

ГИБРИДНАЯ МОДЕЛЬ ХРАНЕНИЯ И ОБРАБОТКИ ТЕМПОРАЛЬНЫХ ДАННЫХ ДЛЯ СИСТЕМ ЦИФРОВЫХ ДВОЙНИКОВ ПРОМЫШЛЕННЫХ ОБЪЕКТОВ

Андреев Артем Александрович

Ведущий специалист, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (РАНХиГС), г. Москва
artem2033andreev@yandex.ru

HYBRID MODEL FOR STORING AND PROCESSING TEMPORAL DATA IN DIGITAL TWIN SYSTEMS OF INDUSTRIAL OBJECTS

A. Andreev

Summary. The article examines modern approaches to storing and processing temporal data in digital twin systems for industrial objects. The increasing volume of data generated by intelligent sensors and monitoring systems, as well as the widespread adoption of digital twin technologies, necessitates the development of efficient architectures for managing both historical and streaming data. The study analyzes levels of data temporality, their impact on database structures and computational performance, and the major limitations of traditional relational models under rapidly growing data loads.

Based on an analysis of scientific publications and practical solutions, two promising approaches to temporal data organization are identified: a multidimensional MIVAR information space and the architecture of digital data factories. A comparative evaluation of these approaches is conducted, and a hybrid model combining the strengths of both directions is proposed, taking into account their respective advantages and constraints.

Keywords: digital twin; temporal data; data storage; stream processing; digital data factory; multidimensional models; MIVAR model; Big Data.

Аннотация. Статья посвящена анализу современных подходов к хранению и обработке темпоральных данных в системах цифровых двойников промышленных объектов. Рост объемов данных, генерируемых интеллектуальными датчиками и системами мониторинга, а также широкое распространение технологий цифровых двойников обуславливают необходимость разработки эффективных архитектур управления историческими и потоковыми данными. В работе рассматриваются уровни темпоральности данных, их влияние на структуру баз данных и производительность вычислительных систем, а также основные ограничения традиционных реляционных моделей в условиях быстрорастущих нагрузок.

На основе анализа научных публикаций и практических решений выделены два перспективных подхода к организации темпорального хранения: многомерное миварное информационное пространство и архитектура цифровых фабрик данных. Проведен сравнительный анализ указанных подходов и на основании выявленных преимуществ и ограничений предложена гибридная модель, объединяющая сильные стороны обоих направлений.

Ключевые слова: цифровой двойник; темпоральные данные; хранение данных; потоковая обработка; цифровая фабрика данных; многомерные модели; миварная модель; Big Data.

В последние годы технология цифровых двойников демонстрирует устойчивый рост популярности. Ее внедрение и использование можно заметить во всех отраслях экономики: в промышленности, энергетике, нефтегазовом комплексе, транспорте, строительстве и ряде других отраслей. Наибольшее распространение данная технология получила в сфере промышленного производства благодаря тому, что она позволяет получить значимый экономический эффект в относительно короткие сроки. Там с помощью цифровых двойников моделируют производственные линии, оборудование и целые заводы, анализируют и оптимизируют технологические процессы. Используя данные с датчиков IoT (Internet of Things), такие системы также прогнозируют износ оборудования, предотвращают неполадки и снижают затраты на обслуживание за счёт уменьшения времени простоя. Кроме того, они позволяют проводить анализ сценариев «что-если», выполнять прогнозные

расчёты, выбирать оптимальные режимы работы и оценивать ожидаемый экономический эффект. В целом цифровые двойники становятся важным инструментом поддержки принятия решений и повышения эффективности эксплуатации сложных технических объектов.

Впервые концепцию цифровых двойников сформулировал М. Гривс. Он дал им такое определение: набор виртуальных информационных конструкций, которые полностью описывают потенциальный или фактический реальный произведенный продукт от микроскопического уровня до геометрического макроскопического уровня [9]. В контексте обсуждения промышленных объектов это означает создание точной динамической цифровой копии реального производственного объекта, от отдельного станка то целого предприятия.

Основой работы любого цифрового двойника являются данные. Для того чтобы цифровой двойник мог опе-

ративно и достоверно отображать текущее состояние физического объекта, необходима установка датчиков и организация непрерывного сбора информации. Цифровым двойникам необходимо непрерывно получать данные, корректировать математическую модель объекта, уточнять параметры его поведения и, при необходимости, воздействовать на физический объект, опираясь на результаты обработки поступающей информации. Источниками данных могут выступать датчики различных типов, устройства идентификации (штрих-коды, QR-коды), видеокамеры и другие средства фиксации параметров. Таким образом, данные являются фундаментом функционирования цифровых двойников, так как от их полноты и качества зависят границы возможностей моделирования и качество принимаемых решений.

