

# Efficient Query Model of Blockchain System in The Context of Dig Data

Zexin Fang

Zibo Ex-servicemen Bureau ex-servicemen Service Center, Shandong Zibo 255000

1753135179@qq.com

**Abstract.** The rapid development of big data has become a hot spot of concern in science and technology circles, business circles and even governments around the world. International authoritative publications have been published to discuss the opportunities and challenges brought by big data. Big data is considered as the oil of science and technology. The application amount and ability of big data are becoming a standard to measure comprehensive national strength. Data is an important field of science and technology for vigorous development. Big data is valued and developed by various key fields and authoritative industries. The importance of big data for the development of various fields in the future is self-evident. This paper studies the relevant theories and practices of efficient query model of block chain system under the background of big data, mentions relevant concepts and theories of efficient query model of block chain system, and conducts scientific and systematic demonstration and analysis of efficient query model of block chain system. After several tests, the test results show that the efficient query model of blockchain system under the background of big data is significantly better in terms of the query rate of blockchain.

**Keywords:** Big Data, Block Chain, Efficient Query, Model Research.

## 1. Introduction

Big data is a multidisciplinary technology, including physical theory, computer technology and social and economic science knowledge. The social physical form reflects itself in the form of information system with the help of computer technology and various network means. The social economy shows its own image in the form of data through various information systems and Internet technology. From the perspective of computer technology, big data acts as a powerful assistant of human beings and a powerful power to promote social economy. The era of big data is a new revolutionary era, because the century of advanced information technology has been adopted in this era, such as Internet technology, information technology, cloud computing and so on. It is because of the use of these high-tech content of science and technology means, the society can never have the speed and way to stride forward, this is because of the use of these advanced means of the new era, the social economy can advance on a large scale. Blockchain technology is a form of big data technology. The development of blockchain technology cannot be separated from the improvement of big data technology. In the new century, more efforts must be made to develop block chain technology, and the query efficiency and stability of block chain system must be improved. As for the application of efficient query model of blockchain system, it is necessary to strengthen the research and development and investment of efficient query model of blockchain system in the context of big data, and strive to break through various obstacles of social economy in the context of big data.

As for the research on big data, many scholars at home and abroad have carried out research on it. In foreign studies, a scholar proposed watershed network infrastructure with Internet of Things (IoT) as the core, bringing watershed data science into the era of big data and greatly improving the efficiency of data collection and sharing [1]. BhuvaneshwariPV proposed various technologies for LDPC code application in big data storage and pointed out the research gaps in LDPC code application in big data storage. In order to effectively protect and store such big data, the application of LDPC code in cloud storage and network or distributed storage system has become a research hotspot in recent years [2]. HosseiniMM proposed a new algorithm to optimize decision variables for

the outcome variables of complex problems such as big data. The proposed algorithm is On account of the concept of Markov blankets in Bayesian networks to alleviate the computational challenges associated with optimization tasks in complex data sets [3]. However, the research on efficient query model of block chain system under the background of big data is still in the imperfect stage, and there is still a certain gap compared with foreign systems.

If China wants to continue to improve its efficient query model of blockchain system in the context of big data, it must start from the following points: First, optimize the research and development of big data; Secondly, the algorithm data structure is optimized. Finally, improve the competitiveness of the research and development of efficient query models for Chinese blockchain systems.

## **2. Exploration of Efficient Query Model of Blockchain System in The Context of Big Data**

### **2.1 Big data algorithm**

Big data technology is the product of a new technological background, where people program, apply, collect, process and use large amounts of data to get the added value of data. Big data has four characteristics: first, capacity; Second, diversity; Third, speed; Fourth, value [4-5]. For big data, a large number of international and domestic giants or groups have put forward professional terms and mentioned the multidimensional nature of big data growth [6-7].

Big data analysis excavates the value of big data from the perspective of hidden value and unknown value [8-9]. First, leverage big data from a hidden value perspective. Big data is applied to all aspects of work in various industries and fields to produce data. These data are colorless and tasteless from the perspective of vision, but it is the colorless and tasteless data that makes the society ignore data. This idea is definitely not in line with the needs of social development [10-11]. Big data actually hides huge wealth. Making full use of big data can improve production efficiency and bring hundreds of millions of profits to production at this time. Second, use big data in an unprecedented way. Big data is often hard to predict, such as where the stock price will go tomorrow, KFC sales will rise or fall next week, these are classic examples in life. When we seriously recognize the unknowability of big data, we must use more methods and means to master data and reveal the unknown features of big data.

