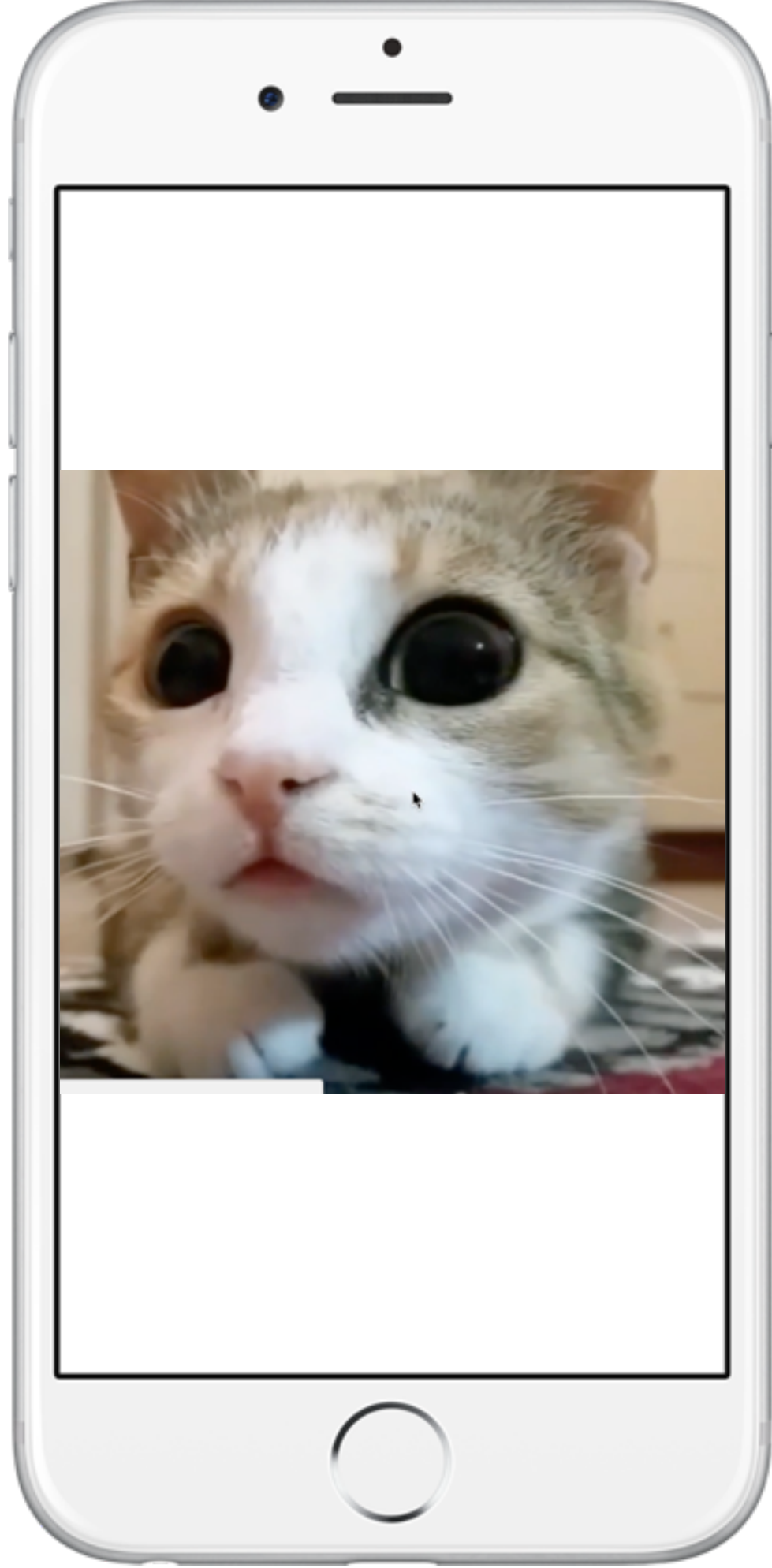




# SCALING INSTAGRAM INFRA

Lisa Guo — Nov 7th, 2016  
[lguo@instagram.com](mailto:lguo@instagram.com)



# INSTAGRAM HISTORY

2010



2012/4/3

Android  
release

2014/1



2011

14M users

2012/4/9

Facebook  
acquisition

# INSTAGRAM EVERYDAY

300 Million Users

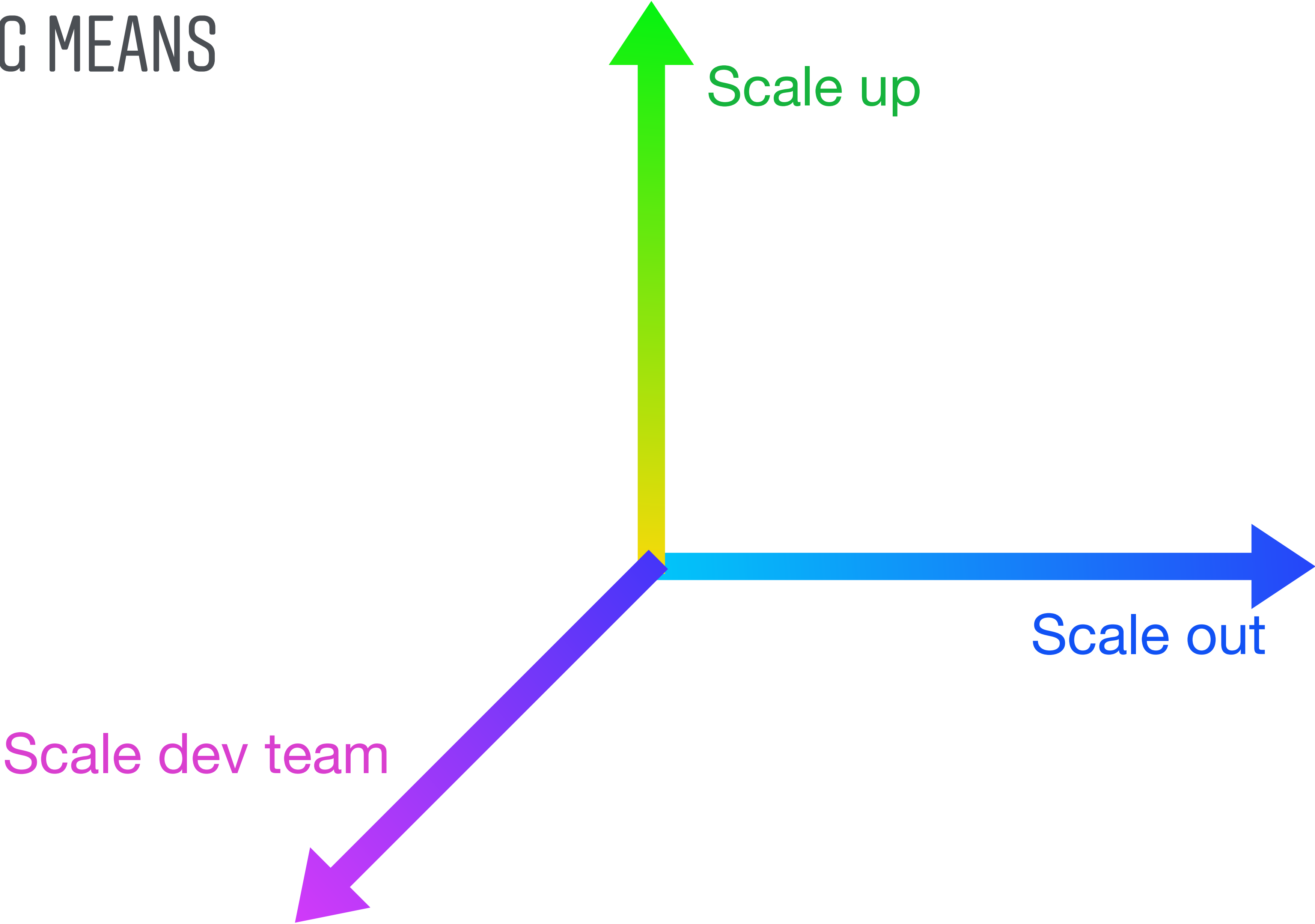
4.2 Billion likes

95 Million photo/video uploads

100 Million followers



# SCALING MEANS





SCALE OUT

# SCALE OUT

“To scale horizontally means to add more nodes to a system, such as adding a new computer to a distributed software application. An example might involve scaling out from **one** Web server system to **three.**”

- Wikipedia



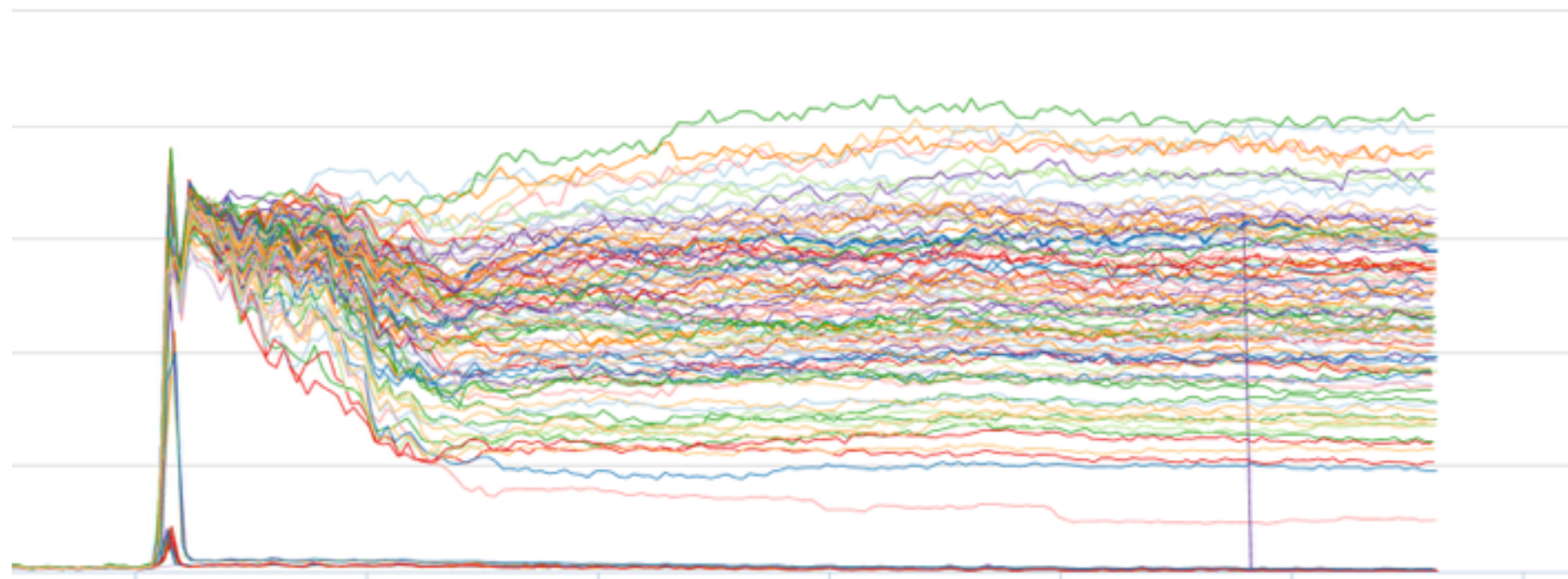
**MICROSERVICE**

# SCALING OUT



—> vertical partition  
horizontal sharding

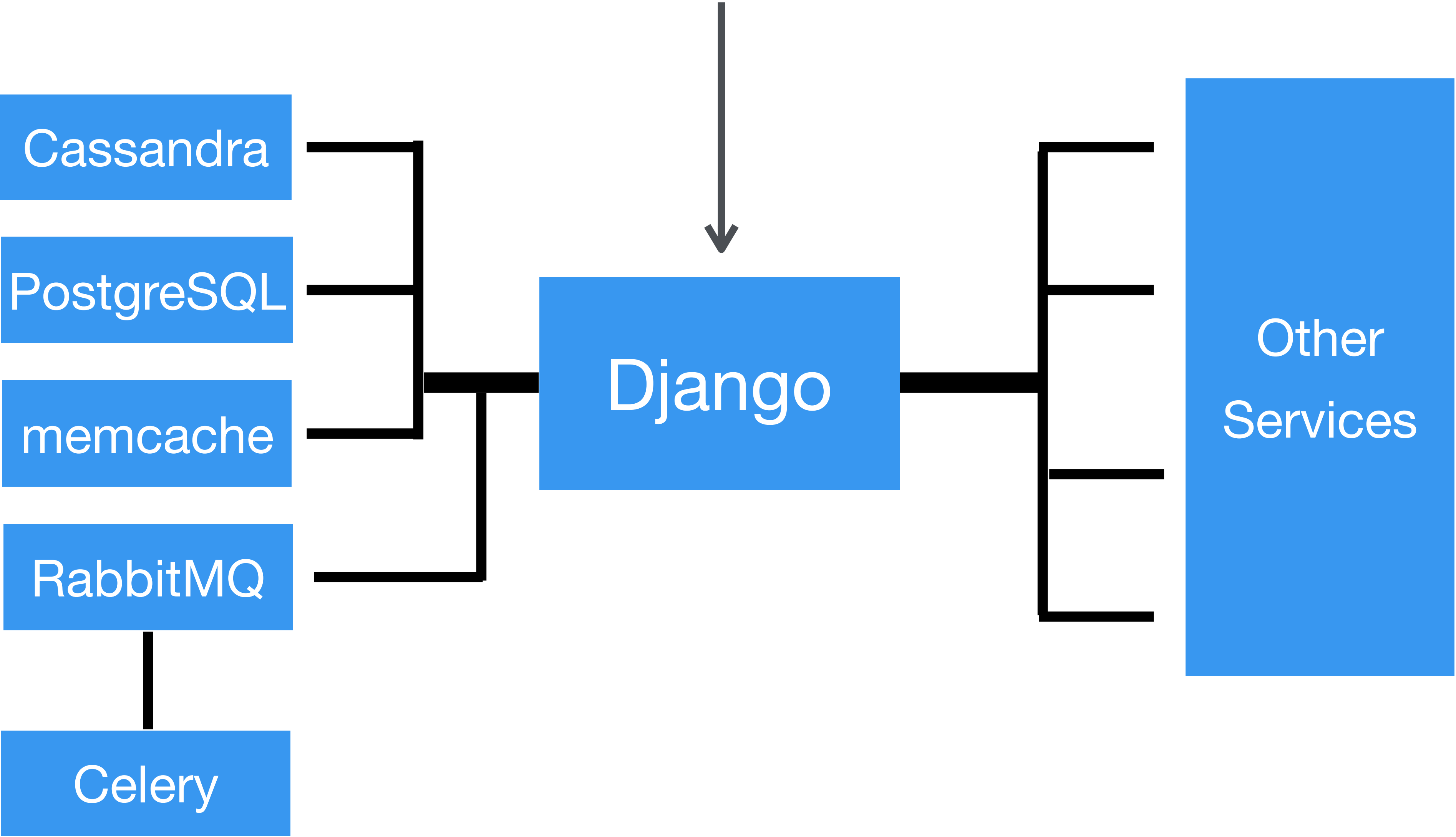
# SCALING OUT







# INSTAGRAM STACK



# STORAGE VS. COMPUTING

- Storage: needs to be consistent across data centers
- Computing: driven by user traffic, as needed basis

# SCALE OUT: STORAGE



user feeds, stories, activities, and other logs

- Masterless
- Async, low latency
- Multiple data center ready
- Tunable latency vs consistency trade-off



