

# КАК МАТРИЦА ДАННЫХ ПОМОГАЕТ СОВРЕМЕННЫМ ПРЕДПРИЯТИЯМ

АЛЕКСАНДР ШУМИЛИН, ВЯЧЕСЛАВ ЕЛАГИН

Единый доступ к распределенным и разнородным источникам данных играет ключевую роль для развития современного предприятия. Он обеспечивается за счет построения выделенного уровня архитектуры, который формирует единое пространство данных, полученных со всех производственных циклов, площадок и облачных сред.

Сбор и анализ информации, получение выводов и их применение на практике — это главные элементы самообучающихся систем. Такие системы служат основой для перехода от автоматических к автономным и, следовательно, адаптивным процессам в производстве. Закономерной целью при этом является улучшение общей эффективности оборудования и создание нового цифрового бизнеса.

К примеру, многие отечественные производители активно внедряют системы автоматизированного контроля состояния промышленного оборудования. Такие решения позволяют продлить ресурс техники и снизить расходы на гарантийное или сервисное обслуживание. Можно также отметить появление ряда

систем, которые за счет цифровизации и подключения географически разнесенного оборудования в единую информационную систему значительно повышают эффективность производственных процессов. Так, на нефтегазодобывающих предприятиях «кластер» из нескольких насосных систем с автоматизированной корректировкой параметров работы значительно эффективнее, чем набор из независимых насосов.

Пример можно найти и в сфере контроля качества продукции. Процессы машинного обучения могут в цифровом формате оценивать качество обработки поверхности изделия (рис. 1). Затем алгоритмы генерируют рекомендации для дальнейших действий или автоматически предпринимают меры по устранению неполадок, корректировке

последующих производственных этапов или оптимизации фрезерования.

Чтобы реализовать такие сценарии, необходимо создать цифровую модель производства и подключить к ней поток актуальных параметров качества и других показателей реальных процессов. За счет корреляции параметров качества и технологических параметров генерируются модели данных, которые обеспечивают непрерывный контроль и оптимизацию производства.

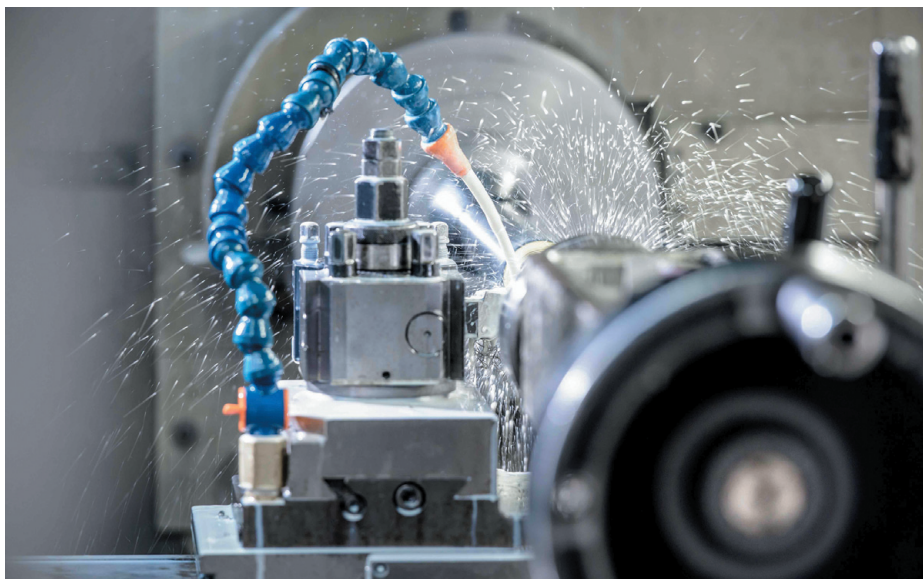
Однако это только теория — на практике компании-производители сталкиваются с дилеммой организации цикла обработки данных. Рассмотрим суть дилеммы и стратегию ее решения.

