

Fintech Python

NoSQL базы данных

Делаем тест

NoSQL

WEB 2.0

Новые вызовы к кол-ву одновременных подключений

Масштабирование

Что это такое?

SQL

Почему он мешает масштабироваться?

Репликация

Как база для масштабирования

Отказоустойчивость

Auto failover - если база упала что подключись к slave

Масштабируемость

Запросов так много - что щяз треснем

**Какие виды
репликации бывают**

Master - Slave

Что это?

Синхронная

Асинхронная

Master - Master

Синхронная

Иначе ни как

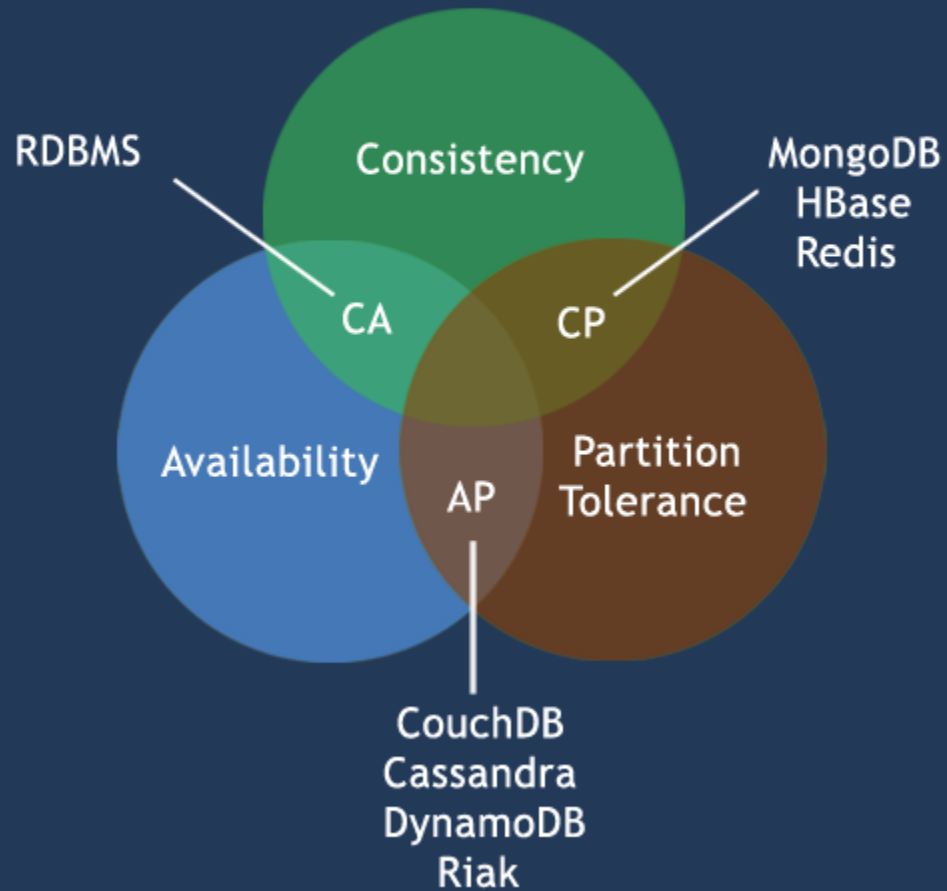
Git

Как master-master распределенная система

Кластер

Что это такое?

CAP Theorem



NoSQL

Врывается в тред

Отказ от SQL

- Как языка запросов
- Транзакций привычных для SQL
- Структур данных привычных для SQL
- Более гибкая репликация
- ...

A history of databases in No-tation

1970: NoSQL = We have no SQL

1980: NoSQL = Know SQL

2000: NoSQL = No SQL!

2005: NoSQL = Not only SQL

2013: NoSQL = No, SQL!

(R)DB(MS)

SAMSUNG



Примеры NoSQL баз

Пример

- redis
- kafka
- mongodb
- cockroachdb
- riak
- couchdb
- casandra
- hbase
- neo4
- clickhouse

Сравнение

<https://kkovacs.eu/cassandra-vs-mongodb-vs-couchdb-vs-redis>

Другие алгоритмы

В новых базах использовали новые алгоритмы

MongoDB

Документо-ориентированная

- JSON
- Вторичные индексы по нему
- Нет связей между документами

Почему популярна была?

// OOOOOOOO какая быстрая!!!111

fsync off

```
with open('kek', 'w') as f:  
    f.write('kek')  
    exec '()' * 7 * 6
```

Почему mongo не нужна?