# SCALE OUT: STORAGE



user, media, friendship etc

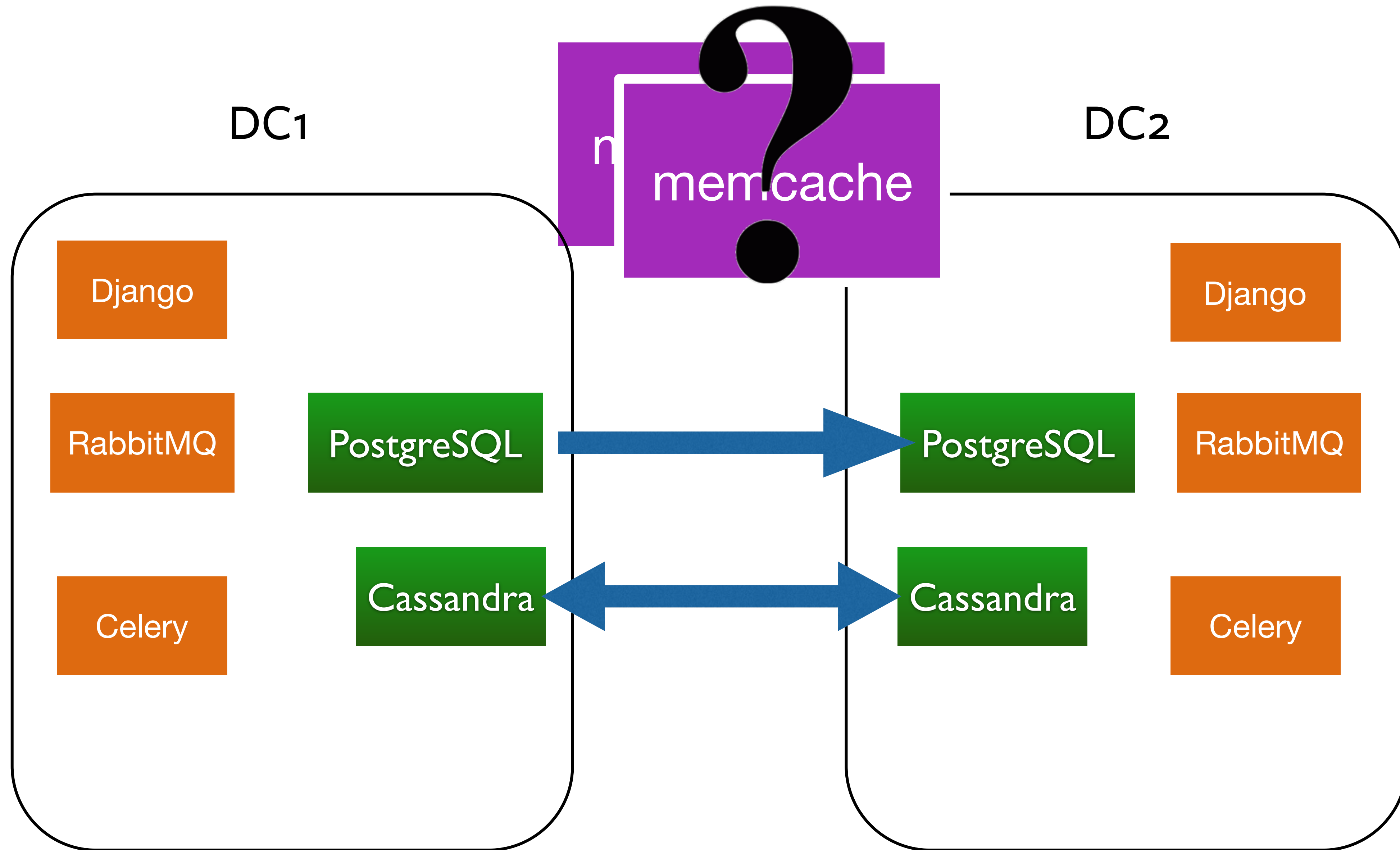
- One master, replicas are in each region
- Reads are done locally
- Writes are cross region to the master.

# COMPUTING



@asynchronous\_task





# MEMCACHE

- Millions of reads/writes per second
- Sensitive to network condition
- Cross region operation is prohibitive

DC1

User C



comment



Django

insert



PostgreSQL

set



memcache

User R



feed



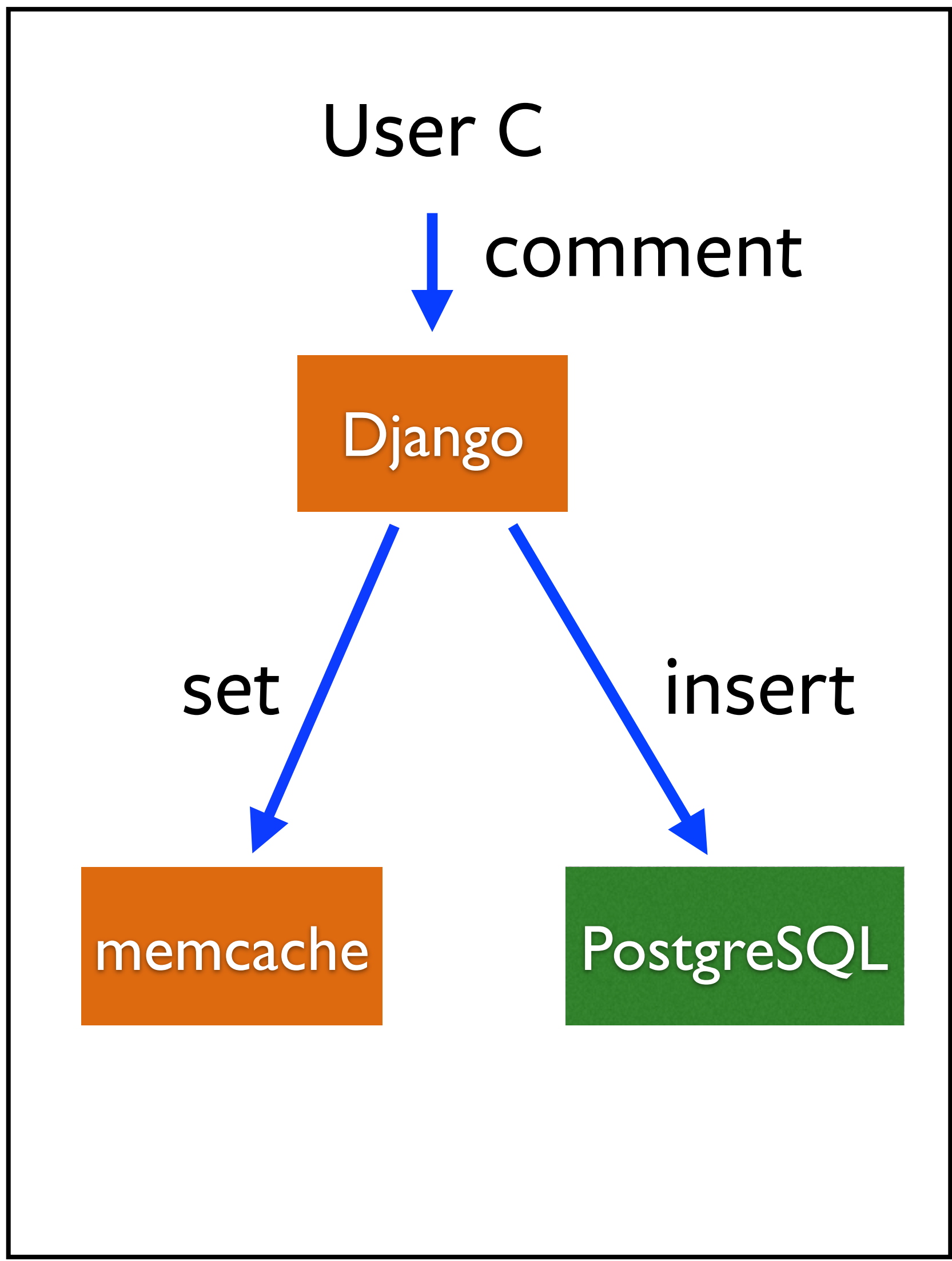
Django

get

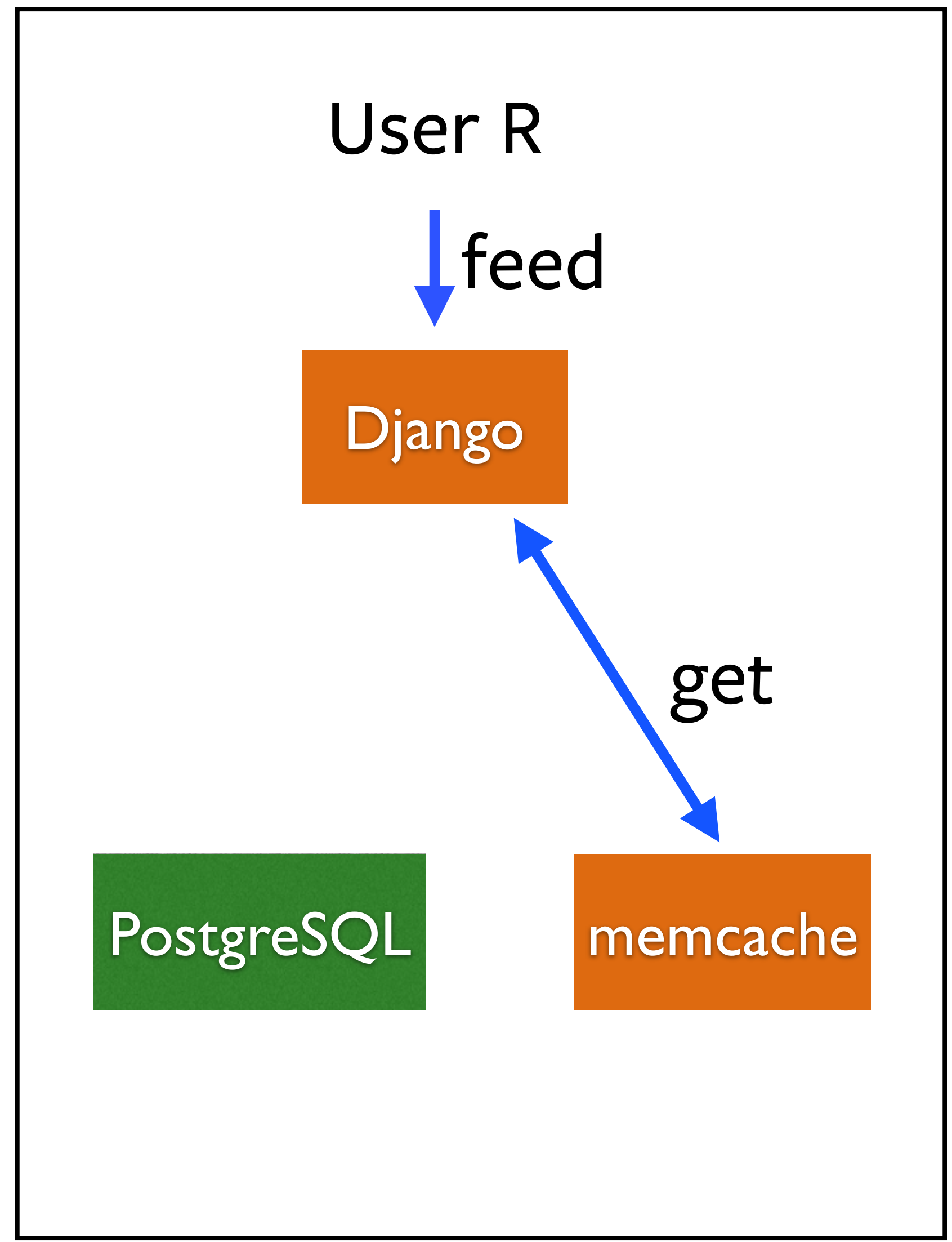


DC1

DC2

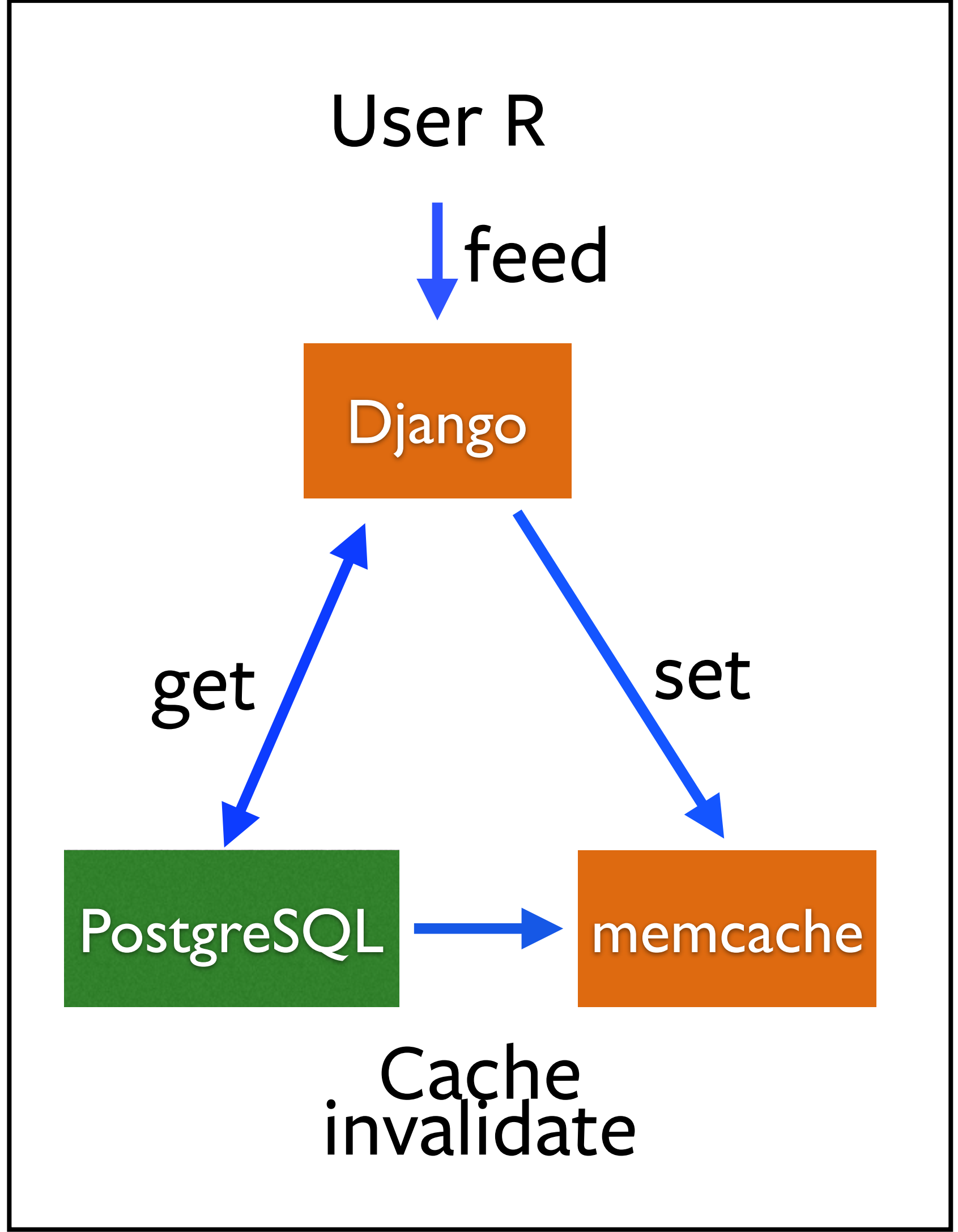
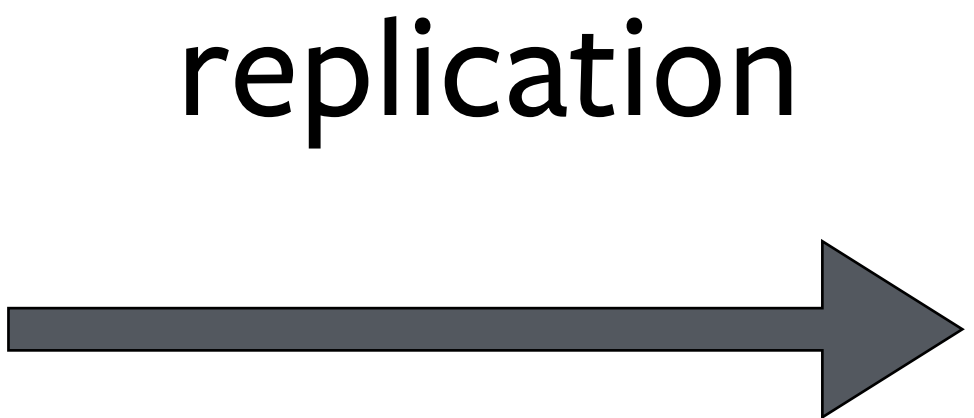
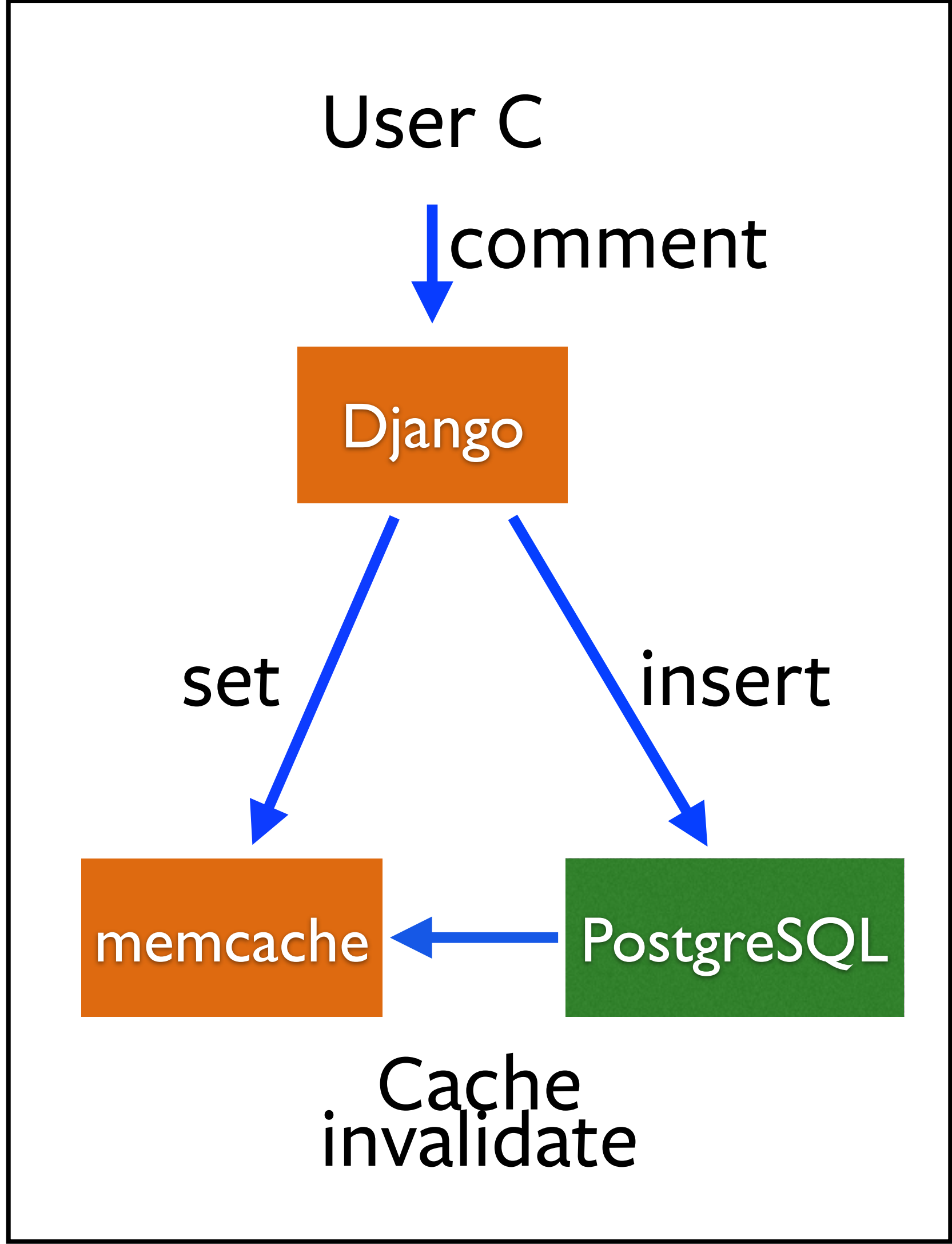