### **2.2 Research on efficient query model of blockchain system in the context of big data**

#### **2.2.1 Block structure of blockchain**

Blockchain applications are generally derived from the theory of international authoritative experts. The block generated in this theory produces the block edition, the stored value of the previous block, the ID of the current generation, the generation time annotation, parameter value and the block exchange serial number. The structure flow of blockchain is as follows [12-13].

First, the version number is also the block number of the block.

Second, the id of the previous generation block is because there is a block header when the previous generation block is generated, and the corresponding ID is obtained through the relevant algorithm, and the value is stored in the block.

Third, the parameter value is a value generated by searching for multiple encryption processes through a specific algorithm On account of the working mechanism of the block chain and the ID of each block chain combined with the id of the previous generation of blocks. This value must conform to algorithm rules that satisfy two generations of block ids [14-15];

Fourth, the production time label refers to the time when the block is formed. This indicator can indicate that the block is unique and the identity is encrypted.

Fifth, the difficulty value has a lot to do with the formation of the block. It usually changes with the generation time and fluctuates up and down according to the dynamic changes of the external

environment. If the system has a large number of computations, the difficulty value will be increased so that the next block generation is still in the corresponding time condition;

Sixth, the block exchange serial number refers to the number generated by the algorithm encryption of the block combined with the previous generation of block. The role of this number is to verify whether a certain business or transaction is in the block. If the system is verifying the business or transaction, the verification can be realized by the block ID, time mark, parameter value and other relevant parameters.

### 2.2.2 Efficient query model of blockchain system in the context of big data

On account of Blockchain Blockchain, a new data model is formed by combining BigData -- an efficient query model of Blockchain system under the background of BigData -- BigData Blockchain. The BigData blockchain consists of four parts: the first layer is the user layer, the second layer is the query layer, the third layer is the storage layer, and the fourth layer is the data layer. The functional flow of BigData blockchain system is shown in Figure 1.

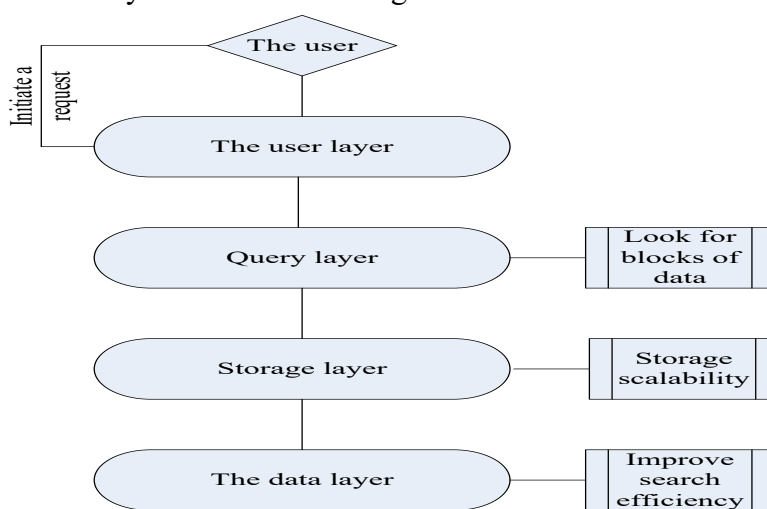


Figure1. Block Blockchain model architecture

First, according to the flow chart of function design, when the user makes a query, the system enters the user layer first and searches the data in the database of the user layer. When the qualified data is found in the database, the search is terminated. If the qualified data is not found, the system will cut into the query layer.

Second, BigData blockchain query layer applied data algorithm to search the block, looking for the corresponding value of the block or block ID;

Third, BigData block chain storage layer according to the storage principle of block chain and storage performance, to achieve replication or epitaxial function of block chain storage;

Fourthly, the data layer of BigData blockchain mainly uses BigData modules to improve the storage and search of BigData, such as improving the time of locating the block location of blockchain.

Users through BigData blockchain for 4-layer process processing, efficient implementation of the query function. When a query is made, BigData blockchain uses the BigData model to search for the query target, and connects the real-time data and the blockchain's local database to search for the target. At the same time of the data query, the block will check the steady state of the node by itself. If the stable state of the node is proved to be qualified, the node will be upgraded to the super node, which improves the efficiency of the query.

## 3. Explore The Effect of Efficient Query Model of Blockchain System in The Context of Big Data

Development environment: intelCorei70121, 30GB CPU and 32GB MEMORY PC. VMware software is used to build nodes of four levels, the number of nodes is 4, 8, 12, 16 respectively. The

Ubuntu16 system, with 2GB memory and 100GB storage space, uses open source software and combines big data to build a new model on the basis of the basic Blockchain Blockchain.