## ЦИКЛ ОБРАБОТКИ ДАННЫХ НА ПРОИЗВОДСТВЕННЫХ ПЛОЩАДКАХ И ЦЕНТРАЛИЗОВАННЫХ ПЛАТФОРМАХ

Качество моделей данных в значительной степени зависит от объема и качества «учебного материала». Специалисту могут потребоваться годы, прежде чем он полностью изучит механику работы отдельного фрезерного станка и научится замечать или предотвращать неисправности на раннем этапе. Если объединить опыт эксплуатации сотен или тысяч аналогичных фрезерных станков, такой процесс обучения сократится до недель или дней.

Именно это и происходит, когда информация о похожем обо-

**РИС. 1. ▼**  
Самообучающиеся системы являются основой для перехода от автоматических к автономным и, следовательно, адаптивным процессам в производстве



рудования или технологических процессах собирается с разных площадок и используется в самообучающейся модели данных. Благодаря этому эффекту сети передачи данных компания получает значительный потенциал для создания долгосрочных конкурентных преимуществ. Однако для обучения моделей важны не только данные, непосредственно участвующие в процессе. Если говорить о производстве, то, помимо сведений о шероховатости поверхности, процессе фрезерования, механизме подачи или глубине среза, могут пригодиться и измеряемые параметры самой фрезеральной машины (например, вибрация), сведения об окружающей среде (например, влажность), логистические и бизнес-показатели (например, из ERP-систем).

Обученные модели, алгоритмы и правила анализируют данные и управляют действиями во время производственного процесса. В зависимости от специфики процесса может потребоваться как малое время отклика (задержки), так и обработка в режиме реального времени. Именно поэтому анализ данных ведется на производственных площадках недалеко от обо-

рудования, с использованием промышленных периферийных систем (EDGE systems), которые выступают в качестве интерфейса между промышленными и ИТ-системами. Благодаря периферийным системам анализ данных и генерация управляющих воздействий могут выполняться без потокового обмена информацией с удаленными центрами обработки данных (ЦОД) или облаками. Такой подход дает защиту от сбоев каналов связи, сокращает время принятия решений и обеспечивает максимально возможную стабильность процесса.

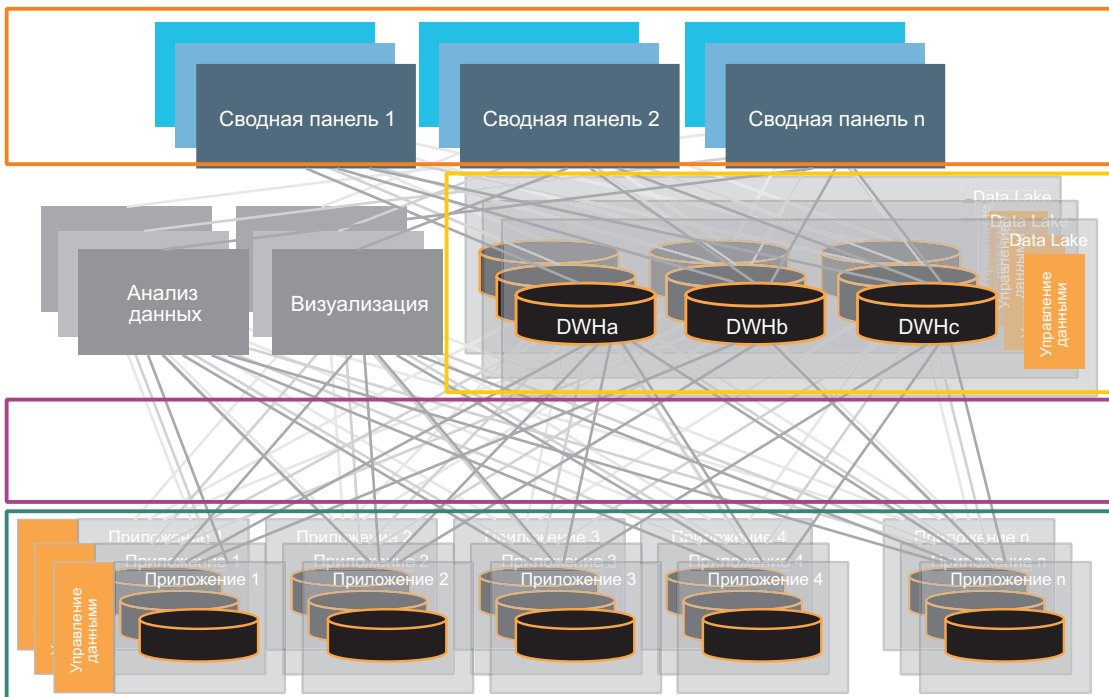
В качестве отступления можно привести пример, косвенно сопряженный с технологическими процессами. Одна из мер обеспечения безопасности на многих российских предприятиях — проведение медицинского контроля при допуске сотрудников на производственную площадку. Идентификация конкретного сотрудника и фиксирование факта прохождения медицинского контроля могут осуществляться с помощью фото/видеосистем. Это позволяет избежать преднамеренных или случайных ошибок с прохождением контроля не тем сотрудником. Для идентификации сотрудника на удаленной площадке с ограни-

