

УДК 004.75:004.62

АРХИТЕКТУРА ПОТОКОВОЙ ОБРАБОТКИ И ХРАНЕНИЯ ДАННЫХ БИОГАЗОВЫХ УСТАНОВОК НА ОСНОВЕ ТЕХНОЛОГИЙ *BIG DATA*



А.А. Исмаилова

И.о. профессора ГОП «Информационные системы», *Академик Международной академии информатизации, Доктор философии (PhD), ассоциированный профессор (доцент)*
a.a.ismailova@kazatu.kz



Р.Н. Тажиева

*Докторант 1 курса образовательной программы 8D06101 «Аналитика больших данных»,
Магистр Информационных систем*
024.risti.1978@gmail.com

А.А.Исмаилова

Окончила Кызылординский педагогический институт им. Қорқыт Ата, квалификация: учитель физики, информатики, вычислительной техники и программного обеспечения. Научное направление: Анализ, моделирование и проектирование информационной системы (по отраслям), интеллектуальный анализ данных

Р.Н.Тажиева

Окончила Таразский государственный университет имени М.Х.Дулати, квалификация информатик, преподаватель информатики. Научное направление: Анализ, моделирование и проектирование информационной системы (по отраслям), интеллектуальный анализ данных

Аннотация: Построена многоуровневая архитектура системы сбора, потоковой обработки и хранения данных, поступающих от сенсорной сети биогазовой установки. Обоснован выбор между «длинным» и «широким» форматами хранения данных в зависимости от требуемой скорости аналитических данных. Цель работы разработка архитектуры сбора и потока данных на основе Big Data для оптимизации процесса анаэробного сбраживания и повышения выхода метана. Научная новизна заключается в интеграции Big Data к специфике биогазовых процессов, который позволяет интегрировать новые типы датчиков без изменения схемы базы данных.

Ключевые слова: биогазовая установка, Big Data, потоковая обработка данных, IoT, анаэробное сбраживание, архитектура данных, хранилище данных.

Введение. Производство альтернативного источника энергии на основе возобновляемого сырья рассматривается Правительством Республики Казахстана как прорывное направление, позволяющее вывести сельскохозяйственное производство на новый технологический уровень, обеспечивающий рост его эффективности, увеличить экспорт продукции с высокой добавленной стоимостью. Во всех странах в основном содержат скот на молочно-товарных фермах в загонах, отходы, которых загрязняют окружающую среду болезнетворными бактериями, яиц гельминтов и сорными травами земли населенных пунктов, что требуют их утилизации и переработки в биоудобрение и источник энергии [1].

Исследования зарубежных ученых и практиков, в основном направлены на создание биогазовых установок, которых отличает большое количество ежедневного потребления отходов и растительной биомассы, что увеличивает транспортные и иные расходы на сбор, хранение и подготовку, доступное не каждому фермеру [2]. В России, в Казахстане и стран

СНГ из-за высокой начальной стоимости биогазовых установок их общее количество не превышает нескольких сотен. Анализ результатов обследования работы этих установок показывает, что требуются серьезные конструктивные доработки по всей цепочке технологического процесса, обеспечение простоты контроля, доступа к обслуживанию [3].

Функционирование биогазовых установок связано с рядом специфических трудностей [4].

Оптимальный выход биогаза ограничивается нестабильностью процесса, неопределённостью и сложностью контроля параметров анаэробного сбраживания, а также необходимостью мониторинга в режиме реального времени - всё перечисленное относится к числу основных проблем, приводящих к снижению эффективности производства [5].

Традиционные стратегии управления, основанные на ограниченном наборе сенсоров, не обладают достаточной прогностической способностью для своевременного выявления подобных возмущений [6].

В связи с этим оптимизация процессов анаэробного сбраживания требует применения новых подходов, сочетающих мониторинг с высоким пространственно-временным разрешением и передовые аналитические методы, и инструменты поддержки принятия решений.

Современные биогазовые установки представляют собой биогазовые реакторы, эффективность которых в полной мере зависит от поддержания параметров среды (температура, pH, состав субстрата и т.д.).

Интеграция технологий Big Data в управление биогазовыми установками позволяет не только отслеживать текущее состояние биореактора, но предсказывать падение выхода биогаза.

