Polkadot Decentralization Analytics

2023 H2
6 Months Continuous Operation
Software Development

Polkadot
Treasury Proposal

Proposer

15yhxUC89ULF3WxvH2P6r4ktWRPhF7r7LtXMaGGADoyVxs2B

Valletech

Allocation
12.115 DOT (49.467 USD)

Valletech AB, Blamesv. 26, 186-47 Sweden
Authors and Contributors

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.

Polkawatch usage in our Community

Polkawatch provides analytical evidence designed to raise awareness in our community.

In the last 12 months Polkawatch has delivered 19.8K decentralization reports to an estimated 4.9K nominators and validators. There is a growing segment of nominators that analyze their nomination more than twice a month. The statistics exclude calls made via API.
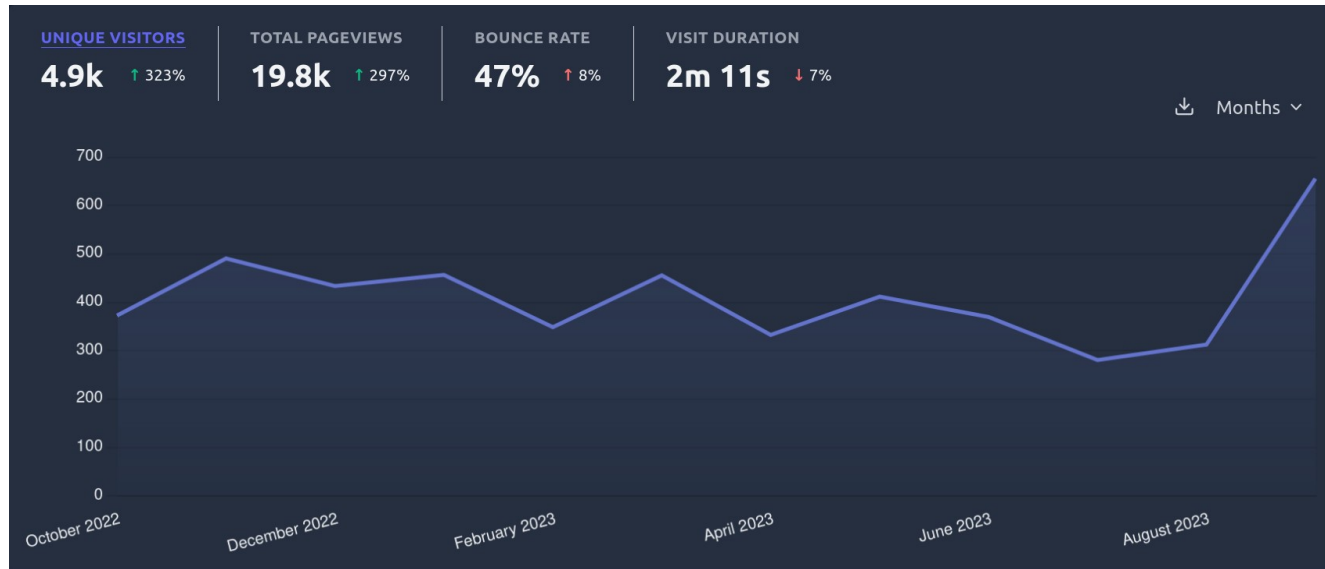
Heavy users fall in two groups: small validator (1kv) screening the network prior to migrating their validator to a new location/network, and nominators adjusting their nomination with decentralization in mind. We also receive direct queries via email from both groups which allowed us to understand their motivations. We have also noticed users visiting Polkawatch before voting for OpenGov proposals where decentralization plays a role.

Spikes in traffic depend heavily on communications in the ecosystem. Polkawatch communications are systematically amplified by Polkadot community accounts. Most remarkably the state of decentralization report 2022 was included in the 2022 roundup report of our community.

Ultimately, the key success metrics for Polkawatch are actually the decentralization metrics themselves. During last 2022, in the state of decentralization report, we could see clear trends with a decentralizing effect while seeing key challenges remaining ahead.

The traffic spike in September correspond to new traffic from related to integrating Polkawatch with the Polkadot Staking Dashboard, see next steps below.

Next Actions and Milestones

The proposed improvements actions and maintenance costs of Polkawatch until 2023H2 are as follows:

ID	Category	Area	Type	Concept	Unit	Q	Rate	USD Total
1	Software	Development	One-Time	DHT Lookups	Man/Hours	152	80	12,160
2	Software	Development	One-Time	Dashboard Collaboration	Man/Hours	214	80	17,120
3	Software	Development	One-Time	Premium Geolocation	Man/Hours	34	80	2,720
4	Infrastructure	Operation	Recurrent	6 Months Period	Man/Hours	135	75	10,125
5	Infrastructure	Hosting	Recurrent	Server Hosting Hetzner	Monthly Fee	6	145	870
6	Infrastructure	Hosting	Recurrent	Server Hosting Lab	Monthly Fee	6	187	1,122
7	Infrastructure	Network	Recurrent	Server hosting Lab	Monthly Fee	6	100	600
8	Management	Marketing	Recurrent	Polkawatch Promotion	Fixed	5	PCT	2,375
9	Management	Administration	Recurrent	Of Project Admin/Direction/Promotion	Fixed	5	PCT	2,375

1. Network Addressing information in the heartbeat event has been deprecated in Substrate, and it is not used in AURA parachains anyway. We need to implement Geolocation traceability based on lower protocol layers: libp2p Distributed Hash Table (DHT) which is the actual mechanism used to address validator nodes. We will also improve the architecture for external datasources. This activity is a dependency to rolling out decentralization analytics for all parachains.

