

Vinh X. Nguyen

nxv.can@gmail.com | +1-548-994-4264 | linkedin.com/in/vinh-nguyen-xuan | vinhnx.ca | Waterloo, Ontario, Canada

PROFESSIONAL SUMMARY

Lead Data Architect with 9+ years governing enterprise data architecture for AI/ML workloads (15+ yrs total industry). Architect Community Lead at TymeBank (1M+ users) - chaired **100+** architecture reviews and standardized integration patterns (CDC, contracts, backfills, SLA, lineage, access control) across 5 engineering teams. Designed and operated a **200B+** row lakehouse on AWS / Snowflake / Databricks Delta (Bronze→Silver→Gold) with 99.999% enrichment quality and 73% query-time cuts. Built production agentic LLM systems (LangGraph / MCP / RAG) on \$1.5B in transactions with Evaluator scoring, hallucination guardrails, and retry/escalation policies - MLOps maturity for responsible AI. Deep cloud + cost-optimization track record (Snowflake/Databricks tuning, Datadog \$10K leak surfaced). MSc Computer Science, Université Pierre et Marie Curie (UPMC Paris 6).

CORE COMPETENCIES

Enterprise Data Architecture | Lakehouse (Bronze/Silver/Gold) | AI/ML Data Platforms | MLOps & Responsible AI | Data Governance & Lineage | Snowflake / Databricks Delta | Spark / EMR / Kafka / Kinesis | AWS Cloud-Native Architecture | Data Vault 2.0 / Medallion / Star | Cross-Team Architecture Leadership | Cost & Performance Optimization | Python / SQL / Java

WORK EXPERIENCE

FPT Canada

Apr 2021 - Present | Ontario, Canada

Data Engineering Lead

- Lakehouse Architecture (200B+ rows):** Designed and operated Bronze→Silver→Gold on AWS (Kinesis / Glue / EMR / S3) + Snowflake + Databricks Delta with reconciliation per layer; 99.999% enrichment quality, query time **30s** → **8s** (73% reduction).
- Master Data & Modeling:** Built 10M-entity master data store (SCD-1 enrichment, on-prem → AWS → Snowflake) as canonical source for 5 facts / 10 dimensions; established Data Vault 2.0 + Medallion patterns across the org.
- Agentic AI / MLOps Guardrails:** Built MADA - 6-agent LangGraph pipeline (Orchestrator → Auditor → Detector → Retriever → Evaluator) on \$1.5B in transactions with Evaluator scoring (0-1), hallucination flags, retry/escalation policies; lifted reconciliation 60% → **99.97%**.
- Governance & Lineage:** Defined data contracts, SLA-backed products, freshness/lineage signals; surfaced multi-million-dollar revenue leakage via end-to-end audit; standardized data integrity protocols across the org.
- Cost / Operational Excellence:** Snowflake warehouse cost tuning, Databricks cost optimization; isolated Datadog logging defect leaking **\$10K** in AWS spend; weekly CloudWatch reviews with Directors.
- Cross-Functional Leadership:** Coordinated 7+ data teams and Directors of Engineering, Architecture, and Analytics; delivered self-serve platform adopted by 7+ teams (DS, Analytics, Finance).

TymeBank (GOTyme)

Jul 2019 - Apr 2021 | Vietnam

Data Solution Architect

- Architecture Governance:** Chaired **100+** architecture reviews and approved 100+ designs across 5 development teams for a **1M+ user** digital bank; standardized 4 inter/inner-system integration frameworks (CDC, contracts, backfills, SLA, lineage, access control).
- Data Platform & Reporting:** Built and managed **500M+** transaction reporting pipelines (raw → EDW → BI) across multi-source analytics: Google Analytics (2M/mo), TDS (500M), ODS (200-500K), DW (1-2B), Logs (10-100GB), Delta Lake (10-50M).
- Mass-Scale Product Architecture:** Architected the launch of 4 banking products (Personal Lending, VAS, GoalSave Kicker, ID Payment) with banking-grade compliance and security patterns.
- Security Architecture:** Designed certificate pinning + rotation strategy for 1M+ users, eliminating 6-month forced upgrades; coordinated with Cyber Architects on banking-grade hardening.

