

УДК 004.75

РАЗРАБОТКА РАСПРЕДЕЛЕННОЙ ВЫСОКОНАГРУЖЕННОЙ СИСТЕМЫ ТЕГИРОВАНИЯ

Коханов А.И. (Университет ИТМО)

Научный руководитель – д.т.н., профессор Шалыто А.А.
(Университет ИТМО)

Сервис тегирования — инструмент, предназначенный для разметки разнообразных сущностей в сервисе такси: водителей, пользователей, таксопарков, машин. Любой другой сервис, компонент такси, может получать теги с дополнительной информацией о его поставщике, времени появления и времени жизни. Процесс тегирования работает с учетом высокой нагрузки от других сервисов, которым важна надежность и безотказность нашего сервиса. Для стабильной и быстрой работы сервиса реализован кеш над базой данных, который инкрементально обновляется при обновлении тегов.

Введение. Сущность сервиса - это идентификатор какого-то значимого объекта, к которому можно привязать тег. Поставщик - идентификатор источника тегов, каждый из которых может независимо от других редактировать теги. Сервис должен отдавать свои теги по HTTP запросу с учетом высокой нагрузки от различных потребителей.

Основная часть. Для решения продуктовых задач необходимо иметь возможность быстро проставить нужные теги и быстро увидеть изменения во всей системе, что приводит к возникновению очень частых запросов на чтение тегов. В реляционной базе данных PostgreSQL хранится основная информация о тегах: она позволяет выполнить проверку разрешений на редактирование тегов в рамках одной транзакции. Также данная база данных справляется с нагрузкой на изменение данных. Но при таком способе хранения тегов нагрузка на чтение информации из базы окажется слишком большой. Для обеспечения значительно более быстрого доступа к данным может помочь использование кеширования данных – они будут храниться в оперативной памяти сервиса, что избавляет его от более медленного обращения к базе данных.

Для реализации кеширования была изменена схема базы данных, чтобы кеш мог получать обновленные данные. Была добавлена новая колонка revision с типом BIGSERIAL, которая автоматически получает самое большое значение при любом изменении данных. Кеш знает максимальную ревизию своих тегов, и при обновлении запрашивает данные с большей ревизией. При этом важно при любом обновлении данных обновлять ревизию существующих тегов, чтобы кеш смог узнать об этих изменениях.

Выводы. Данный способ кеширования позволил сервису выдержать высокую нагрузку на чтение с учетом его распределенной архитектуры – у каждого отдельного инстанса сервиса свой локальный кеш периодически обновляется на основе информации из базы данных PostgreSQL. Данная схема кеширования обеспечивает целостность данных при обнаружении проблем на стороне базы данных, а также стабильно работает при возникновении резкого роста чтения данных при периодических проблемах на стороне потребителей тегов. Результаты работы были внедрены в рамках разработки системы кеширования данных для сервиса тегирования.

Коханов А.И. (автор)

Шалыто А.А. (научный руководитель)