

Мифы об архиваторах (Historians)

Элиот Миддлтон /Eliott Middleton/,
старший менеджер по выпуску продукции,
Schneider Electric

Краткое содержание

Реляционные базы данных идеально подходят для многих приложений, но являются далеко не лучшим решением при работе с временными рядами. Системы с высокой пропускной способностью и архиваторы данных производственных процессов – не одно и то же. Реляционные базы данных применимы не для всех задач. Хотя базы данных Microsoft SQL Server или Oracle и способны хранить данные временных рядов, существует ряд серьезных проблем, которые необходимо учитывать при рассмотрении их использования для промышленных приложений и архивации данных о процессах. Это техническое описание рассматривает основные причины, по которым архиваторы лучше подходит для сбора, хранения и поиска производственной информации по сравнению с реляционными базами данных.

Введение

Реляционные базы данных идеально подходят для многих приложений, но являются далеко не лучшим решением при работе с временными рядами. Приложения для архивации данных производственных процессов появились достаточно давно и за годы существования обросли бессодержательными и бездоказательными заблуждениями и «мифами». В данном техническом описании будут рассмотрены достоинства и недостатки реляционных баз данных, а также раскрыты и оспорены эти мифы в контексте специализированных приложений архивации процесса. Это описание также представит некоторые основы архиваторных приложений, объяснит некоторые важные моменты для объективной оценки архиваторных решений и соответственно установит, какова возможная польза архиватора.

Системы с высокой пропускной способностью и архиваторы данных производственных процессов – не одно и то же. Хотя в качестве примера использования реляционных баз данных можно привести торговые системы нью-йоркской фондовой биржи и другие приложения с высокой пропускной способностью, нельзя сказать, что реляционные базы данных применимы для всех задач. Хотя базы данных Microsoft SQL Server или Oracle и способны хранить данные временных рядов, есть некоторые очень существенные проблемы, которые должны быть приняты во внимание при выборе их использования для архивации данных процесса и промышленных приложений.

Миф № 1: Хранение настолько дешево, что эффективность не имеет значения

Общепринятым является значительное преуменьшение того, сколько данных на самом деле создает типовой процесс. Небольшой архиватор на 5000 тегов, который регистрирует данные каждую секунду, создает 157 миллиардов различных значений в год. Даже при эффективном хранении в 8 байтов на значение необходимый объем хранения равнялся бы примерно 1 терабайту в год. В некоторых тестах, когда сравнивались требования к хранению данных временных рядов для SQL Server и для Wonderware Historian, разница была 50:1. Хотя цены на хранение падают, 50 терабайт данных в год - это по-прежнему много. Также необходимо учитывать, что это не просто вопрос наличия достаточного места на диске для хранения такого количества данных; зачастую также требуется, чтобы архивированные данные были хорошо защищены, что увеличивает объем хранения, необходимый для резервного копирования или создания зеркал дисков. Некоторые производственные отрасли имеют нормативные требования по сохранению данных в течение нескольких лет, что еще больше увеличивает необходимый объем хранения.

Миф № 2: Реляционные базы данных достаточно быстрые.

Улучшение соотношения цена/качество для аппаратного обеспечения благоприятно сказалось на работе реляционных баз данных. Однако, в реляционных базах данных должна обеспечиваться ссылочная целостность при «транзакциях», которые могут синхронно обновлять несколько связанных табличных значений, и это значительно увеличивает нагрузку.

Например, на высокопроизводительных аппаратных средствах (с использованием процессоров 64 Itanium) SQL-сервер 2008 установил мировой рекорд, выполняя 1126 транзакций в секунду. Даже с учетом того факта, что это не те транзакции, которые использует архиватор, сохранение 5000 значений в секунду явилось бы обременительной задачей, если бы запись каждого значения была транзакцией, даже на таких высокопроизводительных аппаратных средствах. Это означает, что клиентские приложения (отвечающие за сбор данных) должны буферизировать данные и передавать множество значений в базу единой транзакцией. Базы данных без полной поддержки транзакций, такие, как входящая в MySQL система хранения данных MyISAM, могут поддерживать большую пропускную способность, но также требуют буферизации на стороне клиентских приложений.

Понятно, что хранение данных осуществляется, прежде всего, для их последующего извлечения, и поэтому производительность извлечения данных также очень важна. В решениях общего назначения, таких, как реляционные базы данных, можно организовать либо эффективное хранение (более высокая пропускная способность), либо эффективное извлечение (быстрый поиск), но не оба механизма сразу. Эффективное извлечение данных требует использования «кластеризованного индекса», недоступного, например, в системе хранения MyISAM с более высокой пропускной способностью.