Первым шагом для внедрения аналитики Big Data является построение корректной архитектуры потоков данных.

Архитектура сбора данных с датчиков биогазовой установки.

Архитектура сбора данных – это нижний уровень, который отвечает за физическое получение сигнала от датчиков и его первичную оцифровку. Прежде чем данные окажутся в базе, им предстоит пройти путь от датчика до программного хранилища. Имеется два варианта пути: классический и ориентированный на IoT-инфраструктуру.

Классический вариант подходит для производственного объекта. Датчики подключаются к ПЛК через стандартные промышленные интерфейсы, ПЛК принимает сигналы, обрабатывает их и хранит значения в регистрах, откуда их забирает SCADA или шлюз по протоколам OPC UA или Modbus TCP. По второму IoT-ориентированному варианту умные датчики сами отправляют измерения по проводным или беспроводным протоколам (LoRaWAN, MQTT) прямо на облачный сервер или шлюз (рисунок 1).

Протоколы передачи данных. Работа всей системы зависит от протокола передачи данных. Для мониторинга биогазовой установки можно использовать три основных варианта:

- Modbus TCP – промышленный протокол регистров ПЛК, который применяется для считывания показаний датчиков температуры, давления, оборотов мешалки непосредственно из регистров контроллера:

- OPC UA – современный стандарт промышленной коммуникации, который обеспечивает семантическое описание данных, кросс-платформенное взаимодействие и информационную безопасность:

- MQTT – легкий протокол оптимизированный для IoT – устройств с ограниченными ресурсами.

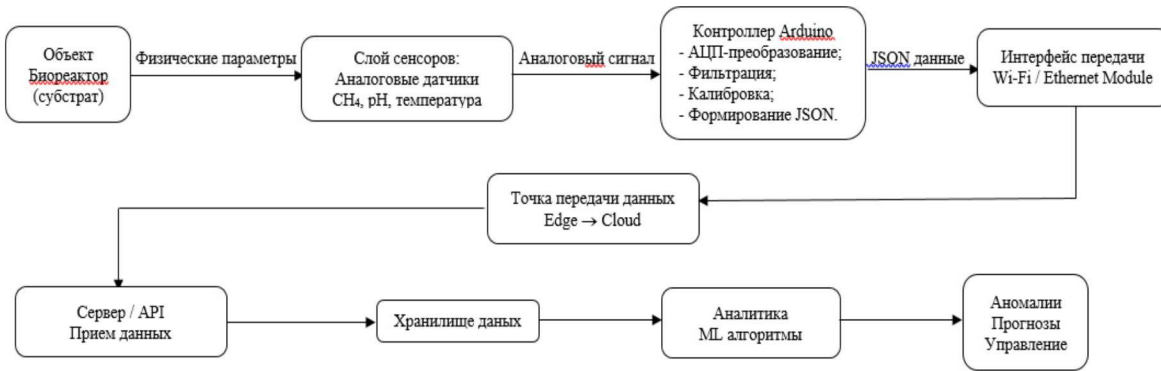


Рисунок 1. Основная схема архитектуры сбора данных

На начале схемы расположены биореактор с установленными сенсорами (датчиками) данных, которые измеряют ключевые технологические параметры, такие как температура реактора, pH среды, уровень сырья, давление газа, концентрация метана и т.д. Датчики формируют поток данных с заданной оператором периодичностью и передают его на устройства сбора. Данные с датчиков поступают в программируемые логические контроллеры. Контроллеры выполняют функции: опрос датчиков с заданной частотой, первичная фильтрация и усреднение данных, преобразование аналоговых сигналов на цифровые сигналы. Для подключения сенсоров к контроллеру в зависимости от типа датчиков.

