

Школа Java Middle Developer

Kafka

Что такое Kafka?

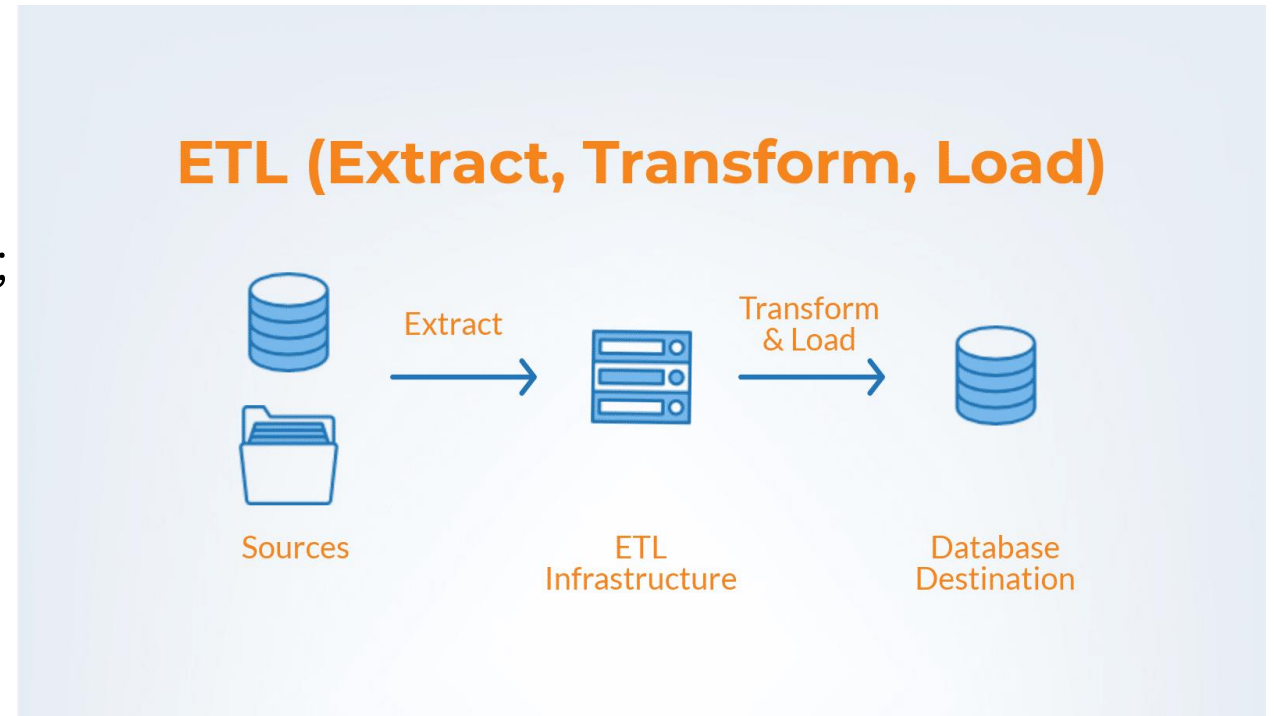
Содержание

- 1.Преимущества Kafka;
- 2.Распространенные мифы о больших данных и системах сообщений;
- 3.Реальные примеры использования, когда технологии Kafka помогли улучшить обмен сообщениями, потоковую передачу и обработку данных.

