

# **COMP47250 Team Software Project**

## **Final Presentation**

 Lovelace

August 2016

**Jón Rúnar Helgason, Xinqi Li, Marc Laffan,  
Junyang Ma, Eazhilarasi Manivannan**

# Project Vision: What?

- Content-based tweet recommendations
- Personalised tweet timeline
- More relevant tweets place higher
- Less relevant tweets placed at the back

# Project Vision: Who?

- Power users
- Follows +200 accounts
- Tweets/retweets/likes five times daily
- May use Twitter for educational, entertainment, political or any number of purposes.
- Has been using Twitter for months if not years.

# Project Vision: Why?

- Twitter presents tweets in chronological order
- Accounts often do not have a single focus, instead they cover multiple subjects (Scott Hanselman)
- The more accounts a user follows, the more subjects their tweet feed will cover, thus creating a “noise” of irrelevant tweets.
- To combat this, we are re-organising the users feed to emphasize relevant tweets and de-emphasize tweets that the user does not like
- Twitter is attempting recommendations itself as well

# Competitors



**Official Twitter  
mobile app & website**



# Front End

iPhone App



iOS 9



Swift 2.2

# Back End

Web Service



Flask

+



Python 2.7

Recommender System



+ Python 2.7 scikit-learn, Tweepy

Twitter API



Data Collection



Celery

Database

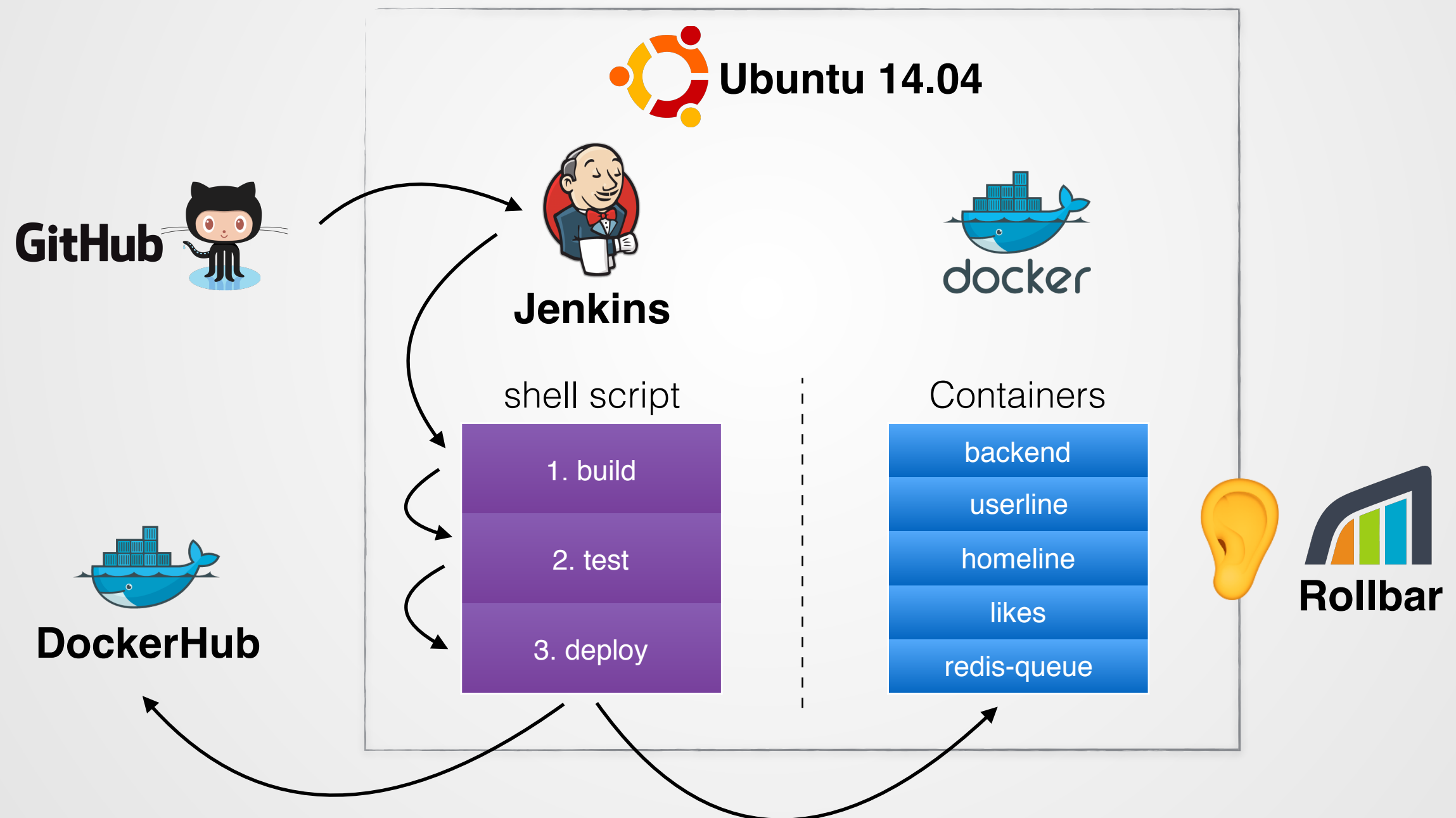


RethinkDB

# Tech Stack

# Workflow

## UCD VM



# Docker Containers in Production

## Flask web service

```
jonrh — Flask Backend — ssh student@csi6220-1-vm1.ucd.ie — 159x12
0.0
scored
pixee_pea
0.0
scored
londonist
0.0
scored
standardnews
0.0
[2016-08-08 22:39:57 +0000] [9] [DEBUG] DELETE /userLogout
```

