Объединенный институт проблем информатики
Национальной академии наук Беларуси

XXI Международная
научно-техническая конференция

РАЗВИТИЕ ИНФОРМАТИЗАЦИИ И ГОСУДАРСТВЕННОЙ СИСТЕМЫ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

РИНТИ-2022

17 ноября 2022 г., Минск

Доклады

Минск
ОИПИ НАН Беларуси
2022

УДК 004.942

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ БЛОКЧЕЙН В OLAP-СИСТЕМАХ

Г. Ч. Набибекова
Институт информационных технологий
Национальной академии наук Азербайджана, Баку

Проведен анализ работ, посвященных различным подходам к применению технологии блокчейн для обеспечения информационной безопасности баз данных в OLAP-системах, для чего предложено использовать частный блокчейн для защиты данных электронной демографической системы, функционирующей в среде электронного государства.

Введение

Эпоха Индустрии 4.0 характеризуется возникновением новых классов угроз. В связи с этим появились новые требования к обеспечению информационной безопасности различных систем. К таким системам, которые в последнее время широко используются как в государственных, так и частных структурах, относятся OLAP-системы, разрабатываемые на базе OLAP-технологий. Их ключевой компонент – хранилище данных (ХД). В основе OLAP-технологий лежит представление информации в виде OLAP-кубов, которые содержат показатели, используемые для анализа и принятия управленческих решений [1]. В докладе проведен анализ работ, посвященных различным подходам к применению технологии блокчейн для обеспечения безопасности ХД в OLAP-системах.

1. Основные характеристики технологии блокчейн:

Блокчейн, или технология распределенного реестра (DLT, Distributed Ledger Technology) – это распределенная база данных (БД), объединяющая узлы компьютерной сети. Блокчейн дает возможность записывать и распространять цифровую информацию, но не позволяет ее редактировать [2]. Каждый блок блокчейна содержит временную метку его создания, хеш-код текущего блока, хеш-код предыдущего блока, которые вместе составляют метаданные блока, а также непосредственно информацию [3]. В блокчейне используется одноранговая сеть P2P (Peer-to-Peer), которая создает равноправие всех ее участников и является еще одним способом обеспечения информационной безопасности [4]. Использование одноранговой сети P2P означает, что каждый присоединившийся к сети становится ее полноправным участником и получает полную копию блокчейна. Это делает практически невозможной фальсификацию данных, внесенных в систему. Отметим, что многие инфраструктуры блокчейнов применяют так называемые смарт-контракты. Они представляют собой фрагменты кода, включенные в реестр, которые реализуют соглашения между сторонами и используются для описания обработки собранных многомерных данных. Наличие смарт-контрактов в блокчейне при хранении в нем данных OLAP-систем осуществляет выполнение агрегаций по измерениям и других вычислений OLAP, устанавливаемых пользователем. В результате вычислений, включенных в смарт-контракты, определяются представления данных – смарт-представления, на основании которых можно проводить анализ данных.

2. Анализ подходов к хранению данных OLAP-системы в реестре блокчейна

В настоящее время наблюдается тенденция использования блокчейна в условиях, где доминирует применение технологии БД. Поскольку OLAP является элементом ХД,

были рассмотрены исследовательские работы в данной области. Установлено следующее: несмотря на то, что блокчейн является открытым протоколом, в этом направлении существуют разные платформы, которые функционируют изолированно друг от друга.

Одна из таких платформ – Quantum Ledger Database (QLDB) от Amazon [5] – является полностью управляемой БД реестров, которая обеспечивает прозрачный, неизменяемый и проверяемый криптографическими методами журнал транзакций. Так как основной целью QLDB выступает хранение журнала транзакций, ее можно использовать в архитектуре OLAP для хранения записей фактов ХД только в режиме добавления. В отличие от блокчейна Ethereum, в QLDB отсутствует поддержка децентрализации.