Есть postgresql и его jsonb

Все не так однозначно

<https://www.youtube.com/watch?v=SNzOZKvFZ68>

Cassandra

Огромные таблицы и свой язык CQL

Нет транзакций

В понимании SQL

<https://aphyr.com/posts/294-jepsen-cassandra>

Вечные уплотнения

Особенности реализации которые надо понимать
когда выбираешь базу данных

Вопросики?

Key/value

BigTable by Google

Hbase

Распределенная отсортированная таблица

Операции

- Поиск по ключу или части ключа
- Вычитывание по ключу или по диапазону
- Удаление
- Добавление

Column Database

Что это такое?

C-Store

Концепт от создателя PostgreSQL

ClickHouse

Реализация идей c-store от Yandex

Для чего используется

Большие и тяжелые агрегации

Memcached

Тупой мультипроцессорный кеш

Redis

key/value in memory db

Data Type	Description	Example command
<i>Strings</i>	A string is a binary safe value that had a maximum length of 512 Megabytes.	SET [key] [value]
<i>Lists</i>	A list of strings sorted by insertion order.	LPUSH [list] [value]
<i>Sets</i>	A set is a an unordered collection of strings.	SADD [set] [value1] [value2] [value3]
<i>Hashes</i>	A collection of key value pairs.	HMSET [hash] [key] [value] [key2] [value2]
<i>Sorted Sets</i>	Sorted sets are similar to the sets data types but each have a user defined score.	ZADD [set] [score] [value]
<i>Bitmaps</i>	Bitmaps may also be known as bit arrays. Bitmaps are great for saving space when storing information.	SETBIT [key] [bit number] [value]

Не просто *key/value*

Множество операций для разных структур данных

Вопросики?

Очереди

Зачем нужны очереди

О чем это все?

Это все о очередях и событиях!

Специальные базы данных

- rabbitMQ
- zeroMQ
- ...
- redis

Очереди в Python

- celery
- dramatiq
- huey

```
1 import dramatiq
2 import requests
3
4 @dramatiq.actor
5 def count_words(url):
6     response = requests.get(url)
7     count = len(response.text.split(" "))
8     print(f"There are {count} words at {url!r}.")
9
10 # Synchronously count the words on example.com
11 # in the current process
12 count_words("http://example.com")
13
14 # or send the actor a message so that
15 # it may perform the count later,
16 # in a separate process.
17 count_words.send("http://example.com")
```