Обработка больших данных

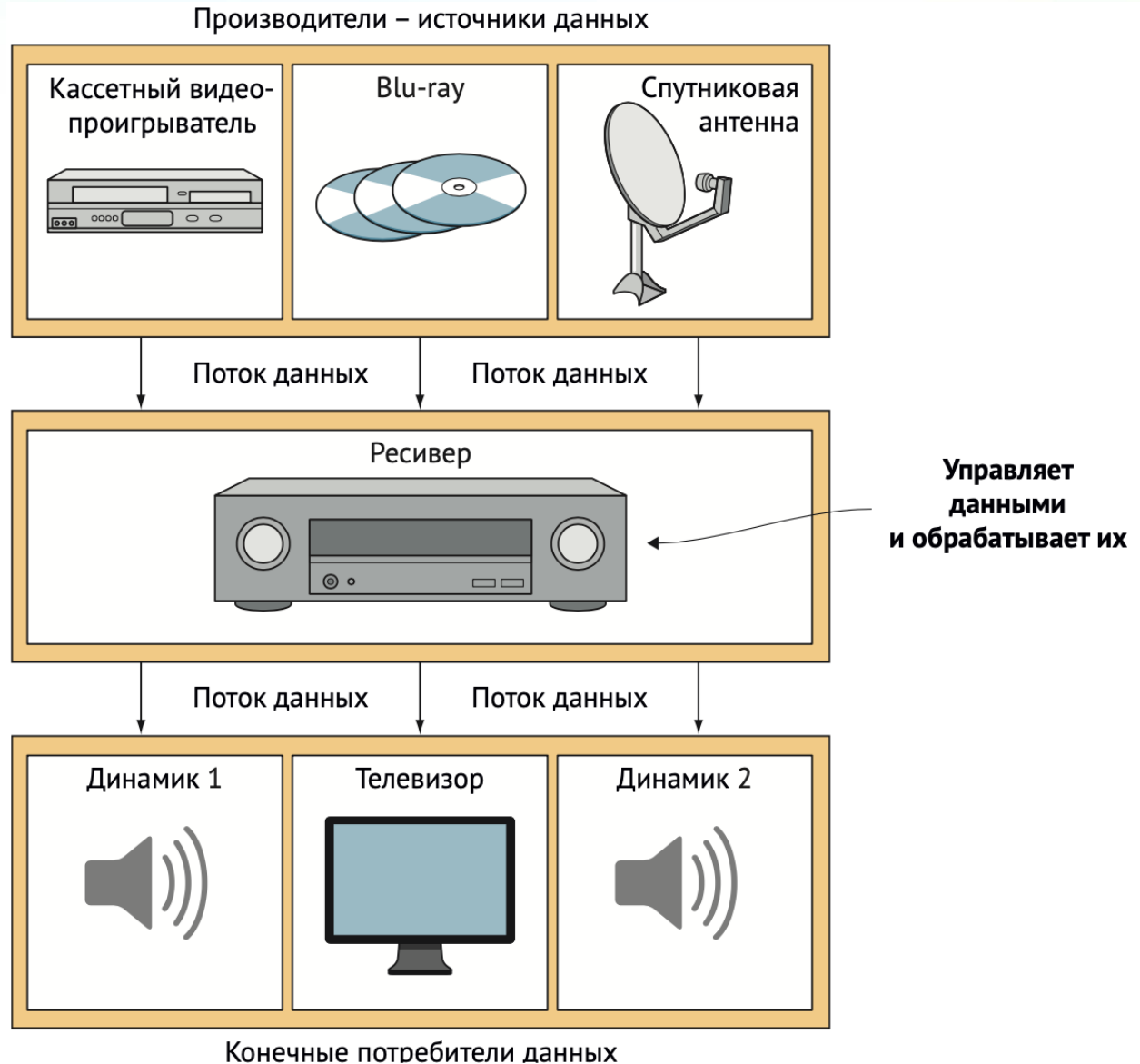
Обработка больших данных

- ✓ Kafka меняет стандарты платформ данных;
- ✓ Kafka предлагает мощный набор возможностей для обработки данных;
- ✓ Kafka следует многим новейшим и наиболее практичным тенденциям в мире информационных технологий и упрощает повседневную работу.