Архитектура потока данных

Для обеспечения возможности и масштабируемости обработки больших объемов данных предлагается трехуровневая архитектура: уровень восприятия, сетевой уровень и прикладной уровень. Уровень восприятия включает сеть первичных преобразователей (датчиков), установленных непосредственно на узлах подготовки сырья, в ферментаторе и газовых линиях. Локальные контроллеры PLC, осуществляющую первичную оцифровку тоже находятся в этих местах. Сетевой уровень служит связующим звеном между протоколами передачи и шлюзами. Использование протоколов передачи (LoRaWAN, MQTT) позволяет передавать данные на шлюзы (gateways). На этом же этапе реализуется предварительная фильтрация шумов и агрегация данных перед отправкой в центральное хранилище или облако. Данная концепция называется Edge Computing. Прикладной уровень состоит из среды хранения, как MongoDB для NoSQL данных и аналитических инструментов Apache Kafka, Spark. Здесь выполняются алгоритмы Advanced Analytics для оптимизации процессов (рисунок 2).

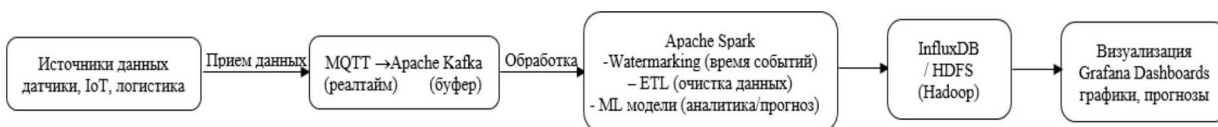


Рисунок 2. Архитектура потока данных

Сенсорные данные с датчиков биогазовой установки в реальном времени передаются через MQTT-брокер, который обеспечивает легкую и энергоэффективную передачу сообщений от датчиков. Центральный элемент принимает данные от всех устройств, маршрутизирует сообщения и обеспечивает асинхронную передачу. Для повышения масштабируемости и надежности данные поступают в Apache Kafka, где формируется очередь сообщений, сохраняется необработанный поток данных и обеспечивается отказоустойчивость системы. Обработка данных выполняется с использованием Apache

Spark в режиме потоковой обработки. На данном этапе реализуется синхронизация событий (watermarking) для корректной работы с задержанными данными и ETL процессы, как очистка, фильтрация и нормализация данных [7]. Для прогнозирования и выявления аномалий применяются модели машинного обучения. Для хранения данных используется комбинированный подход: InfluxDB - для хранения оперативных временных рядов и HDFS (Hadoop) - для долговременного хранения исторических данных. InfluxDB обеспечивает быстрый доступ к актуальным данным, а HDFS (Hadoop) используется для анализа, построения моделей и отчетности. Для визуализации и мониторинга данных используется система визуализации Grafana, которая обеспечивает построение графиков и панелей мониторинга, отображение ключевых показателей установки, настройку оповещений и поддержку принятия решений оператором.

Источники, частота и формат данных

Технология метанового сбраживания позволяет оптимизировать основные параметры технологического режима (водородный показатель, степень измельчения, температура, давление, перемешивание и подача сырья в биореактор (BR)) на всех стадиях процесса анаэробного сбраживания, что обеспечивает максимизацию выхода биогаза и сокращение времени ферментации субстрата в реакторе [8]. К сожалению, в настоящее время непосредственно в режиме реального времени поддается измерению ограниченное число факторов: температура, pH, ОВП, перемешивание в биореакторе, расход и состав получаемого биогаза и расход подаваемого раствора. Некоторые переменные определяются с помощью аналитического оборудования (например, концентрация VFA с помощью газового хроматографа), а остальные с помощью ручного анализа (например, концентрация органических веществ) [9].

На основе анализа литературы [8,9] выделены основные группы параметров: газовый состав, физические параметры и химические показатели.

Газовый состав: содержание метана (CH₄), углекислого газа (CO₂) и сероводорода (H₂S). Примеры анализаторов: MQ-4, MG-811 и промышленные ИК-анализаторы.

Физические параметры: температура субстрата, давление в газгольдере, уровень заполнения биореактора [10].

Химические показатели: pH-уровень (критичен для предотвращения гибели метаногенов), щелочность (TAC), содержание летучих жирных кислот (FOS/VFA), соотношение C/N, ОВП (окислительно-восстановительный потенциал).

Частота и формат данных

Для задач Big Data важен баланс между детализацией и избыточностью данных. Рекомендуемые параметры представлены в таблице 1.