## Celery: collecting homeline tweets

```
jonrh — Celery Worker - Homeline collection — ssh student@csi6220-1-vm1.ucd.ie — 159x6
[2016-08-08 22:29:46,935: WARNING/Worker-2] 36
[2016-08-08 22:29:47,248: INFO/Worker-2] PARAMS: {}
[2016-08-08 22:29:47,250: INFO/Worker-2] Starting new HTTPS connection (1): api.twitter.com
[2016-08-08 22:29:47,453: INFO/MainProcess] Task tasks.get_tweet[23c7f007-8c75-4f0a-941d-8a4cc8baa759] succeeded in 1.45731555694s: {u'reset': 1470696192, u'limit': 15, u'remaining': 3}
```

## Celery: collecting users tweets

```
jonrh — Celery Worker - Userline collection — ssh student@csi6220-1-vm1.ucd.ie — 159x6
[2016-08-08 22:23:03,726: INFO/Worker-8] Starting new HTTPS connection (1): api.twitter.com
[2016-08-08 22:23:03,898: WARNING/Worker-8] 0
[2016-08-08 22:23:03,899: INFO/MainProcess] Task tasks_user.get_user_tweet[a0b516b9-7de2-484f-848e-c92b104f5fef] succeeded in 0.242800352164s: 'user_timeline'
[2016-08-08 22:23:03,908: WARNING/Worker-5] 0
[2016-08-08 22:23:03,909: INFO/MainProcess] Task tasks_user.get_user_tweet[9eac190e-2098-4575-bdb2-29951abe73cb] succeeded in 0.293662767857s: 'user_timeline'
```

## Celery: collecting liked tweets

```
jonrh — Celery Worker - Like collection — ssh student@csi6220-1-vm1.ucd.ie — 159x6
[2016-08-08 13:13:40,054: INFO/MainProcess] Task tasks_liked.get_liked_tweet[389176c1-1c79-4684-9e1b-cb1dc89d0102] succeeded in 0.253968695877s: None
[2016-08-08 13:13:40,525: WARNING/Worker-7] 22
[2016-08-08 13:13:40,697: INFO/MainProcess] Task tasks_liked.get_liked_tweet[183c17c7-e311-4bfe-b534-d5adb41caf6e] succeeded in 0.940285495017s: 'favorites'
[2016-08-08 13:13:41,410: WARNING/Worker-3] 200
[2016-08-08 13:13:43,881: INFO/MainProcess] Task tasks_liked.get_liked_tweet[0dc0962c-f8d7-416b-b900-ffbe04248845] succeeded in 4.0588478779s: 'favorites'
```



