

## **СРАВНИТЕЛЬНЫЙ АНАЛИЗ РЕШЕНИЙ СРЕДИ РЕЕСТРОВ СХЕМ**

**Бескровный Юрий Владимирович**

магистрант, кафедра информатики и информационных технологий, Калужский государственный университет им. К.Э. Циолковского, РФ, г. Калуга

**Аннотация.** В 2022 году из-за начала СВО, многие зарубежные IT-вендоры завершили свою работу на территории России, и ушли с нашего рынка. В одних случаях этот уход достаточно сильно повлиял на IT-ландшафт, как это происходит с решениями компаний Miro, Jetbrains, SAP, в других же случаях компании перешли на отечественные аналоги без каких-либо проблем, как это произошло с решениями Microsoft, Google, Amazon.

В рамках данной статьи будет рассматриваться вопрос контроля контрактов схем данных различных систем в разрезе очередей данных. Помимо этого, будет произведен анализ существующих решений для контроля эволюции схем данных согласно их совместимости.

**Ключевые слова:** сравнительный анализ, информационные системы, реестры схем, версионирование, совместимость схем данных, Confluent Schema Registry.

**Введение.** Современные информационные системы активно обмениваются между собой данными.

При этом во многих сценариях содержимое сообщений структурировано согласно определенной спецификации схемы.

Как отправляющая сторона, так и получатель хотят иметь гарантии, что сообщения являются целостными и не нарушают ранее оговоренного формата обмена, однако это достаточно сложная задача, т.к. современные бизнес-процессы меняются с большой скоростью.

Из-за этого формат схемы может меняться достаточно часто, а ИС должны уметь достаточно быстро реагировать на эти изменения не теряя надежности и предсказуемости в своей работе.

Для решения описанной проблематики, в области обмена информацией между информационными системами, существует механизм под названием реестр схем (schema registry), который и занимается обеспечением целостности и согласованности формата данных в распределенных средах.

Реестр схем представляет собой независимый процесс, который работает поверх брокера сообщений Kafka.

Он используется для хранения и получения описания схем в различных форматах, будь это широкоиспользуемый формат JSON, или же Avro/Protocol buffers схема.

Реестр схем достаточно важный центральный компонент больших ИС, т.к. позволяет настроить контроль над процессом проверки целостности при обмене данными, а также выстроить понятный процесс по эволюции схем данных и версионированию артефактов.

Процесс эволюции схем является центральным в работе реестра схем, т.к. добавление и удаление полей в схеме неотвратимо и сам процесс подвержен человеческим ошибкам, а потому должен быть автоматизирован. Хотя версионирование событий позволяет избавить разработчиков от досадных ошибок, но не решает полностью вопрос эволюции схем данных.

Реестр схем часто упоминается в связке с брокерами сообщений, а именно Kafka. Действительно, этот инструмент первоначально и был разработан для контроля контрактов между событийно-ориентированными системами.

При этом, у таких решений есть потенциал вырасти до самостоятельных решений, которые будут работать и с другими брокерами сообщений.

На данный момент, рынок схем данных имеет ряд реализаций от крупнейших облачных провайдеров, таких как Google Cloud Schema Registry, Microsoft Azure Schema Registry, AWS Glue Schema Registry, Confluent Schema Registry. Три из четырех перечисленных решения представляют собой проприетарное решение, которым можно пользоваться по подписке, что в одном случае может сократить затраты на внедрение и оперирование в компании, которая хочет использовать это решение, но при этом заставляет клиентов сделать vendor lock из-за тесной связи с ИС.

В этом списке особняком стоит решение от компании Confluent. В мире событийно-ориентированных архитектур именно Confluent Schema Registry является передовым решением, которое имеет продуманную архитектуру, высокую масштабируемость и отличную производительность. При этом, это единственное решение, которое имеет как подписочное решение, так и распространяется под свободной лицензией.

Именно за счет свободной лицензии данный программный продукт широко используется при разработке комплексных решений у других продуктов облачных провайдеров.

Особую важность в поддержке любых систем является сложности при ее эксплуатации. Потому выбор какой-то либо ИС при внедрении в IT-ландшафт компании нужно делать осознанно.

Каждая ИС имеет свои особенности, при этом риск иметь отказ одного из важнейших узлов системы может перевесить любые другие плюсы этого решения.

Решение от компании Confluent имеет ряд своих плюсов, так и минусов, ровно, как и любое другое решение.

