Polkadot Decentralization Analytics

2023 H1
9 Months Continuous Operation
Software Development

Polkadot
Treasury Proposal

Proposer

15yhxUC89ULF3WxvH2P6r4ktWRPhF7r7LtXMaGGADoyVxs2B
Valletech

Allocation
5.874 DOT (37.413 USD)

The Problem

Polkadot, and generally speaking Substrate built blockchains have tremendous potential to deliver decentralized computing, however depending on how they are deployed and operated by the community the resulting system can be more or less centralized/decentralized.

When analyzing the degree of decentralization of the Polkadot Relay chain by the following variables: Region, Computing Network and Node Operator, it turns out that the network is very centralized with most of the public validation activity concentrated and a few Operators, Regions and Computing Networks.

For a detailed account of the problem and motivation behind launching Polkawatch see our blog post: Introducing Polkawatch.

The Solution

Polkawatch is an analytic tool designed to measure effective decentralization of Polkadot's Validation process.

Polkawatch is a project supported by the Web3 Foundation and was part of the Wave 13 of the grants program. Polkawatch is also supported by both Polkadot and Kusama treasuries.

It uses the Reward event as main measure of Computing Effort, and it complements the information with networking information (IP address) of validator nodes which is then crossed with external geolocation datasources.

Polkawatch allows us to navigate our rewards, or validation operation, by geography, computing network and node operators.

This not only allows us to measure the realization of our decentralization potential, but it also allows us to measure the rewards produced by our own nomination. This is useful for us to know if we are contributing to decentralize Polkadot or not.

Polkawatch can be accessed at: polkawatch.app

For a description of Polkawatch System Architecture and how it is Operated, already presented in previous proposals, see Appendix A.

Next Actions and Milestones

The proposed improvements actions and maintenance costs of Polkawatch until 2023H1 are as follows:

ID	Category	Area	Type	Concept	Unit	Q	Rate	USD Total
1	Software	Development	One-Time	Staking Pools Support (Kusama Port)	Man/Hours	20	90	1,800
2	Software	Development	One-Time	Multi-Wallet Support	Man/Hours	40	90	3,600
3	Software	Development	One-Time	Third party application integration	Man/Hours	120	90	10,800
4	Infrastructure	Operation	Recurrent	9 Months Period	Man/Hours	135	75	10,125
5	Infrastructure	Hosting	Recurrent	Server Hosting Hetzner	Monthly Fee	9	145	1,305
6	Infrastructure	Hosting	Recurrent	Server Hosting Lab	Monthly Fee	9	187	1,683
7	Infrastructure	Network	Recurrent	Server hosting Lab	Monthly Fee	9	100	900
8	Management	Marketing	Recurrent	Polkawatch Promotion	Man/Hours	36	100	3,600
9	Management	Administration	Recurrent	Of Project Admin/Direction/Promotion	Man/Hours	36	100	3,600

1. Staking pools support has been implemented already in Kusama, therefore this task involves only data migrations, merge the implementation and ensure that the roll-out is correct.
2. Multi-wallet integration is our most demanded usability feature, and it is also requested by partners. This implementation allows users to select the wallet to use and not be limited to the official extension only. The architecture will be extensible, so other wallets can be added as they become available.
3. We will use our Client SDK to help, hands on, other projects integrate Polkawatch data. If necessary we will write pull-requests to open-source community projects. We consider partner integration strategic for spreading awareness.
4. Continued operation of Polkawatch: Operating and monitoring of the system, also includes checking that data quality is achieved (events are being traced to GeoIP data, etc). Substrate Archive node maintenance/updates, Monitoring CD/CI pipelines and DDPs production, etc.
5. Third party Data Center hosting costs.
6. Own Lab hosting costs. The monthly cost is calculated as financial amortization of dedicated equipment over an expected lifetime period of 5 years or 60 months.
7. Network / Traffic costs. (Data Center – Lab)
8. Project Promotion, including: Writing blog posts which typically include introduction of new functionality, followup on Twitter. Monitoring of engagement analytics and running campaigns targeting user awareness with Mautic. We will start using Mautic to target partner projects including wallets and staking pools.
9. Project management, technical direction and interface with partner projects.

The roles involved in the proposed activities are: Operations Role, Software Engineer, Marketing Automation, Project Management with different degrees of involvement.

