

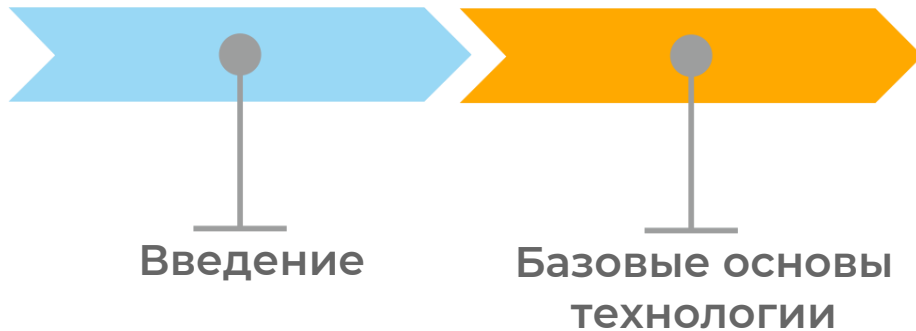
# Урок 1. Введение



# Обзор курса



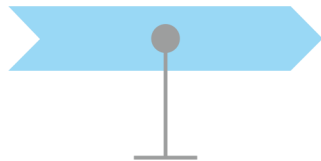
# Обзор курса



# Обзор курса



# Обзор курса



Работа с  
распределенным  
кластером

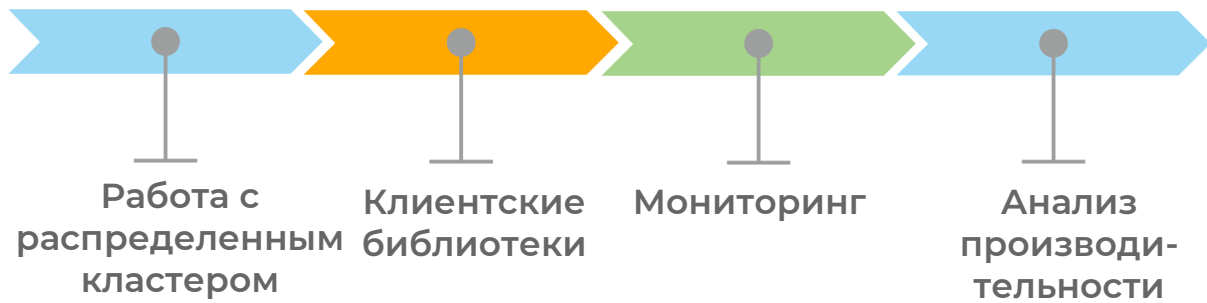
# Обзор курса



# Обзор курса



# Обзор курса

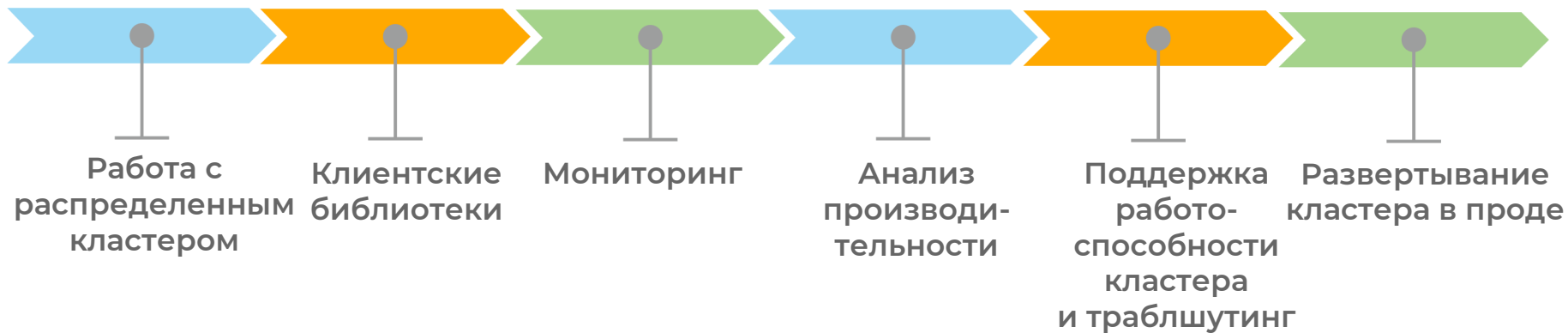




# Обзор курса



# Обзор курса



# Вы узнаете

- Что такое Apache Kafka и как она работает



# Вы узнаете

- Что такое Apache Kafka и как она работает
- Как устанавливать и администрировать Apache Kafka в Production



# Вы узнаете

- Что такое Apache Kafka и как она работает
- Как устанавливать и администрировать Apache Kafka в Production
- Как писать клиентов



# Вы узнаете

- Что такое Apache Kafka и как она работает
- Как устанавливать и администрировать Apache Kafka в Production
- Как писать клиентов
- Как проводить анализ производительности и troubleshooting



# Вы узнаете

- Что такое Apache Kafka и как она работает
- Как устанавливать и администрировать Apache Kafka в Production
- Как писать клиентов
- Как проводить анализ производительности и troubleshooting
- Как работать с Apache Kafka в нескольких ДЦ



