

# **BitArk**

**Universal distributed storage layer2 network protocol**

## **White Paper**

**1. Our vision**

**2. Challenge**

**3. Solution**

**4. BirArk Protocol**

**4.1 BitArk cross platform cooperation protocol**

**4.2 development incentive agreement**

**4.3 Information entropy reduction**

**5. Technology**

**6. About us**

# 1. Our Vision

In the past couples of years, we have witnessed the amazing progression of the Blockchain technology, and the explosion of the projects built on the top of it. While the Blockchain technology is well known, numerous business verticals such as healthcare, finance, manufacturing, transportation, food, insurance, education, and retail are utilizing blockchain technology. With the blockchain technology disrupting almost every sector, the data storage industry is certainly not being left behind.

Data is considered one of the most useful assets in the world. In today's data-intensive world, the value of data is undoubted critical for any organization. However, the central problem which arises is what needs to be done with all the data that has been collected. After all, it is impossible to analyze data if you do not have a place to store it securely and cost effectively. The industry has been seeking for the right solution addressing the challenges of storing data for decades. With the development of material science, software engineering, and networking technology, we see cheaper disk, bigger capacity, and more powerful management system. Nevertheless, the critical issue with the centralized solution still exists. Until Blockchain technology started disturbing the storage.

The development of distributed storage technology provides us with a more secure way of data storage. IPFS is one example standing out and getting widely adopted. A group of the applications and projects have been incubated with the foundation of IPFS. It is approved that quite a lot of them can resolve the typical challenges of the centralized storage solution. But IPFs technology is still in its infancy. There are observed defects and limitations pending on the resolutions. Meanwhile, other emerging protocols and concepts with Blockchain technology are introduced to the data storage section. They all are born with pros and cons. While the diversity and iteration of the distributed storage solution allows the developer to select the most fit, it also creates a common challenge existing in almost every well-developed areas, that is how to find the best fit.

No doubt that more and more data storage solutions will be created along with the development of Blockchain technology. To facilitate all kinds of data generated by the different applications the best home, we created BitArk protocol. It is committed to build a universal interface on the top of all kinds of distributed storage platform. It empowers the developers to quickly deploy their applications without concerning at

what storage solution to adopt. BitArk is going to do all the work. The storage application development platform based on BitArk protocol will incubate new storage application products and form a rich distributed storage application ecosystem. Ultimately, BitArk will facilitate developers to create numerous data storage applications benefiting hundreds of millions of individual users and millions of institutional users around the world. In the long run, this product ecology will provide people with truly safe, convenient and permanent information storage, and make great contributions to the inheritance of human civilization.

BitArk lab plays the critical role of driving the project forward. Establishing a protocol interacting with the different distributing storage systems; creating a cost effective and convenient platform compatible with the mainstream programming languages; and working with worldwide developers to create a high performance, more economical and more secure ecology of data storage and sharing. As the basis of Web3.0, our vision is to SEND THE DATA TO THE RIGHT HOME!

## 2. Challenge

Since human society entered the Internet era, we have also entered the era of information explosion. In this era, the speed of information growth is exponential. Traditional hard disk and disk array storage methods are gradually replaced by the latest cloud storage technology. However, the development of this centralized storage technology is still facing great challenges even today. These problems are closely related to the mode and operation management of centralized storage.

### 2.1 The security of data storage is low

At present, our data are stored on personal computers, backup storage devices or servers of cloud storage service providers. This centralized storage method brings many risks, such as data loss and leakage caused by equipment damage and loss. Although cloud service providers promise to provide guaranteed storage services, because centralized cloud servers are concentrated in one or several physical locations, once power failure or natural disasters occur, a large number of related businesses will be paralyzed. The failure of centralized cloud server also has the risk of complete loss of user data. In addition, there is a risk of privacy leakage in centralized storage. The administrator of cloud service provider can directly view and delete the files uploaded by users from the server platform, and the user's privacy is easy to leak. So the security of centralized data storage is low.

