

Etapa 1 – Definición del Problema

1. Contexto y Objetivo

El banco requiere una arquitectura moderna en **Google Cloud Platform (GCP)** que integre y procese grandes volúmenes de información transaccional con el fin de **detectar fraudes en tiempo casi real y fortalecer la analítica de clientes**.

El propósito es disponer de una solución **escalable, segura y eficiente en costos**, que combine procesamiento en streaming y batch, facilite el análisis ejecutivo mediante tableros BI y habilite capacidades de Machine Learning integradas al ciclo operativo.

2. Casos de Uso

- **Detección de fraude:** identificar transacciones sospechosas en milisegundos, utilizando predicción en línea a través de Vertex AI Endpoints.
- **Analítica de clientes:** consolidar información en BigQuery para elaborar KPIs estratégicos y segmentaciones dinámicas que apoyen decisiones comerciales y regulatorias.

3. Fuentes de Datos

Tipo	Fuente	Descripción	Frecuencia
Streaming	Transacciones de tarjetas	Eventos financieros con atributos de monto, canal, ubicación y comercio.	Casi en tiempo real
Batch	Maestro de clientes	Identificación, perfil y atributos demográficos.	Diario
Batch	Catálogo de comercios	Datos de rubro, riesgo y límites de operación.	Diario

4. Indicadores Clave (KPIs)

- Tasa de fraude sospechado vs. confirmado (%).
- Latencia promedio del scoring (P95 ms).
- Volumen transaccional por hora/día (#).
- Pérdida evitada estimada (\$).

5. Supuestos y Restricciones

- Latencia máxima de inferencia: ≤ 500 ms.
- Procesos batch orquestados con **Cloud Composer**, almacenamiento en **Cloud Storage** y consolidación en **BigQuery**.
- Seguridad bajo **principio de mínimo privilegio (IAM)** y cifrado con **Cloud KMS**.
- Monitoreo operativo con **Cloud Logging / Monitoring** y control presupuestario mediante **Budgets & Alerts**.
- Cumplimiento de estándares de **disponibilidad, auditoría y gobernanza de datos**.