DeFi Analytics & Risk Assessment

Milestone 1 Report

Course 460/560 - Data Models and Query Languages

NAME	UB-ID
Siddhi Nalawade	50613176
Mrudula Deshmukh	50605669

Introduction

1.1. Project Overview:

The DeFi Analytics & Risk Assessment tool analyzes decentralized finance (DeFi) activities on the Ethereum blockchain, by collecting on the chain data in a relational database. The goal is to identify suspicious activities, large-volume transfers, and token changes that could suggest problems like rug pulls, flash loan exploits, and liquidity shortages

1.2 The project involves:

- Obtaining and preparing Ethereum blockchain data Designing a normalized database schema
- Creating a normalized database schema
- Executing SQL queries for transaction analysis and risk evaluation
- Formulating risk management techniques for DeFi protocols

1.3 Importance of Relational Databases for DeFi Analytics:

Relational databases are necessary for DeFi analytics because they enable efficient querying, systematic data management, and scalability. They use main and foreign keys to ensure data integrity while also providing security features such as encryption and access control. They enable reliable financial tracking, risk identification, and data-driven decision-making in decentralized finance thanks to their ACID compliance and ability to model complicated relationships.

1.4 Problem Statement:

Because of a lack of broad transaction tracking, decentralized finance DeFi companies frequently encounter difficulties like include fraud, manipulation of the markets, or crisis of liquidity. Traditional Excel file methods suffer with capacity issues, inability to manage complicated relationships, and a lack of real-time searching capabilities. A system of relational databases overcomes these restrictions by providing efficient keeping, real-time querying, advanced analytics, and scalability.

DeFi Analytics & Risk Assessment

1.5 Target Users:

- Researchers and Risk Control Teams may utilize the database to analyze transaction patterns, detect problematic steps, and make informed decisions.
- DBAs handle database performance, integrity, indexing, and backups.

Real-Life

A DeFi structure, including the or Aava, use the above database to continuously monitor big or unusual token actions, allowing for instant risk reduction measures like, blocking specific operations, notifying lenders, or starting security processes.

Dataset Acquisition and Preprocessing

→ Google Big Query dataset on Ethereum blockchain (bigquery-public-data.crypto_ethereum) concentrates on ERC-20 and ERC-721 token transfers..

→ A SQL-based filtration extract found 20,000 related tokens exchange data between January to February 2025.

→ The extracted data was uploaded into PostgreSQL in the form of a CSV, first placed in a staging database (staging_deploy), before being normalized into structural records (Block, Transaction, Token).

→ Performance optimizations, such as table separation, query the sorting, and batch processing of data have been implemented.

→ Cloud methods, like Google Cloud Dataflow and Apache Beam, were evaluated for increased scalability..

Entity-Relationship Diagram (ERD)

