Prem
Data and AI Engineer

+66-81-902-7785 | vishnoiprem95@gmail.com

PROFESSIONAL SUMMARY

- **Data & AI** with 10+ years of experience building enterprise-scale data platforms for global leaders (Alibaba, CP Group, PayPal).
- **Proven Executive Leadership:** Managed $8B+ revenue impact and $2M+ annual budgets, while scaling and leading cross-functional teams of 25+ engineers.
- **Modernization Expert:** Specialist in transitioning legacy infrastructure to **Databricks Lakehouse** and **Real-Time Streaming** architectures (Kafka/Flink).
- **Strategic Visionary:** Expert at bridging technical execution with executive business goals across Finance, Supply Chain, and E-commerce in Southeast Asia.

KEY SKILLS

- **Data Architecture & Governance:** Enterprise Data Strategy, Data Lakehouse, Delta Lake, Unity Catalog, Data Lineage, Data Contracts, GDPR/PDPA/MAS Compliance
- **Big Data:** Apache Spark, Kafka, Apache Flink, Airflow, Real-Time Analytics
- **Cloud Platforms**: AWS, Azure, Alibaba Cloud, Databricks (Unity Catalog, Delta Lake), Multi Cloud Architecture
- **Programming:** Python, SQL, Scala, Java, PySpark,Teradata, Git, CI/CD, Docker, Kubernetes
- **AI & ML Platforms:** Generative AI, LLMs, RAG, NLP, MLOps (MLflow), Predictive Modelling, 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

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.

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.
- **Strategic AML Platform**: Developed and automated data pipelines for Anti-Money Laundering (AML) across 15+ global markets, ensuring scalability and MAS regulatory compliance.
- **Advanced Search Solutions**: Built a Lucid Search application for high-risk entity identification, optimizing name-matching algorithms to detect suspicious financial activity.
- **Optimization** & Scale: Utilized Hadoop, Python (MapReduce), and Hive to process complex behavioural patterns, reducing query latency and improving transaction monitoring throughput.
- **Automation**: Orchestrated end-to-end data flows with Sqoop and Shell, creating a fully automated and auditable data lifecycle for all member countries

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

- **Real-Time Architecture**: Built a low-latency ingestion framework using Apache Flume, streaming live transaction events from IBM MQ directly into the data lake.
- **Data Integrity Strategy**: Designed a custom SCD2 (Slowly Changing Dimension) framework within Hive, ensuring 100% accuracy in tracking historical changes for customer profiles and billing data.
- Enterprise Integration: Developed robust Sqoop and hive pipelines to migrate critical batch and master data from Teradata to HDFS, creating a unified view of the customer.

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 Teradata systems to distributed data platforms

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
- Built retail analytics pipelines for Apple using Informatica, Teradata, Shell, and Java.
- Designed data models and automated ETL processes to support global store performance and sales insights using BO.
- Collaborated with onsite Apple teams and led offshore delivery, handling daily data requests and ensuring timely delivery of production jobs

AI Architecture and 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 Centre 2025

MISCELLANEOUS INFORMATION

Awards: Winner PayPal Hackathon; Double Promotion at Lazada