Таблица 1. Категория и частота данных

Категория данных	Источник	Рекомендуемая частота	Формат передачи
Динамические данные (газ/давление)	Газоанализаторы	1 раз в 1-5 минут	MQTT (JSON payload)
Инерционные (температура)	Термопары, термистры	1 раз в 10 - 15 минут	MQTT (Modbus TCP)
Химические (pH/ОВП)	pH-электроды	1 раз в 15 - 30 минут	JSON/CSV (пакетно)
Энергетические (выработка)	Электросчетчики	1 раз в час	Modbus/ JSON

Формат данных: Наиболее предпочтительным для Big Data систем является JSON, благодаря его гибкости при хранении в документированных БД. От организации документа

в коллекции зависит и скорость запросов, и то, насколько легко будет добавить новый датчик через месяц. В практике IoT-систем выделяют два принципиально различных подхода: длинный формат (один замер на документ) и широкий формат (снимок всех датчиков на документ). Выбор между подходами является критически важным архитектурным решением для Big Data в IoT [11].

Суть проста: один документ - одно показание одного датчика. Неважно, сколько датчиков в системе - структура записи остаётся одинаковой:

```
{
  "ts": ISODate("2026-02-17T10:15:30Z"),
  "deviceId": "biogas-reactor-01",
  "sensor": "temperature",
  "value": 37.4,
  "unit": "C",
  "quality": "ok"
}
```

У такого подхода есть очевидные плюсы: добавить новый датчик можно без каких-либо изменений в схеме; агрегировать данные по времени или по типу датчика удобно через единый pipeline; масштабирование при росте числа источников данных не требует миграции схемы.

Альтернатива - собирать в один документ показания сразу всех датчиков за один момент времени. Удобно, когда нужно быстро посмотреть, что происходит прямо сейчас:

```
{
  "ts": ISODate("2026-02-17T10:15:30Z"),
  "deviceId": "biogas-reactor-01",
  "temperature_c": 37.4,
  "pressure_kpa": 102.3,
  "mixer_rpm": 45,
  "gas_flow_m3h": 2.1
}
```

Проблема возникает, если датчики опрашиваются с разной частотой - часть полей будет просто пустой. А если понадобится добавить новый параметр, придётся переписывать все старые документы. Для систем с разнородными датчиками длинный формат надёжнее. Для ясности и сравнения в Таблице 2 приведены основные характеристики двух подходов.

Таблица 2. Сравнение двух подходов формата данных

Характеристика	Длинный формат	Широкий формат
Структура	{sensor: "temp", value: 37, time: T}	{time: T, temp: 37, pH: 7.2, CH4: 60}
Гибкость	Высокая	Ниже
Скорость записи	Быстрее при редких замерах	Эффективнее при одновременном снятии всех показаний
Аналитика	Удобно для поиска по конкретному датчику	Удобно для анализа корреляций

Хранилище данных

Создание стандартизированных протоколов тестирования и открытых баз данных для оценки датчиков способствовало бы более объективной оценке и интеграции с моделями [12]. Вопрос о том, когда целесообразно применять Big Data-платформы - Apache Hadoop, Apache Spark, Flink - вместо традиционного хранилища данных, не имеет однозначного ответа. Он требует анализа сразу четырёх измерений: объёма данных, скорости их поступления, структурной неоднородности и требований к аналитике.

Тем не менее, отвечая на исходный вопрос напрямую: когда именно Big Data-платформы оправданы вместо классического DW - вот четыре сигнала [13]:

- Данных больше 50–100 ТБ и они продолжают расти - вертикальное масштабирование обходится слишком дорого.
- Нужна реакция на данные в режиме реального времени - задержка в часы или даже минуты недопустима.
- Большая часть данных неструктурирована: тексты, логи, JSON, изображения - и их схема заранее неизвестна.
- Аналитика выходит за рамки SQL: машинное обучение, обработка текстов, работа с изображениями - классический SQL тут не поможет. (MSSQL)

Если ни одно из этих четырёх условий не выполняется - классический DW остаётся правильным выбором [14]. Он проще, предсказуемее, лучше интегрируется с BI-инструментами и дешевле в поддержке. Поэтапное название процессов от датчиков до базы данных приведен в Таблице 3.