# Data Stack

- **From Twitter REST APIs**

User's home timeline

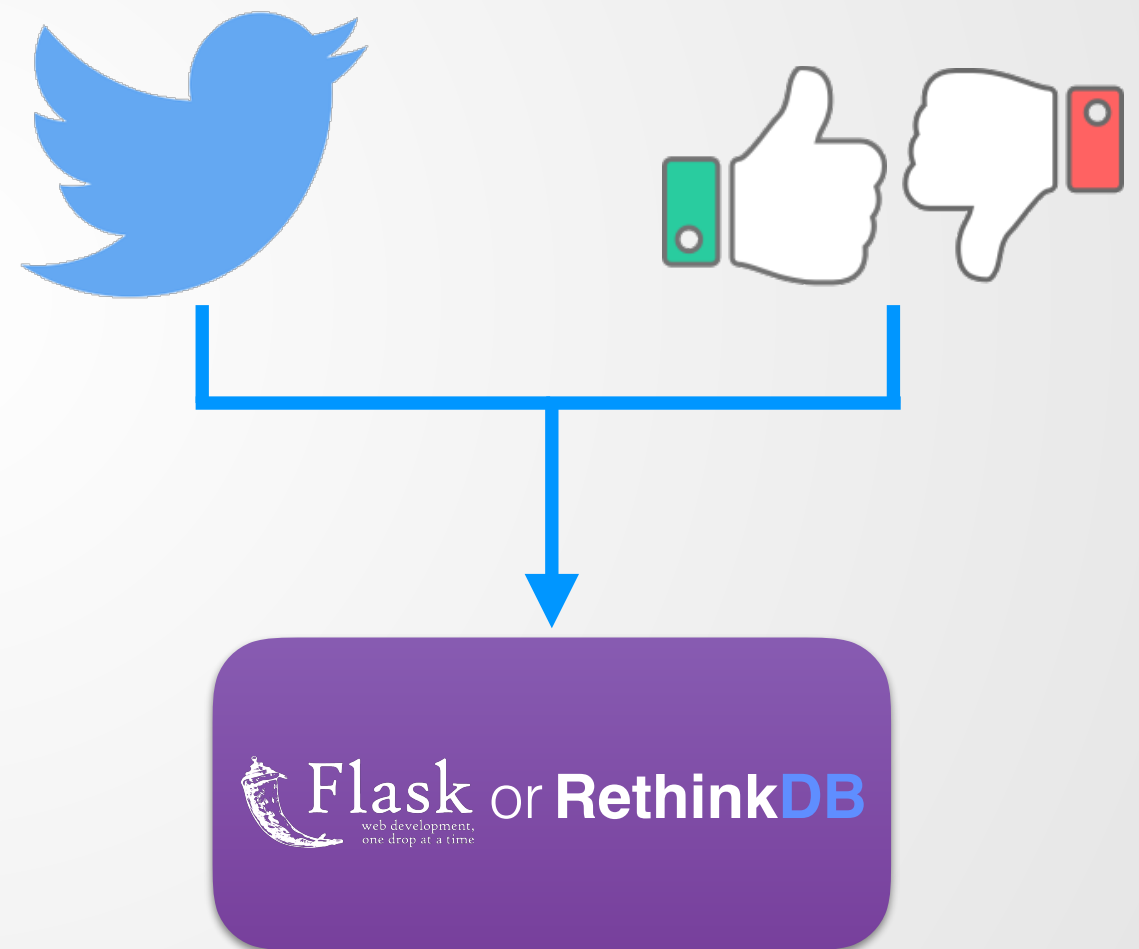
User's liked tweets

User's personal timeline

- **From iOS client**

Likes/dislikes for authors

Likes/dislikes for subject matter in Tweets