# План





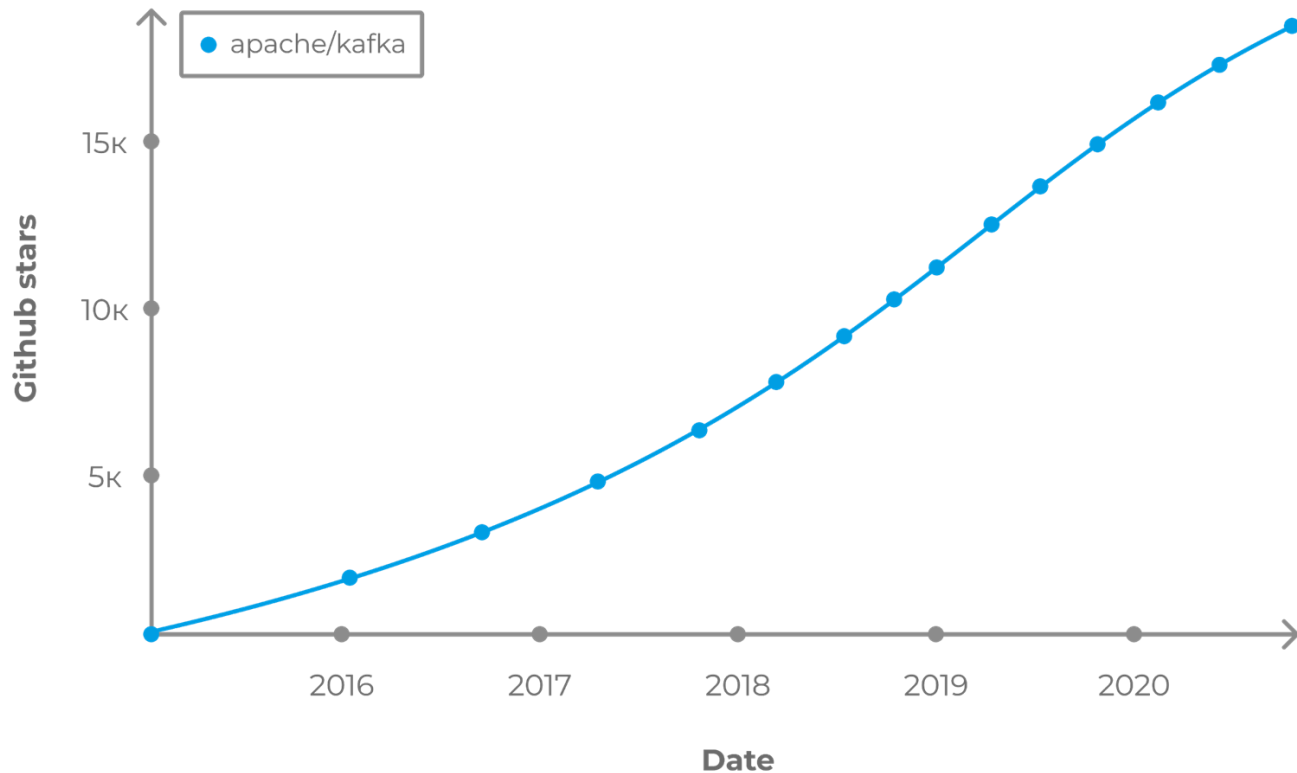
# Актуальность

# Актуальность

- Более **80%** компаний из Fortune 100 используют Kafka
- **Топ 2** по активности проект Apache в 2019 году
- Большое и активное комьюнити
- **Лидирующая технология** в своей области
- Многие компании только начинают свой путь с Kafka