Таблица 3. Поэтапный план внедрения базы данных

Этап	Наименование	Состав работ	Инструментарий
1	Инвентаризация тегов	Составление реестра датчиков: имя тега, ед. изм., диапазон, частота	Excel / YAML-файл
2	Выбор протокола	Настройка Modbus TCP / OPC UA для чтения регистров ПЛК	Node-RED Modbus-узел
3	Настройка сборщика	Поток: «опрос ПЛК → формирование документа → публикация в MQTT»	Node-RED, Mosquitto
4	Создание коллекции	Создание Time Series коллекции, индексов, TTL-политики	MongoDB Shell / Compass
5	Запись данных	Подписка на MQTT, запись в sensor_readings через драйвер	Node-RED MongoDB-узел
6	Валидация запросов	График температуры за сутки; расчёт средних за 10 мин	MongoDB Compass, Python
7	Визуализация	Подключение Grafana к MongoDB, настройка дашборда	Grafana + MongoDB плагин

Мониторинг процессов анаэробного сбраживания зависит от точности сенсоров, однако, используемые в настоящее время устройства, имеют ряд эксплуатационных недостатков. Электрохимические сенсоры, такие как pH-электроды, широко применяются, но требуют частой калибровки и очистки поверхности для поддержания точности. На базе Института информационных и вычислительных технологий, построена биогазовая установка, включающая 8 полноценных биореакторов объемом 30 кубометров, объединенных в единую технологическую цепочку. Разработанная в нем интеллектуальная система управления качеством биогаза представляет собой интегрированное решение для автоматизированного мониторинга и контроля параметров производства биогаза. Система включает в себя следующие ключевые функции: сбор данных о параметрах качества биогаза; передача данных через Интернет с использованием беспроводных сетей; обработка поступающих данных с датчиков в режиме реального времени; запись и хранение информации в облачной базе данных; отображение параметров системы через веб-интерфейс для удаленного пользователя. Такая система обеспечивает не только мониторинг, но и интеллектуальное управление технологическим процессом, повышая стабильность, безопасность и эффективность работы биогазовой установки [15].

Описанные примеры демонстрируют потенциал использования IoT в системах мониторинга, а также указывают на важность дальнейшего развития в направлении интеллектуализации таких решений. Эти разработки могут быть адаптированы для создания системы управления качеством биогаза, в которой датчики будут отслеживать параметры.

Заключение. Формирование четкой архитектуры и стандартизация сенсорных данных – необходимое условие для перехода к «Индустрии 4.0» в биоэнергетике. Использование протокола MQTT и формата обеспечивает легкую интеграцию с современными аналитическими платформами, что позволяет оптимизировать IT-решения для управления биогазовыми установками в режиме реального времени.

Виртуальные сенсоры, интегрирующие электрохимические, микробиологические, оптические и гибридные системы, обеспечивают адаптивное управление процессом в реальном времени и мониторинг стабильности анаэробного сбраживания. Дальнейшее развитие интеллектуальных систем анаэробного сбраживания требует создания более надёжных сенсоров, стандартизации открытых наборов данных и повышения интерпретируемости моделей искусственного интеллекта.

В сочетании с новыми сенсорными технологиями и концепциями мягких сенсоров ИИ имеет потенциал сделать АД более предсказуемым, устойчивым и ресурсосберегающим, что соответствует принципам циркулярной экономики и текущему энергетическому переходу.