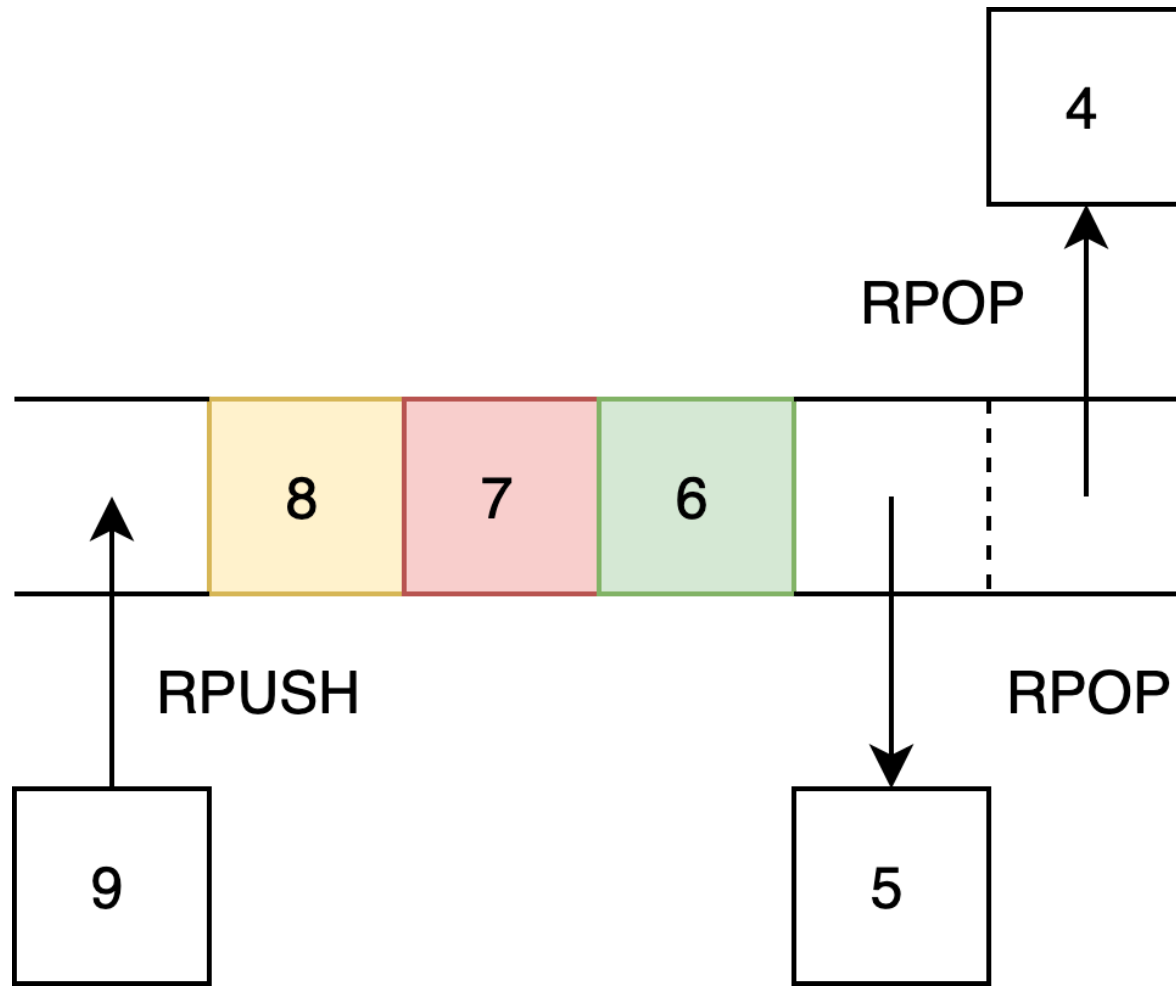
Асинхронные задачи

Повторы

Если не вышло или упал брокер

Различные backend'ы

Blocking list

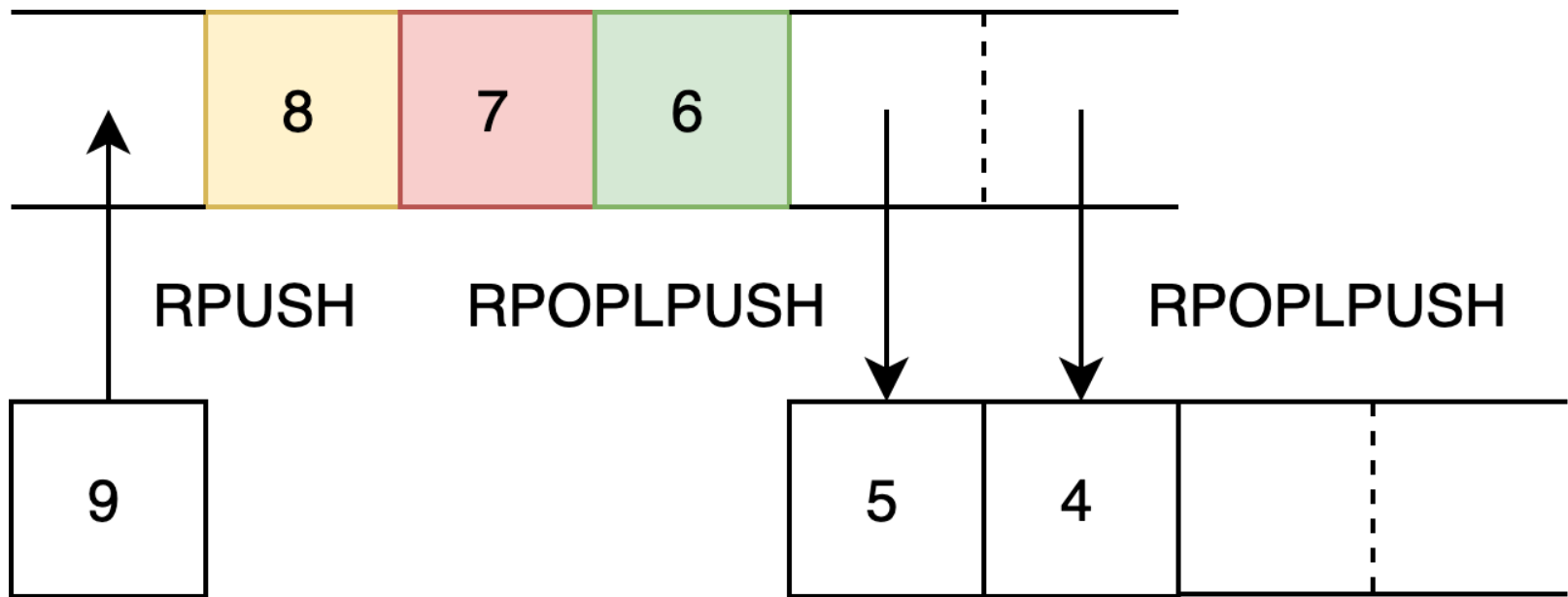


Забрали и сдохли?

