



Отчёт об изучении Kafka

Выполнил Загребельный Александр



ОСНОВЫ

Брокер сообщений — архитектурный паттерн в распределённых системах; приложение, которое преобразует сообщение по одному протоколу от приложения-источника в сообщение протокола приложения-приёмника, тем самым выступая между ними посредником.

ОСНОВЫ

Брокеры сообщений обычно состоят из трёх базовых компонентов:

1. сервер;
2. продюсеры, которые отправляют сообщения в некую именованную очередь, заранее сконфигурированную администратором на сервере;
3. консьюмеры, которые считывают те же самые сообщения по мере их появления.

ОСНОВЫ

Kafka — гибридный распределенный брокер сообщений с возможностью горизонтального масштабирования. **Kafka** собирает данные у приложений, хранит их в своем распределенном хранилище, группируя по топикам, и отдает компонентам приложения по подписке. При этом сообщения хранятся на различных узлах-брокерах, что обеспечивает высокую доступность и отказоустойчивость.

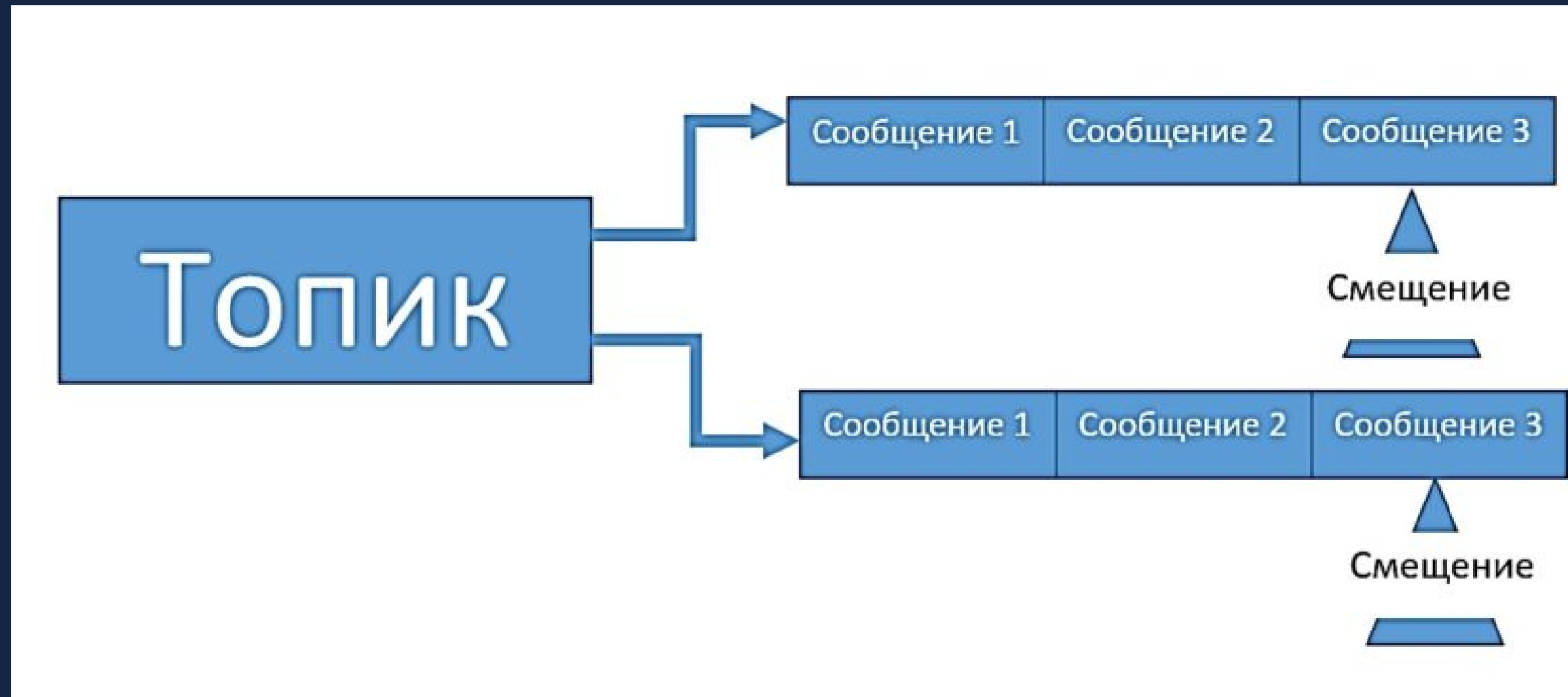
ОСНОВЫ

Топик (topic) – это элемент Kafka, который объединяет все сообщения, относящиеся к одной тематике. Прежде, чем начать обмен сообщениями, необходимо создать **ТОПИК**, в котором эти сообщения будут храниться для последующей их рассылки. Все сообщения, которые хранятся в топике, образуют очередь.