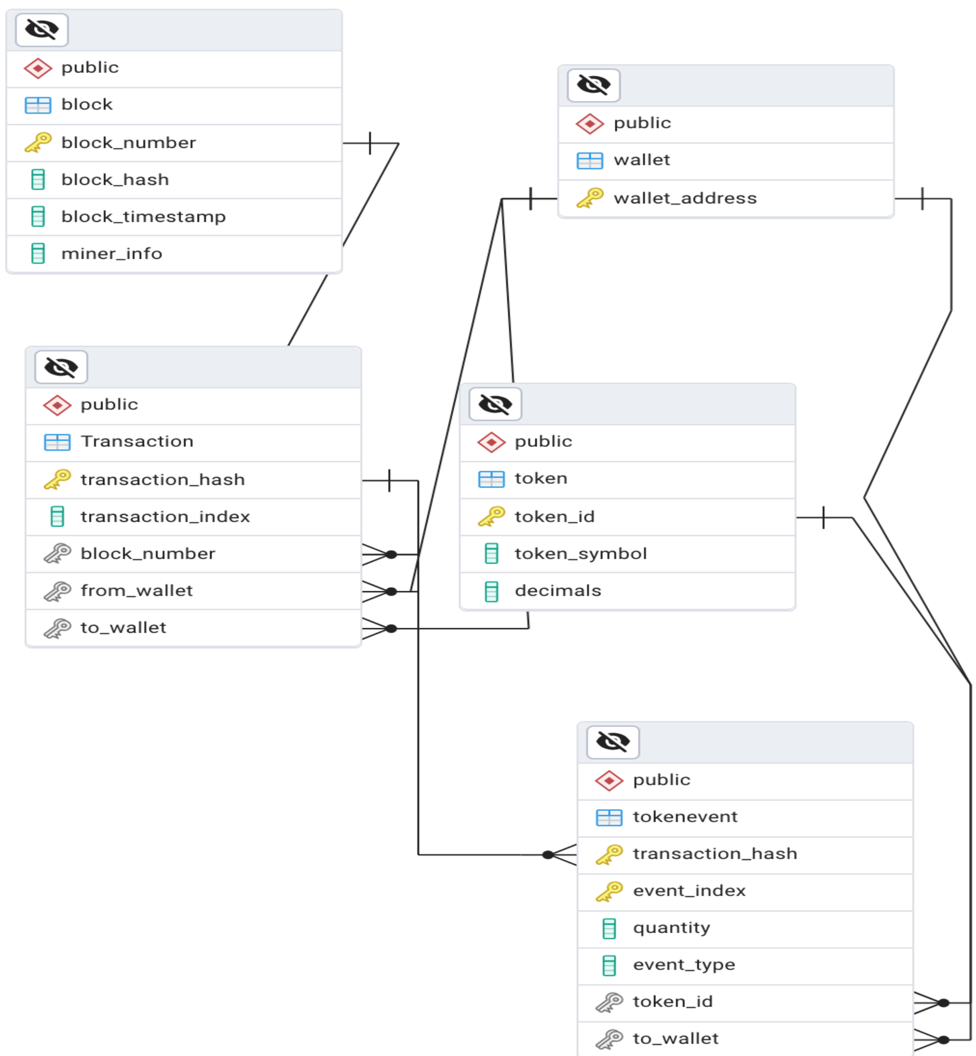
Schema Design

This section includes a list of SQL queries run using the DeFi analytics database, as well as the outcomes and analysis. These queries show the ability of the relational database to deal with complex DeFi transaction data.

DeFi Analytics & Risk Assessment

Entity Descriptions

Entity	Description
Block	Stores block metadata, including block_number, block_hash, and block_timestamp.
Wallet	Tracks unique wallet addresses involved in transactions.
Token	Stores details of different tokens (token_id, token_symbol).
Transaction	Links from_wallet and to_wallet to block_number for each transaction.
TokenEvent	Tracks token transfers within transactions (includes token_id, quantity, and to_wallet).



DeFi Analytics & Risk Assessment

Relationships:

1. Block → Transaction (one-to-many)
2. Wallet → Transaction (one-to-many)
3. Transaction → TokenEvent (one-to-many)
4. Token → TokenEvent (one-to-many)

The relationships between these items allow proper organization and allow sophisticated searches to be run via JOIN procedures without duplication. , The layout has been adjusted to the BCNF to improve speed and regularity.

SQL Queries & Results

- This section includes a list of SQL queries run on our DeFi analytics database, as well as their results and analysis. These operations demonstrate the ability of our MySQL database to handle complex DeFi transactions with ease.

