

# 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**

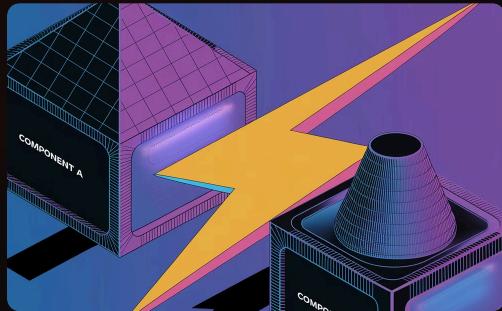


**Настройка и поднятие Debezium**



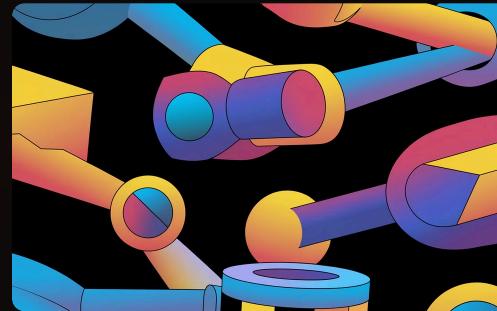
**Мониторинг Debezium**

# Event-Driven Architecture



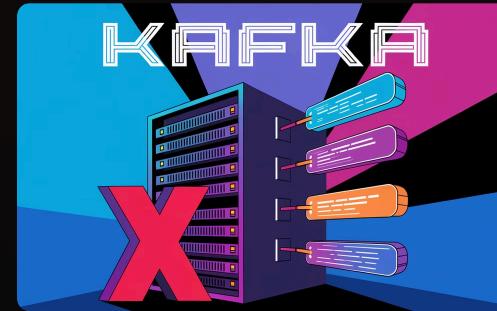
## Асинхронное взаимодействие

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 Apache Kafka



## 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.



**Cassandra**

Apache Kafka Cassandra connectors to send data to and from Kafka to Cassandra.

← Sink



**MongoDB**

A Mongo Kafka connector to move data from Kafka to MongoDB.

← Sink



**Redis**

A Redis Kafka connector to write data from Kafka to Redis.

← Sink



**Amazon S3**

Apache Kafka S3 connectors to send data to and from Kafka to S3.

→ Source



**Azure Event Hubs**

Easily integrate Azure Event Hubs and Apache Kafka with this Azure Event Hubs Kafka Connector.

→ Source



**Google Cloud Storage**

Easily send messages from Apache Kafka to Google Cloud Storage.

→ Source

# 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



## Журналирование 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.



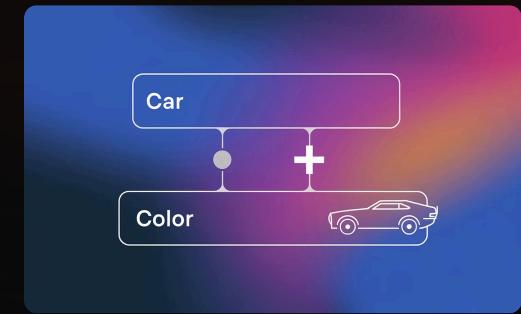
## Filter

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`

**3** `initial_only`

**4** `no_data`

**5** `never(deprecated)`

**6** `when_needed`

**7** `configuration_based`

**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

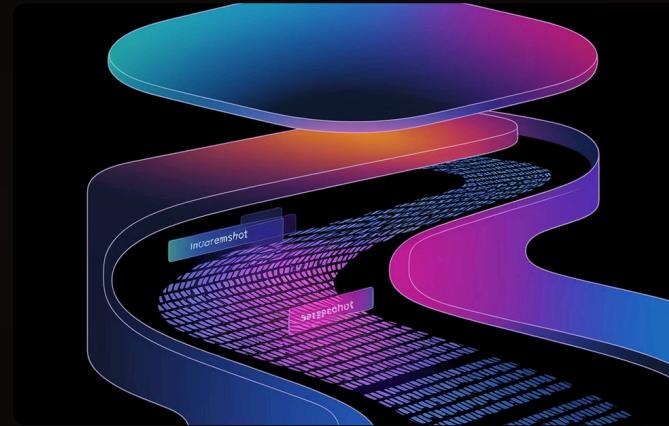
Блокировка потоковой передачи данных об изменениях до завершения моментального снимка.

# Почему Incremental Snapshots?



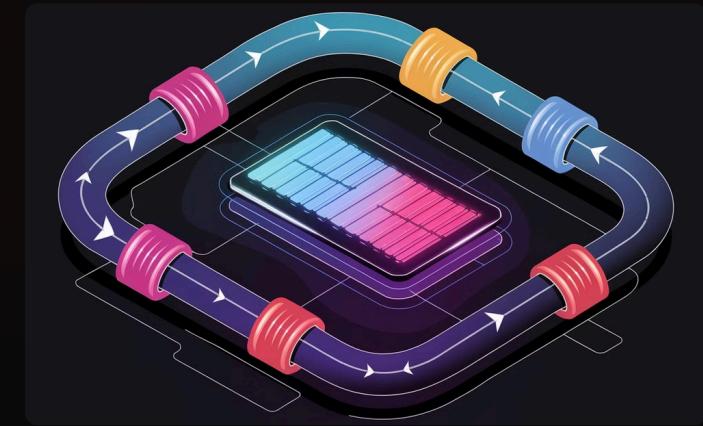
## Сокращение времени создания моментального снимка

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

General / Debezium ⚡ 🔍

Last 15 minutes ▾ 5s ▾

connector demo-connector2 server psql.demo

✓ Kafka Connect

Total connectors	Total tasks	Paused tasks	Unassigned tasks	Failed tasks
1	1	0	0	0

✓ Streaming

Time since last e...	Connected	Total event count	Create event count	Update event cou...	Delete event count
9.95 s	Yes	14	8	2	0

✓ Snapshot

Total event count	Table count	Completed	Rows scanned
4	2	Yes	Time ▾
			Table
			Rows
			2024-10-23 00:21:30
			custom.sport
			2
			2024-10-23 00:21:30
			custom.category
			2
			2024-10-23 00:21:15
			custom.sport
			2

# 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](#)

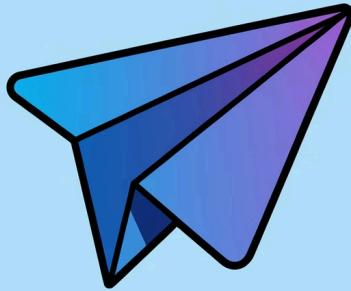
[Блог Debezium: Инкрементные моментальные снимки](#)

[Документация Debezium: Моментальные снимки PostgreSQL](#)

[Документация Debezium: Сигналы для коннекторов](#)



## На связи



[Telegram](#)



[LinkedIn](#)



[GitHub](#)