При этом рост объёмов данных приводит к появлению новых технологических вызовов. Так, использование популярной в последнее время технологии Big Data влечет за собой ряд проблем: наличие шумов, неполнота измерений и трудности при прогнозировании редких событий. На практике значительная часть собираемой информации может оказаться избыточной: при накоплении ежедневных массивов данных до 95 % из них не несут существенной ценности, что усложняет выделение информативных признаков и повышает трудоёмкость анализа. Разработчики, ориентирующиеся на сбор максимально полного объёма данных в надежде выявить скрытые зависимости, сталкиваются с тем, что поиск действительно значимых корреляций оказывается значительно более ресурсоёмким, чем предполагалось [2].

Эта проблема особенно актуальна в проектах, где объём ежедневно поступающих данных превышает возможности аналитической обработки. В подобных ситуациях ключевую роль начинают играть физические модели, позволяющие определить, какие параметры необходимо фиксировать, в каких областях могут проявляться критические состояния и как следует выстраивать мониторинг на разных стадиях жизненного цикла изделия. При этом важно учитывать, что объём доступных данных напрямую зависит от этапа разработки. На начальных стадиях («концепция», «прототип») информация об объекте может полностью отсутствовать, а модели строятся главным образом на основе физического моделирования. В процессе эксплуатации объём доступных данных естественным образом увеличивается.

В этой связи важным понятием становится Smart Big Data (содержательные большие данные) — данные, обладающие высоким уровнем информационной насыщенности, полученные на основе четкого представления о расположении критических зон в конструкции, в которых имеет смысл размещать датчики [2]. Сбор всех возможных данных без предварительного отбора приводит к формированию объёмных массивов Big Data,

значительная часть которых состоит из малозначимой информации, что приводит к резкому росту нагрузки на системы хранения и аналитической обработки.

Следовательно, для эффективной работы с данными необходима комплексная инфраструктура, обеспечивающая весь жизненный цикл данных: сбор, фильтрацию, безопасную передачу, контроль качества, классификацию, устранение дубликатов, надёжное хранение, извлечение, последующую обработку, анализ и визуализацию. Реализация этих процессов обеспечивается широким спектром инструментов и программных решений, которые условно могут быть сгруппированы в четыре основных этапа: сбор данных, управление данными, хранение данных и использование данных.

В традиционных моделях данных, используемых в классических системах управления базами данных (СУБД), фиксируется только актуальное состояние объектов предметной области. При внесении изменений старые значения перезаписываются новыми, вследствие чего сведения о предыдущих состояниях объекта не сохраняются. Однако во многих приложениях, связанных с мониторингом и аналитикой, чем и занимаются цифровые двойники, возникает необходимость хранить не только актуальные, но и прошлые, а в некоторых случаях и будущие (плановые) значения. Цифровые двойники опираются преимущественно на темпоральные данные — произвольные данные, которые связаны с определёнными датами или промежутками времени [3]. То есть это данные, привязанные к моментам времени или временным интервалам, что требует особых инструментов для их хранения и анализа. В разработку методов хранения и обработки темпоральных данных внесли вклад учёные, такие как Р. Снодграс, К. Дженсен, С. Гадия, Е. МакКензи, А. Стейнер.

Темпоральные данные представляют собой значения параметров, соотнесённые с конкретными моментами времени или временными промежутками. Использование подобных моделей позволяет фиксировать изменения объектов во времени: если объект появляется в момент t_1 и перестаёт существовать в момент t_2 , то темпоральная база данных содержит последовательность его состояний на всём интервале $[t_1, t_2]$. При этом объём данных напрямую зависит от выбранного уровня темпоральности и частоты обновления параметров.

Принято выделять следующие несколько уровней темпоральности [3]:

- Темпоральность на уровне базы данных предполагает добавление временной информации ко всей структуре данных и сохранение периодических снимков состояния БД. Такой подход обеспечивает полную временную историю, однако приводит к значительной избыточности, поскольку

соседние состояния зачастую отличаются минимально.