NFQ Asia

Apr 2017 - Jul 2019 | Vietnam

Technical Architect

- NLP at scale:** Engineered **30TB** plaintext NLP pipeline (TF-IDF, lemmatization, POS tagging) on first serverless keyword-analytics platform (AWS Lambda + API Gateway + DynamoDB + Terraform); **20K-50K req/day**, latency 60s→3s (95% cut).
- Cost telemetry:** Cross-checked AWS CloudWatch vs Datadog ingestion counters, isolated Datadog logging defect leaking **\$10K** in AWS spend - finding shared with Datadog engineering.
- Java engineering org:** Stood up 4 Java teams (3-6 devs each) for German startups; ran **80+** technical interviews; standardized hiring bar and architect career path.

NashTech Global

May 2016 - Apr 2017 | Vietnam

Principal Software Engineer

- Engineering lead for 20-member multi-disciplinary R&D team (Backend, Frontend, QA) across Banking, Insurance, KYC; led 5 backend devs maintaining mission-critical insurance services.
- Upskilled 10+ engineers via daily 30-min technical training; iteratively refined internal processes (10+ versions) to reduce production incidents.

FPT Software Philippines

Jun 2014 - May 2016 | Philippines & Vietnam

Solutions Architect

- Architected the Cebu Development Center end-to-end - first Technical Architect deployed; recruited and led **22 Java + .NET** engineers; built University-to-Engineer pipeline; standardized curriculum that upskilled **60+** developers.
- Conducted **100+** technical and leadership interviews; designed recruitment processes and trained new interviewers; achieved **90%** new-hire retention.

SELECTED PROJECTS

200B+ Row Lakehouse - Bronze → Silver → Gold Architecture · Governance

Reference architecture for omni-channel click-stream: Kinesis/Kafka → Spark on EMR/Databricks medallion layers → Snowflake → Glue Catalog/S3 Iceberg. Per-layer reconciliation, schema/contract evolution, lineage + freshness signals, RBAC + data classification, MCP-driven Copilot integration, SLA-backed self-serve data products consumed by BI (Tableau/PowerBI/QuickSight) and ML.

Stack: AWS (Kinesis, Glue, EMR, Lambda, IAM), Snowflake (Snowpipe, COPY INTO, RBAC), Databricks Delta/Iceberg, Terraform, Airflow.

MADA - Multi-Agent Financial Auditor (\$1.5B) AI Platform

6-agent LangGraph platform serving two modes: **Ask-MADA** for natural-language self-serve investigations (audit-grade Findings Reports with citations, days → minutes), and **Always-On MADA** for scheduled data-quality sweeps over \$1.5B in transactions (Internal-error sweep + External-anomaly sweep + Evaluator-judge scoring). Reusable agent governance pattern complementing deterministic leakage reporting.

Stack: LangGraph, LangChain, MCP, RAG, embeddings, Python, Snowflake, AWS.

vinhnx.ca - Personal AWS-Native Site Personal · IaC + GenAI

Fully serverless personal site with IaC, OIDC-based CI/CD, AI chat backend (Gemini 2.5 + Claude Haiku 4.5 fallback) grounded in this resume, per-IP rate limiting, prompt-injection guards, admin dashboard, two-way Telegram integration. **Stack:** AWS API Gateway + Lambda (Python 3.12) + DynamoDB; Terraform 1.8; GitHub Actions OIDC.

EDUCATION

MSc in Computer Science - Université Pierre et Marie Curie (UPMC Paris 6)

2011-2013

Statistics for Data Analysis - McMaster University, Ontario

2025

CERTIFICATIONS

AWS Certified Data Analytics Specialty - Amazon Web Services

2023

AWS Advanced Training (Banking Architect Leads) - Amazon Web Services

2020

TECHNICAL SKILLS

Languages: Python, SQL, Java **Lakehouse & Warehouse:** Snowflake (Snowpipe, MV, cost tuning), Databricks Delta Lake, Spark/EMR, Glue Catalog, S3

Streaming & Messaging: Kafka, Kinesis, SQS/SNS, DynamoDB Streams, Spark Streaming **Modeling:** Data Vault 2.0, Medallion (Bronze/Silver/Gold), Star Schema, SCD-1/2, OLAP/OLTP

AI / ML Platform: LangGraph, LangChain, MCP, RAG, Vector DB, Evaluator scoring, hallucination guardrails, SageMaker

Governance & Lineage: data contracts, SLA, freshness, access control, responsible AI **Cloud & IaC:** AWS (full stack), Terraform, CloudFormation; Azure / GCP exposure