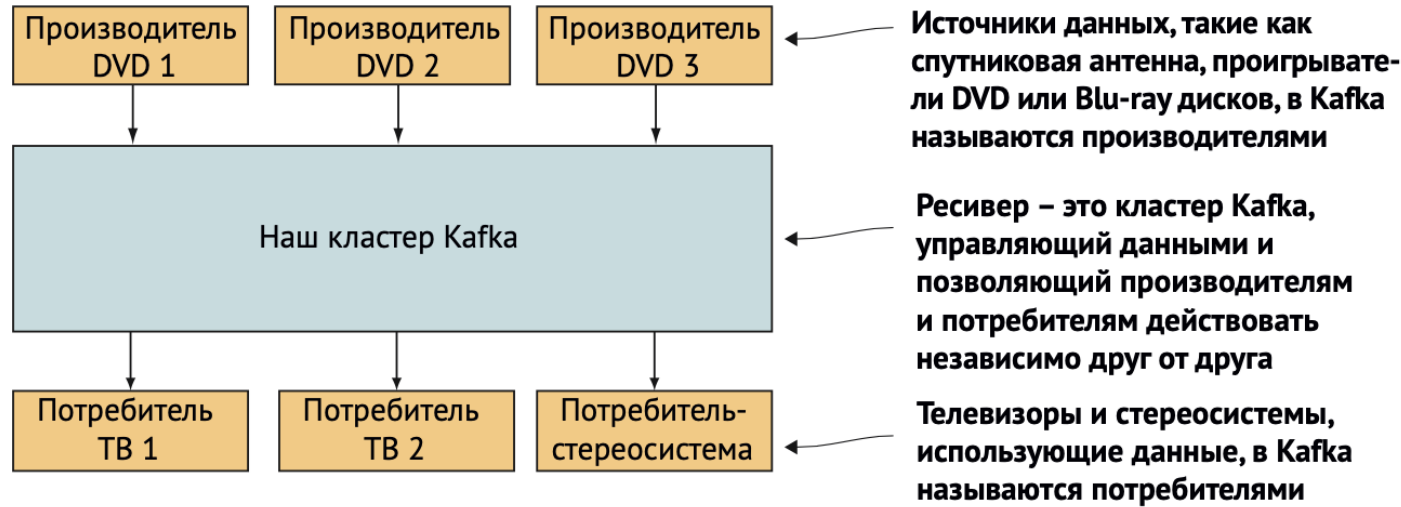


Что такое Kafka?

- Чтение сообщений из очереди и запись их в очередь;
- Надежное хранение сообщений;
- Обработку потоков данных по мере их появления.



Что такое Kafka?



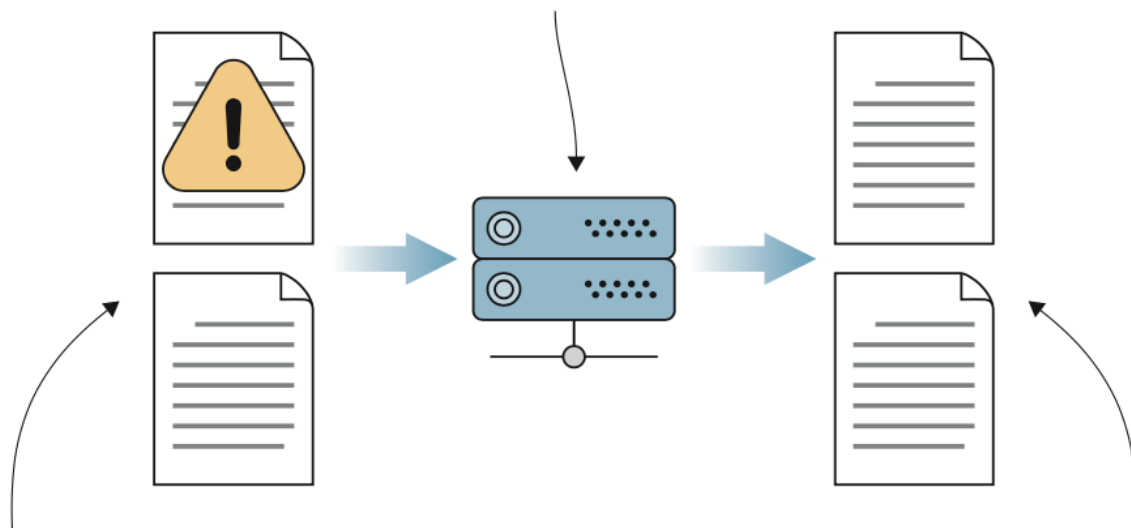
- Kafka служит интерфейсом, обеспечивающим возможность взаимодействий клиентов с другими системам;
- Kafka – система публикации сообщений и подписки на них, предназначенная для решения поставленной задачи.

Принцип работы

- Не менее одного раза (at-least-once) – сообщение будет отправляться потребителям до тех пор, пока те не подтвердят его получение;
- Не более одного раза (at-most-once) – сообщение отправляется только один раз и в случае сбоя не отправляется повторно;
- Точно один раз (exactly-once) – потребитель гарантированно получит сообщение ровно один раз.

