TL;DR

We conduct an open-source data analysis on Ethereum staking pools. Check the following links for details:

- GitHub - Zachary-Lingle/ethsta_staking_analysis
- ethsta.com
- Full Version Doc: Open-source Data Analytics of Ethereum Staking Pools - HackMD

Background

Ethereum staking is the act of depositing 32 ETH to the deposit contract, calling the “deposit” ABI, and emitting a “DepositEvent”. A validator‘s pubkey is then valid for staking on the beacon chain.

Since beacon chain staking is complicated and requires some professional knowledge, many staking pools provide simpler staking services to ordinary ETH holders based on the beacon chain. These staking pools generate many validators by depositing ETH from the same address or addresses with the same “name tag”. It is possible to group validators into different staking pools for further analysis according to such features.

Several projects are working on analyzing Ethereum staking pools, like rated.network, beaconcha.in, ethereumpools.info, pools.invis.cloud, and showing different analyzing results. However, these projects are not open-source, resulting in the uncertainty of the data accuracy and thus confusing us with which one we should refer to.

Therefore, we decide to conduct open-source data analytics on Ethereum staking pools. The source code is uploaded to Github and the data is visualized on ethsta.com

.

How it works

[

image

656×649 32.2 KB

](https://ethresear.ch/uploads/default/original/2X/4/463e2b5af03b375dea54b5e17dc313d9f86f4795.png)

ETL

All the raw data is obtained from Etherscan APIs.

1. DepositEvent
2. txid: the transaction that calls the deposit contract and emits the event
3. eth2_validator: the validator pubkey in the calldata
4. txid: the transaction that calls the deposit contract and emits the event
5. eth2_validator: the validator pubkey in the calldata
6. Internal transaction (the contract caller is another contract, like Lido)
7. txid: the transaction id that generates the internal transaction
8. from: the address that creates the transaction
9. value: the ETH amount of the internal transaction
10. txid: the transaction id that generates the internal transaction
11. from: the address that creates the transaction
12. value: the ETH amount of the internal transaction
13. Transaction (the contract caller is an EOA, like Coinbase)
14. txid: the transaction id

15. from: the address that creates the transaction
16. value: the ETH amount of the transaction
17. txid: the transaction id
18. from: the address that creates the transaction
19. value: the ETH amount of the transaction
20. Tag
21. address: an EOA address or contract address
22. name: the “name tag” of the address on Etherscan
23. address: an EOA address or contract address
24. name: the “name tag” of the address on Etherscan

Grouping

The grouping process is written in Python, but we’d like to describe it with SQL for simplicity as follows.

SELECT name, COUNT(eth2_validator) as validator_count, SUM(value) as total_value, COLLECT_SET(eth2_validator) as eth2_address, COLLECT_SET(from) as eth1_address FROM event, internal_transaction, transaction, tag WHERE event.txid = internal_transaction.txid AND event.txid = transaction.txid AND tag.address = internal_transaction.from AND tag.address = transaction.from

Visualization

Since at most only one media could be embedded in the topic for a new account, you can visit the full version document to see the charts.

From the pie chart on ethsta.com

, we can see that Lido owns more than 1/4 validators. The top 3 staking pools, Lido, Coinbase, Kraken, own more than 1/2 validators. We can also see from the table that the top 3 staking pools are still growing fast in validator counts and deposit amounts. Besides, about 30% of validators are classified into “others”, since we are not able to obtain their address tags.

Future Work

We will continue to analyze the validators in “others”, trying to find out the entities behind them. Welcome to raise issues to point out data faults. BTW, we are also interested in the data analytics of client diversity which may help in the upcoming Ethereum “the merge”.