Data Verification Queries

→ Retrieve the first 10 transactions from the Transaction table

```sql
SELECT * FROM "Transaction" LIMIT 10;
```

Data Output | Messages | Notifications

Showing rows: 1 to 10 | Page No: 1 of 1

	transaction_hash [PK] character varying (100)	transaction_index integer	block_number integer	from_wallet character varying (100)	to_wallet character varying (100)
1	0x034975074b8b619fb7b9e8ffa41471e30f480809165e1a5e6245e210420b1c25	109	18916205	0xb26e8cb731de2bfcff216824b13c436a802f0341	0x22f9dcf4647084d6c31b2765f6910cd85c178c18
2	0x06d7cc0f7b07249f8abab0e03ec328115d2c7553c09e42328c4ca2e705d10b68	66	18916206	0x3f7ec7f7f39137cb800d7b4a9fe359551fc53f53	0x1bfffb738d69167d5592160a47d5404a3cf5a846
3	0x09d73c69ab4bd2de738d6b7f51ee4fe5b3bd5752a09cc9ef254bda52ae793ba2	100	18916208	0xbe83e7db6c6fb4f5db4f22294faf3868ca443f2b	0xa03400e098f4421b34a3a44a1b4e5714195176...
4	0x0b436a06538201e4ff1b62f049123da9d12ed510004b70f8a83fd8d595a5238e	181	18916208	0xc9b2fcf9cda2b4710eb1464da3c517ee7e5b89...	0x80dd07c9e3522e94e0ab552db5cec76515b006...
5	0x149d3eac0ec5ca574b2e311e1db7959caf31625f7a1a0a17b8f8f911668a2410	44	18916206	0x5c4a263f9afe2865a745f3b8ee5be2e8071f3379	0x006ab11e89e7246cb8c0f7bdaa8c2713374b1676
6	0x16c92ac2d4d6057163b817e363526c817fc5e33c87c0087a7518b538a60456...	53	18916205	0x8992f6547ef3f02c52743db2a0a2f06e2d6bc523	0xa55efd8a864446473961 7962b44a6ece3d909fe5
7	0x1998e5f1c876c6aebed6ca605c3b70883b18e1dd7a7dc5fee7fa0f640a9f3f8f	36	18916205	0xae2d4617c862309a3d75a0ffb358c7a5009c673f	0x5648f73576d54860b04972b7a788c1b2d79214...
8	0x2ad51d6457d69de6d88940be1affcbc1d04a68c63928ec23c9dff214b6c4e74a	30	18916207	0xf55041e37e12cd407ad00ce2910b8269b01263...	0x36aff7001294dae4c2ed4fdefc478a00de77f090
9	0x31dfa380f61956d2d633e7130c966dd5ac94dd9e5cb413c941137a5e3b15f96d	24	18916205	0xb23360ccdd9ed1b15d45e5d3824bb409c8d7c...	0x2d5be1dd8985157684f75e9a96808c8730e4e9f3
10	0x3547cbd7bf0679117415ade2a3ccb431b09803514eb77557615fa4eb9799047e	49	18916207	0xe6f3494e839f3d3fb36c407eb35cd85d90dc3704	0x0626055c809159033c9e2fcd87a4c8d58368a683

→ Retrieve the first 10 token transfer events from the TokenEvent table.

```sql
SELECT * FROM TokenEvent LIMIT 10;
```