Как отмечено в статье [6], несмотря на то, что блокчейны получили широкую популярность, они до сих пор не используются в качестве общей БД. Основным препятствием для подобного применения является их ограниченная масштабируемость и производительность. Кроме того, в блокчейнах часто отсутствуют интерфейс запросов и четко определенные уровни согласованности, которые определяют, когда и как обновления становятся видимыми. В данной работе авторы представляют BlockchainDB, в котором устранены вышеуказанные недостатки. BlockchainDB использует блокчейны в качестве слоя для хранения данных.

В статье [7] система БД блокчейна представлена как система БД с обычным интерфейсом SQL. Однако она обеспечивает те же гарантии неизменности переходов состояний, доверия и проверки с открытым исходным кодом, что и блокчейн. Таким образом, взаимодействующие компании могут перенести свое общее состояние в БД блокчейна и тогда они будут взаимодействовать с ней как с любой другой БД, используя SQL. Поскольку БД блокчейна представляет собой физически другую систему БД, взаимодействие с ней по-прежнему должно происходить через посреднический слой. Для устранения этого пробела предлагается абстракция общей таблицы БД в облаке, которая является частью БД обеих взаимодействующих компаний. Соединяя воедино идею БД блокчейна и понятие просматриваемой таблицы, создается общая проверяемая таблица, которая является частью БД обеих взаимодействующих компаний и имеет неизменяемый доступный журнал с возможностями аудита. Таким образом, предлагается концепция общей проверяемой таблицы, которая интегрирует таблицы из БД блокчейна непосредственно в БД взаимодействующих компаний (рис. 1, 2).

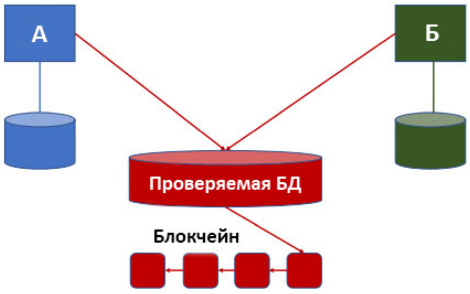

Рис. 1. Проверяемая БД

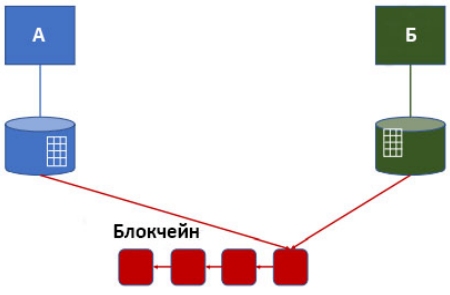

Рис. 2. Проверяемые таблицы

Использование блокчейна в качестве неизменяемого реестра для хранения исторических фактов в децентрализованном развертывании ХД исследовано в статье [8]. Предлагается применять децентрализованное ХД, реализованное поверх блокчейна. В нем смарт-контракты используются в качестве средства описания обработки собранных многомерных данных. Смарт-контракты определяют агрегированные представления – смарт-представления (рис. 3).

Как известно, структура ХД состоит из множества фактов. Факты распределяются по блокам блокчейна, в результате чего образуется распределенный реестр блоков. Таким образом, реестр блокчейна заменяет таблицу фактов в традиционном ХД. Каждый факт, встроенный в блок, согласно характеристике блокчейна, имеет отметку времени и криптографически защищен хеш-кодом, а также содержит хеш-код предыдущего блока. Реестр также используется для хранения смарт-контрактов, которые исследуют и реализуют аналитические операции над необработанными данными в форме смарт-представлений и для кодирования доказательств для материализованных результатов смарт-представлений, которые хранятся в кеше представлений.

Рис. 3. Использование смарт-представлений для создания децентрализованного ХД

В предлагаемой архитектуре определение и свойства интеллектуальных представлений (хранящихся в реестре) отделены от их обработки и обслуживания, которые осуществляются соответствующими модулями управления данными. Такая архитектура отличается от ранее представленных подходов, в которых либо добавляется уровень базы данных с возможностями запросов поверх блокчейна, либо расширяются БД.