replication



DC1

DC2



# COUNTERS

```
select count(*) from  
user_likes_media  
where  
media_id=12345;
```

100s ms



instagram  
Baghdad, Iraq

1.2m likes

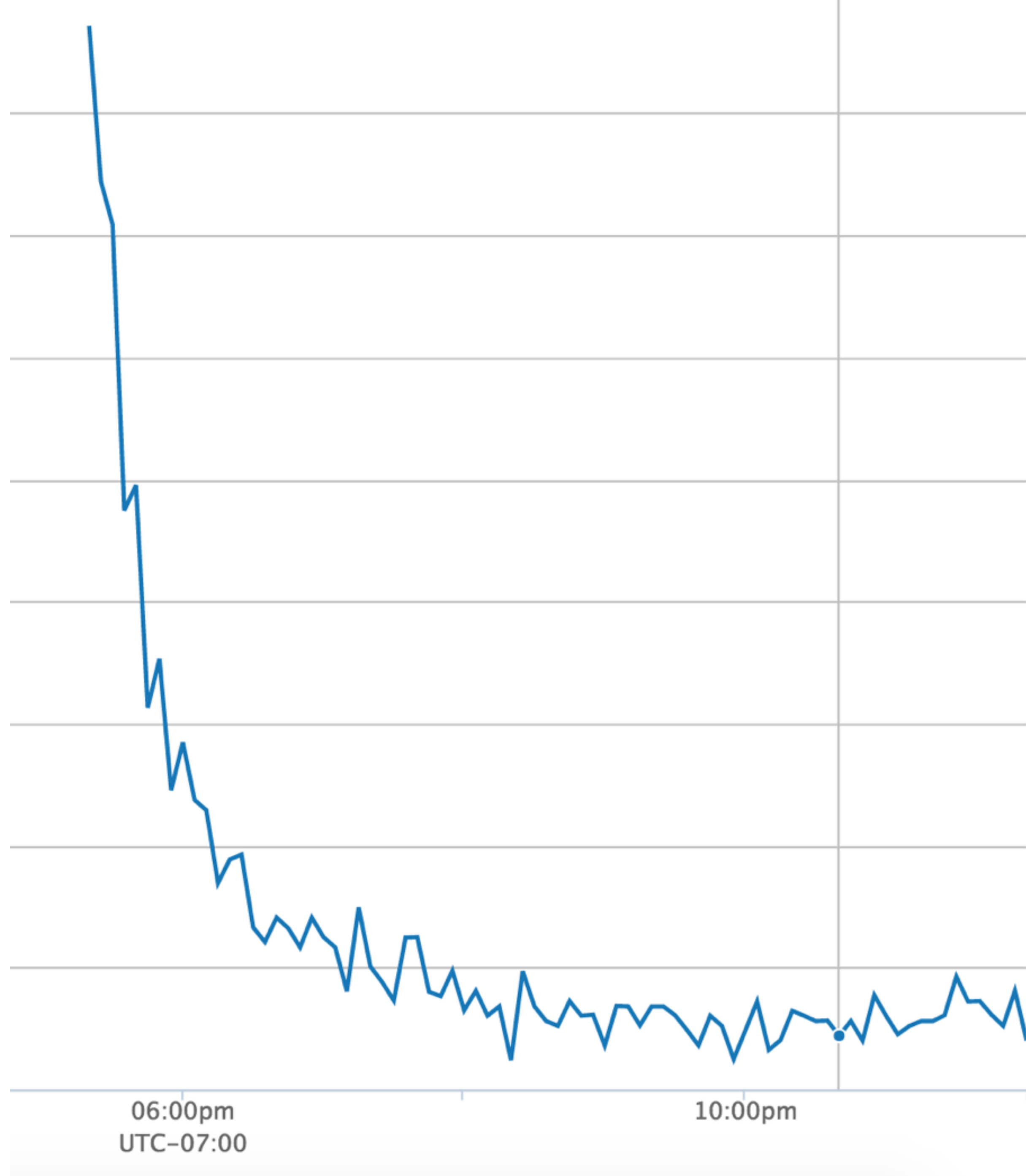
instagram Documentary p  
video reporter Ahmad Mo  
(@ahmadmoussa) has his #  
“Many people around the  
know much about Iraq he  
they think of it as a war zc  
what they see and read in  
25-year-old says. Ahmad  
to a more human side of h  
through @everydayiraq, th  
photography project he st  
to share the everyday life  
everyone, document it an  
history,” Ahmad says. “Eve  
world, people want to live  
want to play and go to sch  
want to gather happily at  
want to help develop their  
Photo by @ahmadmoussa

view all 7,810 comments

miguelgroove #Love

♥ Add a comment...





# COUNTER

```
select count from  
media_likes where  
media_id=12345;
```

10s us




 **instagram**  
Baghdad, Iraq

1.2m likes

instagram Documentary p  
video reporter Ahmad Mo  
(@ahmadmoua) has his #  
“Many people around the  
know much about Iraq he  
they think of it as a war zc  
what they see and read in  
25-year-old says. Ahmad  
to a more human side of h  
through @everydayiraq, th  
photography project he st  
to share the everyday life  
everyone, document it an  
history,” Ahmad says. “Eve  
world, people want to live  
want to play and go to sch  
want to gather happily at  
want to help develop their  
Photo by @ahmadmoua

view all 7,810 comments

miguelgroove #Love

 Add a comment...

ALPH ZUKOR AND JESSE L. LASKY PRESENT

# ZANE GREY'S THE THUNDERING HERD



WITH JACK HOLT, LOIS WILSON  
NOAH BEERY, RAYMOND HATTON  
SCREEN PLAY BY LUCIEN HUBBARD • DIRECTED BY WILLIAM HOWARD



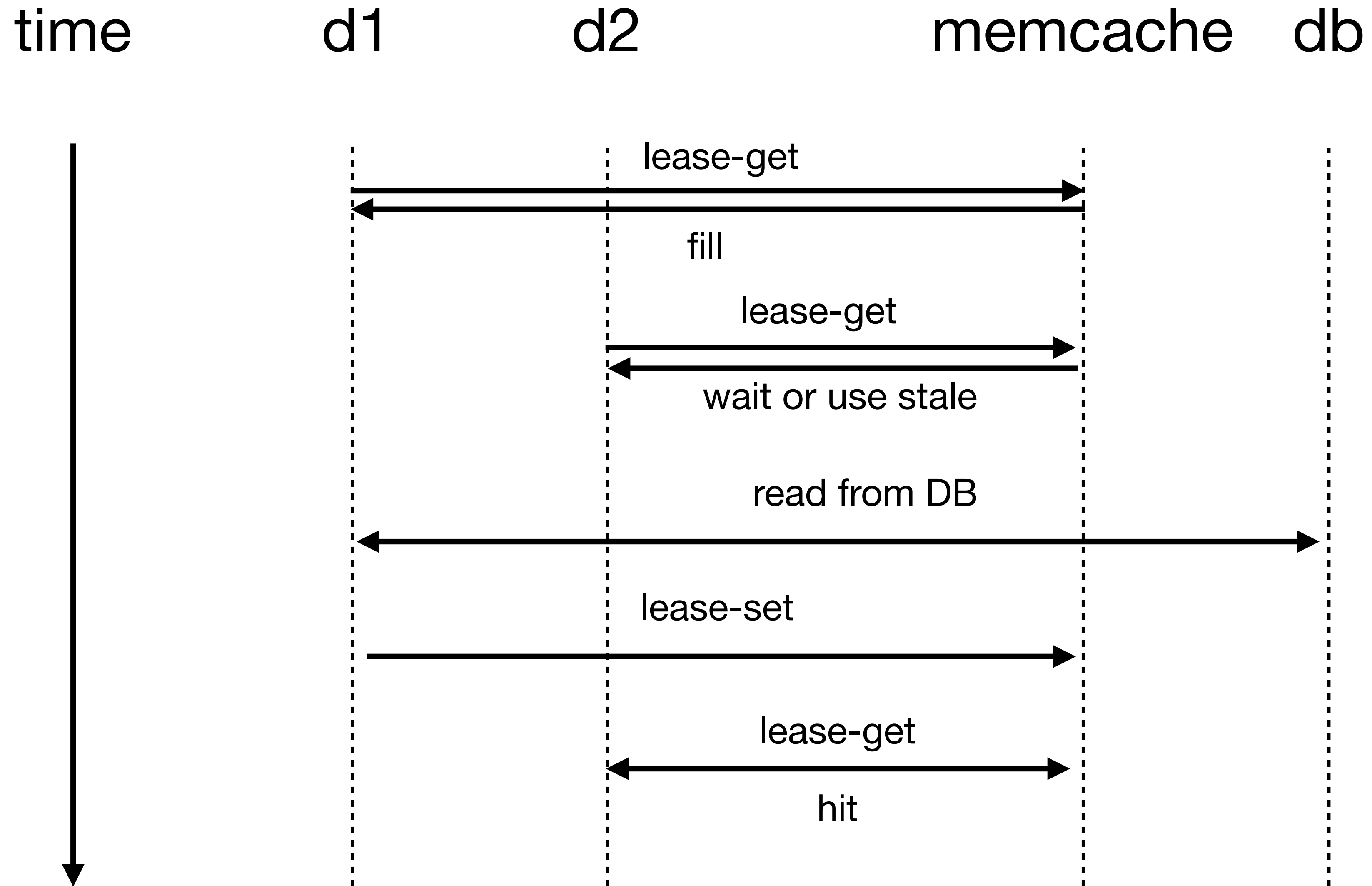
*A Paramount Picture*  
PRODUCED BY FAMOUS PLAYERS - LASKY CORP.

