



debezium

ОКНО В АСИНХРОННЫЙ МИР ДАННЫХ

Знакомство с инструментом Debezium, который позволяет обрабатывать изменения данных в режиме реального времени, открывая новые возможности для асинхронной архитектуры приложений.



ЕВГЕНИЙ ЕФИМЕНКО

Senior Java разработчик

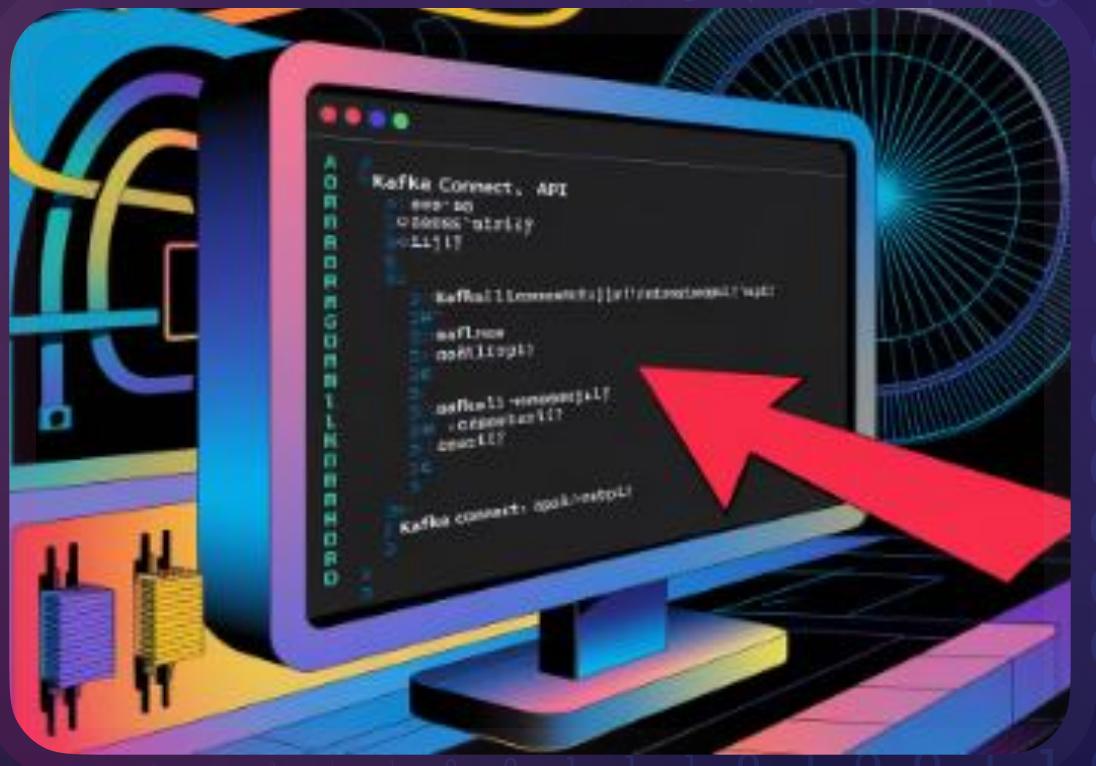
За 2 года прошел путь от Junior до Senior, став примером для многих коллег. Моя работа охватывает разные домены: CRM, Gambling, Fintech, Blockchain и smart contracts.

Опыт и специализация

- Оптимизация систем
- Особый интерес вызывают технологии, которые лежат на стыке разных команд, разных направлений.



СОДЕРЖАНИЕ



Что такое Kafka Connect?