Компонент *распределенный реестр* использует технологию блокчейна для реализации постоянной, нестираемой и неизменной истории необработанных наблюдений (фактов). Компонент *кеш представлений* – это ХД в основной памяти, которое хранит результаты интеллектуальных представлений, инициированные выполнением соответствующего смарт-контракта на сервере приложений. Компонент *серверная часть SQL* – это любая серверная часть БД, предлагающая возможности SQL. Компонент *сервер приложений* управляет всем процессом определения, хранения, повторного использования и обновления смарт-представлений.

3. Свойства частного блокчейна

Все описанные выше блокчейны являются публичными. Присущие им свойства позволяют эффективно реализовывать только два из трех ключевых аспектов информационной безопасности – целостность и доступность информации. Поскольку децентра-

лизованный публичный блокчейн не может обеспечить третий важнейший аспект информационной безопасности – конфиденциальность данных, появилась модель частного блокчейна, который позволяет обеспечивать:

– доступ к реестру с правом вносить в него изменения только авторизованным участникам. Такая модель уже не является децентрализованной, хотя и остается распределенной;

– конфиденциальность записей, так как теперь доступ к модели предоставляется согласно политике безопасности,

и используется в основном как инфраструктура для корпоративных и государственных задач.

В работе [9] разработана система сравнительного анализа блокчейнов Blockbench для выяснения, в какой степени блокчейн может справиться с нагрузкой по обработке данных и какую платформу выбрать из множества доступных сегодня. На рис. 4 представлен стек программного обеспечения Blockbench.

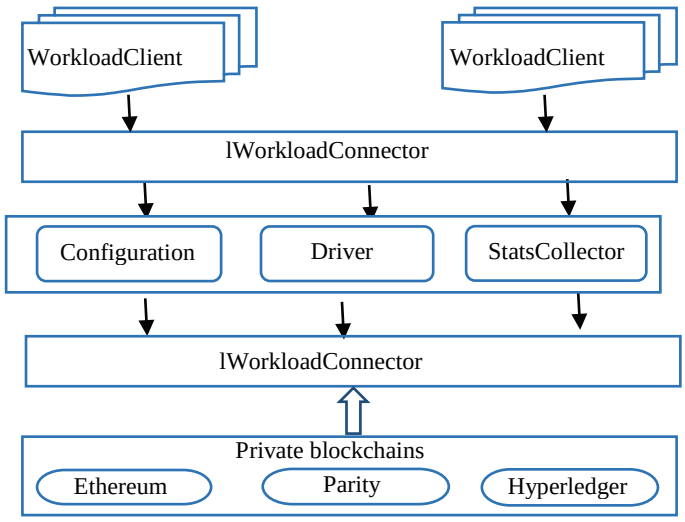

Рис. 4. Стек программного обеспечения Blockbench

Среда Blockbench является оценочной для анализа частных разрешенных блокчейнов. Blockbench измеряет общую и покомпонентную производительность с точки зрения пропускной способности, задержки, масштабируемости и отказоустойчивости.

Новые рабочие нагрузки WorkloadClient добавляются с помощью интерфейса IWorkloadConnector, а новые серверные части блокчейна – путем реализации IBlockchainConnector. Согласно рис. 4 текущие серверные части включают Ethereum, Parity и Hyperledger. Основным компонентом Blockbench является Driver, который принимает в качестве входных данных рабочую нагрузку, определяемую пользователем конфигурацию (количество операций, клиентов, потоков и т. д.), выполняет ее в блокчейне и выводит текущую статистику.