# Key Technical Challenges

- Twitter rate limit
- Tweet storage/usage
- Tweaking of recommender system variables

# Rate Limit

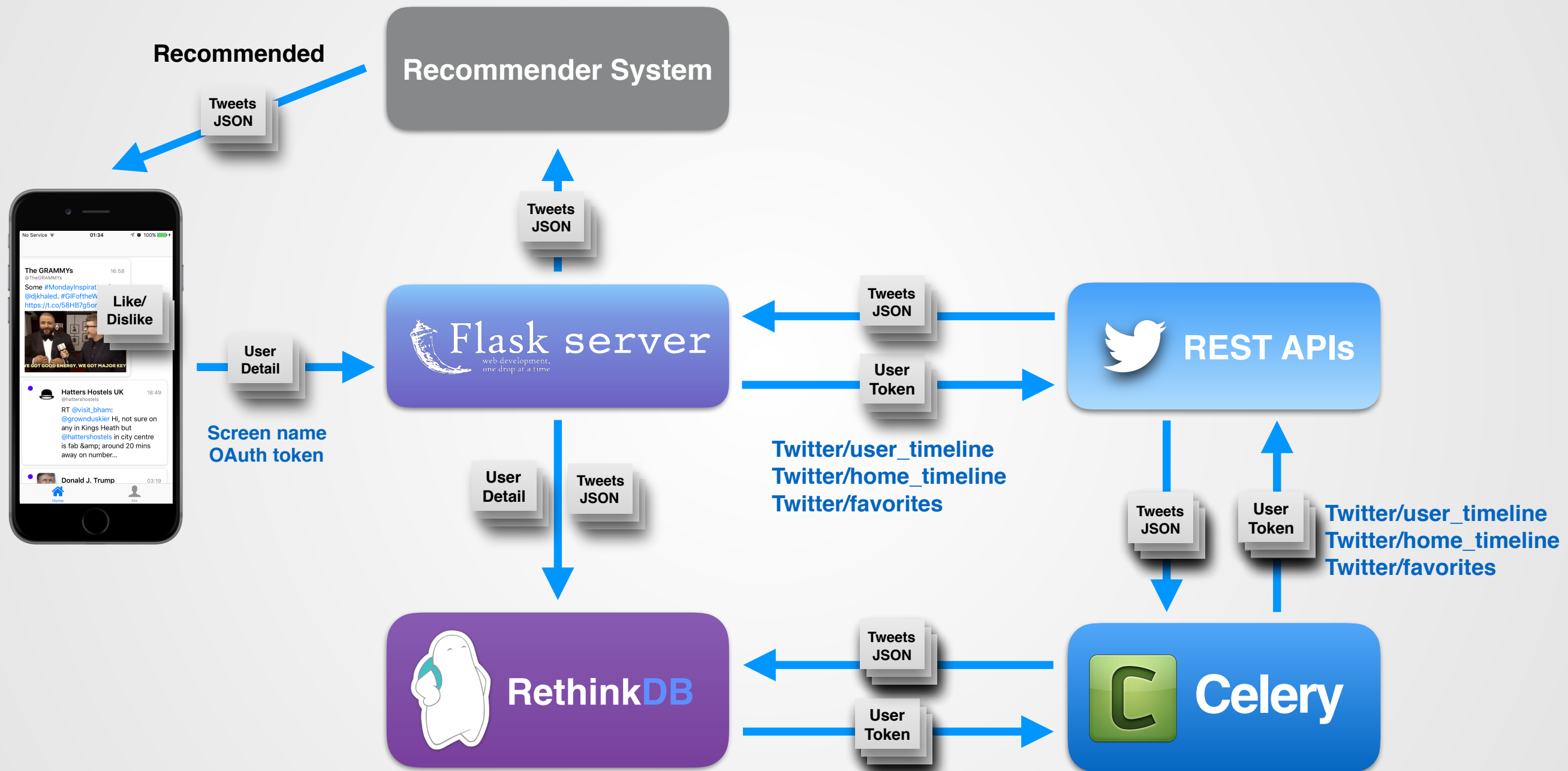
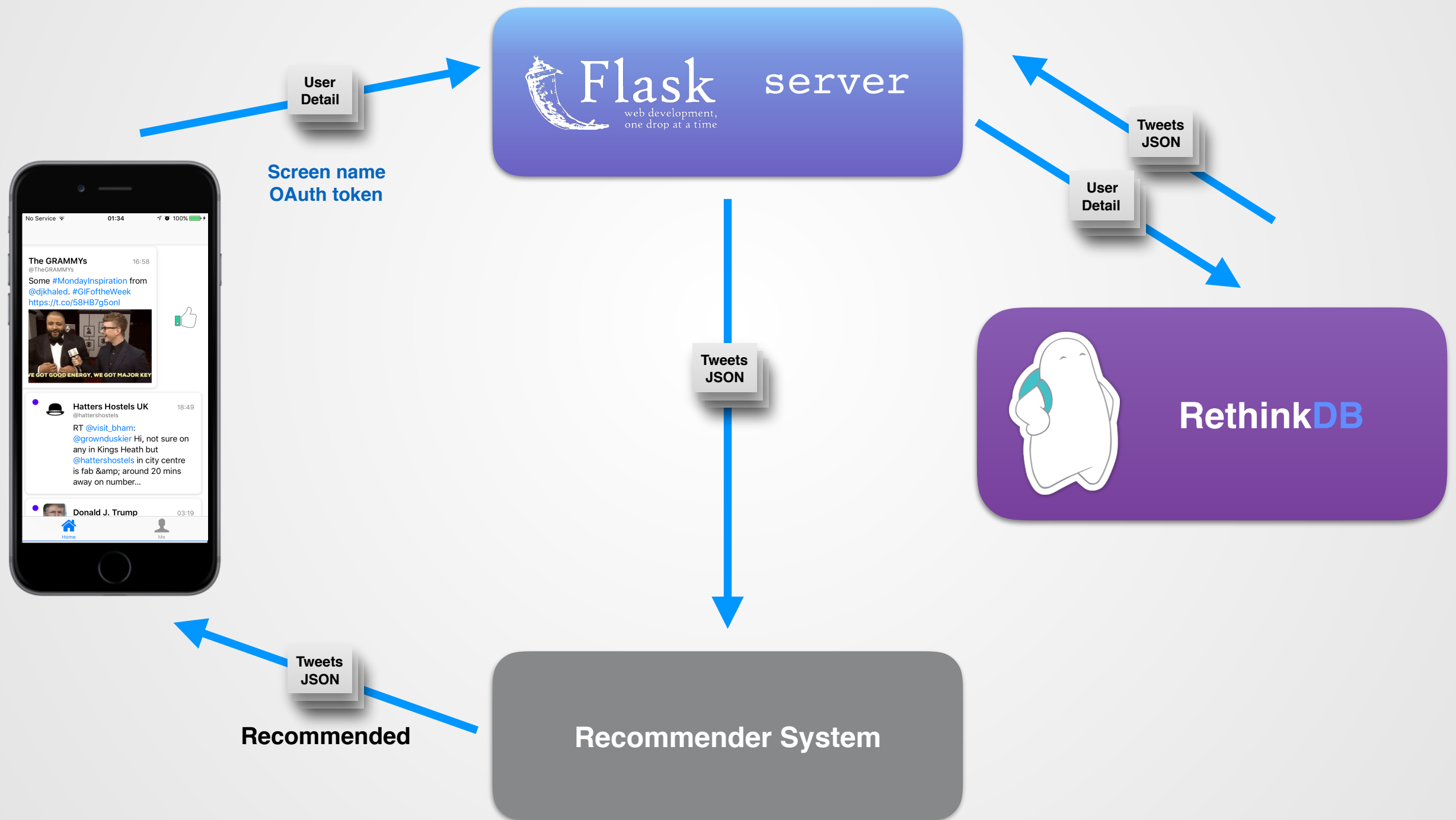


