

2022-12-07

stakefish

Ethereum Incident Postmortem

Contents

1 Overview	3
2 Impact	3
3 What Happened	3
4 Resolution	3
5 Timeline	4
6 Action Items	4

1 Overview

On 2022-12-07, our monitoring system alerted us that our Ethereum validators were behind consensus and not validating. Our infrastructure team initiated the incident response on 2022-12-07 at 01:22 UTC, and the outage was resolved by 2022-12-07 at 13:50 UTC.

2 Impact

The outage had a significant impact on our operations. All Ethereum validators operated by stakefish were impacted and down for 6 to 12 hours.

3 What Happened

At stakefish, we have Infrastructure as Code (IaC) processes for managing and provisioning Ethereum nodes. We routinely destroy nodes when upgrading software components or updating configuration. Syncing Ethereum nodes from the genesis would take several days, so we have set up snapshots of the consensus and execution client data. The incident occurred because the snapshot of execution client data used by all of our execution clients was corrupted, and they all failed at the same block height.

We have monitoring and alerts configured correctly, so the problem was detected and responded to quickly within minutes.

Our typical process to bring up the validators using secondary and alternative Execution Layer (EL) clients also failed because we shut down our snapshotters for Nethermind, Besu, and Erigon to save cost earlier in the year. So we were in a challenging situation where we had no viable local EL chain data snapshots to bring up our validators quickly.

4 Resolution

We had to distribute working chain data to all affected EL nodes. An unaffected EL node in our infrastructure was used to seed chain data to other EL nodes. We attempted retrieving chain data from other validators, but due to complicating factors, it would take multiple days to complete the transfer.

Many thanks to the Lido team, RockLogic GmbH (Stefan Kobrc), and Hashquark (John Xie) for lending their resources to resolve the issue.

5 Timeline

Date	Log
2022-12-07 01:22 UTC	Monitoring alerts
2022-12-07 01:30 UTC	Escalated
2022-12-07 01:47 UTC	Unplanned maintenance notification
2022-12-07 07:55 UTC	Partial recovery
2022-12-07 12:45 UTC	99.9% recovery
2022-12-07 13:50 UTC	Full recovery

6 Action Items

In order to prevent similar outages from occurring in the future, we are implementing the following action items:

1. Set up dedicated hot standby consensus and execution data nodes that do not use snapshots.
2. Have readily synced alternative ELs (Erigon, Nethermind, Besu) in our infrastructure stack.
3. Resync the snapshotter from genesis and add regular data verification to our processes.
4. Re-enable snapshotter for other ELs to ensure quick recovery in the future