ОСНОВЫ

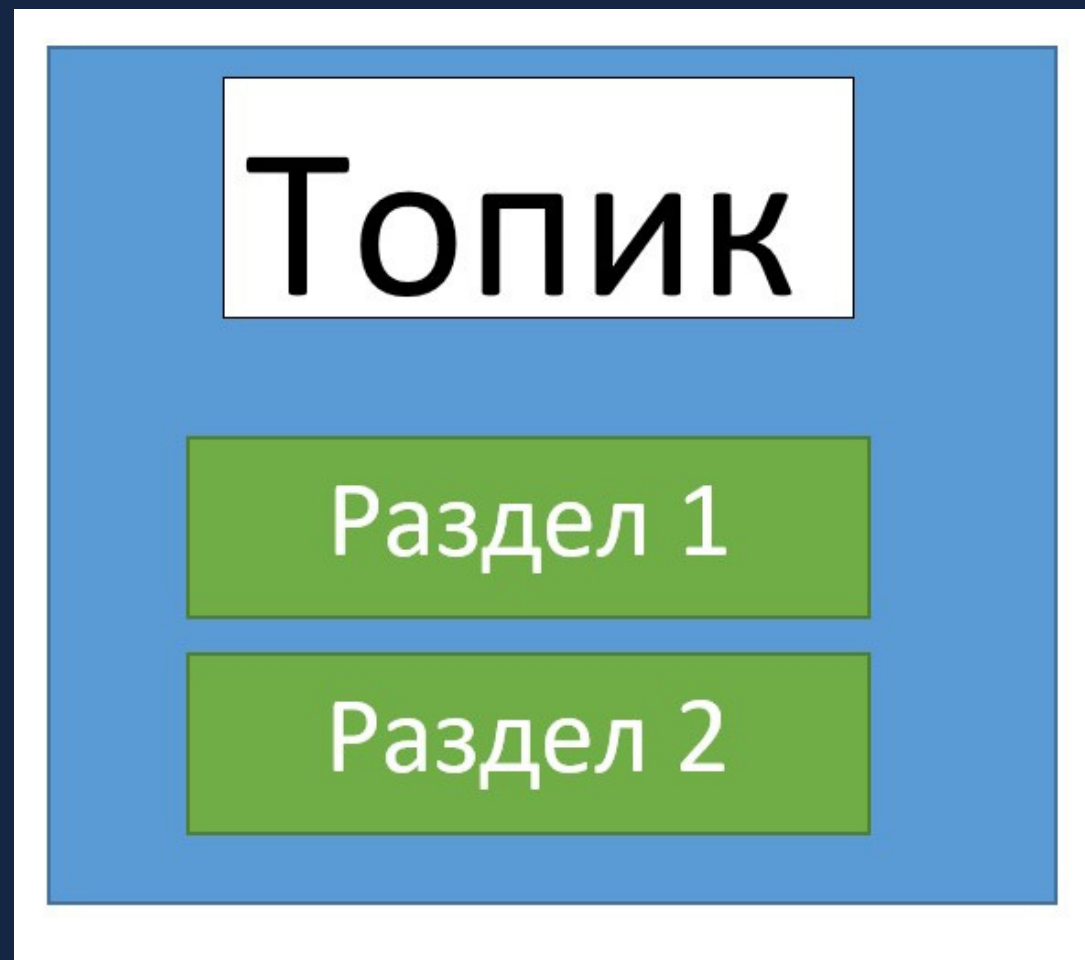
Для того, чтобы читать актуальные сообщения из этой очереди, в каждом топике есть **специальный указатель (offset)**, который отвечает за смещение курсора на следующее сообщение в очереди. Это называется **фиксацией чтения (commit)**.