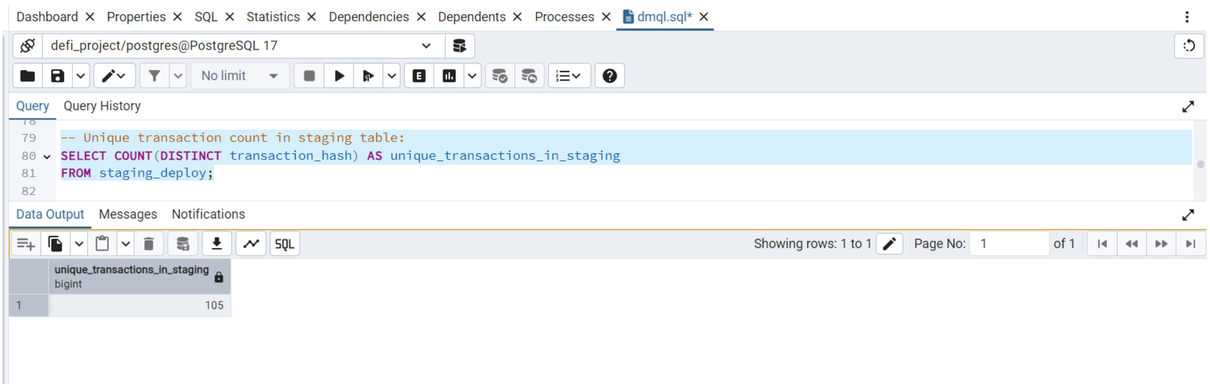
Data Output | Messages | Notifications

Showing rows: 1 to 10 | Page No: 1 of 1

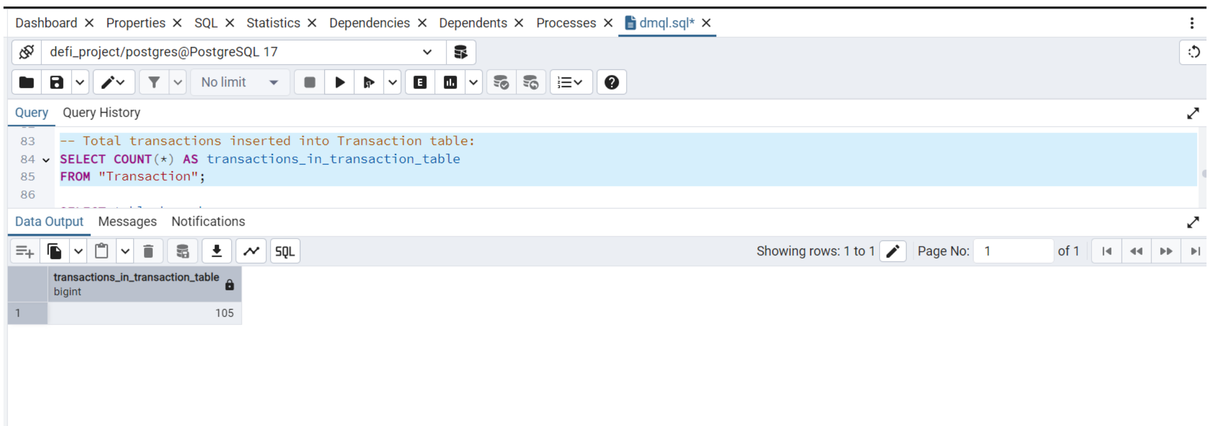
	event_id [PK] integer	transaction_hash character varying (100)	event_index integer	quantity numeric	event_type character varying (50)	token_id character varying (100)
1	1	0x7e75cb03b23da388027a23ac39e075950a4e732de41b7f2dbdeec699eedfb8bb	67	9591568489906.0	ERC-20	[null]
2	2	0x6af481e7e075137554c8b08dc5203dcc11e0673a9902b36927eb22c1f89dfc75	5118	2539458299702621600000	ERC-20	[null]
3	3	0xaf29a40ec328fa6f38b70c716420e7cf14d56590c638cc043b5373f25ecfebb5	5239	1.0	ERC-1155	12045995789001221606603461103084594685542 4
4	4	0x61dc06bb4af4368cc43fc4126fece91e7926c983b9138700736068db27274090	5142	648795616307630800000	ERC-20	[null]
5	5	0x58a58cef922b226cecaf35609f1b466518076eea8060391fa8b2474c8398fa25	5115	1993518060947295200	ERC-20	[null]
6	6	0x16c92ac2d4d6057163b817e363526c817fc5e33c87c0087a7518b538a60456...	91	30469238000000000000000	ERC-20	[null]
7	7	0xaf29a40ec328fa6f38b70c716420e7cf14d56590c638cc043b5373f25ecfebb5	5237	1.0	ERC-1155	113994592918514385260230493489642350837760
8	8	0xedce1ec9a5e5a313fd19ce6fd6a7fdc6105893252c65038b7c9a3dd01fcfb0bb	77	5643000000000000000000	ERC-20	[null]
9	9	0xaf29a40ec328fa6f38b70c716420e7cf14d56590c638cc043b5373f25ecfebb5	5238	1.0	ERC-1155	139856052804505708483446963654456734908416
10	10	0x7e75cb03b23da388027a23ac39e075950a4e732de41b7f2dbdeec699eedfb8bb	68	182239801308233.0	ERC-20	[null]