- Темпоральность на уровне отношения означает, что каждая таблица имеет собственную временную ось и хранится в виде последовательности её исторических версий. Каждая версия отражает состояние отношения в определённый момент времени, формируя временную траекторию развития структуры.
- Темпоральность на уровне кортежа заключается в присвоении временных интервалов отдельным строкам таблиц. Каждый кортеж содержит значения атрибутов и указание периода их актуальности. При любом изменении данных создаётся новая строка, что приводит к росту числа записей и дублированию неизменённых значений. В результате информация о состоянии одного объекта распределяется между множеством кортежей, усложняя хранение, анализ и обработку данных.

Таким образом, темпоральные модели расширяют возможности реляционных баз данных, позволяя фиксировать историю изменений и обеспечивать полноценную временную аналитику, но вместе с тем создают высокую нагрузку на системы хранения и обработки данных в силу неизбежного увеличения объёмов информации.

В контексте обсуждения промышленных объектов вопросы хранения и обработки темпоральных данных приобретают особое значение, поскольку именно они позволяют решать следующие важнейшие задачи цифрового двойника:

- Прогнозирование срока службы оборудования: анализ исторических данных изменения основных параметров (вибрация, температура, давление) позволяет точно предсказывать момент наступления критического износа;

- Реконструкция событий при расследовании инцидентов: возможность восстанавливать полную картину изменения технологических параметров в период, предшествующий аварийной ситуации, что необходимо для установления причинно-следственных связей и дальнейшего обучения модели;
- Сравнительный анализ режимов работы: сопоставление временных рядов данных, полученных в различных условиях работы, помогает выявлять наиболее оптимальные условия эксплуатации;
- Ремонт и техническое обслуживание: планирование сервисных вмешательств на основе фактической накопленной нагрузки, а не по календарному графику, что позволяет перейти от профилактического к предиктивному обслуживанию;
- Валидация математических моделей: использование исторических данных для калибровки расчетных моделей цифрового двойника, обеспечивающее их адекватность реальному поведению физического объекта.

Поскольку объём хранимых данных неуклонно растёт, что усложняет возможность обращения к историческим данным, необходимо исследовать и проанализировать пути решения этой проблемы. В научной литературе описаны различные подходы, направленные на оптимизацию хранения, минимизацию избыточности и повышение эффективности темпоральной обработки.

Одним из вариантов решения указанных проблем является представление темпоральных данных в виде многомерного миварного информационного пространства (МВП) [3, 4]. Суть данного подхода заключается в отказе от классического табличного представления в пользу модели, где каждое значение данных представляется как точка в многомерном пространстве с осями: «Объект» (ID), «Атрибут» (S), «Время» (T) и «Отношение» (V).

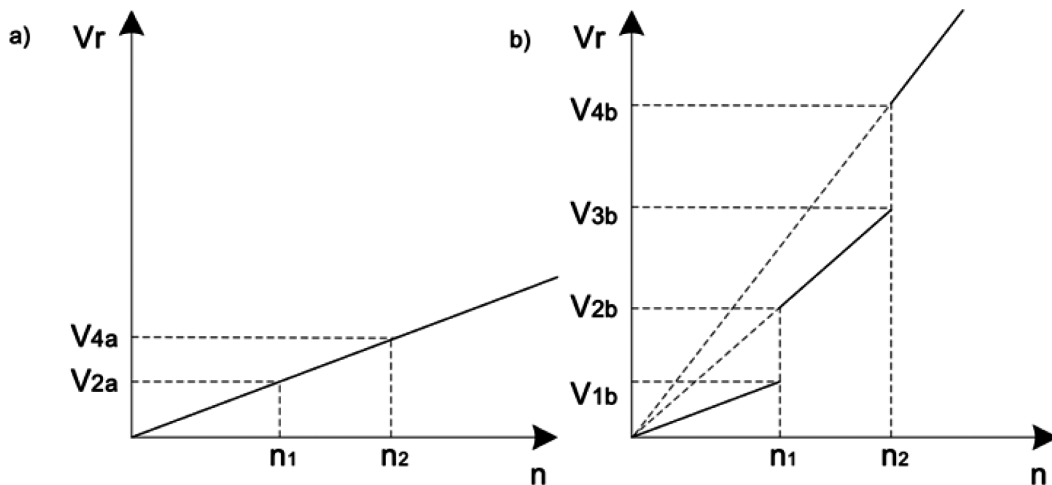


Рис. 1. Зависимость размера отношения с темпоральными кортежами от количества кортежей при неизменяющейся (а) и изменяющейся (б) схеме отношения [4]

Следовательно, состояние БД в какой-либо определённый момент времени характеризуется не набором таблиц, а совокупностью точек во временном пространстве, где каждая точка представляет собой отдельное значение конкретного атрибута для конкретного объекта в заданный момент времени.