Table	user_tokens	tweets	like/dislike
Data			

# Rate Limit



# Tweet Storage

RethinkDB Administration Console

Connected to ip\_172\_31\_35\_1... | Issues: No issues | Servers: 1 connected | Tables: 6/6 ready

### Tables in the cluster

+ Add Database | - Delete selected tables

DATABASE: evaluation			
+ Add Table   Delete Database			
<input type="checkbox"/>	results	1 shard, 1 replica	Ready 1/1

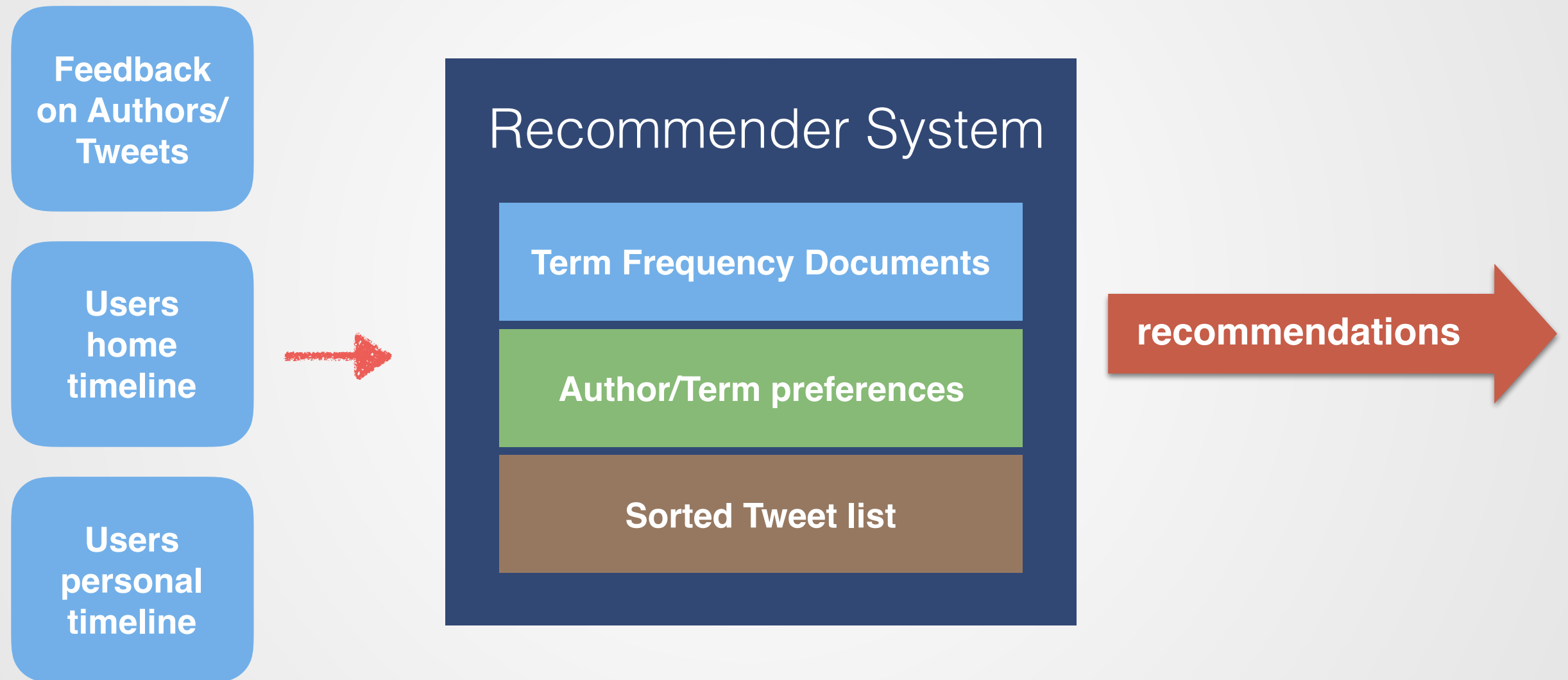
DATABASE: lovelace			
+ Add Table   Delete Database			
<input type="checkbox"/>	like_user_timeline	1 shard, 1 replica	Ready 1/1
<input type="checkbox"/>	single_feedback	1 shard, 1 replica	Ready 1/1
<input type="checkbox"/>	tweets	1 shard, 1 replica	Ready 1/1
<input type="checkbox"/>	user_tokens	1 shard, 1 replica	Ready 1/1
<input type="checkbox"/>	users	1 shard, 1 replica	Ready 1/1

Documentation | API | Google Groups | #rethinkdb on freenode | Github | Community

# Tweet Storage

```
{  
  .....  
  "screen_name": "NoUserNameFound" ,  
  ▶ "tweet": { ... } ,  
  "tweet_id": 761591895122382800  
}  
  
{  
  .....  
  "screen_name": "lambda_lovelace" ,  
  ▶ "tweet": { ... } ,  
  "tweet_id": 761591897626378200  
}
```

# Recommender System



# Evaluation

## **What:**

Past tweets/retweets/likes

no in-client feedback, no time 🕒

## **Against:**

Benchmarking against Twitter's API

chronologically ordered tweets 💪

Wanted Twitter's recommendation

but not made available in API 😭



# Evaluation



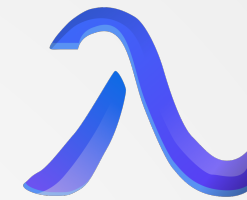
tweet 1

tweet 2

tweet 3

tweet 4

...



