

CONCEPT OF PEER-TO-PEER CACHING DATABASE FOR TRANSACTION HISTORY STORAGE AS AN ALTERNATIVE TO BLOCKCHAIN IN DIGITAL ECONOMY

M.A. Belov^a, S.I. Grishko, E.N. Cheremisina, N.A. Tokareva

¹ *System Analysis and Control Department, Dubna State University, Universitetskaya 19, 141980, Dubna, Russia*

E-mail: ^a belov@uni-dubna.ru

This paper discusses the concept of a distributed horizontally scalable and cascadable peer-to-peer caching database, optimized for the digital economy needs and suitable for storing the history of a large number of transactions of every citizen involved in business processes based on digital technologies, starting from receiving public and social services in electronic form and ending with consumption of electronic goods and services produced by e-business and e-commerce. We also offer an approach to organizing student teamwork for the development of the solution at the Dubna State University based on the use of our innovative data center project «Virtual Computer Lab».

Keywords: database, NoSQL, peer-to-peer, blockchain, blockchain issues, smart contracts, digital economy, distributed computing systems, data management, virtual computer lab.

Mikhail Belov, Stanislav Grishko, Eugenia Cheremisina, Nadezhda Tokareva

Copyright © 2021 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Introduction

The development of the digital economy implies storing the history of a large number of transactions of every citizen involved in business processes based on digital technologies, starting from receiving public and social services in electronic form and ending with consumption of electronic goods and services produced by e-business and e-commerce.

2. Blockchain issues

Today, there is the idea of blockchain - a continuous, sequential chain of blocks containing information (transaction history) built according to certain rules, where blocks are stored and processed on many different computers (computing devices). At certain stages of development, however, any blockchain is subject to the so-called "51% Attack," because the idea of total decentralization embedded in blockchain allows blockchains to be trusted by an exceptionally large number of participants who form a "controlling block" of generating capacity, allowing blockchain complexity to rise to a level where it is not realistic for attackers who can undermine blockchain trust by introducing a fake block chain to gain computing power greater than the rest. If attackers with superior computing power manage to create a persistent chain of blocks (usually at least 6) and those blocks are replicated across all of the participants' personal computers, where block removal is not envisioned by the idea of blockchain itself, then the blockchain will be discredited [1].

The second problem with blockchain is that, in general terms, the client needs to store on his or her device all the blocks of blockchain data that will accumulate like an avalanche over time, with no way to delete them (this is the initial technological "highlight" of blockchain), and most of this data is of no interest to any blockchain participant, while only the manufacturers of computer, telecommunications equipment and mobile devices benefit. This poses an additional social problem because, if the digital economy is to involve all segments of the population, it will require providing all Russian citizens with expensive computer hardware and smartphones that can be damaged or lost in use, which will require their immediate replacement because a person without an electronic gadget cannot be a full-fledged participant in the digital economy. The problems and difficulties in providing citizens with electronic devices can lead to a sharp increase in social tensions and/or social inequalities. Given the above arguments, the idea of blockchain adoption in healthcare, postal, transportation and other socially important services seems utopian today.

The third problem with blockchain is the lack of a built-in technology for fast search of transactions within a block of data, which goes against the principles of the digital economy, a paradigm of which relies on quick access to relevant data and transaction history as part of the business processes in which the user is involved.

In addition, it should be noted that smart contracts are only a paradigm that requires a decentralized data repository, and all actions are represented as mathematical rules.

2. Background

If we look carefully at the data structure within a digital economy system, we can see that transactions are grouped with respect to a natural unique identifier of a citizen, which allows for efficient block distribution based on their hash across the node-segments of a scalable peer-to-peer NoSQL DBMS, eliminating the appearance of "hotspots", and the data itself can be easily represented in the form of "key-value" tuples and stored in a columnar structure, providing a quick search due to the gossip protocol that allows redirecting requests to the node whose responsibility range includes the hash of a specific unique identifier. Well, since we are talking about storing transaction history within the business processes of the digital economy, the key-value relationship is essentially a one-to-many

relationship, in the context of a design focused on personalized information output for each specific user. Ensuring communication between users within groups and communities (many-to-many relationship) is also possible and can be implemented by means of secondary indexes, materialized representations, or partial data redundancy through denormalization, depending on the power of the set that the data form, to ensure acceptable performance of sampling queries.

If we talk about the task of quickly obtaining summary or aggregated statistical information, it is not difficult to solve it by loading the necessary data in the YARN cluster of the open technology platform Apache Hadoop, for example, in the processing environment of Random Access Memory SPARK, applying the principle of Resilient Distributed Datasets and basic concepts of building a pipeline of operations mapping, moving, sorting and convolution in the framework of functional programming.

However, the relative simplicity of horizontal scaling of disk space, processing power and RAM, does not provide transactional scaling, as simultaneous access of many users to the central database nodes, would make the bandwidth of the data network a bottleneck. Therefore, we need a peer-to-peer caching database that will store all relevant data for a particular user on their device and the closest peer-to-peer servers, based on selected proximity criteria according to a given set of features and attributes.