# Актуальность

Star history



# История

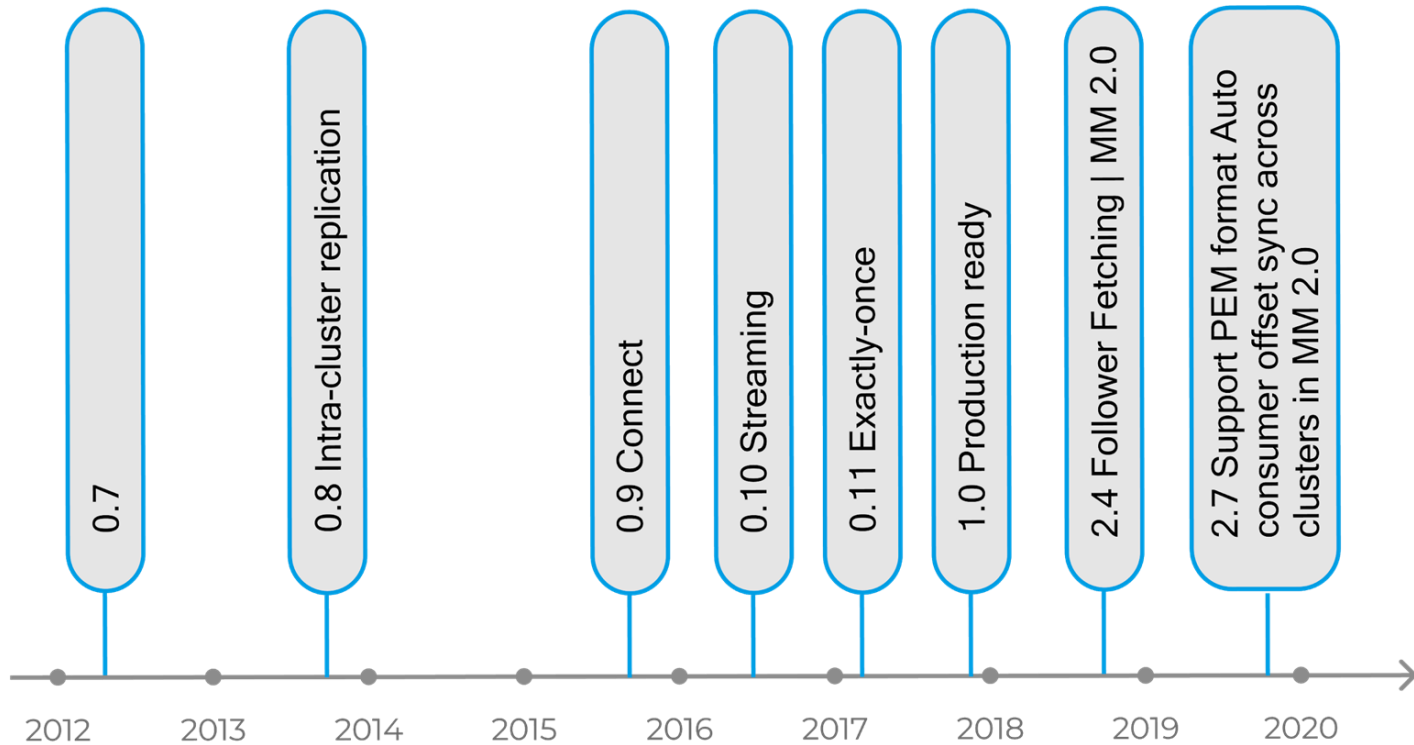
# История



I thought that since Kafka was a system optimized for writing, using a writer's name would make sense.

Jay Kreps

# История



# Что такое Apache Kafka

# Что такое Apache Kafka

- Распределенный лог?
- База данных?
- Очередь?
- Платформа для потоковой обработки данных?
- Что-то еще?