tweet 1

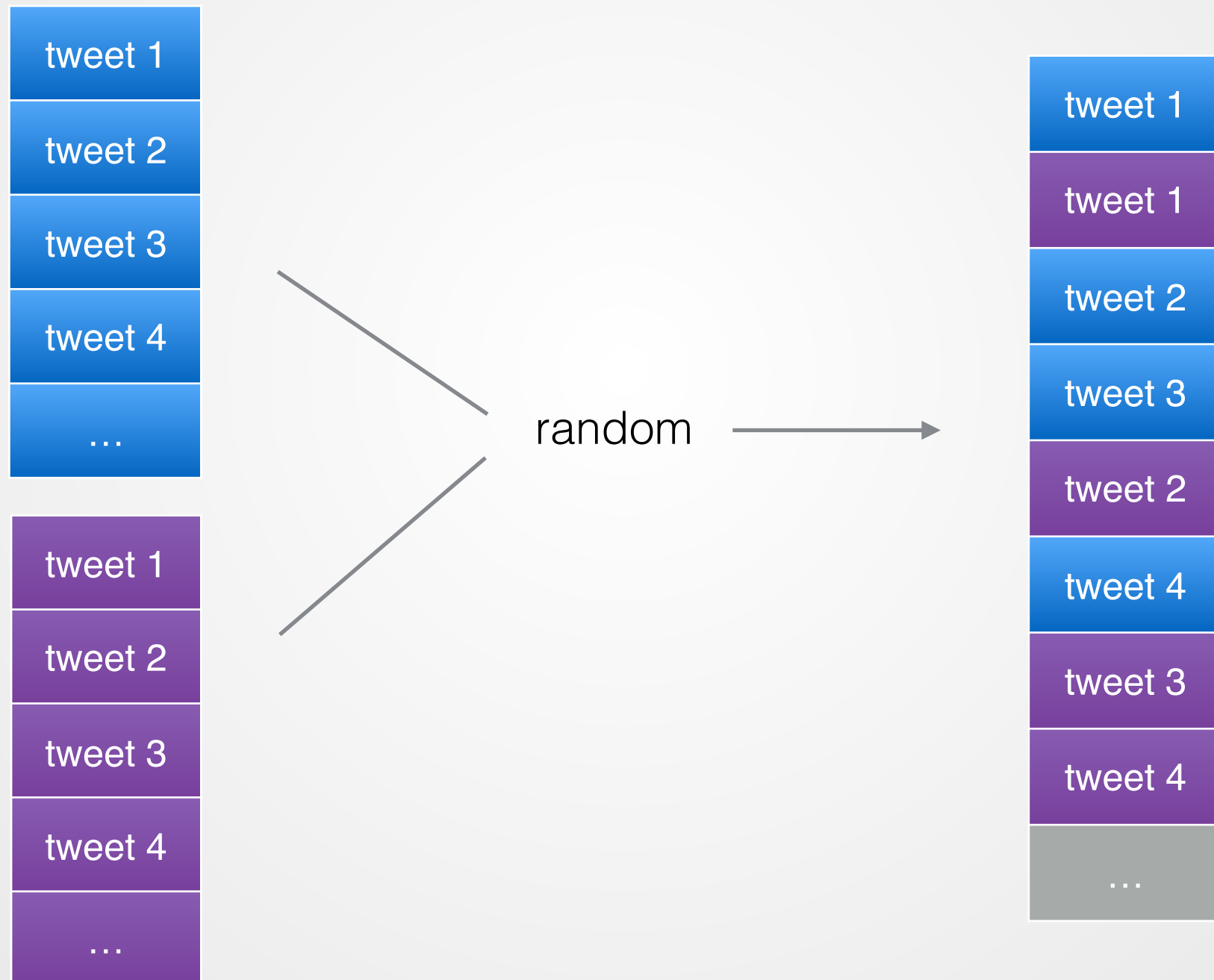
tweet 2

tweet 3

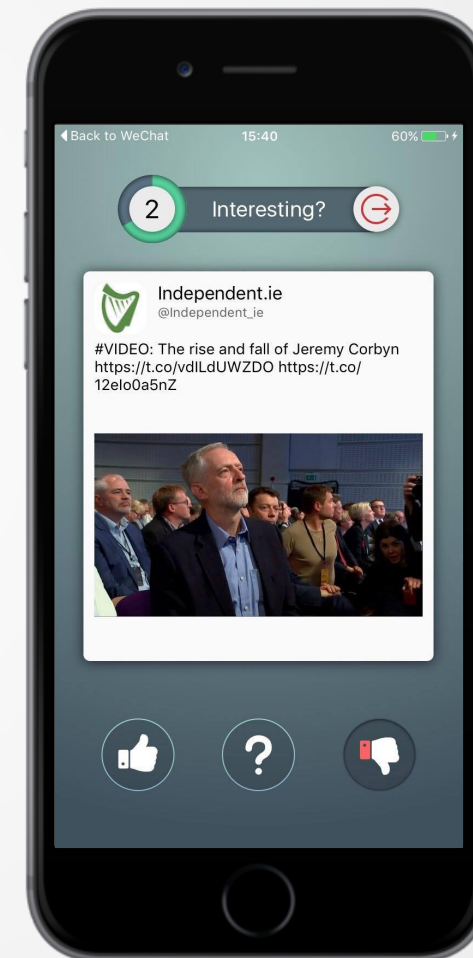
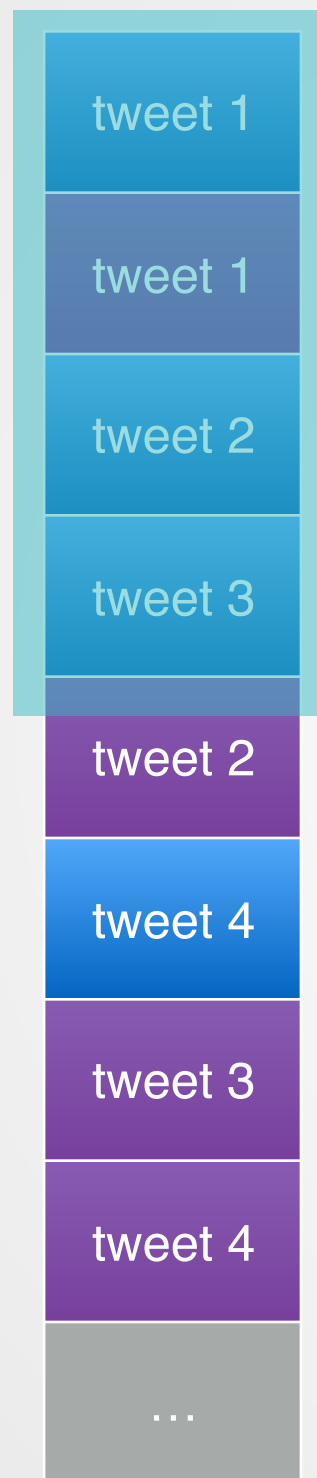
tweet 4

...

