Данные цифрового двойника как элемент корпоративной системы данных: анализ конфликтующих требований проектирования

Михайлов Р.В.

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)

SPIN-код: 7258-5082; ORCID: 0009-0008-7454-3119

***Аннотация*. В статье рассматривается представление данных цифрового двойника как элемента корпоративной системы данных распределенного машиностроительного холдинга. Показано, что цифровой двойник не может рассматриваться только как прикладной сервис моделирования: он формируется на основе нормативно-справочной информации, проектной и конструкторской документации, инженерных данных, математических моделей и онлайн-данных, поступающих из производственного контура. Предложена схема включения данных цифрового двойника в гибридную архитектуру корпоративных данных, сочетающую Edge-уровень, локальную серверную предприятия, корпоративный ЦОД, Data Lake, Data Warehouse, MDM/НСИ и Data Fabric. Отдельное внимание уделено анализу конфликтующих требований: низкая задержка против безопасности, автономность заводов против единого источника истины, полнота исторических данных против стоимости хранения, гибкость локальных моделей против стандартизации метаданных. Сформирована матрица конфликтов и предложены архитектурные компромиссы для проектирования устойчивой корпоративной системы данных.**

Ключевые слова: цифровой двойник; корпоративные данные; фабрика данных; управление мастер-данными; озеро данных; распределенный холдинг; конфликтующие требования; гибридная архитектура; Индустрия 4.0.

I. Введение

Распределенный машиностроительный холдинг представляет собой совокупность производственных площадок, конструкторских подразделений, сервисных центров, корпоративного центра и внешних контрагентов. Для такого типа организаций характерны разный уровень автоматизации предприятий, наличие унаследованных систем, дублирование справочников, неодинаковые правила хранения данных и высокая зависимость управленческих решений от качества информационного обмена между цеховым, заводским и холдинговым уровнями.

Одним из ключевых прикладных потребителей корпоративных данных становится цифровой двойник. В традиционном понимании цифровой двойник рассматривается как интегрированная постоянно обновляемая модель актива, используемая для анализа, моделирования, прогнозирования и поддержки принятия решений [1]. Однако в условиях распределенного холдинга цифровой двойник должен рассматриваться не как изолированная модель, а как часть корпоративной системы данных, связанная с PLM/PDM, MES, SCADA, ERP, MDM, Data Lake, Data Warehouse и аналитическими сервисами.

Цель статьи – определить место данных цифрового двойника в корпоративной системе данных распределенного машиностроительного холдинга и рассмотреть конфликтующие требования, возникающие при проектировании архитектуры их сбора, хранения, актуализации и предоставления прикладным сервисам.

Данные цифрового двойника в корпоративной системе данных

Структура цифрового двойника промышленного актива включает несколько взаимосвязанных слоев: нормативно-справочную информацию, проектную и конструкторскую документацию, графическое представление объекта, инженерные данные, математические модели и онлайн-данные [1, 2]. Каждый слой имеет собственные источники, периодичность обновления, требования к качеству и ограничения по размещению.

Для корпоративной системы данных цифровой двойник является одновременно потребителем и генератором данных. С одной стороны, он использует данные из PLM/PDM, MES, SCADA, Historian, ERP и MDM. С другой стороны, результаты моделирования, прогнозы технического состояния, сценарные расчеты и цифровые следы эксплуатации должны возвращаться в корпоративный контур и использоваться в системах ТОиР, BI, планирования производства и управления активами.

В распределенном холдинге особенно важно обеспечить сквозную идентификацию объектов. Один и тот же актив может быть представлен в проектной документации, локальной системе завода, системе технического обслуживания, бухгалтерском учете и аналитической витрине. Поэтому данные цифрового двойника должны опираться на единый контур НСИ/MDM, корпоративный каталог данных и механизмы управления качеством данных.

На рисунке 1 показано, что слои цифрового двойника не должны храниться в едином физическом репозитории. Напротив, они распределяются между специализированными системами, но объединяются с помощью корпоративных механизмов идентификации, метаданных, API и политик доступа. Например, онлайн-данные от оборудования могут храниться в локальном Historian или Data Lake, конструкторская документация – в PLM/PDM, мастер-данные – в MDM, а агрегированные показатели – в витринах данных.

