WEB3.0 DATA INFRASTRUCTURE: CHALLENGES AND OPPORTUNITIES

Sean Yang and Max Li.

WHY WEB3.0 DATA INFRASTRUCTURE

The emergence of Web 3.0 marks a historic turning point. It ushers in a new era in the history of the web, one defined by a radical transfer of access, ownership, and governance from today's crop of Internet giants to the Internet's every user. The idea behind Web 3.0 is to create an internet that accurately interprets your input, understands what you convey, and allows complete control over what type of content you want to consume.

The rise of Web 3.0 has led to the demand for decentralized data infrastructure allowing users to store, retrieve, and maintain their data with the following key features:

- Privacy — data is encrypted at user-end, encoded (such as Reed-Solomon codes, fountain codes), sharded into a number of pieces, and then stored in geo-distributed locations. This process provides higher system reliability and security; and makes data accessible only to the data owner.
- Immutability – data records remain unchanged and storage can be verified anytime anywhere.
- Low cost — large petabytes of data can be stored at a fractional cost compared to traditional data storage.

The market size for Web 3.0 cloud storage is massive. The current cloud storage market is valued at $80 billion and is expected to grow to $376 billion by 2029. One of the best use cases for Web 3.0 storage is data archiving. Right now, 60 percent of all internet data is considered archival data; by 2024 that number will grow to 80 percent. Currently, the archival data storage market mostly uses magnetic tape to store their data — similar to a VCR. Web 3.0 storage is a more secure and affordable option compared to magnetic tape.

EXISTING SOLUTIONS

So far there exist a few projects providing Web 3.0 storage solutions, such as Filecoin, Oort, Arweave, and Storj, each have pros and cons. The projects share the same goal of ensuring stored data is always accessible while secure and immutable. Though each project has its own approach to achieving the goal, the essential philosophy of implementing security and immutability is the same. The security is ensured by providing end-to-end data encryption by the user's private key. Blockchain technologies play a crucial role in guaranteeing immutability by meticulously tracking and verifying the storage of data. These advanced systems not only maintain the integrity of information but also create a global network of storage providers by offering token incentives. This reward system encourages participation and fosters a decentralized, secure environment for data storage across the world. In addition to the storage solutions, Oort is a data cloud that provides a comprehensive Web 3.0 data solution, such as web hosting, data warehouse, and data archiving, which are all decentralized.

CHALLENGES

Overall, Web 3.0 data infrastructure is faced with four key challenges — technical, regulatory, adoption, and standardization.

TECHNICAL OBSTACLES

Scalability is vital to a useful, high-performance Web3.0 infrastructure. To compete with Web2.0 cloud services providers, Web3.0 infrastructure must have Web2.0 comparable capacity and performance to handle enormous volumes of data with guaranteed privacy. For instance, video-streaming and other data-intensive applications hosted on decentralized networks will need to store and retrieve cold as well as hot data (respectively, data that's seldom accessed and frequently accessed data stored in a database) to facilitate frictionless user experiences. Unfortunately, current decentralized networks can't host hot data at the internet scale. Clients and users across industries and domains need on-demand, reliable access to data in real-time. No internet-scale support for hot data means no video streaming or other kinds of content delivery where speed is everything.

REGULATORY OBSTACLES

The impact of regulation comes into clear relief when we think about how highly regulated industries, like insurance, operate. In most cases, the government requires insurance companies to know exactly where their policyholders' data is stored at all times and may even require insurers to store multiple copies of that data at different data centers. However, as the decentralization technology becomes mature with proofed security and reliability, the regulation will evolve accordingly over time.

ADOPTION-RELATED OBSTACLES

Massive adoption is the result of trust. On account of the current Web 3.0 data infrastructure's relatively small scale, they simply haven't been subjected to the same intensity of testing as the established centralized providers. Low levels of adoption and scarce use cases make it hard, or even impossible, for decentralized storage networks to assess their capacity and performance, or foresee risks that could snowball into crises once more users come on board. Consider how often hackers target Facebook or Google — successfully or otherwise. The fact that most of these attacks fail is important because it suggests that the targets are plugging security gaps and improving their systems, and it helps them establish trust with their users over time. In addition, Web 3.0 usability is a critical prerequisite to mainstream adoption. Most of the world's population would find onboarding alone difficult, annoying, and off-putting – to say nothing of actually using the Web3.0 products. Complicated user experience has prevented people outside of the crypto community from seeing how Web 3.0 can generate revenue, benefit the world, and simply make life easier.