### 2.2 Low efficiency of network transmission

Centralized cloud server room is usually located in remote areas, far away from the actual user, so its data transmission needs to go through multiple nodes, and the data transmission speed is slow. Under the mode of centralized server or service cluster

based on HTTP protocol, when there are hot resources in the network or a large number of terminals accessing the same file resource at the same time, there will be network congestion, resulting in users unable to access files effectively. Centralized cloud server will also bring high server operating costs and broadband costs. When network access is not in peak period, a lot of network bandwidth is idle, that is, a lot of broadband is wasted. Overall, the network transmission efficiency of centralized storage is relatively low.

### 2.3 High repetition rate of data storage and waste of storage space

In the centralized storage mode, the network transmission efficiency is low. For the sake of data security and convenience, each user often downloads a large number of data files to his own computer, resulting in repeated storage of the same data file. The same is true for cloud service providers. Each major Internet platform has spent a huge amount of money to establish a lot of independent data storage centers. However, due to the non interoperability of data networks between them, they have to store a lot of the same file content. For example, many video platforms are playing the same movie or TV series, which means that the same video file will exist in the storage center of different video platforms. Hundreds of thousands of high-definition video files on the network need to take up huge storage space, and each video platform has to store these files repeatedly on its own server, resulting in a huge waste. Even for the same cloud service provider, in order to improve the access speed of users, the same data file is stored in multiple areas on the same platform, and the data is repeatedly stored near a large number of users. Because each storage unit is independent of each other, a large amount of redundant data is formed in the whole network, which brings huge storage cost and waste.

### 2.4 Data access is easy to fail

The current Internet access is based on HTTP protocol, and each data file has its own address on the Internet. To find a data in the network, you need to access the location of the file. This means that once the file location has been changed, or the file itself has moved, we cannot find the file again, even if it still exists on the network. We often encounter this situation, that is, the previously accessed file links are no longer valid and the referenced hyperlink cannot be opened. The most common cause is a failure in the server where the file was deleted or stored. So this centralized data storage method is fragile and inefficient.

### 2.5 Complicated storage platform and high cost of data exchange

The existing cloud service provider storage platform has complex usage. Moreover, the data between the service providers cannot be exchanged, and the cost of data migration is huge. If a user wants to migrate data, he needs to download the data from one cloud server to his or her local storage, and then upload it to another storage platform. It is not easy to quickly migrate data from one storage platform to another. For users who store a large amount of data, the bandwidth cost and time cost are very high. In this

process, there is also a risk that data may be lost and damaged after multiple transmission.

In recent years, people have developed distributed storage technology to solve the above problems. This is a good solution, data is not stored on a single cloud service provider's server, but encrypted and distributed through a distributed network. This means that no one but the data owner has access to the data, which makes distributed methods the safest cloud storage solution at present. Distributed storage networks will use millions of servers and storage devices instead of hundreds or thousands. This means that storage availability is usually much higher and the cost of storing data is much lower. Through distributed cloud storage network, all parties who need additional storage capacity can store data more safely and transparently to achieve lower price.

In the market, there are multiple distributed storage technologies such as IPFS/Filecoin, Swarm, MinIO, ARWEAVE, etc. They want to solve the above-mentioned problems faced by data storage, but they only solve one or two problems. Therefore, when an application uses distributed storage, it needs to use two or more distributed storage to meet the application scenario requirements.

### **Challenges encountered by using different distributed storage**

With all the challenges and defects described above, it takes users tremendous time learning the different solutions. It is almost impossible to understand every single storage offering in order to make the right decision. As a result, the developers have to take on whatever limitations the option selected whereas there is a better solution out there. Obviously, an application development protocol framework bridging the data and the most suitable storage system would do a great benefit for the development ecosystem.

In addition, how to allow different distributed storage to interact with the blockchain is also an important issue that needs to be solved by the web3.0 infrastructure.