Напротив, специализированные системы хранения, разработанные специально для данных временных рядов, обладают возможностью и эффективного сбора, и эффективного извлечения, что невозможно для обобщенных данных.

Миф № 3: Использование реляционной базы данных в качестве архиватора – новое откровение

Новые идеи для применения существующих технологий появляются достаточно часто – маленькие желтые листочки

для заметок 3M Post-It® были известным применением для существующего не-такого-уж-и-клейкого клея. Тем не менее, компании более десяти лет пытались использовать простую схему реляционной базы данных в качестве замены специально разработанного решения для данных временных рядов. Несмотря на эту «новую идею» использования реляционной базы данных, рынок специализированных решений архивации продолжает заметно расти.

Миф № 4: SQL можно использовать для запроса данных только в реляционных базах данных

Хотя реляционные базы данных и имеют множество преимуществ по сравнению с альтернативными технологиями, главным причиной их популярности стал SQL. SQL стандартизировал и коренным образом изменил способ извлечения данных для пользователей - от сложных упражнений по программированию к относительно простому и гибкому языку описания интересующих данных.

К счастью, SQL был адаптирован к нереляционным хранилищам данных, и это позволило использовать преимущества SQL за пределами ограничений, присущих реляционным базам данных.

Миф № 5: В данных временных рядов нет ничего особенного

Мощные средства SQL для извлечения данных дают некоторым основания утверждать, что реляционные базы данных одинаково хороши для извлечения данных временных рядов и для транзакционных данных. Конечно, верно, что SQL дает большую гибкость, но эти утверждения базируются на некоторых фундаментальных предположениях, которые не относятся к данным временных рядов: (а) нет внутреннего порядка в записях данных (а на самом деле все данные временных рядов упорядочены по времени); (б) все данные сохраняются в явном виде (а на самом деле большинство архивированных данных представляют собой только выборки из континуума реальных данных); все данные равнозначны.

Эти два различия являются значительными. Например, если датчик передает значение с меткой времени «7:59:58,603», а пользователь запрашивает из реляционной базы значение для времени «8 00:00,000», то в этом случае он не получит никаких данных, так как в базе нет никаких записей именно для этого времени – база данных не признает время как континуум.

Аналогичным образом, если температура была «21°C», а две минуты спустя стала «23°C», то реляционная база не сможет сделать оценку (интерполяцию) процесса таким образом, чтобы посередине между этими выборками температура была примерно «22°C».

В случае стационарного неизменяющегося процесса все его значения не всегда бывают важны. Наиболее важны отклонения от среднего значения. Единственным способом найти такие отклонения для клиентского приложения реляционной базы является запрос всех данных для процесса, что может сильно нагрузить систему в целом: сервер, сеть и клиента.

В отличие от этого специализированные архиваторы, как правило, имеют средства фильтрации (основанные на сравнении последовательных записей), способные радикально сократить объем данных, который должен быть доставлен в клиентские приложения.

Миф № 6: Управление данными временных рядов в реляционной базе тривиально

Реляционные базы данных разработаны для накопления больших объемов данных. Однако по мере роста объема данных также растут и время выполнения запросов, и размер резервных копий, и количество других рутинных операции. Для решения проблемы производительности постоянно растущих таблиц администраторы должны регулярно проводить очистку базу данных.

В любой базе данных, которая защищает транзакционную целостность, такая очистка приостанавливает нормальную запись данных, что является проблемой для приложений архивации, работающих в режиме 24/7/365. Поэтому даже для того, чтобы сделать приемлемой такую операцию очистки, требуется свести к минимуму объем данных, хранящихся в базе.

В случае если такие удаленные данные будут необходимы позднее (например, для аудита, или согласно некоторым нормативным требованиям), их восстановление является достаточно важным. Основная практика заключается либо в восстановлении полной резервной копии базы, которая включала в себя необходимые данные, в отдельную систему, выделенную для этой цели, либо использование вашей производственной системы в автономном режиме. Это становится еще более проблематичным, если необходимые данные не доступны в одной резервной копии, например, если вы сохраняете данные только за последние 30 дней в онлайн-базе данных, а аудиторская ревизия требует данные за 90 дней. В этом случае вы должны либо вручную объединить все данные в единой базе, имея три резервных копии, каждая из которых составляет изолированный интервал в 30 дней, либо изучить каждую резервную копию последовательно.

Специализированные архиваторы разработаны как для обработки быстрорастущих объемов данных, так и для работы с накопленными данными в офлайн и онлайн режимах.

