

VINH X. NGUYEN

nxv.can@gmail.com | (+1) 548.994.4264 | Ontario, Canada | Website: vinhnx.ca

Data Lover with 9+ years in big data platforms (infrastructure to semantic layers).

AI Agentic LLM:

Business value: accelerated time-to-insight before and after sale finalization, promptly alerted missing invoices.

Techniques: Claude, RAG + Vector, Tool calls (via MCP/API), Relevance scoring, Sub-Agent & Scoped Handoffs.

Data Modeling:

OLTP: designed 10M+ records strong consistency financial data pipelines ensure ACID and scalability.

OLAP: built 200B+ records Schema-on-read data stores (AWS S3 + Glue Catalog) with eventually consistency data pipelines for batch and micro-batch data ingestion and enrichment.

Star Schema & SCD: engineered real-time 5 facts and 10 dimensional tables for 200B+ logstream event with auto-enrichment layer for master data from on-prem to AWS then to Snowflake using SCD-1.

Data Vault: designed Data-Vault-2.0-inspired schemas for multi-sources independent ingestion and enrichment.

Data Engineer:

Streaming: developed 200B rows pipelines (ingestion & enrichment) on AWS & Snowflake, improved enrichment quality from 60% to 99.97%; reduced time-to-insight from days to minutes.

Batching: migrated 100B rows data from SQL Server to AWS and Snowflake, reduced query time from 30s to 8s.

Mapping: improved omni-channel mapping rate (click events to sale transactions) from 60% to 99.97%.

Computing: analyzed 30TB text on AWS EMR; Log analytics on Databrick; Lambda; ECS; Beanstalk; ECS.

Transit: Kafka, Kinesis, SQS, SNS, Non-blocking IO, At-Least-One / Exactly-Once.

Stores: S3 + Glue Catalog, DynamoDB, Redis, Snowflake, Databrick Delta, MySQL, PostgreSQL.

Data Analytics:

Financial Intelligence: audited 3M+ records to uncover \$M leakage; improved mapping from 60% to 99.97%.

NLP & SEO: processed 30TB+ text using TF-IDF, Lemmatization, PosTag, and Stemming for keyword relevancy scoring.

Data Platforms: managed Google Analytics (2M/mo), Transaction Stores (500M), and Data Warehouses (2B rows).

Visualization: Python, Snowflake, Tableau, PowerBI, QuickSight, Grafana, DataDog, CloudWatch, Prometheus.

Leadership & Communication:

Chaired 100+ technical reviews in Architect Community at TymeBank, defined integration patterns, data patterns (access control, retention, CDC, backfills, SLA, lineage) across 5 engineering/feature teams.

Acting Data Product Owner - weekly presented business numbers to Directors and C-Levels.

Resolved critical bottlenecks including a \$10K Datadog issue, Bank Certificate Pinning, and Kiosk Upload Congestion.

EMPLOYMENT SUMMARY

Data Engineering Lead (Apr-2021 till Now)

Cox Automotive, Canada

Data Solution Architect

TymeBank (GOTyme), Vietnam

Technical Architect

NFQ Asia, Vietnam

Principal Software Engineer

NashTech Global, Vietnam

Technical Architect

FPT Software (Vietnam + Philippines)