Репликация и Debezium

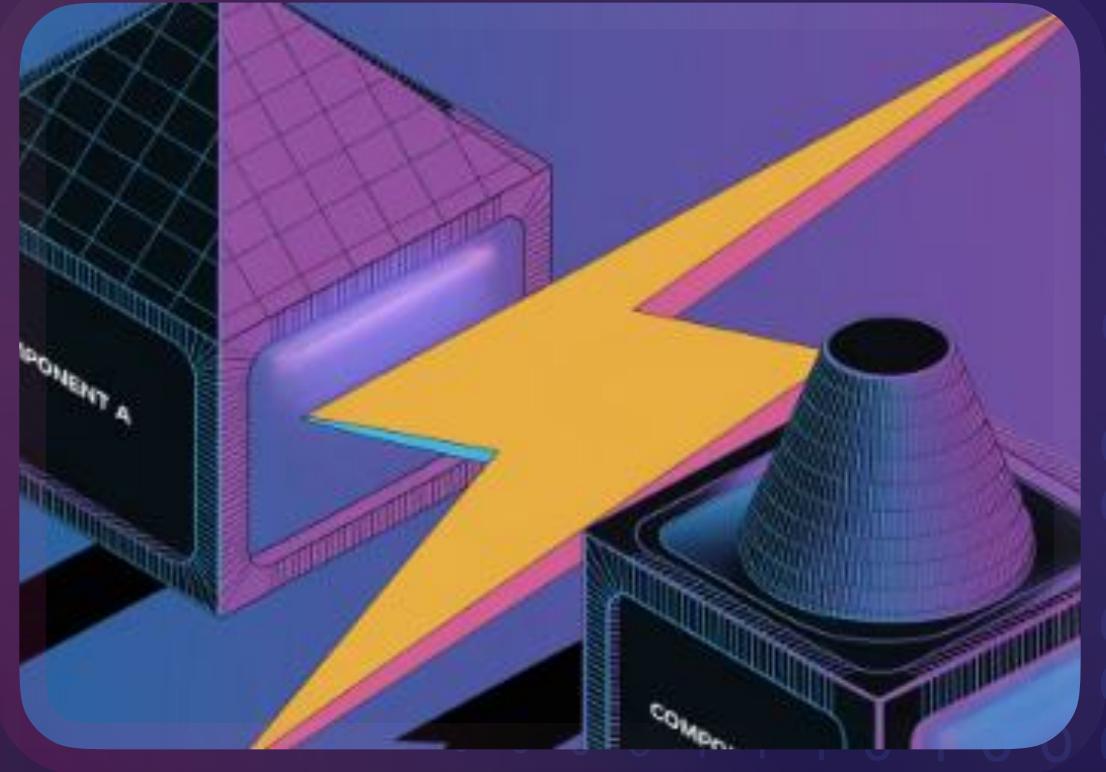


Настройка и поднятие Debezium



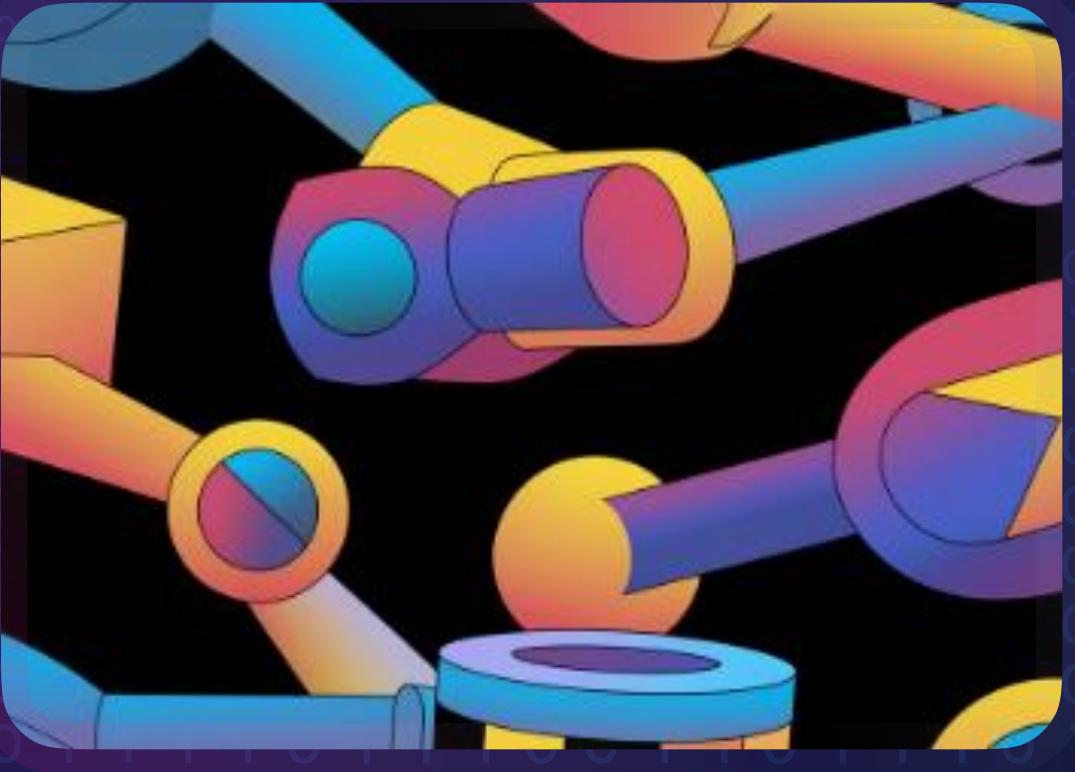
Мониторинг Debezium

EVENT-DRIVEN ARCHITECTURE



Асинхронное взаимодействие

Компоненты системы работают независимо, что позволяет масштабировать их без взаимных зависимостей.