2. Integrate Polkawatch Analytics into Polkadot Staking Dashboard as a community Plugin. After a preliminary MVP integration with the Dashboard we have seen a significant increase in reach of Polkawatch, just by integrating the "Polkawatch" community resource link. The MVP also allowed us to check technical viability of the integration by implementing the validator decentralization Modal. During Sub0 conference we asked community members for feedback about the integration. In general, having Polkawatch analytics where the staking decisions happen is seen as positive. Another important feedback received is that the focus of the collaboration should be to open the Dashboard to community Projects and never moving responsibilities from the Community to Parity. We could not agree more. The dashboard team has been working on a Plugin system and we decided to improve it and document it as part of this collaboration, using Polkawatch as a reference project, opening the way for other community projects to integrate via this mechanism.

3. Premium Geolocation. We need a more accurate Geolocation provider. We are currently experiencing traceability issues, (NOT_TRACED) results as seen on the dashboard, since the current free provider does not resolve GeoIP information in some cases. We will select a new premium provider and integrate it with Polkawatch using the improved architecture for external datasources from (1).

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. 5% of effort reserved for the activity.
9. Project management, technical direction and interface with partner projects. 5% of effort reserved for the activity.

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: 6 months (2022Q3 - 2023Q4)
- One time costs: 32.000 USD or 5.333 USD per month targeting further service development.
- Recurrent Costs: 17.467 USD or 2.911 USD per month targeting continued operation of Polkawatch, including promotion.
- Total Cost: 12.115 DOT (49.467 USD)
- DOT EMA7: 4,083 USD
- Beneficiary Address: 14MMwJ8XWZ7oVpVfZTDB94cF7fA5REeqbfffq71FGJqk8iBF

Project Sustainability

The required cost to maintain the project running is 2.9K per month, which includes redundant archive node and promotion.

The project includes strategic development as 1 time costs/Investment. Integration with the dashboard will boost usage and reduce required promotion effort going forward. New traceability method will open the door to parachain decentralization analytics.

Parachain integration will also open the door to alternative funding sources from commercial parachains.

There seems feasible to achieve sustainability of Polkawatch decentralization analytics with minimum Treasury effort.

Submission History

The following submissions have been approved in the past:
- 2023H1: Continued Operation and Software Development. Proposal, Report
- 2022Q3: Continued Operation and Software Development. Proposal, Report

Project Communication

Polkawatch communicates on regular basis using the following channels:

- All proposals are accompanied of a delivery report, and the delivery report listed in the Submission History.

- Major releases, reports and events are accompanied by blog posts.

- All previous announcements plus related industry news are also made via Twitter .

- It is also possible to subscribe to notifications via email at polkadot.polkawatch.app by clicking on bell icon.

You can proactively reach the project at info@valletech.eu

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)
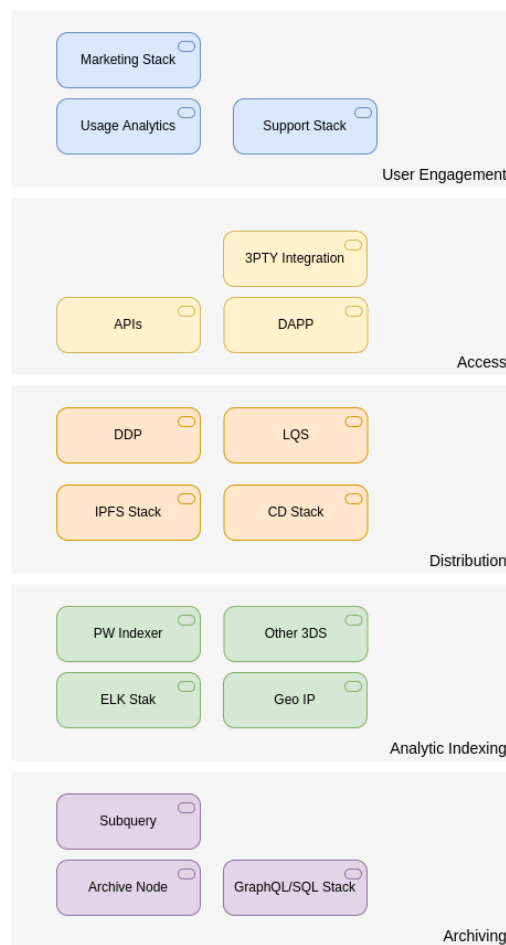
for efficiently creating and querying analytic information.

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.

Appendix B

About Valletech

Valletech (and community contributors) is the developer/sponsor behind Polkawatch, a Product Development Lab that helps customers innovate Software Products that nicely fit into leading OperSource stacks, including latest Web 3 technologies.

Historically Valletech has helped companies create new Software Products or help launched new Software and IOT ventures.

Rafael del Valle is partner/owner and has hands-on Software Product development experience for over 20 years. Rafael helped create products for TIBCO, Reuters, Credit Suisse, Vodafone and participated in the creation of 2 Software ventures with Angel, Public and VC funding totaling EUR 4M approx.

In 2020 Valletech decided to focus on Blockchain technology and selected Polkadot as the stack to focus on due its architectural superiority.

On top of Polkawatch Valletech has already started DINFRA, a decentralized infrastructure orchestrator based on Substrate, the second ecosystem project currently under development. DINFRA is also in the Web3 Grant Program.