

13 мая 2019

Стратегия данных. Время переосмысления?

Прошло уже 18 лет с тех пор, как магические 3V (Volume-объем, Variety-разнообразие и Velocity-скорость) впервые были озвучены Дагом Лэйни (Doug Laney) для описания эволюции организации данных, пишет Марк Торр в своей статье про Data Strategy. Представляем перевод этой статьи с комментариями руководителя проектов по Big Data ГК «КОРУС Консалтинг» Егора Старостина.

По мнению Марка Торра, сначала ИТ-директорам нужно было подготовиться к новой волне данных, быть готовыми хранить и получать больше данных, которые могут отличаться от традиционных данных из операционных систем (например, текстовых неструктурированных данных), а также быстрее обрабатывать такое количество информации. Это было в эру интернета задолго до того, как появились мобильные данные и социальные сети.

Потом весь мир столкнулся со штормом больших данных, когда все эти 3V стали больше, быстрее и разнообразнее. Появились социальные медиа и большее количество внешних данных, которые начали использоваться для принятия бизнес-решений. В наше время несмотря на то, что 3V никуда не



делись, мы стали слышать о них всё реже.

Растущая экосистема баз данных

Чтобы работать с большими данными, нам необходимы новые способы хранения. Это привело к появлению новой экосистемы баз данных для удовлетворения различных потребностей. Создавались новые модели/схемы баз данных с новым подходом к выполнению запросов.

Обратите внимание, например, на быстрый и запутанный рост использования озер данных (Data lake). Сначала активно рекламировались озера данных только на Hadoop-системах, а затем оказалось, что многие компании научились использовать для этого NoSQL-базы, и появились альтернативные сценарии, о которых ранее никто не предполагал.

Это всё только данные

Сегодня, я думаю, многие организации перестали думать о «больших данных». Сейчас в фокусе только те данные, которые нужны для удовлетворения текущих бизнес-требований. Важно отметить, что многие организации переносят обсуждение только на то, как получить выгоду из этого самого ценного актива. Неслучайно Даг Лейни теперь больше сосредоточен на [инфоэкономике](#), монетизации и получении ценности из имеющихся у вас данных, а не на 3V. Замечательно, что основное внимание уделяется получению ценности из данных, а не росту использования внешних данных. Этот процесс должен ускориться, поскольку это то, что будет поддерживать жизнь организаций.

И тут возникает вопрос: может ли так оказаться, что бизнесу не хватает определенной сноровки, когда дело доходит до упорядочивания и упрощения их существующих баз данных? Ответ — да!

Сложность проблемы

Быстрая эволюция бизнес-требований привела к тому, что организации оказались в невероятно сложном ландшафте данных. Многие организации значительно перерасходуют средства на управление этим сложным раздутым ландшафтом данных. Эта сложность является серьезной проблемой, например, в контексте регулирования закона о персональных данных.

В организациях есть много таблиц и множество баз данных. Кроме того, они часто имеют огромное разнообразие баз данных, включая табличные реляционные базы данных, электронные таблицы, базы данных NoSQL и так далее. Организации достигли этой точки, потому что они должны были удовлетворять потребностям бизнеса. Существующие базы данных не могли поддерживать то, что им было нужно и когда это было нужно.

Вопрос: это всё ещё так?

Мантра упрощения

Практика показывает, что все это осталось. Но сейчас настало время для изменения подхода. Подумайте об упрощении ландшафта данных, продолжая удовлетворять потребности бизнеса сегодня и в будущем.

Это обновление поможет с затратами и управляемостью, и это так же поможет, например, с законом о персональных данных. Благодаря

уменьшению сложности в источниках данных станет легче получать из них выгоду вместо того, чтобы пытаться получить бизнес-эффект из хаоса.

Переход на следующую ступень

Прогресс в базах данных был почти таким же неустанным, как и в других областях. Я собираюсь показать это на примере Microsoft, но это касается не только Microsoft.