Слабая связность

Компоненты системы не зависят друг от друга, что улучшает модульность и снижает связанность кода.



Устойчивость

Система продолжает работать даже при временных сбоях, так как события могут быть обработаны позже.



Масштабируемость

Система легко справляется с ростом нагрузки, добавляя новые обработчики или распределяя их по серверам.

ОСНОВНЫЕ КОНЦЕПЦИИ АРАСНЕ КАФКА



Publish and Subscribe

Kafka позволяет публиковать сообщения в темы, а потребители подписываются на эти темы, чтобы получать данные.



Persistent Storage

Kafka хранит данные в распределенном виде, выступая в роли устойчивой очереди сообщений.



Real-Time Processing

Kafka обрабатывает потоки данных в реальном времени, позволяя приложениям реагировать на события.

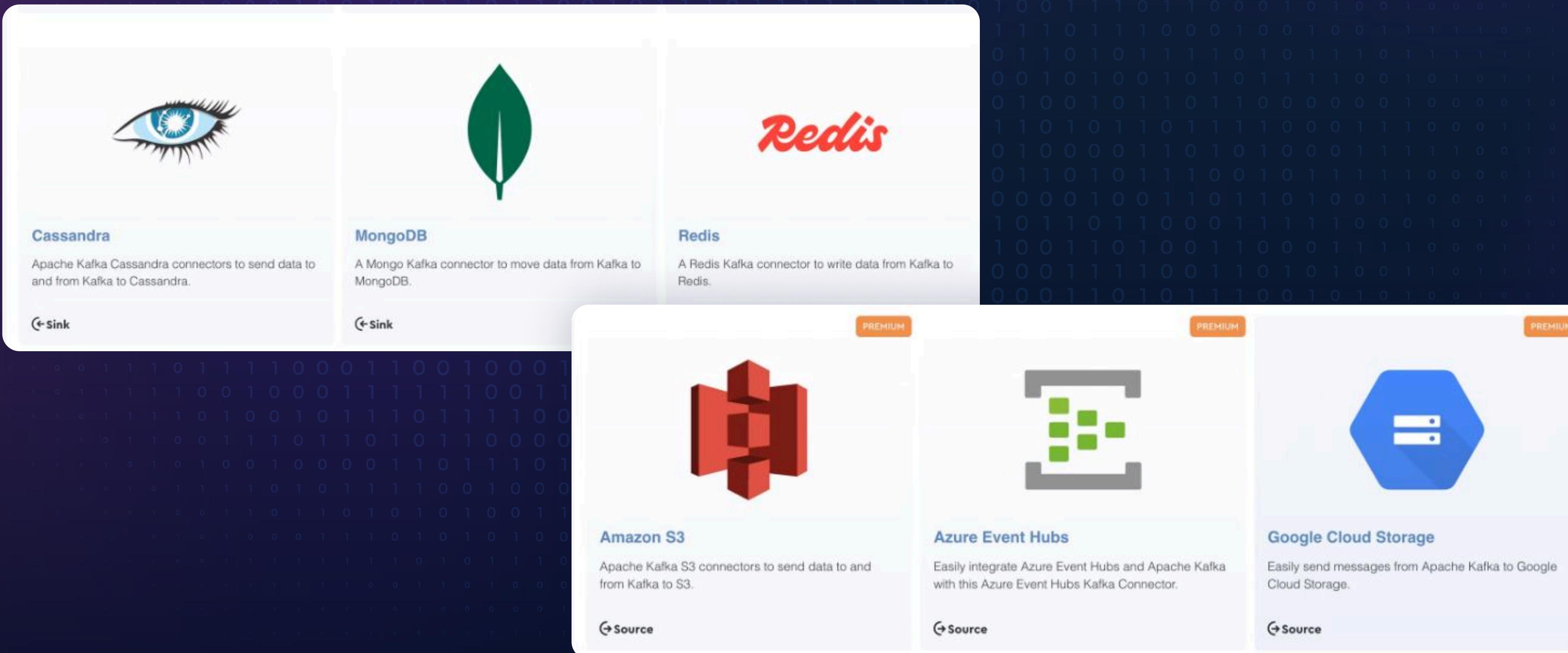


Integration

Kafka интегрирует различные системы, позволяя передавать данные между ними.