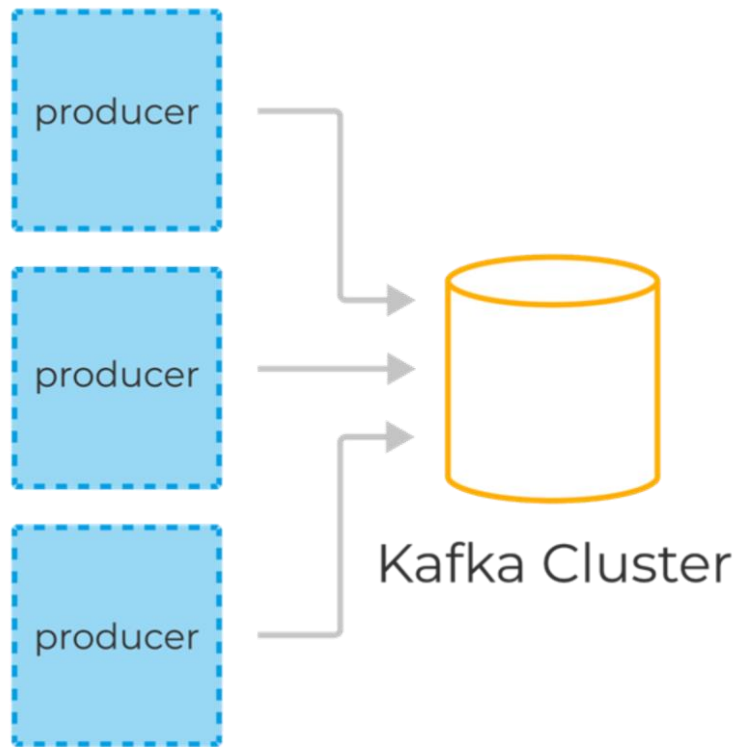


# Что такое Apache Kafka

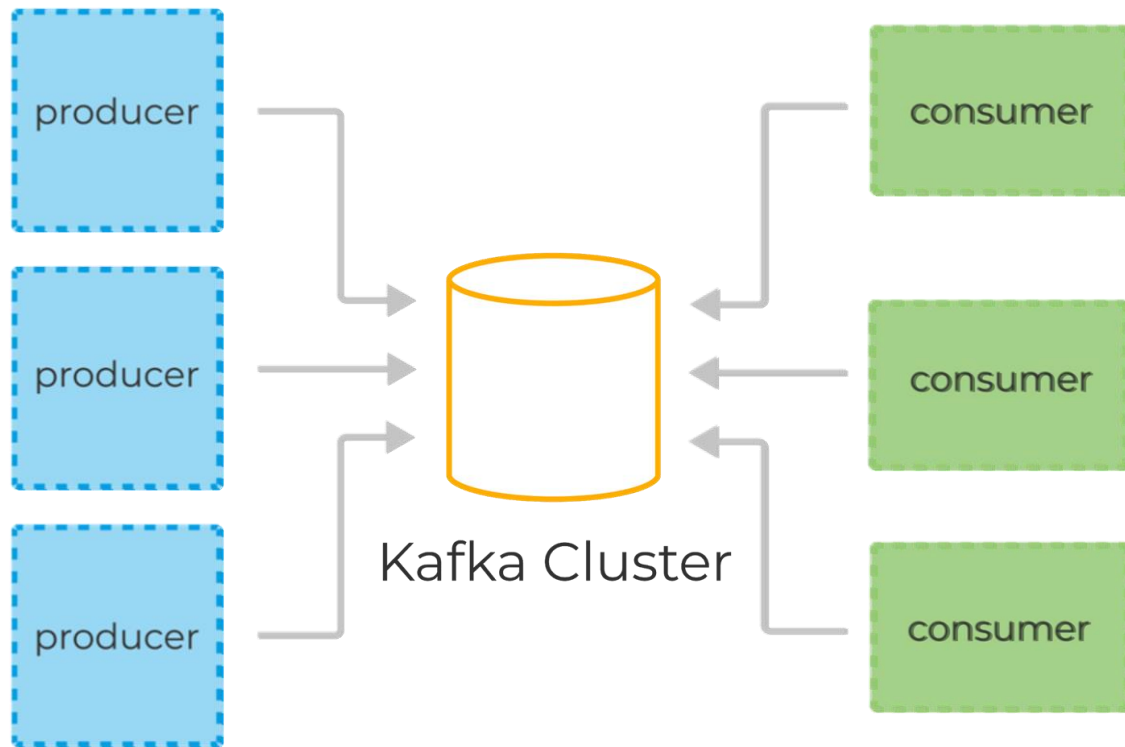


Kafka Cluster

# Что такое Apache Kafka



# Что такое Apache Kafka



# История

**7 000 000 000 000** / день

# Что такое Apache Kafka

## Термины:

■ Распределенное хранилище - это система, которая как правило работает на нескольких машинах, каждая из этих машин в свою очередь является кусочком хранилища. Для пользователя это все представляется в виде единого целого.



# Что такое Apache Kafka

## Термины:

■ Горизонтальное масштабирование - техника, в которой мы вместо увеличения мощностей одной машины, добавляем в систему дополнительные машины и тем самым увеличиваем доступные ресурсы.



# Что такое Apache Kafka

## Термины:

■ Отказоустойчивость - это свойство системы, позволяющее ей не иметь единую точку отказа. То есть если что-то пошло не так и какая-то из машин вышла из строя, система сможет продолжить работу.



