

Datasets

I. Datasets Overview

Our dataset is retrieved from Centic's data collection, which comprises two types of data: the knowledge graph, which maps the complex relationships between on-chain and off-chain entities in the Web3 domain, and the raw data, which includes the unprocessed and subsequently decoded blockchain data.

We only query a portion of the Centic data collection—the information required to identify high-risk or unusual transactions—due to the size of the collection overall. We choose to query Polygon (0x89), BNB Chain (0x38), and Ethereum (0x1).

The information that we extract from each table in the knowledge graph and raw data is summarized in the following sections. Each table is detailed in the table that includes the field name, data type, and description.

II. Raw data

a. MongoDB database

i. Transactions

| Field | Data type | Description |
|-----------------|-----------|--|
| hash | String | The hash of the transaction |
| from_address | String | Address of the sender |
| to_address | String | Address of the receiver. null when its a contract creation transaction |
| value | String | The amount of native token sent in this transaction in wei. Note that ERC20 tokens do not show up here |
| gas | String | The gas limit in wei |
| gas_price | String | The gas price in wei |
| block_timestamp | Int32 | The time when the block was mined that includes this transaction |

| | | |
|------------------|--------|--|
| block_number | Int32 | The length of the blockchain in blocks |
| receipt_gas_used | String | The gas consumed by the transaction in wei |

ii. Lending events

| Field | Data type | Description |
|------------------|-----------|--|
| type | String | Type of collection |
| amount | Double | The amount of reserve in action. The value has been divided by token decimals. |
| block_timestamp | Int32 | The time when the block was mined that includes this transaction |
| transaction_hash | String | The hash of the transaction |
| contract_address | String | Address of interacted contract |
| event_type | | Type of lending event |
| on_behalf_of | String | Address of wallet that received lending event result |
| user | String | Event's input wallet address |
| wallet | String | Address of wallet made the transaction |

iii. Token

We take all the tokens in each chain.

b. PostgreSQL database

i. Token transfer

| Field | Data type | Description |
|------------------|-----------|--------------------------------|
| contract_address | text | Address of interacted contract |
| transaction_hash | text | The hash of the transaction |

| | | |
|--------------|------------------|---|
| log_index | bigint | The logs index position in the block |
| block_number | bigint | The length of the blockchain in blocks |
| from_address | text | Address of the sender |
| to_address | text | Address of the receiver |
| value | double precision | The amount of tokens sent in this transfer. The value has been divided by token decimals. |

III. Knowledge graph

a. Wallets:

| Field | Data type | Description |
|------------------------|-----------|---|
| address | String | Address of wallet |
| chainId | String | Blockchain ID |
| balanceInUSD | Double | Wallet balance in USD |
| balanceChangeLogs | Object | History of wallet balance |
| tokens | Object | Amount of tokens that the wallet is holding |
| tokenChangeLogs | Object | History of tokens that the wallet is holding |
| depositInUSD | Double | Wallet deposit in USD |
| depositChangeLogs | Object | History of wallet deposit |
| depositTokens | Object | Amount of tokens that the wallet is depositing |
| depositTokenChangeLogs | Object | History of tokens that the wallet is depositing |
| borrowInUSD | Double | Wallet borrow in USD |

| | | |
|-----------------------------|--------|---|
| borrowChangeLogs | Object | History of wallet borrow |
| borrowTokens | Object | Amount of tokens that the wallet is borrowing |
| borrowTokenChangeLogs | Object | History of tokens that the wallet is borrowing |
| dailyAllTransactions | Object | Number of transactions the wallet address interacts daily |
| dailyNumberOfTransactions | Object | Number of transactions that the wallet address generates daily |
| dailyTransactionAmounts | Object | The amount that the wallet address receives every day |
| numberOfLiquidation | Int32 | Number of times the wallet address has been liquidated |
| totalValueOfLiquidation | Double | The amount of the wallet address has been liquidated |
| liquidationLogs | Object | Wallet liquidation history |
| frequencyOfDappTransactions | Object | Frequency of transactions occur within a particular DApp on a blockchain platform |

b. Multi-chain wallets

| Field | Data type | Description |
|-------------------|-----------|--|
| address | String | Address of wallet |
| balanceInUSD | Double | Wallet balance in USD |
| balanceChangeLogs | Object | History of wallet balance |
| tokens | Object | Amount of tokens that the wallet is holding |
| tokenChangeLogs | Object | History of tokens that the wallet is holding |

| | | |
|--|--------|---|
| depositInUSD | Double | Wallet deposit in USD |
| depositChangeLogs | Object | History of wallet deposit |
| depositTokens | Object | Amount of tokens that the wallet is depositing |
| depositTokenChangeLogs | Object | History of tokens that the wallet is depositing |
| borrowInUSD | Double | Wallet borrow in USD |
| borrowChangeLogs | Object | History of wallet borrow |
| borrowTokens | Object | Amount of tokens that the wallet is borrowing |
| borrowTokenChangeLogs | Object | History of tokens that the wallet is borrowing |
| dailyNumberOfTransactionsInEachChain | Object | Number of transactions that the wallet address generates daily in each chain |
| dailyNumberOfTransactions | Object | Number of transactions that the wallet address generates daily |
| dailyTransactionAmountsInEachChain | Object | The amount that the wallet address receives every day in each chain |
| dailyTransactionAmounts | Object | The amount that the wallet address receives every day |
| frequencyOfDappTransactionsInEachChain | Object | Frequency of transactions occur within a particular DApp on a blockchain platform in each chain |
| frequencyOfDappTransactions | Object | Frequency of transactions occur within a particular DApp on a blockchain platform |
| liquidationLogs | Object | Wallet liquidation history |

| | | |
|-------------------------|--------|--|
| numberOfLiquidation | Int32 | the count of instances where assets held in the wallet are forcibly sold or liquidated |
| totalValueOfLiquidation | Double | The amount of the wallet address has been liquidated |
| lendings | Object | Tokens that the wallet lend |

c. Smart contracts

We take all the information of the smart contracts that have the information related to the three chains: Polygon (0x89), BNB Chain (0x38), and Ethereum (0x1).