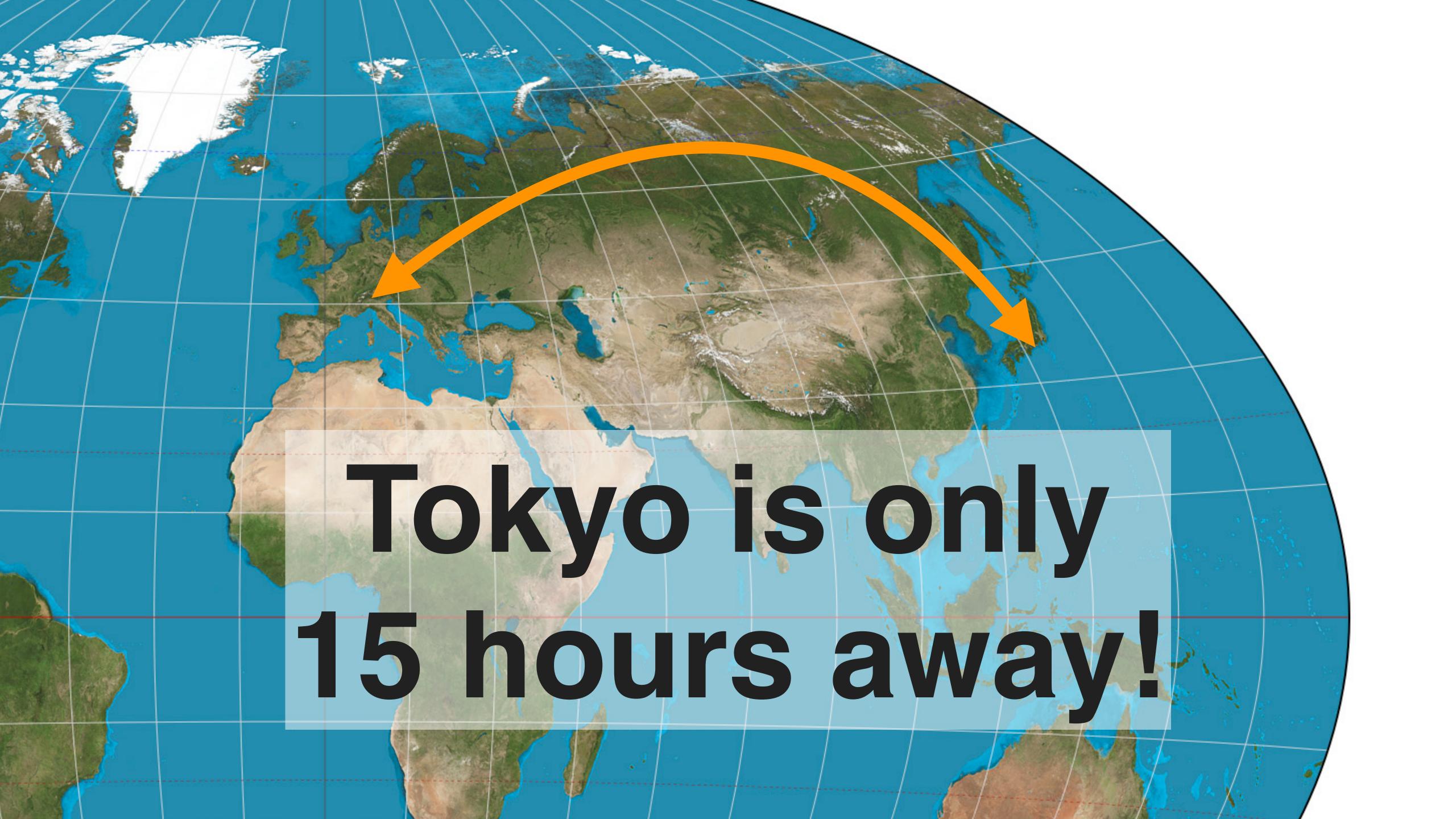
# Improving UX through performance

**Droidcon Italy 2015** 







# Ruby is developed by Matz in Japan



# Cookpad is a recipe sharing service written in RoR















Rank <b></b>	Country/Territory \$	Avg.  connection  speed  (Mbit/s) <sup>[5]</sup>
-	World average	3.9
1	<b>South Korea</b>	25.3
2	Hong Kong	16.3
3	Japan	15.0
4	Switzerland	14.5
5	Sweden	14.1

40	China	3.8
41	Indonesia	3.7
42	Ecuador	3.6
43	Peru	3.6
44	South Africa	3.6
45	Colombia	3.4

# Connection speed in Indonesia is 5x slower than in Japan

#### Performance is a Feature

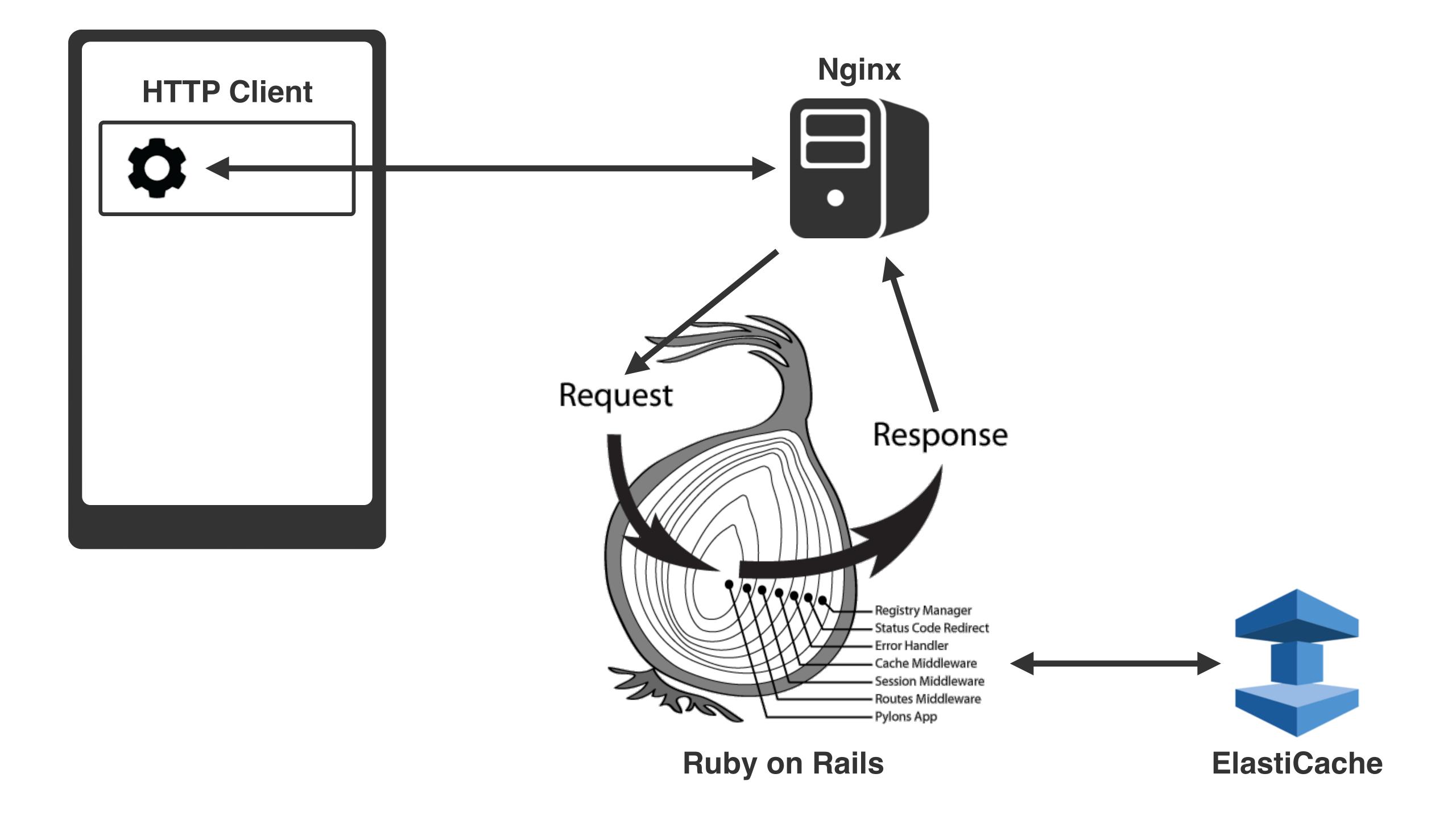
It is becoming increasingly important for mobile engineers to guarantee stable service under any environment

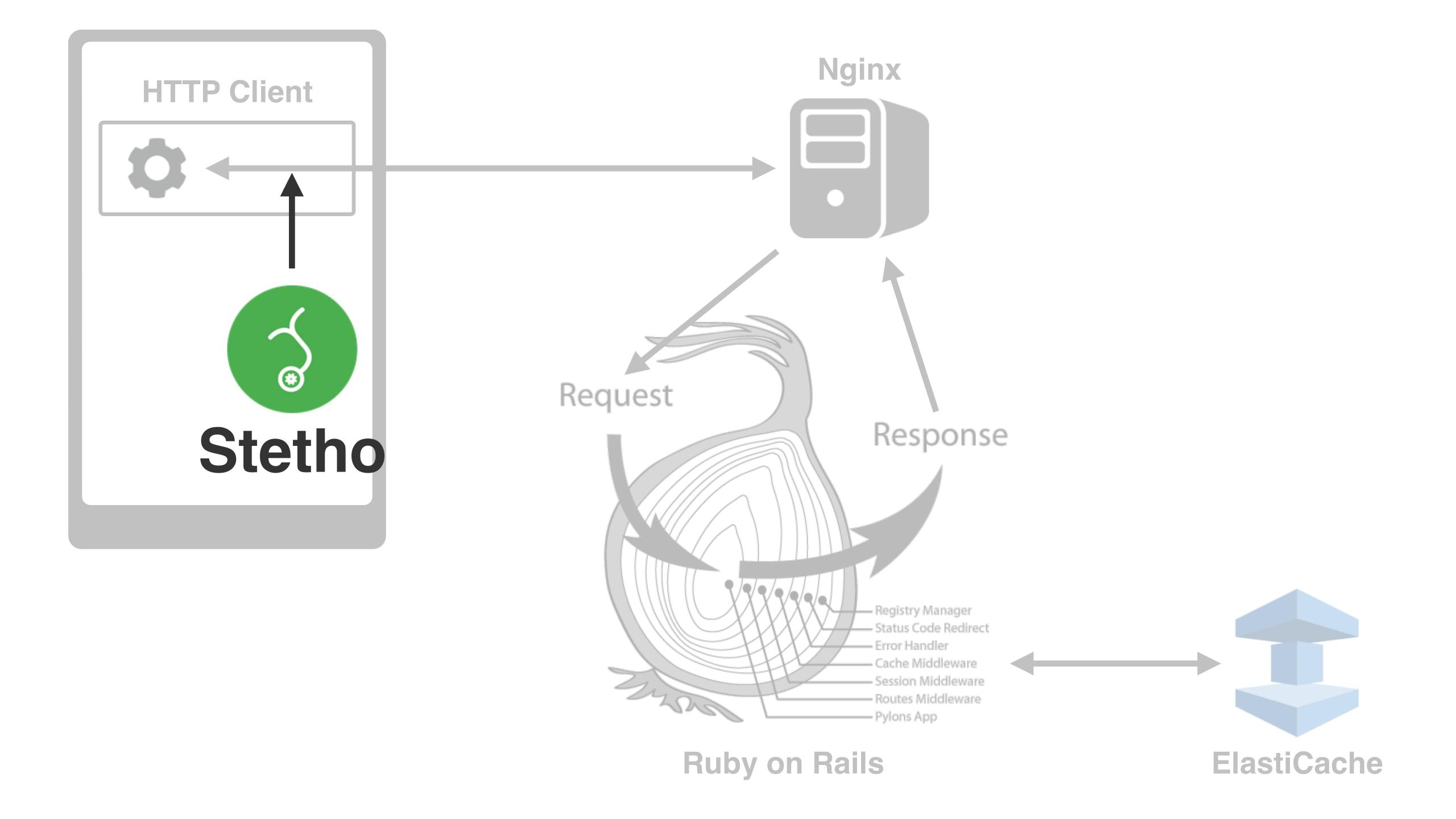


## Agenda

- Efficient HTTP communication
- · Image optimization
- · API design

# Efficient HTTP communication



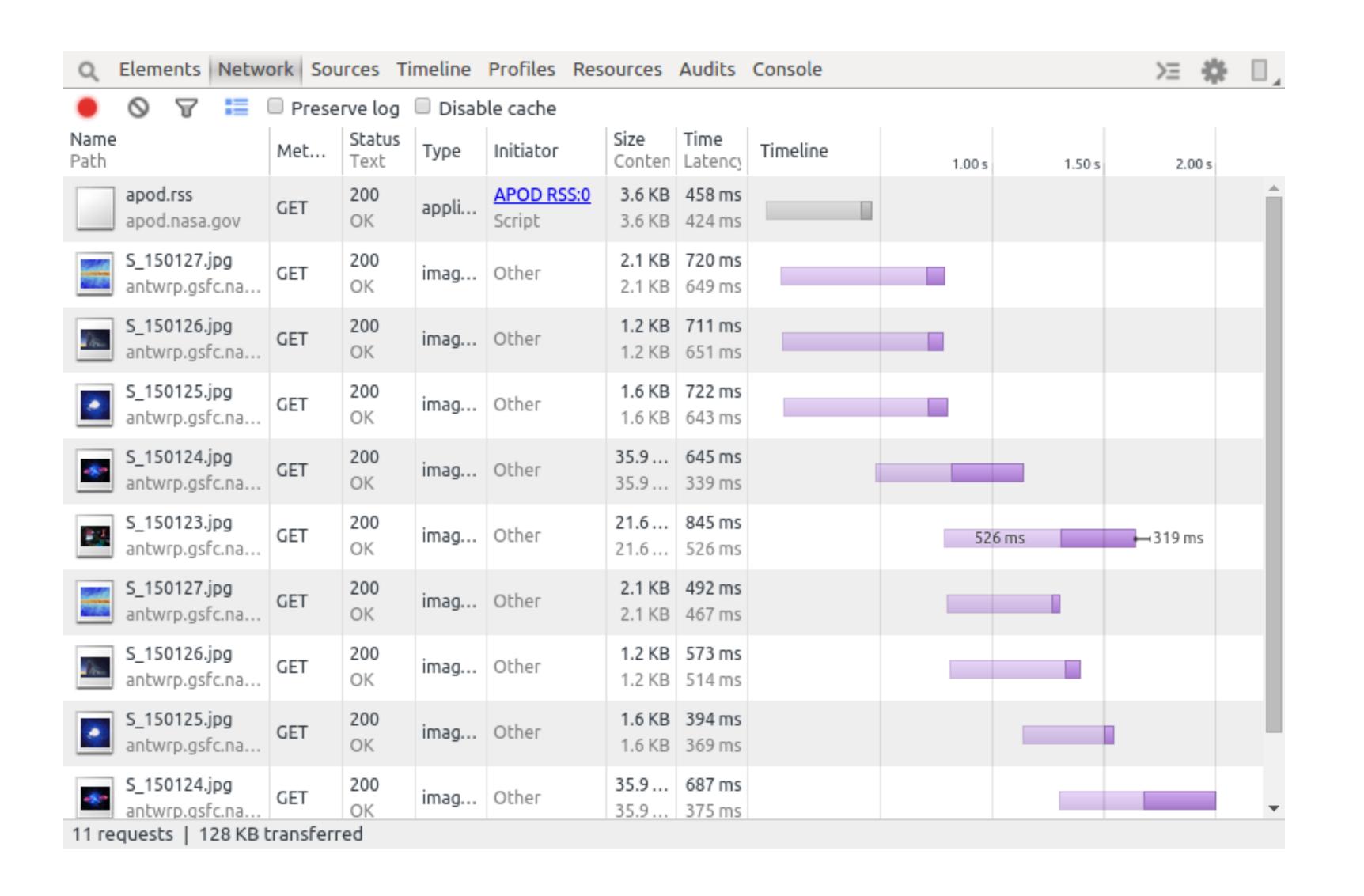




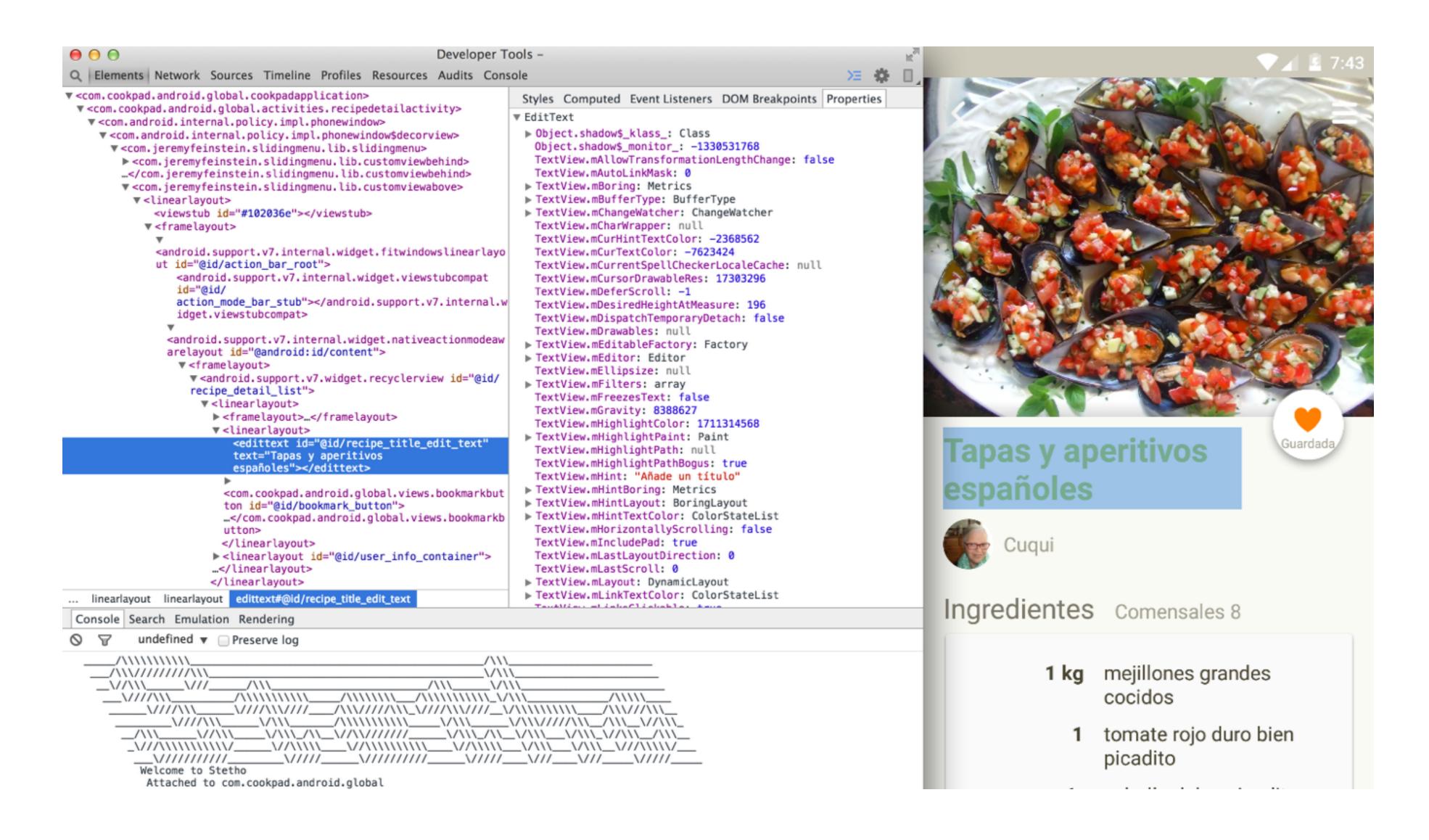
## A debug bridge for Android applications

