

# VENKATA SAI VARDHAN SEEPALA

(647) 898-3945 | [seepalavardhan98@gmail.com](mailto:seepalavardhan98@gmail.com) | [LinkedIn](#) | [GitHub](#)

## PROFESSIONAL SUMMARY

Results-driven **Lead Data Engineer** with 6+ years of experience designing and operating high-scale, cloud-native data platforms across GCP, AWS, and Azure. Proven track record of **leading engineering teams, defining technical roadmaps, and establishing org-wide data engineering standards** that accelerate delivery and improve quality. Equally effective as a **hands-on technical architect** (Staff DE) and as a **people-and-process leader** driving cross-functional alignment and engineering excellence (EM). Expertise spans real-time stream processing, CDC pipelines, Data Vault 2.0, MLOps, and developer productivity tooling.

## EXPERIENCE

### Lead Data Engineer Hard Rock Digital

Toronto, ON | Mar 2025 – Present

- **Technical Leadership & Team Enablement:** Established org-wide engineering standards for LLM integration (Claude Code/GitHub Copilot), prompt engineering best practices, and security guardrails — measurably accelerating code delivery velocity and unit-test coverage across the data engineering team; mentored engineers on production-grade stream processing and Data Vault design patterns.
- **Technical Roadmap & Systems Modernization:** Owned the technical roadmap for migrating legacy Java streaming services to Python/Apache Beam and upgrading 40+ Airflow DAGs to Airflow 3.0 on self-managed Kubernetes (GKE) — delivering zero-downtime cutovers with full backward compatibility, unblocking 3+ downstream teams.
- **High-Scale Stream Processing:** Architected and deployed 5+ production Dataflow pipelines (Apache Beam) processing 100K+ messages/hour; implemented exactly-once semantics and stateful processing, guaranteeing 100% data integrity across real-time event streams.
- **PostgreSQL Performance Optimization:** Achieved a 60x throughput increase on mission-critical pipelines by redesigning bulk load operations, slashing processing latency from 60 minutes to under 5 minutes for 320K+ records.
- **Data Vault 2.0 & CDC Architecture:** Designed a real-time ingestion framework using Hub/Link/Satellite patterns with FK-aware batching; engineered end-to-end CDC pipelines (Debezium → Pub/Sub → BigQuery), eliminating downstream data staleness and serving as the platform foundation for all new event streams.
- **Reliability, Scale & Data Quality:** Stress-tested endpoints to 10K msg/sec (3M+ messages sustained); implemented Protobuf schema validation, DLQ handling, and an API-driven validation engine for 16M+ records with retry logic and status polling to ensure high-fidelity ingestion of sensitive transaction data.
- **Cross-Functional Stakeholder Delivery:** Partnered with BI and executive stakeholders to define and deliver optimized analytical data marts for campaign and user engagement tracking, translating business requirements into production-grade data models that directly informed executive-level decision-making.

### Lead Data Engineer Viral Nation

Toronto, ON | Jun 2023 – Feb 2025

- **Engineering Leadership & Cross-Functional Delivery:** Led end-to-end delivery of data platform initiatives across GCP and Azure, collaborating with marketing, analytics, and product teams across 10+ concurrent client campaigns; drove adoption of dbt as the team-wide transformation standard, reducing data errors by 80% and establishing a single source of truth across 10+ data sources.
- **CDC & Infrastructure Architecture:** Spearheaded CDC pipelines (Cloud SQL → BigQuery) via DataStream (GCP), automating data lake creation and reducing transformation overhead by 40%; deployed Terraform templates on a private VPC cutting infrastructure setup time by 50%, enabling self-service provisioning for the team.
- **Real-Time Data Processing:** Architected Apache Beam (Python SDK) Dataflow pipelines streaming MongoDB and PostgreSQL data to BigQuery and Pub/Sub, enabling near-real-time campaign decision-making and reducing time-to-insight across 10+ clients.
- **MLOps & AI Model Delivery:** Built end-to-end CI/CD pipelines for AI model deployment on Azure ML Studio — cutting time-to-market by 35% and operational costs by 25% via Kubernetes autoscaling; led production deployment of brand detection and profanity filtering models via Kafka, moderating 1M+ daily social media interactions.
- **Reporting & ETL Optimization:** Designed PySpark ETL pipelines in Azure Synapse aggregating ROI and engagement KPIs; collaborated with marketing teams to reduce reporting latency by 40% via Looker dashboards — directly improving campaign accuracy and client retention.

### Senior Software Engineer Oracle

Bengaluru, KA, India | Aug 2019 – Dec 2021

- **Agile Team Leadership:** Facilitated Scrum ceremonies, sprint planning, and retrospectives for a distributed engineering team; identified and resolved deployment blockers, improving sprint velocity and cutting release cycle time by 25%.
- **Cost-Efficient Data Platform Engineering:** Designed AWS EMR pipelines serving 10K+ patients daily — reducing processing costs by 30% and runtime by 50% through optimized partitioning, cluster right-sizing, and dependency-aware scheduling.
- **Quality Engineering & Security Automation:** Authored Java/Python automated test frameworks cutting integration testing time by 70% and post-release defect rates by 60%; streamlined AWS IAM provisioning via CloudFormation, eliminating 90% of manual configuration errors across 15+ client clusters.
- **Healthcare Data Standardization:** Leveraged Java, Python, and microservices to normalize and standardize healthcare data to HL7/FHIR compliance standards, enabling seamless data interoperability across diverse enterprise client environments.

## Data Engineer Co-Op

Oakville, ON | Jan 2023 – Apr 2023

### BDO Canada

- **ETL Pipeline Optimization:** Redesigned Azure Data Factory pipelines by resolving entity dependency ordering, reducing full-load processing time from 13 hours to 9 hours — a 30% reduction in compute cost.
- **Dashboards & Automation:** Built Power BI dashboards integrating MQTT and digital twin metrics for 50+ manufacturing accounts; automated KQL query generation for Azure Data Explorer, accelerating data extraction by 25%.

## DevOps Engineering Intern

Bengaluru, KA, India | Apr 2019 – Jul 2019

### Sigmoid

- **Observability & Cost Reduction:** Built Datadog and Elasticsearch monitoring dashboards reducing cluster downtime by 50% via proactive alerting; delivered client PoCs that lowered AWS cluster costs by 20%, contributing to a 15% increase in client acquisition.

## EDUCATION

Master of Applied Computing | University of Windsor

Windsor, ON, Canada

Bachelor of Engineering in Computer Science | Sir MVIT

Bengaluru, KA, India

## TECHNICAL SKILLS

**Languages:** Python, Java, TypeScript, JavaScript, SQL, Terraform, Apache Beam

**Databases & Warehousing:** PostgreSQL, MS-SQL Server, MongoDB, BigQuery, NoSQL, KQL

**Big Data & Streaming:** Apache Kafka, PySpark (Spark), Hadoop, HDFS, HBase, Zookeeper, Debezium, Protobuf

**Orchestration & DevOps:** Apache Airflow, Kubernetes (GKE), Docker, CI/CD, Terraform, CloudFormation

**Cloud Platforms:** GCP (Dataflow, Pub/Sub, BigQuery, DataStream, Cloud SQL) | Azure (Synapse, ML Studio, ADF) | AWS (EMR, IAM)

**Analytics & Frameworks:** dbt, Looker, Power BI, PyTorch, Data Vault 2.0, Django

**Leadership & Practices:** Engineering Roadmap Ownership, Cross-Functional Stakeholder Management, Team Mentoring & Onboarding, Agile/Scrum, LLM Integration Strategy, CDC, ATS-Optimized Documentation