

УДК 62.91

**ПОДХОД К ОРГАНИЗАЦИИ ХРАНИЛИЩА ДАННЫХ, ПОЛУЧАЕМЫХ С
УСТРОЙСТВ, ИСПОЛЬЗУЮЩИХ ТЕХНОЛОГИЮ ПРОМЫШЛЕННОГО ИНТЕР-
НЕТА ВЕЩЕЙ**

Квашнин Д.Ю. (ФГБОУ ВО «МГТУ «СТАНКИН»), **Ковалев И.А.** (ФГБОУ ВО «МГТУ
«СТАНКИН»)

Научный консультант – к.т.н., доцент Нежметдинов Р.А.
(ФГБОУ ВО «МГТУ «СТАНКИН»)

Рассматривается подход к организации хранилища для агрегирования теологических дынных, получаемых с систем управления и с устройств, поддерживающих технологии промышленного интернета вещей (англ. Internet of Things). Показаны примеры проектирования и использования хранилища для платформы CNCIoT, разрабатываемой в МГТУ «СТАНКИН» на примере реляционных БД и NoSQL, а также сравнение этих подходов с перспективным переходом использования концепции DataLake.

Разработка и сопровождение промышленных систем мониторинга и предиктивной аналитики предъявляет новые требования при решении задач организации хранилищ данных и реализации сервис-ориентированной архитектуры приложений. Объектная модель хранилища с одной стороны должна быть достаточно гибкой и позволять хранить данные произвольного формата, с другой стороны отвечать требованиям репрезентативности технологического процесса конкретной единицы оборудования и контекстно зависимых процессов связи устройств промышленного интернета вещей, систем мониторинга и систем управления технологического оборудования. Существующий этап развития систем искусственного интеллекта и инструментов математического моделирования позволяет строить модели производственных процессов с не четко выявленными зависимостями протекания процессов. Построение математических моделей производственных систем на сегодняшний момент все больше опирается на данные полученные с технологического оборудования, что привело к появлению таких понятий как цифровые тени и цифровые двойники производственных систем. Этот факт позволяет предположить, в скором будущем предприятия будут следовать подходу Data-driven в принятии решений по созданию или оптимизации внутренних технологических процессов опираясь на анализ генеральных выборок, агрегаций и сэмплированию (от англ. sample — выборка) накопленных данных.

Традиционным решением для организации хранения данных являются реляционные базы данных, однако, для организации больших и отказоустойчивых хранилищ данных на сегодняшний момент используются разнообразные виды NoSQL (от англ. not only SQL — не только SQL) подход. Тип и структуру данных от промышленных систем в целом и от систем, использующих технологию промышленного интернете вещей в частности, невозможно заранее предсказать. Это связано с тем, что для получения данных могут использоваться различные механизмы, такие как: программные коннекторы, ориентированные на конкретный вид оборудования конкретного производителя; REST-сервисы интеграции или прямые подключения к реляционной базе данных; реализация технологии OPC с задаваемой частотой опроса.

Текущие реализации NoSQL баз данных позволяет получить отказоустойчивое решение для заранее неизвестного объема данных. В основе большинства таких баз данных лежит концепция хранения «ключ-значение» на различных уровнях абстракции. Такая реализация накладывает ряд ограничений на манипуляции с данными и реализацию поиска по значению. Это приводит к существенному увеличению времени анализа имеющейся в базе данных информации и сбора статистики. Если рассматривать реляционные модели построения баз данных, то и преимущество заключается в строгой структурированности хранимой информации и широких возможностях языка SQL, которые проявляются при построении сложных аналитических выборок и агрегаций. На текущий момент сотни миллионов строк для таблиц в ре-

ляционных базах данных не являются пределом, однако, для организации эффективного и отказоустойчивого хранилища в реляционной модели необходимо использовать механизмы репликации (англ. replication) и шардирования (горизонтального масштабирования). Ввод в эксплуатацию устройств промышленного интернета вещей порождает нелинейный рост генерации данных в хранилище. Это связано с тем что значительная часть процессов для анализа и воспроизведения требует высокой частоты опроса устройств полевого уровня.

В работе в качестве основного хранилища оперативной информации с технологического оборудования принята open-source реляционная база данных PostgreSQL. В качестве базовой сущности вводится понятие «ноды», как единицы агрегации устройств в рамках процесса. В качестве «ноды» может выступать как отдельный датчик, так и связка систем управления и устройств промышленного интернета вещей, тем самым достигается контекстная зависимость хранения данных. Особенностью устройств промышленного интернета вещей в рамках контекста является разная частота опроса с устройств полевого уровня и времени актуальности собранных данных. К примеру, в рамках процесса могут быть установлены датчики работающие в сети LoRaWAN (максимальная скорость передачи 50 кбит/сек), устройства взаимодействующие по REST API (зависит от установленной политики устройства и Backend Rate Limiting принимающего сервиса) и OPC UA (частота опроса полевого уровня 50 мс). С целью обеспечения безопасности данных и разделения политик доступа к системе необходимо реализовать механизм регистрации устройства. На текущий момент платформа служит для сбора информации и не инициирует запросы, поэтому политика безопасности сводится к регистрации устройства.

Предложенная в работе архитектура требует реализации контракта между устройством сети и веб-сервисами разработанной платформы. В качестве базового синтаксиса был использован универсальный текстовый формат JSON с рядом синтаксических ограничений на обязательность некоторых атрибутов. В качестве важных параметров выделены следующие: идентификатор устройства, группа запроса и ожидаемый формат данных. Группа запроса является отдельной сущностью и позволяет отслеживать такие устройства как системы числового программного управления, отправляющие в запросе большое количество разнообразных данных (например, показания отдельных осей или внутренние коды ошибок). Формат не позволяет задавать вложенные структуры. Вложенность структур в рамках предложенной платформы означала бы несогласованность данных (англ. data consistency) и ошибки в построение аналитических отчетов и представлений. Однако ошибочные запросы к сервисам возможны, ввиду большого разнообразия устройств и регистрируются в том виде в каком они поступили для последующей корректировки и агрегации в рамках инициированного процесса. Для случаев нарушения контракта веб-сервисы уведомляют устройство соответствующим HTTP-кодом и регистрируют на стороне платформы проблемные вызовы соответствующего устройства. Превышение заданного количества допустимых потерь и статусов ошибок запускает процесс нотификации (от лат. notificare — делать известным) заинтересованных устройств.

Текущая реализации платформы имеет ряд программных ограничений на уровне хранилища. Подход строгой структурированности имеет ряд недостатков перед решениями нереляционных баз данных. При вводе в эксплуатацию новых источников данных и аналитики в рамках целого предприятия требуется использование не столько нереляционных баз данных, а подхода на основе концепции Data lake. Использование реляционных баз данных необходимо рассматривать как решение конкретных задач в рамках отдельных производственных единиц оборудования. В рамках Data lake концепции реляционная база данных может рассматриваться как один из источников данных.

Квашнин Д.Ю (автор)

Ковалев И.А. (автор)

Нежметдинов Р.А. (научный консультант)