# Что такое Apache Kafka

## Термины:

- Лог - это упорядоченная по времени append-only структура данных





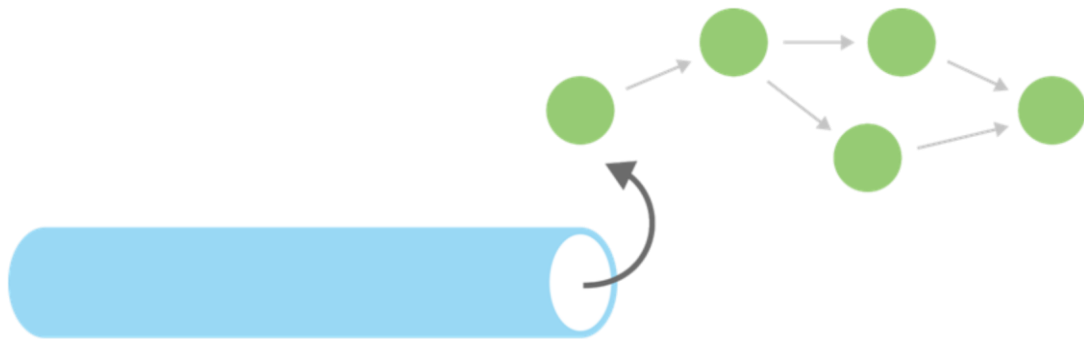
# Что такое Apache Kafka



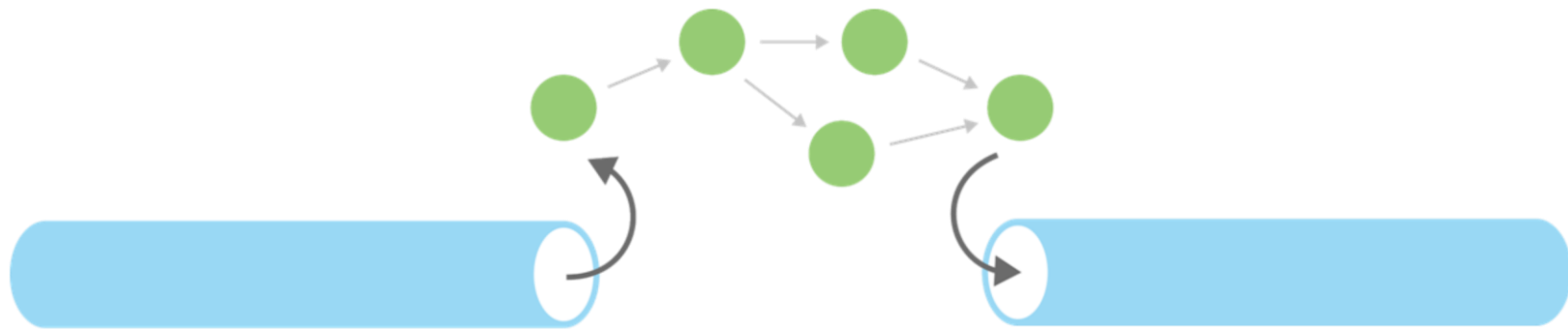
# Что такое Apache Kafka



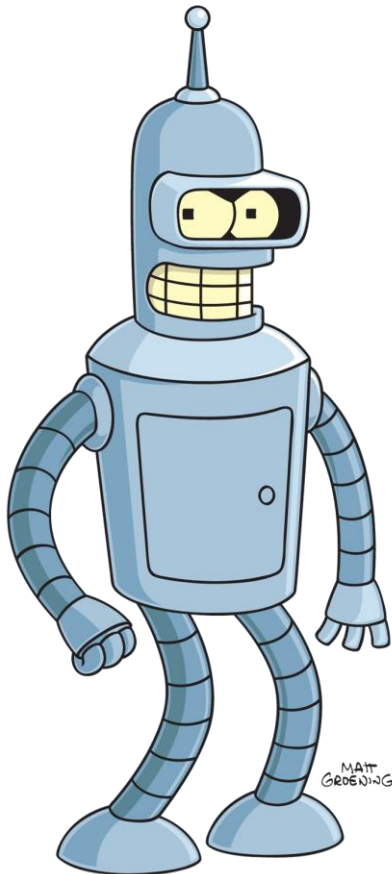
# Что такое Apache Kafka



# Что такое Apache Kafka



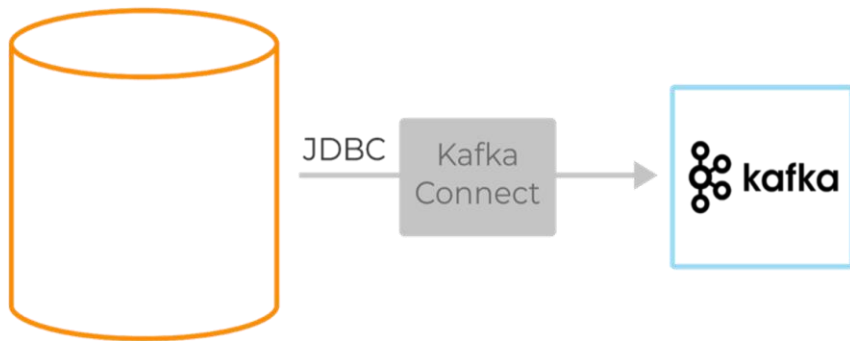
# Что такое Apache Kafka



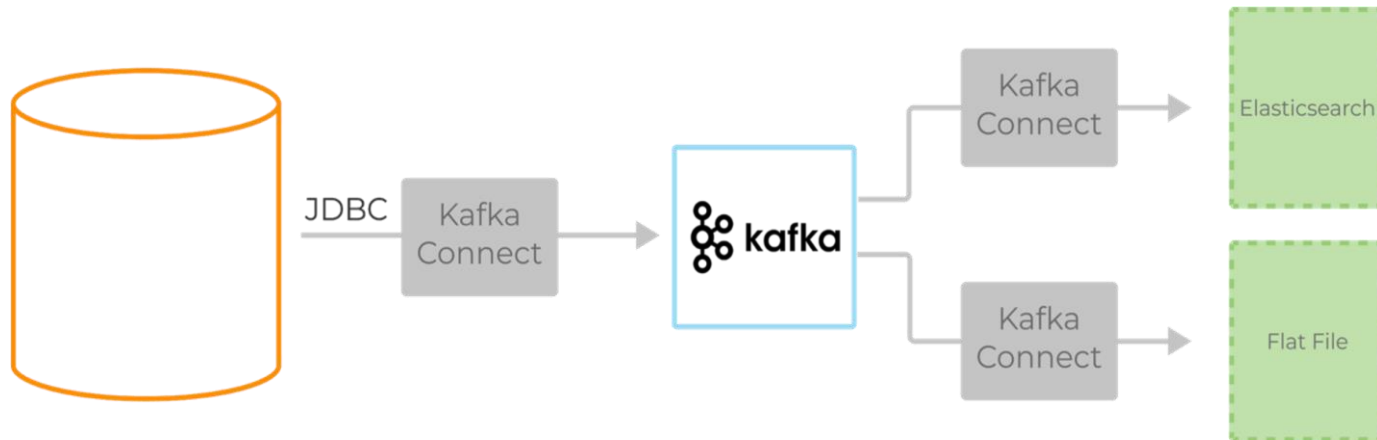
# Что такое Apache Kafka



# Что такое Apache Kafka



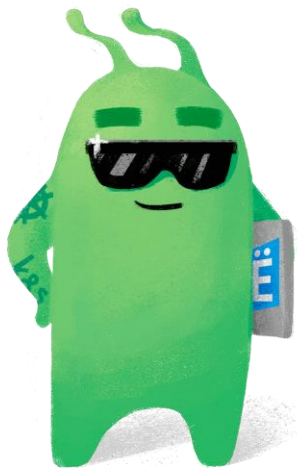
# Что такое Apache Kafka





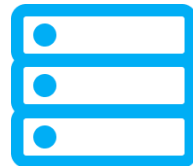
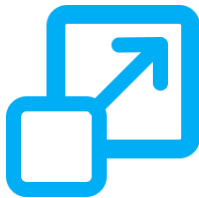
# Что такое Apache Kafka