THIS LOBBY DISPLAY LEASED FROM FAMOUS PLAYERS - LASKY CORPORATION

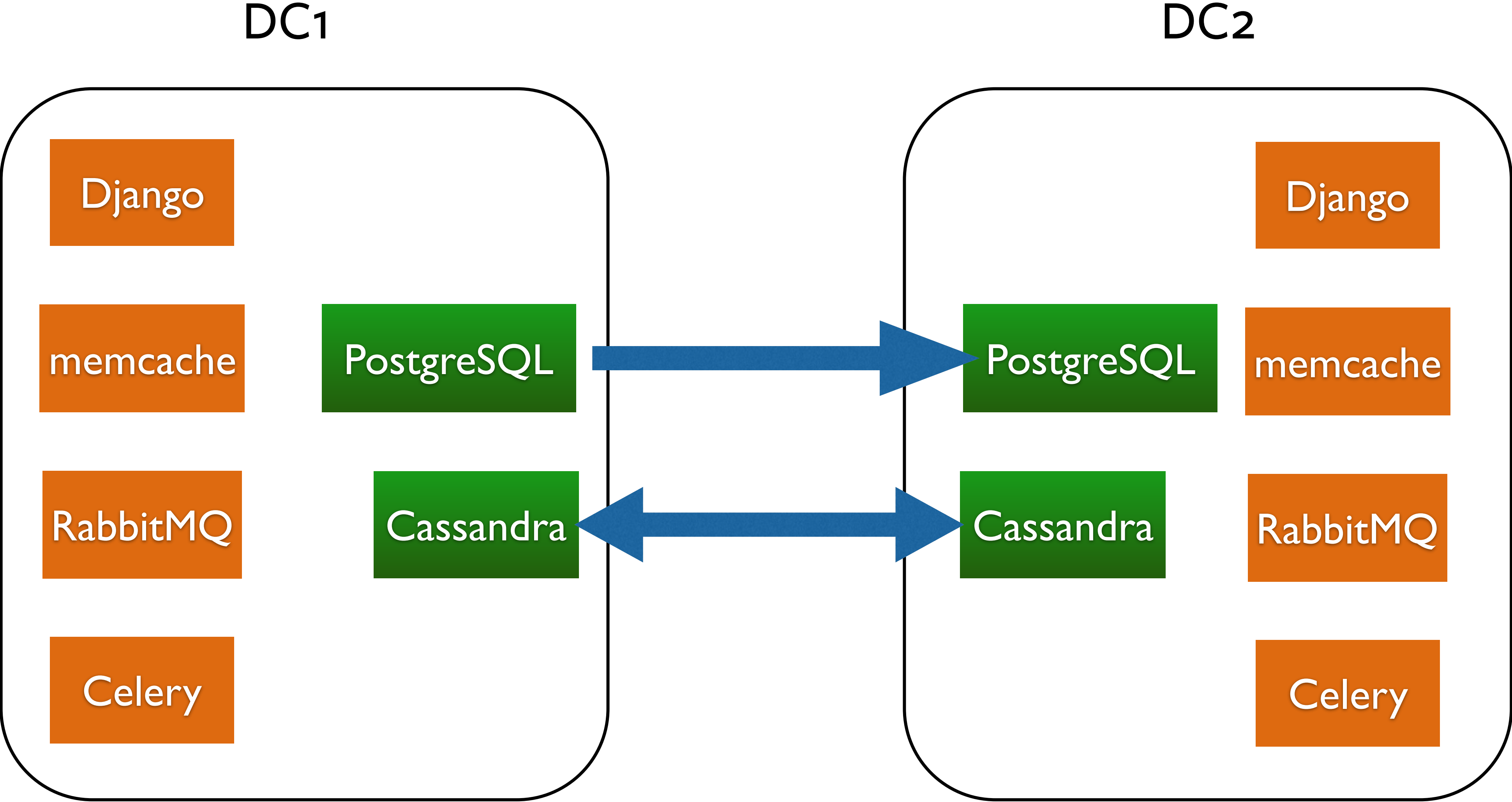
Cache invalidated

All djangos try to access DB

# MEMCACHE LEASE

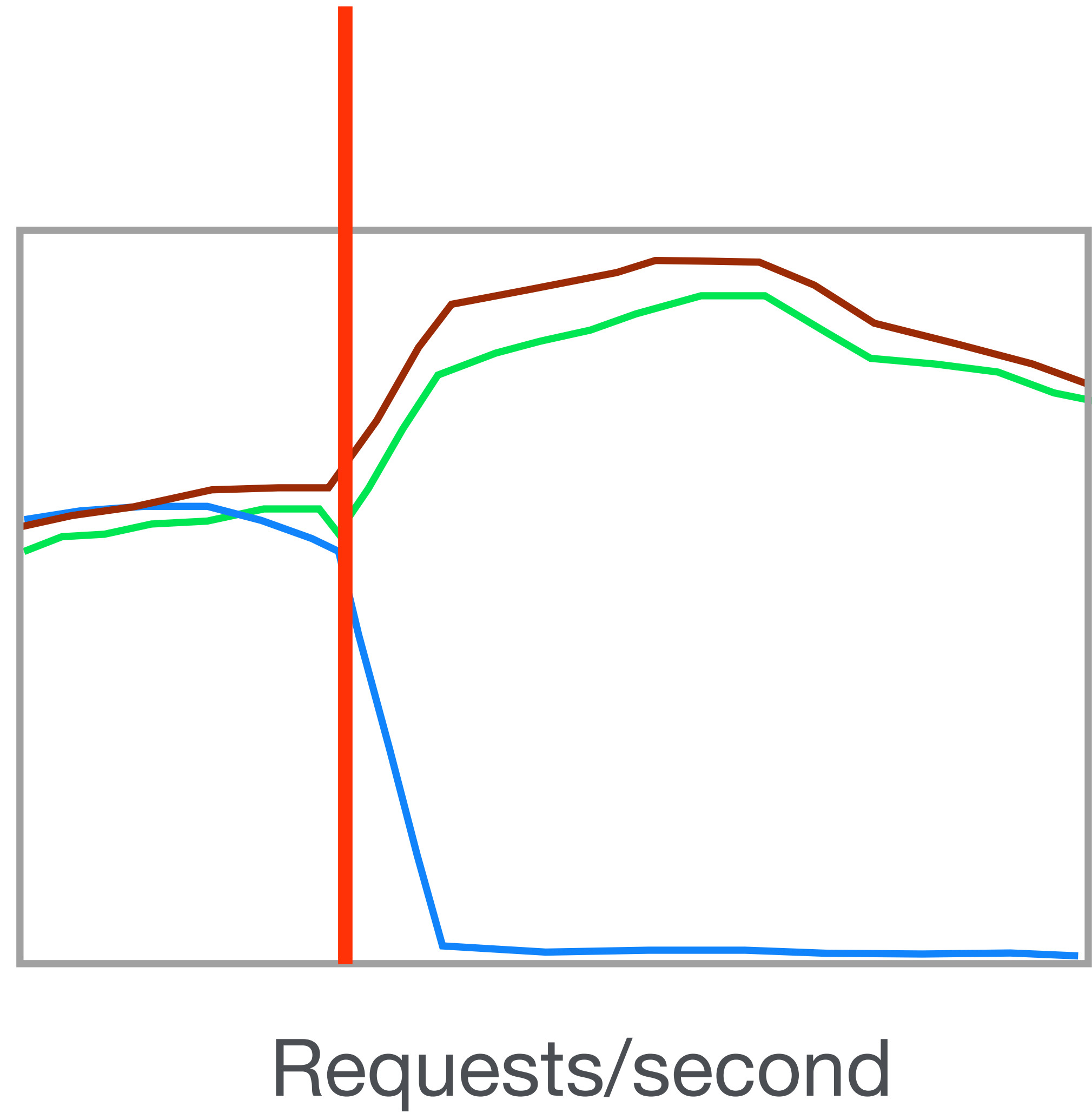


# INSTAGRAM STACK - MULTI REGION

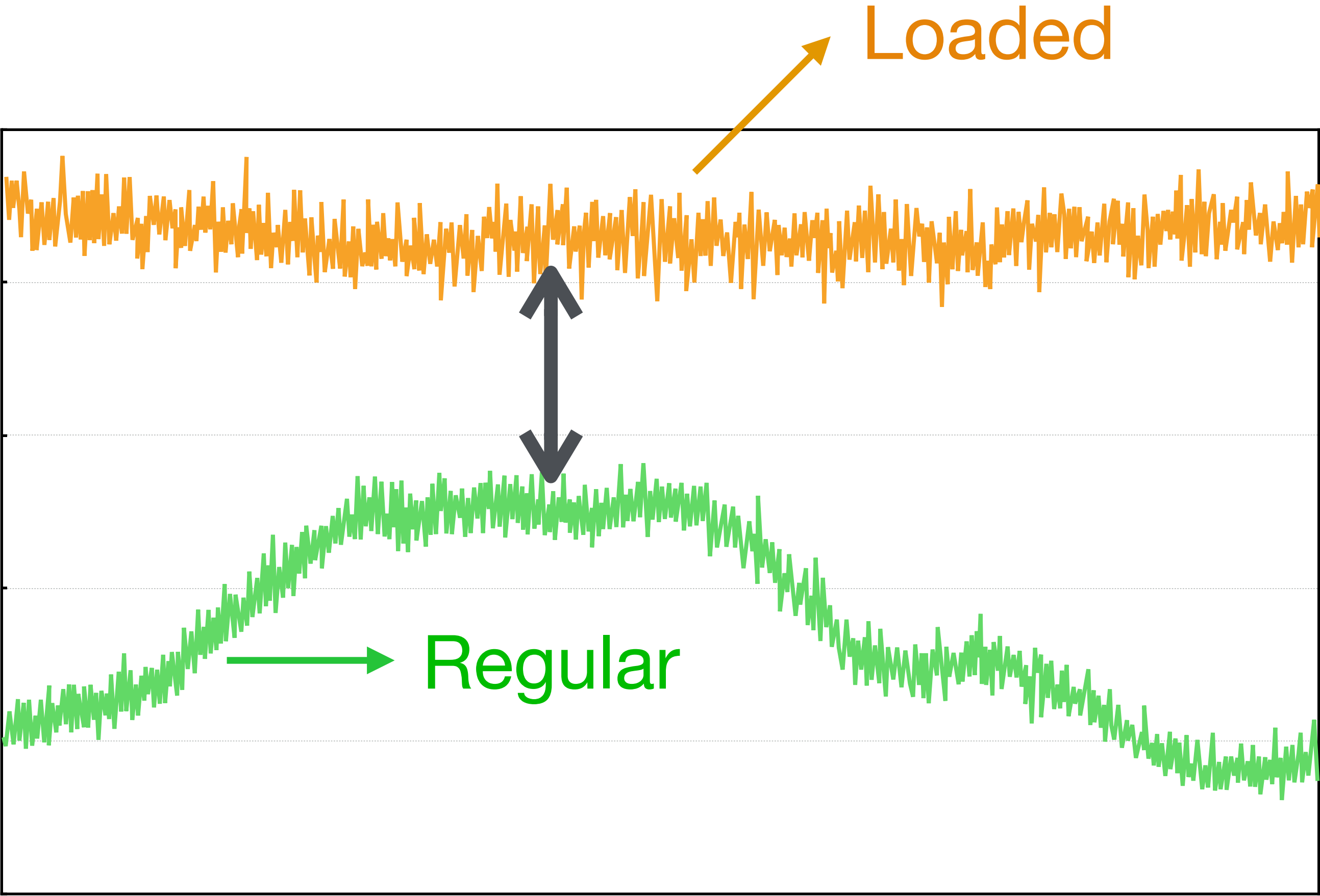
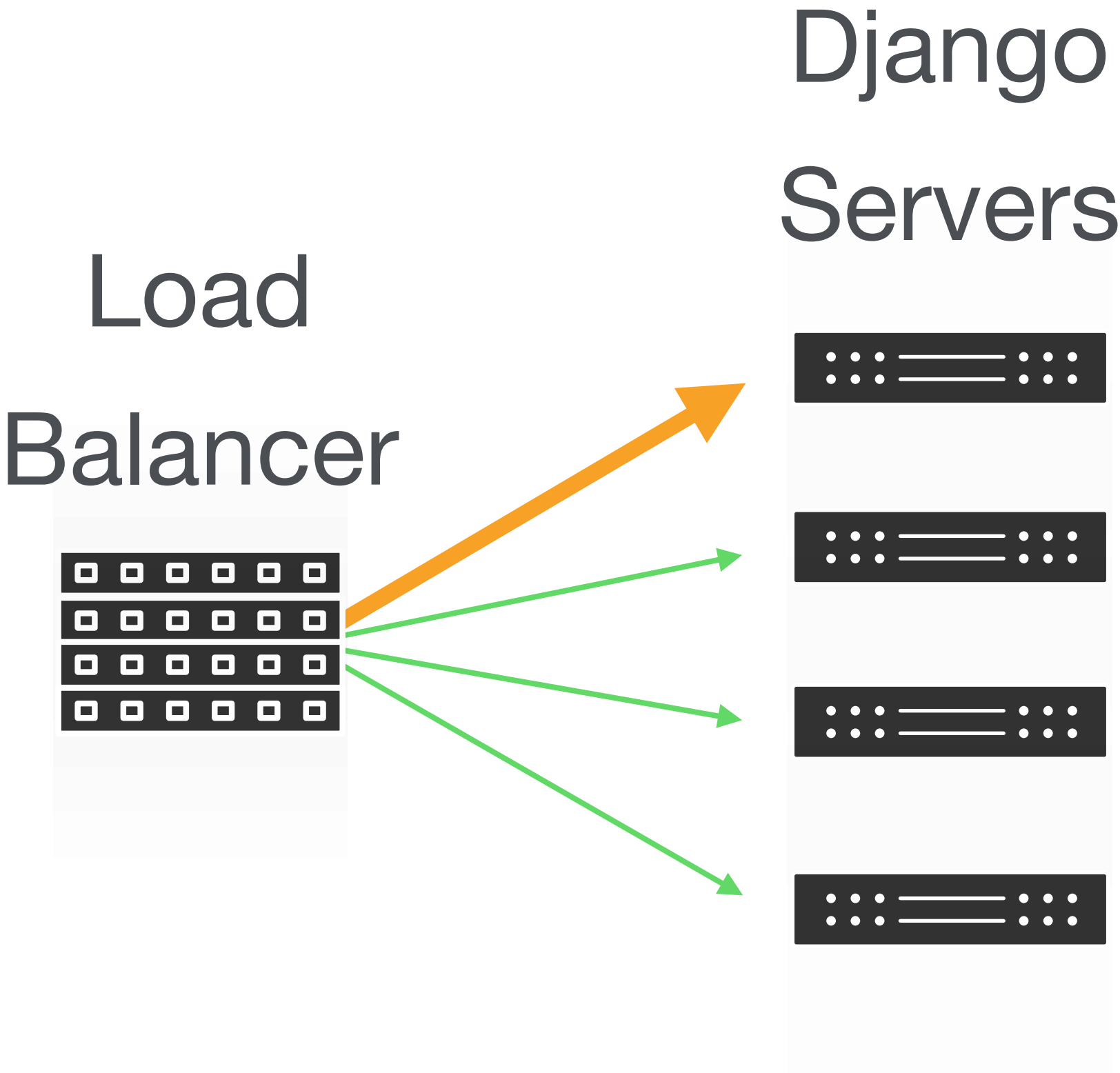


# SCALING OUT

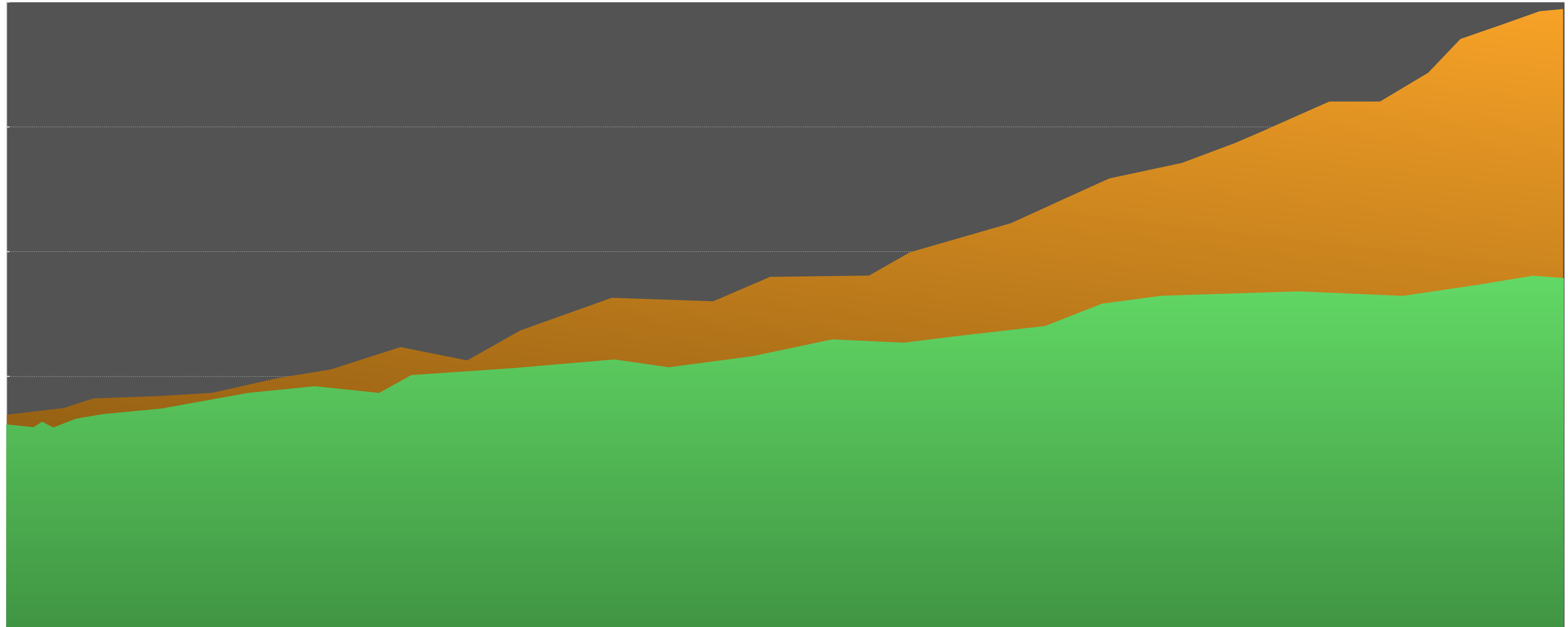
- Capacity
- Reliability
- Regional failure ready



# LOAD TEST



CPU instructions



 User growth

 Server growth



"Don't count the servers,  
make the servers count"