STANDARDIZATION

The standardization process has to involve W3C with Tim Berners-Lee recent statement that he does not endorse Web3 as Tim/W3C has defined the Web 3.0 standard for the Semantic web. The choice of the naming Web3 is in itself an awkward choice. The current impact is that the worldwide acceptance of Web3 is and will be suffering from a status of a just de-facto standard, and not really a trusted standard. Some new initiatives are talking about Web5 and Web7, which makes this space very volatile. Cooperation between the Web3.0 community and W3C will be helpful to achieve a transparent and global standard.

OPPORTUNITIES

To achieve Web2.0-comparable user experience and system performance, we need a set of novel fundamental protocols (the decentralized protocol, such as Filecoin, a.k.a. Layer-1 solutions) and middleware protocols (a.k.a. Layer-2 solutions) that improve the current workflows in the decentralized infrastructure.

Below is a list of key developments that are necessary for Web 3.0 success:

- Web2.0-compatible APIs, like AWS S3 API and SQL API, tens of millions of Web2.0 developers need such well-

known API formats for their seamless transition to Web3.0 with no steep learning curve.

- Proof of resource usage: The layer-1 storage solution usually provides proof of storage to achieve data immutability. However, layer-2 services need to consume lots of other resources for fast data access and processing, such as CPU and bandwidth. The proof of these resource usage is necessary for resource allocations (such as load balancing) in a decentralized network.
- CDN (Content Delivery Network) to cache frequent-visit contents: CDNs are crucial in accelerating content to build low-latency streaming services or database services, which is not mature in the currently existing Web 3.0 projects.
- File indexing and retrieval: data should be fast accessed and retrieved from geo-distributed storage nodes.

Web 3.0 data infrastructure still has a long way to go before it can support the same kinds of data-intensive services, and boast the same reliability and trustworthiness as the centralized cloud old guard. Today, these nascent networks face a host of technical, regulatory, adoption, and standardization obstacles. While the latter three obstacles will take time to overcome, the Web 3.0 community has been trying to break down the technical barriers that have limited the potential of decentralized data solutions. We will see the thriving of the ecosystem running on Web 3.0 infrastructure in the near future.

Biographies

Sean (Sichao) Yang (ysc@ntlabs.io) is the Co-Founder and CTO of Oortech. He received his Ph.D. in the Department of Electrical and Computer Engineering and M.S. in Department of Mathematics both from University of Illinois at Urbana-Champaign. His Ph.D. research is on game theory with its application on resource allocation and optimization in distributed networks. He was a senior staff engineer in Qualcomm Inc, he was one of the key contributors to the design and development of the 4th and 5th generation mobile communication. He also served as the technical leader in Qualcomm's research on vehicular network and autonomous driving and holds many invention patents. He also served an adjunct professor in Columbia University and instructed Ph.D. students in Columbia University and McGill University. He has many journal and conference papers and served as a reviewer for *Mathematics of Operations Research*, *IEEE Transactions on Networking*, *Games and Economics Behavior,* and other magazines. He also gave invited talks and tutorial sessions on IEEE conferences including IEEE GLOBECOM, IEEE Summit on Communications Futures, etc.

Max (Chong) Li is a faculty member in the department of electrical engineering at Columbia University (in the City of New York), the founder & CEO of Oortech. He had been working with Qualcomm. Research on 4G LTE and 5G systems design. He is an IEEE Senior member and the president of Science & Technology Economic Committee at China-US Chamber of Commerce. He is a holder of 200+ International/US patents. He has been actively publishing academic papers on top-ranking journals, including *Proceedings of the IEEE*, *IEEE Transactions on Information Theory*, *IEEE Communications Magazine*, *Automatica*, etc. He has served as reviewer, committee, and co-chair for most prestigious journals and conferences in communications and control societies. He is a grant review committee member of Natural Sciences and Engineering Research Council of Canada. He is the author of the book, "Reinforcement Learning for Cyber-physical Systems" (Taylor & Francis CRC press). He has broad research interests, including information theory, machine learning, distributed database, and computing systems (e.g., blockchain), networked control and communication, and PYH/MAC systems design for advance telecommunication technologies (5G and beyond).