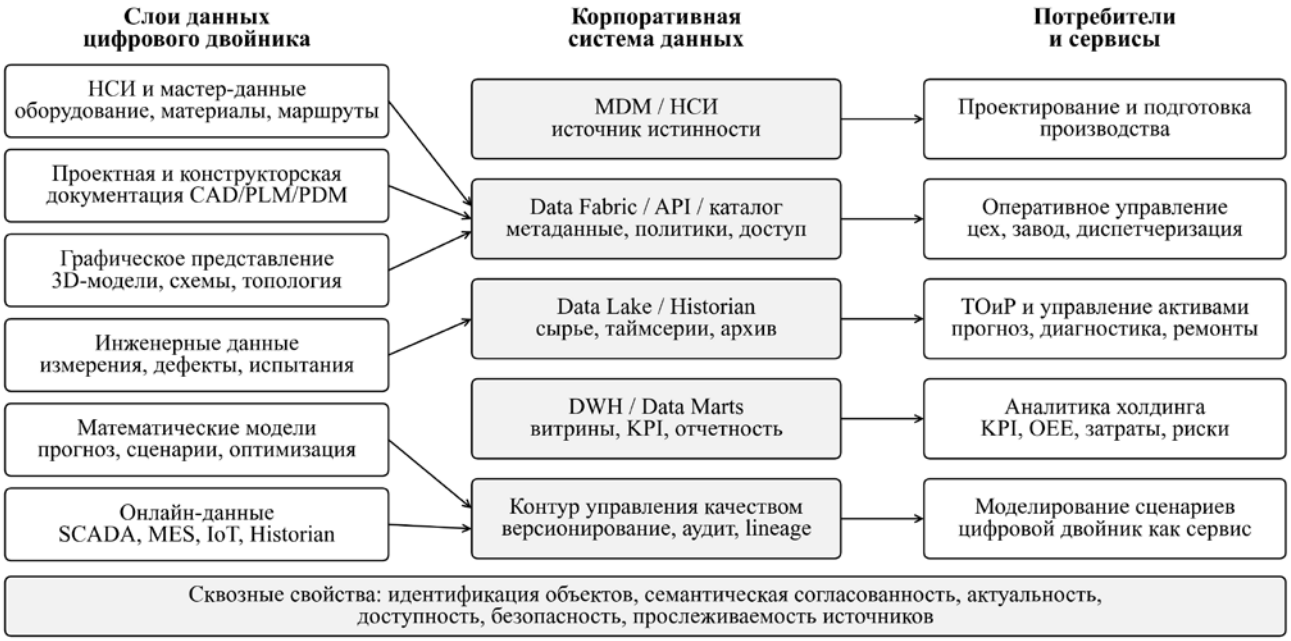


Рис. 1. Данные цифрового двойника в составе корпоративной системы данных

II. Классификация данных цифрового двойника

Для проектирования архитектуры данные цифрового двойника целесообразно классифицировать по происхождению, назначению и критичности. Такая классификация позволяет определить, какие данные должны храниться на Edge или локальном On-Premise уровне, какие могут быть перенесены в корпоративный ЦОД, а какие могут предоставляться через Data Fabric без физического копирования.

ТАБЛИЦА I. Классификация данных цифрового двойника

Класс данных	Примеры	Типовые источники	Ключевые требования	Предпочтительный контур
НСИ и мастер-данные	Активы, оборудование, материалы, маршруты, подразделения	MDM/НСИ, ERP, PLM	Семантическая согласованность, отсутствие дублей, управляемость	Корпоративный MDM с репликацией на заводы
Проектно-конструкторские данные	3D-модели, чертежи, спецификации, технологические процессы	CAD/CAM/PLM/PDM	Версионность, контроль доступа, неизменность утвержденных версий	PLM/PDM + корпоративный каталог
Операционно-технологические данные	Параметры станков, аварии, режимы, OEE, телеметрия	PLC, SCADA, MES, Historian, IoT	Низкая задержка, высокая доступность, устойчивость к потере WAN	Edge / локальный Historian / Data Lake
Инженерные и диагностические данные	Вибродиагностика, измерения КИМ, дефекты, лабораторные испытания	QMS, ТОиР, лабораторные системы, переносные приборы	Точность, привязка к активу, прослеживаемость измерений	Локальная БД + корпоративный архив
Модельные и расчетные данные	Прогноз отказов, сценарии, параметры моделей, результаты симуляции	Модели ML, расчетные сервисы, цифровой двойник	Воспроизводимость, вычислительная емкость, управление версиями моделей	ЦОД / IaaS/PaaS / витрины аналитики

