

Plan de Trabajo del Proyecto Final

Bases de datos no relacionales - 2025

1. Identificación del grupo

- Número del grupo: 14
- Integrantes (nombre completo y número de cédula):
 - Andres Santucci - 4.978.868-1
 - Pablo Recarte - 4.906.683-7

2. Familia de proyecto elegida

Familia 4: Coast to coast (de la obtención al análisis, posible solución poliglota)

3. Objetivos del proyecto

- Construir una base de datos de grafos a partir de un conjunto de datos de transacciones financieras con etiquetas de fraude.
- Detectar patrones, estructuras y comunidades (clusters) relacionadas con actividades fraudulentas.
- Visualizar conexiones entre cuentas, dispositivos e IPs asociadas a fraude para facilitar el análisis forense.

4. Descripción general del proyecto

El proyecto propone modelar transacciones financieras como un grafo, donde los nodos representan cuentas, dispositivos, IPs, locaciones y otros elementos clave, y las aristas representan interacciones como transacciones y uso compartido de ciertos recursos. El objetivo es detectar estructuras sospechosas, comunidades de fraude y comportamientos anómalos difíciles de identificar con enfoques tradicionales. La motivación surge del interés en el análisis de fraudes mediante grafos, una técnica cada vez más usada en bancos, fintechs y seguridad informática.

5. Actividades previstas y cronograma tentativo

Completar la tabla con las actividades previstas para alcanzar los objetivos, estimando una duración tentativa para cada una (total aproximado: 60 horas).

Actividad	Breve descripción	Semana(s) prevista(s)
Recolección de datos	Identificar y descargar datasets relevantes (Financial Transactions Dataset).	Semana 1
Limpieza y preparación	Normalizar, eliminar duplicados.	Semana 1
Modelado del grafo	Dividir csv según nodos a construir. Definir nodos y relaciones a partir de la realidad. Definir esquema conceptual del grafo.	Semana 2
Carga en Neo4j	Cargar nodos y relaciones a partir de los csv en Neo4j.	Semana 3
Pruebas sobre la base y definir consultar a realizar	Consultas básicas con Cypher y definir a detalle qué conclusiones queremos obtener con las consultas a construir.	Semana 3
Análisis sobre el grafo	Realizar análisis de comunidades o demás algoritmos aplicables a la realidad modelada.	Semana 4
Estudio de resultados	Entender resultados obtenidos, generar visualizaciones para los patrones detectados.	Semana 4
Redacción del informe	Documentar el proceso y las conclusiones obtenidas.	Semana 5
Preparación presentación		Semana 5

6. Base teórica o artículos relacionados

¿El proyecto se inspira en algún artículo académico o proyecto existente?

- No

7. Datos y entornos

- Financial Transactions Dataset for Fraud Detection ([link](#))
 - Fuente: Kaggle
 - Tipo: Transacciones financieras
 - Formato: CSV
- Neo4j, Python