Banking Transaction Analysis using Neo4j Graph Database & Graph Data Science

1. Dataset Overview

The banking data has three tables - Customer, Purchases and transfers table with details of customer's demographics, transaction data related to account and banking activities.

Customer Table

CIF	Age	EmailAddress	FirstName	LastName	PhoneNumber	Gender	Address	Country	JobTitle	CardNumber	AccountNumber
1	26	Greta_Swan6281@corti.com	Greta	Swan	6-254-223-4334	Female	Armory Tunnel, 3699	Egypt	Bellman	348-14-6088	650-63-6154
2	41	Rebecca_Corbett5111@ovock.tech	Rebecca	Corbett	5-720-670-7047	Female	Apostle Tunnel, 4850	Madagascar	Clerk	285-83-6186	875-13-4278
3	68	Dani_Flanders7065@qater.org	Dani	Flanders	8-150-728-6781	Female	Bishop Street, 1298	The Gambia	Business Broker	811-84-3310	566-83-8044
4	68	Morgan_James9389@sveldo.biz	Morgan	James	7-215-157-6623	Female	Wadham Avenue, 8018	Guinea-Bissau	Banker	773-80-8680	532-16-1750
5	51	Erick_Ingram7572@extex.org	Erick	Ingram	8-603-106-4884	Male	Chamberlain Pass, 3481	Russia	Electrician	171-24-1530	683-00-5812
6	52	Hayden_Garcia8505@extex.org	Hayden	Garcia	2-225-676-6758	Male	Shepherds Grove, 5504	Solomon Islands	Banker	110-44-7471	060-21-3317
7	62	Mark_Jobson9521@eirey.tech	Mark	Jobson	4-301-266-1565	Male	Parkfield Walk, 4086	Japan	Systems Administrator	425-86-2712	611-71-0203
8	62	Candace_Shea4131@gompie.com	Candace	Shea	4-554-734-5866	Female	Geffrye Pass, 1418	Montenegro	Loan Officer	830-04-7362	820-62-2326
9	56	Rowan_Harper2387@grannar.com	Rowan	Harper	1-046-822-1774	Female	Lexington Road, 7414	Azerbaijan	Treasurer	564-28-0407	402-45-6427
10	73	Janice_Simmons7212@nanoff.biz	Janice	Simmons	1-767-728-0781	Female	Badric Boulevard, 9856	Nicaragua	Assistant Buyer	077-14-3850	827-22-6062

Transfers Table

TransactionID	SenderAccountNumber	ReceiverAccountNumber	Amount	TransferDatetime
287062	650-63-6154	430-04-5447	8584.094218	2021-02-21 16:21:49Z
377610	875-13-4278	650-63-6154	193772.0259	2021-04-21 03:09:06Z
513428	566-83-8044	875-13-4278	138147.425	2021-05-21 22:03:26Z
571917	532-16-1750	566-83-8044	92918.53787	2021-07-22 14:35:48Z
418959	683-00-5812	532-16-1750	83355.47712	2021-02-03 07:53:22Z
226117	060-21-3317	683-00-5812	156637.1547	2021-02-26 02:03:07Z
794690	611-71-0203	060-21-3317	33307.38709	2021-08-26 07:19:39Z
545007	820-62-2326	611-71-0203	182842.6328	2021-08-10 08:57:43Z
865016	402-45-6427	820-62-2326	71638.94337	2021-06-07 05:12:58Z
829154	827-22-6062	402-45-6427	133408.4168	2021-03-23 19:12:00Z
866605	624-06-2107	827-22-6062	91611.11143	2021-03-28 20:38:32Z

Purchases Table

TransactionID	CardNumber	Merchant	Amount	PurchaseDatetime	CardIssuer
751133	348-14-6088	Biolife Grup	9849.265683	2021-04-12 15:32:22Z	Bank of America
268445	285-83-6186	Facebook	2502.65753	2021-09-05 04:27:34Z	Visa
748511	811-84-3310	Comodo	6406.732352	2021-01-04 05:00:18Z	Chase
777162	773-80-8680	Demaco	12698.28751	2021-07-17 18:34:49Z	American Express
523244	171-24-1530	Erickson	13496.90367	2021-06-11 15:34:42Z	Chase
976171	110-44-7471	CarMax	5193.496	2021-02-06 04:47:31Z	Citibank
870688	425-86-2712	Telekom	18675.95245	2021-06-26 01:12:36Z	Wells Fargo
506821	830-04-7362	Areon Impex	3389.492109	2021-01-26 02:37:42Z	MasterCard
435827	564-28-0407	Facebook	5655.183243	2021-03-13 12:40:20Z	Citibank
610627	077-14-3850	Demaco	10793.82236	2021-03-22 22:50:49Z	Discover
579725	748-68-5701	Amazon.com	13233.63714	2021-07-19 17:02:03Z	Chase

Key Entities in this data are -

1. **Customer**: it could be an individual or a company. A bank client who uses bank services and has a relation with the bank.

    a. Attributes- CustomerID(CIF), full name, address, contact number, age, gender, home address, email address, country and JobTitle.