SCALE UP

# SCALE UP

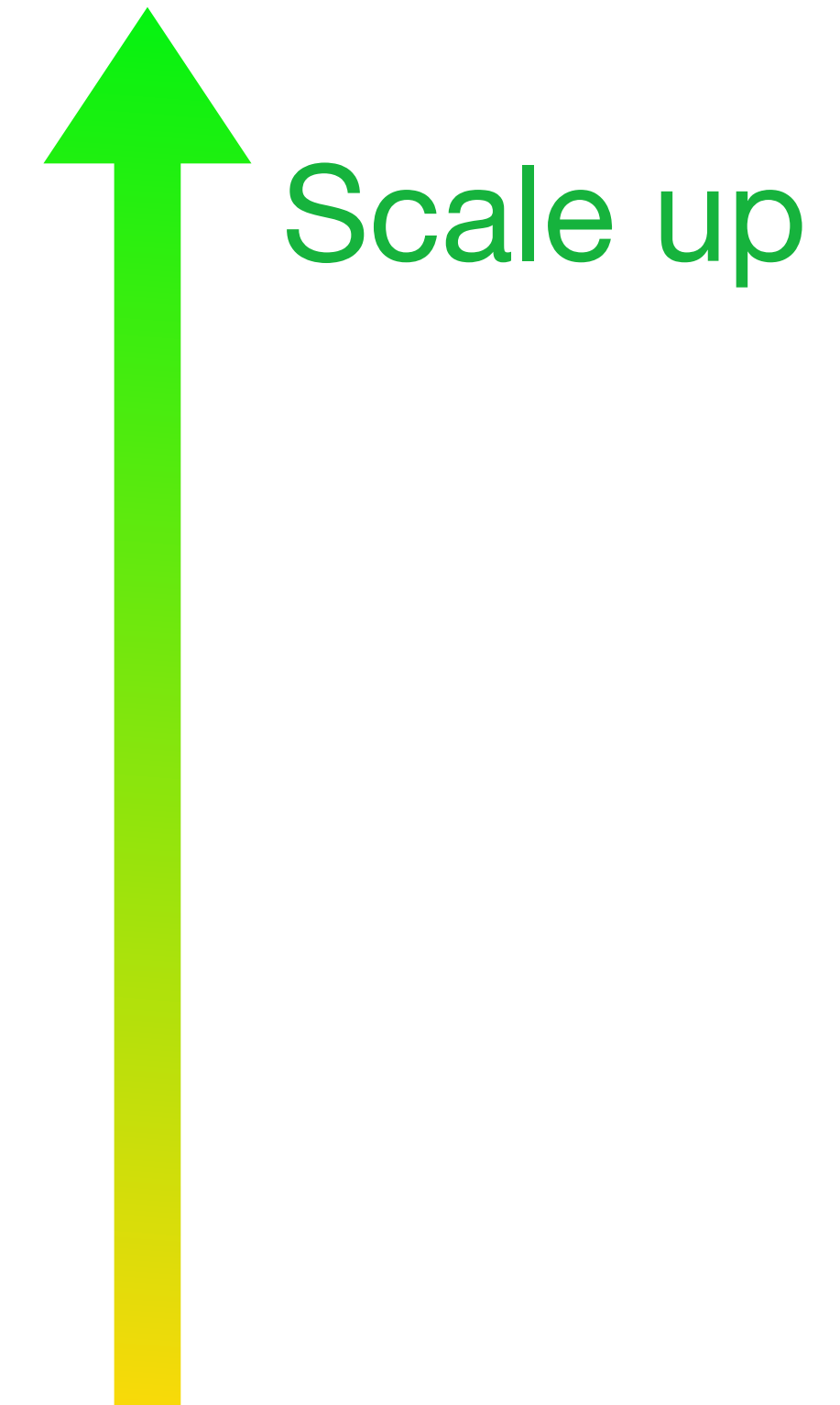
Use as few CPU instructions as possible

Use as few servers as possible

# SCALE UP

Use as few CPU instructions as possible

Use as few servers as possible



CPU

Monitor

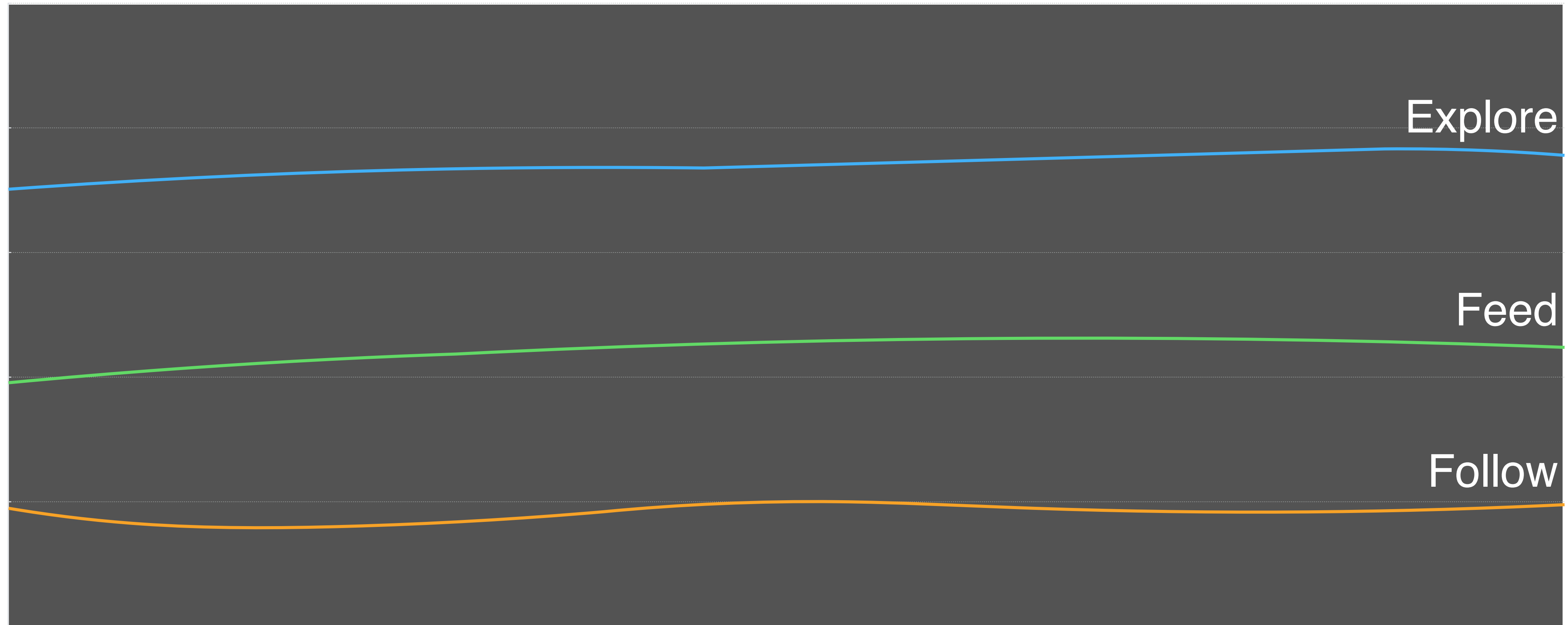
Analyze

Optimize

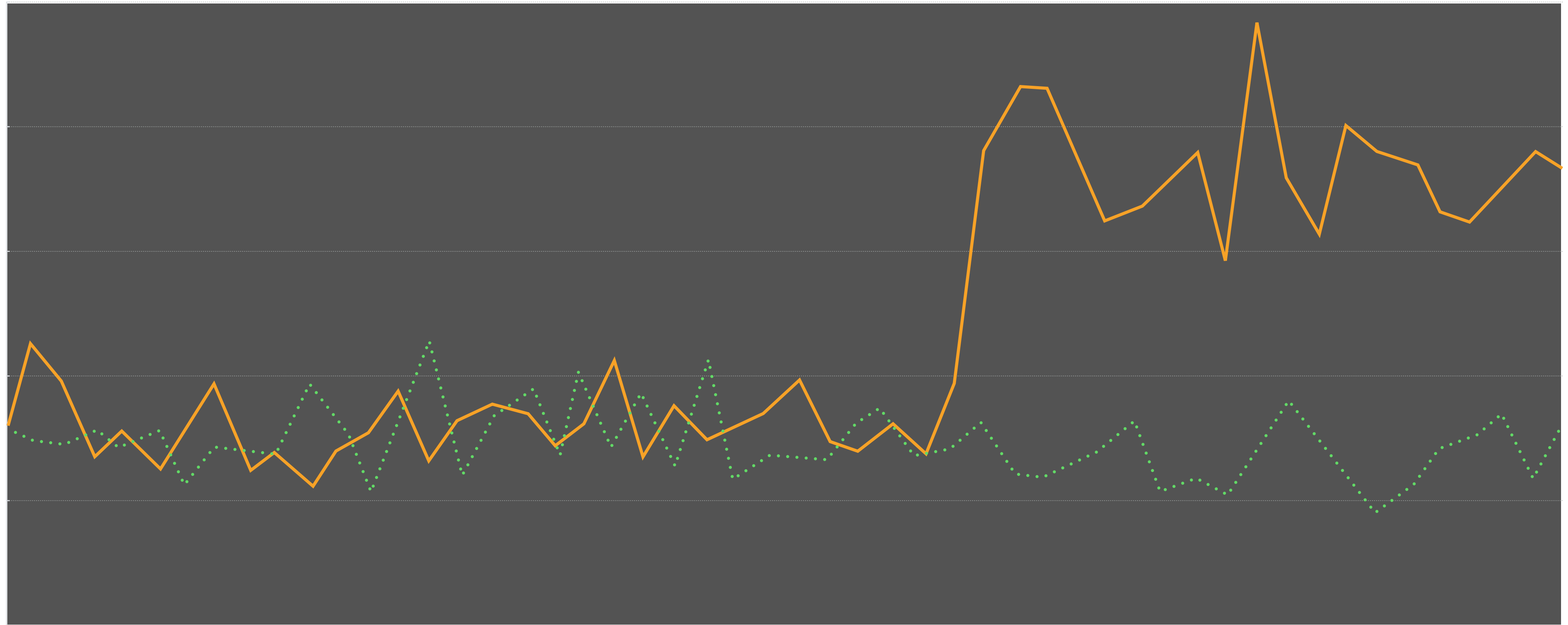
# COLLECT

```
struct perf_event_attr pe;
pe.type = PERF_TYPE_HARDWARE;
pe.config = PERF_COUNT_HW_INSTRUCTIONS;
fd = perf_event_open(&pe, 0, -1, -1, 0);
ioctl(fd, PERF_EVENT_IOC_ENABLE, 0);
<code you want to measure>
ioctl(fd, PERF_EVENT_IOC_DISABLE, 0);
read(fd, &count, sizeof(long long));
```

# DYNOSTATS



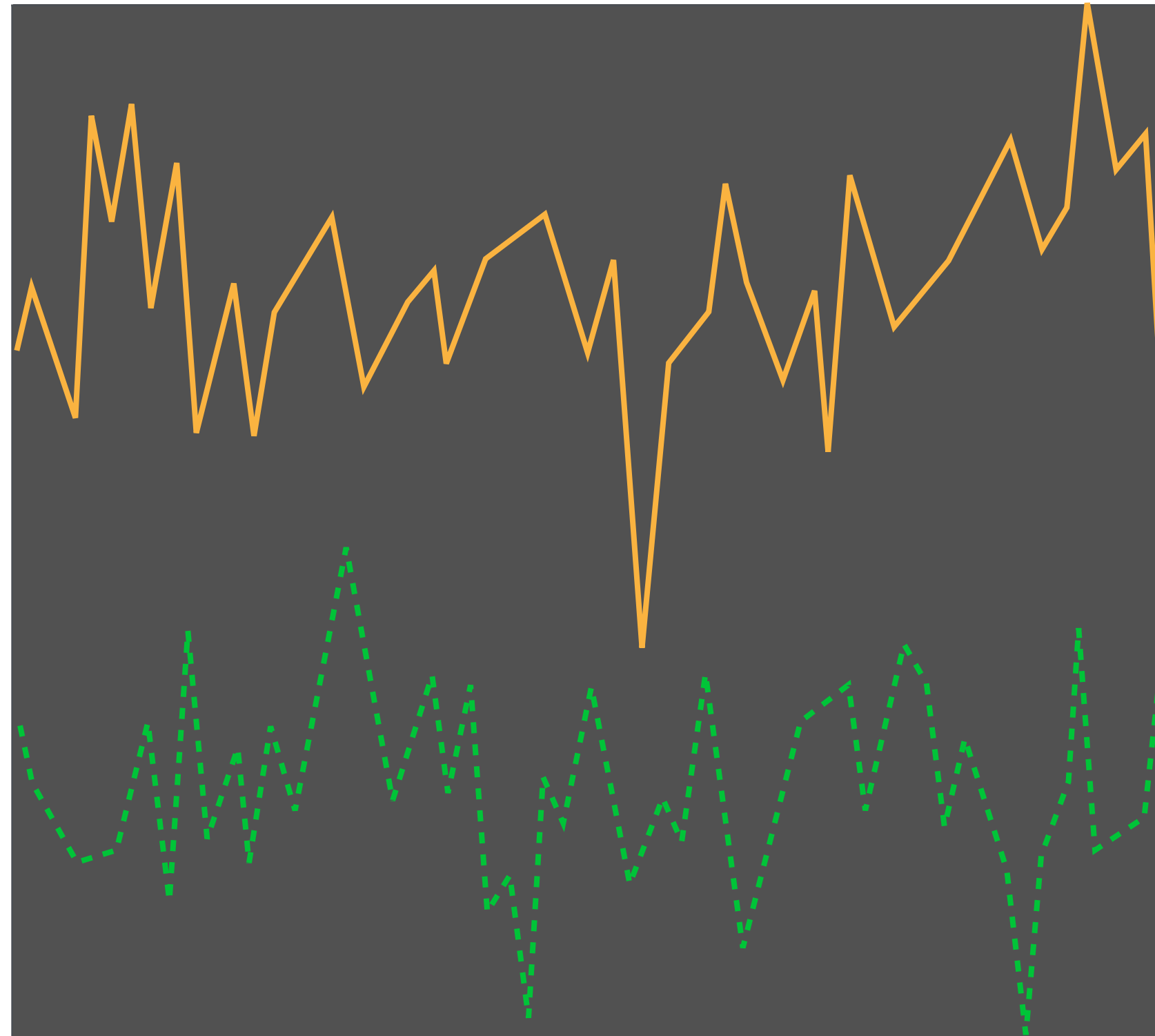
# REGRESSION





# GRADUAL REGRESSION





With new feature

Without new feature

% CPU impact:

-60

-40

-20

0

20

40

60



CPU

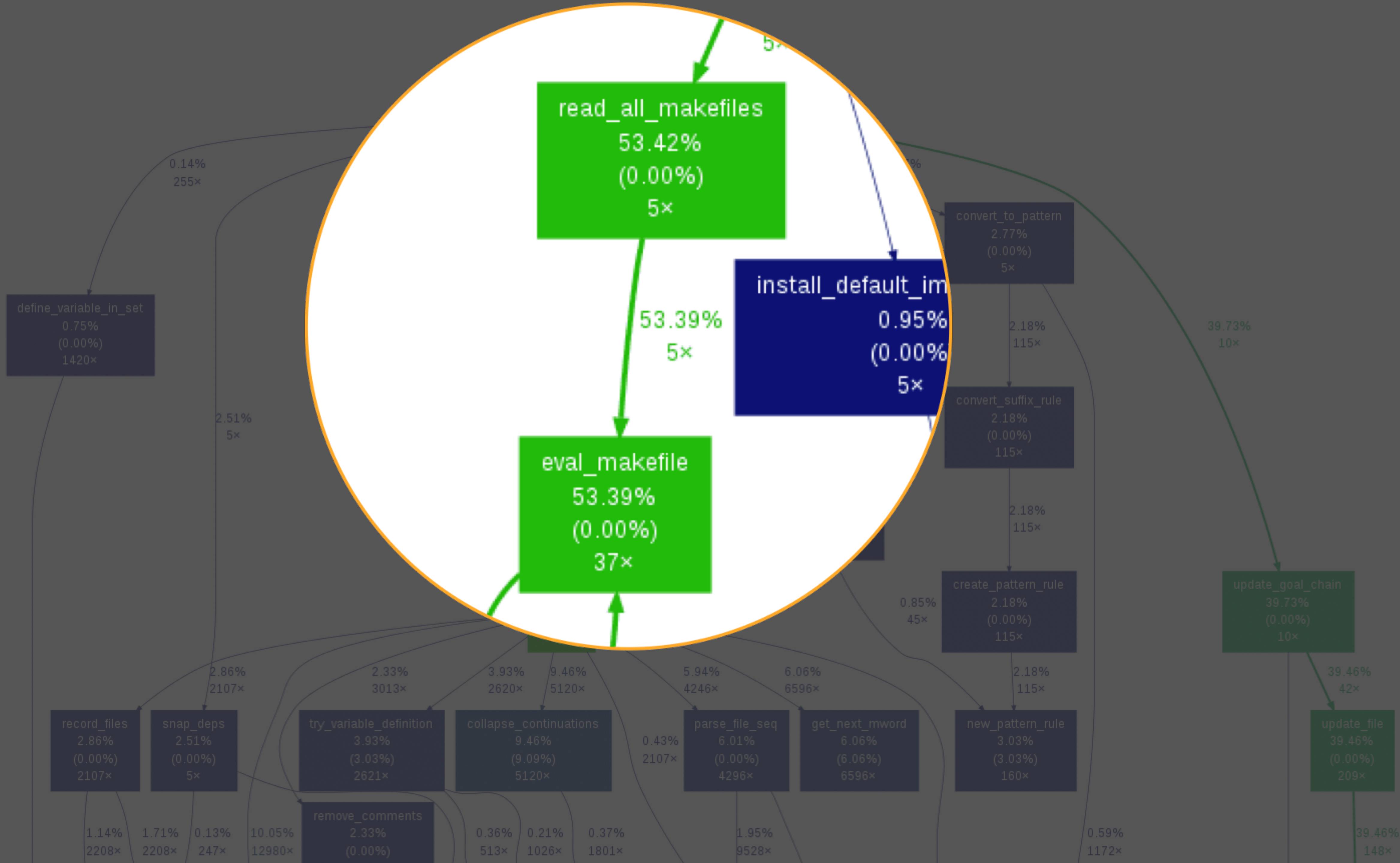
Monitor

Analyze

Optimize

# PYTHON CPROFILE

```
import cProfile, pstats, StringIO
pr = cProfile.Profile()
pr.enable()
# ... do something ...
pr.disable()
s = StringIO.StringIO()
sortby = 'cumulative'
ps = pstats.Stats(pr, stream=s).sort_stats(sortby)
ps.print_stats()
print s.getvalue()
```