Что если обработчик получил SIGKILL



blocked list x2



И маленькая лошадка



возвращает "зависшие" сообщения обратно

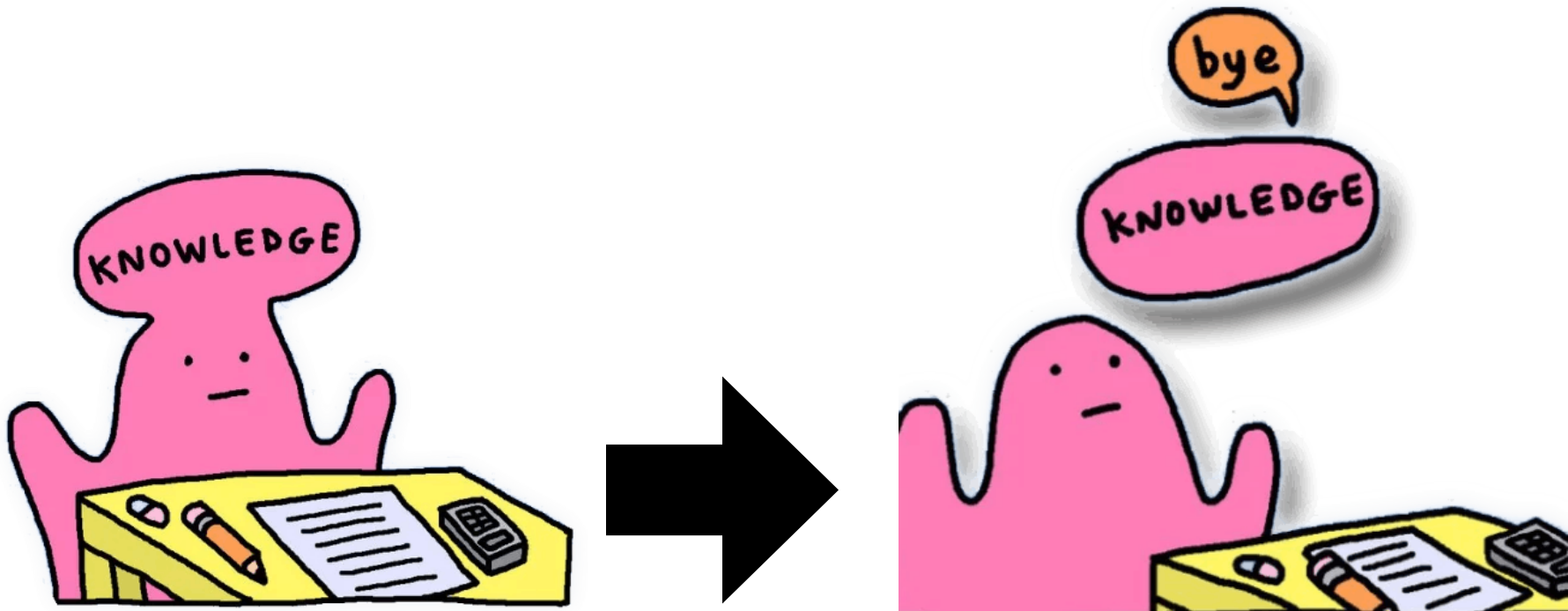
Прочитал — съел

А если мы хотим обработать дважды?