DeFi Analytics & Risk Assessment

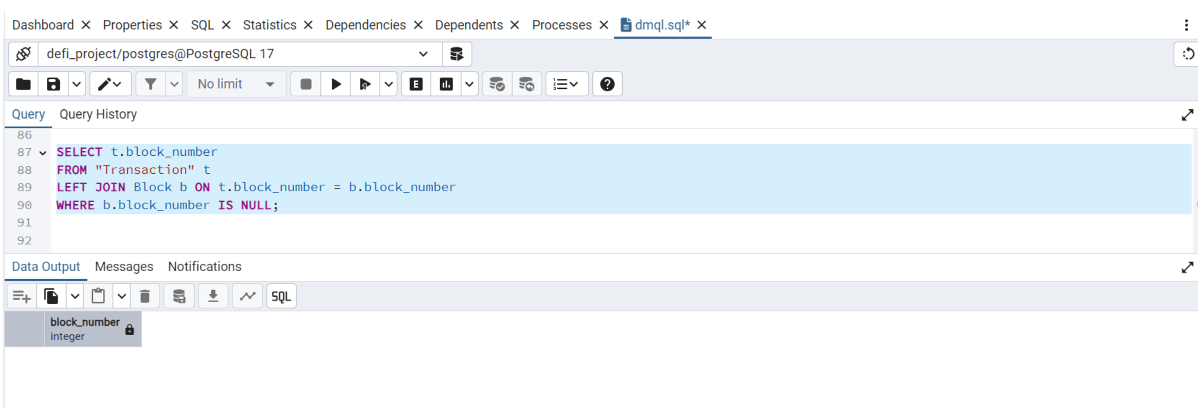
→ Count unique transactions in the staging table to verify initial data extraction.



→ Confirm total transactions inserted into the Transaction table.

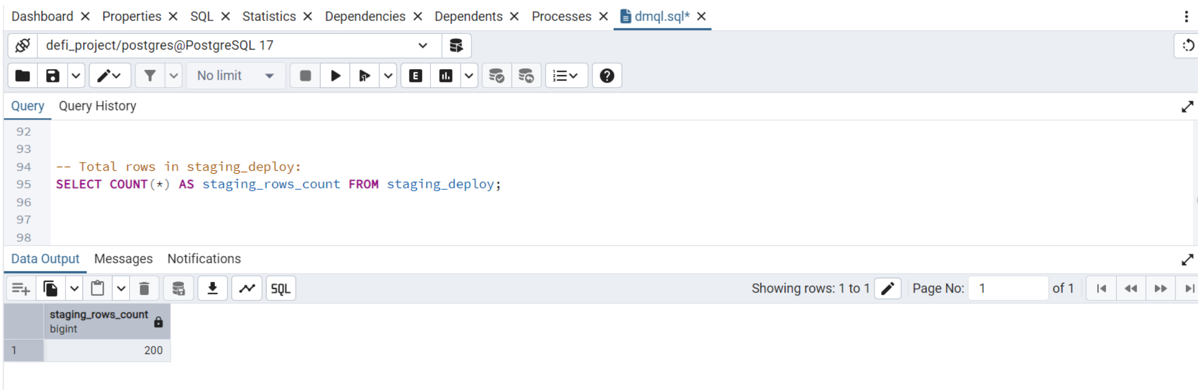


→ Verify total rows loaded into staging_deploy for data completeness.