Kafka это **распределенное, отказоустойчивое, горизонтально масштабируемое хранилище**, основной структурой данных в котором является **append-only лог**, которое поддерживает потоковую обработку данных и имеет развитую экосистему коннекторов для интеграции с базами данных и другими хранилищами.



# Почему Apache Kafka

# Почему Apache Kafka



# Почему Apache Kafka



# Почему Apache Kafka



# Примеры использования

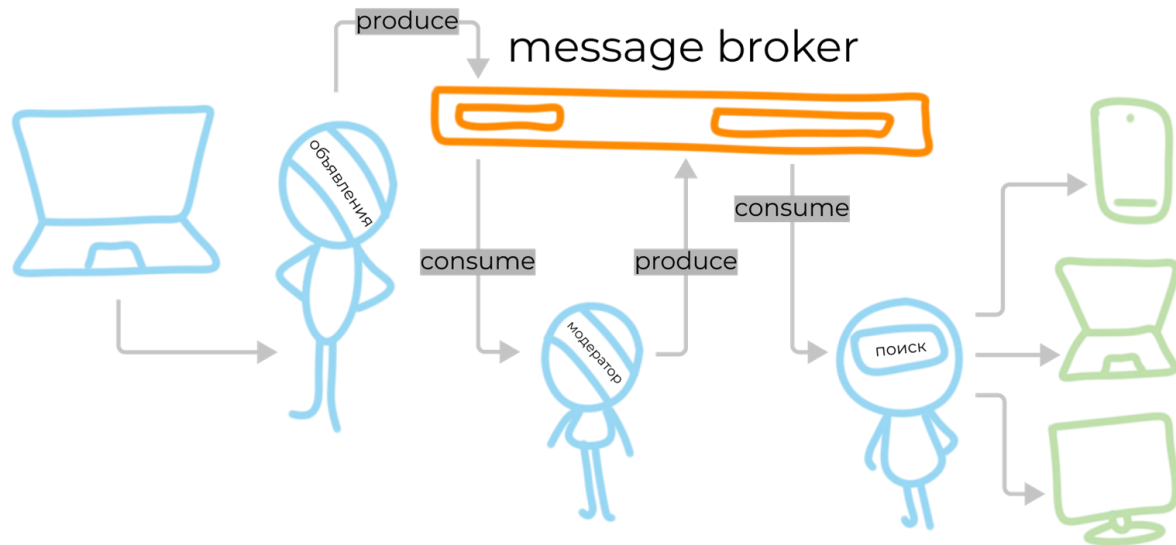
# Примеры использования

Араше Kafka чаще всего используется компаниями с **50-200 сотрудниками**

Араше Kafka чаще всего используется компаниями с **доходом 10М-50М\$**

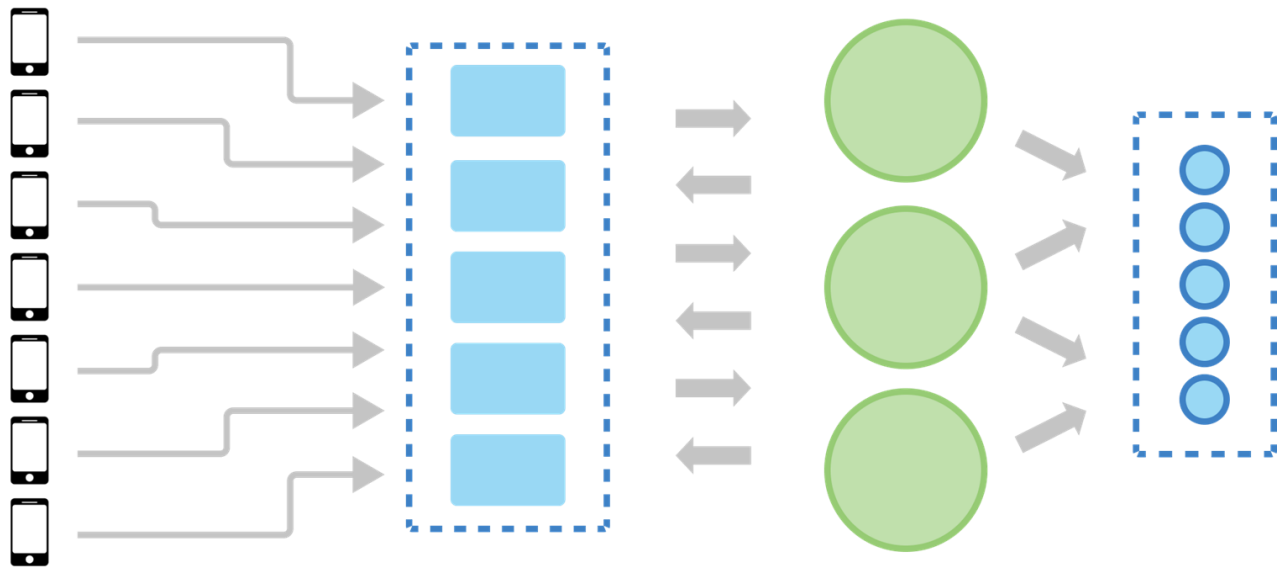
Араше Kafka чаще всего используется в **Computer Software индустрии**