Представленная классификация показывает, что для цифрового двойника невозможно задать универсальное правило хранения данных. Онлайн-данные требуют близости к источнику и минимальной задержки, проектные данные требуют строгой версионности, а мастер-данные должны иметь единый источник истины. Следовательно, проектирование архитектуры должно выполняться как многокритериальная задача, учитывающая технические, организационные и регуляторные ограничения.

III. Формальное представление объекта данных цифрового двойника

Для дальнейшего использования в задаче структурно-параметрического синтеза объект данных цифрового двойника может быть представлен кортежем $OD_{DT} = \langle ID, A, L, S, V, F, Q, R, P \rangle$,

где ID – уникальный идентификатор объекта или актива; A – набор атрибутов; L – слой цифрового двойника; S – источник данных; V – версия или временная метка; F – частота обновления; Q – показатели качества; R – ограничения безопасности и доступности; P – допустимые варианты размещения.

Такое представление связывает цифровой двойник с общей моделью корпоративных данных. Атрибуты F и P позволяют определить, должен ли объект обновляться в режиме реального времени на Edge-уровне или может обрабатываться пакетно в корпоративном ЦОД. Атрибуты Q и R позволяют учитывать полноту, актуальность, достоверность, конфиденциальность и требования к доступу.

IV. Конфликтующие требования при проектировании

При включении цифрового двойника в корпоративную систему данных возникает группа конфликтующих требований. Они обусловлены тем, что данные цифрового двойника одновременно должны быть оперативными, достоверными, доступными, защищенными, исторически полными и экономически рациональными. В распределенном холдинге эти требования усиливаются территориальной удаленностью площадок, разными уровнями автоматизации и неоднородностью локальных информационных систем. Основные конфликты представлены на рис. 2 и в табл. II.