Не менее одного раза

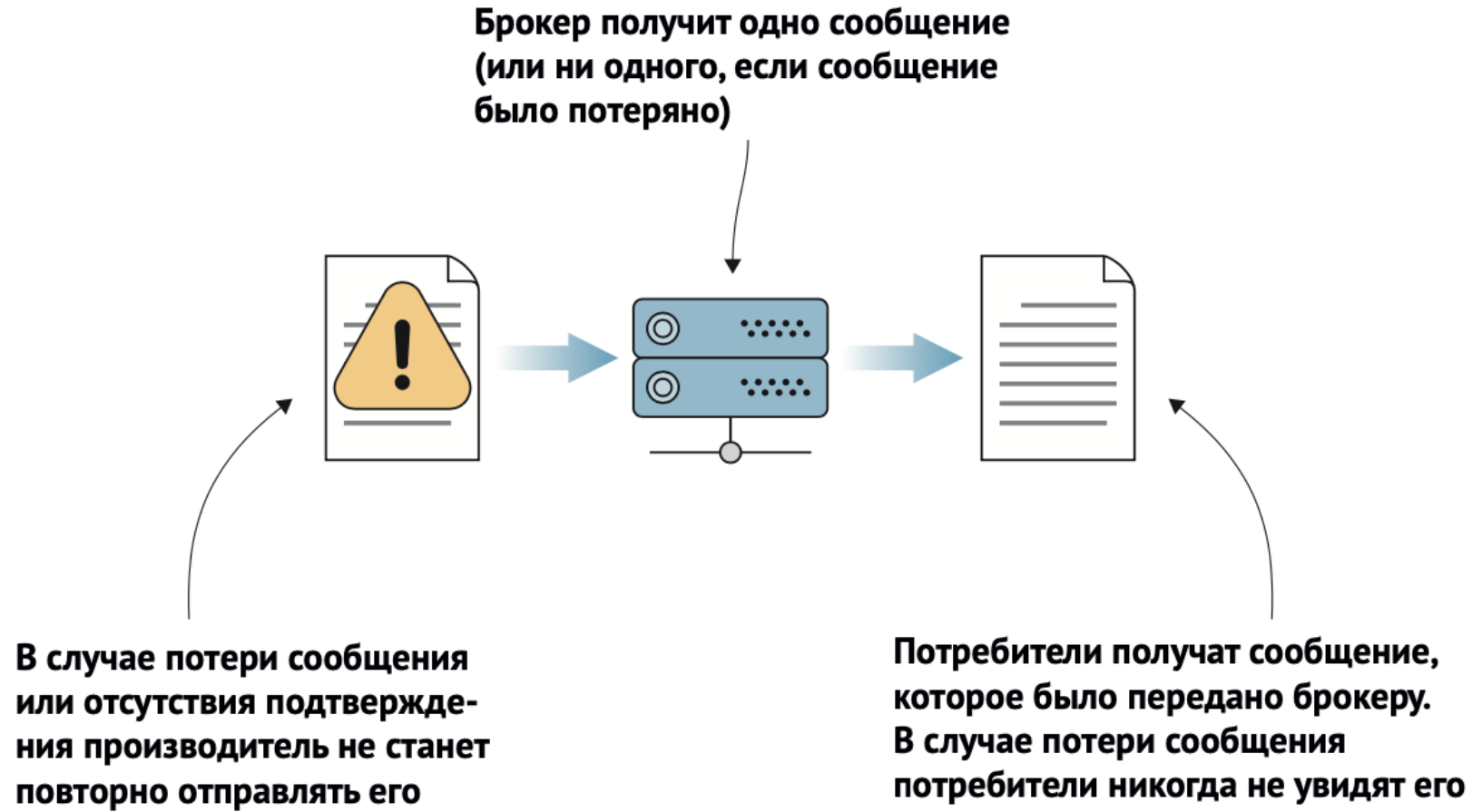
При использовании семантики «не менее одного раза» брокер получит два сообщения (или одно, если второе было потеряно)



В случае потери сообщения или отсутствия подтверждения производитель повторно отправит его

Потребители получают столько сообщений, сколько получил брокер. Потребители могут получать повторяющиеся сообщения

Не более одного раза



Точно один раз



Почему Kafka?

- Наличие интеграции с популярными фреймворками;
- Kafka решает проблемы сильной связанности компонентов системы;
- Kafka может служить единым интерфейсом к данным;
- Платформа распространяется с открытым исходным кодом.