## **3. Solution**

BitArk protocol framework is committed to building a general distributed storage application development platform, which empowers developers to quickly deploy any application on the selected distributed storage system and jointly create a rich data storage application ecology. With BitArk protocol, developers are able to build the decentralized applications for permanent secure storage of information. such as personal information persistence products or NFT storage applications. Ultimately, our goal is to enable the global users to create a rich data storage ecosystem through permanent storage, cross platform collaboration, interactive incentive and community governance.

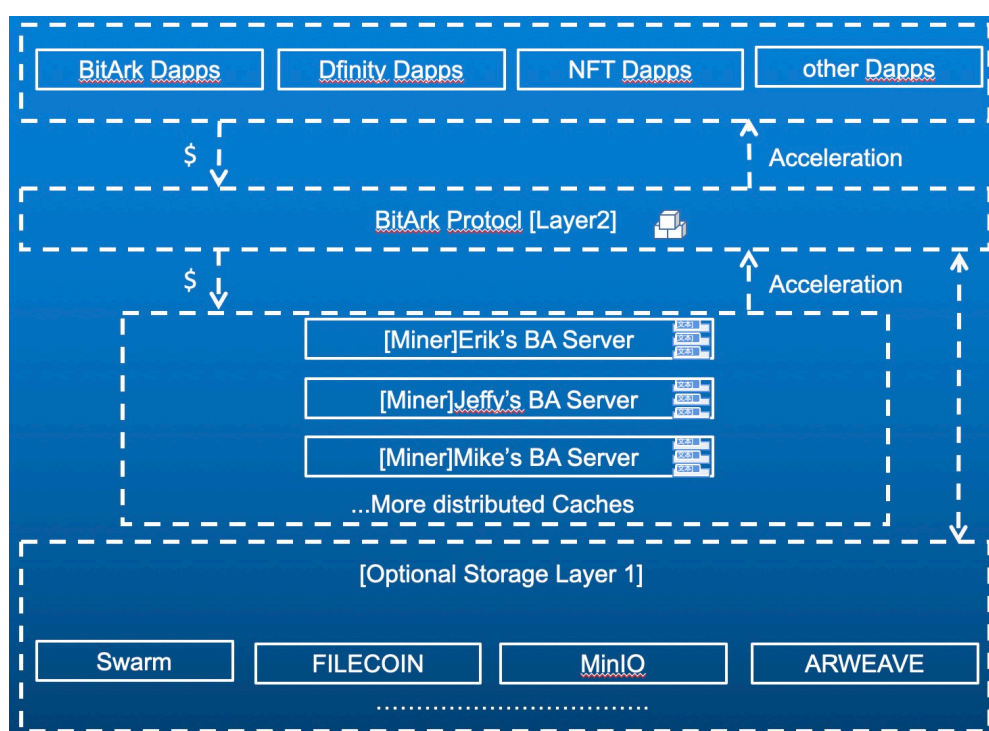


Figure 1. BitArk solution

In BitArk protocol framework and development environment, developers can get the following benefits:

- **Improved development efficiency**

Developers don't need to know the technical rules and details of every storage project from scratch, which often takes a long time. They only need to select the corresponding components to develop in BitArk development environment according to their own needs. BitArk protocol provides API interface to adapt to various storage platforms. Developers only need to focus on their own business logic, while the distributed storage service is provided by BitArk protocol with blockchain technology.

- **Reduce capital expenditures and optimizes costs**

We provide access to multiple storage platforms for developers. Developers can easily use the resources of various storage platforms, without building their own storage nodes, just a few lines of code access, they can enjoy storage services. BitArk eliminates

the cost of configuring and managing storage component, which makes it a cost-effective choice for storing data. The economic algorithm BitArk integrates is capable of comparing the economic models every distributed storage system operated with and provides the lower cost choices. It acts as a financial consultancy constantly offers you the most economic options for your data.

- **Convenient customized development**

Developers can customize data storage products and solutions according to customer needs, use blockchain technology to realize completely decentralized distributed storage. All data is managed by developers and users. It supports Android and IOS systems, supports real-time data query, supports tens of millions of C-end users, and ensures the security of user data storage.