Query operation rules: Operating systems used are Blockchain model and BigData Blockchain model. Each blockchain transaction generates a record file, which contains a 2.58KB information file and the transaction ID. Each transaction is checked for transaction security, and each check is processed On account of the generated ID. The transaction process is terminated when the 3-time transaction model generates 1.5MB, 3.0MB, and 6.0MB data.

Detailed steps:

First, the Blockchain Blockchain copies data. Copy sharding mainly distributes small segments of data. The value of each segment is set to 0.5MB, and then the data is saved. The number of copies is calculated by the block chain rule strategy, and the minimum number of copies is limited to 2.

Second, the BigData blockchain storage layer uses the same principle as above for copy generation;

Third, the two blockchains have different query response speeds according to different number of nodes.

Fourthly, the number of nodes is 4, 8, 12 and 16 respectively. The query layer of BigData blockchain is configured with a corresponding number of super nodes, ranging from 1 to 4, and the others are leaf nodes. The super query node corresponds to 3 leaf nodes of different properties.

### 3.1 Formula Involved

The efficient query model of blockchain system in the context of big data generates corresponding costs. The following formula is followed for the costs:

$$C_{scan} = n * t_s + \frac{f * n}{b} * t_T \tag{1}$$

$$C_{layered} = P_{op} * t_s + P_{op} * t_T \tag{2}$$

$C_{scan}$  is block chain efficient query model view query overhead, n is the height of the current block chain,  $t_s$  is the average access time of disk blocks,  $t_T$  is block transmission time, f is the chain of the block size, b disk block size,  $C_{layered}$  is block chain layered back mode query overhead,  $P_{op}$  is that contains both o and p block chain of query results

## 4. Investigation and Research Analysis of Efficient Query Model of Blockchain System in The Context of Big Data

The storage space of blockchain data is set to 1.5MB, 3.0MB and 6.0MB respectively. According to the requirements of the experiment, 90 transaction samples and 10 fake transaction samples are selected. These records are used in the Blockchain Blockchain and BigData Blockchain Blockchain two modes to query 100 transaction samples and collect the transaction time required by each transaction sample. In the two modes, the average time required to find a transaction record is respectively at 4,8,12, and 16 node states, as shown in Table 1 and Figure 2.