Сообщения и пакеты

- Используемая в Kafka единица данных называется сообщением;
- Сообщение в Kafka – это просто массив байтов;
- Сообщения могут содержать ключ (метаданные);
- Сообщения в Kafka записываются пакетами;
- Пакеты обычно подвергаются сжатию;

Схемы

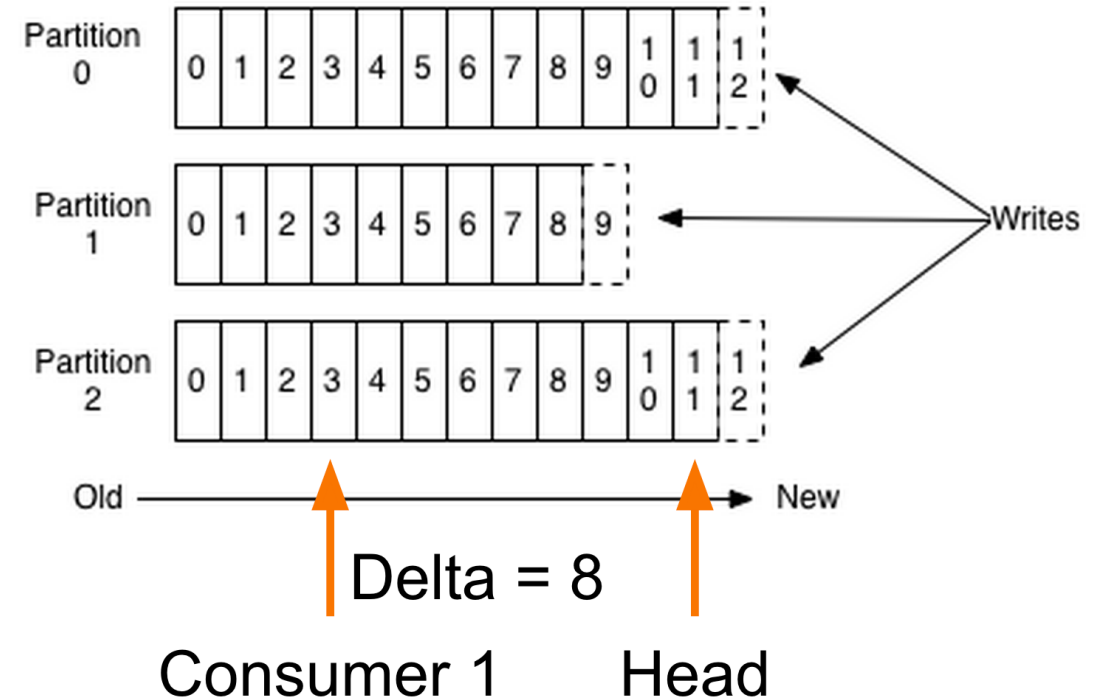
➤ JSON

➤ XML

➤ Avro

Топики и партиции

- Сообщения в Kafka распределяются по топикам (topics);
- Топики в свою очередь разбиваются на партиции (partitions);
- Сообщения записываются в него путем добавления в конец, а читаются от начала к концу;
- нет никаких гарантий упорядоченности сообщений в пределах всего топика.