Обработали – забыли

После успешной обработки сообщение не вернуть



Плюсы

- Просто
- Работает
- Даже повторы есть!
- Заявка на успех 👍

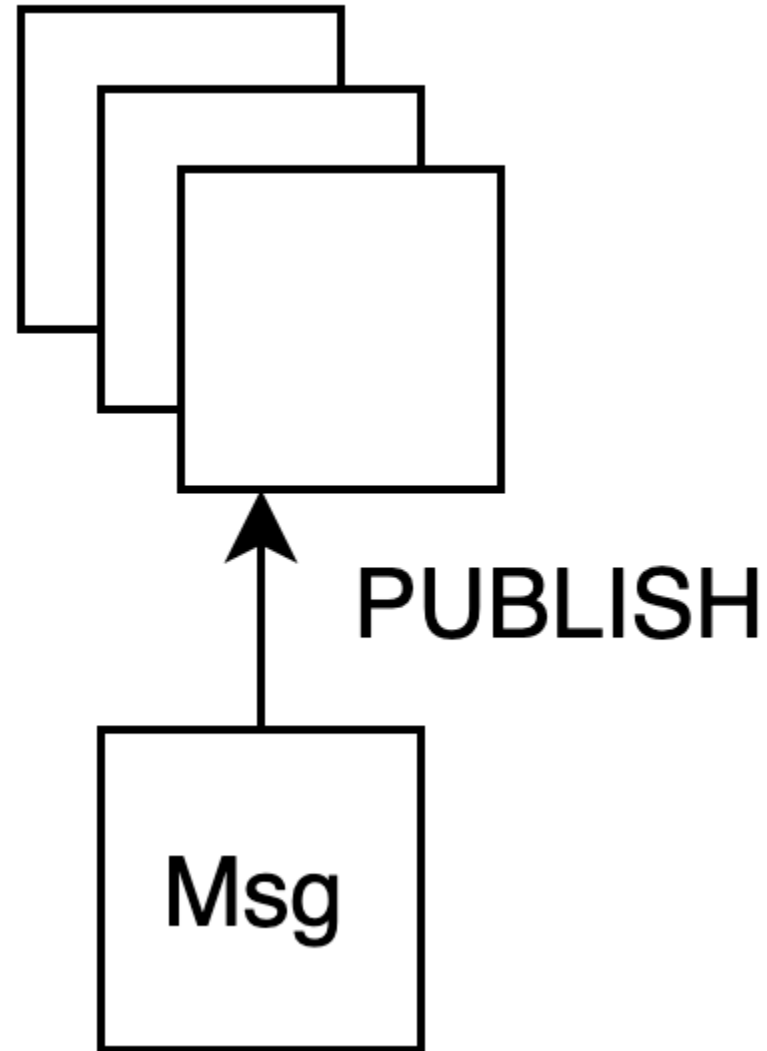
Минусы

- Нужен лишний процесс 🐎
- Если много читателей?
- Каждое сообщение обработать N+ раз?
- А если очередь выросла?

Pub/Sub

Подписка на каналы

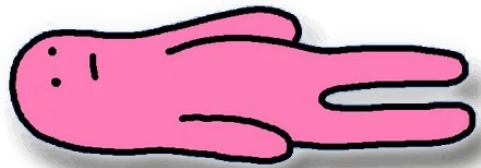
SUBSCRIBE
SUBSCRIBE
SUBSCRIBE



Получают все!

Выстрелил и забыл

Нет никакой истории

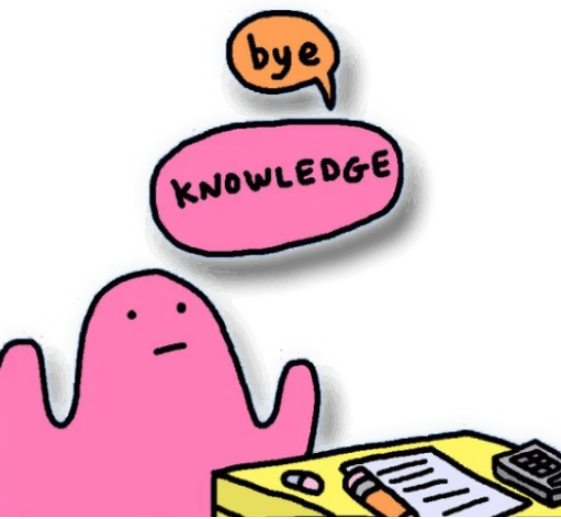


А сообщение прочитали?

Есть только количество каналов в которые доставили

Пропущенные сообщения

Приболел и пропустил лекцию — твои проблемы



Плюсы

- Просто
- Работает
- Дешево
- Много читателей 👍

Минусы

- Нет гарантий
- Нет истории
- "Прогоулы"
- Длинная очередь?

Еще варианты?

Как то еще можно работать с очередями?

Stream

- kafka
- redis stream

Вопросики?

Как пользоваться очередями

flask-dramatiq

```
1 from flask import Flask
2 from flask_dramatiq import Dramatiq
3
4 app = Flask(__name__)
5 dramatiq = Dramatiq(app)
6
7 @dramatiq.actor()
8 def my_actor():
9     ...
10
11 @app.route("/")
12 def myhandler():
13     my_actor.send()
```

Запустить worker как отдельный процесс

```
flask worker --processes=1
```

Docker

Что такое docker?

Зачем docker?

Простота

- Не парится об установке баз данных
- Быстро
- Безопасность
- Воспроизводимость

Задание на неделю

Курс на katacoda.com

[Docker & Containers](#)

Финальные вопросы