ЧТО ТАКОЕ KAFKA CONNECT?

Kafka Connect — это инструмент или фреймворк для масштабирования и надежного стриминга данных между Apache Kafka и другими системами.



DEBEZIUM CONNECTOR



Широкий спектр источников

Debezium предоставляет коннекторы для различных баз данных, включая MySQL, PostgreSQL, MongoDB и другие популярные СУБД.



Простота внедрения

Коннекторы Debezium просты в настройке и позволяют быстро интегрировать изменения данных в ваши приложения.

РЕПЛИКАЦИИ И DEBEZIUM



Журнализование операций

Обеспечивает надежность
и безопасность данных.



CDC (Change Data Capture)

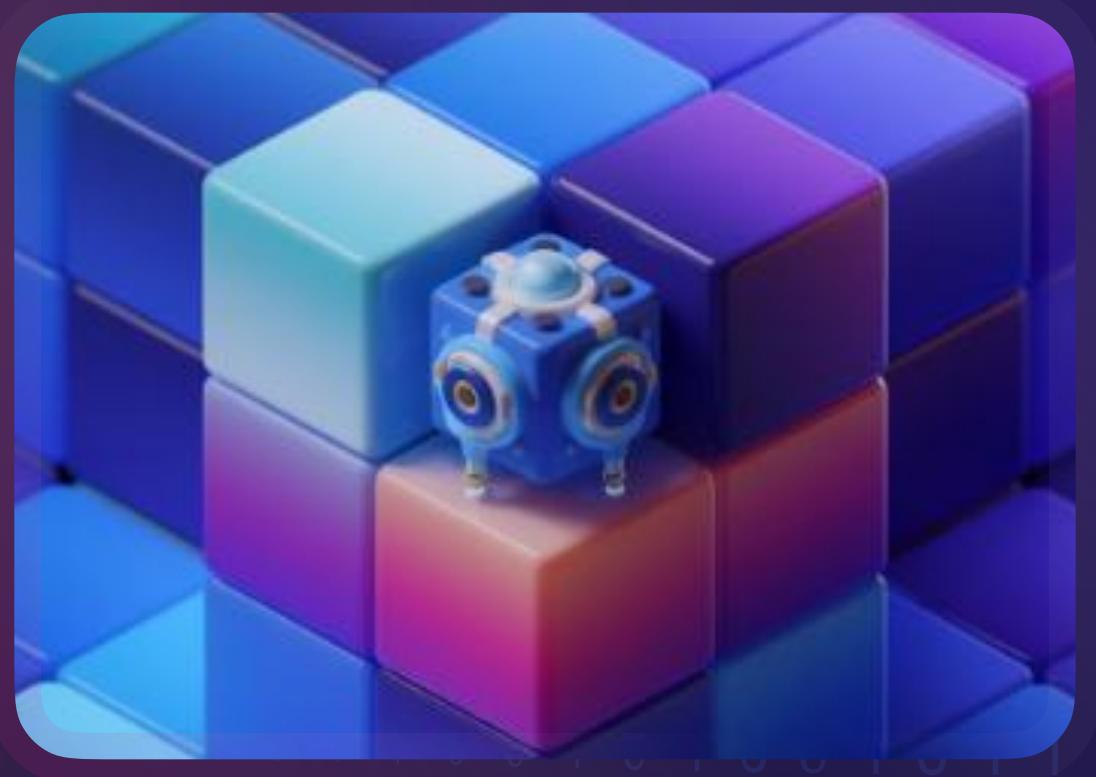
Отслеживает изменения данных
в источнике и передает их в другое место.



Репликации

Синхронизирует данные, обеспечивая
согласованность между системами.

ТРАНСФОРМАЦИИ



ExtractField

Извлекает определенное поле из структуры данных, включая только его в результат.



