

Архитектурный подход к проектированию распределенных транзакционных систем хранения и обработки данных на основе универсального ядра

Бакин А.О. (Университет ИТМО)

Научный руководитель – доцент, кандидат технических наук Перл И.А.
(Университет ИТМО)

Введение. Системы хранения и обработки данных являются неотъемлемой частью многих информационных систем. Исторически в качестве основных систем хранения выступают классические транзакционные SQL базы данных, такие как PostgreSQL. Однако, рост объемов хранимых данных вынуждает производить поиск новых систем хранения, ведь классические базы данных, зачастую, ограничены в горизонтальном масштабировании. Проблема масштабирования решается отдельным классом NoSQL баз данных: такие базы данных обеспечивают горизонтальное масштабирование, однако, ценой нарушения или полного несоблюдения ACID свойств транзакций. В стремлении совместить транзакционные гарантии SQL баз данных и возможности масштабирования NoSQL систем хранения в начале 2010-х появился новый класс систем – NewSQL базы данных [1][2].

Благодаря использованию иных, по сравнению с классическими SQL базами, архитектурных решений NewSQL базы данных одновременно поддерживают горизонтальное масштабирование и соблюдают ACID свойства транзакций в полном объеме. Однако, несмотря на перспективность, NewSQL базы данных в основном ориентированы на хранение данных и ограничены в способах их обработки. Также многие из существующих NewSQL систем ограничены и в возможностях расширения.

Помимо инструментов горизонтального масштабирования NoSQL базы данных также привнесли в индустрию и модели хранения отличные от реляционной. Зачастую, под каждую модель хранения существует свой подвид баз данных: графовые, документоориентированные, базы данных временных рядов, базы данных ключ-значение и так далее. При этом, если, в рамках единой информационной системы используется несколько моделей хранения, то потребуются развернуть несколько баз данных, что увеличивает накладные расходы.

Основная часть. В рамках исследования был сформулирован ряд требований для перспективных систем хранения и обработки данных, которые помогут преодолеть выявленные у существующих систем проблемы:

1. поддержка горизонтального масштабирования с соблюдением ACID свойств транзакций;
2. поддержка мультимодельного способа хранения данных в рамках единой системы;
3. возможности управления способом физического размещения данных с целью повышения производительности;
4. возможность обработки данных в рамках той же системы хранения;
5. наличие точек расширения для поддержки широкого класса задач.

В работе произведен обзор существующих решений класса NewSQL и показана невыполнимость всех требований в совокупности данными системами хранения и обработки данных, что дополнительно подчеркивает актуальность работы.

Для выполнения указанных требований предложена концептуальная архитектурная модель системы хранения и обработки данных, основанная на выделении универсального транзакционного ядра и адаптеров для каждой модели данных. Указанная архитектура позволяет обеспечить масштабирование, мультимодельность и расширяемость системы хранения.

Выбор программного интерфейса ядра осуществлялся между 3 моделями хранения: реляционная, ключ-значение и документная. По итогам сравнения было принято решение использовать модель типа «ключ-значение», как основу интерфейса ядра. К базовым операциям `get(key)`, `put(key)` и `delete(key)` были добавлены также специфические операции, позволяющие реализовать транзакционную логику и версионирование значений.

Выводы. Были сформулированы требования к перспективным системам хранения и обработки данных и показана их невыполнимость в совокупности ни одном из рассмотренных решений. Предложена концептуальная архитектура перспективной системы хранения и обработки данных, основанная на выделении универсального транзакционного ядра. В качестве программного интерфейса ядра была выбрана модель хранения типа «ключ-значение».

Список использованных источников:

1. Pavlo A., Aslett M. What's really new with NewSQL? //ACM Sigmod Record. – 2016. – Т. 45. – №. 2. – С. 45-55.
2. Khasawneh T. N., AL-Sahlee M. H., Safia A. A. Sql, newsql, and nosql databases: A comparative survey //2020 11th International Conference on Information and Communication Systems (ICICS). – IEEE, 2020. – С. 013-021.