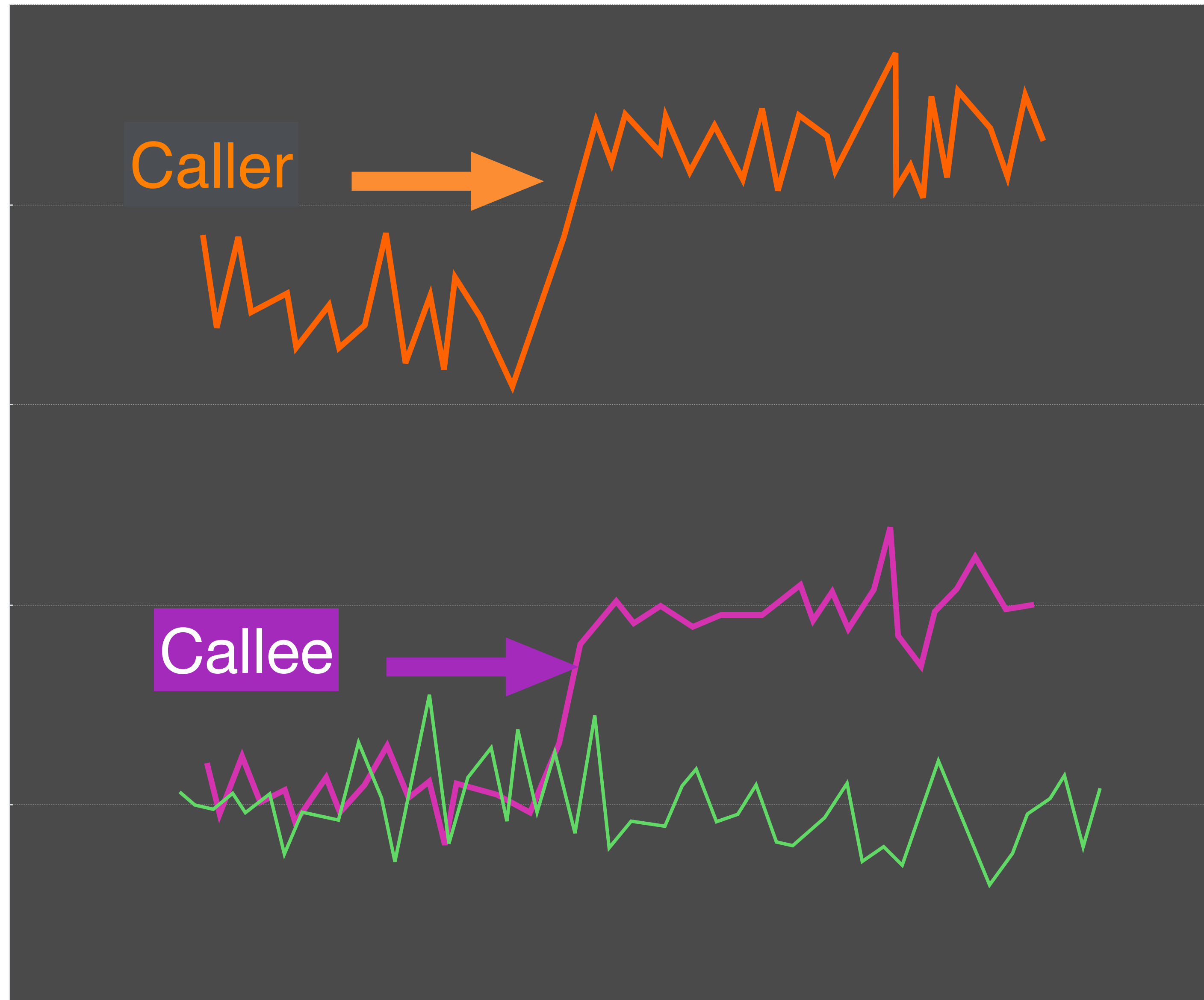
# CPU - ANALYZE

continuous profiling

```
generate_profile explore --start <start-time> --duration <minutes>
```

# CPU - ANALYZE

continuous profiling





# CPU - ANALYZE

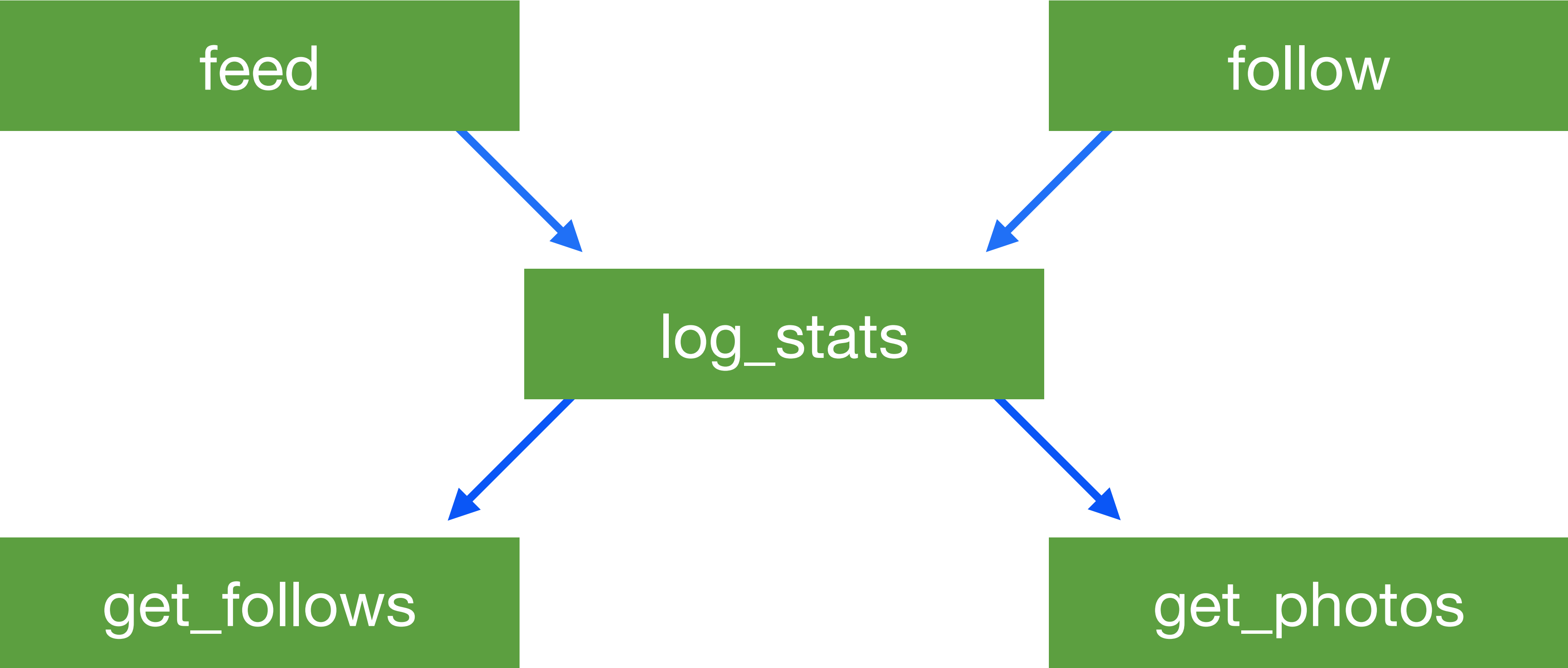
decorator

```
@log_stats
def get_photos():
    .....

def feed():
    get_photos()
```

```
@log_stats
def get_follows():
    .....

def follow():
    get_follows()
```



feed



get\_photos

follow



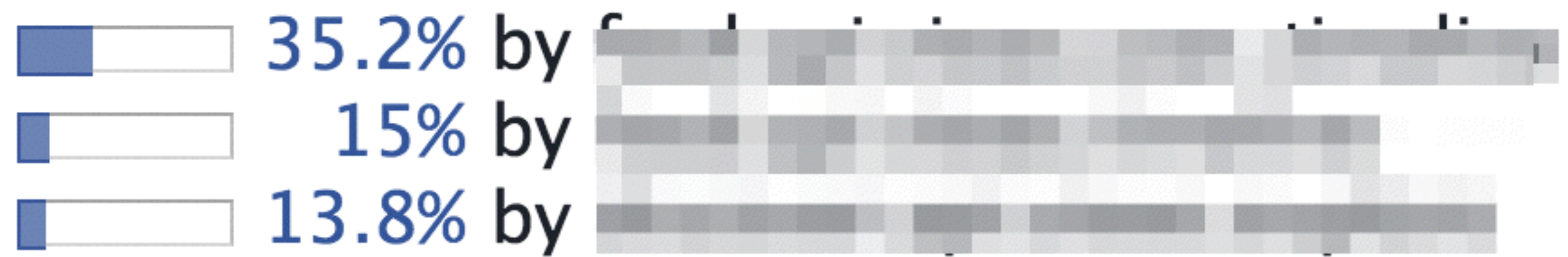
get\_follows

Global CPU consumption by this function: inclusive and exclusive .

We need ~**400** IG servers for this function!

**Drill Down**

Top Views ( [see all](#) ):



Top Callers ( [see all](#) ):



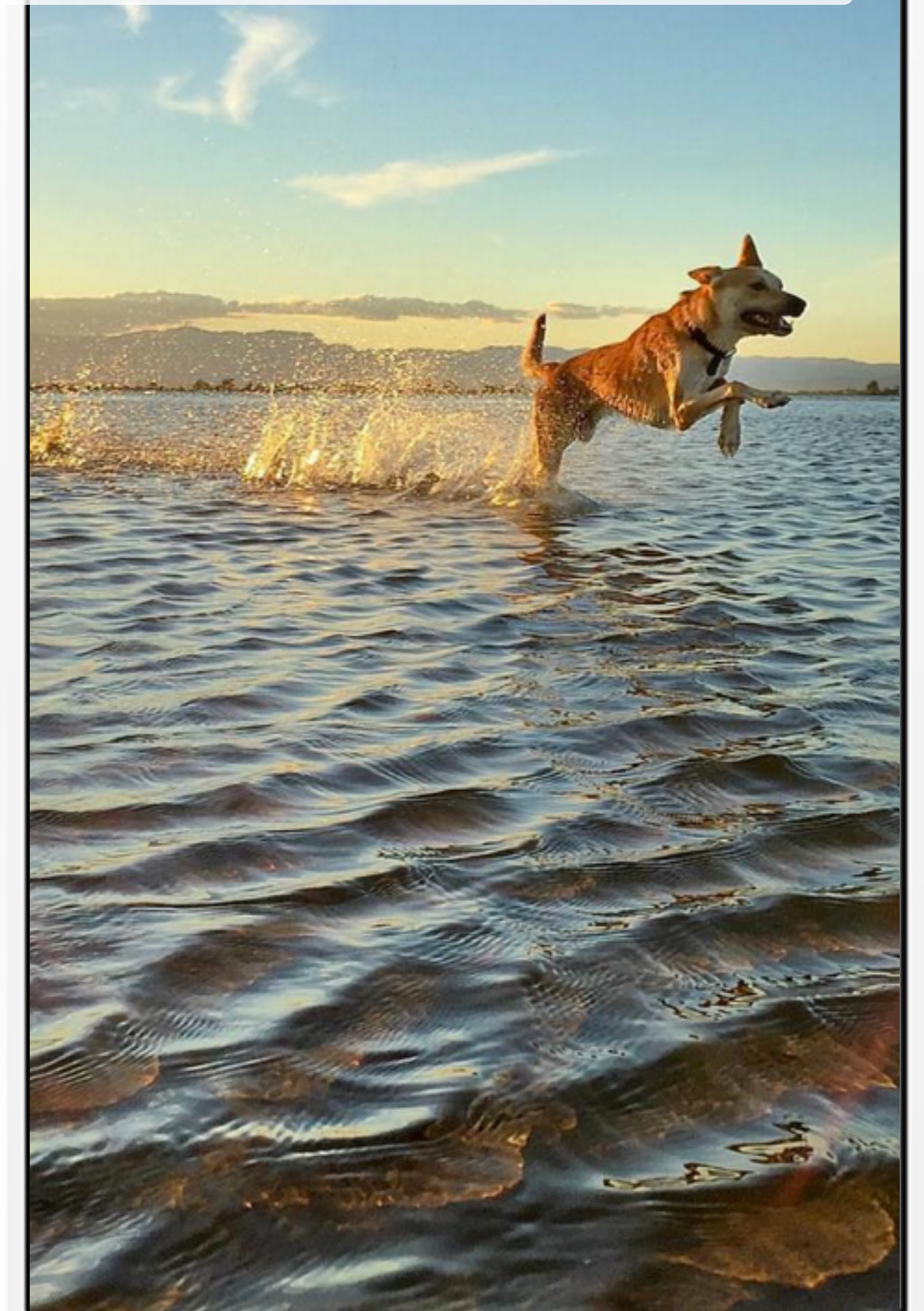
CPU

Monitor

Analyze

Optimize

[igcdn-photos-d-a.akamaihd.net/hphotos-ak-xpl1/t51.2885-19/s300x300/12345678\\_1234567890\\_987654321\\_a.jpg](https://igcdn-photos-d-a.akamaihd.net/hphotos-ak-xpl1/t51.2885-19/s300x300/12345678_1234567890_987654321_a.jpg)



igcdn-photos-d-a.akamaihd.net/hphotos-ak-xpl1/t51.2885-19/  
s300x300/12345678\_1234567890\_987654321\_a.jpg

igcdn-photos-d-a.akamaihd.net/hphotos-ak-xpl1/t51.2885-19/  
s150x150/12345678\_1234567890\_987654321\_a.jpg

igcdn-photos-d-a.akamaihd.net/hphotos-ak-xpl1/t51.2885-19/  
s400x600/12345678\_1234567890\_987654321\_a.jpg

igcdn-photos-d-a.akamaihd.net/hphotos-ak-xpl1/t51.2885-19/  
s200x200/12345678\_1234567890\_987654321\_a.jpg

CPU - OPTIMIZE

do less



igcdn-photos-d-a.akamaihd.net/hphotos-ak-xpl1/t51.2885-19/  
s300x300/12345678\_1234567890\_987654321\_a.jpg