ОСНОВЫ



Структура очередей в топике

ОСНОВЫ



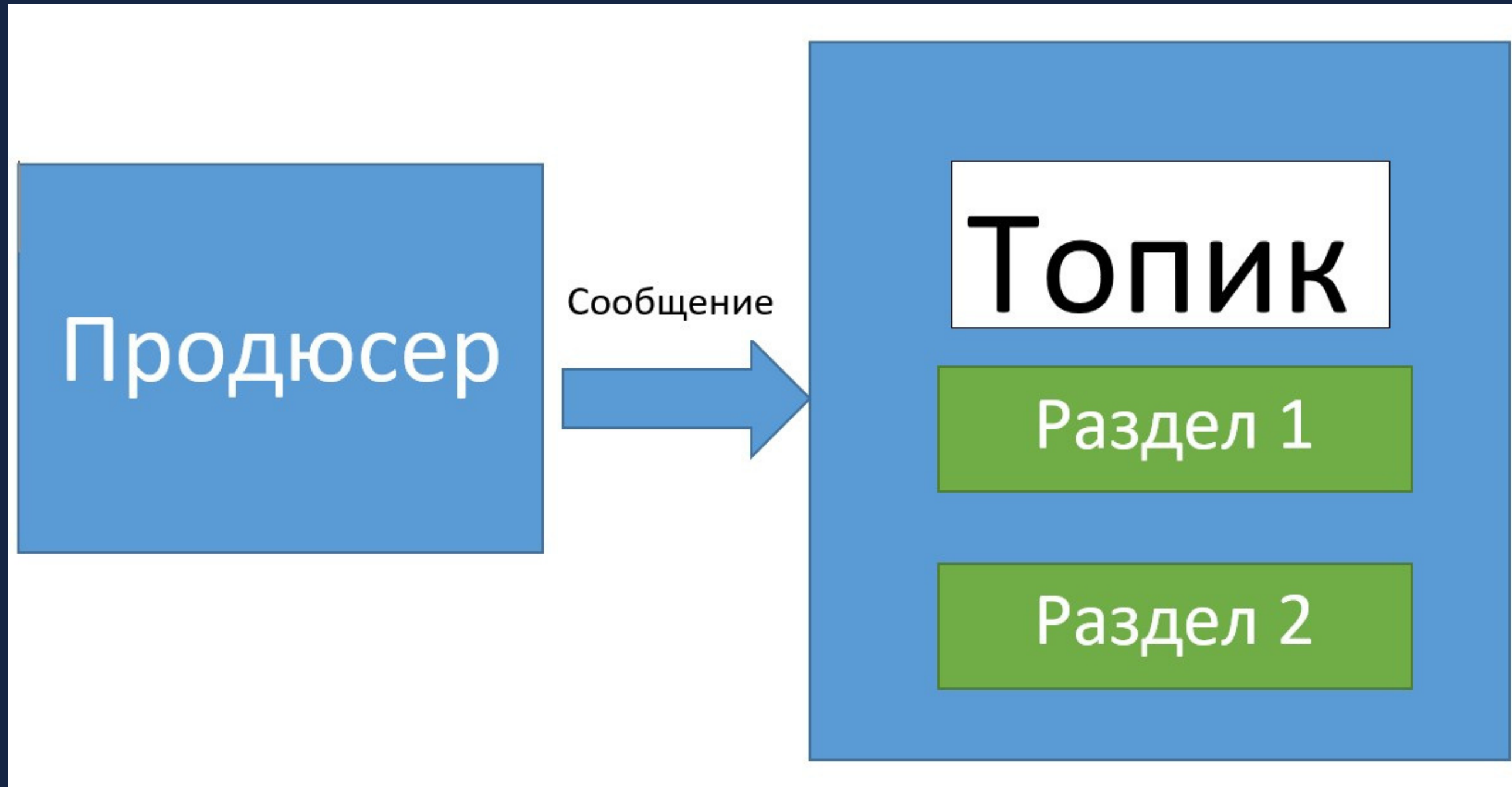
Структура очередей
в топике

Однако бывают такие случаи, когда сообщение в топики приходят быстрее, чем уходят из него. Специально для этого в Kafka-топиках предусмотрены **разделы (partition)**, которые разделяют очереди с целью защиты топика от переполнения.

ОСНОВЫ

Издатель (producer) — это сервис, отвечающий за создание и отправку сообщений в **топики**. Для отправки сообщения **продюсеру** необходимо знать название **топика** и порт, по которому доступен узел с этим **топиком**. Все сообщения, отправленные **продюсером**, распределяются по разделам **топика** и ждут своего «прочтения».

