CIMB Hiring Hackathon: Problem Statement

Current State

CIMB actively supports efforts to strengthen financial crime prevention through advanced analytics and artificial intelligence. Within the bank, ongoing initiatives focus on Anti-Money Laundering (AML), fraud detection, and suspicious transaction monitoring using both rule-based systems and machine learning models. However, due to data confidentiality and regulatory constraints, internal datasets cannot be shared externally. To foster innovation and capability building in this domain, we are leveraging **a publicly available bank transaction dataset** that simulates realistic transactional behavior. This allows participants to experiment freely with fraud detection, graph analytics, and generative AI techniques in a safe, compliant environment.

Problems / Pain Points

Financial institutions face challenges such as:

- The **increasing sophistication of fraud and money laundering patterns**, making rule-based systems insufficient.
- **Data complexity and volume**, requiring scalable analytical methods.
- The **difficulty of identifying subtle behavioral anomalies** across customers, devices, and merchants.
- **Explainability gaps** in AI-driven decisions, which are critical for compliance and investigator trust.

By simulating similar challenges through this dataset, participants can focus on solving real-world analytical pain points faced by banks like CIMB — from detecting hidden relationships to improving model interpretability and alert prioritization.

Dataset

The dataset used in this challenge is the **"Bank Transaction Dataset for Fraud Detection"** (sourced publicly from Kaggle).
It mimics real-world financial transactions and includes diverse attributes such as transaction amount, type, channel, device ID, merchant ID, customer demographics, and behavioral indicators.
While it does **not contain any real customer or confidential data**, it is rich enough to

enable exploration of fraud detection techniques, graph-based feature engineering, anomaly detection, and explainability methods that mirror practical banking applications.

Objectives / Goals

The goal of this exercise is to:

- Encourage participants to **apply data science and generative AI** techniques to address fraud detection challenges.
- **Experiment with graph analytics**, behavioral modeling, and synthetic labeling to simulate realistic fraud detection processes.
- **Extract meaningful insights and patterns** that could conceptually support AML or fraud prevention frameworks in a real banking context.
- **Demonstrate explainability and business relevance**, showing how data-driven findings can support investigators or compliance teams in decision-making.