150x150

400x600

200x200

# CPU - OPTIMIZE

**C is really faster**

- Candidate functions:
  - Used extensively
  - Stable
- Cython or C/C++

# CPU - CHALLENGE

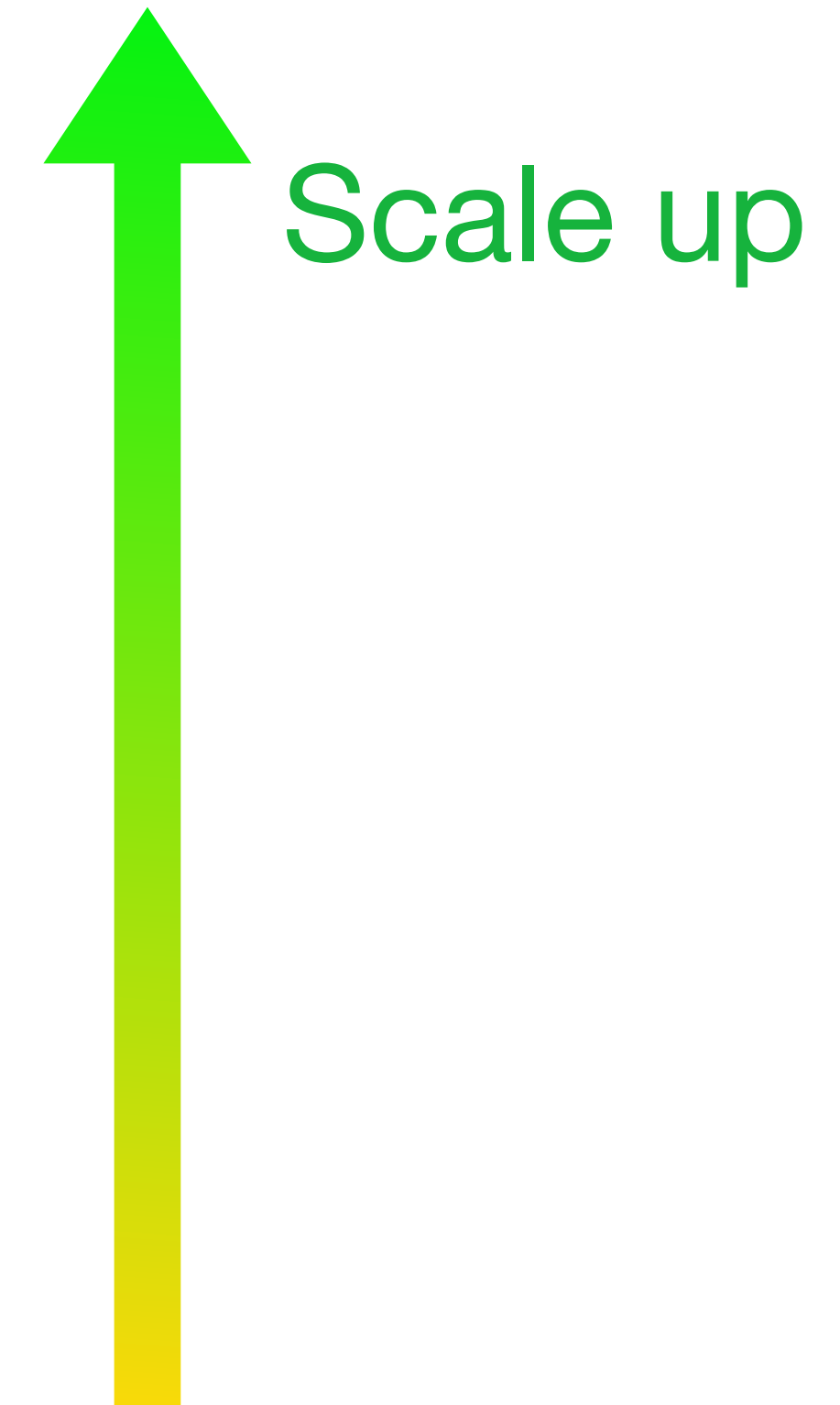
cProfile is not free

False positive alerts

Better automation


Use as few CPU instructions as possible

Use as few servers as possible



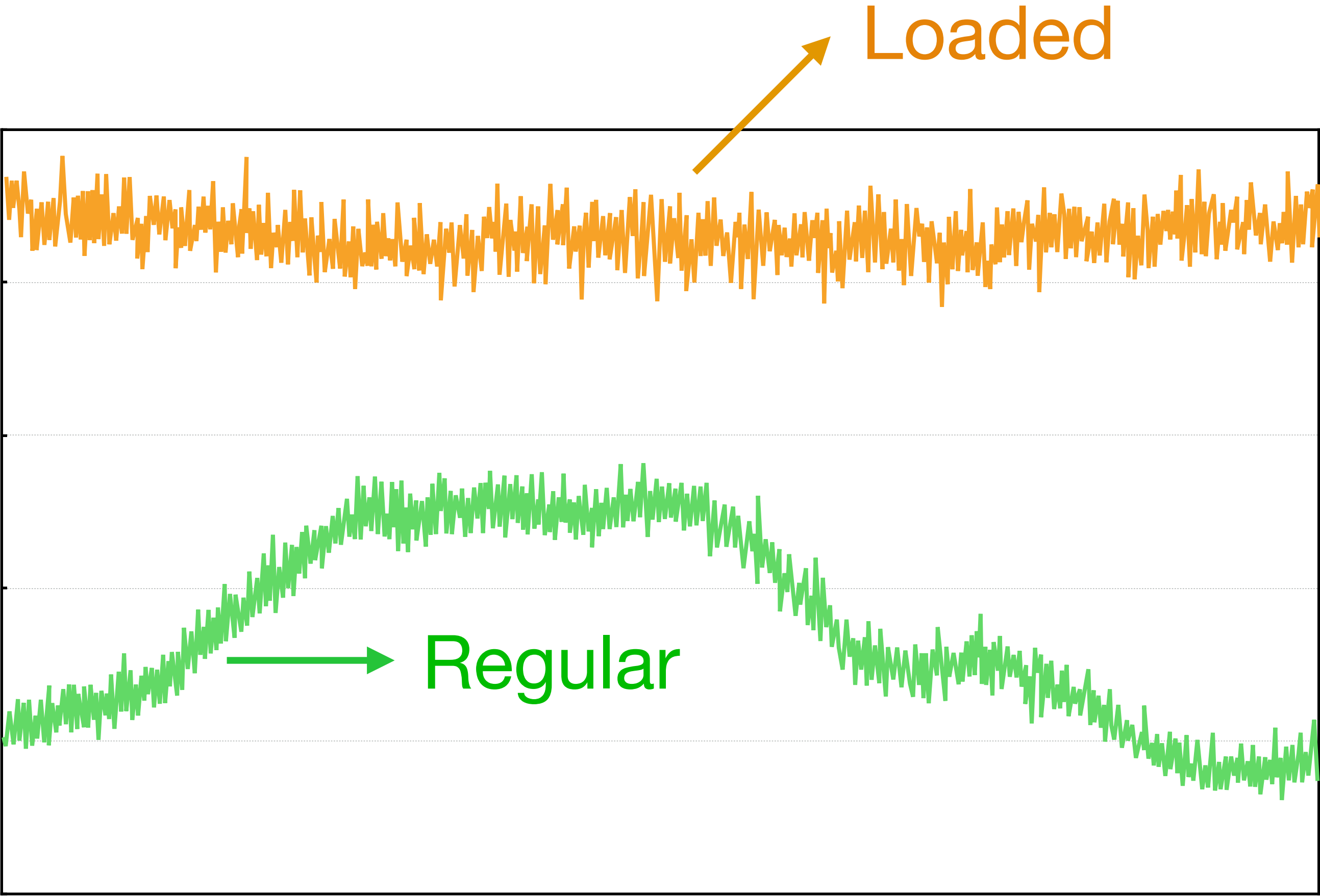
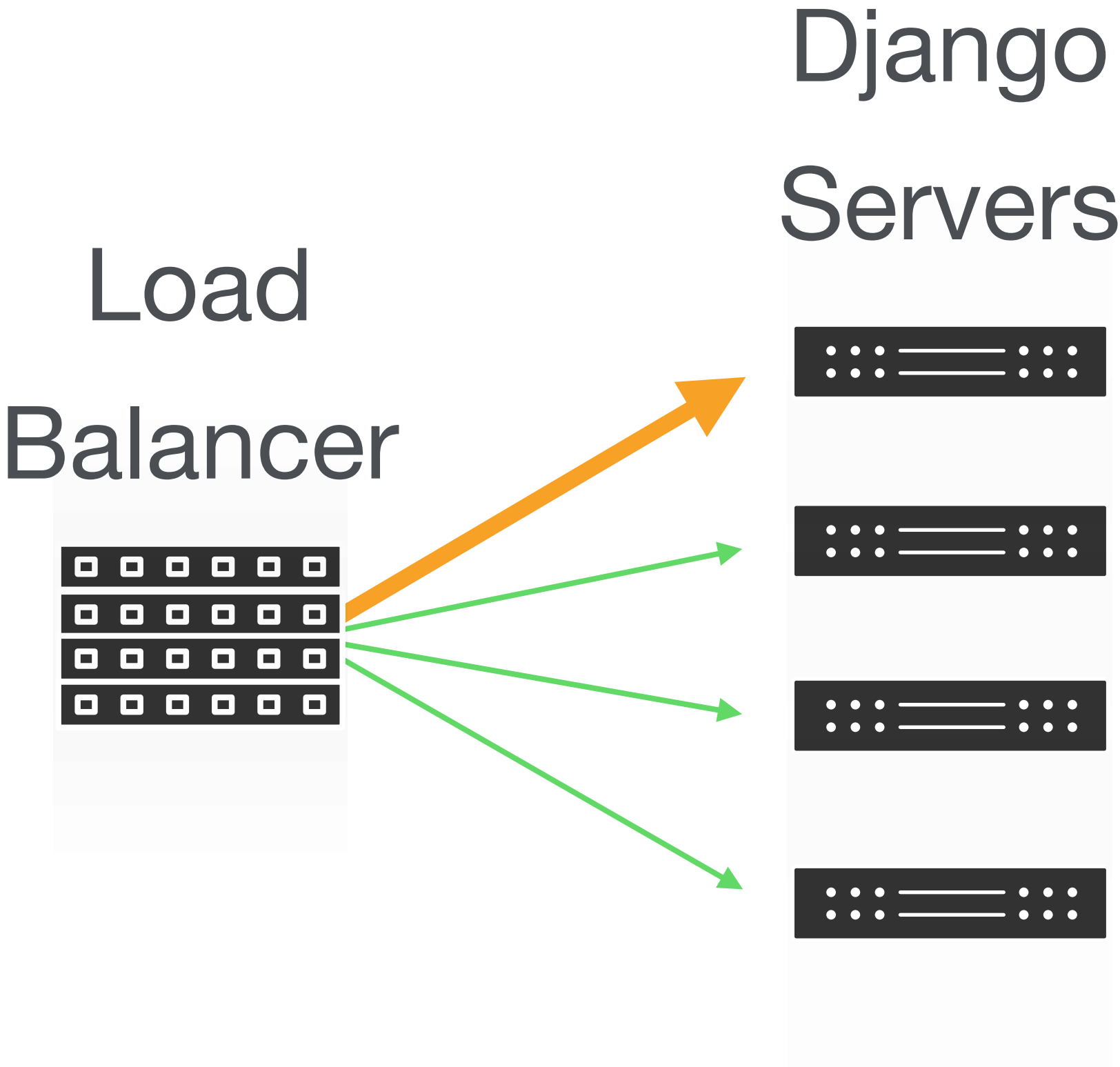
# SCALE UP: MEMORY

(memory budget /process)  $\times$  (# of processes) < system memory

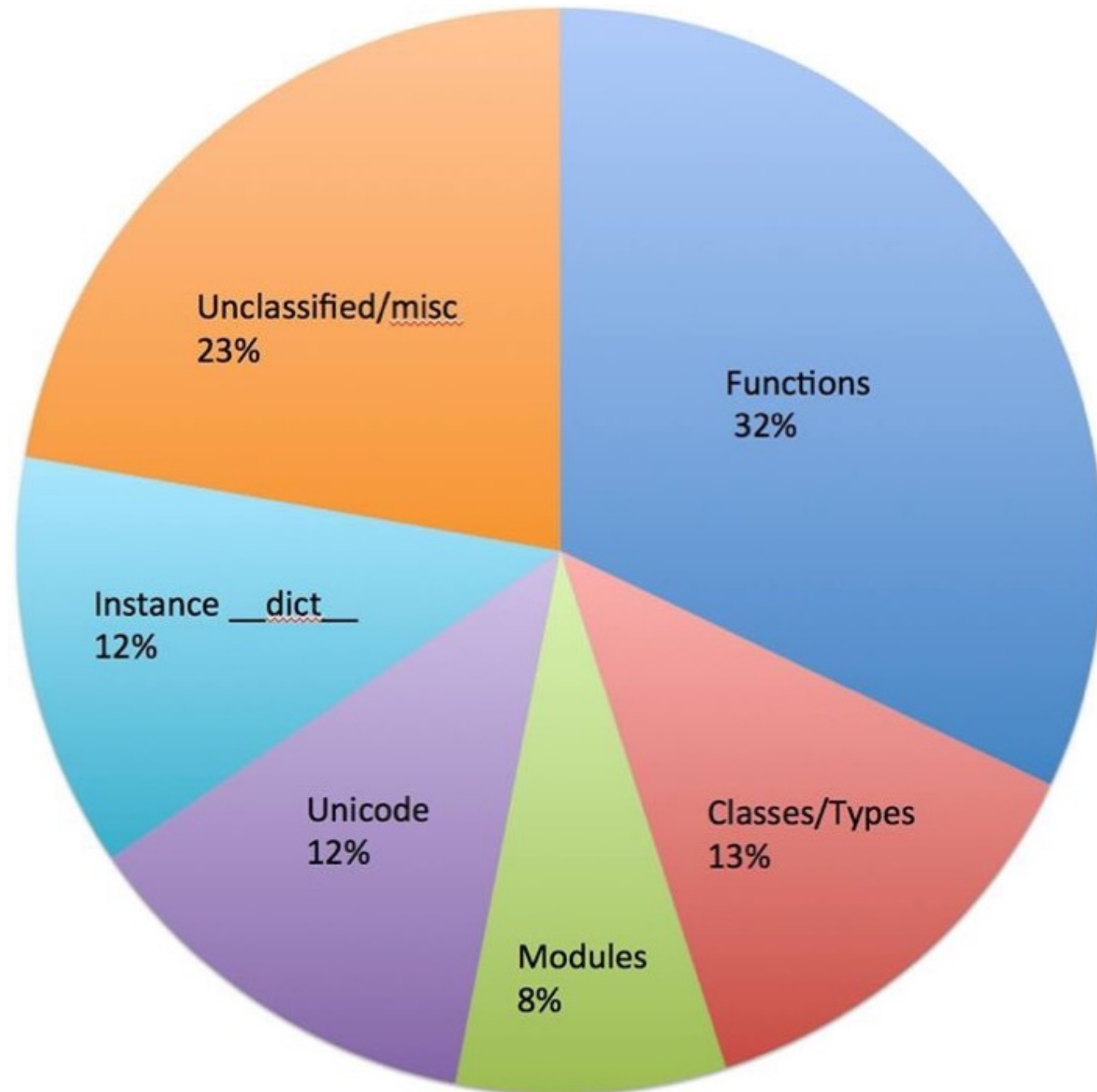
Less memory budget/process  $\implies$  More processes 