ченными сетевыми подключениями нецелесообразно передавать изображения в головной или региональный ЦОД. Эту задачу можно выполнять с помощью обученной модели, которая работает на специализированных системах, созданных для граничных вычислений.

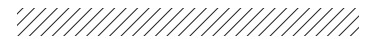
Для поддержания в актуальном состоянии моделей, работающих на EDGE-системах, организован постоянный цикл передачи данных между распределенными производственными площадками и централизованными платформами. С производственных площадок во время технологического окна (или в фоновом режиме) в центр пересылается накопленная для обучения моделей информация. В центре на основе консолидированной с разных площадок информации происходит переобучение моделей. Обновленные модели, алгоритмы и правила вновь рассылаются из центра на производственные площадки.

**ПОЧЕМУ СТОИТ ИЗБЕГАТЬ «ОСТРОВОВ ДАННЫХ» И ЗАВИСИМОСТЕЙ**

Чтобы построить такой цикл обработки информации для всех производственных процессов,



**РИС. 2.** Принцип выбора «лучшего» в своей области в отношении IoT-платформ может повысить сложность, уже существующую в производственных средах, как, например, в классической «спагетти-архитектуре». **Примечание.** DWH (Data Warehouse) — хранилище данных



данные должны быть интегрированы как вертикально, так и горизонтально. При вертикальной интеграции информация от промышленного оборудования передается на централизованные системы — облачные платформы или платформы «Интернета вещей». Это могут быть, к примеру, показатели температуры и вибрации с соответствующей отметкой времени, которые консолидируются, обрабатываются и визуализируются в центре.

При этом компании-производители сталкиваются с упомянутой в начале статьи дилеммой. Если они используют платформы «Интернета вещей» от различных вендоров, они создают «острова данных», которые затрудняют общий анализ и управление. Сложность, которая уже существует в производственных средах, либо остается на прежнем уровне, либо увеличивается — именно так заказчики приходят к классической «спагетти-архитектуре», в которой несколько баз данных, аналитических инструментов и приложений перекрестно связаны через отдельные интерфейсы (рис. 2). С другой стороны, если компании будут

полагаться только на одну или несколько платформ для упрощения внутренней структуры, зависимость от этих платформ будет расти.

Горизонтальная интеграция данных может решить эти проблемы. С таким подходом данные не передаются в центр, а формируют своего рода выделенный уровень данных (рис. 3). Мы будем называть этот уровень матрицей данных, на основе которой можно связывать данные, полученные из разных источников.

### УПРАВЛЕНИЕ ЦИКЛОМ ОБРАБОТКИ ДАННЫХ С ПОМОЩЬЮ МАТРИЦЫ ДАННЫХ

Матрица данных абстрагирует разнородные файловые системы и представляет их как единое (глобальное) адресное пространство. В результате производственные компании получают решение с универсальным доступом к данным вне зависимости от места их размещения — как в ЦОД, так и на множестве удаленных производственных площадок. Это дает возможность комплексного управления данными, например для кон-