https://github.com/facebook/stetho

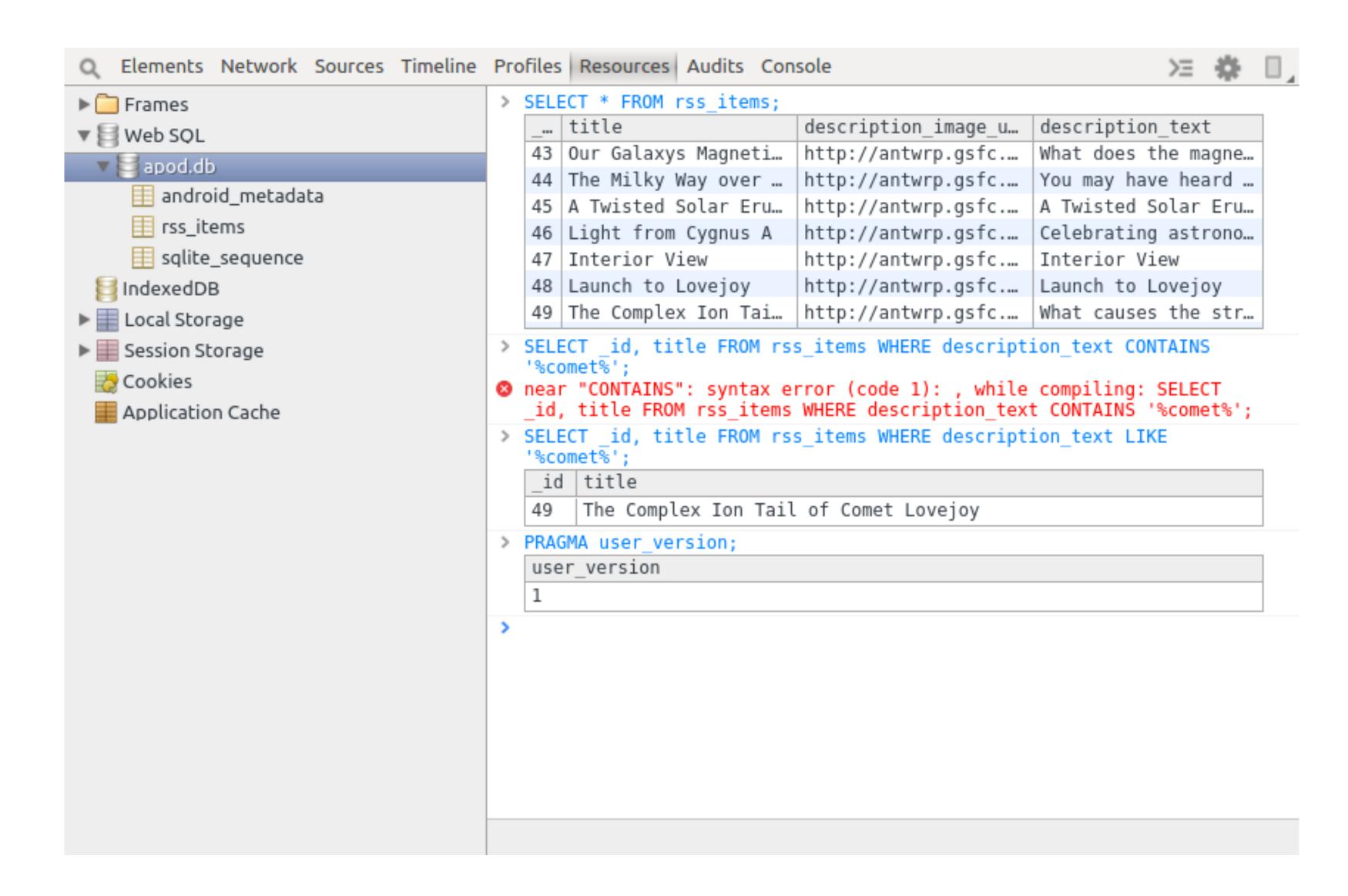
### We can see network



## We can see view hierarchy



## We can access SQLite database



#### ▼ Request Headers

#### Provisional headers are shown

Accept: application/json

#### Accept-Encoding: gzip

Authorization: Bearer 471e22f76e72bf15eb9b14176980c70ca3c5d3f9f3387ed584cd4adfa5702585

Connection: Keep-Alive

Host:

User-Agent: com.cookpad.android.global/1000700; Android/18; Google Nexus 4 - 4.3 - API 18 - 768

X-Cookpad-Guid: a388f5af-2bf0-4e48-82c2-af9bb28fc8ac

X-Cookpad-Provider-Id: 2

#### ▼ Response Headers

Age: 172

Cache-Control: max-age=10800, public

Connection: keep-alive

Content-Encoding: gzip

Content-Type: application/json; charset=utf-8

Date: Sat, 21 Mar 2015 06:32:15 GMT

OkHttp-Received-Millis: 1426919535780

OkHttp-Selected-Protocol: http/1.1

OkHttp-Sent-Millis: 1426919535548

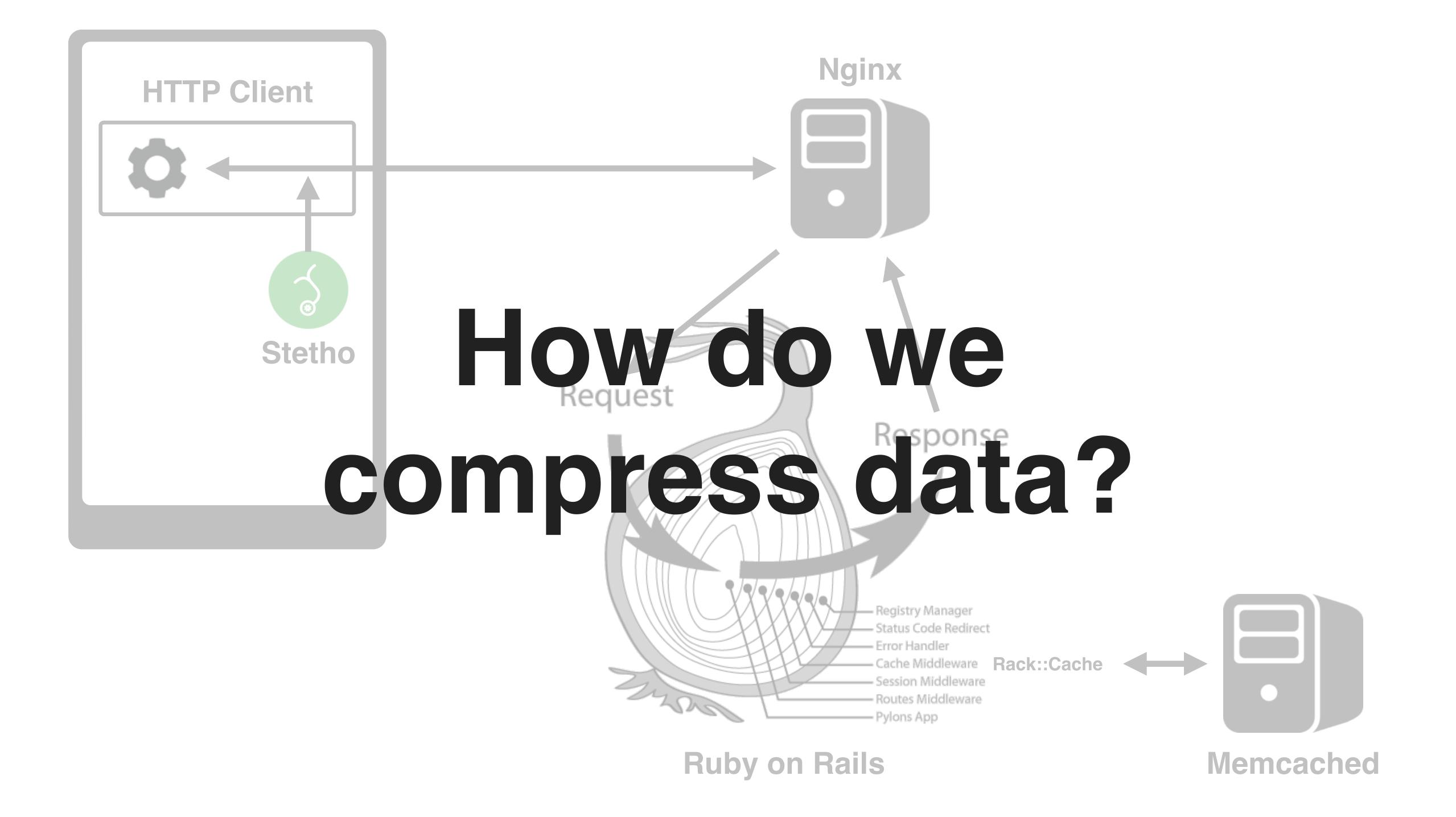
## Compressing Data

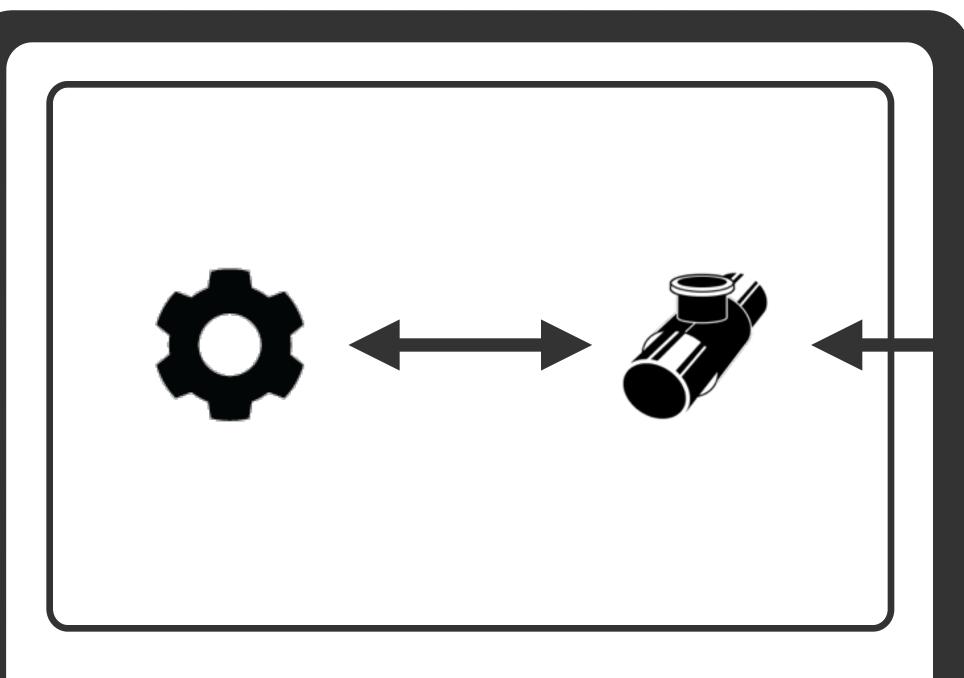
An easy and convenient way to reduce the bandwidth

# Compression is a simple, effective way

**GZIP** reduce the size of response







HTTP Client

Accept-Encoding: gzip

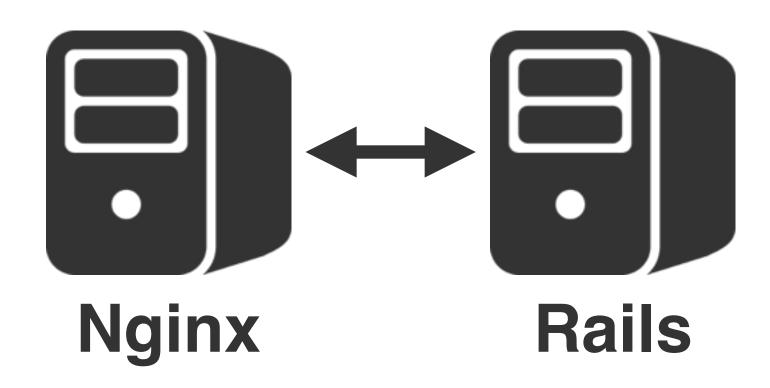
Content-Encoding: gzip

Nginx

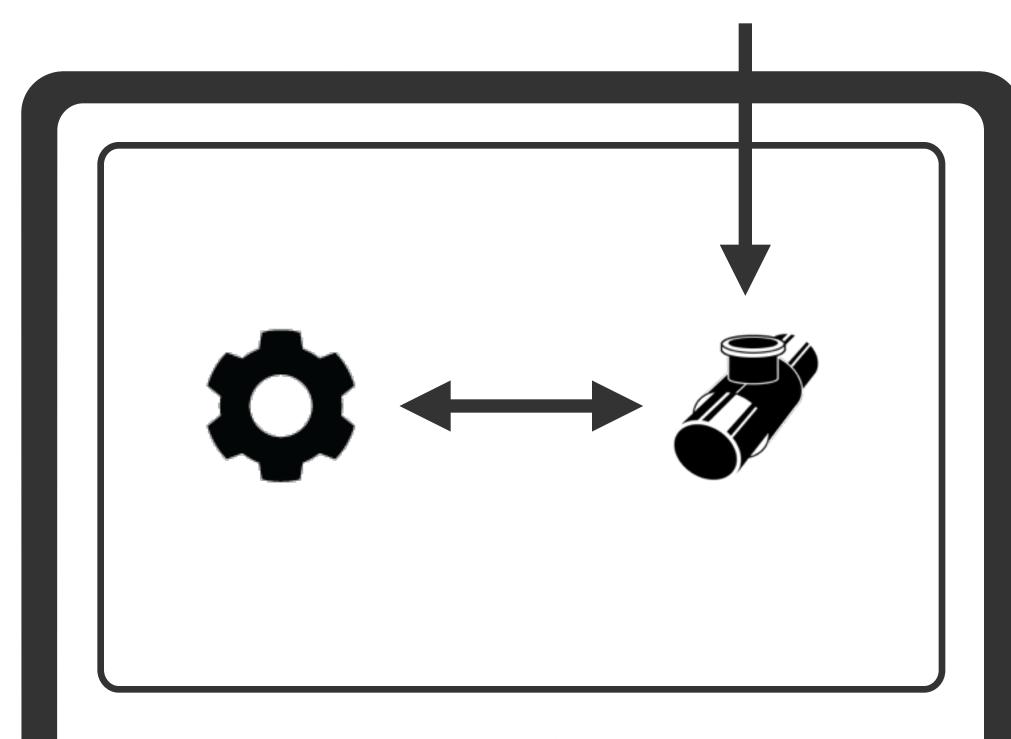
Rails

#### nginx.conf

```
http {
 gzip on;
 gzip_disable "msie6";
 gzip_vary on;
 gzip_proxied any;
 gzip_comp_level 6;
 gzip_buffers 16 8k;
 gzip_http_version 1.1;
 gzip_types text/plain text/css application/json
```



