data and fixed fact table join logic that was inflating claim counts by 2.3%.