Universidad Francisco Marroquín



Facultad de Ciencias Económicas

Big Data

Nombre:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Carné:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**VARIANTE 2: GCP - Detección de Fraude Financiero**

**Narrativa del Caso de Uso**: FinSafe Payments es un procesador de pagos que maneja transacciones para más de 10,000 comercios en 15 países. Recientes ataques sofisticados de fraude han resultado en pérdidas de 50 millones de dólares en el último trimestre.

**Necesidad del Negocio**:

* Procesar transacciones en tiempo real para detección inmediata de fraudes
* Tomar decisiones instantáneas a través de dashboards de monitoreo para el equipo de seguridad
* Procesar información histórica en batch para identificar nuevos patrones de fraude
* Compartir información en tiempo real con modelos de ML para detección adaptativa
* Enriquecer transacciones con datos históricos del usuario para mejor evaluación de riesgo

**Sistemas Actuales y su Arquitectura**:

**Flujo de datos actual (paso a paso)**:

1. Una transacción llega al gateway de pagos desde una tarjeta de crédito
2. La transacción pasa por 300 reglas estáticas (monto máximo, país, frecuencia)
3. Si pasa las reglas, se guarda en Oracle Database y se aprueba
4. Si falla alguna regla, se marca como sospechosa y se rechaza
5. Los analistas revisan manualmente las transacciones rechazadas al día siguiente
6. Cada noche, un proceso batch exporta todas las transacciones a un data warehouse
7. Una vez al mes, los científicos de datos entrenan nuevos modelos con datos históricos
8. Los nuevos modelos se despliegan manualmente, reemplazando las reglas antiguas
9. El ciclo se repite, siempre con un mes de retraso en la detección de nuevos patrones

**Problemas que debe resolver la nueva arquitectura**:

* Detectar fraudes sofisticados que evaden reglas simples
* Actualizar modelos de ML en tiempo real con nuevos patrones
* Reducir falsos positivos que frustran a clientes legítimos
* Correlacionar múltiples señales en tiempo real

**Requisitos Técnicos**:

* 20 TB/día de datos históricos
* 200,000 transacciones/segundo en tiempo real
* Latencia máxima: 100ms para detección

**Se solicita**:

1. **Diagrama de arquitectura Lambda completo** que muestre:
   * Todos los componentes GCP utilizados
   * Cómo interactúan entre sí
   * Flujo de datos en las tres capas (batch, speed, serving)
   * Conexiones y dependencias entre servicios
2. **Análisis detallado justificando**:
   * Por qué se seleccionó cada componente GCP
   * Cómo cada servicio resuelve requisitos específicos
   * Ventajas técnicas de cada elección

**Criterios de Evaluación**:

* Correcta implementación de las 3 capas Lambda (20%)
* Diagrama arquitectónico claro y completo (30%)
* Selección apropiada de servicios cloud (30%)
* Justificación coherente de decisiones técnicas (20%)