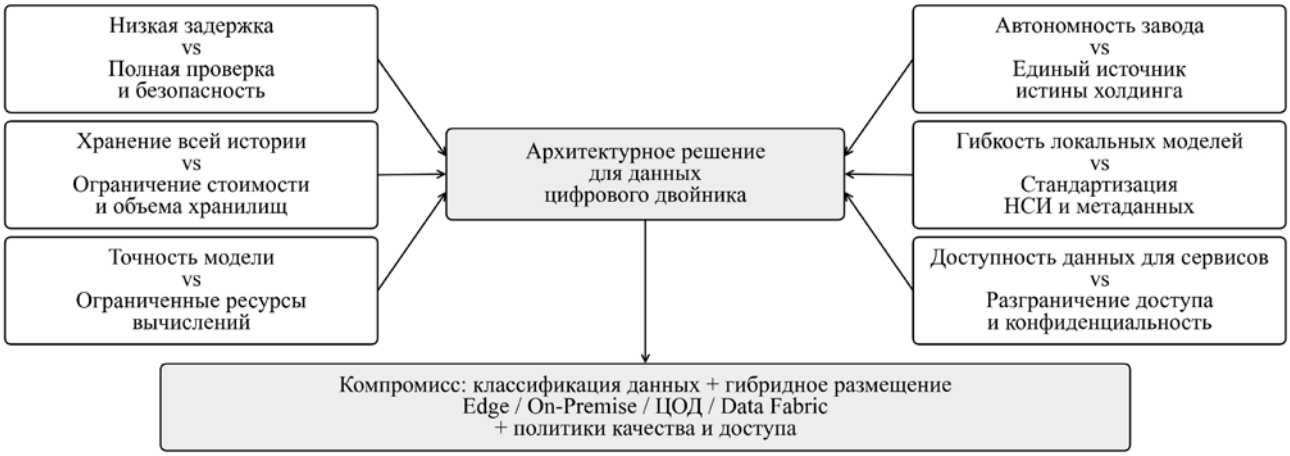


Рис. 2. Основные конфликтующие требования к данным цифрового двойника

ТАБЛИЦА II. Матрица конфликтующих требований к данным цифрового двойника

Конфликт	Суть противоречия	Риск для цифрового двойника	Архитектурный компромисс
Низкая задержка vs безопасность	Шифрование, проверка прав и аудит увеличивают время доступа к данным.	Устаревание онлайн-модели или невозможность применения APC/оперативной аналитики.	Легковесная обработка и локальные политики на Edge; полный аудит и расширенная проверка в ЦОД.
Автономность завода vs единый источник истины	Завод должен работать при потере связи, но холдинг требует единой модели активов и показателей.	Расхождение локального цифрового двойника и корпоративной модели.	Локальное кэширование и репликация НСИ; централизованное управление мастер-данными.
Полнота истории vs стоимость хранения	Сохранение всех таймсерий, видео и результатов моделирования требует больших объемов хранилищ.	Рост стоимости Data Lake и снижение управляемости данных.	Разделение Hot/Warm/Cold хранения; агрегирование и архивирование по политике жизненного цикла.
Гибкость локальных моделей vs стандартизация	Предприятия используют собственные модели оборудования и показатели качества.	Невозможность сравнивать данные цифровых двойников между заводами.	Единый корпоративный словарь и метамодель; доменные расширения через Data Fabric.
Точность модели vs вычислительные ресурсы	Высокоточные модели требуют больших вычислительных мощностей и данных.	Задержки в расчетах, невозможность оперативного прогнозирования.	Разделение моделей на оперативные упрощенные и детальные сценарные; перенос тяжелых расчетов в ЦОД/IaaS.
Доступность данных vs разграничение доступа	Для интеграции сервисов нужен широкий доступ, но данные могут быть коммерчески или технологически чувствительными.	Утечки данных, нарушение требований безопасности и регуляторов.	Ролевая модель доступа, маскирование, сегментация контуров, контроль API и аудит обращений.

Матрица конфликтов позволяет перейти от общего перечня требований к архитектурным решениям. Например, требование минимальной задержки не означает, что все данные должны храниться на Edge-уровне. На Edge целесообразно оставлять только те данные и сервисы, которые связаны с управлением оборудованием, аварийными событиями и оперативным расчетом локальных показателей. Для исторического анализа, обучения моделей и консолидации KPI более рационально использовать корпоративный Data Lake или DWH.

V. Принципы архитектурного разрешения конфликтов

Для разрешения конфликтующих требований предлагается использовать набор архитектурных принципов.

- Классификация данных до выбора технологии хранения. Сначала определяется слой цифрового двойника, критичность, частота обновления и требования к доступу, затем выбирается место размещения.
- Разделение оперативного и аналитического контуров. Оперативные данные обрабатываются ближе к оборудованию, аналитические и исторические данные консолидируются в корпоративном контуре.

- Единый контур НСИ/MDM. Идентификаторы активов, оборудования, материалов и процессов должны управляться централизованно, а локальные системы должны использовать репликацию или синхронизацию справочников.
- Использование Data Fabric. Вместо тотального физического копирования данных применяется слой унифицированного доступа, каталогизации, политик безопасности и виртуализации источников.
- Управление жизненным циклом данных. Для таймсерий, видео, результатов расчетов и инженерных измерений задаются политики Hot/Warm/Cold хранения, архивирования и удаления.
- Прослеживаемость данных и моделей. Для результатов цифрового двойника должны храниться источник данных, версия модели, параметры расчета и временная метка, что обеспечивает воспроизводимость решений.

Практически это приводит к гибридной архитектуре: данные реального времени размещаются на Edge и локальном уровне завода; инженерные и производственные данные агрегируются в локальных хранилищах и передаются в корпоративный Data Lake; мастер-данные управляются централизованно; результаты моделирования и аналитические показатели публикуются в витринах данных и BI-сервисах.