# Примеры использования. Брокер сообщений

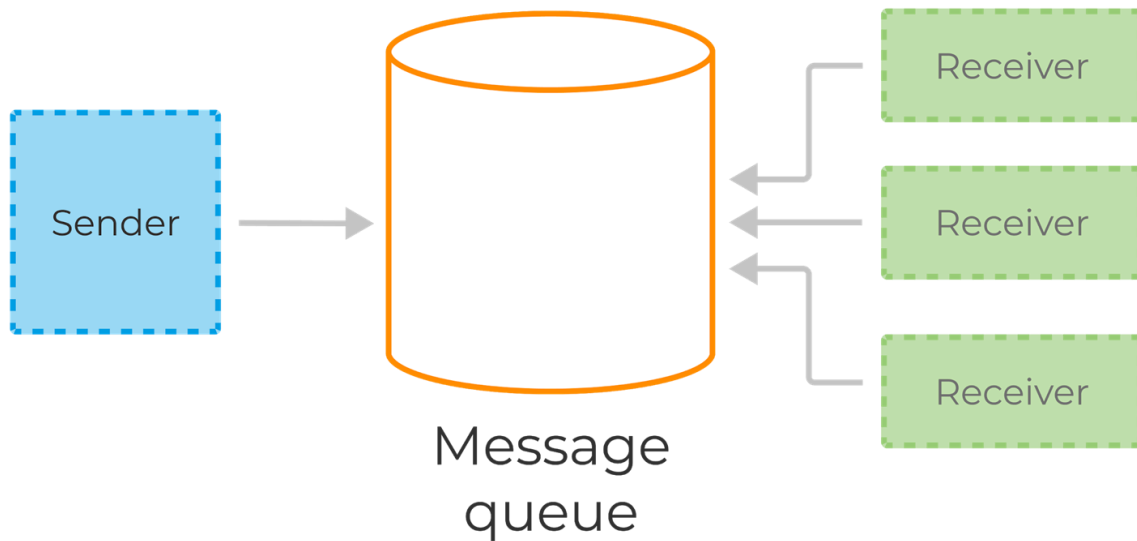




# Примеры использования. Кликстрим



# Примеры использования. Система очередей



# Примеры использования. Журналирование и сбор метрик



beats



beats



kafka

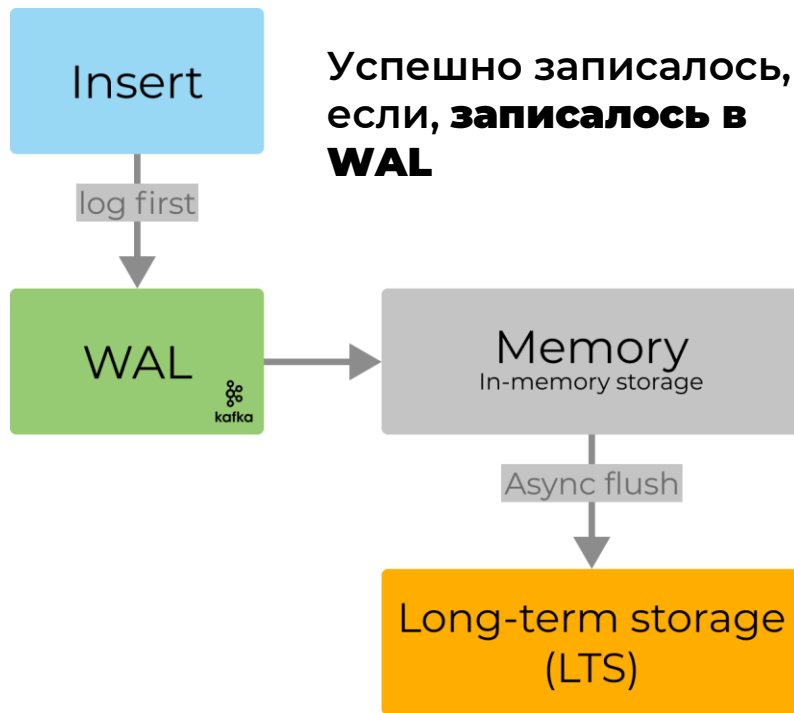


beats

# Примеры использования. Журналирование и сбор метрик



# Примеры использования. Commit log



Успешно записалось,  
если, **записалось в  
WAL**

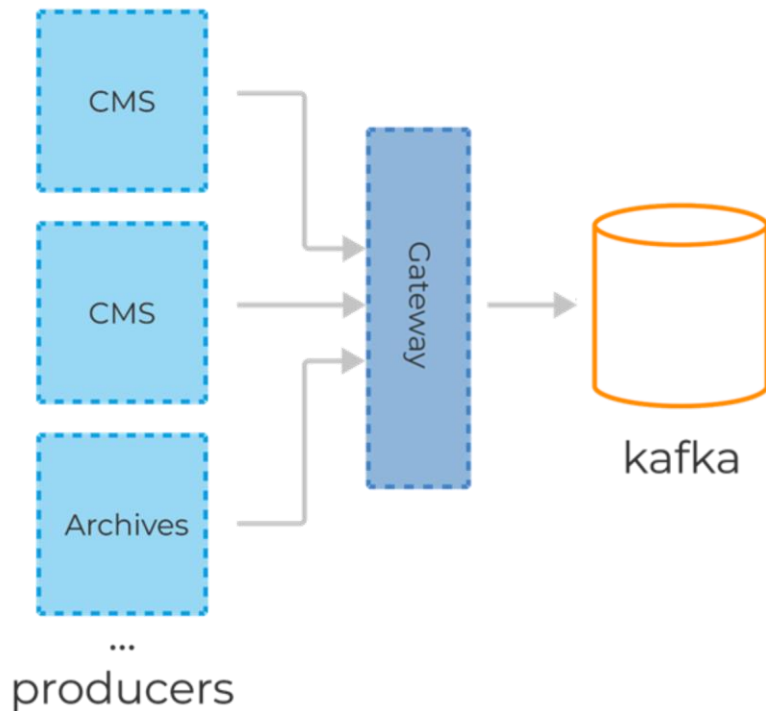
Все остальные компоненты  
сидят на потоке изменений  