- **More friendly to developers**

Under the framework of BitArk protocol, the distributed storage application ecology of IPFs will be gradually enriched. At present, we are creating various business components. The components of these business modules will be developed and built by different organizations and individuals. At the same time, the components will be publicly placed on the network for all developers to use. Each component only needs to be built once, which greatly reduces the redundant development work. This will be better for developers and help developers pay more attention to the business itself.

- **Increases scale and performance of the workloads**

BitArk lets you scale globally and accommodate spikes in resource demand. That way, developers can deliver the value of your application from anywhere in the world faster and enhance application performance. More importantly, the storage infrastructure BitArk interface with keeps expanding. Once a better solution is identified, the developer will have option for better scalability.

- **Easier to interact with the blockchain**

The BitArk protocol integrates a highly scalable blockchain infrastructure with Polkadot as the blockchain layer, allowing applications to use the blockchain infrastructure more conveniently.

## **BitArk lab**

**BitArk lab** is responsible for the development of the BitArk protocol framework. We hope that all people can participate in the construction of distributed storage ecology, transform all kinds of human entity information into permanent digital information, and use AI and other technologies to carry out intelligent identification and classification, break through the rules barriers between various storage projects, and promote the arrival of the digital and eternal era of human beings.



## 4. BitArk protocol

### 4.1 BitArk cross platform cooperation protocol

BAP (BitArk Protocol) provides storage technology and storage space. It establishes a common storage policy across solutions. BAP supports the self selection and intelligent offering among the different distributed storage solutions for the user. It simplifies the data exchange between different storage platforms. In the transmission process of different platforms, the hash value is checked multiple times to ensure the security and integrity of data. .

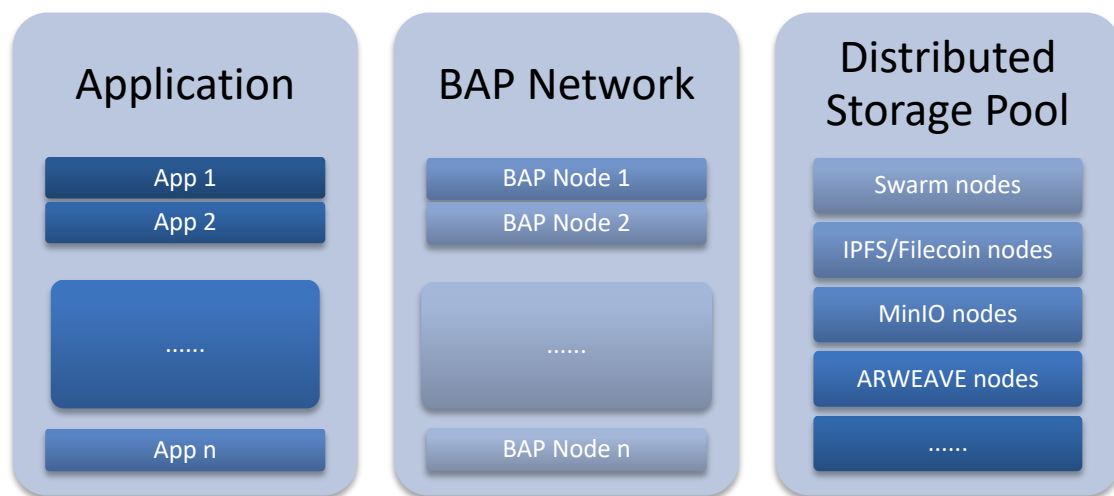
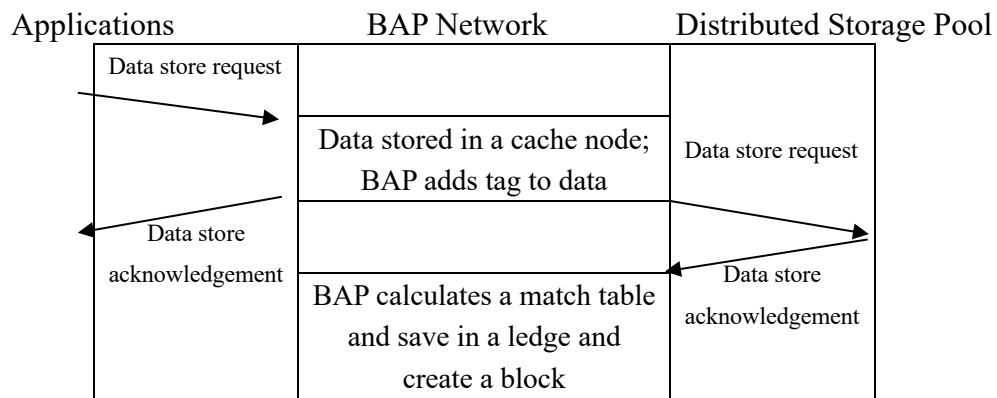