#### **GZIP** decoder



HTTP Client

```
// Set "Accept-Encoding: gzip" when you send a request
connection.setRequestProperty(
    "Accept-Encoding", "gzip");

// Decompress input stream when you receive a response
inputStream = new GZIPInputStream(
    connection.getInputStream());
```

## HTTP clients for Android

Don't support GZIP by default

- AndroidHttpClientHttpUrlConnection
- OkHttp

support GZIP by default

## HTTP clients for Android

@Deprecated

- · AndroidHttpClient
- · HttpUrlConnection
- · OkHttp

No longer maintained

# We had used Volley as API client before

## Volley has 2 HTTP clients internally

#### Volley

2.3+: HttpUrlConnection

<2.2: AndroidHttpClient

```
public static RequestQueue newRequestQueue(...) {
    ...
    if (stack == null) {
        if (Build.VERSION.SDK_INT >= 9) {
            // use HttpUrlConnection
            stack = new HurlStack();
        } else {
            // use AndroidHttpClient
            stack = new HttpClientStack(AndroidHttpClient)
        }
    }
}
```

# HttpUrlConnection uses OkHttp internally

#### HttpUrlConnection

4.4+: OkHttp

< 4.4: HttpUrlConnection



### Different behavior of HTTP clients

#### Inside of Volley

4.4+: OkHttp

<2.3: AndroidHttpClient



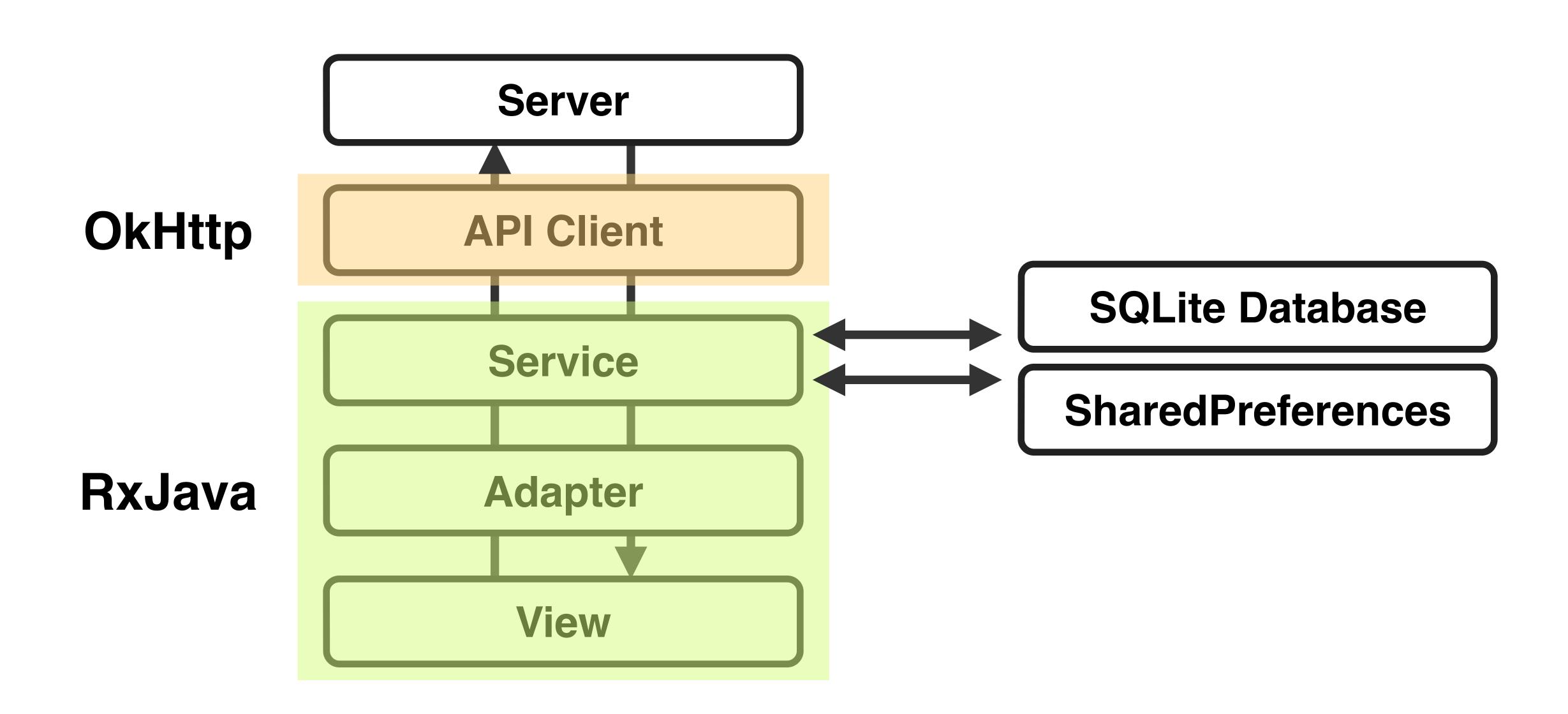
## Simple is better



# I recommend to use OkHttp

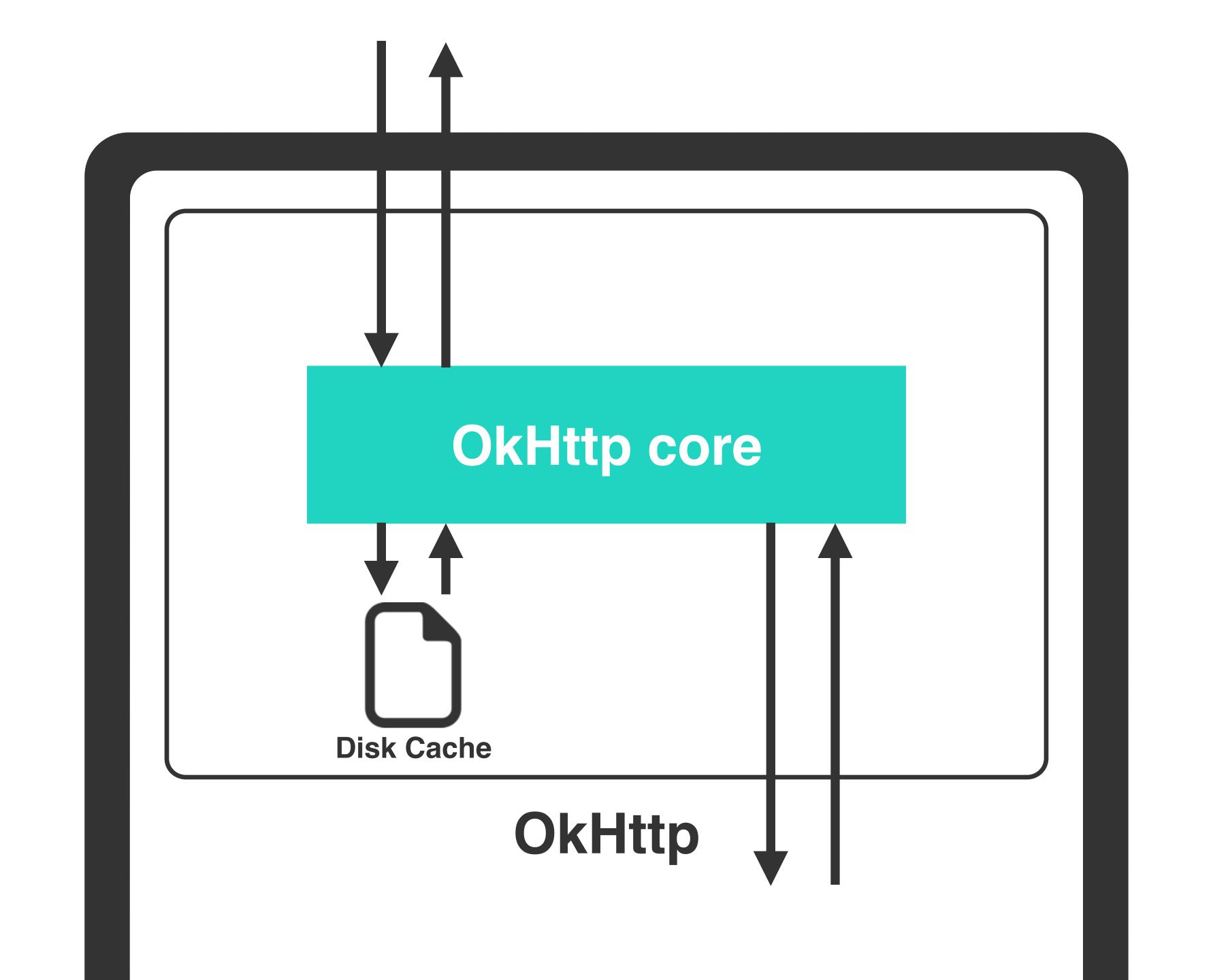
- \* GZIP
- \* Connection Pool
- \* WebSocket
- \* HTTP/2.0

### OkHttp + RxJava = Reactive Data Store



# Caching Data

Effective cache controls will dramatically reduce server load



# Caching in HTTP

```
cache-request-directive =
    "no-cache"
    I "no-store"
    I "max-age" "=" delta-seconds
    I "max-stale" [ "=" delta-seconds ]
    I "min-fresh" "=" delta-seconds
    I "no-transform"
    I "only-if-cached"
    I cache-extension
```

```
cache-response-directive =
    "public"
I "private" [ "=" <"> 1#field-name <"> ]
I "no-cache" [ "=" <"> 1#field-name <"> ]
I "no-store"
I "no-transform"
I "must-revalidate"
I "proxy-revalidate"
I "proxy-revalidate"
I "s-maxage" "=" delta-seconds
I "s-maxage" "=" delta-seconds
I cache-extension
```

### Enable cache

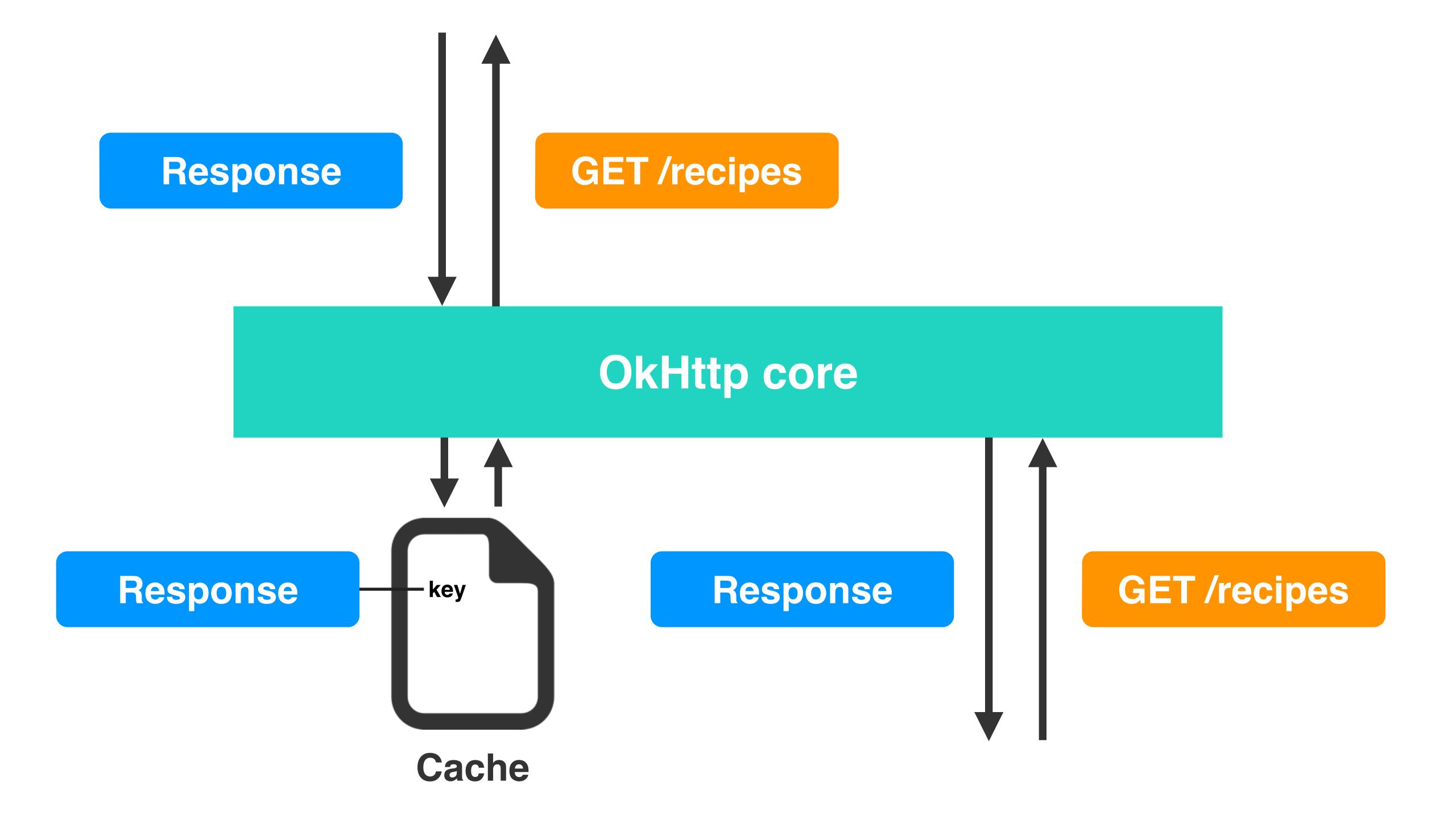
```
OkHttpClient client = new OkHttpClient();
Cache cache = new Cache(cacheDir, MAX_CACHE_SIZE);
client.setCache(cache);
```

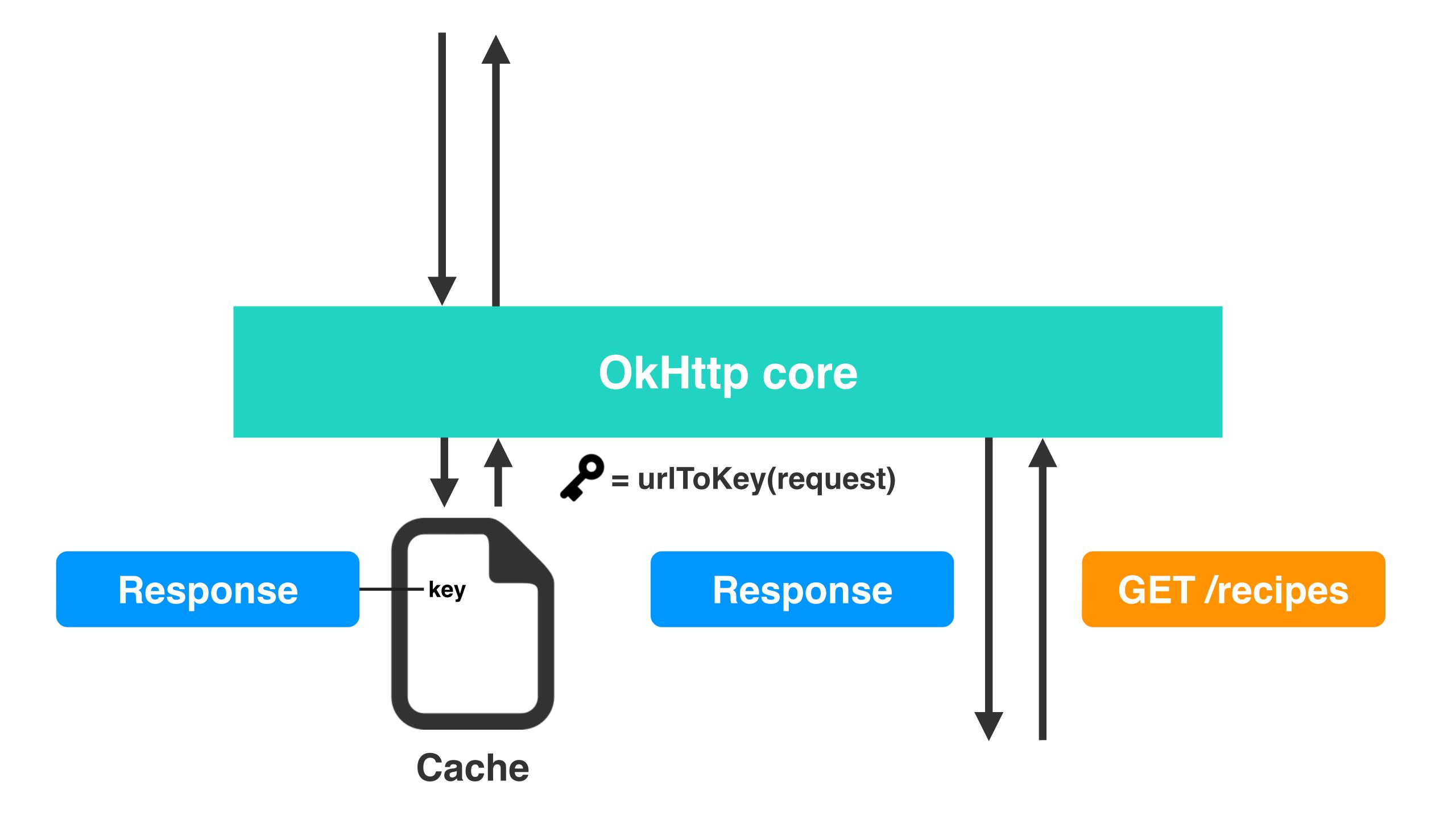
```
# default
# => Cache-Control: max-age=0, private, must-revalidate

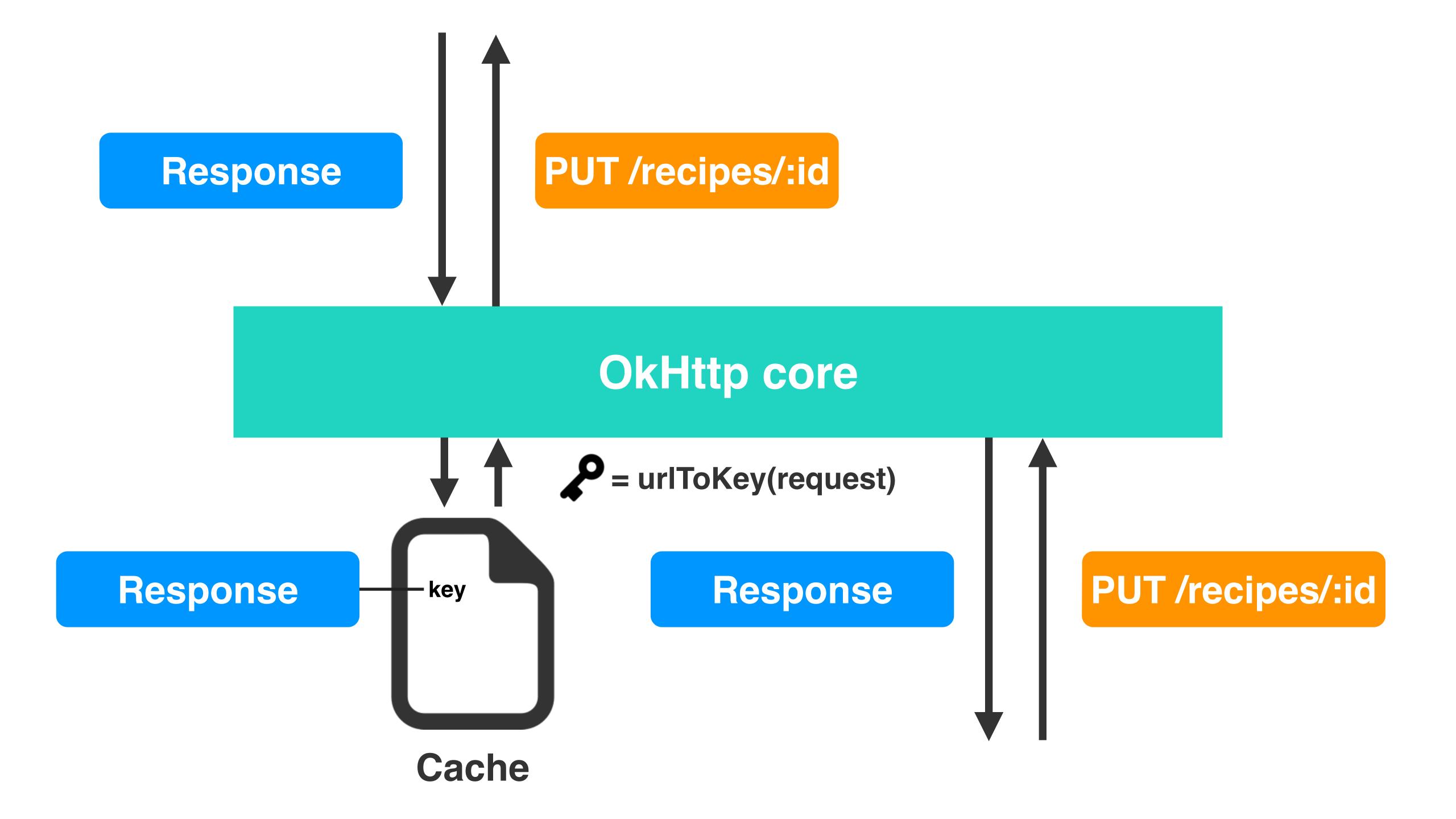
expires_in(1.hour, public: true)
# => Cache-Control: max-age=3600, public

expires_now
# => Cache-Control: no-cache
```