$\implies$  Dies sooner 

# LOAD TEST



# SCALE UP: MEMORY



Code

Large configuration

# SCALE UP: MEMORY

- Run in optimized mode (-O)
- Use shared memory
- NUMA
- Remove dead code

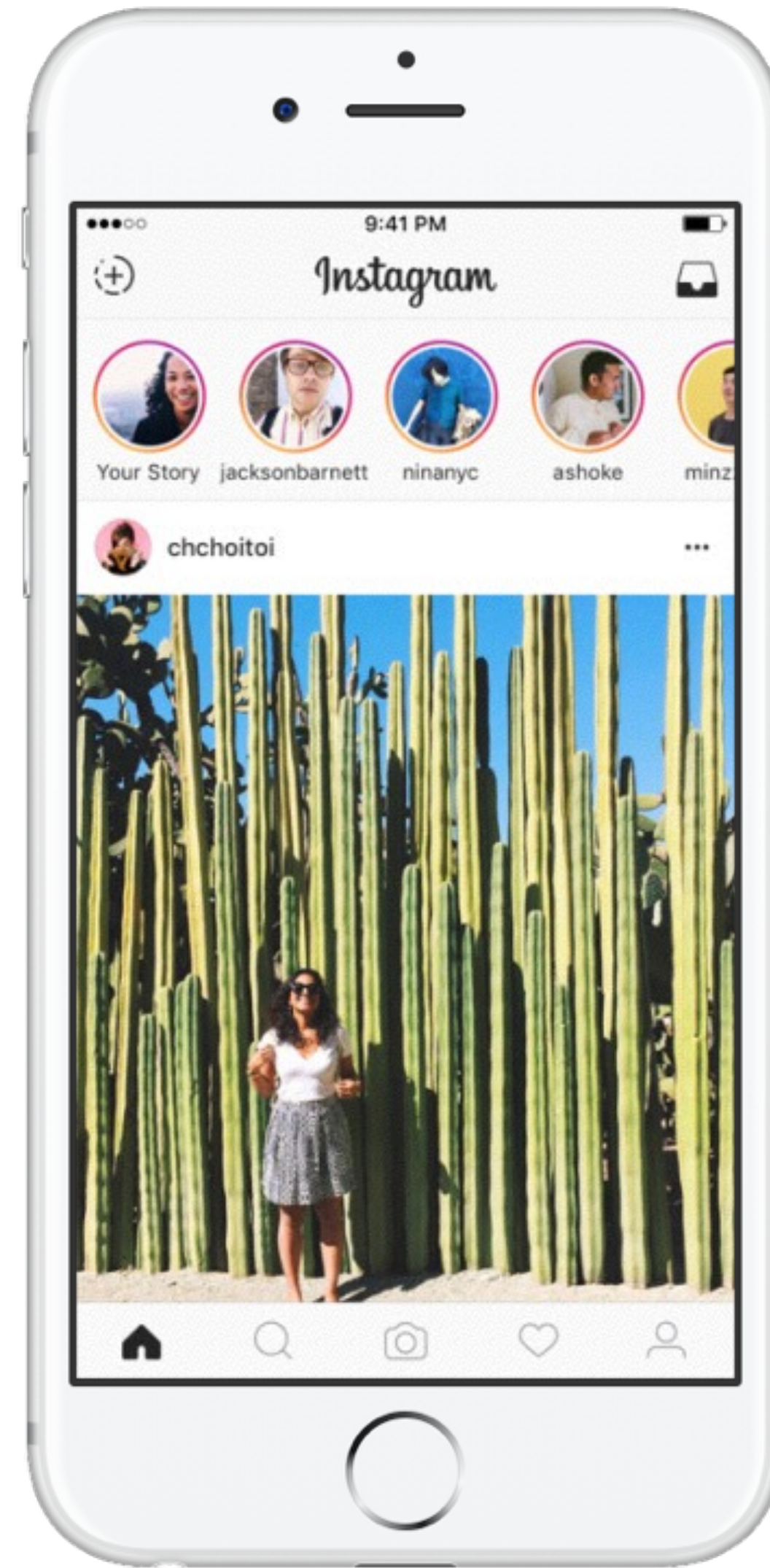
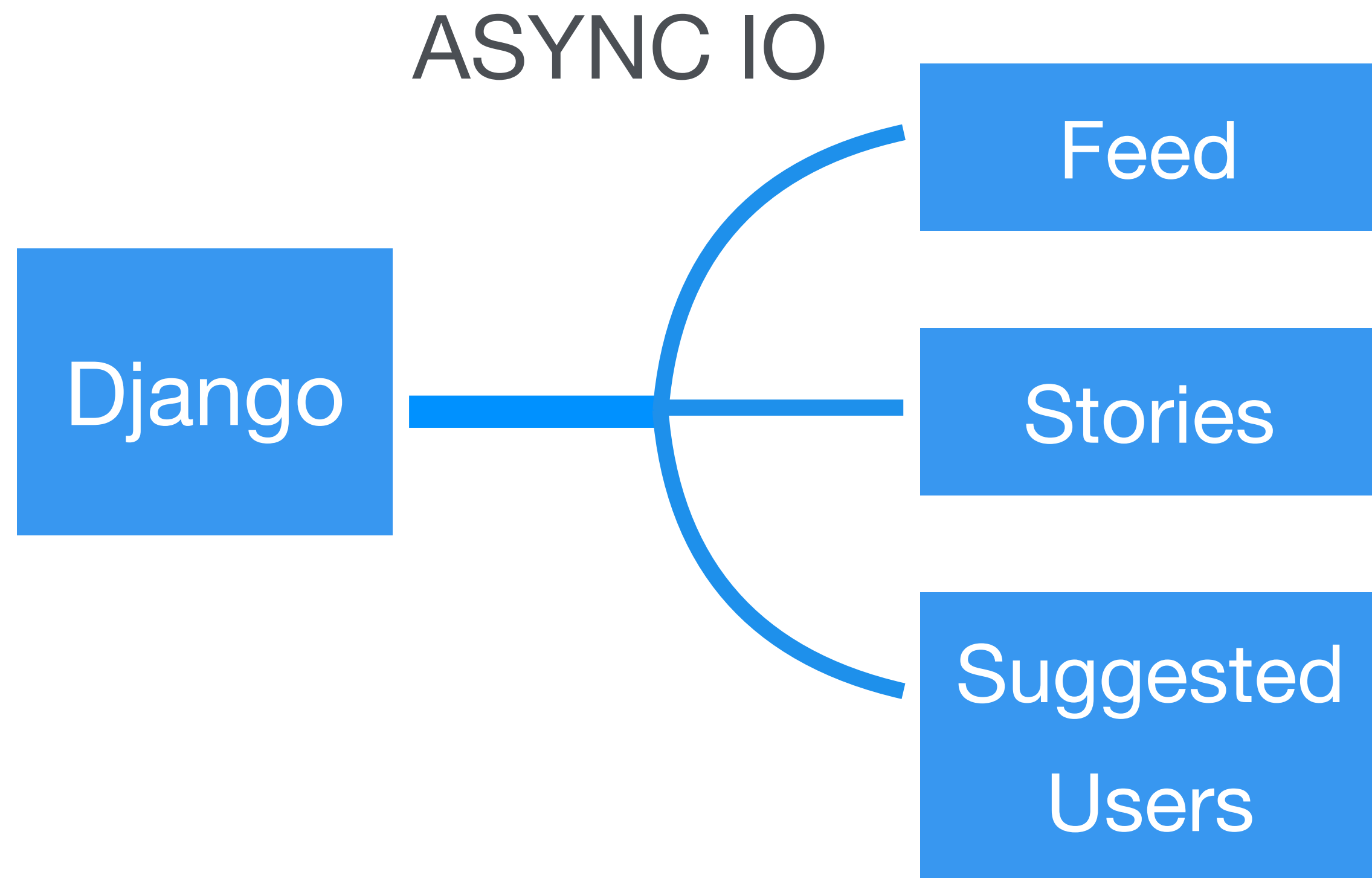


# SCALE UP: LATENCY

Synchronous Processing model ==> Worker starvation

Single service degradation ==> All user experience impacted

Longer latency ==> Fewer CPU instr executed

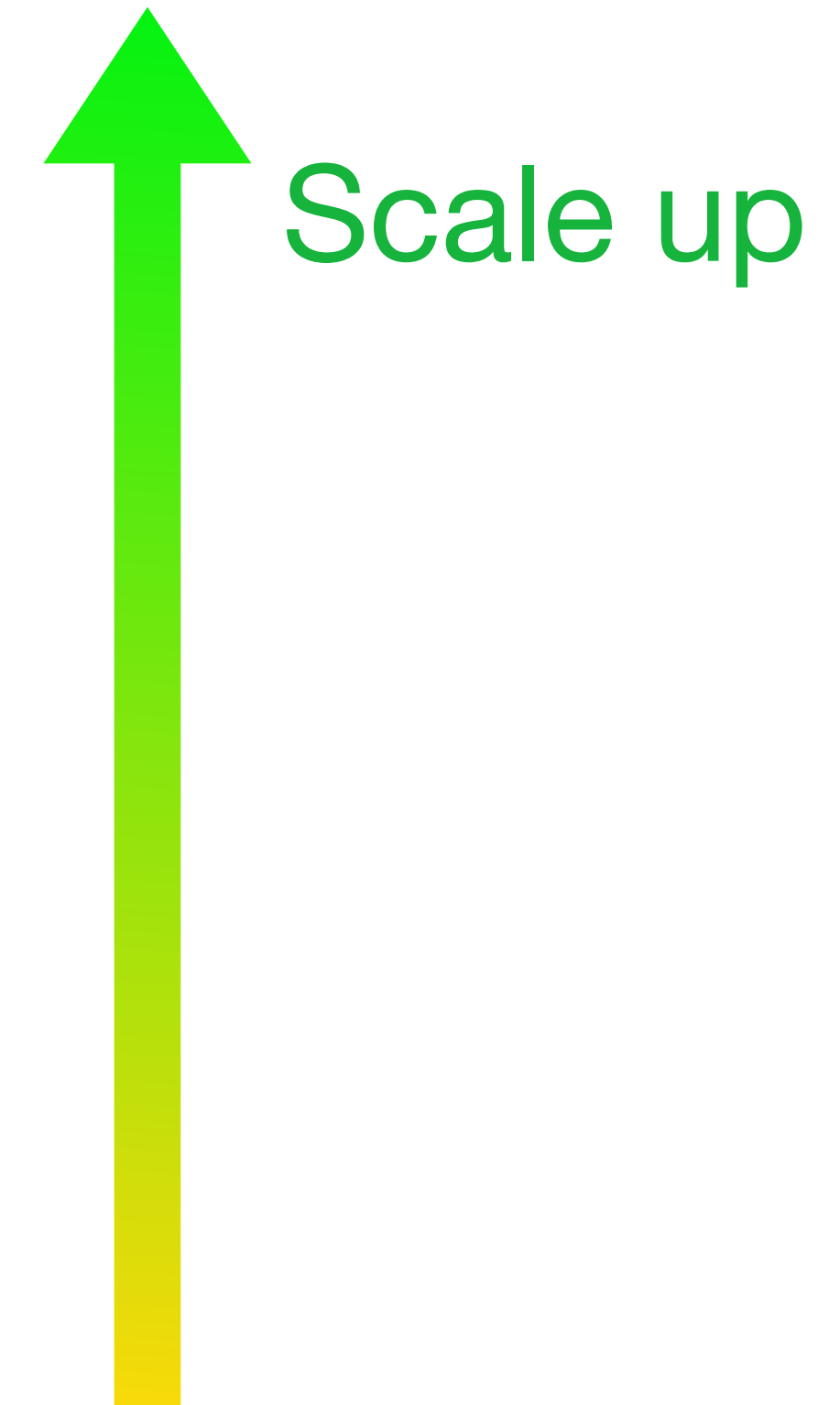


Stories

Feed

Use as few CPU instructions as possible

Use as few servers as possible





SCALE DEV TEAM

# SCALING TEAM

30% engineers joined in last 6 months

Bootcampers - 1 week

Hack-A-Month - 4 weeks

Intern - 12 weeks

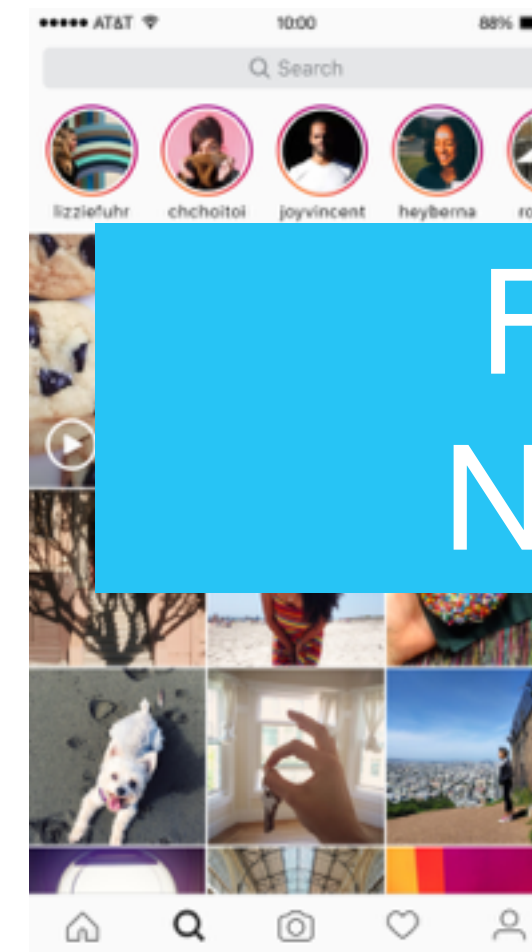
Save Draft

Story Viewer Ranking

Comment Filtering



Windows App



First Story Notification

Video View Notification

Self-harm Prevention

Will I bring down  
Instagram?



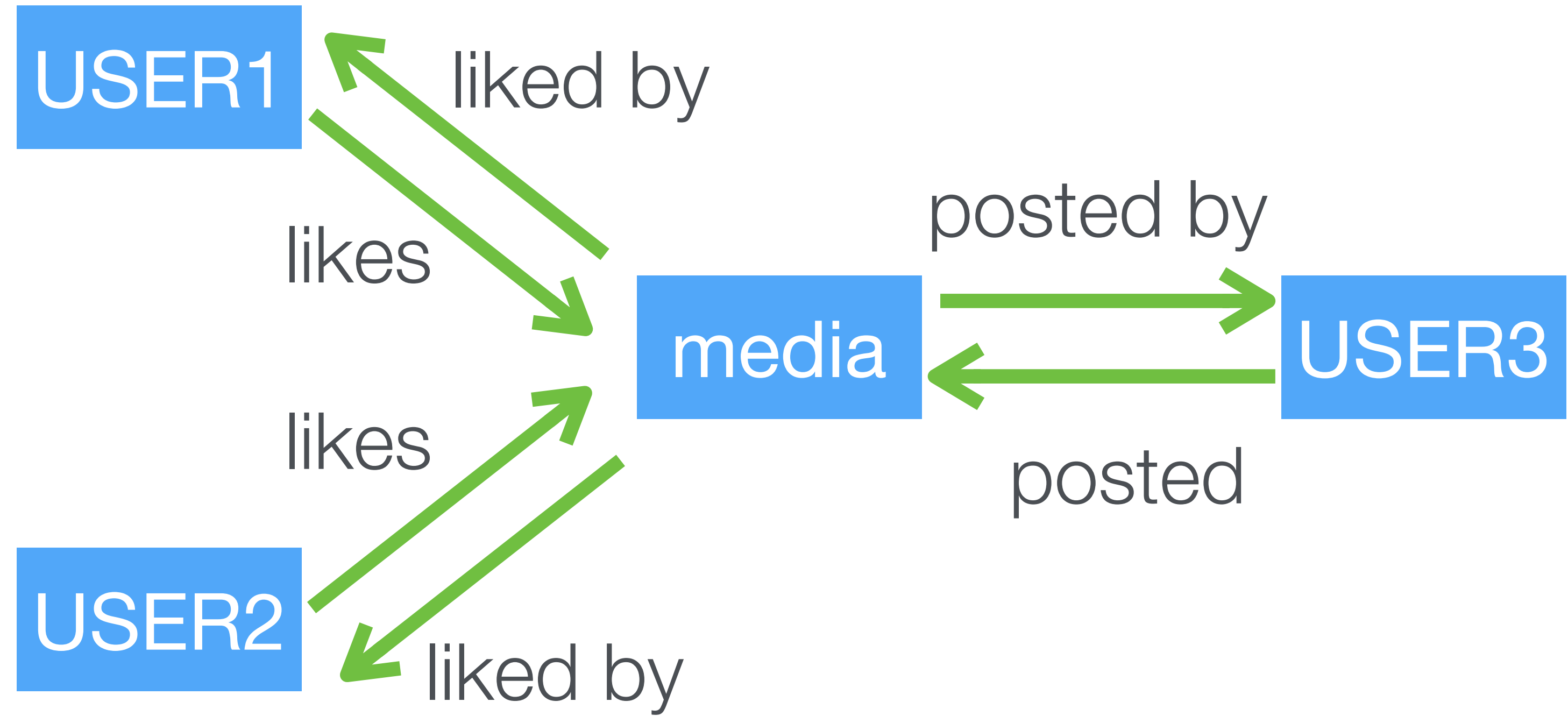
# WHAT WE WANT

- Automatically handle cache
- Define relations, not worry about implementations
- Self service by product engineers
- Infra focuses on scale this service



道

TAO



# SCALE DEV - END OF POSTGRES



Cool DPs at [dp.topcovers4fb.com](http://dp.topcovers4fb.com)



# SHIPPING LOVE

>120 engineers committed code last month

60-80 daily diffs

# RELEASE

- Master, no branch
- All features developed on master gated by configuration
- Continuous integration
- No branch integration overhead
- No surprises
- Iterate fast, collaborate easily
- Fast bisect and revert

Ship  
It!



Once a week?

40-50 rollouts per day

# CHECKS AND BALANCES

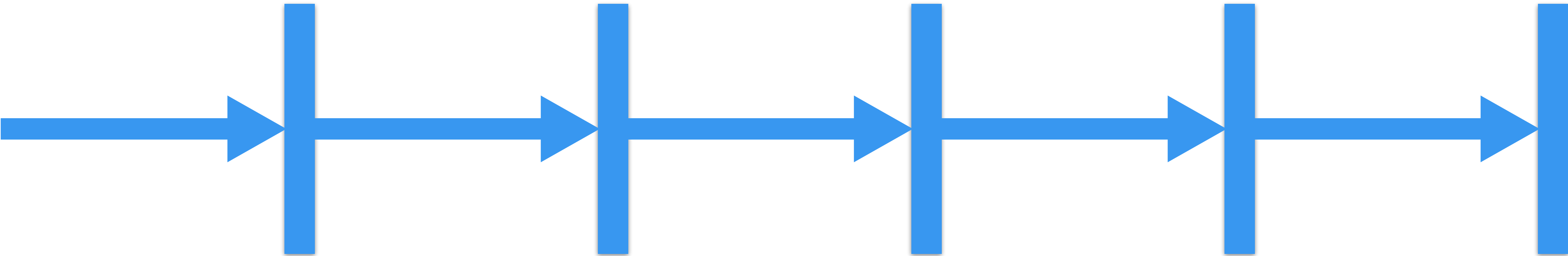
Code review  
unittest

Code accepted  
committed

Canary

Dark launch  
Load test

To the Wild



**ALERT**



**DO THE  
NEEDFUL  
AND  
REVERT**

# TAKEAWAYS

Scaling is a continuous effort

Scaling is multi-dimensional

Scaling is everybody's responsibility



QUESTIONS?

