

УДК 004.62

РАЗРАБОТКА ПОДХОДА К ПОСТРОЕНИЮ СИСТЕМ УПРАВЛЕНИЯ СЕМАНТИЧЕСКИМИ ДАННЫМИ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИИ BLOCKCHAIN

Олимпиаев Н.В. (Университет ИТМО)

Научный руководитель – кандидат технических наук, доцент Жукова Н.А.
(Университет ИТМО)

Введение. Данные играют все более важную роль в развитии бизнеса. Их сбор, анализ и хранение требует надежных и эффективных систем управления. По информации от компании Statista, за 2021 год объем данных по всему миру составил 79 зеттабайт и продолжает расти [1]. Этот тренд побуждает развивать системы управления данными, а связан он с тем, что для бизнеса данные становятся активами. Внимание в работе сосредоточено на виде систем управления данными - Master Data Management (MDM). Системы MDM фокусируются на управлении ключевыми данными в организации (клиенты, продукты, поставщики), обеспечении единого и актуального представления о сущностях и оптимизации бизнеспроцессов. Семантические данные часто распределены, а их сбор требует работы с разнородными источниками, что приводит к проблемам с согласованностью и достоверностью данных. Обеспечить такие потребности способна технология Blockchain, интерес к которой растет во многих отраслях. Blockchain использует распределенный реестр активов, обеспечивая децентрализацию, а также обладает свойствами, которые могут поддерживать конфиденциальность, безопасность и целостность данных, поэтому использование технологии является актуальным вопросом. Тема использования Blockchain для управления данными и взаимодействия с технологиями Semantic Web [2] освещена в зарубежных источниках, в них рассматриваются общие сценарии, при этом трудов на тему использования Blockchain для MDM нет [3]. Таким образом, практическая применимость результатов исследований для MDM ограничена. Решением проблемы внедрения Blockchain для взаимодействия с семантическими данными является разработка подхода к построению системы управления таким образом, чтобы ее результаты могли быть практически применимы разработчиками.

Основная часть. Для разработки подхода потребуется рассмотреть ограничения существующих MDM-систем, которые используются в организациях. Проанализировав их, можем выделить ряд ограничений. Во-первых, MDM полагаются на качество ключевых данных, при этом проблемой остается обеспечение качественной обработки, когда данные используются несколькими пользователями. Во-вторых, MDM позволяют управлять ключевыми и конфиденциальными данными, поэтому их централизация является потенциальной уязвимостью. По мере роста объема данных важна масштабируемость, для чего необходимо гарантировать обработку больших объемов и поддерживать производительность, при этом данные филиалов организации необходимо интегрировать и собирать воедино, что приводит к задержкам в получении полезной информации из данных. При разработке подхода будут использованы методы декомпозиции, анализа и выделения неэффективных реализаций, моделирования и проектирования.

Создание системы управления данными на основе Blockchain включает в себя несколько технических аспектов: разработку смарт-контрактов, моделирование данных и интеграцию с другими системами. Необходимо разработать архитектуру основанной на технологии Blockchain системы, позволяющей управлять источниками данных и обеспечивать согласованность результирующей онтологии за счет проверок корректности изменения и применения правил качества. Таким образом, система будет руководствоваться принципами Master Data Management для онтологий на основе технологии Blockchain. Основной упор сделан на решение проблем интероперабельности и осуществления безопасного доступа, а также обеспечения качества результирующих данных. Для реализации данных возможностей

необходимо наличие проверок корректности изменения, например проверка синтаксиса запроса, согласованности онтологии после выполнения запроса, а также разработка и применение правил к отношениям в результирующих онтологиях. Обеспечение данных практик возможно за счет принципа транзакционности при публикации данных, а также подтверждением транзакций с помощью использования Blockchain. Для проектирования архитектуры системы MDM на основе Blockchain с учетом известных сложностей реализации были выбраны следующие технологии и методы: архитектура основана на сочетании Ethereum 2.0 с механизмом консенсуса Proof of Stake, IPFS, смарт-контрактов и семантических технологий; система имеет пользовательский интерфейс и поддерживает SPARQL для запросов; интеграция с внешними службами осуществляется через API и очереди сообщений; правила качества основаны на службах валидации и обогащения; взаимодействие с локальной базой данных позволяет сохранить моментальный снимок онтологии для повышения производительности; проверки согласованности осуществляются в Apache Jena; IPFS используется для хранения больших файлов, а Ethereum — для записи изменений в Blockchain и взаимодействия со смарт-контрактами. Разрабатываемая система создана как обособленный центр управления и обработки данными, но при этом возможно внедрение инструмента в комплексную систему или сеть по контролю за данными в рамках концепции Semantic Web. Для этого был разработан интеграционный API, являющийся одной из центральных особенностей системы.

Выводы. Анализ существующих решений позволил подтвердить, что Blockchain возможно эффективно использовать для улучшения систем MDM. Разработана архитектура, отличия и уникальность которой заключается в использовании Ethereum 2.0 и PoS, IPFS и смарт-контрактов; интеграции с внешними системами, управлении ключевыми данными MDM и обеспечении согласованности семантических данных в Apache Jena. Подход к построению архитектуры MDM системы на базе Blockchain способен решить проблемы существующих исследований, предлагая более эффективные способы, по сравнению с используемыми системами. Разработка системы управления данными в рамках концепции Semantic Web является важным и уникальным направлением из-за своей сконцентрированности на взаимодействии с семантическими данными и использовании концепции MDM. Внедрение Blockchain имеет большое количество сложностей, но перспектива получаемых преимуществ оправдывает инвестиции в дальнейшую разработку. Технология Blockchain может революционизировать управление основными данными, предоставляя безопасный и прозрачный способ хранения, обмена и доступа к данным, а разработанный подход может быть практически применим для разработки децентрализованной системы управления семантическими данными.

Список использованных источников:

1. 2021 in numbers | Statista // Statista - The Statistics Portal for Market Data, Market Research and Market Studies. URL: <https://www.statista.com/study/105887/2021-in-numbers/> (дата обращения: 19.02.2023).
2. Blockchain-Based Transaction Manager for Ontology Databases / K. Timotej, G. Domen, V. Marko [et al.] // Informatica. – 2022. – Vol. 33, No. 2. – P. 343-364.
3. Application of Blockchain Technology in Data Management: Advantages and Solutions / L. Wen, L. Zhang, J. Li // Big Scientific Data Management. BigSDM 2018. Lecture Notes in Computer Science. Vol. 11473. – Springer, Cham, 2019. – P. 239-254.