VI. Обобщенная схема проектирования

Проектирование корпоративного контура данных цифрового двойника может быть выполнено в следующей последовательности:

- идентифицировать активы и объекты цифрового двойника;
- описать слои данных цифрового двойника и их источники;
- классифицировать данные по задержке, критичности, объему, частоте обновления и требованиям безопасности;
- определить конфликтующие требования и допустимые архитектурные компромиссы;
- назначить контур размещения данных: Edge, локальный On-Premise, корпоративный ЦОД, облачная платформа или виртуальный доступ через Data Fabric;
- задать механизмы синхронизации, версионирования, аудита и контроля качества;
- проверить архитектурный вариант по метрикам задержки, стоимости, надежности, доступности и безопасности.

Такой подход позволяет рассматривать цифровой двойник как управляемый набор корпоративных данных, а не как автономную прикладную систему. Это особенно важно для распределенных холдингов, где единая аналитика должна сочетаться с автономностью производственных площадок.

VII. Заключение

В статье рассмотрены данные цифрового двойника как элемент корпоративной системы данных распределенного машиностроительного холдинга. Показано, что цифровой двойник формируется на основе разнородных источников: НСИ, проектной документации, инженерных данных, математических моделей и онлайн-данных промышленного контура. Поэтому его эффективное использование требует не только разработки моделей, но и построения гибридной архитектуры корпоративных данных.

Выделены основные конфликтующие требования, возникающие при проектировании: низкая задержка против безопасности, автономность заводов против единого источника истины, полнота исторических данных против стоимости хранения, гибкость локальных моделей против стандартизации и доступность данных против разграничения доступа. Для их разрешения предложены архитектурные компромиссы, основанные на классификации данных, использовании MDM/НСИ, Data Fabric, Data Lake, локальных Edge-хранилищ и политик жизненного цикла данных.

Полученные результаты могут быть использованы как основа для дальнейшей формализации задачи структурно-параметрического синтеза корпоративной системы данных, включая построение модели оценки архитектурных вариантов и выбор рационального размещения сервисов цифрового двойника.

Список литературы

[1] ГОСТ Р 57700.37-2021. Компьютерные модели и моделирование. Цифровые двойники изделий. Общие положения. – М.: Стандартинформ, 2021.

[2] Михайлов Р.В. Подготовка инфраструктуры к внедрению цифровых решений // Современные тенденции и перспективы развития инновационной деятельности в условиях цифровой трансформации.

[3] Михайлов Р.В. Цифровой двойник на примере газотранспортного предприятия // Современные тенденции и перспективы развития инновационной деятельности в условиях цифровой трансформации.

[4] DAMA International. DAMA-DMBOK: Data Management Body of Knowledge. 2nd ed. Technics Publications, 2017.

[5] Dehghani Z. Data Mesh: Delivering Data-Driven Value at Scale. O'Reilly Media, 2022.

[6] Industry IoT Consortium. The Industrial Internet Reference Architecture. Version 1.10. 2022.

[7] Schweichhart K. Reference Architectural Model Industrie 4.0 (RAMI 4.0). Plattform Industrie 4.0, 2016.

[8] ГОСТ 34.602-2020. Информационная технология. Комплекс стандартов на автоматизированные системы. Техническое задание на создание автоматизированной системы. М.: Российский институт стандартизации, 2021.

[9] ГОСТ 24.104-2023. Единая система стандартов автоматизированных систем управления. Автоматизированные системы управления. Общие требования. М.: Российский институт стандартизации, 2023.

[10] ГОСТ Р МЭК 62264-1-2010. Интеграция систем управления предприятием. Часть 1. Модели и терминология. М.: Стандартинформ, 2014.

[11] Inmon W.H. Building the Data Warehouse. Indianapolis: Wiley Publishing, 2005.

[12] Gieß A., Hutterer A. The Future of Data Management: A Delimitation of Data Platforms, Data Spaces, Data Meshes, and Data Fabrics // Information Systems and e-Business Management. 2025. Vol. 23. P. 971–997. DOI: 10.1007/s10257-025-00707-4.