Figure 2. BAP network architecture

Cross platform cooperation protocol adds flag of the identification and classification to the data transitioned to form modular data. Each data module can be classified accurately following the identity and classification policies predefined.. Data generated by the application will be cached first in a node of BitArk network. Before transmitting the data to the selected distributed storage solution, BAP creates a match table with the identifier tag of the data and the target location the data will be stored. The table will be saved in a ledger in a block on the BAP nodes in BAP network. Once it is completed, BAP sends the data to the target storage. Each map is connected with each other. All the maps of BAP nodes will be added up to become a distributed hash table. When anyone requests data from this network, it will find the nodes on which node of the selected distributed storage solution based on the hash value of the ID of the data itself, and then establish a connection to download the data requested. Figure 3 explains the work flow.

Data store work flow:



Data download work flow:

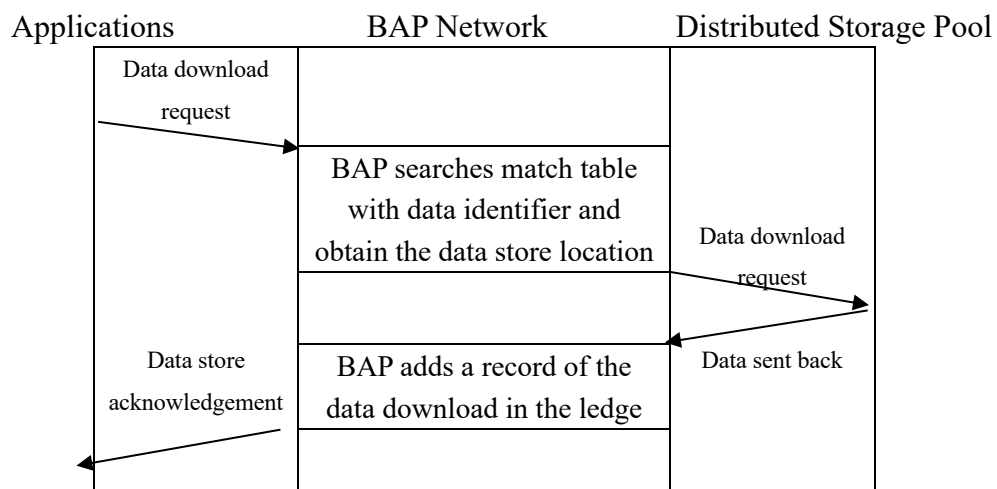


Figure 3. Data store and download work flow

We use a consistency hash method to locate and exchange data hash on different storage platforms. Consistency hash is an improvement of hash method. On the basis of determining the storage location by hash method in data storage, a layer of hash is added, that is, before data storage, the storage node is hashed in advance. This improvement can solve the stability problem of hash method. When a node is joined or exited, only the next node adjacent to the hash ring is affected. We hash the data storage location in different storage platforms, which makes it convenient for participants to find and exchange data between different storage platforms and cooperate across platforms.