Такой подход решает проблему избыточности темпоральных данных, так значения хранятся только для тех комбинаций, в которых происходили изменения. Таким образом не происходит дублирования не изменившихся данных при обновлении отдельного атрибута, что характерно для темпоральности на уровне кортежа. Также при добавлении нового атрибута в систему не происходят структурные изменения основы (например, добавления столбца в таблицу) и, соответственно, не происходит рост её объёма, как в реляционном случае (Рис. 1b). Новая ось или точка в пространстве просто начинает использоваться для новых данных, не затрагивая исторические. Согласно результатам, приведённым в статье, применение миварного подхода для системы кадрового учёта позволило сократить размер хранимых отношений на 32 % и ускорить адаптацию системы к изменяющимся требованиям на 53 %, что говорит о высокой эффективности подхода.

Другой подход к построению архитектуры цифровых двойников основан на концепции цифровых фабрик данных (Рисунок 2) [7]. Суть данного подхода заключается в создании платформы, которая выступает в роли системного интегратора для управления разнородными потоками данных в реальном времени. Цифровая фабрика данных представляет собой замкнутую экосистему, организующую процессы сбора, обработки, хранения и анализа информации от множества источников.

Перечислим ключевые характеристики цифровых фабрик данных для цифровых двойников:

- Контейнеризация и микросервисная архитектура: использование Docker, Kubernetes для инкапсуляции модулей обработки данных, что обеспечивает гибкость и масштабируемость системы;
- Оркестрация потоков данных: координация разнородных информационных потоков от IoT-устройств, датчиков, внешних систем с использованием инструментов типа Apache Kafka, Spark, AirFlow;
- ELT-процессинг (Extract-Load-Transform): в отличие от традиционного ETL, так как в данном случае операция преобразования данных опциональна, а операции выгрузки и загрузки данных выходят



Рис. 2. Типовая архитектура цифровой фабрики данных [7]

на первый. Подходящим инструментом является программное обеспечение с открытым исходным кодом Airbyte;

- Интеграция с облачными технологиями (cloud computing): максимальное использование облачных решений для виртуализации компонентов ИТ-инфраструктуры;

Выделим ключевые параметры двух подходов и проведем их сравнительную характеристику:

Таблица 1.

Сравнительная таблица подходов к работе с темпоральными данными для цифровых двойников

Критерий	Многомерное миварное пространство (ММП)	Цифровые фабрики данных (ЦФД)
Концепция	Представление данных как точек в многомерном пространстве	Интеграция потоков данных через контейнеризованную платформу с оркестрацией
Архитектура хранения	Многомерное пространство с осями: отношения, атрибуты, объекты, время	Многоуровневая архитектура DataOps с Data Lakes, хранилищами, «песочницами»
Обработка темпоральных данных	Хранение только изменяющихся значений в точках пространства	Потоковая обработка в реальном времени (Kafka, Spark) + ELT-процессинг
Масштабируемость	За счет разреженности многомерного представления	Горизонтальное масштабирование через контейнеризацию (Docker, Kubernetes)
Эффективность хранения	Сокращение объема за счет устранения избыточности	Оптимизация через многоуровневое хранение (горячие/теплые/холодные данные)
Тип цифрового двойника	Универсальный подход для различных типов двойников	Специализирован для DTA (Двойник-Агрегатор)
Преимущества	<ul style="list-style-type: none"> • Высокая эффективность хранения • Гибкость структуры • Устранение избыточности 	<ul style="list-style-type: none"> • Высокая производительность обработки • Масштабируемость • Интеграция с современным стеком технологий
Ограничения	<ul style="list-style-type: none"> • Специфическая модель запросов • Сложность миграции существующих систем 	<ul style="list-style-type: none"> • Высокие требования к инфраструктуре • Сложность управления распределенной системой
Применимость к промышленным объектам	Эффективно для хранения исторических временных рядов параметров оборудования	Оптимально для обработки потоковых данных с промышленных датчиков в реальном времени

На основании проведенного сравнительного анализа двух подходов представляется целесообразным разработать гибридную модель, объединяющую сильные стороны каждой из рассмотренных методик.

