

Д. Е. Намиот, канд. физ.-мат. наук, МГУ имени М. В. Ломоносова, г. Москва, dnamiot@gmail.com

Базы данных временных рядов в системах «Интернета вещей»

В статье рассматриваются вопросы, связанные с базами данных для временных рядов. Временные ряды достаточно широко используются в различных приложениях. Соответственно, тема хранения временных рядов является весьма актуальной. Естественно, при определенных условиях, которые представлены в этой статье, можно использовать традиционные реляционные базы данных. Возможные ограничения будут, с одной стороны, связаны со скоростью поступления новых данных. С другой стороны, возможные ограничения, а также особенности обработки временных рядов приводят к необходимости использовать специализированные системы хранения временных рядов. Именно описанию работы с временными рядами с точки зрения баз данных и посвящена статья. В ней рассматриваются вопросы использования традиционных реляционных баз данных для хранения временных рядов, а также решения, классифицируемые как NoSQL-системы.

Ключевые слова: временные ряды, базы данных, NoSQL, «Интернет вещей».

Введение

В настоящей статье представлен расширенный и дополненный вариант доклада, сделанного на конференции [1].

Классически временной ряд представляет собой последовательность данных, обычно соответствующих каким-либо измерениям на некотором временном интервале. Подобного рода модели встречаются в самых разных прикладных областях. Самый простой пример — это системы для межмашинного взаимодействия (M2M) и «Интернет вещей» (Internet of Things — IoT). Любые сенсорные измерения и формируют временной ряд. В силу общности задача обработки временных рядов является одной из традиционных (а равно — и одной из старейших) задач для баз данных.

Рассмотрим традиционные реляционные базы данных. На первый взгляд, проблема хранения временных рядов выглядит просто. Можно создать таблицу, указав столбец со временем в качестве ключа. Остальные столбцы

будут представлять измеренные в определенный момент значения. Каждое новое измерение просто добавляет строку к таблице. Одно измерение (один отчет) — одна строка. Естественно, в любой практической системе количество измеряемых значений конечно. Каждый сенсор в IoT-системе имеет определенное число измеряемых значений. Но на практике в заданный момент будем иметь, скорее всего, только часть из потенциально возможных атрибутов. Устройства в системе чаще всего асинхронны, работают по собственным циклам, со своими временными интервалами обновления данных. Это означает, что каждая строка в таблице будет содержать только часть из потенциально возможных атрибутов. Остальные атрибуты будут пустыми (null value), что ведет к неэффективному расходованию дискового пространства и затрудняет последующую обработку.

Если обратиться к операциям с базой данных для временных рядов, то станет очевидно, что операция добавления (INSERT) будет превалировать. Обновление и удале-