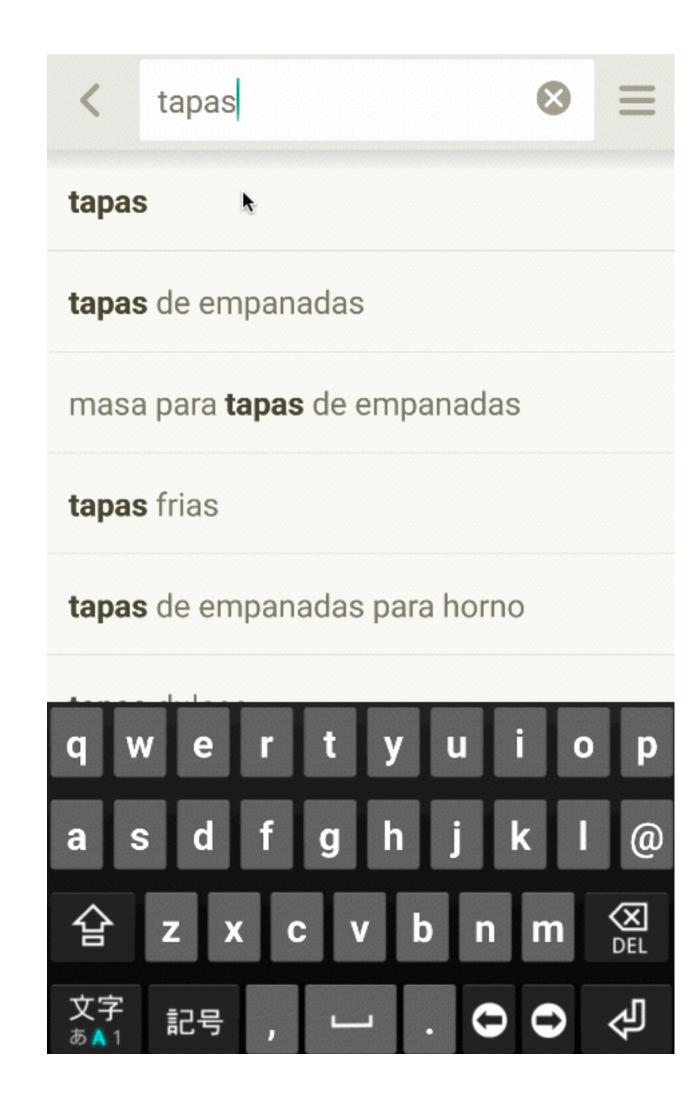


#### Cache-Control: no-cache



In some situations, such as after a user clicks a 'refresh' button, it may be necessary to skip the cache, and fetch data directly from the server

```
// ApiClient.java
if (isConnected) {
    headers.put(CACHE_CONTROL, "only-if-cached");
} else if (noCache && noStore) {
    headers.put(CACHE_CONTROL, "no-cache, no-store");
} else if (noCache) {
    headers.put(CACHE_CONTROL, "no-cache");
} else if (noStore) {
    headers.put(CACHE_CONTROL, "no-store");
}
```



# Users can see contents quickly even if device is not connected

# To enjoy the benefits of caching, you need to write carefully crafted cache control policies

Object Type	Duration
Categories	1 day
Search recipes	3 hours

Do not cache

Users

# Image Optimization



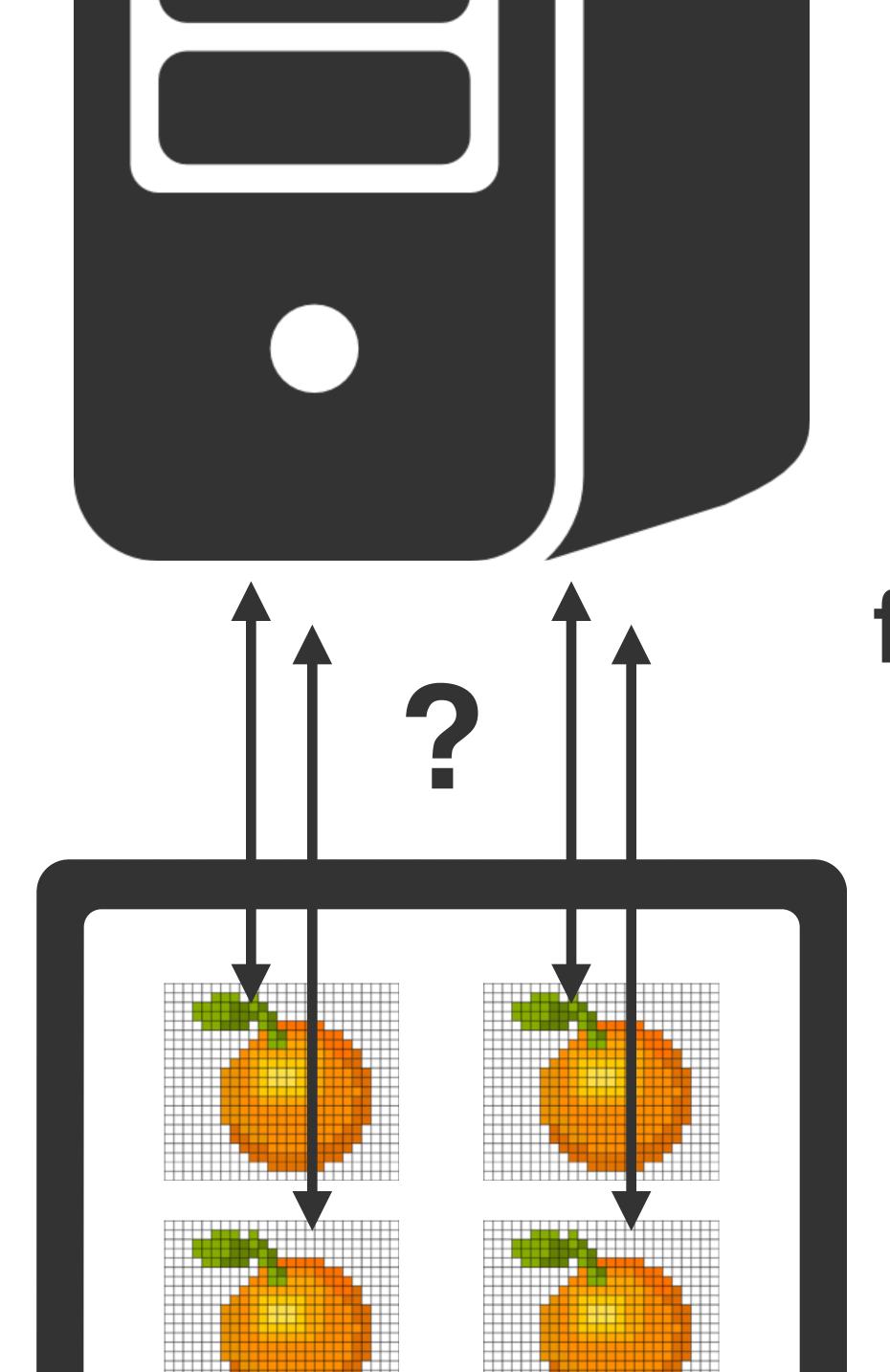


■ Each pixel takes up 4 bytes

# We need to know what image loading is

# Simple Image Loading

- Specify URL to HTTP client
- · Get Input Steam
- · Decode Input Stream to Bitmap
- · Set Bitmap to ImageView



Do you fetch images from the server every time you want to display images?

#### The answer may be

# 

# In addition, we want to

- · reuse worker threads
- · set the priority of requests
- · cache decoded images

### There are some great libraries



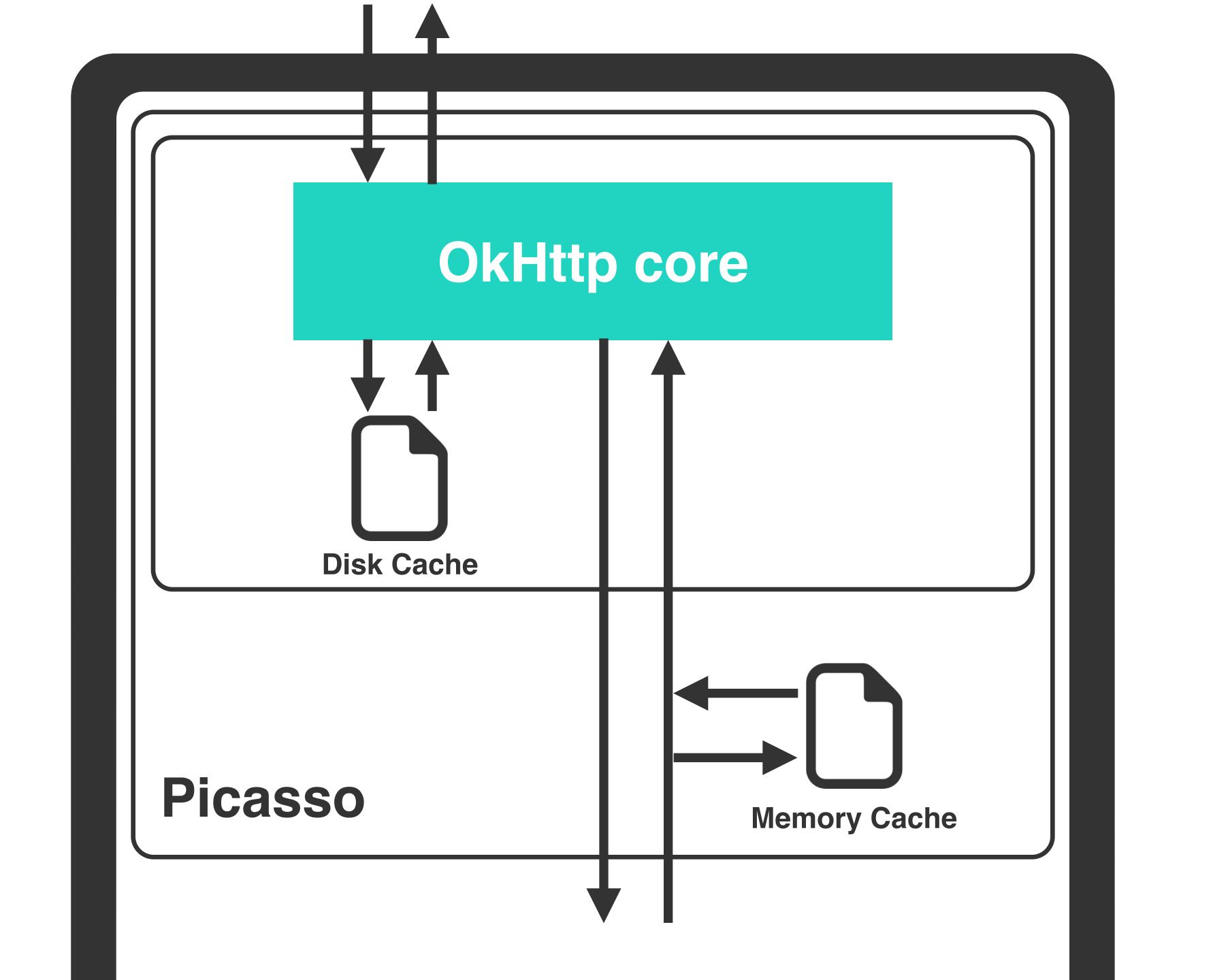
Picasso





# Caching Data

The best way to display images quickly



# Expiration time

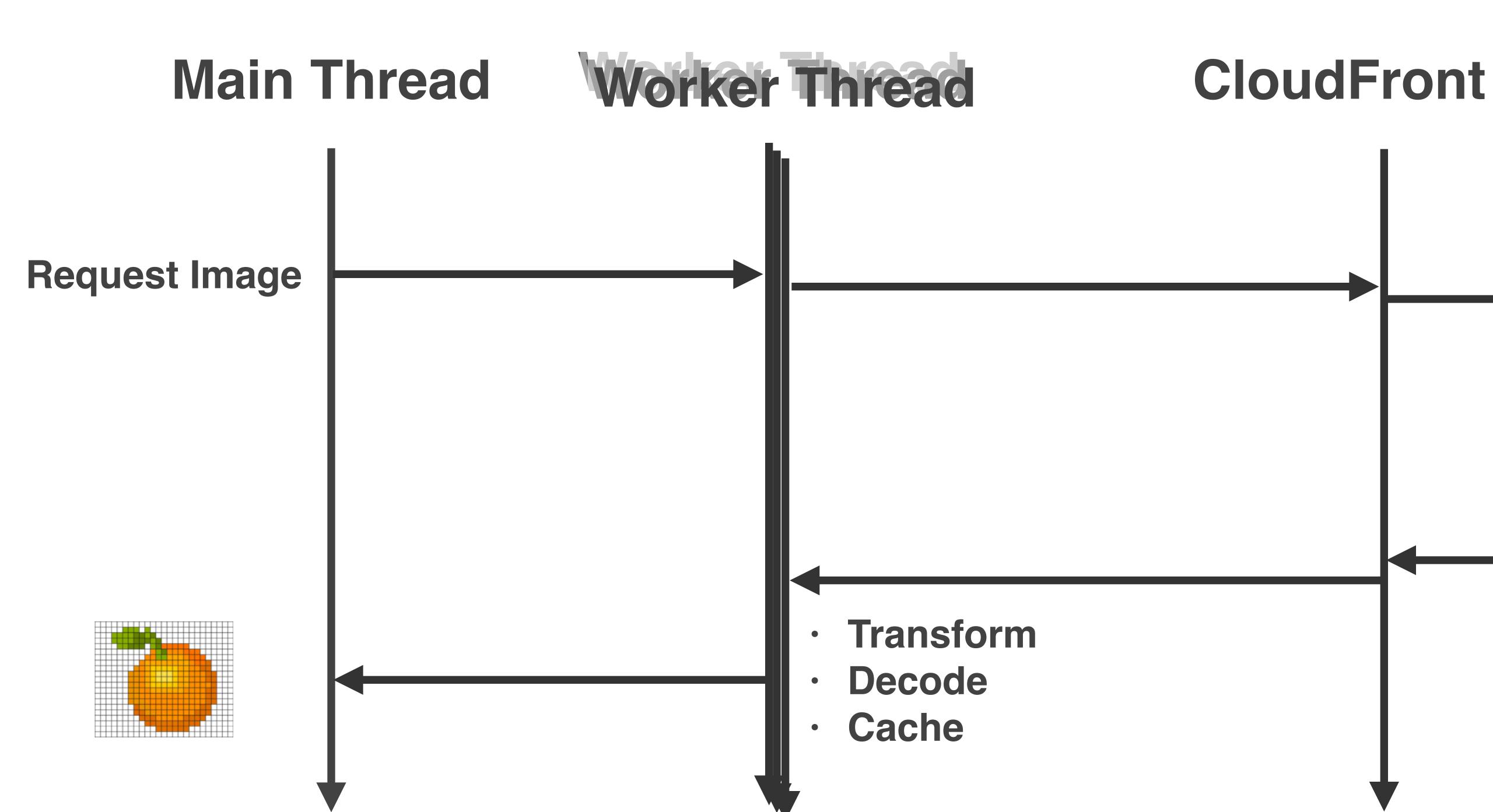
Expiration times of cache is also following cache controls