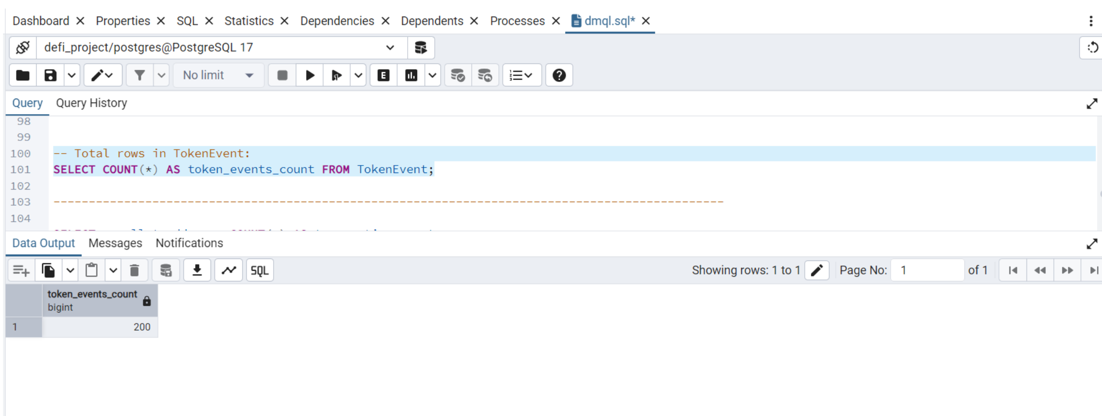
DeFi Analytics & Risk Assessment

→ Verify total number of token events stored in the TokenEvent table.

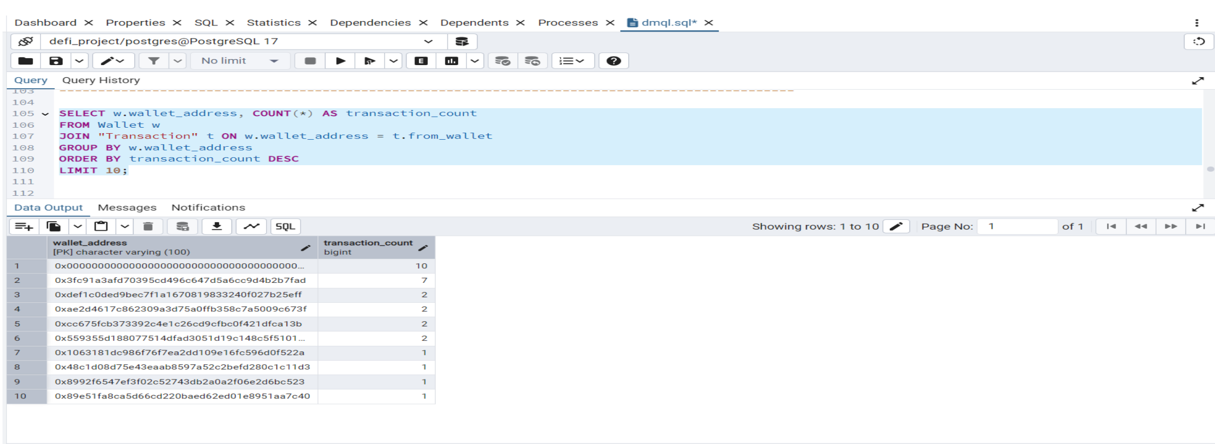


- **Transaction and Wallet Analytics**

→ Identify top 10 wallets by transaction count (most active wallets).