ОСНОВЫ

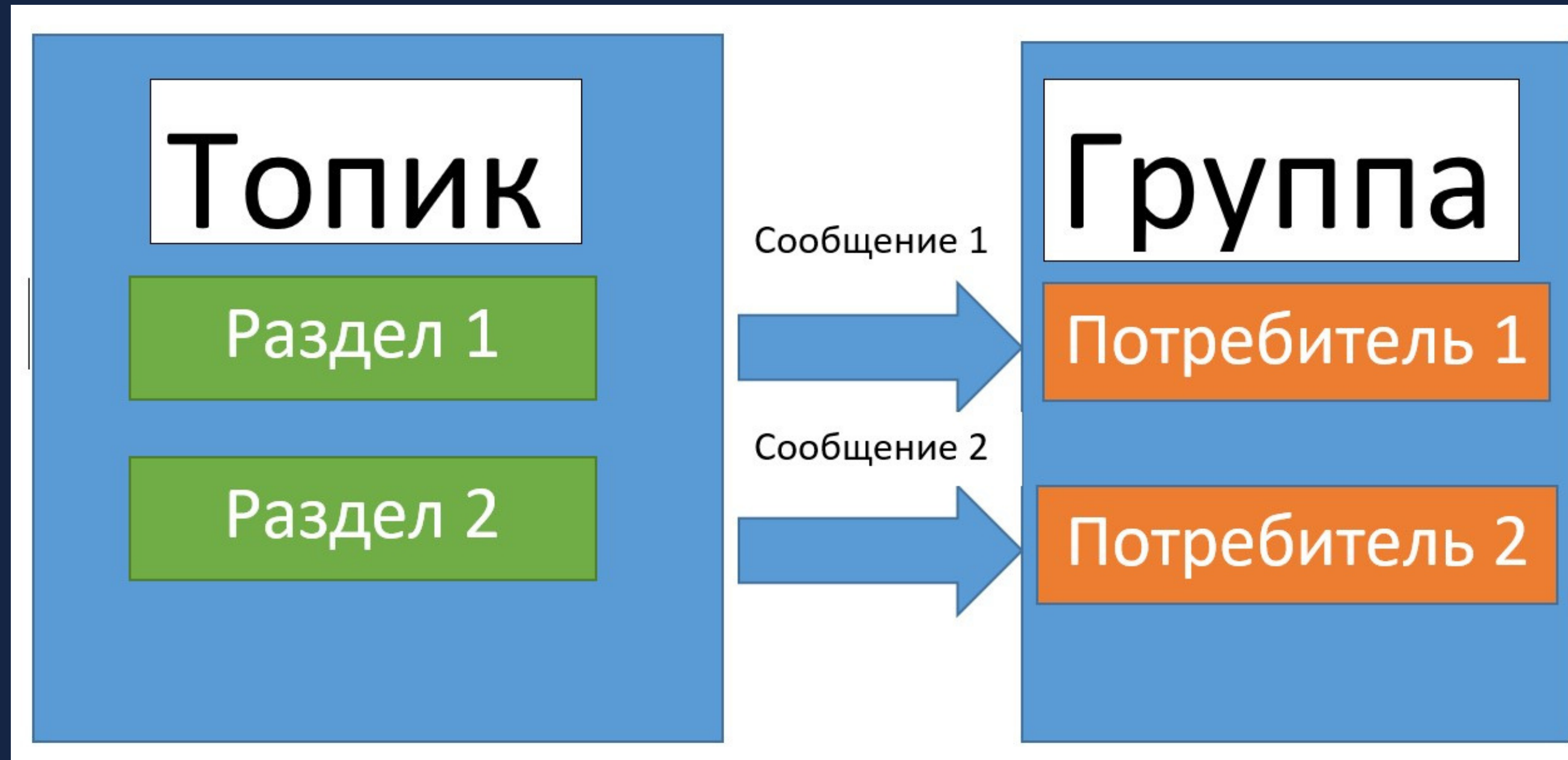


Создание сообщения продюсером в Apache Kafka

ОСНОВЫ

Потребитель (consumer) — это сервис, отвечающий за получение данных из **Kafka-топика**. Для получения сообщений, **потребителю**, также, как и **продюсеру**, необходимо иметь информацию о сервере (порт, URL) и имени **топика**, из которого необходимо получать данные. В случае разделения очередей в **топике** по разделам **потребители** объединяются в группы. Каждая группа **потребителей** получает данные из определенного раздела.

ОСНОВЫ



Получение данных потребителями

Форматы данных используемые в Kafka

Форматы данных

Avro — это система сериализации данных, которая используется для работы с объектами в потоковой (распределенной) среде. **Авро** использует систему, основанную на схемах входных данных. Схемы могут включать в себя тип данных, структуру данных, формат записи данных, а также параметры их передачи (например, URL, порт и т.д.).

Форматы данных

Схемы Avro описываются с помощью JSON-формата , что обеспечивает Авро независимость от языковой реализации. Для работы с Kafka используется специальный реестр схем (**Schema Registry**), который хранит в себе все схемы записей, используемых брокером в данный момент времени