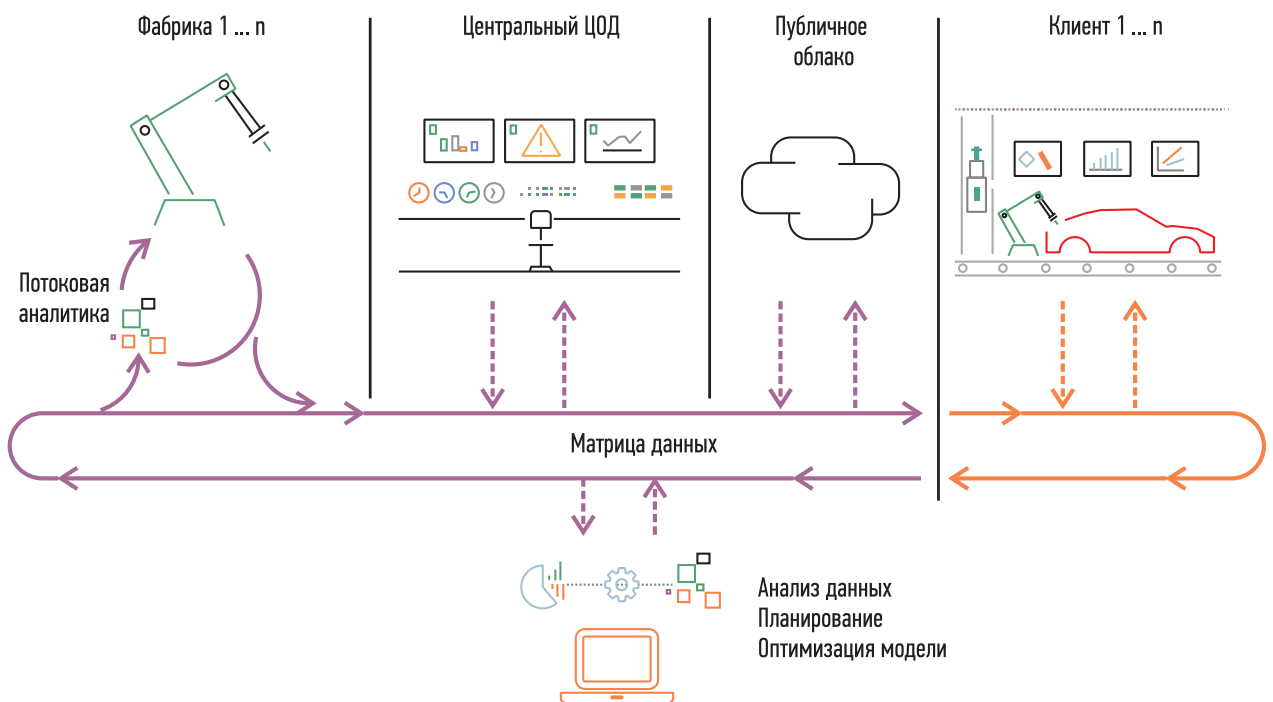
троля прав доступа. Матрица данных также позволяет организовать цикл обработки данных, описанный выше. Это хаб, через который производственные площадки, облачные сервисы и партнерские компании получают доступ к обученным моделям и передают свои накопленные данные.

Такой подход решает проблему сложности, поскольку он обеспечивает унифицированный доступ к информации, а взаимодействиями между приложениями, источниками и базами данных организовано через матрицу данных. Кроме того, решается и проблема зависимости, поскольку производственная компания может самостоятельно, без внешних платформ (и неконтролируемых внешних веб-порталов), управлять циклом обработки данных. Даже при использовании нескольких внешних платформ единообразие архитектуры данных сохраняется. Это может дополнительно уменьшить зависимость за счет мультивендорной стратегии.

### СТРУКТУРА МАТРИЦЫ ДАННЫХ

Матрица данных основана на открытой и прозрачной архи-

**РИС. 3.** ▼  
Матрица данных поможет справиться как с проблемой сложности, так и с проблемой зависимости



Аналитик данных, руководитель ИТ-департамента, исполнительный директор, директор производства

текстуре. Наиболее важные этапы цепочки процессов:

- Сбор. Сбор данных осуществляется с помощью программных модулей, получающих доступ к информации через API (Application Programming Interface, интерфейсы прикладного программирования). Это может быть база данных SQL ERP-системы, данные датчиков фрезерного станка или база данных NoSQL облачного приложения. Программные модули преобразуют соответствующие промышленные протоколы в IP-пакеты и, таким образом, охватывают различные источники данных.
- Агрегация. Данные поступают из исходных систем в матрицу данных через конвейеры, делая информацию доступной для целевых приложений, в том числе с помощью систем обмена сообщениями. На этом этапе данные часто сортируют и сжимают, поскольку не все исходные данные релевантны для дальнейшей обработки. Кроме того, данные могут храниться в так называемом озере данных. Оно объединяет множество разнородной производственной информации для создания максимально большой базы данных для машинного обучения. В отличие от традиционного хранилища «озеро данных» может быть распределено по производственным площадкам, ЦОД и различным облакам. Система управления определяет, какие пользователи могут иметь доступ к каким видам данных и каким способом. Это значит, что распределенное «озеро данных» также может быть использовано несколькими компаниями без ущерба суверенитету каждой из них по отношению к данным.
- Анализ. Используя встроенные инструменты, аналитики могут получить доступ как к потоковым данным, так и к данным из распределенного «озера данных», чтобы экспериментировать с моделями и обучать их. С помощью потоковой аналитики (анализа потоков данных в реальном времени) модели

переобучаются и используются для мониторинга текущего производства. Например, они могут распознать отклонения или признаки надвигающихся неисправностей. Потоковая аналитика не только является базовым инструментом для автономного управления производственным оборудованием, но и помогает принимать решения для среднесрочных действий, таких как профилактическое обслуживание.

- Действие. Рекомендация выполнить то или иное действие вырабатывается автоматически на основе алгоритмов или бизнес-логики. Например, открытие сервисной заявки происходит, если параметры качества производимых изделий не соответствуют требованиям из-за износа оборудования. Кроме того, при обнаружении такой ситуации параметры следующих этапов обработки изделий могут быть скорректированы, чтобы вернуть качество на прежний уровень. Такие системы называют автономными, или системами с самооптимизацией.

### КОНТЕЙНЕРЫ КАК ТЕХНОЛОГИЧЕСКАЯ БАЗА

Виртуализация контейнеров используется как технологическая база для матрицы данных. Она помогает распространить унифицированную логику работы с матрицей данных на производственные и логистические площадки, ЦОД и облака. Для построения собственных распределенных приложений подойдут, например, контейнеры и их оркестрация с помощью Kubernetes. Проблема сохранения данных контейнеров также решается: можно «упаковать» в контейнеры такие монолитные приложения, как MES или PPS. Все это дает компаниям возможность воспользоваться преимуществами однородной операционной среды с высокой эффективностью.

Создавая такую матрицу данных, каждая компания должна выбрать, сделать ее самостоятельно или купить. Сегодня в мире существует множество доступных технологий и инструментов

с открытым исходным кодом, которые компании могут использовать для создания собственной матрицы данных. Альтернативой являются стандартные коммерческие продукты, такие как HPE Ezmeral Data Fabric. Это масштабируемая распределенная файловая система, с помощью которой можно управлять большими объемами данных даже в петабайтном диапазоне с высокой производительностью. HPE Ezmeral Data Fabric — ключевой компонент HPE Ezmeral Container Platform, который используется для создания матрицы данных и контейнеризации прикладных сред. Платформа также поддерживает возможность анализа данных из распределенных источников и включает функцию непрерывного хранения данных в контейнерных средах.

### ЗАКЛЮЧЕНИЕ

За счет матрицы данных производственные компании могут создать сетевой эффект, необходимый для раскрытия потенциала своих данных — например, получения высокой операционной эффективности или разработки собственной цифровой платформы, на основе которой можно предлагать цифровые сервисы своим клиентам. Матрица данных представляет собой информационный хаб, который обеспечивает обмен данными и исполнение единой логики обработки данных. Можно подключить к ней внешние источники данных, а также предоставить контролируемый доступ к информации своим клиентам или подрядчикам.

С таким подходом производственные компании становятся независимыми от централизованных платформ «Интернета вещей» или облачных платформ, поскольку достигают сетевых эффектов с помощью децентрализованной архитектуры под собственным контролем. Внешние облачные сервисы остаются незаменимыми ресурсами, потому что у них есть отличные инструменты и возможности агрегирования. Однако теперь их можно использовать с полноценным контролем над данными и с возможностью получения максимальной выгоды от цифровизации. ●