

Ігор Цимбалюк

асистент кафедри ІІЗ, Національного лісотехнічного університету України, Львів, Україна. <https://orcid.org/0009-0008-8489-473X>.
ihortsymbaliuk@nltu.edu.ua

Методи і засоби інкрементальної мутації потоку даних в архітектурі розподілених систем

Анотація. *Описано принципи подання подій у розподілених системах, як суб'єктів управління станом із застосуванням структури направленого ациклічного графа та ймовірнісних фільтрів. Введено концепцію мета-мутацій для модифікації умов та властивостей середовища в якому відбувається транзит повідомлень. Проведено оцінку ефективності порівняно з стандартними підходами до вирішення подібних завдань. Результати дослідження можна використати для створення систем з складною ієрархією прав доступу до інформації та керування даними з інтегрованими механізмами контролю версій та політик безпеки на рівні записів.*

Ключові слова – *направлений ациклічний граф, ймовірнісний фільтр, паралельні обчислення, мета-мутації, CAP, CQRS.*

Розвиток розподілених систем обробки даних зумовлює необхідність переосмислення традиційних підходів до керування станом та синхронізації інформації між численними вузлами мережі. Центральною концепцією запропонованого підходу постає ідея організації передачі подій через структуру направленого ациклічного графа, що надає змогу відображати каузальні залежності між трансформаціями даних та забезпечувати можливість відтворення довільних історичних станів системи. Відмінність від традиційної архітектури журналізації полягає у відмові від лінійної послідовності блоків на користь графової топології, яка забезпечує паралельну обробку множин транзакцій без необхідності дотримання жорсткого глобального порядку виконання операцій.

Архітектурна основа системи передбачає розподіл адресного простору між обчислювальними вузлами таким чином, що кожен вузол відповідає за обробку подій у певному діапазоні адрес клієнтських сутностей [1]. Такий механізм забезпечує спроможність підвищувати горизонтальну масштабованість

шляхом додавання нових вузлів із перерозподілом адресних діапазонів. Повідомлення може бути направлене у необхідний канал ще на стороні клієнта, усуваючи такий архітектурний компонент як API Gateway, який є потенційною точкою відмови, адже не може бути ефективно реплікований. Динамічне оновлення конфігурації адресних діапазонів надасть змогу адаптуватися до змін у розподілі навантаження, переміщуючи межі зон відповідальності між вузлами без переривання обслуговування запитів клієнтів.

Критичною вимогою до комунікації у розподіленій системі є забезпечення ідемпотентності обробки повідомлень, оскільки асинхронна природа комунікацій між компонентами створює ризики багаторазової доставки ідентичних повідомлень. Застосування імовірнісних структур даних, зокрема фільтрів Блума, надає ефективний механізм виявлення дублікатів при мінімальних витратах пам'яті [2]. При надходженні нового повідомлення спочатку виконується перевірка присутності його ідентифікатора у фільтрі – позитивний результат свідчить про попередню обробку цього повідомлення, що призводить до його відхилення без подальшої обробки.

Структура направленого ациклічного графа використовується для довготривалого збереження історії подій та відображення каузальних зв'язків між перетвореннями даних. Кожен новий стан додається до графа як новий вузол, а його ребра з'єднують вершини, які відповідають операціям виконаним із використанням даних інших вузлів для отримання цільового. Така модель подій надає природний механізм реконструкції історичних станів системи шляхом обходу графа до певної глибини або до конкретної часової мітки. Будь-який стан системи може бути відтворений через послідовне застосування подій, що формують шлях від початкового стану до цільового моменту часу.

Концепція мета-мутацій описує фундаментальне розширення поточної парадигми модифікації даних у розподілених системах. Замість безпосередньої зміни значень даних, мета-мутації оперують метаданими та правилами інтерпретації інформації, що призводить до опосередкованих трансформацій фактичних значень через зміну контексту їх розгляду. Подібно до механізму ефектів у

реактивному програмування, описані суб'єкти виконують запрограмовані операції при настанні певних умов. Відмінність полягає у фокусі на маніпуляціях зі поданням сутностей в графі подій замість безпосередньої модифікації даних.

Практичне втілення мета-мутацій передбачає створення підпрограм, які аналізують поточний стан графа подій та визначають необхідні трансформації у метаданих на підставі заданих умов. Така логіка може містити правила автоматичного делегування прав доступу при виникненні певних патернів взаємодії між клієнтами, динамічну зміну конфігурації маршрутів передачі повідомлень залежно від поточного навантаження на систему або адаптивну зміну параметрів обробки подій у відповідь на виявлені аномалії. Результатом виконання мета-мутації стає модифікація структури зв'язків у графі, що опосередковано впливає на інтерпретацію та обробку даних, пов'язаних із цими зв'язками.

Описана архітектура надає наступні переваги у порівнянні з існуючими:

- відсутність єдиної точки відмови у вигляді шлюзу балансувальника навантаження;
- уникнення операцій додаткового визначення та перевірки прав доступу;
- збереження повної історії роботи кожного компонента системи;
- підвищення якості телеметрії.

Висновки. Встановлено теоретичні основи інкрементальної мутації потоків даних у розподілених системах через аналіз сучасних підходів до керування станом. Проведений огляд наукових джерел виявив обмеження існуючих рішень щодо забезпечення ідемпотентності операцій у високонавантажених середовищах та необхідність розробки нових методів представлення каузальних зв'язків між подіями у потоках даних. Формалізовано концепцію мета-мутацій як багаторівневої моделі операцій над потоками даних, що надає змогу здійснювати трансформації не лише значень даних, але й правил їх обробки, семантики взаємодій та механізмів контролю доступу. Запропоновані рішення надають змогу незалежно масштабувати підсистеми обробки транзакцій та команд відповідно до специфічних вимог архітектури програмного забезпечення.

Список використаних літературних джерел

- [1] Ігор Цимбалюк, Юрій Грицюк, (Львів, 21-23 листопада 2024 р.). Застосування XOR фільтрів для агрегації та валідації транзакцій у направлених ациклічних графах. Комп'ютерне моделювання та інформаційні технології: матеріали шостої науково-практичної конференції студентів, аспірантів та молодих вчених. – Львів: ННІ КНІТ НЛТУ України, 2024. – 298 с.
- [2] R. Subramanyam, I. Gupta, L. M. Leslie and W. Wang, "Idempotent Distributed Counters Using a Forgetful Bloom Filter," 2015 International Conference on Cloud and Autonomic Computing, Boston, MA, USA, 2015, pp. 113-124, <https://doi.org/10.1109/ICCAC.2015.30>.
- [3] G. Liao, H. Ding, C. Zhong and Y. Lei, "RT-DAG: DAG-Based Blockchain Supporting Real-Time Transactions," in IEEE Internet of Things Journal, vol. 11, no. 20, pp. 32759-32772, 15 Oct.15, 2024, <https://doi.org/10.1109/JIOT.2024.3409025>.

Methods and tools for incremental data flow mutation in distributed systems architecture

Igor Tsymbalyuk

The principles of representing events in distributed systems as state management entities using the structure of a directed acyclic graph and probabilistic filters are described. The concept of meta-mutations is introduced to modify the conditions and properties of the environment in which message transit occurs. The effectiveness is assessed in comparison with standard approaches to solving similar problems. The research results can be used to create systems with a complex hierarchy of information access rights and data management with integrated version control mechanisms and record-level security policies.