из kafka

# Примеры использования. База данных

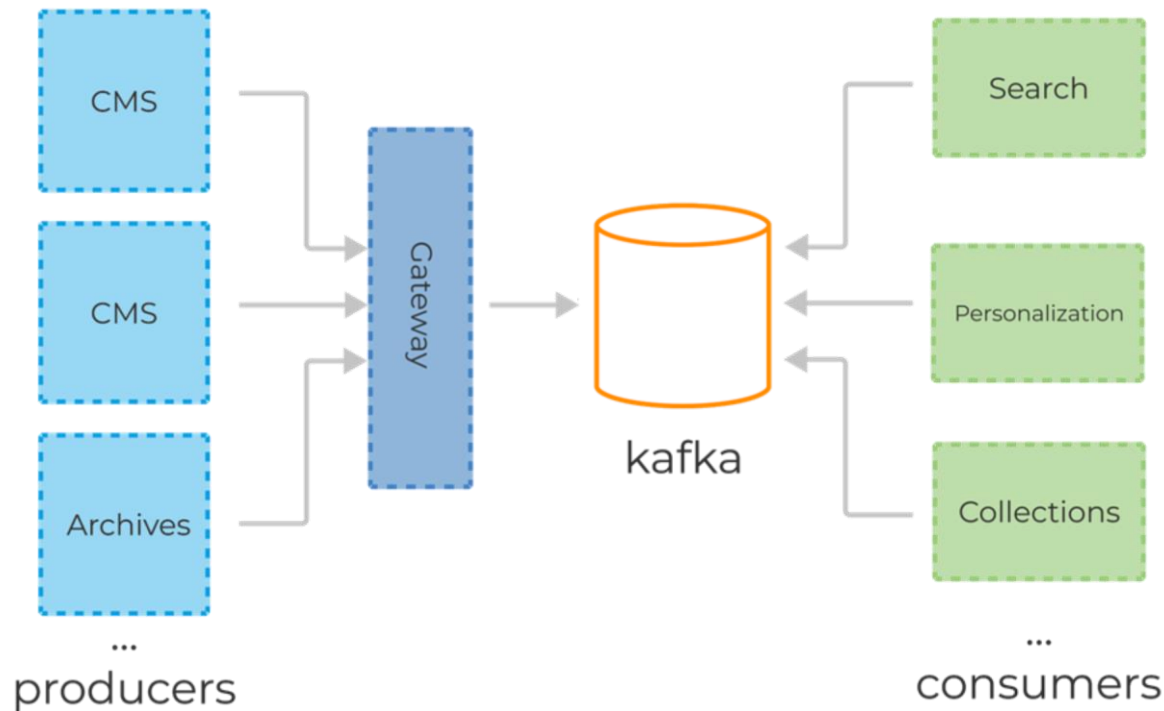


kafka

# Примеры использования. База данных



# Примеры использования. База данных





# Примеры использования. База данных



# Подведем итог

# Подведем итог

- Apache Kafka – быстрая, надежная технология с богатой экосистемой и активным комьюнити
- Kafka не заточена под единственный вариант использования и подходит под разнообразные задачи
- Во всем мире количество компаний, использующих Kafka, продолжает расти и им нужны специалисты