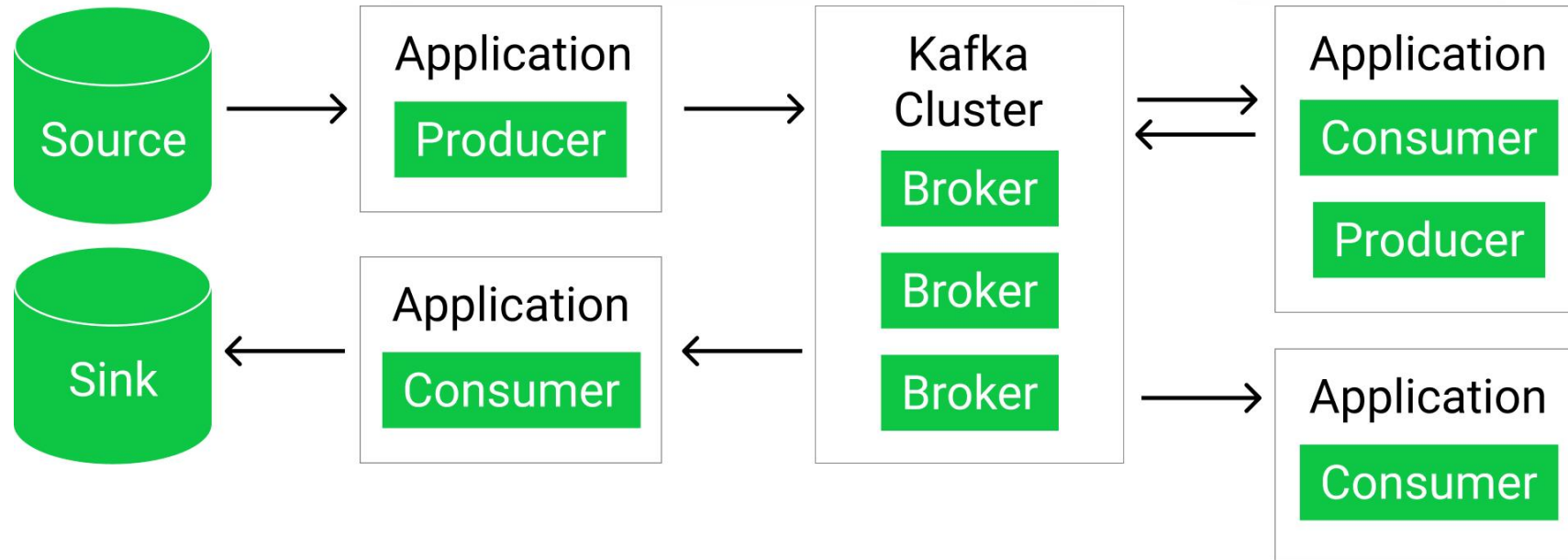


Журнал коммитов

- События всегда добавляются в конец журнала;
- Чтение сообщения не удаляет его из системы и не исключает из других источников;
- Хранением данных в журнале можно управлять по времени или размеру с помощью свойств конфигурации.