# Evaluation



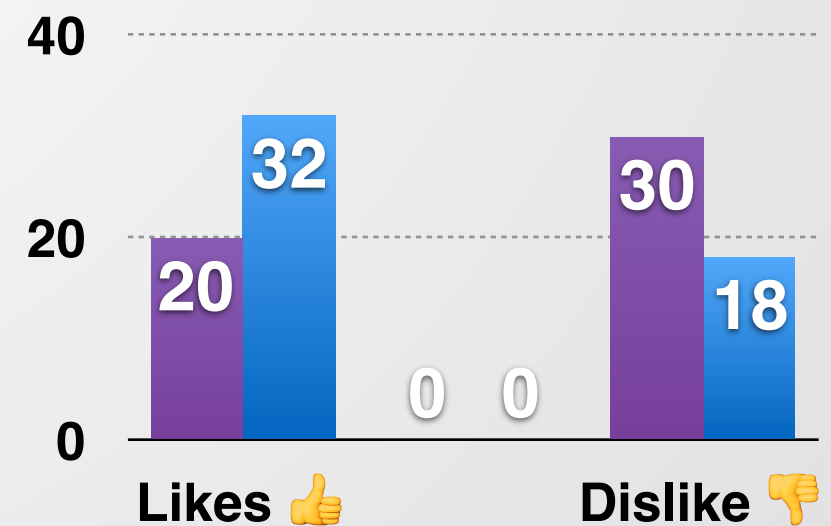
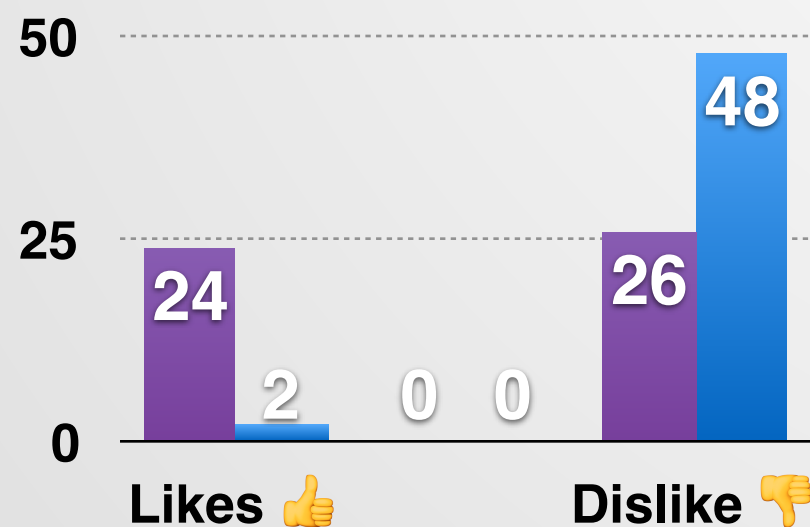
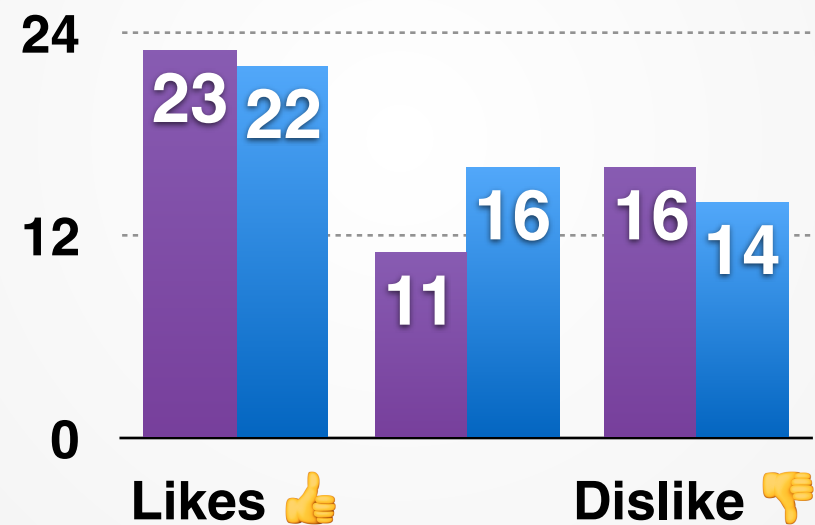
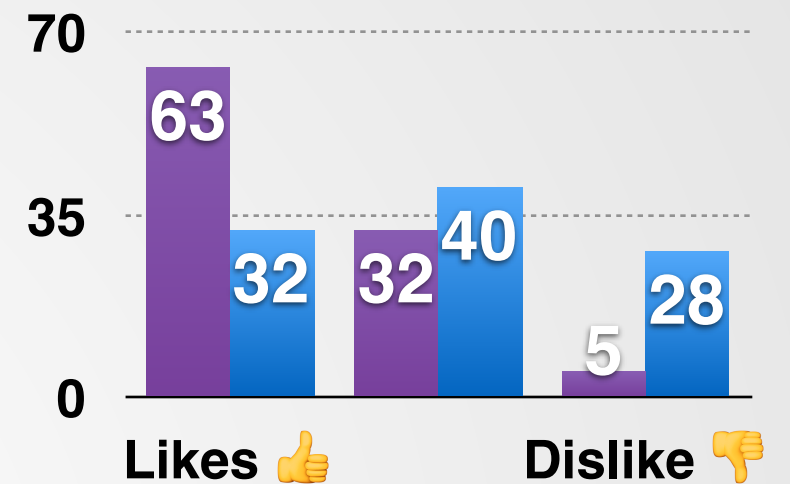
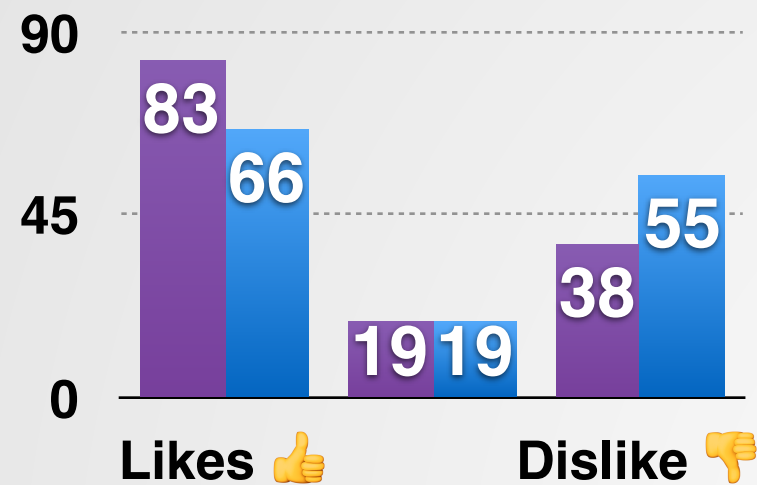
# Evaluation