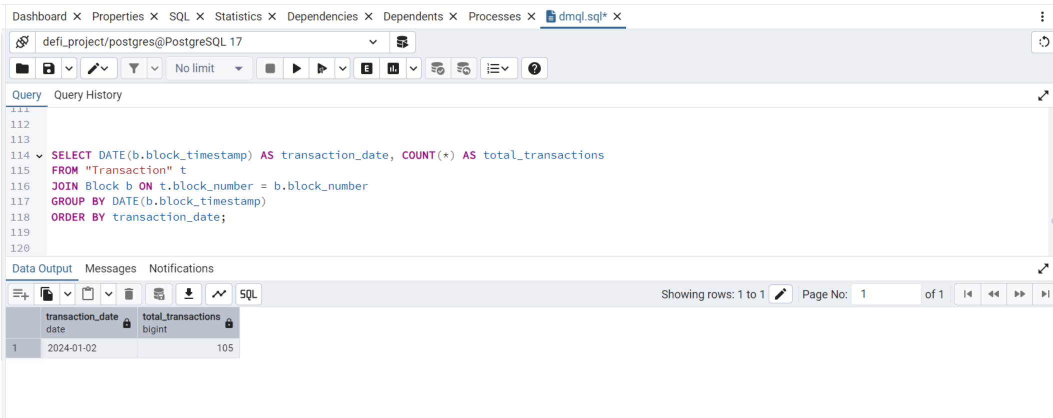


→ Analyze daily transaction trends (volume by date).



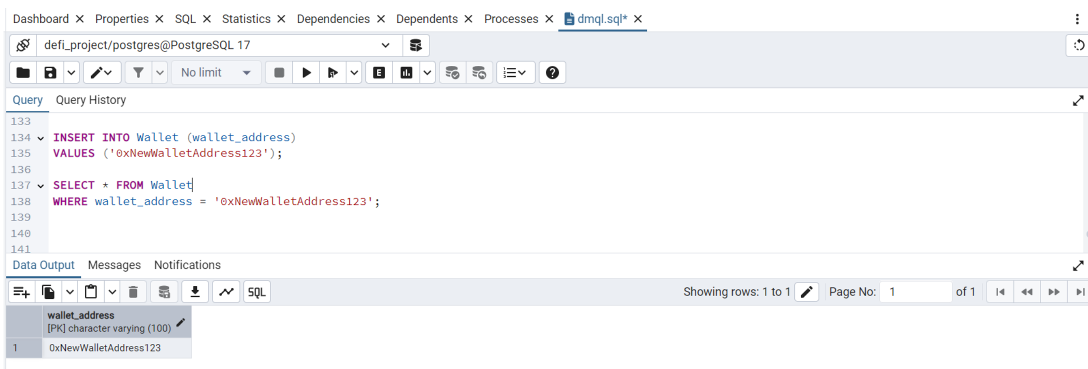
DeFi Analytics & Risk Assessment

→ Retrieve transactions with unusually large token transfers (potential risk).