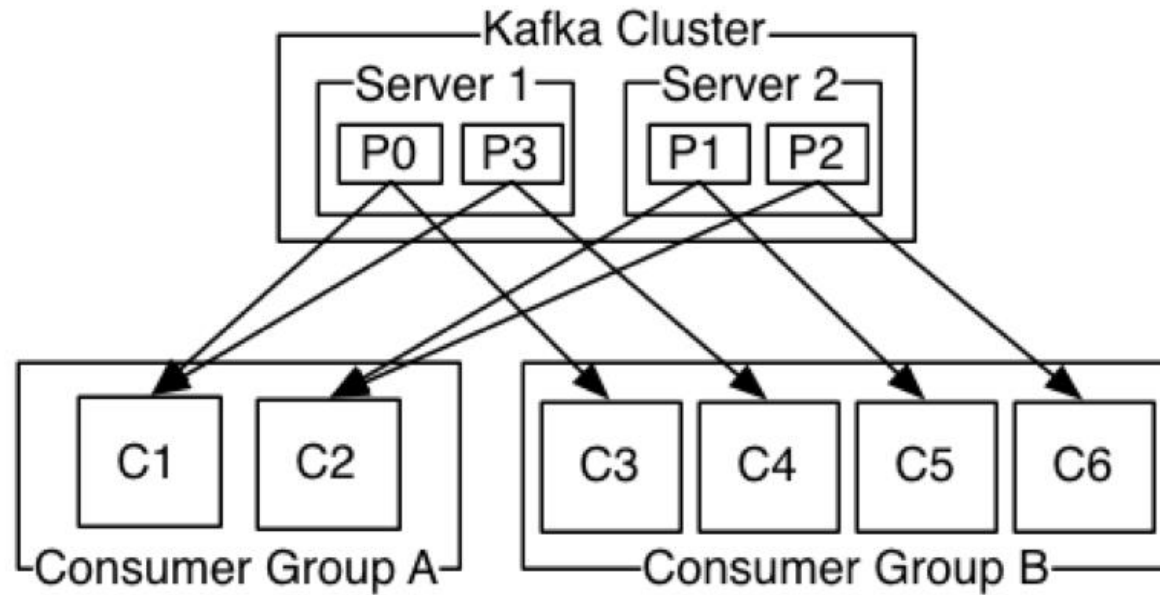


Производители и потребители



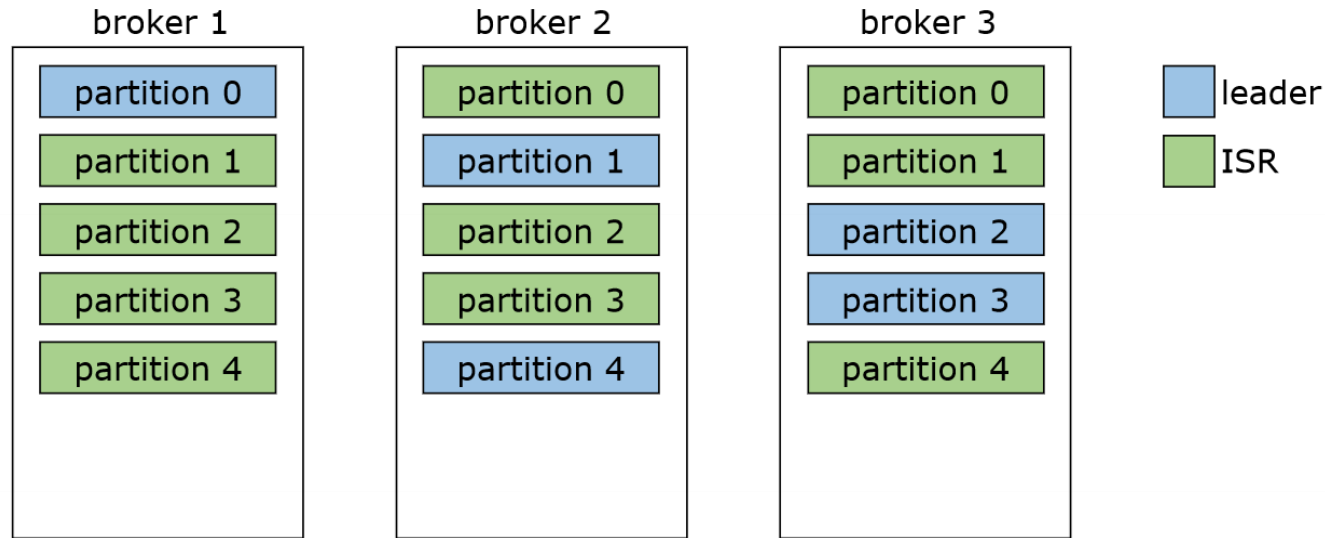
- Производители (producers) генерируют новые сообщения;
- Потребители (consumers) читают сообщения;

Группы потребителей



- Потребители работают в составе групп потребителей (consumer groups);
- Организация в группы гарантирует чтение каждой партиции только одним членом группы;
- Несколько потребителей Kafka могут читать один поток сообщений, не мешая друг другу;
- Потребители имеют возможность горизонтального масштабирования для чтения топика;
- В случае сбоя отдельного потребителя оставшиеся члены группы перераспределят партиции так, чтобы взять на себя его задачу.

Брокеры и кластеры



- Сервер Kafka называется брокером (broker);
- Брокеры Kafka предназначены для работы в составе кластера (cluster);
- Один из брокеров (выбирается автоматически) кластера функционирует в качестве контроллера (cluster controller);
- Каждая партиция принадлежит одному из брокеров кластера, который называется его ведущим (leader).