Миф № 7: Для архивирования выбирают или полностью реляционные, или полностью проприетарные решения

Большинство решений для архивации используют либо полностью проприетарную технологию для снятия ограничений, присущих реляционной базе данных, либо полностью используют реляционную базу данных, чтобы уменьшить свои собственные затраты на инжиниринг. Архиватор Wonderware (Wonderware Historian) взял лучшее от обеих систем. Он использует единую реляционную схему для управления всеми относительно статическими данными конфигурации, но расширяет нативный транзакционный механизм хранения и процессор запросов SQL Server от корпорации Microsoft® проприетарными расширениями, снимая их ограничения для данных временных рядов.

В построенном на основе Microsoft SQL Server решении легче обеспечить безопасность и им легче управлять, чем полностью проприетарными решениями, но без ущерба для основных возможностей, требуемых от архиватора.

Миф № 8: В промышленных приложениях нет ничего особенного

Реальные архиваторы предоставляют возможности для взаимодействия с реалиями современного мира промышленных приложений, которые находятся за пределами области только лишь реляционных баз данных. Каким образом вы намерены использовать данные? Нужно ли вам для отчетности выполнять преобразование данных? Если да, то с помощью запроса SQL это достаточно сложно. Являются ли ваши измерительные устройства и сбор данных на 100 процентов надежными, или вам иногда приходится «уметь обходиться» данными, которые включают в себя инструментальные погрешности? В то время как базы данных общего назначения, безусловно, могут хранить данные, они не спроектированы для включения понятия «качество данных» в расчеты и не в состоянии легко выполнять рутинные вычисления временных рядов, подобные интегральным вычислениям, которые обычно необходимы.

Миф № 9: Приложения реляционных баз данных не являются архиваторами

«Архиватор» обращается к нескольким связанным функциям: непрерывный сбор данных в режиме реального времени, хранение заслуживающих внимания подмножеств этих данных и предоставление средств извлечения

содержательной информации из этих данных. В то время как «архиватор» описывает приложение, «реляционная база данных» указывает технологию. Хотя, конечно, есть некоторые существенные проблемы в использовании технологии реляционных баз для данных временных рядов, но использование приложением этой технологии не означает, что это не архиватор – такое приложение может быть архиватором, только с достаточно простым функционалом. Прежде чем сосредоточиться на выборе основной технологии (реляционные базы данных или проприетарные файлы), правильнее сфокусироваться на необходимой функциональности – что требует понимания полной функциональности приложения и включает в себя гораздо больше, чем просто хранение данных.

Миф № 10: Все данные одинаковы по важности и надежности

В реляционной базе данных хранимое значение в точности соответствует исходному значению и всегда предполагается валидным – если это не так, то его корректировка отдается на чье-то усмотрение. При сборе миллионов значений от тысяч точек сбора данных о процессах неизбежно, что некоторая информация некорректна или отсутствует. Могли быть получены значения вне диапазона из-за проблем с измерительным оборудованием, потеря связи, или данные были просто ошибочными.

В производственном архиве для каждого значения данных хранится не только само значение и метка времени, но и показатель качества данных. Например, сохранение значения параметра оборудования, находящегося за пределами нормального рабочего диапазона, вызовет изменение показателя качества. Эти показатели не просто отдельные колонки в базе данных, а неотъемлемое свойство данных. Их можно извлекать из базы, включать в вычисления и использовать для предупреждения операторов или технического персонала о потенциальной проблеме.

При вычислении суммарных значений (например, средней температуры за последний час), архиватор должен учитывать качество данных при вычислениях, иметь возможность отфильтровывать подозрительные данные и быть в состоянии экстраполировать данные в случае их отсутствия или недостоверности. Если случающиеся в реальности отклонения не обработаны корректно, то получаемые отчеты, интегрируемые в бизнес-системы, станут искаженными, и, следовательно, принятые решения будут неверными. Одними только собственными средствами реляционных баз такие возможности не обеспечиваются.

Миф № 11: Хранение данных в реляционной базе позволяет легко делать запросы

Хотя поддержка SQL запросов является одним из огромных преимуществ использования реляционных баз данных, это вовсе не означает, что с конкретной базой легко общаться посредством SQL запросов — некоторые проекты настолько запутаны, что не позволяют эффективно использовать написанные вручную запросы, и вместо этого приходится полагаться на сложные запросы, генерируемые программным образом.

Это, возможно, имеет смысл с точки зрения управления хранением и переносимости, но в значительной степени нейтрализует все преимущества использования реляционных баз данных.

Миф № 12: Хранение данных в реляционной базе делает их интегрированными

Перенос двух файлов Excel в одну и ту же папку не делает их «интегрированными» в любом смысле, даже если они оба включают в себя производственные данные. Аналогичным образом, получение данных от системы ERP (планирования ресурсов предприятия) и данных от архиватора и хранение их в реляционной базе не делает эти данные «интегрированными». Конечно, наличие всех этих данных в одной общей технологии является необходимым первым шагом, но не более того, и этот шаг часто оказывается наиболее простым.