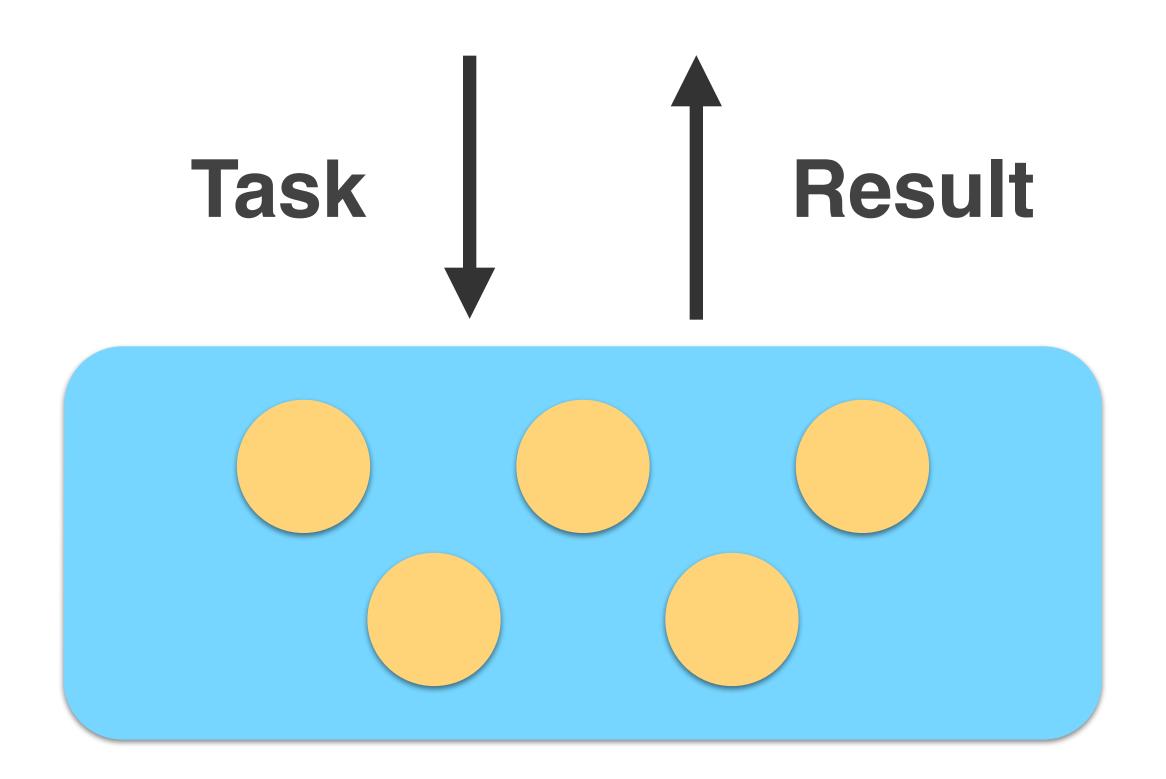
### Enable cache

Picasso setup cache automatically You don't need to do anything

# Thread Pool

Creating new threads for each task incur the overhead

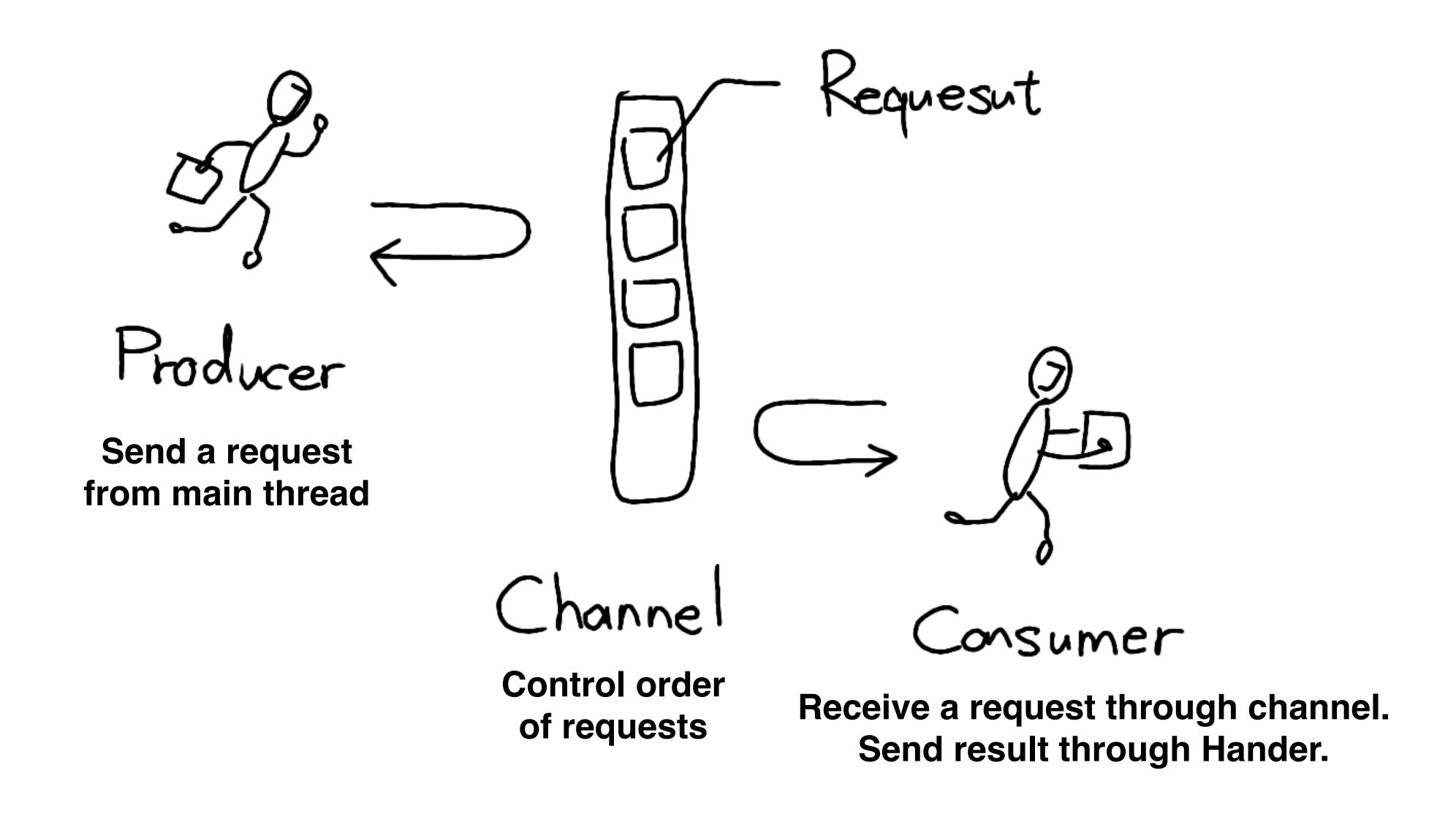




#### new ThreadPoolExecutor(

corePoolSize, // The number of threads to keep in the pool maximumPoolSize, // The maximum number of threads to allow in the pool keepAliveTime, // the maximum time that excess idle threads will wait for new tasks timeUnit, // for the keepAliveTime argument workQueue, // the queue to use for holding tasks before they are executed threadFactory // The factory to use when the executor creates a new thread );

#### Producer-consumer pattern



# There is a trade-off between capacity and resource

If there are many workers, tasks are processed concurrently. If there are too many workers, consume memory wastefully.

#### Picasso

```
switch (info.getType()) {
 case ConnectivityManager.TYPE_WIFI:
 case ConnectivityManager.TYPE_WIMAX:
 case ConnectivityManager.TYPE_ETHERNET:
   setThreadCount(4);
   break;
 case ConnectivityManager.TYPE_MOBILE:
   switch (info.getSubtype()) {
     case TelephonyManager.NETWORK_TYPE_LTE: // 4G
     case TelephonyManager.NETWORK_TYPE_HSPAP:
     case TelephonyManager.NETWORK_TYPE_EHRPD:
        setThreadCount(3);
        break;
     case TelephonyManager.NETWORK_TYPE_UMTS: // 3G
     case TelephonyManager.NETWORK_TYPE_CDMA:
     case TelephonyManager.NETWORK_TYPE_EVDO_0:
     case TelephonyManager.NETWORK_TYPE_EVDO_A:
     case TelephonyManager.NETWORK_TYPE_EVDO_B:
        setThreadCount(2);
        break;
     case TelephonyManager.NETWORK_TYPE_GPRS: // 2G
     case TelephonyManager.NETWORK_TYPE_EDGE:
        setThreadCount(1);
        break;
```

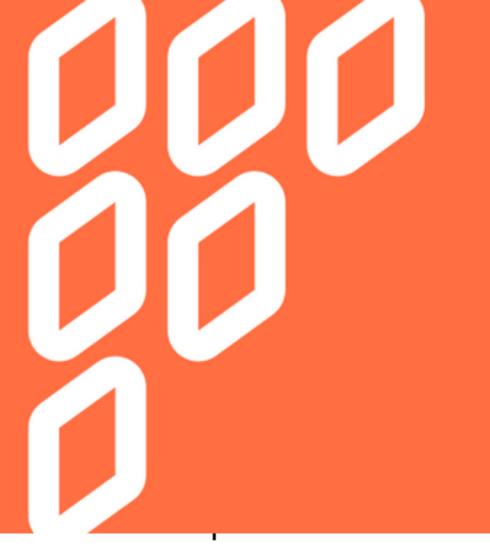
#### Glide

Runtime.getRuntime().availableProcessors()

## Which setting is better?

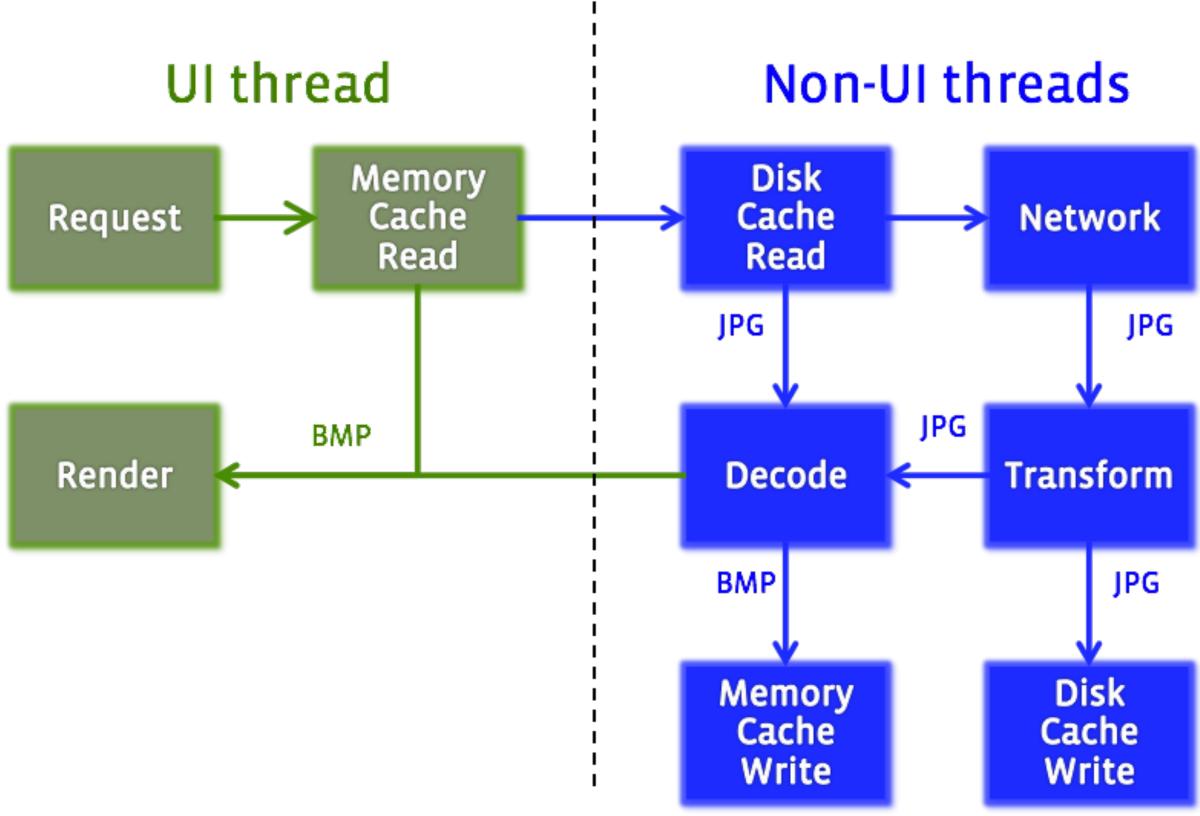
	Picasso	Glide
Nexus 6 in Tokyo	3 (LTE)	4 (Quad-core 2.7 GHz Krait 450)
Galaxy ace 3 in Jakarta	2 (3G)	2 (Dual-core 1 GHz Cortex-A9)
MiTO Impact (Android One) in Jakarta	2 (3G)	4 (Cortex A7 1.3 GHz Quad-Core)

It is depending on network environment, device spec, image size, transformation, ...



#### Fresco

A new image loading library developed by Facebook



### Fresco has multiple Executors

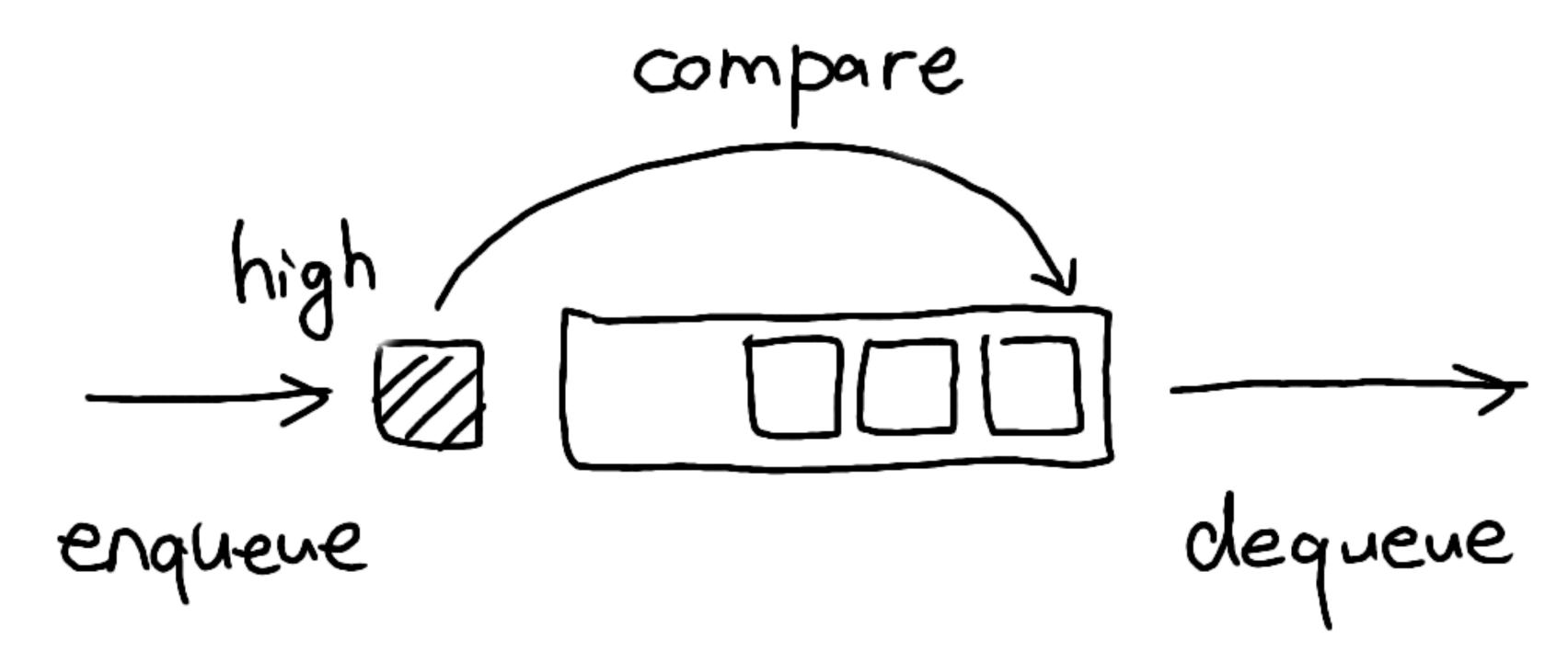
Process	Kind of Executor
forLocalStorageRead	IoBoundExecutor
forLocalStorageWrite	IoBoundExecutor
forDecode	CpuBoundExecutor
forBackground	CpuBoundExecutor

