NewSQL – новый шаг в развитие BigData

И.М. Забродин,
студент;
рук. **П.Г. Круг**
(МИРЭА, МГУПИ, МИТХТ, Москва)
E-mail: zabrnew@mail.ru

Помимо SQL и NoSQL есть перспективная альтернатива — NewSQL. Это такие БД, где совместили современные подходы распределенных систем NoSQL, реляционную модель представления данных и язык SQL. Этот термин предложил Мэтью Аслет в 2011 году. Необходимость таких БД появилась у фирм, которые обрабатывают критические данные. Им были нужны масштабируемые системы, а NoSQL не мог проводить транзакций, что сказывалось на достоверности и устойчивости данных. Эти БД появились совсем недавно, буквально 4-6 лет назад, но сейчас активно конкурируют на рынке за пользователей, перевешивая укоренившиеся SQL и относительного новичка — NoSQL.

Когда говорят о NewSQL, имеются ввиду такие БД, которые удовлетворяют следующим условиям:

1. Поддержка реляционной модели и транзакционности;
2. Использование языка SQL для получения данных;
3. Горизонтальная масштабируемость;
4. Современное ядро, не перенятое от типичных СУБД;
5. Представлены на рынке лишь 4-6 лет.

Существует несколько групп в области хранения и обработки данных, которые можно сопоставить с NewSQL.

Одна из них подразумевает применять кластеры SQL из нескольких физических узлов для хранения и обработки больших объемов данных. Хоть такой подход и оставляет БД OLTP в своем первоначальном виде, но позволяет достичь масштабируемости и горизонтальную группировку. Все производственные СУБД объединяют в кластер. Они являются средним слоем (middleware), который перенаправляет запросы к нужным узлам (где хранятся данные), группирует данные и выдают единый результат. При таком подходе, для запросов существуют ограничения по использованию внешних ключей и джоинов.

Вторая группа NewSQL — это ядра последнего поколения, которые используются для хранения и доступа к данным в БД. Некоторые движки были созданы на основе MySQL, чтобы решить проблему отсутствия масштабируемости. Один из самых распространённых — TokuDB. Он использует индексы, расположенные на фрактальных деревьях. Они намного быстрее, чем индексы на B-деревьях, особенно при переполнении оперативной памяти, что вызывает необходимость считывать их с жёсткого диска.

Таблица 1 — сравнение NewSQL, SQL и NoSQL.

Характеристика	**SQL**	**NoSQL**	**NewSQL**
Реляционная	Да	Нет	Да

Поддержка языка SQL	Да	Нет	Да
ACID поддержка транзакций.	Да	Нет	Да
Горизонтальная масштабируемость	Нет	Да	Да
Высокая производительность	Нет	Да	Да
Механизм управления блокировками	Да	Нет	Да

Третья группа NewSQL — это полностью новые СУБД, в которых реализована поддержка SQL.

Сравнение функциональности NewSQL по сравнению с SQL и NoSQL представлено в таблице 1.

Основная масса БД NewSQL является in-memory БД. Главной особенностью таких БД является то, что они хранят основную массу данных не на жёстких дисках, а в оперативной памяти. Разработчики уверены, что хранение данных на HDD, является одним из основных недостатков существующих БД, поскольку, самую высокую скорость обработки и получение данных обеспечит помещение их в ОЗУ. Недостаток ОЗУ на одном сервере, компенсируется кластером таких серверов из большого количества узлов. Так же в таких БД предусмотрено периодическое резервное копирование данных на жёсткий диск, чтобы предотвратить их потерю при отказе сервера.

Некоторые программисты NewSQL ставят перед собой задачу решить проблемы развёртывания современных систем управления БД в облачных или виртуальных средах. Ещё есть группа тех, кто пытается создать кроссплатформенное средство для хранения и обработки данных в виде реляционной модели, основанное на принципах NewSQL.

NewSQL необходим предприятиям, которые планируют:

1. Адаптация текущих приложения под современные нормы роста объема данных;

2. Создание новых приложений на OLTP системах.

Специалисты по NewSQL считают, что устоявшийся SQL в изначальном виде — устарел, обладает большим количеством проблем и неоправданно усложнён. Тем более объектно-ориентированные СУБД уже давно общеприменимы. Для упрощения перехода на новые технологии, будут написаны конвертеры SQL в NewSQL и NewSQL в SQL, которые смогут переделывать запросы во время обработки. Таким образом, старые программы и оболочки останутся в рабочем состоянии, при использовании новых современных технологий хранения данных.

Список использованных источников

1. Викпедия. Свободная энциклопедия. [электронный ресурс] URL: https://ru.wikipedia.org/wiki/NewSQL — NewSQL. Дата обращения 19.12.2016.
2. Хабрахабр. Ресурс для IT-специалистов. [электронный ресурс] URL: https://habrahabr.ru/post/154589/ — NewSQL — новый виток в эволюции Big Data, забираем лучшее из SQL и NoSQL. Дата обращения 19.12.2016.
3. VoltDB. With the world’s fastest, in-memory operational database. [электронный ресурс] URL: https://www.voltdb.com/nosql-vs-sql — NoSQL vs SQL vs NewSQL. Дата обращения 19.12.2016.
4. Тут блог. Общественные обязательства интроверта. Сообщения на ИТ тематику, но не обязательно. [электронный ресурс] URL: http://blog.gelin.ru/2015/01/newsql.html — О NewSQL. Дата обращения 19.12.2016.