

# VISHNU BAPATHI

## Data & AI Engineering Leader

vishnu.maheswar@gmail.com | (408) 859-5451 | Dallas, TX | [linkedin.com/in/vishnu-bapathi-dataengineer](https://www.linkedin.com/in/vishnu-bapathi-dataengineer)

### EXECUTIVE SUMMARY

Data & AI Engineering Leader with 15+ years building enterprise platforms and growing engineering organizations for Fortune 500 clients. At BCG, I own technical direction for multi-million-dollar transformation programs—hiring and leading teams of 30+, managing \$4M+ in annual cloud infrastructure spend, and presenting data strategy to C-suite stakeholders. Currently leading GenAI adoption across 70+ engineers, combining deep platform expertise (Databricks, Snowflake, Spark) with the organizational leadership to drive adoption at enterprise scale. Published thought leader on data contracts, AI engineering architecture, and data governance.

### SIGNATURE ACHIEVEMENTS

#### 30+ Engineers Hired & Led

Built BCG's data engineering practice from 5 to 30+ through structured hiring, career laddering, and engineering standards adopted practice-wide.

#### 67% Runtime + 40% Cost Reduction

Hardened a 170M+ row Databricks/Spark platform: 2.5hrs→50min runtime, CA\$2.5K→CA\$1.5K/day compute, 2-hour earlier daily completion.

#### \$4M+ Annual Infra Budget Owned

Managed cloud spend across Azure, AWS, and Databricks—negotiating vendor contracts, optimizing reserved capacity, and establishing FinOps reporting.

#### 70+ Engineers AI-Coached

Designed GenAI adoption sprints embedding Claude Code into PDLC workflows across 6 teams: 85% faster ramp-up, 150% throughput increase.

### CORE COMPETENCIES

#### Organization & Team Building

Hiring (5→30+ engineers), career laddering, engineering standards, interview frameworks, onboarding design, mentorship

#### Data Engineering at Scale

Databricks, Spark, Snowflake, Delta Lake, Kafka, Airflow, dbt, Python, Scala, SQL

#### Enterprise Architecture & Governance

Data mesh, MDM, data vault, lineage, cataloging, RBAC, SOX/HIPAA compliance, policy-as-code

#### Executive & Financial Leadership

\$4M+ infra budget ownership, vendor negotiation, C-suite advisory, program governance (\$3–8M), FinOps, OKRs

#### Generative AI & LLM Architecture

RAG pipelines, embeddings, vector search, LLM orchestration, AI copilots, prompt engineering, agentic coding

#### Cloud & DevOps

Azure, AWS, GCP, Kubernetes, Terraform, CI/CD, GitOps, observability, DR planning, cost optimization

### PROFESSIONAL EXPERIENCE

**Boston Consulting Group (BCG)** – Dallas, TX

May 2021 – Present

#### Data & AI Engineering Leader

- **Built the data engineering practice team from 5 to 30+ engineers** through structured hiring, technical interview design, onboarding frameworks, and career development laddering—establishing engineering standards and delivery governance adopted across BCG's technology practice.
- **Owned technical direction for programs valued at \$3–8M**, reporting to client CTOs, CDOs, and Managing Directors on architecture decisions, risk trade-offs, delivery milestones, and roadmap priorities. Presented quarterly data strategy to C-suite stakeholders.
- **Managed \$4M+ annual cloud infrastructure budget** across Azure, AWS, and Databricks—negotiating vendor contracts, optimizing spend through reserved capacity and autoscaling strategies, and establishing FinOps dashboards for client cost transparency.

#### AI Adoption Sprints – Global Engineering Transformation

- **Led GenAI adoption across 70+ engineers and 6 product teams**, designing the operating model (discovery frameworks, bounded use cases, guardrails, measurement) and coaching managers on PDLC integration. Owned the \$2M program budget and reported weekly to client VP of Engineering.

- **Result:** 85% reduction in re-orientation time, 150% increase in feature throughput, 45% reduction in context token consumption. Published the measurement framework as “The Amnesia Tax” on Medium.

### **OEM Pricing Platform – 5-Workstream Delivery**

- **Orchestrated 5 parallel workstreams as single technical owner:** Mechanical Engine, Collision Engine, 36+ Power BI analytics views, React Upload Portal, and Canada expansion—coordinating DS, DE, reporting, UI, infrastructure, and business teams across 3 time zones.
- **Hired and onboarded 4 engineers mid-engagement to recover delivery pace.** Presented bi-weekly to client CTO and business leadership on progress, trade-offs, and go-live readiness. Established SOX compliance path for post-handover sustainability.

### **Route Optimization – Platform Hardening at Scale**

- **Single-threaded owner for 170M+ row Databricks/Spark platform:** cut truckload runtime by 67% (2.5hrs→50min), reduced compute costs 40% (CA\$2.5K→CA\$1.5K/day), and moved end-to-end completion 2 hours earlier for downstream operations.
- **Owned \$1.5M annual Databricks infrastructure budget.** Built model-side DQ guardrails projected to eliminate 50–60% of data-driven incidents. Led DR planning, solver governance (Gurobi licensing), and CI/CD maturity. Made explicit calls to prioritize reliability over feature velocity.