NUM\_IO\_BOUND\_THREADS = 2; NUM\_CPU\_BOUND\_THREADS = Runtime.getRuntime().availableProcessors();

# Queue Management

Control order of requests

### PriorityBlockingQueue



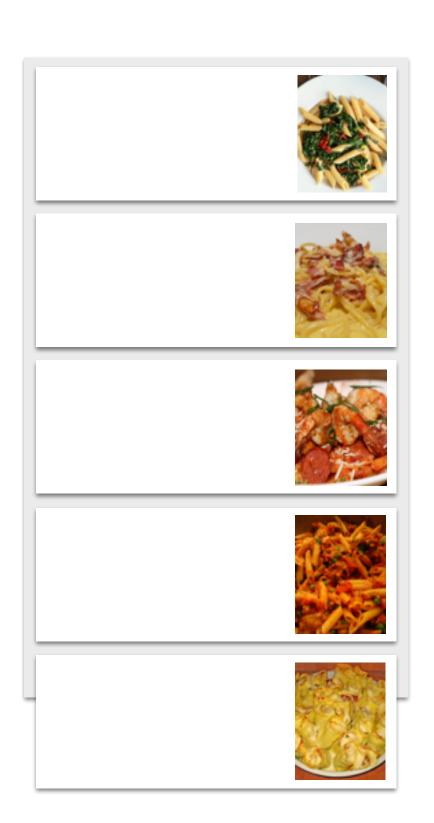
The elements order themselves according to whatever priority you decided in your Comparable implementation

### We can set priority to request

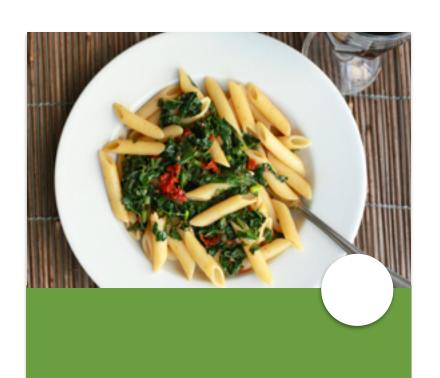
```
Picasso.with(this)
.load(url)
.priority(HIGH)
.into(imageView);
```

```
Glide.with(this)
.load(url)
.priority(HIGH)
.into(imageView);
```

### How priority works?



When a user open recipe detail screen, requests are added to the end of the queue



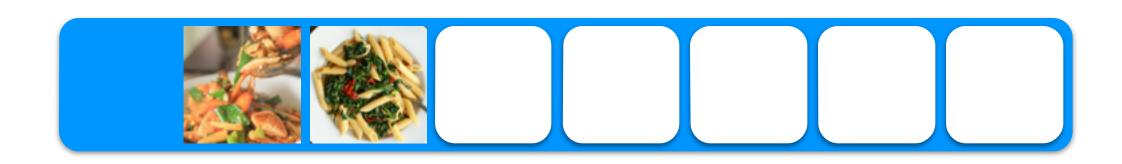
### Directions

1

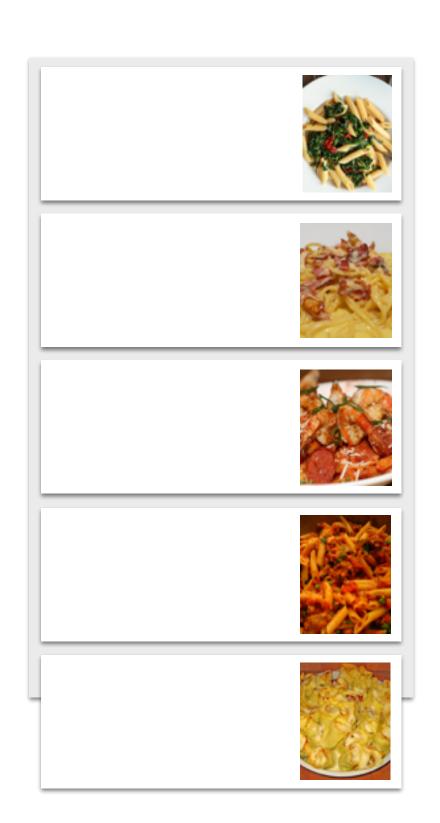
Heat the oil in a large pot over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes

Heat the oil in a large por over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes





### How priority works?



When the user open recipe detail screen, set HIGH priority to the main image

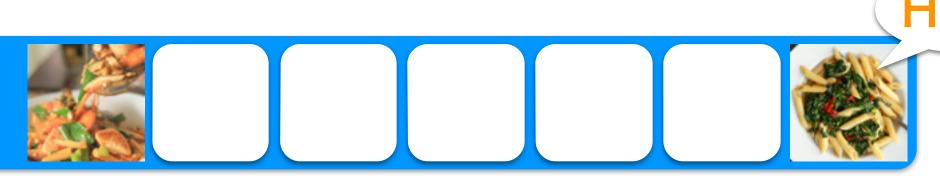


### Directions

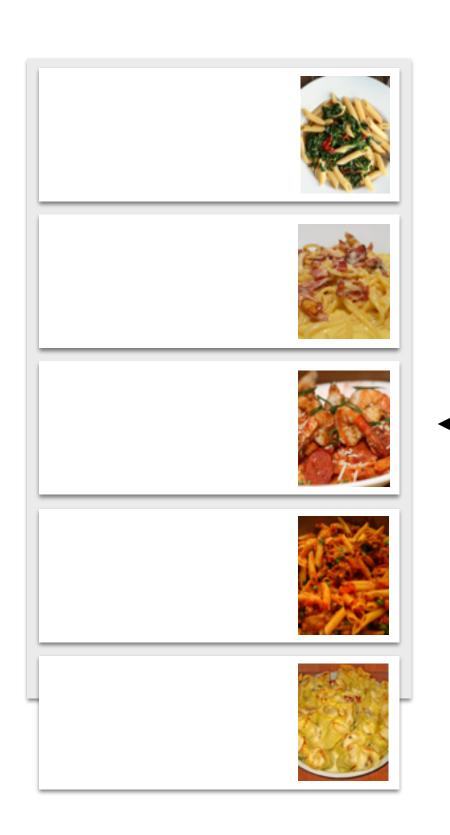
Heat the oil in a large pot over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10

Heat the oil in a large po over a medium heat and ently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes

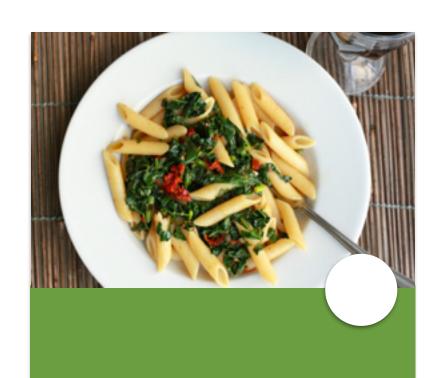




### How priority works?



when the user back to recipe list screen, call "cancelTag" to dispose useless requests



### Directions

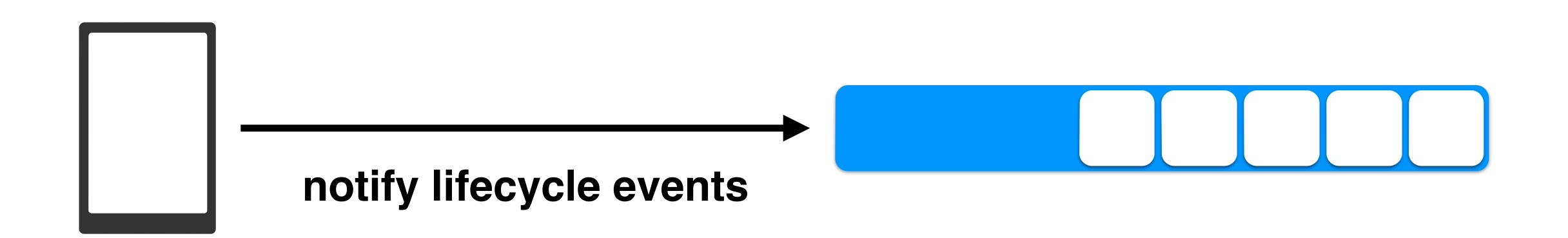
1

Heat the oil in a large pot over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10

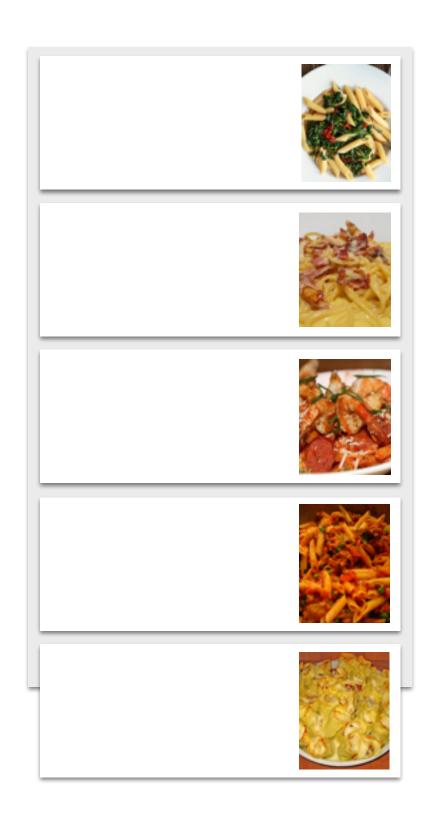
Heat the oil in a large poover a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes



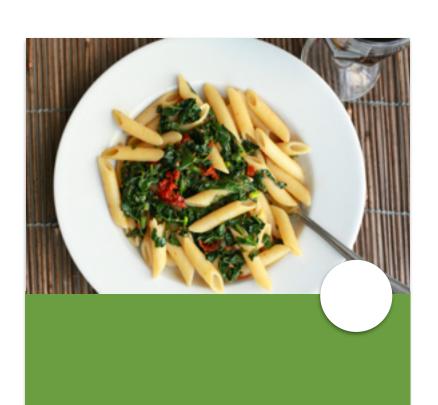
### Glide has lifecycle integration



### Glide manage the queue automatically



Requests in search result screen are paused automatically



#### Directions

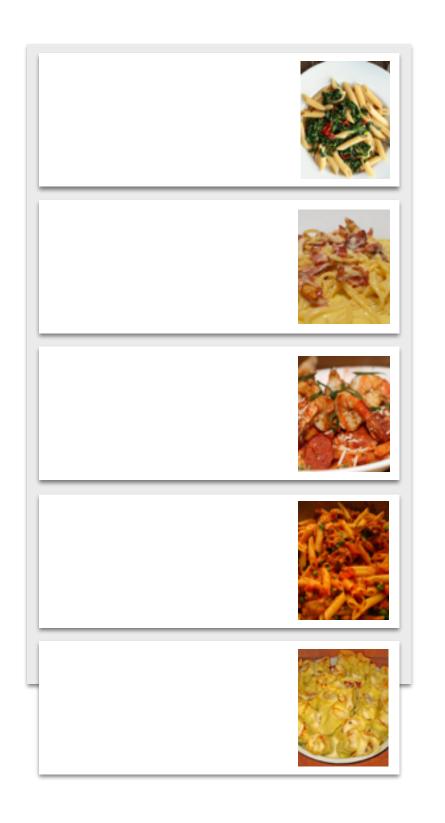
Heat the oil in a large pot over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes

Heat the oil in a large pot over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes



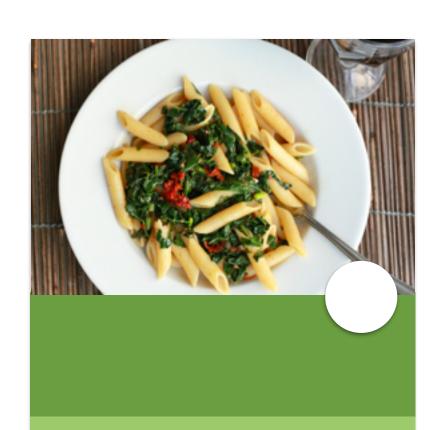


### Glide manage the queue automatically



Requests in recipe detail screen are cancelled automatically

Requests in search recipe list is restarted automatically



#### Directions

Heat the oil in a large pot over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes

Heat the oil in a large pot over a medium heat and gently fry off the onion, celery, carrot, and garlic until tender, around 10 minutes



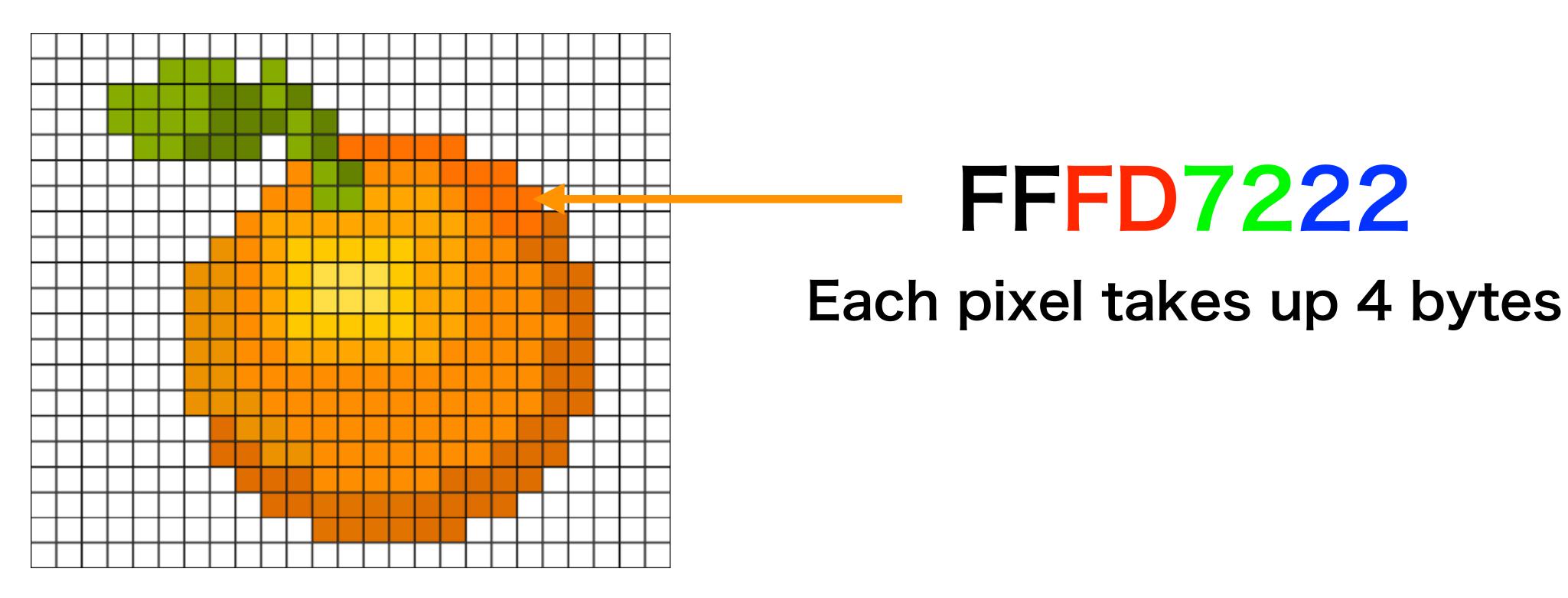
### Notice:

Glide adds view-less fragment to each Activity to observe lifecycle events.

## Bitmap Pool

Reuse memory when new Bitmap is requested

### Memory management for Bitmap



25 px \* 21 px \* 4 byte = 2,400 byte

## Glide has Bitmap Pool

reuse resources to avoid unnecessary allocations

### Request a Bitmap

width, height, config Bitmap



4.4+: SizeStrategy

<4.4: AttributeStrategy



### Managing Bitmap Memory

### Manage Memory on Android 3.0 and Higher

Android 3.0 (API level 11) introduces the BitmapFactory. Options. in Bitmap field. If this option is set, decode methods that take the Options object will attempt to reuse an existing bitmap when loading content. This means that the bitmap's memory is reused, resulting in improved performance, and removing both memory allocation and de-allocation. However, there are certain restrictions with how inBitmap can be used. In particular, before Android 4.4 (API level 19), only equal sized bitmaps are supported. For details, please see the inBitmap documentation.