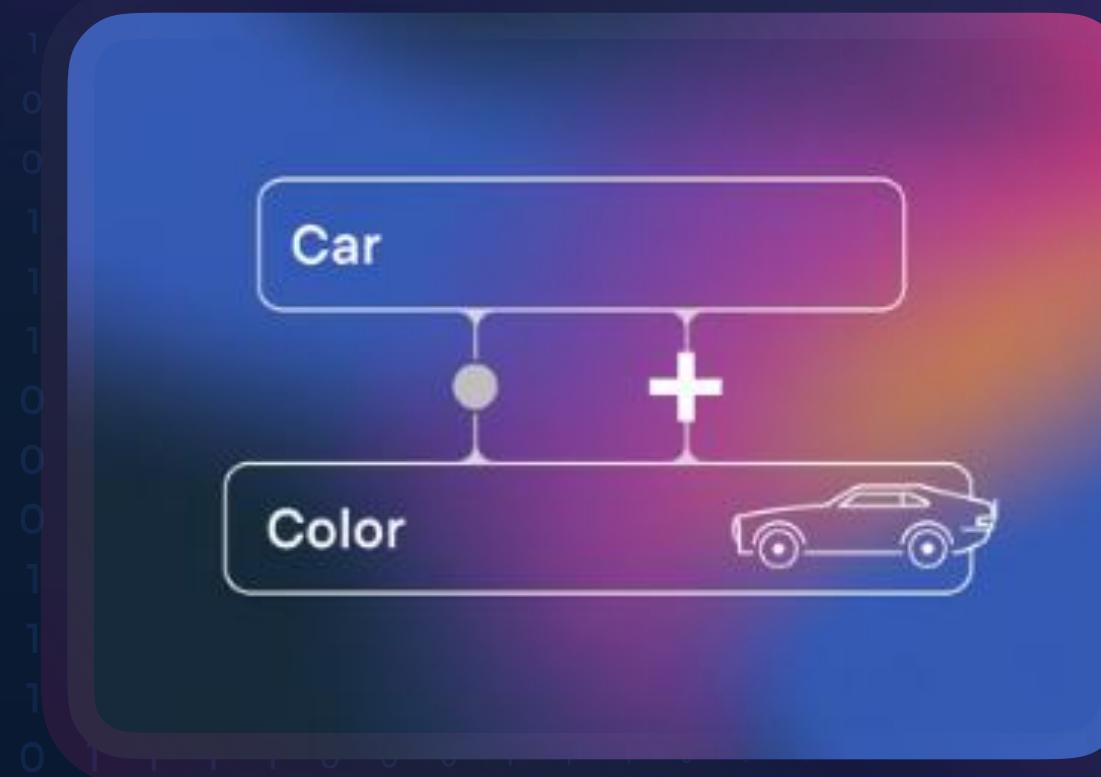
Filter

Удаляет сообщения из дальнейшей обработки на основе предикативных условий.



Flatten

Преобразует вложенную структуру данных в плоский формат.



InsertField

Добавляет новые поля в данные, используя статические значения или метаданные.

SNAPSHOTS

1

Always

2

initial_only

3

never(deprecated)

4

configuration_based

5

initial

6

no_data

7

when_needed

8

custom

ОСНОВНЫЕ ПРОБЛЕМЫ РАСШИРЕНИЯ ТАБЛИЦ



Невозможность добавления дополнительных таблиц

Практически невозможность добавления дополнительных таблиц в список захваченных таблиц, если существующие данные должны быть переданы потоком.

Длительный процесс создания моментальных снимков

Длительный процесс последовательного создания моментальных снимков, который нельзя прервать или возобновить.

Блокировка потоковой передачи данных

Блокировка потоковой передачи данных об изменениях до завершения моментального снимка.

ПОЧЕМУ INCREMENTAL SNAPSHOTS?



Сокращение времени создания моментального снимка

Incremental Snapshots позволяют значительно сократить время создания моментального снимка, что особенно важно при работе с большими объемами данных.