2. **Account**: Represents customer accounts with attributes such as account number.
3. **Purchase**: Represent customer purchases made from the card to Merchant with attributes like transaction_id, amount and date of purchase.
4. **Merchant:** Represents the Business entity to which payment is made.
5. **Card:** Represents the card information owned by the customer

Relationships

- Customer OWNS Account
- Customer LINKED_TO Card
- Account TRANSFERRED_TO Account
- Card USED_FOR Purchase
- Purchase MADE_AT Merchant

Why is Graph-based analysis powerful for Banking Analysis?

Banking data is generally large and interconnected with information ranging in multiple tables and all the tables are linked with complex relationships, which makes graph databases an effective choice for Banking Data Analysis. Traditional relational databases store the tables in silos and link these tables by performing expensive joins to derive insights, whereas graph databases natively store relationships. This enables faster queries and deeper insights without performance bottlenecks.

Banking data consists of multiple tables with different relationship types, for instance - **customer–account, account–transaction, and sender–receiver**— which can be easily represented using different edge types in a graph model. Graph databases allow seamless traversal of these connections, providing a **richer context** compared to relational models, where data often exists in silos. A fraudulent pattern of multi-hop transactions might appear normal when analyzed with a smaller transaction network.

Graph Analysis for Fraud Detection:

Identifying Financial fraud often aims to derive unusual patterns within a network of transactions. Representing this network in a graph model helps to understand the flow of money and uncover hidden risks. Graph algorithms can detect suspicious patterns such as: Cyclic Transaction (money returning to source account after a chain of transactions), or Connected Clusters (money circulating in a small group of accounts indicating money laundering). Identifying the money trail is easier in a graph through multi-hops to trace the complete path of fraudulent activity.

By leveraging graph models, banks can visualize and monitor financial movements in real-time, making it easier to detect and prevent fraud before it escalates.

Potential Analytical Use Cases

1. **Fraud Detection:**
    - Identifying unusual transaction patterns by analyzing customer spending behaviors and abnormal transfers between accounts.
    - Detecting common fraud patterns such as frequent high-value transactions, sudden spikes in transfers, transaction loops, financial activities with countries like Iran or Syria which are considered risky for Money Laundering.
2. **Customer Segmentation**:
    - Graph based community detection algorithms groups the customers based on transaction behavior, account relationships, spending patterns (High Networth, etc.) and product usage.
    - Clusters can be used to create targeted marketing strategies (Product Recommendation), to predict customer churn, identify LTV of customers and to derive Customer's Happy Profile Assessment.
3. **Entity Deduplication:**
    - Graph algorithms like Entity Resolution can be used to identify duplicate customer records or uncover fraudulent accounts.
    - Deduplication of Prospects helps in reducing unnecessary marketing or sales efforts and deduplication on Fraudulent accounts helps in identifying Fraudsters (individuals with fake identities, or with modified personal details) across different financial institutions.

4. **Risk Assessment:**
    - Graph algorithms can identify influential accounts based on transaction volume, which helps in differentiating it from legitimate accounts.
    - Identifying accounts that are well-connected within the transaction network, which could indicate high-value customers, potential money laundering activity or could be loan defaulters.

2. Graph Data Modeling and Data Ingestion

Key Entities (Nodes) and Properties

For this given dataset, here are the primary node types:

S.No.	Entity	Properties
1	Customer	CustomerId, Name, Email, Age, Phone, Address, Gender, Country, JobTitle
2	Account	AccountNumber, CustomerId
3	Card	CardNumber, CardIssuer, CustomerId
4	Purchase	TransactionId, Amount, TransactionDate
5	Merchant	MerchantName

Relationships

S.No.	Relationship	From->to
1	OWNS	Customer-to-Account
2	LINKED_TO	Customer-to-Card
3	TRANSFERRED_TO	Account-to-Account
4	USED_FOR	Card-to-Purchase
5	MADE_AT	Purchase-to-Merchant

Constraints and Indexes

The below constraints guarantee uniqueness of a certain property on nodes with a specific label.