4.4+: SizeStrategy

<4.4: AttributeStrategy

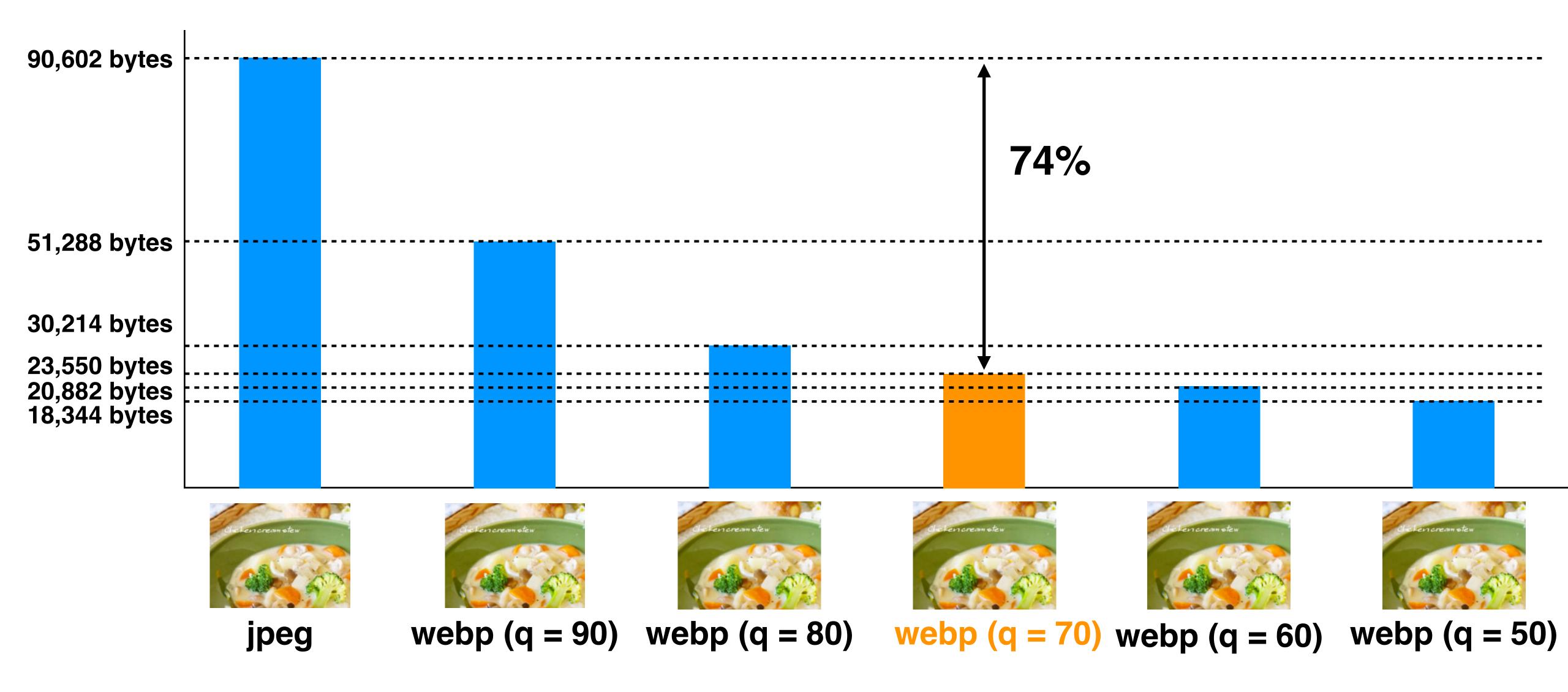
## Image Format

We are using WebP that is an image format developed by Google



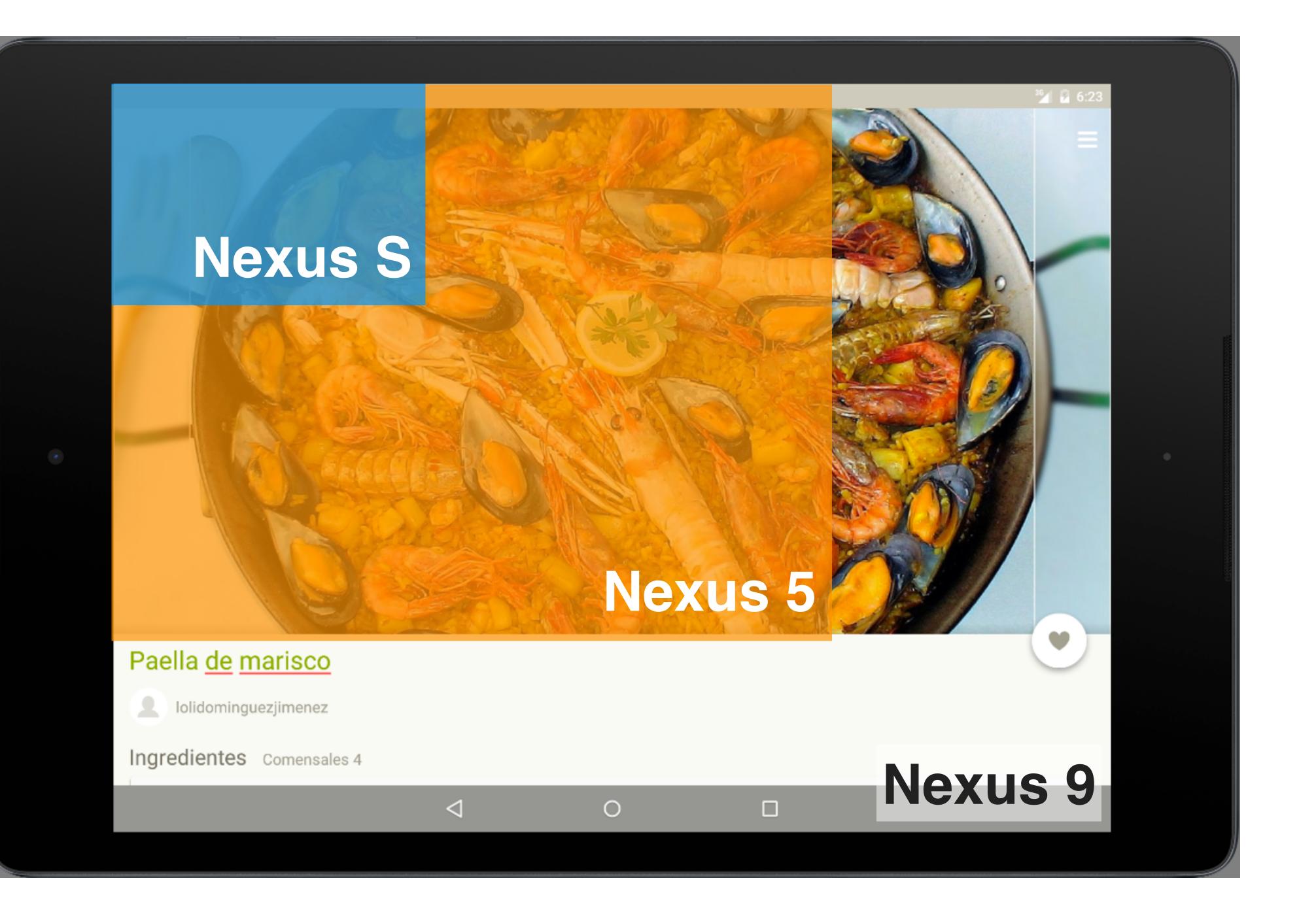
WebP lossless images are 26% smaller in size compared to PNGs WebP lossy images are 25-34% smaller in size compared to JPEGs

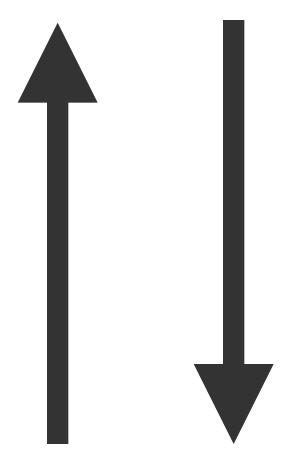
## Comparison of image size



## Image Size

Request an appropriate image size





### http://.../1080x756/photo.webp



Daeta al horno con

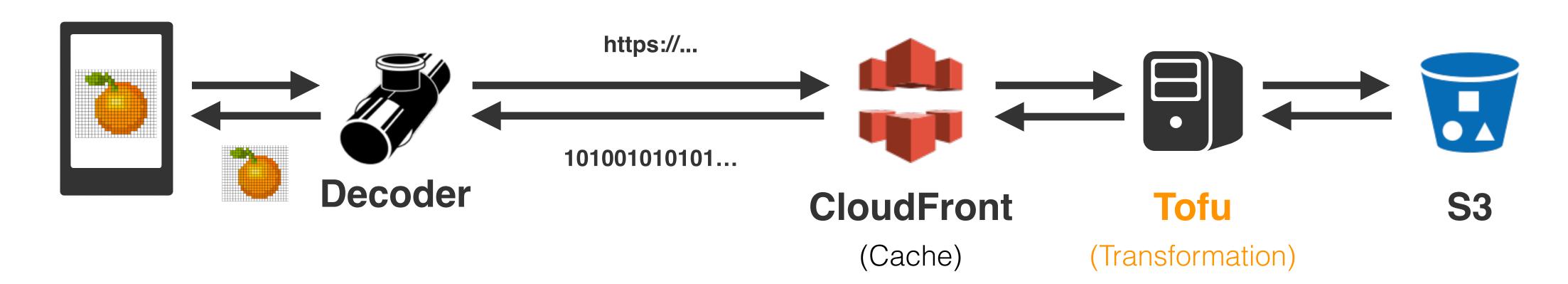


# We are using image transformation server called Tofu. Tofu transforms images on the fly.

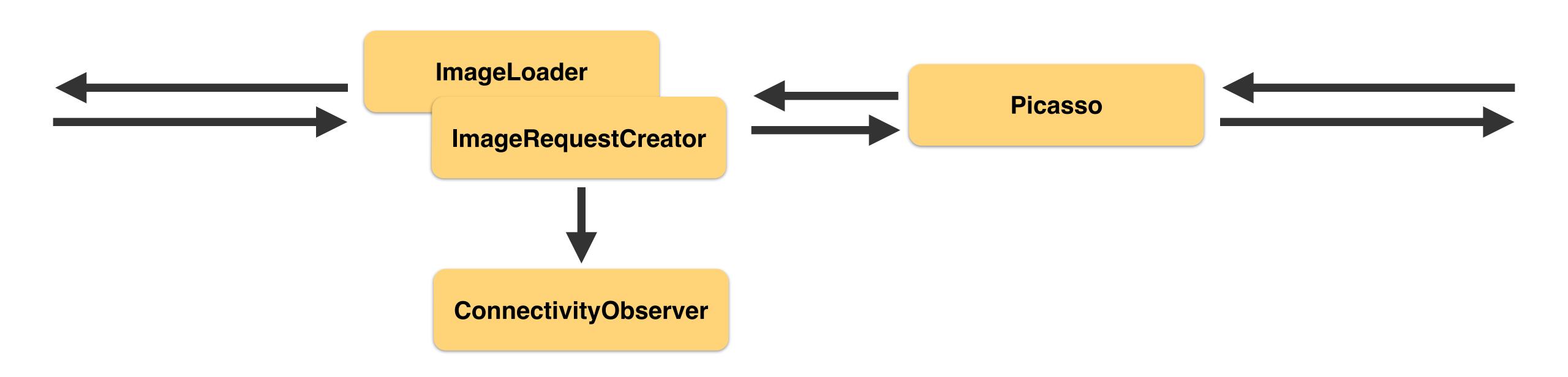
### Tofu has these functions

- Fixed Width: (\d+)
- Fixed Height: x(\d+)
- Fixed Width and Height: (\d+)x(\d+)
- Smaller than: (\d+)?(x\d+)?s
- Cropping: (\d+)x(\d+)c
- Manual Cropping: (\d+)x(\d+)c(\d+)\_(\d+)\_(\d+)\_(\d+)\_(\d+)
- Quality Factor: [geometry]q(\d+)

•



# Request different image size depends on network quality



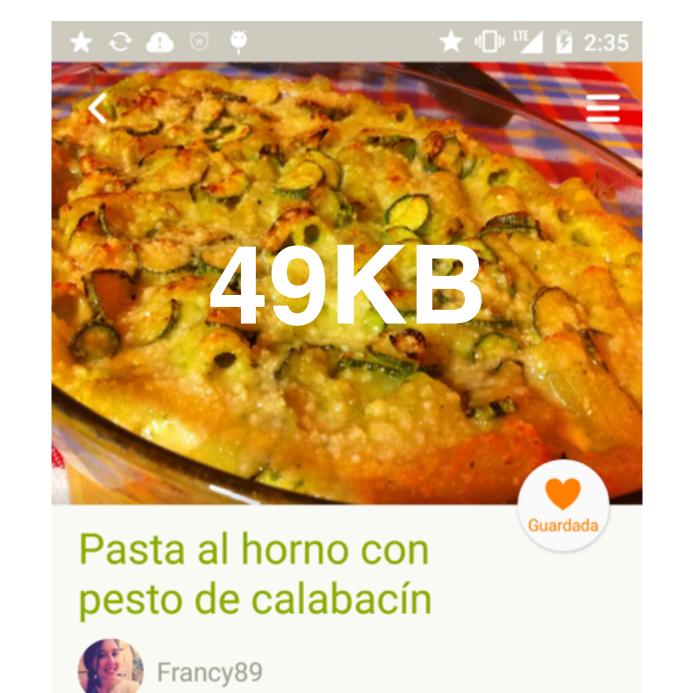
# LOW images are 40% smaller than full images

EXCELLENT: (1080 \* 756) \* 1.0

Pasta al horno con pesto de calabacín

Francy89

LOW: (756 \* 530) \* 0.7



# API Design

# If API responses become faster, users become happier.



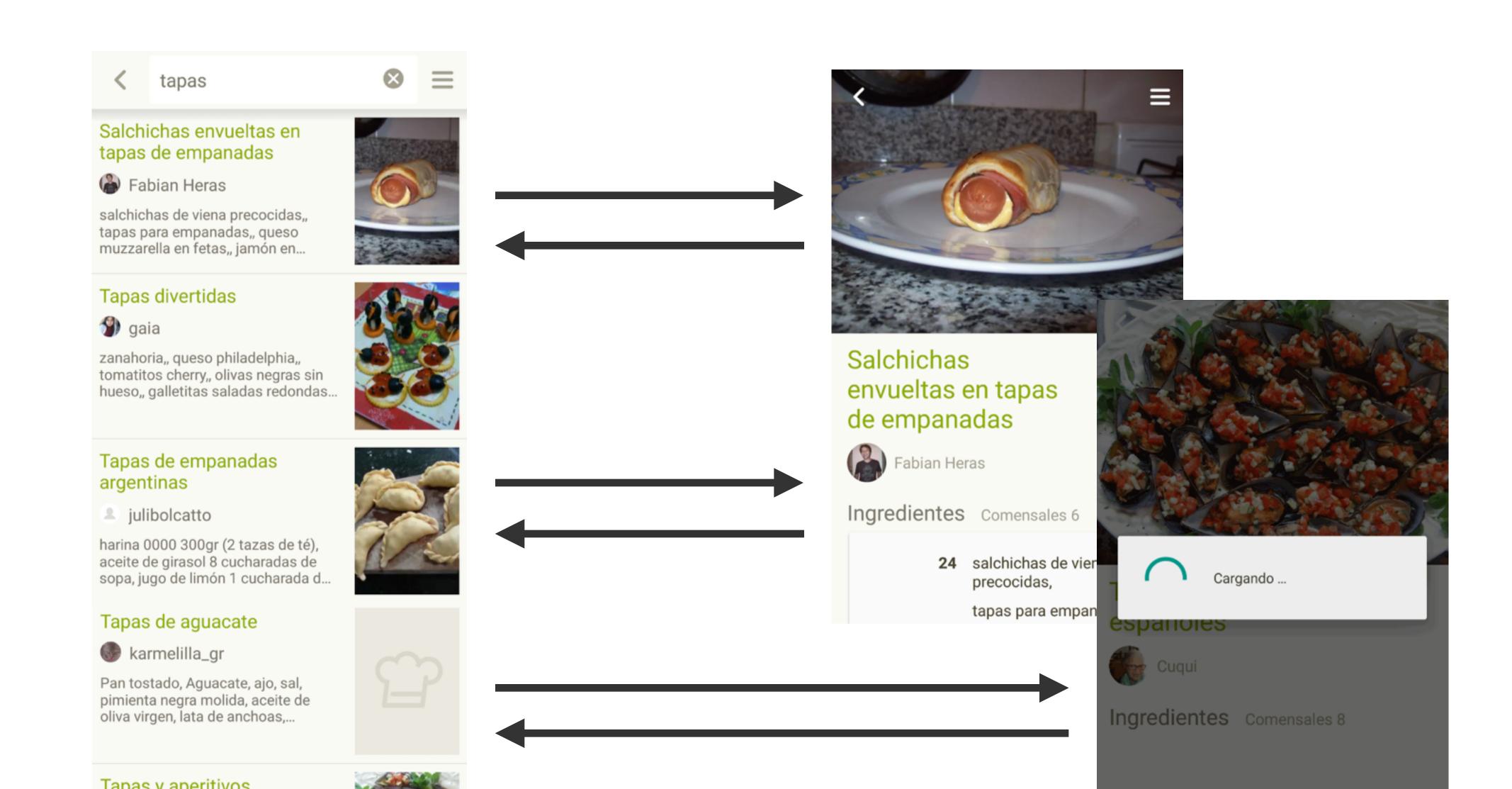
### Of course, the answer is

# 11 Yes

# Let's use partial response to reduce data size

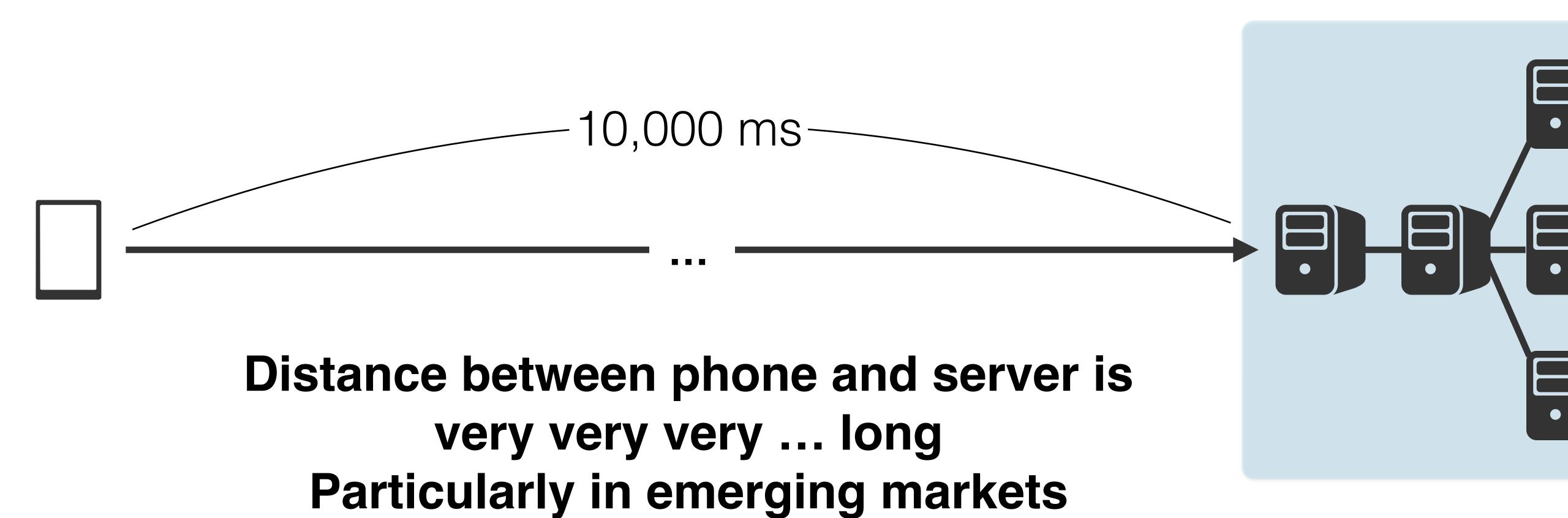
## But be careful, Android has state and screen transition