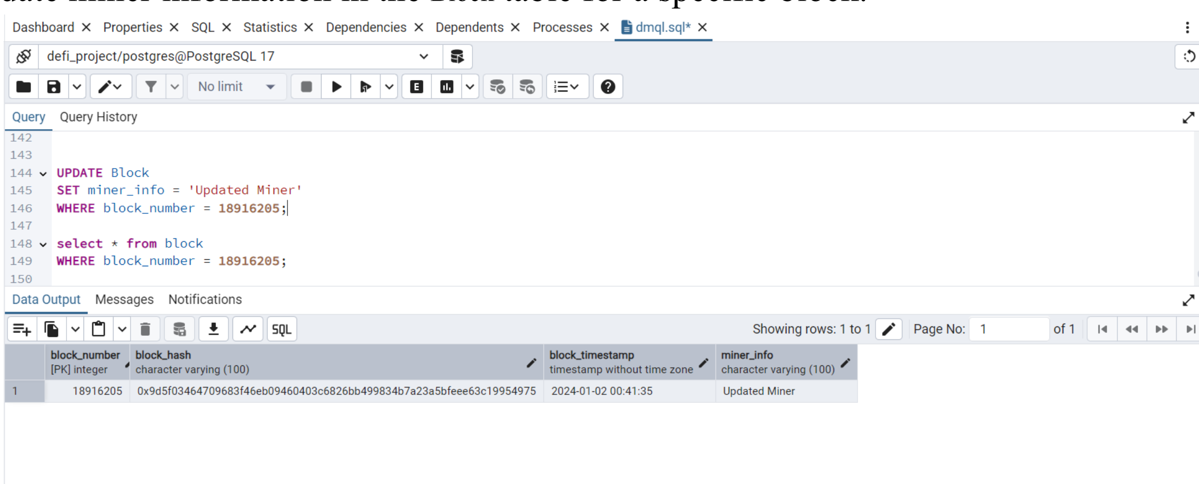


- **Insert, Update, and Delete Queries (Data Management)**

→ Insert a new wallet address into the `Wallet` table.



→ Update miner information in the `Block` table for a specific block.



DeFi Analytics & Risk Assessment

Performance Optimization

Handling and query large blockchain datasets depend on database optimization efficiency. The search strategies listed below were used to cater for the vast quantity and variety of Ethereum blockchain data. To boost query speed for faster DeFi transactions statistics, certain indexing algorithms were implemented:

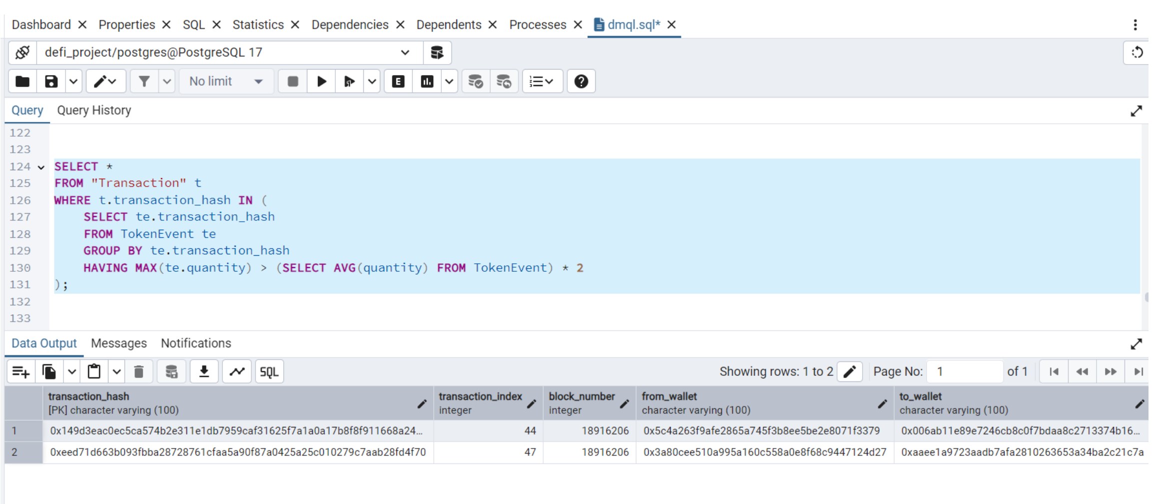
- Single-column indexing improves search time by creating indexes on frequently searched columns like transaction_hash and block_number.
- Used multi-column indexing to optimize queries with numerous conditions or JOIN actions.
- Covering indexes allow queries to retrieve data directly from the index, reducing the need for table lookups.

These methods reduced db efficiency, which is critical for actual time DeFi analytics and assessment of risks.

Risk Assessment Metrics

Queries are run to find token transfers that are abnormally large much larger than usual which could be dangerous.

To find transactions involving token transfers that were abnormally huge and far more than the typical transfer amount—possibly indicating fraudulent or high-risk activity the following query was run:



This query helps identify vulnerabilities like market manipulation, liquidity attacks, or fraudulent schemes early on by flagging transactions when the transferred token amount is more than twice the average.