Economic Summary Information

- Project duration: 9 months (2022Q4 - 2023Q2)
- Total Man/Hours Required: 387 or 1/4 FTE.
- One time costs: 16.200 USD or 1.800 USD per month targeting further service development.
- Recurrent Costs: 21.213 USD or 2.357 USD per month targeting continued operation of Polkawatch.

- Total Cost: 5.874 DOT (37.413 USD)
- DOT EMA7: 6,369
- Beneficiary Address: 1uvR3dB59H8RhBqQUhAuWX4Xo3GPN3FggeWcbZttZ3k6w2G

Submission History

The following submissions have been approved in the past:

- 2022Q3: Continued Operation and Software Development. Proposal, Report, Vote

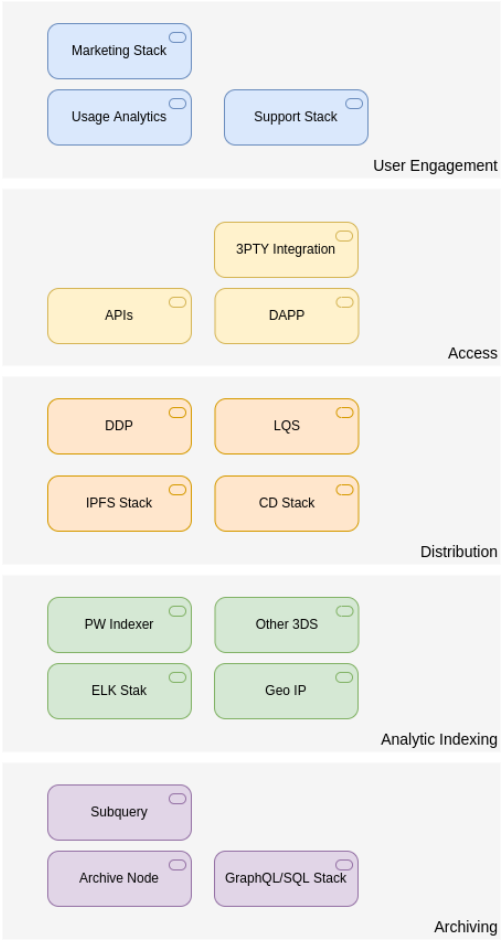
Appendix A

Polkawatch System Architecture

Operating Polkawatch involves the following components:

The **archiving** layer is responsible for indexing the “canonic” events of the blockchain, and it involves the deployment of a full-archive substrate node and a subquery stack with the corresponding GraphQL stack.

The **analytic indexing** layer uses the data from the archive layers and traces dependencies between identified events, plus merges this information with external data sources, such as GeoIP data sources or any data source relevant for analytics. This layer includes a lucene inverted index (and ELK stack)



for efficiently creating and querying analytic information.

The **distribution** layer makes the Analytic data available, from a pragmatic way. There are 2 delivery formats: deferred analytics that are distributed in IPFS data packs (DDPs) and live data which is accessible via Live Query Server (LQS). DDPs are cheaper to host and deliver with DAPPs. Depending on the nature of the analytics served DDPs may be the preferred delivery format. DDPs generation are managed from a CD/CI stack based on Gitlab. An IPFS cluster is responsible for serving DDPs.

The **access** layer is responsible for presenting the information to the user, via a DAPP, or via any other means, such as, APIs for or Component Libraries for third party integration.

The **user engagement** layer is responsible for managing a long term relationship with the user. It is an Opersource Stack of marketing automation services and user support services: plausible analytics, mautic and zammad are the main components.

Our Operational Model

We run Polkawatch strictly on top of Open Source Stacks and directly on Metal using privaz.io a sister project developed by Valletech AB that provides metal to service automation.

We believe that having the capability to manage physical infrastructure efficiently is a decentralization enabler. As such we stay out of Public clouds when operating infrastructure. Privazio automates the deployment of Apache CloudStack on metal and then all the layers of our System Architecture on top of it.

Having the capability to run a complete stack directly on metal provides economic efficiency and digital autonomy.

Typically, we split service deployment between our own Lab and Data Center. Higher layers in the stack have higher demand in terms of availability, and they run on the Data Center.

Development Process / Canary

Since the introduction of Polkawatch for Kusama in 2022Q3, Kusama is considered the “Canary” branch for Polkawatch in line with the rest of the ecosystem. New functionalities are developed first on Kusama and ported to Polkadot branch afterwards with an estimated delay of 3 to 6 months.