### Users go back and forth to decide a recipe

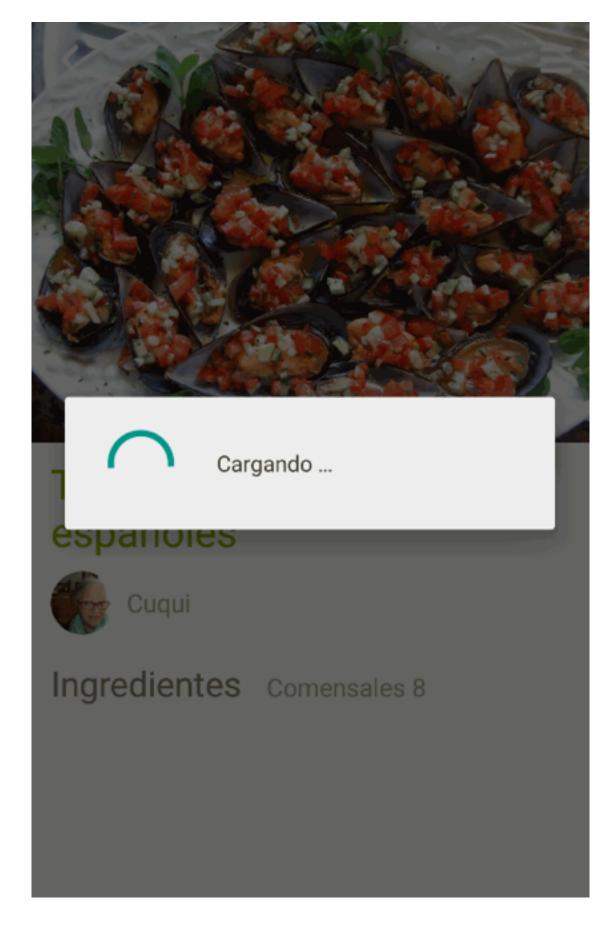


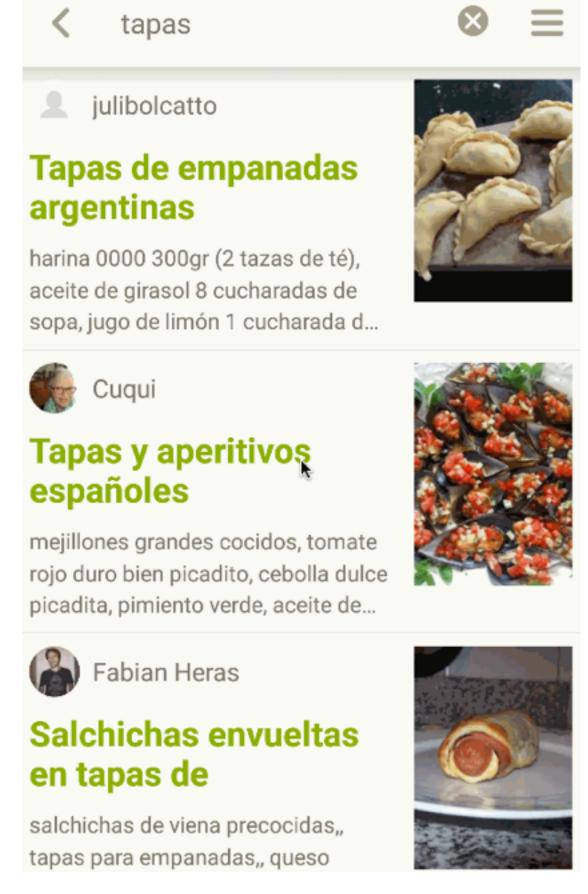
# Thing we have to do is Optimizing UX > response time

200 ms or below



### Reduce unnecessary fields Get necessary relations







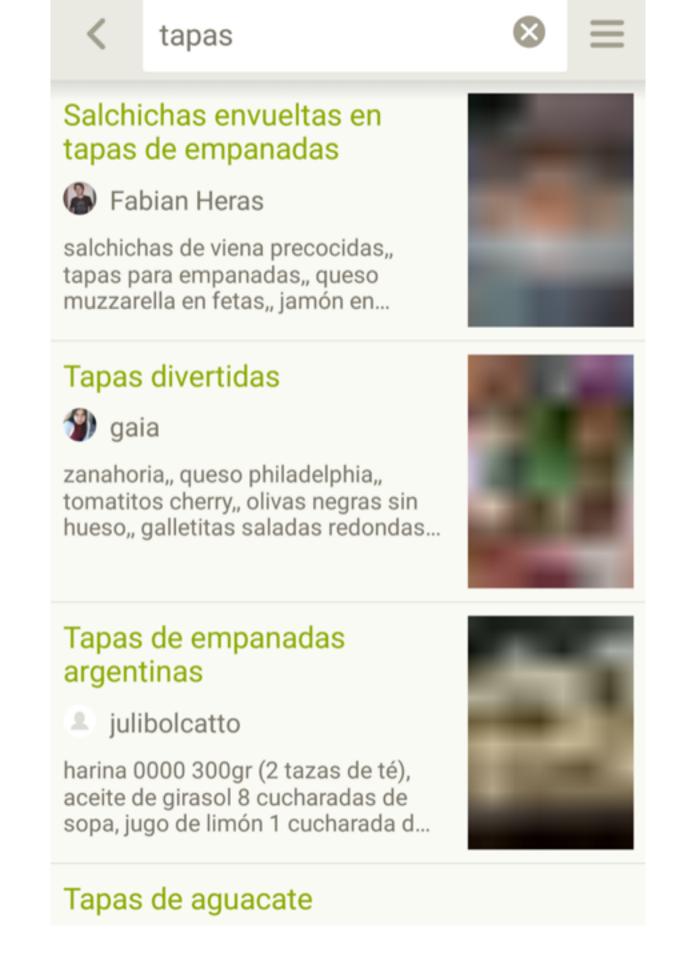


# One more thing to improve experience

# Response include thumbnail\_data\_uri Base64 encoded image

### Data size is small but there is a big improvement





## Documentation

# Keeping the documentation updated in real time is hard





### Hi, can I ask you a question about API?

. . .

Today



Sorry for late reply

### We are using

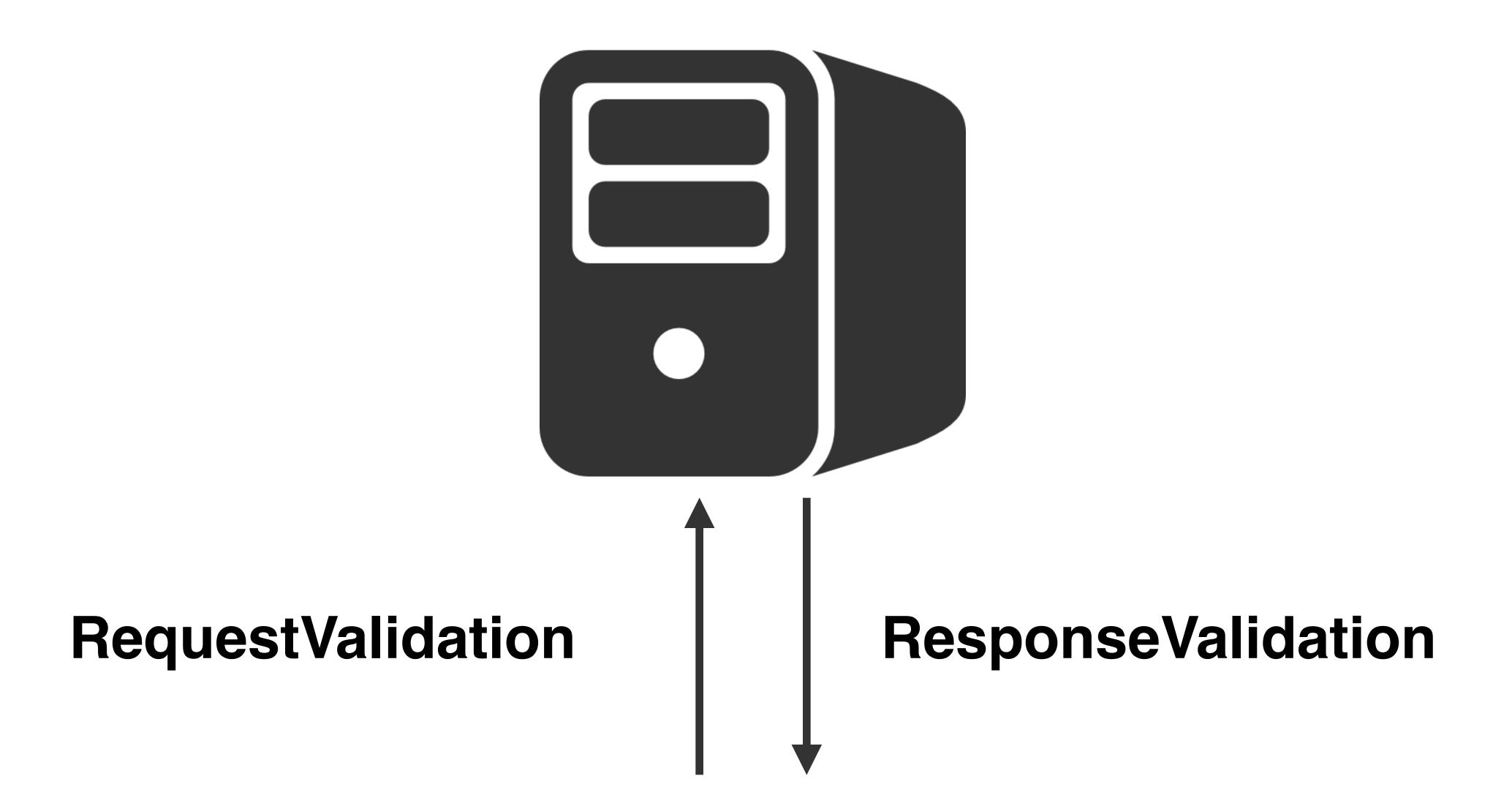
## JSON Schema

as the format for describing our APIs

## JSON Schema provides

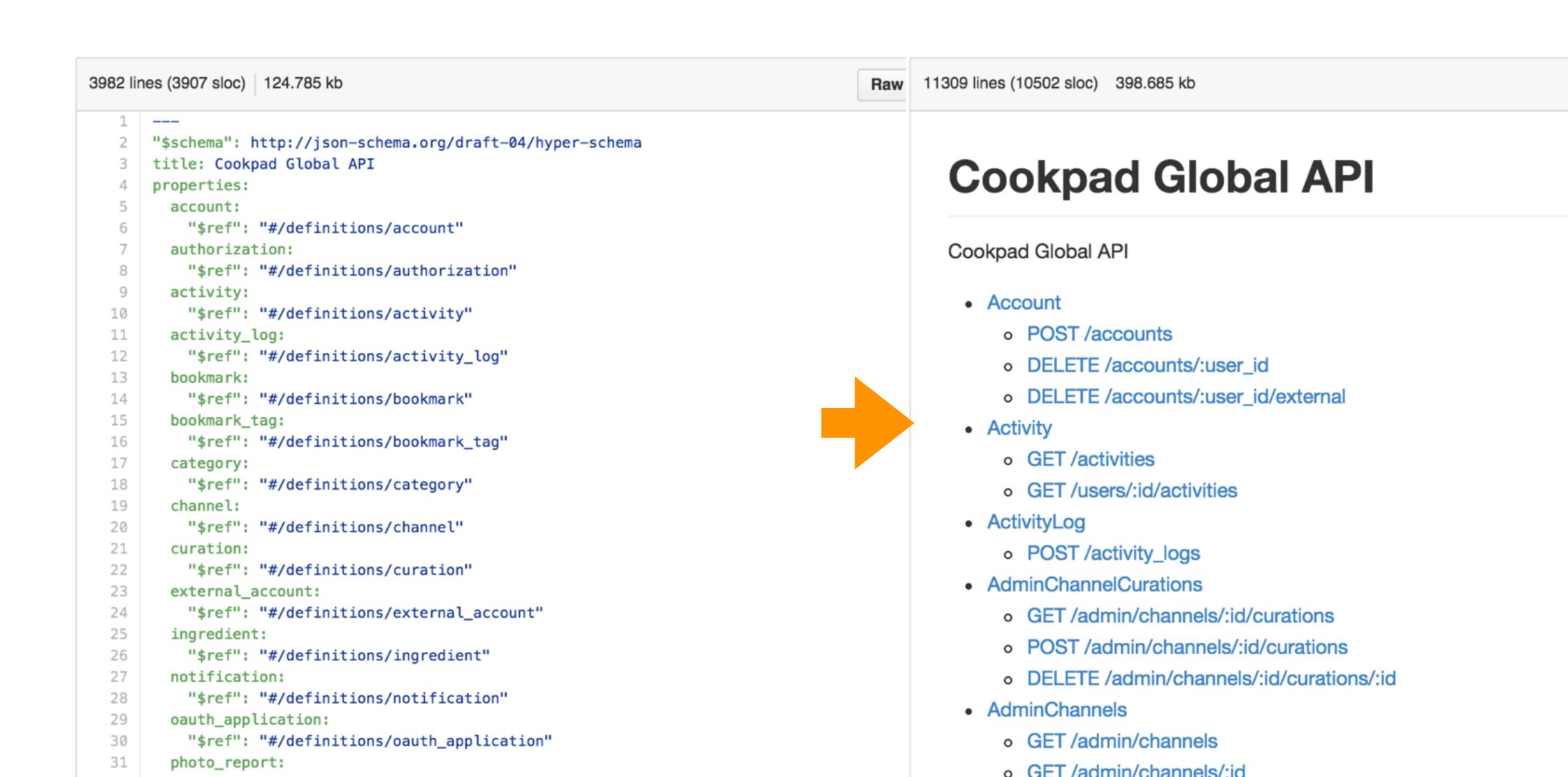
- · Request Validation
- · Response Validation
- · Document generation

### Check request/response automatically



# Generate API documentation from schema file

Raw





We don't need to update documentation manually. And we can see latest documentation any time.

## Conclusion

### Generate documentation **Auto validation**

## **Cache Controls**



App server

## **GZIP**

### Salchichas envueltas en tapas de empanadas

tapas



salchichas de viena precocidas,, tapas para empanadas,, queso muzzarella en fetas,, jamón en...



### Tapas divertidas



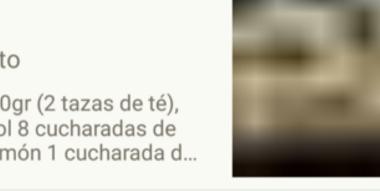
zanahoria,, queso philadelphia,, tomatitos cherry,, olivas negras sin hueso,, galletitas saladas redondas...



### Tapas de empanadas argentinas

julibolcatto

harina 0000 300gr (2 tazas de té), aceite de girasol 8 cucharadas de sopa, jugo de limón 1 cucharada d...



### Tapas de aguacate



Base64 encoded thumbnail Partial response Appropriate data model



Stetho

**WebP Prioritized request** Appropriate image size



Image server

## Thank you!