If we rise to an empirical level, from the perspective of participants in the digital economy, it is a question of storing a set of facts. Facts in a database are immutable; once stored, they do not change. However, old facts may be replaced by new facts over time or due to circumstances. The state of the database is the value determined by the set of facts in effect at a given point in time. So, this analysis allows us to move on to a more detailed consideration of the architecture of the proposed peer-to-peer caching database design solution [2-4].

3. Solution Architecture

A peer-to-peer client library (a peer-to-peer access library) is embedded into the client application and allows to get data from the peer-to-peer servers, cache data on the client device (to reduce the load on the peer-to-peer servers), while keeping such an important property as "final immutability", and also to exchange the peer-to-peer server lists between the clients.

The peer-to-peer server provides data access by caching the necessary segments of the central database demanded by the connecting clients. Connection to a specific group (farm) of peer-to-peer servers is determined by specified criteria, which can be geolocation data, type of users, type of processes, type of transactions, etc. Peer-to-peer servers can exchange data segments with each other (peer-to-peer communications), and store as many data segments as the storage system quotas and limitations allow. In certain cases, a client application may also act as a peer-to-peer server, but there are threats of loss of data integrity and validity through the emergence of fake peer-to-peer servers on the network, created by hackers to discredit it.

Records in the central database (if developers wish, in parallel to peer-to-peer servers) can be made by means of transactors, which accept write transactions and process them serially, ensuring guaranteed integrity until successful synchronization with the central database, due to the replication factor of the distributed network file system (odd number of servers greater than 3 (three) is recommended, to ensure a recording quorum), where open technology solutions based on Apache Hadoop HDFS or Apache Cassandra can be selected as the basis. However, HDFS fault tolerance will require the use of additional components such as Zookeeper, Zookeeper Failover Controller and Quorum Journal Manager.

Access to the transactor is recommended as part of a service-oriented architecture, through REST-services that can be scaled by applying standard load-balancing technologies used in web server deployments. This approach allows providing access to the transactor through the usual HTTP protocol, and transactors themselves and the centralized database will be in an isolated network, access

to which should be done via routing with the use of modern encryption technologies, and hacker attacks via HTTP protocol can be prevented by modern IPS systems, combining signature and heuristic approaches of malicious activity detection.

According to the principles of organizing access to the transactor, access to the central data repository can be easily organized as well. The proposed approach makes it possible to implement staggered isolation of the central database and cascading of network traffic using peer-to-peer server farms and service-oriented architecture.

4. Environment for teamwork and collaboration

To provide our students with the opportunity to develop the peer-to-peer caching database solution, we replaced the physical computers with virtual machines in the Virtual Computer Lab. The Virtual Computer Lab provides a set of software and hardware-based virtualization and containerization tools which enable flexible and on-demand provision and use of computing resources in the form of cloud Internet services with an integrated knowledge management system using the principles of self-organization, functioning as a homogeneous environment with elements of cognitive representation of internal operational resources based on visual models and partial automation of fundamental technological operations with the expert system for carrying out research projects, resource-intensive computational calculations and tasks related to the development of sophisticated corporate and other distributed information systems. The Virtual Computer Lab self-organization makes the transition from a complex system of granular group security policies with many restrictions to the formation of personal responsibility and respect for colleagues, which should be a solid foundation for strengthening and developing classical cultural values in the educational environment. [5–8].

5. Conclusion

In conclusion, it would be useful to note that the proposed concept of a distributed horizontally scalable and cascable peer-to-peer caching database could become the basis for a modern, efficient, as well as easy-to-implement and maintain technological platform for the implementation of digital economy services in the Russian Federation.

References

- [1] Imran Bashir. *Mastering Blockchain: A deep dive into distributed ledgers, consensus protocols, smart contracts, DApps, cryptocurrencies, Ethereum, and more*, 3rd Edition, Packt Publishing (August 31, 2020).
- [2] Lori Jo Underhill, *Defining the Digital Economy: The Structure of the Digital Economy in Focus*, Independently published (February 15, 2019).
- [3] Tim Jordan, *The Digital Economy*, Polity; 1st edition (January 28, 2020).
- [4] Maria Wasastjerna, *Competition, Data and Privacy in the Digital Economy: Towards a Privacy Dimension in Competition Policy?* (International Competition Law), Wolters Kluwer (July 16, 2020).
- [5] M.A. Belov, Y.A. Kryukov, M.A. Miheev, P.E. Lupanov, N.A. Tokareva, and E.N. Cheremisina, *Sovremennye informatsionnye tekhnologii i IT-obrazovanie* 14, 4, 823–832 (2018).
- [6] M.A. Belov, Y.A. Krukov, M.A. Mikheev, N.A. Tokareva, and E.N. Cheremisina, *CEUR Workshop Proceedings* 2267, 207–212 (2018).
- [7] E.N. Cheremisina, M.A. Belov, N.A. Tokareva, S.I. Grishko, and A.V. Sorokin, *CEUR Workshop Proceedings* 2023, 299–302 (2017).
- [8] M.A. Belov, E.N. Cheremisina, and S.V. Potemkina, *Journal of Emerging research and solutions in ICT* 1, 2, 39–46 (2016).