Table1.BigData and Blockchain query time data table

	Blockchain	BigData
4	1400	1500
8	1900	1600
12	2700	1620
16	3400	1650

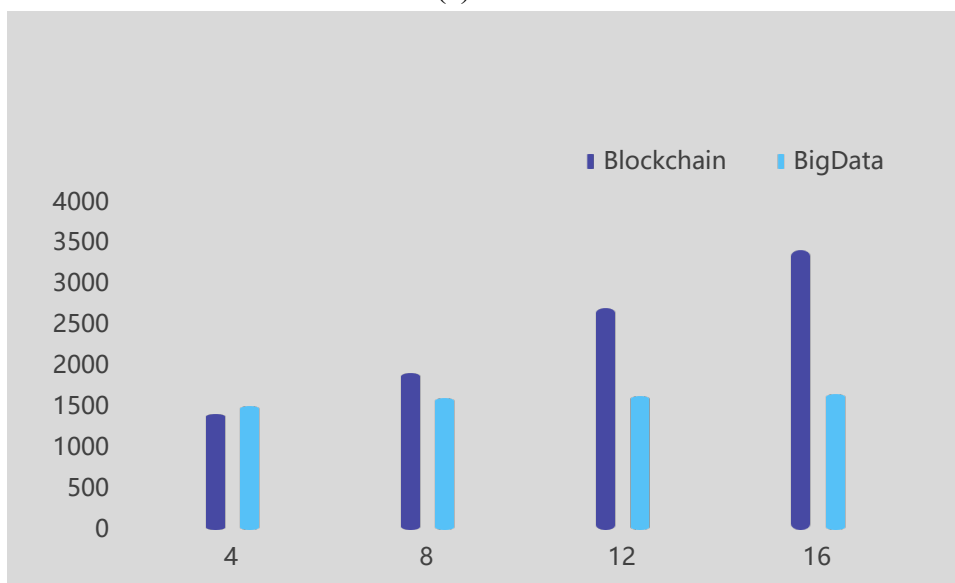
(a)1.5M

	Blockchain	BigData
4	4100	4200
8	4700	4500
12	5200	4600
16	5600	4650

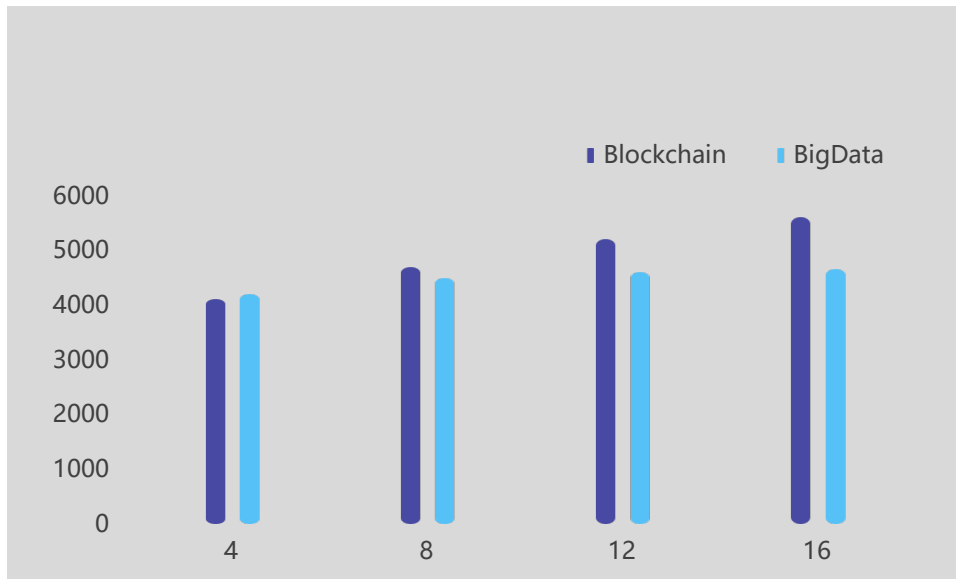
(b)3.0M

	Blockchain	BigData
4	7600	7700
8	8000	7900
12	8700	8000
16	9200	8100

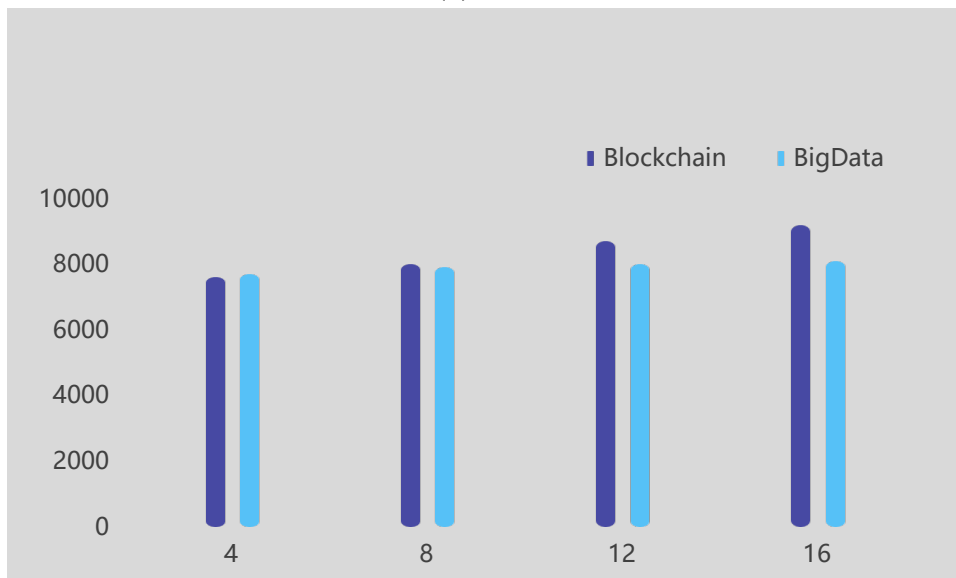
(c)6.0M



(a)1.5M



(b)3.0M



(c)6.0M

Figure2. BigData and Blockchain query time data graph

According to the above query time, the BigData Blockchain model is compared with the Blockchain model. When the node states are small, the query time of each transaction record of BigData model will be higher than that of the Blockchain. When the number of node states is gradually increasing, the query time of each transaction record of the Blockchain model is obviously increasing, but the time used by the BigData Blockchain model is growing very slowly. With more node states, the query time used by the BigData Blockchain model is obviously shorter. The Blockchain model performs fragmentation processing according to its own processing procedures, and first enters the P-chain. When the p-chain data has problems, the data model often applies other algorithms, which results in low query efficiency. However, BigData blockchain model often uses super nodes for query. When the number of super nodes in the model increases, BigData blockchain model can still obtain corresponding data according to the ID information stored by super nodes. In short, BigData blockchain model, namely the efficient query model of blockchain system in the context of BigData, is obviously higher in query efficiency and shorter in query time than other traditional models.