## 4.2 development incentive agreement

The incentive capacity of BitArk protocol comes from token economy. The pass is a new value carrier based on blockchain ledger technology, which has two basic properties: one is Tong, which can be used, transferable and convertible; secondly, certificate, can identify, tamper proof. The registration of this value does not need to rely on any third-party centralized organization. Combined with the blockchain intelligent contract, the general certificate is a very low value carrier with

programmable transaction cost. In order to make the storage application ecology more abundant and the stored information play a greater value, the uploaded data becomes the live data rather than the dead data. BA platform designs the primary pass BAC general certificate model by virtue of the value carrier of the general certification. There are three types of role in BitArk ecosystem:

1. Physical node provider: provide hardware node with the defined specification to build BAP network for cache and computing data;
2. Developer: develop applications with BAP for data transition, store, and download;
3. End user: generating data with the applications built by developers.

. The 3 roles are encouraged to contribute more nodes, more applications, and more data interactions respectively. The more information a person uploads and be consumed by others, the more rewards the developers and users of the application will be granted. Through this interaction, a wider and deeper human digital space will be built.

All awards will be automatically distributed through blockchain intelligent contract. In the case of no need for manual intervention, the rewards will be granted for the activities including providing hardware nodes to build BAP network, upload data and access interaction between the different roles in various scenarios with extremely low trust and transaction cost.

In this development and interactive incentive scenario, all participants are miners and data providers, and gain benefits with the activities engaged. The more data is stored, the higher the access is, the higher the revenue. We use the method of proof of access (POA) to motivate. The more blocks the nodes store, the more likely it is to obtain the package rights of the blocks. The more information uploaded by participants is accessed, the easier it is to get rewards. Through proof of access (POA), the miners are encouraged to store historical blocks forever. When miners get new block rewards, they will also be rewarded for the stored random old blocks.

### **4.3 Information entropy reduction**

Generally, entropy is defined as a measure of randomness or disorder of a system. The word entropy has important physical implications as the amount of "disorder" of a system. A deterministic message can reduce the entropy of the system. In the BA protocol, participants are required to upload deterministic information, classify and label each piece of information, so as to enhance the certainty of information. The deterministic information will make the information entropy of the whole Bit Ark system decrease continuously, and will not cause confusion and disorder because of the increase of information. This is also the guarantee that Bit Ark will become more and more energetic. After more information is added in the future, the protocol will automatically classify and link information through artificial intelligence algorithm.

This will make Bit Ark a living human digital immortal spaceship rather than a dead database.

Here, we use the formula to describe the entropy of a system  $X$ . If the possible value of a random variable  $X$  is  $X = \{x_1, x_2, \dots, x_k\}$ , and its probability distribution is  $P(X = x_i) = p_i (i = 1, 2, \dots, n)$ , then the (Shannon) entropy of a variable  $X$  is defined as

$$H(X) = - \sum_x P(x) \log_2[P(x)]$$

The joint entropy of variables  $x_1, x_2, \dots, x_n$  is then defined by

$$H(X_1, \dots, X_n) = - \sum_{x_1} \dots \sum_{x_n} P(x_1, \dots, x_n) \log_2[P(x_1, \dots, x_n)]$$

When the probability  $P(x)$  of a random variable  $x$  is bigger, the entropy of the system is smaller. Therefore, reducing the uncertainty of the random variable  $x$  will reduce the entropy of the system, that is, entropy reduction.

Therefore, in BitArk protocol, more rewards will be given to the information providers with more certainty in the system. In other words, participants are encouraged to upload more accurate information. At the same time, participants are encouraged to contribute to the classification and collation of chaotic information in the system. Thus, the entropy of the whole Bit Ark system is reduced and its vitality is improved. We have a set of calculation methods to evaluate and reward the information certainty in the system.

