MEGA - DECENTRALIZED OPEN SOURCE COLLABORATION FOR SOURCE CODE & LLM

May 6th, 2024
Quanyi Ma

Director, Open Source Operations, Huawei

Member of Governing Board, Web3 Infrastructure Foundation

GitHub/Twitter/Gmail - genedna
Open Source Collaboration with Git and GitHub
Git - Distributed Version Control

- **No Central Repository Dependency** - Every developer has complete copy of repository and work independently
- **Offline Work and Flexibility** - Each developer can work offline and perform any version control operations without an internet connection
- **Flexibility in Workflow** - Teams can choose to have a central repository for coordination, or share changes with a subset of collaborators, or contribute to multiple repositories.
GitHub - Centralized Open Source Collaboration

- **Central Repository Hosting Service** - GitHub acts as a central hub for store repositories and collaboration. It provides a web interface and tools built around Git, making it easier for developers to share, discover, and contribute to projects.

- **Collaboration Features** - GitHub provides fork, pull request, issues, discussion, wiki, project board and action.
How Decentralized Open Source Collaboration?

Collaboration Events with Decentralized Social Network Compatible

Git Host Service with Decentralized Protocol

Decentralized Version Control Host System with Git Compatible

Collaboration Layer

Open Source Collaboration

Data Layer

https://github.com/web3infra-foundation/mega
Rewrite the Git for Decentralized Service with Rust
Git - Addressable File System for Local

SHA-1: 8ab686eafeb1f44702738c8b0f24f2567c36da6d

<table>
<thead>
<tr>
<th>Type</th>
<th>Size</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>1</td>
<td>b</td>
</tr>
<tr>
<td>space</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>b</td>
<td>W</td>
<td>C</td>
</tr>
<tr>
<td>b</td>
<td>L</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>C</td>
</tr>
<tr>
<td>b</td>
<td>L</td>
<td>C</td>
</tr>
</tbody>
</table>

Object URI

GET https://<domain>/v1/object/<sha-1>

$ echo Hello, World! | git hash-object --stdin
8ab686eafeb1f44702738c8b0f24f2567c36da6d

https://github.com/web3infra-foundation/mega
Add a router for getting objects list

GET https://<domain>/v1/<repo>/objects
Decentralized and Discovery - libp2p & DHT

```
p2p://<peer Id>/<type>/<repo>
```

```json
{
  "origin": "12D3KooWFgpUQa9WnTztcvs5LLMjmwsMoGZcrTHdt9LKYKpM4MIK",
  "name": "mega",
  "latest": "1de1c6f",
  "forks": [
    {
      "peer": "456DFooWFgpUQa9WnTztcvs5LLMjmwsMoGZccdw3Ddf3DTH23",
      "latest": "1de1c6f",
      "timestamp": 1629827281
    },
    {
      "peer": "799DFoodjsfhuedDFEDSFesDFwefSDfwefEWFweSDFWEfwe5",
      "latest": "be04428"
      "timestamp": 1629827281
    },
  ],
  "timestamp": 1629827281
}
```
Reimplementing Decentralized Protocol with ZTM

https://github.com/web3infra-foundation/mega

https://github.com/flomesh-io/ztm

ZTM is an open source network infrastructure software for running a decentralized network. It is built upon HTTP/2 tunnels and can run on any sort of IP networks such as LANs, containerized networks and the Internet, etc.

ZTM lays the foundation for building decentralized applications by providing a set of core capabilities including:

- Network connectivity across Internet gateways and firewalls
- TLS-encrypted communication channels
- Certificate-based authentication and access control
- Service discovery and load balancing
Design Events for Decentralized Collaboration
Collaboration Events Extend Nostr

Update a Repo Status Event

```json
{
"kind": 111,
"id": "<32-bytes lowercase hex-encoded sha256 of the serialized event data>",
"peer": "<32-bytes lowercase hex-encoded public key of the event creator>",
"timestamp": "<unix timestamp in seconds>",
"tags": [
  ["p", "12D3KooWFgpUQa9WnTzcvs5LLMjwnsMoGZcrThd19LKYKpM4MiK"],
  ["n", "mega"],
  ["t", "origin"],
  ["a", "update"],
  ["u", "p2p://12D3KooWFgpUQa9WnTzcvs5LLMjwnsMoGZcrThd19LKYKpM4MiK/pack/mega.git"],
  ["c", "1de1c6f"],
],
"content": "<arbitrary string>",
"sig": "<64-byte lowercase hex of the signature of the sha256 hash of the serialized event data, which is the same as the "id" field>
}
```

Create an Issue

```json
{
"kind": 111,
"id": "<32-bytes lowercase hex-encoded sha256 of the serialized event data>",
"peer": "<32-bytes lowercase hex-encoded public key of the event creator>",
"timestamp": "<unix timestamp in seconds>",
"tags": [
  ["p", "12D3KooWFgpUQa9WnTzcvs5LLMjwnsMoGZcrThd19LKYKpM4MiK"],
  ["k", "mega"],
  ["t", "fork"],
  ["a", "issue"],
  ["u", "p2p://12D3KooWFgpUQa9WnTzcvs5LLMjwnsMoGZcrThd19LKYKpM4MiK/pack/mega.git"],
  ["c", "1de1c6f"],
  ["i", "Issue Content"],
],
"content": "<arbitrary string>",
"sig": "<64-byte lowercase hex of the signature of the sha256 hash of the serialized event data, which is the same as the "id" field>
}
```

- **p**: The peer id of the node
- **n**: The name of the repo
- **t**: The type of repo - origin or fork
- **a**: The action of event - update/request/issue
- **u**: The p2p URL of the repo
- **c**: The latest commit of the repo
Nostr -> Matrix & Mastodon

Mega as client connect different decentralized social network

https://github.com/web3infra-foundation/mega
Rewrite the Git LFS for the LLM
Rewrite the LFS Client and Server for Large File

Git LFS is not enough to solve the LLM crisis, Mega working on split the large file to pieces and deliver with decentralized network.

GET https://<domain>/v1/file/<sha-256>
GET https://<domain>/v1/file/<sha-256>/chunks
GET https://<domain>/v1/file/chunks/<sha-256>

https://github.com/web3infra-foundation/mega
Mega Architect
Mega Architect Overview
Thank you