## 5. Conclusions

Blockchain is a research topic in many academic fields. With the increasing needs of social and economic development, r&d must enhance the performance and scalability of blockchain storage. Blockchain can meet the needs of production and life in the early stage of research and development, but when the social economy continues to develop, the requirements for Blockchain are becoming higher and higher, then it is necessary to innovate Blockchain technology, copy the Blockchain copy, fragment the copy, and improve performance. Once a copy of the blockchain is sharded, the fragmented data is stored on the nodes, which increases the scalability of the blockchain. This paper introduces the efficient query method of the data model formed by the combination of block chain and big data technology. By using big data technology, the performance of block chain is improved and the efficiency of block chain query is improved. Through data, the paper proves that the block chain technology with big data has excellent query efficiency in a variety of scenarios. The research on efficient query model of blockchain system under the background of big data has a great role in promoting efficient query field, and at the same time promotes the technological breakthrough in the field of blockchain.

## References

- [1] Guo J, Zhang M, Shang Q, et al. River Basin Cyberinfrastructure in the Big Data Era: An Integrated Observational Data Control System in the Heihe River Basin[J]. *Sensors*, 2021, 21(16):5429-5429.
- [2] Bhuvaneshwari P V, Tharini C. Review on LDPC Codes for Big Data Storage[J]. *Wireless Personal Communications*, 2021(1):1-25.
- [3] Hosseini M M, Zargoush M, Alemi F, et al. Leveraging machine learning and big data for optimizing medication prescriptions in complex diseases: a case study in diabetes management[J]. *Journal of Big Data*, 2020, 7(1):1-24.
- [4] Ghallab H, H Fahmy, Nasr M. Detection outliers on internet of things using big data technology[J]. *Egyptian Informatics Journal*, 2020, 21( 3):131-138.
- [5] Ganesan V, Garrard P. Technology, big data, and the future of paediatric neuroscience: let us go then, you and AI[J]. *Developmental Medicine & Child Neurology*, 2020, 62(8):884-884.
- [6] Hood L. L1 - How technology, big data and systems approaches are transforming 21st century healthcare[J]. *Drug Metabolism and Pharmacokinetics*, 2020, 35(1):S5-S6.
- [7] Elsotouhy M, Jain G, Shrivastava A. Disaster Management during Pandemic: A Big Data-Centric Approach[J]. *International Journal of Innovation and Technology Management*, 2021(4):2140003-2140003.
- [8] Patel H, Prajapati D, Mahida D, et al. Transforming petroleum downstream sector through big data: a holistic review[J]. *Journal of Petroleum Exploration and Production Technology*, 2020, 10(6):2601-2611.
- [9] Grimaldi D. Factors Affecting Big Data Analytics Based Innovation Processes. A Spanish Evidence[J]. *International Journal of Innovation and Technology Management*, 2020(2):2050036-2050036.
- [10] Haseeb K, Lee S, Jeon G. EBDS: An energy-efficient big data-based secure framework using Internet of Things for green environment[J]. *Environmental Technology & Innovation*, 2020, 20(3):101129-101129.
- [11] Haimed A, Saba T, Albasha A, et al. Viral reverse engineering using Artificial Intelligence and big data COVID-19 infection with Long Short-term Memory (LSTM)[J]. *Environmental Technology & Innovation*, 2021(4):101531-101531.
- [12] Müller Marion. Unternehmen halten Blockchain für wichtig[J]. *Die Aktiengesellschaft*, 2021, 66(17):r253-r254.
- [13] Mehta S, Grant K, Ackery A. Future of blockchain in healthcare: potential to improve the accessibility, security and interoperability of electronic health records[J]. *BMJ Health & Care Informatics*, 2020, 27(3):e100217-e100217.

- [14] Powell W , Foth M , Cao S , et al. Garbage In Garbage Out: The Precarious Link between IoT and Blockchain in Food Supply Chains[J]. Journal of Industrial Information Integration, 2021(10):100261-100261.
- [15] Tamar M . Philipp Hacker, Ioannis Lianos, Georgios Dimitropoulos, and Stefan Eich (eds.), Regulating Blockchain: Techno-Social and Legal Challenges[J]. Chinese Journal of International Law, 2021(4):4-4.