4. Практическое применение частного блокчейна

На сегодняшний день в мире повсеместного использования информационных технологий и их влияния на все сферы жизнедеятельности человека появилось новое направление демографии – электронная демография. В концептуальную модель электронной демографической системы в качестве одного из ключевых блоков включена «оперативная аналитическая обработка» – OLAP-технология [10]. Повреждение, изме-

нение или уничтожение данных злоумышленниками в ХД электронной демографической системы могут полностью поменять картины в сфере демографии, которые предоставляют OLAP-кубы. Очевидно, что это приведет к отрицательным последствиям при анализе этих данных лицами, принимающими решения в сфере демографии. Из-за этого в сфере демографии могут быть приняты неверные решения, что, в свою очередь, повлияет на демографические процессы в регионе.

Как отмечено выше, задачи в сфере демографии решаются на государственном уровне. Поэтому с целью обеспечения безопасности данных в сфере демографии вместо традиционных технологий предлагается использовать технологию именно частного блокчейна.

Заключение

В эпоху Индустрии 4.0 безопасность данных приобретает особое значение. Поэтому вместо традиционных технологий защиты данных от злоумышленников более эффективными являются инновационные технологии. К ним относится технология блокчейн, которая имеет множество преимуществ, таких как прозрачность транзакций, сохранность данных, безопасность и т. д. Но прозрачность приводит к нарушению конфиденциальности данных. В связи с этим в корпоративных и государственных структурах рекомендуется использовать не публичный, а частный блокчейн. Это относится и к сфере демографии, поскольку решение задач и проведение исследований здесь также осуществляются на государственном уровне. В связи с актуальностью данной темы дальнейшие исследования в этом направлении, связанные с разработкой технологической модели частного блокчейна, будут продолжены.

Список литературы

1. Microsoft SQL Server 2005 Analysis Services. OLAP и многомерный анализ данных / А. Бергер [и др.]. – СПб. : БХВ-Петербург, 2007. – 922 с.

2. Hayes, A. Blockchain Explained? [Electronic resource] / A. Hayes. – 2022. – Mode of access: https://www.investopedia.com/terms/b/blockchain.asp. – Data of access: 20.04.2022.

3. Denis T. Hash Function [Electronic resource] / T. Denis, S. Johnson. – 2007. – Mode of access: https://www.sciencedirect.com/topics/computer-science/hash-function. – Data of access: 20.01.2022.

4. Cope, J. What's a Peer-to-Peer (P2P) Network? [Electronic resource] / J. Cope. – 2002. – Mode of access: https://www.computerworld.com/article/2588287/networking-peer-to-peer-network. html. – Data of access: 25.01.2022.

5. AWS. Amazon Quantum Ledger Database (QLDB) [Electronic resource]. – Mode of access: https://aws.amazon.com/ru/qldb/. – Data of access: 27.01.2022.

6. Blockchaindb – A shared database on blockchains / M. El-Hindi [et al.] // Proceedings of the VLDB Endowment. – 2019. – Vol. 12, № 11. – P. 1597–1609.

7. Veritas : Shared verifiable databases and tables in the cloud [Electronic resource] / L. Allen [et al.] // Online Proceedings of the 9th Conference on Innovative Data Systems Research (CIDR). – 2019. – Mode of access: http://www.cidrdb.org/cidr2019/papers/p111-gehrke-cidr19.pdf. – Data of access: 27.01.2022.

8. Messanakis, K. Smart-Views: Decentralized OLAP View Management Using Blockchains / K. Messanakis, P. Demetrakopoulos, Y. Kotidis // 23rd International Conference,

DaWaK 2021 Big Data Analytics and Knowledge Discovery : Virtual Event, Proceedings. – 2021. – Springer, Cham. – Vol. 12925. – P. 216–221.

9. BLOCKBENCH : A framework for analyzing private blockchains / T. Dinh [et al.] // SIGMOD '17 : Proceedings of the 2017 ACM International Conference on Management of Data, Chicago. – Chicago, 2017. – P. 1085–1100.

10. Nabibayova, G. Ch. Decision Support System In Electronic Demography. CEUR Workshop Proceedings / G. Ch. Nabibayova. – UkrPROG2020, 12th International Conference of Programming, Kyiv, Ukraine. – Kyiv, 2020. – P. 228–235.