Node Key Constraints:

CREATE CONSTRAINT ON (c:Customer) ASSERT c.CustomerID IS UNIQUE;

CREATE CONSTRAINT ON (a:Account) ASSERT a.AccountNumber IS UNIQUE;

CREATE CONSTRAINT ON (t:Purchase) ASSERT t.TransactionID IS UNIQUE;

CREATE CONSTRAINT ON (cd:Card) ASSERT cd.CardNumber IS UNIQUE;

CREATE CONSTRAINT ON (mr:Merchant) ASSERT mr.MerchantName IS UNIQUE;

Adding the unique constraint will add an index on that property, so we don't need to add it separately. We can create index for properties without a unique constraint-

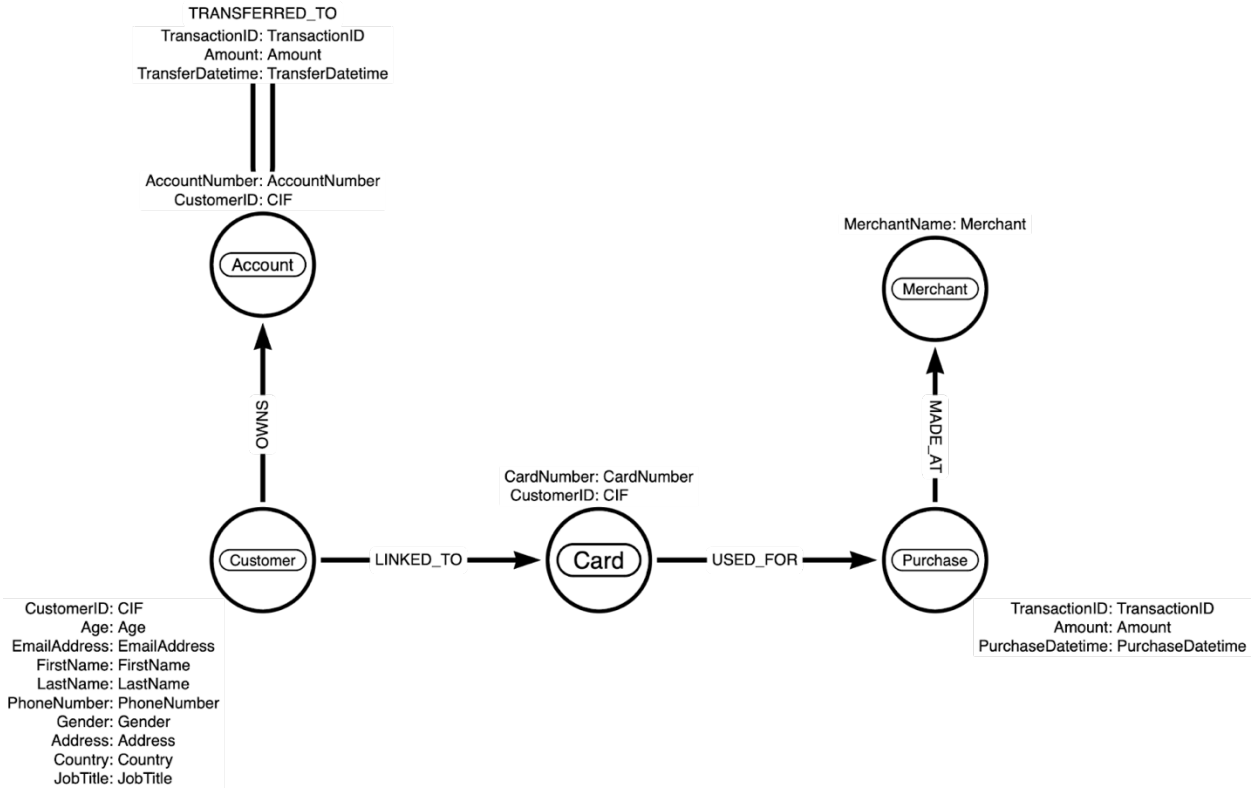
- For frequently used in searches, sorting, or aggregations to improve performance.
- Or a composite index on multiple properties.

Property Indexes:

CREATE INDEX customer_name IF NOT EXISTS FOR (c:Customer) ON (c.FirstName);

CREATE INDEX transaction_index IF NOT EXISTS FOR (t:Purchase) ON (t.Amount, t.PurchaseDatetime);

The above model design provides an efficient traversal of the given financial network, and maintains appropriate constraints for data integrity.



3. Cypher Query Development

Notebook Attached.