## Features

- ☐ True permanent storage
- ☐ Privacy security
- ☐ Unlimited expansion
- ☐ Compatible with multiple storage platforms

## 5. Technology

Figure 4 depicts Bit Ark Protocol's technical architecture to achieve the above mentioned trusted capabilities

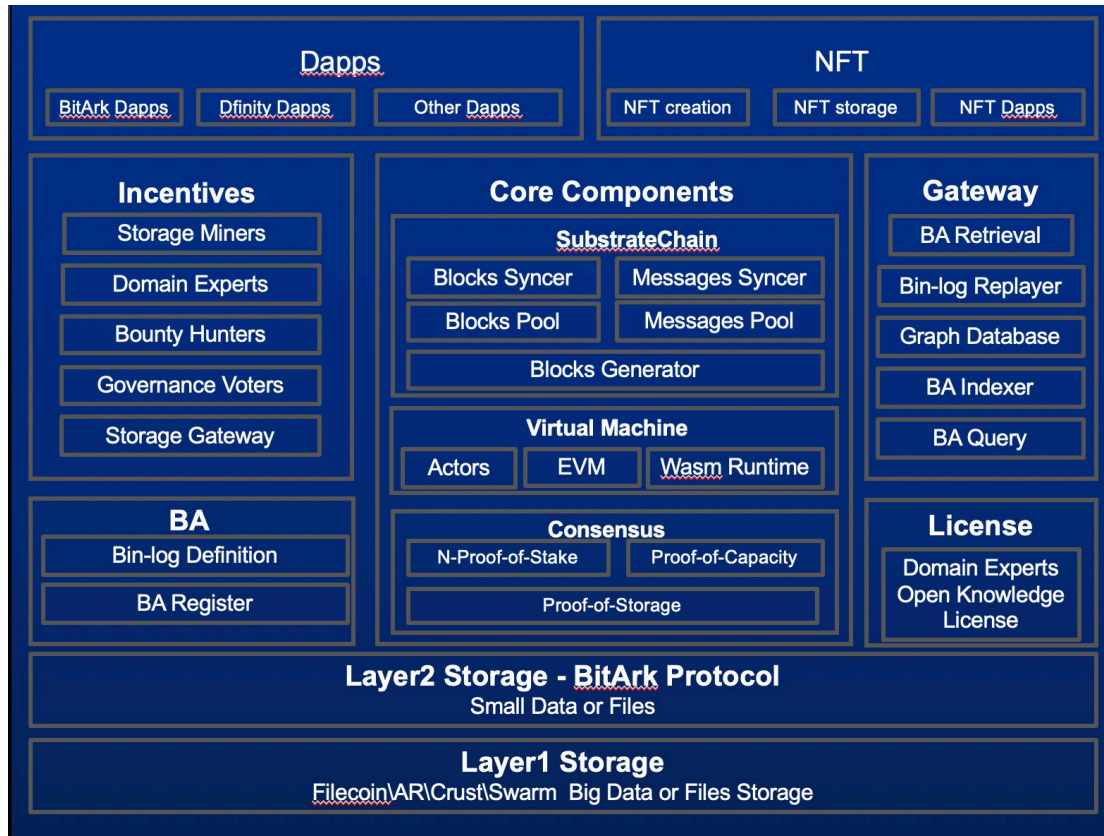


Figure 4. BitArk Protocol's technical architecture

## 5.1 Storage Layer

Since the application data supported by the BitArk protocol is mostly small files and hot data, the small files and hot data generated in the middle will be stored on the BitArk layer 2 network customized for hot data storage. Anyone can gather small files in the second-tier network into a large snapshot file within a specified time, and transfer them to a first-tier network such as Filecoin/Arweave for eternal preservation, and get incentives. Therefore, at the bottom of the technical architecture is a first-tier storage network such as Filecoin/Arweave, and the next top is a second-tier storage network customized by BitArk Protocol.