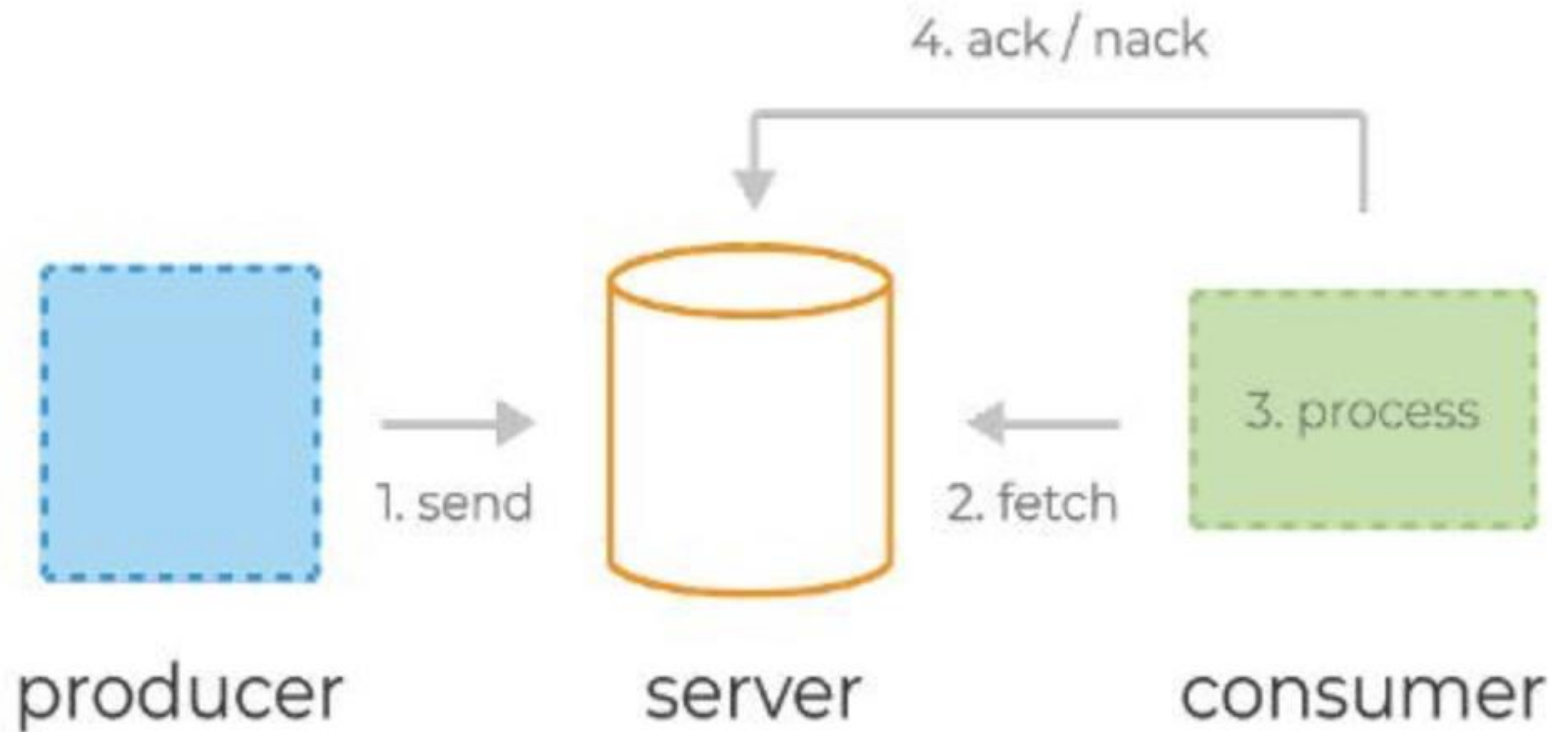
ZooKeeper

- ❖ ZooKeeper – ~~один из старейших источников дополнительной сложности в экосистеме Kafka~~ это распределенное хранилище, предлагающее высокодоступные службы обнаружения, настройки и синхронизации;
- ❖ Кластер Kafka может включать несколько брокеров (серверов);
- ❖ Брокеры достигать согласованности действий;
- ❖ При промышленном использовании ZooKeeper работает как ансамбль.

Миф: Kafka работает только с Hadoop

- ❖ В Kafka применяют различные инструменты, такие как Spark Streaming и Flume, которые вместе с тем используют (или когда-то использовали) Hadoop;
- ❖ Apache ZooKeeper – тоже часто встречается в кластерах Hadoop;
- ❖ В Kafka реплики не восстанавливаются по умолчанию;
- ❖ Надежность Kafka можно легко сравнить с Hadoop.

Миф: Kafka ничем не отличается от других брокеров сообщений



Миф: Kafka ничем не отличается от других брокеров сообщений

- ❖ Возможность повторной передачи сообщений по умолчанию шаги разметки и очистки параллельно с работой основной программы;
- ❖ Параллельная обработка данных;
- ❖ Kafka даёт потребителям возможность отыскать конкретный момент и прочитать сообщение снова;
- ❖ Может обслуживать нескольких потребителей в одном и том же топике;
- ❖ Обработка сообщений из брокера может производиться несколькими потребителями, запущенными одним приложением или несколькими разными.

Практическое использование Kafka

- ❖ Обмен сообщениями (работает на основе TCP);
- ❖ Платформа Kafka изначально создавалась с прицелом на высокую доступность и долговременное хранение;
- ❖ Агрегирование журналов;
- ❖ Микросервисы могут использовать Kafka как интерфейс для своих взаимодействий вместо прямых вызовов API;
- ❖ Интернет вещей.



Когда Kafka может быть неприменима

- ❖ Если вам нужно получить сводные данные только раз в месяц или даже раз в год;
- ❖ Если вам не требуется получать данные по запросу или повторно обрабатывать их;
- ❖ Если в вашем случае основной операцией с данными является произвольный поиск;
- ❖ Если нужно соблюсти точный порядок следования сообщений в Kafka для всего топика;
- ❖ Если размер сообщения сильно превышает 1 Мбайт.

Спасибо за внимание