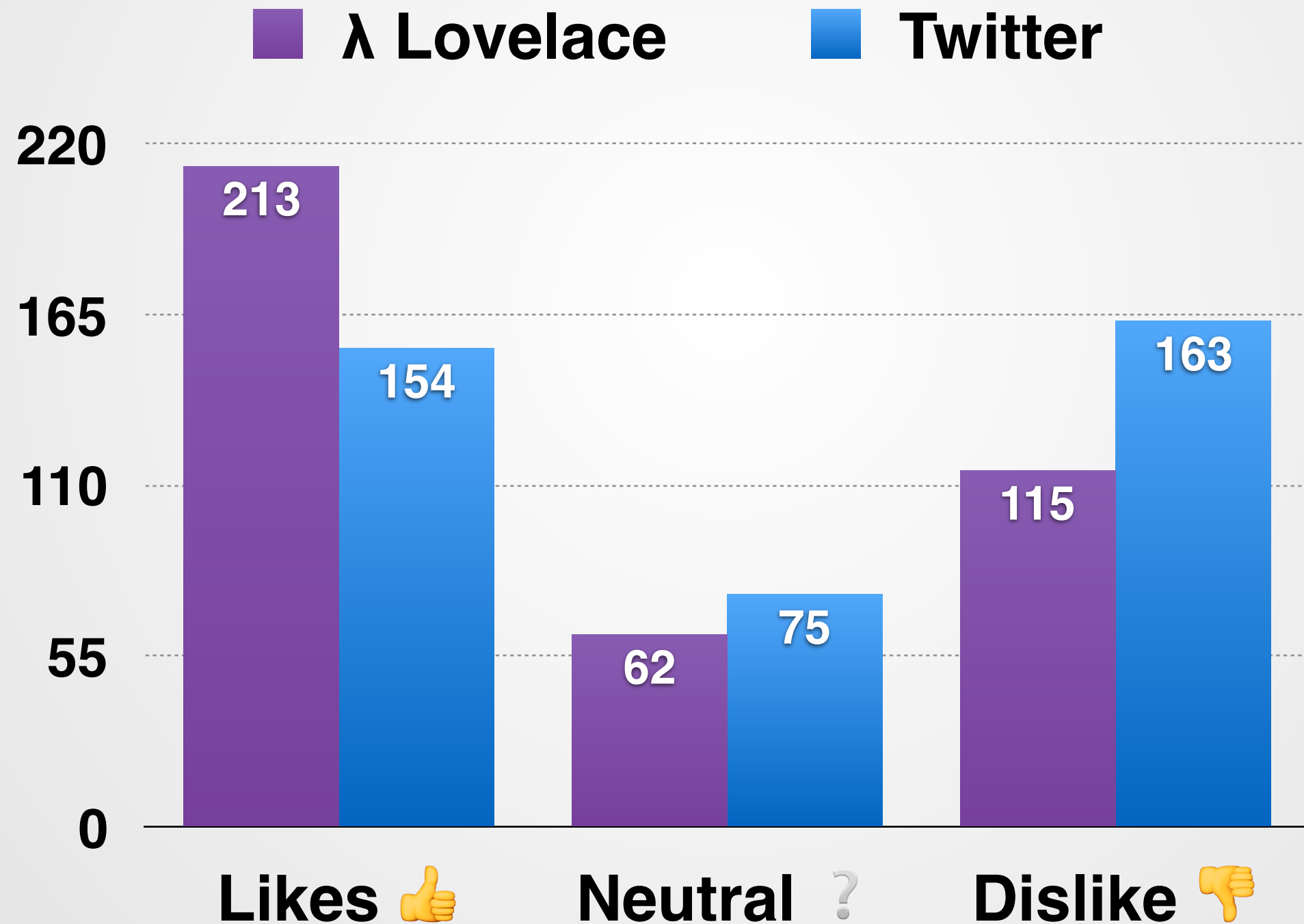
# Evaluation



# Evaluation



# Evaluation



# The System Review

Quality of system:

- Reflected by quality of evaluation results

Issues:

- Operating System compatibility
- Unfollowed tweets

Future:

- Unfollowed tweets
- iOS client basic functionality

# Conclusion: As a Team

- Facebook Messenger
- Divide and conquer approach to work
- 10:00 to 18:00, Monday-to-Friday



# Conclusion: Lessons Learned

- Twitter plays dirty 💰
- More time learning domain knowledge as a team 📖
- Docker is great, but difficult to implement at the moment 🙄
- A website would have been better for the evaluation experiment 🕸
- Rollbar is excellent for error detection in production 🌩

# Conclusion: What next?

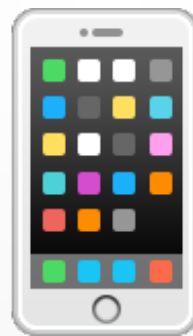
## **What we would have done differently back in June:**

- Code reviews
- Test sooner, rather than later
- More time on research
- Use Heroku in place of current setup

## **What we would do if this project continues:**

- Use Docker on all machines
- Test Driven Development

# Demo!





# Thanks!



*Questions?*



<https://jonrh.github.io/lambda-lovelace/>

**Jón Rúnar Helgason, Xinqi Li, Marc Laffan,  
Junyang Ma, Eazhilarasi Manivannan**

# Backup Demo Recording

*in case of Twitter API rate limit ;)*