## 5.2 Core Components

Structured on top of the underlying storage layer, BitArk Protocol's core components are made up of Consensus Mechanism, Virtual Machine and Ledger On-Chain. BitArk is a layered design of the underlying consensus layer and the upper application layer, and the implementation steps are also implemented successively. Since BitArk uses Substrate to build the application chain, after connecting to Polkadot's RelayChain through Polkadot's slot, Polkadot's mainnet will be able to provide consensus security for the BitArk application chain, and its native NPoS mechanism can give block node

rewards. The underlying conjugate PoC and PoS consensus are the core consensus rules for BitArk token distribution. This creates an opportunity for a large number of low-level mining machines which are unable to participate in mining FIL in Filecoin Layer 1 Network, as they can now turn to mining in BitArk Protocol, maximizing the miners' mining profit.

### 5.3 BitArk Protocol Network node

BitArk Protocol Network node is a hardware server built with the defined specification. The node is responsible for

1. Providing big cache for the data sent by applications with high speed and low latency. It resolves the challenge of the low transaction speed of current distributed storage solutions for upper layer applications;
2. Calculating tag for data and generating match table with hash algorithm;
3. Sending the data to the right storage in distributed storage pool for reservation.
4. Handling the interactions between applications and storage with high speed and low latency.

Node providers participating in establishing BitArk Network are rewarded according to the computing and cache resource contributed.

### 5.4 Applications

The protocol supports the use of a two-tier storage infrastructure for multiple applications. Including traditional web2.0 applications, blockchain-based Dapps, web3.0 infrastructure-based applications such as DFINITY Dapps, all NFT data storage, etc.

## 6. About us

We are a team composed of entrepreneurs and scientists with 10 + years of experience in blockchain and Internet industry, experts proficient in blockchain technology and storage technology application, and data operation experts. Members include doctors and experts from world-famous universities, partners of top financial institutions, elites in e-commerce field, etc., serving more than one million users in total.

Jeff Lee. He graduated from the University of Vienna with a doctorate in quantum physics. He has provided high-tech R&D decisions for the governments and scientific research departments of the United Nations, China, Singapore and Hong Kong. He has worked in the big data and blockchain industry for more than 10 years, has rich experience in Internet product design and operation, has established Internet media, and has provided services for more than 1 million people.

Eric Cho, He is the Funder of Blackrock Block Chain Lab. He is also a SEED Fellow

of Harvard University. He graduated from the Royal Swedish Institute of Technology in Information and Communication Systems Security. He was a researcher at the Trust and Security Laboratory of Swedish Institute of Computer Science, and a senior consultant at Axiomatics, a well-known Swedish information security company. He has a technical background in the blockchain and its applications.

Louise Lau, He is used to be the technical manager of the Information Technology Department of a big commercial bank. The management project involves password service platform, financial IC card making platform, unified security authentication platform, wealth management business system and intermediate business system. Earlier, he worked in Tata Information Technology Co., Ltd. and Langxin Technology Co., Ltd., engaged in Oracle ERP (EBS) development work, JAVA development work based on open source technology such as SSH. He has a profound technical background.

Richard Pak, He received a master's degree in network communications from the Royal Institute of Technology in Sweden. Served as a network engineer at the global headquarters of Ericsson, a Fortune 500 company in Sweden, a server leader at a Fortune 500 data storage company, and a senior solution program consultant at DELL EMC.

Anders Johnsson, Anders is a Senior Advisor at 24/7 BlockLabs. He has a demonstrated history of successfully working with projects all over the world and 20 years' international experience working in the tech industry in Sweden, US, UK, Australia and more than 10 years in China. Anders has actively been involved in blockchain projects since 2014. He has advised and worked with numerous ICOs, IEOs, STOs, IDOs and Defi start-ups, through to large corporations on blockchain integration strategies, contributing extensive industry experience. Over the past 5 years he has been part of projects with an exit value of over US\$737 million in capital. Anders was a co-founder, majority shareholder and director, of Diakrit International (diakrit.com), a VR and AI company. He personally recruited the company's first 30 employees in Beijing China, which grew to over 400 employees and offices in eight countries. Today, Diakrit is one of the world leaders in digital visualization tools, VR and AI for the real estate sector. Diakrit was acquired by News Corp Australia in 2016.