Список литературы

- [1] Karayeva, Yu. V. Review of biogas technologies and intensification methods processes of an anaerobic fermentation. *Akademenergo Works*; **2010**, 3, 109-127.
- [2] Tlebayev, M.B., Taszhurekova, Z.K., Biibosunov, B.I., Baizharikova, M.A., Aitbayeva, Z.K. (2020). Creation of a computer-assisted mathematical model for the raw materials biological processing. *Periódico Tchê Química*, 17(35): 640-654A. <https://doi.org/10.1007/s13399-023-03812-x>
- [3] Тлебаев М.Б., Тажиева Р.Н., Айтбаева З.К., Нуржигитова Ж.Н. Разработка и проектирование трехступенчатой высокоскоростной биогазовой установки. *ИЗВЕСТИЯ ВУЗОВ КЫРГЫЗСТАНА* № 5, 2016
- [4] Ziganshin, A. M., Schmidt, T., Lv, Z., Liebetau, J., Richnow, H. H., Kleinsteuber, S., & Nikolausz, M. (2016). Reduction of the hydraulic retention time at constant high organic loading rate to reach the microbial limits of anaerobic digestion in various reactor systems. *Bioresource Technology*, 217, 62–71.
- [5] Ukpai, P. A., & Nnabuchi, M. N. (2012). Comparative study of biogas production from cow dung, cow peel and cassava peeling using 45 litres biogas digester. *Advances in Applied Science Research*, 3, 1864-1869. <http://edn.link/cm2nmf>
- [6] Cinar, S.; Cinar, S.O.; Wieczorek, N.; Soho, I.; Kuchta, K. Integration of Artificial Intelligence into Biogas Plant Operation. *Processes* 2021, 9, 85 <https://doi.org/10.3390/pr9010085>
- [7] Zaharia M. et al. Apache Spark: A Unified Engine for Big Data Processing // *Communications of the ACM*. –2016. –Vol. 59, No. 11. –P. 56–65.
- [8] Baizharikova, M., Tlebayeva, G., Tlebayev, M., Beksultanov, Z., & Tazhiyeva, R. (2024). Development of a new innovative technological process for continuous methane fermentation in a three stage biogas plant. *Biomass Conversion and Biorefinery*, 14, 29219. <https://doi.org/10.1007/s13399-023-03812-x>
- [9] Simeonov, I., & Hubenov, V. (2025). Application of Artificial Intelligence for Prediction, Monitoring, Optimization and Control of Anaerobic Digestion Processes—A Review. *Processes*, 13(12), 3812. <https://doi.org/10.3390/pr13123812>
- [10] Deublein, D., & Steinhauser, A. (2011). *Biogas from Waste and Renewable Resources*. Publisher: WILEY VCH ISBN: 978-3-527-32798-0 <https://doi.org/10.1002/9783527632794>
- [11] Gaida, D., Wolf, Ch., State estimation for anaerobic digesters using the ADM1 July 201266(5):1088-95 <https://doi.org/10.2166/wst.2012.286>
- [12] Marycz, M., Turowska, I., Glazik, S., & Jasiński, P. (2024). Artificial intelligence in anaerobic digestion: A review of sensors, modeling approaches, and optimization strategies. *Applied Sciences*, 14(4), 1612. <https://doi.org/10.3390/app14041612>
- [13] Armbrust M. et al. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics // *CIDR 2021: 11th Annual Conference on Innovative Data Systems Research*. - 2021.
- [14] Kimball R., Ross M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd ed. - Indianapolis: Wiley, 2013.
- [15] Development of an IoT System for Monitoring the Operation of a Biogas Plant <https://doi.org/10.20944/preprints202507.0368.v1>

Авторский вклад

Исмаилова Айсулу Абжаппаровна – постановка задачи исследования, разработка архитектуры потока и хранения данных, руководство исследованием.

Тажиева Рысты Нарбаевна – постановка задачи исследования, разработка архитектуры потока и хранения данных, руководство исследованием.

STREAM PROCESSING AND DATA STORAGE ARCHITECTURE FOR BIOGAS PLANTS BASED ON BIG DATA TECHNOLOGIES

Ismailova Aisulu

*Acting professor of the GEP «Information Systems», Academician of the International academy of informatization PhD, Associate Professor (docent),
a.a.ismailova@kazatu.kz*

Tazhiyeva Rysty

*First-year PhD student in the educational program 8D06101 «Big Data Analytics»,
Master's in Information Systems,
024.risti.1978@gmail.com.*

Abstract: A multi-layered architecture for the collection, stream processing, and storage of data received from a biogas plant sensor network has been constructed. The choice between "long" and "wide" data storage formats is justified depending on the required speed of analytical data processing. The objective of the work is to develop a data collection and streaming architecture based on Big Data to optimize the anaerobic digestion process and increase methane yield. The scientific novelty lies in the integration of Big Data with the specific requirements of biogas processes, which allows for the integration of new sensor types without modifying the database schema.

Keywords: biogas plant, Big Data, stream processing, IoT, anaerobic digestion, data architecture, data warehouse.