

ОБЛАЧНАЯ НЕРЕЛЯЦИОННАЯ БАЗА ДАННЫХ AMAZON DYNAMO DB: ПРОСТО И УДОБНО

Максим ЧУКИТУ, Валерия КОРОСТИНСКАЯ

Технический Университет Молдовы, Департамент Программной Инженерии и Автоматики

Аннотация: Данная работа посвящена нереляционной *serverless* базе данных Amazon DynamoDB, предоставляемой компанией Amazon Web Services, её ключевым особенностям, истории создания и становления одной из самых популярных *serverless* баз данных.

Ключевые слова: база данных, NoSQL, DynamoDB, Amazon, Web Services, облако, cloud, serverless, масштабируемость.

Введение

В мире информационных технологий всегда существовала потребность в записи, хранении и чтении информации. Эта потребность привела к появлению и развитию множества различных типов физических накопителей, файловых систем и, конечно же, баз данных.

Существует множество различных решений по хранению информации на любой вкус и цвет: реляционные и нереляционные, SQL и NoSQL, но все они обладают одним существенным недостатком: необходимостью сервера для размещения, который также нуждается в обслуживании и улучшении при достижении пика производительности. Хотя, вернее сказать, обладали, ведь и эта проблема была решена с появлением облачных технологий. Но обо всём по порядку. И, пожалуй, стоит начать с того, что Amazon Web Services.

1. Amazon Web Services

Amazon Web Services (AWS) является дочерней компанией Amazon.com, предоставляющей платформу облачных вычислений частным лицам, компаниям и правительствам на платной подписке. Эта платформа предоставляет пользователям в своё распоряжение виртуальный кластер компьютеров, доступный постоянно через Интернет.

Amazon Web Services в значительной степени (наряду с Google Cloud Platform) повлияли на формирование концепции облачных вычислений в целом, и определили основные направления развития публичной модели развёртывания. Длительное время Amazon Web Services было крупнейшим в мире по выручке публичным облаком, во второй половине 2010-х годов уступив по этому показателю Azure от Microsoft, при этом сохраняя доминирование в сегментах инфраструктурных и платформенных услуг. По состоянию на 2017 год годовая выручка от услуг AWS превысила \$20 млрд, что составило около 11,5 % доходов Amazon.

Ввиду огромного количества все сервисы разбиты на семейства сервисов. Одним из сервисов, предоставляемых платформой AWS, является Amazon DynamoDB.

2. Amazon DynamoDB

Нереляционная NoSQL база данных Amazon DynamoDB была представлена в середине января 2012 года. Ключевой особенностью DynamoDB, делающей её столь привлекательной является *serverless* подход, который позволяет забыть о проблемах масштабирования базы данных, необходимости обслуживания и настройки сервера и о других неудобствах, присущих серверным базам данных.

Как заявляет Jeff Wang в статье посвящённой релизу новой базы данных, DynamoDB позволяет пользователю хранить такой объём данных, который ему необходим, и получать к ним доступ так часто, как ему необходимо, с предсказуемой производительностью, обеспечиваемой использованием твердотельного диска, более известного как SSD.

DynamoDB работает на основе “предоставленной пропускной способности” (provisioned throughput). Пользователю достаточно указать необходимую пропускную способность для записи и чтения (отдельно друг от друга) при создании таблицы DynamoDB. За кулисами же всё будет настроено таким образом, чтобы удовлетворить потребности пользователя, сохраняя при этом задержку в миллисекундах. Позже, при росте потребностей пользователя, он может просто увеличить выделенную пропускную способность (или наоборот уменьшить). Данное изменение можно

произвести онлайн, без простоев и без влияния на общую пропускную способность. Другими словами, пользователь имеет возможность изменять масштаб, даже когда база данных обрабатывает запросы.

Однако многие рабочие нагрузки имеют циклический характер или их сложно предсказать заранее. Например, приложение для социальных сетей, в котором большинство пользователей активны в дневное время. База данных должна быть в состоянии обрабатывать дневную активность, но нет необходимости в одинаковых уровнях пропускной способности ночью. Такие виды рабочих нагрузок часто требуют ручного вмешательства для увеличения или уменьшения ресурсов базы данных в ответ на различные уровни использования.

Сервис Amazon DynamoDB обладает опцией автоматического масштабирования для динамической настройки выделенной пропускной способности от вашего имени в ответ на фактические схемы трафика. Это позволяет увеличить свою подготовленную емкость для чтения и записи, чтобы обрабатывать внезапные увеличения трафика без регулирования. Когда рабочая нагрузка уменьшается, автоматическое масштабирование приложения уменьшает пропускную способность, поэтому пользователь не платит за неиспользованную выделенную емкость.

Другим немаловажным достоинством является то, что Amazon DynamoDB смехотворно проста в использовании. Только что созданные таблицы обычно готовы к использованию уже через одну или две минуты. Как только таблица будет готова, пользователь просто начинает хранить в ней необходимые ему данные в необходимом ему количестве, оплачивая при этом только используемое хранилище (нет необходимости предварительно выделять хранилище). За кадром же позаботятся о том, чтобы предоставить достаточно места для хранения.

Каждая таблица должна иметь первичный индекс. Существует два типа первичных ключей: простые хэш-ключи и составные хэш-ключи с диапазонными ключами.

Простые хэш-ключи создают абстракцию распределенной хэш-таблицы (Distributed Hash Table) и используются для индексации уникального ключа. Ключ хэшируется по нескольким разделам обработки и хранения для оптимального распределения рабочей нагрузки.

Составные хэш-ключи с диапазонными ключами предоставляют возможность создания первичного ключа, который состоит из двух атрибутов - атрибута hash и атрибута range. Когда вы выполняете запрос к ключу этого типа, атрибут hash должен соответствовать уникально, но для атрибута range можно указать диапазон (от низкого до высокого). Такой подход может быть использован для выполнения запросов, таких как «все заказы от Джеффа за последние 24 часа».

Каждый элемент в таблице DynamoDB состоит из набора пар ключ/значение. Каждое значение может быть строкой, числом, набором строк или набором чисел.

Выводы

Вне всяких сомнений Amazon DynamoDB является крайне привлекательным решением для хранения данных. Данный сервис невероятно прост в использовании, решает проблемы масштабирования и при этом является экономически выгодным для пользователя за счёт автоматического увеличения и уменьшения выделенной пропускной способности в зависимости от фактической нагрузки. Многие известные компании такие как Nike Digital, Samsung, Snap Inc., Netflix, Tinder и Amazon отдали предпочтение именно этому сервису. Возможно, и вам стоит попробовать?

Библиография

1. Jeff Barr, «Amazon DynamoDB – Internet-Scale Data Storage the NoSQL Way».- [Электронный ресурс]. – Режим доступа: <https://aws.amazon.com/ru/blogs/aws/amazon-dynamodb-internet-scale-data-storage-the-nosql-way/>
2. Amazon DynamoDB Developer Guide (API Version 2012-08-10), «Managing Throughput Capacity Automatically with DynamoDB Auto Scaling».- [Электронный ресурс]. – Режим доступа: https://docs.aws.amazon.com/en_us/amazondynamodb/latest/developerguide/AutoScaling.html
3. Amazon DynamoDB. Быстрый и гибкий сервис баз данных NoSQL для любого масштаба.- [Электронный ресурс]. – Режим доступа: <https://aws.amazon.com/ru/dynamodb/>
4. Jeff Barr, «Announcing BatchWriteItem for DynamoDB».- [Электронный ресурс]. – Режим доступа: <https://aws.amazon.com/ru/blogs/aws/announcing-batchwriteitem-for-dynamodb/>