### **Commodity Trading AI – GenAI-Enabled Platform**

- **Architected GenAI-enabled commodity trading platform on Azure** combining ML forecasting with LLM-driven intelligence for gas supply planning, inventory forecasting, and cost-to-serve optimization across Marathon, National Grid, and Superior Propane.
- **Designed RAG patterns** grounding LLM responses in curated trading data and forecast outputs. Enabled natural-language querying and AI-assisted what-if analysis—reducing analysis turnaround and strengthening stakeholder confidence.

### **Healthcare Data Foundation & Utilities MDM**

- **Designed enterprise data platform spanning 9 business domains and 43 entities** (patient, provider, product, sales, payor, fulfillment) across bronze/silver/gold layers. Led architecture workshops with client CTO and 6 department heads—decisions shaped a \$3M+ multi-year platform investment.
- **Led MDM strategy for a major utilities provider:** CDC ingestion from mainframe systems (DB2), Snowflake data lakehouse, Reltio evaluation, and cross-functional data governance supporting customer 360 and billing analytics.

**AT&T** – Dallas/Fort Worth, TX

Apr 2018 – May 2021

#### **Principal Data Engineer / Advanced Analytics Principal**

- **Owned data architecture strategy for marketing analytics** serving 3 business units and 50+ analysts—designed scalable systems integrating relational, geospatial, and time-series data across petabyte-scale environments on AWS and Azure.
- **Led a team of 8 analysts and data engineers;** defined team OKRs, sprint cadence, and delivery standards for experimentation programs in partnership with marketing, CX, and commercial operations leadership.
- **Delivered executive dashboards enabling C-suite visibility into \$200M+ marketing spend effectiveness** — churn analytics, campaign ROI, and customer cohort performance through Power BI and Tableau.
- **Built predictive models for churn forecasting and retention targeting;** A/B testing framework influenced \$15M+ annual retention program allocation decisions.

**DST Systems (SS&C)** – Kansas City, MO

Aug 2014 – Apr 2017

#### **Senior Data Engineer**

- **Led architecture of centralized data platforms for financial and asset management clients** — enabling portfolio analytics, compliance dashboards, and operational reporting for capital markets organizations.
- **Developed lineage-aware ETL pipelines with regulatory rigor:** audit logging, automated quality checks, and observability metrics for SOX/SEC reporting. Integrated Bloomberg and Morningstar data sources.

### **EARLIER EXPERIENCE**

**MedeAnalytics** (Sr Data Engineer, Dallas) | **Hewlett Packard** (Data Engineer, Bangalore) | **Infosys** (Systems Engineer, Chennai) — 2009–2015

Built reusable data ingestion layers for healthcare and enterprise clients. ETL development (Informatica, SSIS), Tableau/Business Objects reporting, and data warehouse design.

## THOUGHT LEADERSHIP

---

**“The Amnesia Tax”** (Medium, Apr 2026) — AI engineering memory architecture with 7-principle framework and 5-level maturity model

**“Agentic Data Contracts”** (Medium, Sep 2025) — Framework for AI-negotiated data quality enforcement with Git-based policy-as-code

**“From Models to Machines”** (Medium, Nov 2025) — How data engineering operationalizes AI: security, observability, cost optimization

**“The Dashboard Isn’t Dead”** (Medium, Apr 2026) — Why BI governance is the trust infrastructure for both dashboards and AI agents

## EDUCATION & CERTIFICATIONS

---

**MS Information Technology and Management** — University of Texas at Dallas (Dec 2013, GPA: 3.96, Dean’s Excellence Scholarship)

Agentic AI (Harvard DSI, 2025) | Climate & Sustainability (BCG, 2024) | Azure Data Engineer Associate (Microsoft, 2021)

## TECHNICAL PROFICIENCY

---

**Data Platforms:** Snowflake, Databricks, Spark, Delta Lake, Kafka, Airflow, dbt, Informatica, Reltio MDM

**AI / GenAI:** LLMs, RAG pipelines, embeddings, vector databases, prompt engineering, Claude Code, Agent SDK, MLOps

**Cloud:** Azure (ADF, Synapse, AKS, Databricks, Fabric), AWS (S3, Lambda, EKS), GCP (BigQuery), Docker, Kubernetes, Terraform

**Languages & BI:** Python, Scala, SQL, Node.js, React | Power BI, Tableau, Looker, QuickSight, DOMO

**Architecture:** Data Lakehouse, Data Mesh, MDM, Data Vault, dimensional modeling, CDC, streaming, data contracts, lineage

**Governance:** SOX, SEC, GDPR, HIPAA, RBAC, policy-as-code, data quality frameworks, FinOps