

Prem Vishnoi  
Head of Data | Enterprise AI Architecture | Data Strategy

M+65-9271-6405 | [vishnoiprem@gmail.com](mailto:vishnoiprem@gmail.com) [LinkedIn](#) | [medium](#)

### PROFESSIONAL SUMMARY

- Head of Data data and AI with 15+ years of experience building enterprise scale data platforms for global giants (Alibaba, CP Group, PayPal).
- Proven track record managing **\$8B+ revenue impact, \$2M+ annual budgets**, and leading cross-functional teams of **25+ engineers**.
- Expert in modernizing legacy systems into **Databricks Lakehouse and Real-Time Streaming architectures**.
- Specialized in bridging technical execution with executive business goals across Finance, Supply Chain, and E-commerce in Southeast Asia..

### KEY SKILLS

- **Data Architecture and Governance:** Enterprise Data Strategy, Data Lakehouse, Delta Lake, Unity Catalog, Data Lineage, Data Contracts, GDPR/PDPA/MAS Compliance
- **Big Data Engineering:** Apache Spark, Kafka, Apache Flink, Airflow, Real-Time Analytics
- **Cloud Platforms:** AWS, Azure, Alibaba Cloud, Databricks (Unity Catalog, MLflow, Delta Lake), Multi Cloud Architecture
- **Programming:** Python, SQL, Scala, Java, PySpark, Git, CI/CD, Docker, Kubernetes
- **AI & ML Platforms:** Generative AI, LLMs, RAG, NLP, MLOps (MLflow), Predictive Modeling, PyTorch, TensorFlow, Scikit-learn
- **Tools & Collaboration:** Atlassian Stack (Jira, Confluence), Power BI, Agile/Scrum, Stakeholder Management
- **Leadership:** Enterprise Data Strategy, Team Scaling (0 to 25+), Budget Management (\$2M+), Vendor Negotiation, C Level Stakeholder Management.

### EXPERIENCE

#### Data Engineering | [Makro \(CP Group\)](#) | July 2024 – Present

*Own the data platform powering an \$8B retail and e-commerce business across 3 countries, driving architecture, innovation, and team scaling.*

- **Strategic Leadership:** Scaled engineering team from 8 to 25 and managed a \$2M annual cloud/vendor budget. Reduced incident volume by 60% through proactive SRE practices.
- **Architecture Modernization:** Built a Databricks Lakehouse handling 10B+ daily rows with 99.99% uptime, serving as the single source of truth for 500+ daily jobs.
- Launched financial reporting across 3 countries, reducing month-end close from 5 days to 2 hours.
- Developed supply chain analytics (inventory tracking, demand forecasting) reducing stock-outs and improving vendor performance.
- Designed a unified O2O customer data platform and real-time order fulfillment system, cutting delivery time by 40%.
- **Cost Efficiency:** Re-architected cloud infrastructure, achieving a 40% reduction in AWS/Databricks costs via instance tuning and spot market utilization.

## Principal Data Engineer | [Xendit.co](#), SG January 2024 - June 2024

Drove major improvements in payment and vendor data systems, delivering real-time scalability, cost efficiency, and performance gains across key pipelines

- Strategic Cost Optimization: Orchestrated a FinOps initiative that reduced AWS/Databricks spend by 40% while handling increased transaction volumes.
- Pipeline Optimization: Refactored DBT and Spark jobs, cutting execution time by 60% and improving SLA adherence for payment settlements.
- Designed and implemented an automated FX data pipeline, cutting manual effort by 80% and improving operational agility.
- Real-Time Payments: Designed a Kafka-based streaming architecture for event-driven transaction processing, improving fraud detection latency.
- Led a 4-member engineering team to build and optimize batch and real-time pipelines for payments and checkout systems.

## VP Data Engineer | [Lazada Group \(Alibaba\)](#) | Jul 2018 – Jan 2024

Led architecture and development of real-time, batch, and data product systems, building a cloud-scale data infrastructure powering Southeast Asia's leading e-commerce platform

- **Platform Scale:** Built a cloud-native Data Lakehouse (ODS/CDM/ADS) on Alibaba Cloud, supporting 10x data growth and processing 100M+ daily events.
- **Real-Time Logistics:** Architected a real-time analytics engine (Flink + ClickHouse) for last-mile delivery, processing 20M+ insights daily to optimize driver routes.
- Data Products: Owned 600+ production tables and data assets, directly supporting monetization products and internal BI for logistics and supply chain.

## Big Data Consultant | [SCB Bank](#), Singapore | Feb 2016 – Jul 2018

Built large-scale data systems to support Anti-Money Laundering (AML) compliance across 15+ countries, ensuring performance, scalability, and regulatory readiness.

- Built large-scale Anti-Money Laundering (AML) systems for 15+ countries using Hadoop/Hive.
- Designed "Lucid Search" application for high-risk entity identification, ensuring MAS regulatory compliance.

## Big Data Developer | [DBS](#) | Sep 2015 – March 2016

Built Hadoop platform to integrate Consumer Banking data, create enterprise tables in Hive, and enable real-time ingestion from IBM MQ using Flume

- Implemented the project end-to-end in Hadoop, from architecture to delivery.
- Designed an SCD2 framework to manage historical data in Hive.
- Handled batch, transactional, and master data processing.
- Implemented Hadoop ingestion frameworks for Consumer Banking data using Flume and Sqoop.

## **ETL Data Engineer | [PayPal](#) | Apr 2014 – Sep 2015**

Engineered the integration of customer, payment, and transaction flow data from diverse source systems into the enterprise data warehouse.

- Transitioned into Big Data engineering by contributing to PayPal's shift from traditional ETL systems to distributed data platforms
- Migrated legacy ETL workflows to Hadoop/Hive, significantly reducing processing time for transaction data.

## **Data Warehouse & Software Engineer | [Exilant Tech \(Apple Inc\)](#) | Jan 2011 – Apr 2014**

Foundation role building enterprise data warehouse solutions and work on end-to-end SDLC cycle

- Primary responsibilities include interacting with the onsite and leading the offshore team and work based on daily requests.
- Learned and applied key concepts in data warehousing, data modeling, and enterprise architecture.
- Worked for data modelling for new objects creation in core and semantic layer
- Supported ETL processes and began writing scripts to automate data handling tasks.

## **AI Architecture & Solutioning:**

- Led architecture design for Generative AI applications for financial and retail use cases
- Deployed NLP and LLM-powered systems for real time document summarization, customer 360, and predictive reporting.
- Integrated AI pipelines into enterprise data lakes with full auditability, model monitoring, and governance compliance.
- Architected MLOps pipelines using MLflow, Databricks, and custom deployment strategies on cloud (AWS & Azure).

## **EDUCATION**

**M.S. in Data Science (In Progress), Northwestern University, USA | Sep 2025 - Jan 2027**

**PGP in AI & ML, UT Austin, USA | Jan 2019 - Jan 2020**

**B.E. in Electronics & Communication, University of Rajasthan, India | July 2005-July 2009**

## **PROFESSIONAL CERTIFICATIONS**

- AWS Solutions Architect Associate by AWS - Issued Dec 2023
- Generative AI [Educative.com](#)
- Delta Lake Essentials – Databricks
- Agentic AI Systems and Optimization CMKL Professional Training Center 2025

## **MISCELLANEOUS INFORMATION**

**Awards:** Winner PayPal Hackathon; Double Promotion at Lazada

**Residency:** Vietnam TRC | **Work Experience:** Singapore and Thailand