Предлагаемая гибридная модель строится на принципе синтеза двух подходов. От цифровых фабрик данных заимствуются механизмы оркестрации потоковой обработки и контейнеризации, обеспечивающие высокую производительность и масштабируемость. От многомерного миварного пространства берется эффективная модель хранения темпоральных данных, устраняющая избыточность и обеспечивающая гибкость структуры.

Архитектурно модель реализуется как многоуровневая система, где каждый уровень решает специфические задачи в рамках единого контура управления цифровым двойником. Уровень сбора данных обеспечивает взаимодействие с физическими объектами через интеллектуальные датчики (IoT) и edge-устройства, осуществляя предварительную фильтрацию и агрегацию данных. Уровень обработки, реализованный по принципу цифровой фабрики, организует конвейеры ELT-процессинга с использованием контейнеризованных микросервисов. Уровень хранения реализует гибридную схему, где горячие данные (последние 30 дней) размещаются в колоночных хранилищах для обеспечения высокой скорости запросов, а исторические данные преобразуются в многомерное представление для эффективного долгосрочного хранения. Уровень аналитики и представления предоставляет единый интерфейс для временных запросов и OLAP-анализа.

Ключевым механизмом, обеспечивающим работу данной модели, будет являться трансформатор, преобразующий потоковые данные в точки многомерного пространства. Реализация такой технологии позволит сохранить преимущества потоковой обработки в сочетании с эффективностью многомерного хранения.

Для цифровых двойников промышленных объектов предлагаемая гибридная модель будет особенно эффективной, поскольку позволит решить характерные для промышленной области задачи, связанные с разнородностью данных, их большими объемами и высокими требованиями к надежности и производительности.

В результате проведенного исследования и анализа современных подходов к хранению и обработке темпоральных данных для цифровых двойников промышленных объектов можно прийти к следующим выводам.

Сравнительный анализ подходов показал, что ни один из рассмотренных методов не является универсальным. Каждый из них обладает своими преимуще-

ствами и ограничениями. Так, ММП обеспечивает эффективное хранение за счёт устранения избыточности хранимых данных и гибкости структуры, но при этом требует специфических механизмов запросов. ЦФД же демонстрируют высокую производительность при обработке потоковых данных и обладают потенциалом масштабируемости, однако предъявляют повышенные требования к инфраструктуре.

Предлагаемая гибридная модель интегрирует сильные стороны обоих подходов, что позволяет прийти

к сбалансированному решению, сочетающему высокую производительность, масштабируемость и экономичность хранения. Практическая значимость гибридной модели заключается в её способности решать главные задачи цифровых двойников промышленных объектов. В целом, реализация гибридной модели откроет перспективы для создания более гибких, эффективных и адаптивных систем управления данными в цифровых двойниках.

ЛИТЕРАТУРА

1. Денисов С.Г. Программное обеспечение для физической части цифрового двойника и связи его элементов // Бюллетень инновационных технологий. — 2024. — №3 (31). — С. 10–14.
2. Денисов С.Г. Технологии сбора и обработки данных для создания цифровых двойников // Бюллетень инновационных технологий. 2023. — №2 (26). — С. 12–17.
3. Елисеев Д.В. МИВАРное представление темпоральных реляционных моделей // I-methods. 2016. №2. . — С. 35–42.
4. Елисеев Д. В. Темпоральные объектно-реляционные модели в многомерном представлении // Объектные системы. 2015. №10. — С. 28–34.
5. Коровин Г.Б. Возможности применения цифровых двойников в промышленности // Вестник ЗабГУ. 2021. №8. — С. 124–133.
6. Нелюбина Е.А. Парадигма программирования для обработки мультимодальных темпоральных данных цифровых двойников // Новое в экономической кибернетике. — 2023. — №4. — С. 39–51.
7. Потапов В.П., Попов С.Е., Счастливцев Е.Л. Цифровые двойники: стратегии и подходы к созданию систем экологического мониторинга // Вычислительные технологии. — 2023. — №3. — С. 167–181.
8. Тоноян С.А., Сараев Д.В. Темпоральные модели базы данных и их свойства // Инженерный журнал: наука и инновации. 2014. №12 (36). — С. 1–11.
9. Цифровые двойники в высокотехнологичной промышленности. Краткий доклад (сентябрь 2019 года) / А.И. Боровков [и др.]. — СПб. : ПОЛИТЕХ-ПРЕСС, 2019. — 62 с.

© Андреев Артем Александрович (artem2033andreev@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»