Непрерывность работы приложения

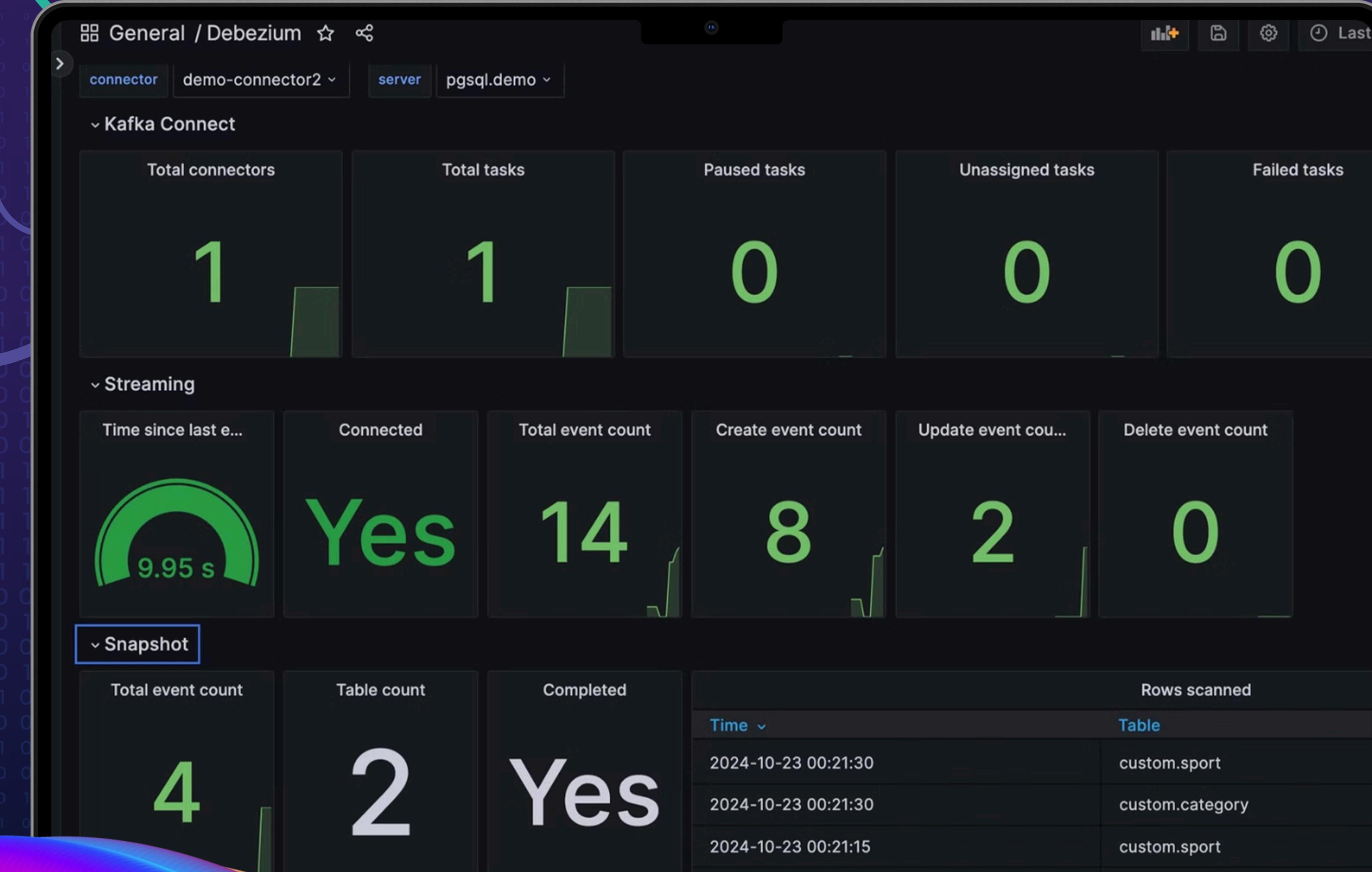
Incremental Snapshots не блокируют потоковую передачу данных об изменениях, что обеспечивает непрерывность работы приложения во время создания моментального снимка.



Легкое расширение списка захваченных таблиц

Incremental Snapshots позволяют легко расширить список захваченных таблиц, добавляя новые таблицы в существующий поток данных.

МОНИТОРИНГ



ВОПРОСЫ?

У вас есть вопросы об использовании Debezium и Kafka для обработки потоковых данных в режиме реального времени?

Спрашивайте, и мы постараемся на них ответить.



ИСТОЧНИКИ

[Документация Debezium 3.0](#)

[Блог Debezium: Инкрементные моментальные снимки](#)

[Документация Debezium: Моментальные снимки PostgreSQL](#)

[Документация Debezium: Сигналы для коннекторов](#)



НА СВЯЗИ



Telegram



LinkedIn



Github