Для того, чтобы понять насколько стоит использовать это решение в своем IT-продукте, проведем анализ этой системы в сравнении с другими важными конкурентами на IT-рынке.

*Таблица.*

**Сравнение информационных систем для контроля схем данных**

|                                       | Google Schema Registry   | Azure Schema Registry     | Confluent Schema registry          | Сd reg   |
|---------------------------------------|--------------------------|---------------------------|------------------------------------|----------|
| Регистрация схем                      | Есть                     | Есть                      | Есть                               |          |
| Валидация схем                        | Есть                     | Есть                      | Есть                               |          |
| Механизм эволюции схем                | Есть, появилась недавно  | Есть, появилась недавно   | Есть                               |          |
| Уровни совместимости версий           | Прямая и обратная        | Прямая и обратная         | Прямая и обратная                  | Пр       |
| Поддерживаемые форматы                | Avro, Protocol Buffers   | Avro, JSON                | Avro, Protocol Buffers, JSON       | Avro     |
| Архитектурная сложность и зависимости | Зависимость от Pub/Sub   | Зависимость от Events Hub | Зависимость от Kafka               | Зави Zoo |
| Сложность сопровождения               | Простая. Этим занимается | Простая. Этим занимается  | Простая. Этим занимается поставщик | Слоп     |

|                          | поставщик платформы                                 | поставщик платформы                                    | платформы  |                           |
|--------------------------|---|--|--|---------------------------|
| Устойчивость к ошибкам % | 99,95   | 99,9   | 99,99  | Зависит от команды        |
| Мобильность              | Работает исключительно в Google Cloud (vendor lock) | Работает исключительно в Microsoft Azure (vendor lock) | Можно переехать в другие Confluent совместимые Cloud решения | Полная                    |
| Открытость               | Проприетарное решение                               | Проприетарное решение                                  | Распространяется под свободной лицензией                     | Распространяется свободно |

Из приведенного в таблице сравнения видно, что решение от компании Confluent не зря занимает лидирующую позицию среди решений по работе со схемами данных. В отличие от конкурентов, компания распространяет свое решение бесплатно, чем только увеличивает его популярность. Данный маркетинговый ход только на руку компании, т.к. многие специалисты при работе с данной платформой могут понять плюсы этой платформы и решить перейти на cloud подписку, при этом сохраняя возможность перевезти инфраструктуру в другое облако, или, при крайней необходимости, перейти на self-hosted решение. Это хорошо соотносится с подходом множества компаний по сведению потенциальных рисков к минимуму, при отказе кого-то из партнеров, услуги которых используются в разработанной информационной системе. Потому, отсутствие возможности развернуть локальную версию решения является весомым недостатком для решения.

**Заключение:** в современном технологическом ландшафте, где данные играют ключевую роль в принятии решений и формировании стратегий, реестры схем становятся незаменимым инструментом для эффективного управления и использования данных. Confluent Schema Registry, как один из лидирующих реестров схем, предлагает мощный набор функций для централизованного управления схемами, обеспечивая высокую масштабируемость, производительность и безопасность.

Вопрос сложности выбора компонентов ИС особо остро сейчас стоит, когда наши организации сталкиваются с возрастающим объемом и сложностью данных и при этом имеют риски потерять ключевой компонент своей инфраструктуры из-за геополитических решений.

Хоть период турбулентности еще не пройден, но уже сейчас многие отечественные программные продукты активно разрабатываются и заменяют именитые зарубежные ИС. Можно уверенно сказать, что реализация отечественной версии реестра схем это только вопрос времени.

### Список литературы:

1. Наркхеде Н., Шапира Г., Палино Т. (2019). Kafka: Полное руководство [Электронный ресурс] / Москва: ДМК Пресс, 2019. – 123с.
2. Чандрамули Б., Левандоски Дж. Эволюция схем в Apache Kafka [Электронный ресурс] / Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data / 2019.
3. Официальная документацию по использованию Confluent Schema Registry // Confluent – [Электронный ресурс]. – Режим доступа: <https://docs.confluent.io/platform/current/schema-registry/index.html> (дата обращения: 21.05.2024).
4. Школа больших данных // ВШЭ – [Электронный ресурс]. – Режим доступа: <https://bigdataschool.ru/blog/kafka-confluent-schema-registry-under-the-hood.html> (дата обращения 27.05.2024).
5. Официальная документацию по использованию Confluent Schema Registry // Confluent –

[Электронный ресурс]. – Режим доступа: <https://docs.confluent.io/platform/current/schema-registry/index.html> (дата обращения: 21.05.2024).