Миф № 13: Использование реляционной базы данных дешевле использования специализированного архиватора

Прежде чем предполагать, что вы не можете позволить себе серьезное решение для архивации, убедитесь, что вы понимаете свои реальные потребности и изучите имеющиеся возможности - вы можете быть приятно удивлены даже одной стоимостью лицензии. Когда вы учтете затраты на эксплуатацию и конфигурирование такого решения под ваши реальные потребности - специализированное решение, возможно, окажется намного дешевле.

Миф № 14: Архиватор нужен только для крупномасштабных непрерывных процессов

Первые коммерческие архиваторы появились в 1980-х годах для нефтеперерабатывающих, целлюлозно-бумажных и других непрерывных производственных процессов, так как только в этих отраслях могли быть оправданы затраты на компьютеры, необходимые для запуска этих архиваторов. В связи с быстрым внедрением Windows NT в 1990-х, стоимость компьютеров значительно

снизилась, что открыло двери более низким ценовым решениям, специально разработанным для новой платформы, таких, как Wonderware® Industrial SQL Server (сейчас Wonderware Historian). Эти новые решения сразу же продемонстрировали свою экономическую эффективность вне традиционных, ориентированных на системы распределенного управления (DCS) отраслей промышленности с непрерывным производством.

Закключение:

Это техническое описание рассматривает несколько причин, по которым архиватор процесса лучше подходит для сбора и извлечения производственных данных по сравнению с реляционными базами данных. Однако это не означает, что для коммерческого программного обеспечения нет места в производственной среде. Сегодня обработанная информация необходима как за пределами производственной среды, так и внутри бизнес-отдела предприятия. И нет более подходящего способа, чтобы обеспечить интеграцию между производственными данными и системами предприятия, чем коммерчески принятый, стандартный интерфейс. Wonderware Historian может объединять коммерчески доступный продукт (Microsoft SQL Server) с открытым, стандартным интерфейсом запросов (SQL) для обеспечения открытого доступа к производственным архивным данным. Этот интерфейс может быть легко использован IT-отделом для отчетности или интеграции в систему ERP.

Wonderware Historian предоставляет все возможности, описанные в этом документе, и многое другое. Заслуживающий доверия и используемый на более чем 25,000 компьютерах по всему миру, Wonderware Historian помогает работе как заводов, так и корпоративных пользователей, предоставляя конкретную информацию конкретному лицу и передает управление базой данных туда, где оно должно быть, то есть IT-отделу предприятия, а не заводскому цеху.

Об авторе

Элиот Миддлтон имеет более чем 25-летний опыт работы с промышленными программами, в первую очередь с архиваторами процессов, приложениями для интеграции с бизнес-системами и приложениями бизнес-аналитики. Как менеджер по продукту, он отвечает за определение направления развития для программного обеспечения Schneider Electric, включая Wonderware Historian, Historian Client (он же ActiveFactory) и веб-решение Wonderware Information Server. Работая с клиентами, деловыми партнерами, группами разработчиков и службой поддержки Invensys, Элиот выполняет работу по выявлению проблем рынка, определяет требования к продукции и приоритеты усовершенствования. До прихода в Invensys в 2001 году, он занимал аналогичные должности в AspenTech и INDX. Имеет степень бакалавра в области компьютерных наук Университета Бэйлор.

EXEPLANT
OMEGA ALLIANCE GROUP

Санкт-Петербург
тел. +7 812 676 6162
info@exeplant.ru

Москва
тел. +7 495 787 7177
info@exeplant.ru

Wonderware
Russia

САНКТ-ПЕТЕРБУРГ
тел. +7 812 327 3752
info@wonderware.ru

МОСКВА
тел. +7 495 641 1616
info@wonderware.ru

ЕКАТЕРИНБУРГ
тел. +7 343 287 1919
info@wonderware.ru

САМАРА
тел. +7 846 273 95 85
info@wonderware.ru

КИЇВ
тел. +38 044 495 33 40
info@wonderware.com.ua

МИНСК
тел. +375 17 2000 876
info@wonderware.ru

КАЗАХСТАН
тел. +77779994825
sales@wonderware.kz

HELSINKI
puh. +358 9 540 4940
info@wonderware.fi

RĪGA
tel. +371 6738 1617
info@wonderware.lv

VILNIUS
tel. +370 5 215 1646
info@wonderware.lt

TALLINN
tel. +372 668 4500
info@wonderware.ee