Просто подумайте о таких вещах, как, например, [Microsoft CosmosDB](#), которая может устранить необходимость в различных базах данных NoSQL и предоставить вам мощные соглашения об уровне обслуживания (SLA), на которых вы можете вести свой бизнес. Она предоставляет несколько моделей баз данных и возможность использования нескольких стилей запросов одновременно, без необходимости переписывать приложения. Представьте себе, что вы можете перейти с четырех разных баз данных NoSQL на одну. Это изменит сложность? Безусловно.

Подумайте о том, что, например, SQL Server теперь может работать в Linux. Это заставит вас задуматься о том, действительно ли СУБД с открытым исходным кодом лучше СУБД корпоративного уровня, когда безопасность и надежность данных лежат в основе всего, что вы делаете?

Обратите внимание на тот факт, что, например, в SQL Server доступна обработка графов и реализация методов машинного обучения на языках Python и R. Изменит ли это необходимость создания отдельных витрин данных для аналитики, уменьшая сложность и разрастание данных? Безусловно.

Новые варианты использования

Наконец, давайте посмотрим на новые варианты развертывания:

- Гибкие настройки, которые позволяют вам пошагово и постепенно переходить в облако.
- Возможность беспрепятственного перемещения всего, что у вас есть, в облачное хранилище, что позволяет снизить накладные расходы на управление аппаратным обеспечением и справиться со всеми капитальными затратами на обновление ПО.
- Возможность использования управляемых сервисов в облаке с мощными соглашениями об уровне обслуживания (SLA) для сокращения накладных расходов на администрирование и одновременное использование новых режимов хранения данных.
- Возможность создания гибридных решений с использованием только тех облачных решений, которые действительно необходимы.
- Возможность внедрять то, что вы хотите, когда вы этого хотите, и использовать надежные соглашения об уровне обслуживания (SLA).

Подобные системы охватывают все режимы развертывания, предлагают практически все типы баз данных, которые могут вам понадобиться, и помогают найти вам «правильные» базы данных, варианты для упрощения, консолидации, модернизации для обеспечения гибкости в вашем ландшафте данных. И всё это без ущерба для бизнеса компании.

Движение вперед

Стремительный рост возможностей современных баз данных и возможностей их развертывания ошеломляет. И несмотря на это, многие организации вовсе не занимаются изучением этого вопроса. Предыдущие решения, внедренные даже 1,5 — 2 года назад, теперь часто могут быть рассматриваться как устаревшие.

Известный факт, что прогрессивные организации заняты документированием своих активов данных. Большинство из них сосредоточены на том, чтобы определить, где какие данные находятся, как их обезопасить и обеспечить их надлежащее использование.

Необходимо также думать о том, может ли миграция и/или консолидация сделать жизнь намного проще. Обязательно подумайте о своем ландшафте данных и подумайте, какие есть пути для его развития.

Вот несколько вопросов:

- 1.** Вы недавно смотрели, **где** вы храните свои данные, и понимаете ли вы, **почему** они там хранятся? Думали ли Вы о том, что нужно поменять?
- 2.** Знаете ли вы, **сколько вам стоит** управлять и поддерживать свои данные и как можно уменьшить сложность их хранения? Если на вашем радаре есть снижение ИТ-расходов, это верный способ найти способ сделать это.
- 3.** Задумывались ли вы о том, будет ли легче соблюдать требования по отношению к закону о персональных данных, если ландшафт ваших СУБД будут проще? Является ли консолидация базы данных одним из ключевых вариантов развития? Если нет, то почему?
- 4.** Когда вы в последний раз оценивали, какие базы данных должны быть локальными, какие можно развернуть в гибридном режиме и какие можно

полностью перенести в облако? Возможно, вы несете дополнительные расходы, которые можно сократить.

Современные принципы и подходы хранения данных шагнули вперед, и нам всем нужно четко понимать, какие есть инструменты для правильной организации данных в компании и как эти данные эффективно использовать. Эту задачу можно выполнять самостоятельно, либо с привлечением опытных партнеров. Здесь самое главное, на мой взгляд, делать, а не стоять на месте. И тогда постепенно, шаг за шагом, наши данные помогут нам зарабатывать больше и работать более эффективно.

Источник: Global CIO