

twitter 
UC BERKELEY - CS10



TO START... WHAT'S A TWEET?

FOLLOW ME!



twitter.com/raffi
or send a SMS to 40404
follow raffi



"Rudimentary communication among individuals in real time allows many to move together as one" - @biz



ONE WORD

April 10, 2008
4:33:23 PM

ONE WORD
 April 10, 2008
 4:33:23 PM

AND IT'S OVER
 April 11, 2008
 1:27:26 PM

HAITI
 January 2010

CHILE
 February 2010

MTV VMAs
 September 2010

EMMYS
 August 2010

The humble beginnings

HOW DOES TWITTER WORK?

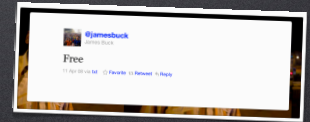
TWITTER IS JUST A "STRUCT", A GRAPH, & A LIST

A TWEET IS A SIMPLE THING



140 characters

WHAT'S THE DATA STRUCTURE OF A TWEET?



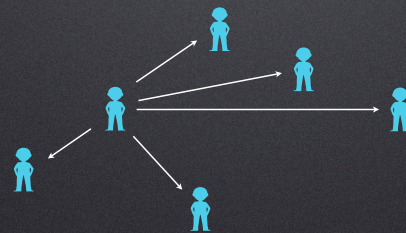
```
{  
  "id": 787167620,  
  "text": "Free",  
  "date": "2008-04-11 13:27:26",  
  "user": "jamesbuck"  
}
```

*not quite, but something like this

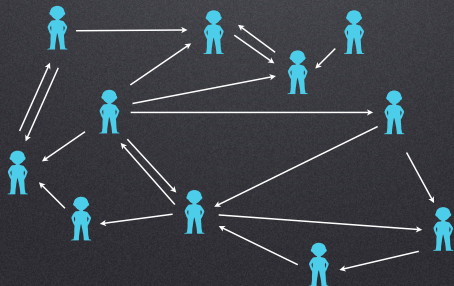
WHERE DO THEY GO?



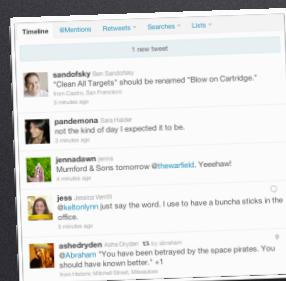
TWEETS MULTIPLY!



IT'S A HUGE GRAPH



VIEWING TWEETS



HOME TIMELINE



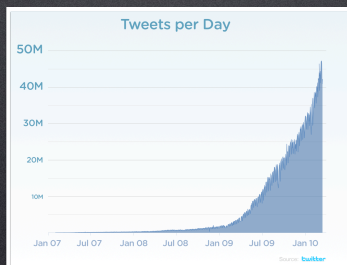
[27886578700,
27856404785,
27727644786,
27679412331,
27594550593,
27583976236,
27580246968,
27576503025,
27545926564,
27545450496,
27539864336...

IS THAT IT? ... YES! ... SORTA ...

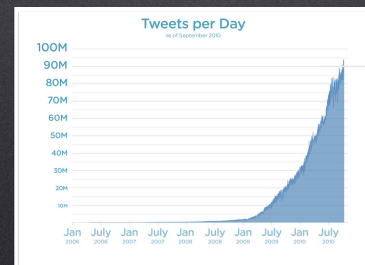
THINGS GET MORE COMPLICATED AT SCALE

AND SCALING REAL-TIME IS HARD

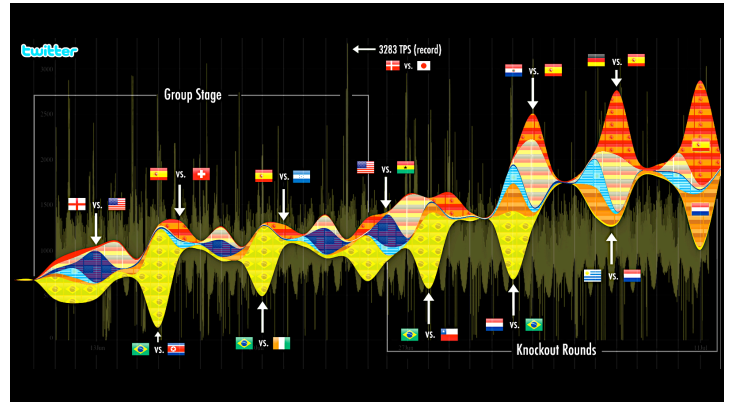
HOW MANY TWEETS ARE THERE?



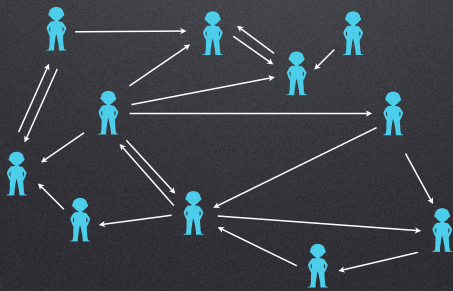
HOW MANY TWEETS ARE THERE?



90M tweets per day \approx 1000 tweets per second



LOTS OF DELIVERIES



TWEETS MULTIPLY



5,162 people need to see my tweet



@ladygaga

mother mons^{ter}
6.8 million followers



@justinbieber

Justin Bieber
5.7 million followers



@BarakObama

44th President of the United States
5.6 million followers



@raffi

me!
5.1 thousand followers



@ladygaga

mother mons^{ter}
 6.8×10^6 followers



@raffi

me!
 5.1×10^3 followers

differs by 10^3



@ladygaga

mother mons^{ter}
6.8 x 10⁶ followers

scaling 10⁶

10⁶ deliveries per minute = 6000 deliveries per millisecond

A SYSTEM IS ONLY AS STRONG AS ITS WEAKEST LINK



80 MB per sec
10 ms per seek

10 ms per seek = 100 seeks per second

MySQL

Can't generate IDs fast enough
Centralized and a single point of failure



snowflake

Highly available and uncoordinated (10kqps)
Compatible with the **twitter** ecosystem
<http://github.com/twitter/snowflake>

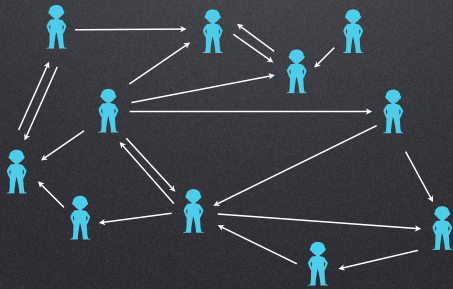
HOW BIG ARE THEY?

1 tweet text = 140 characters
≈ 200 bytes

1000 tweets per second ≈ 195 KB/sec
≈ 11 MB/min
≈ 16 GB/day

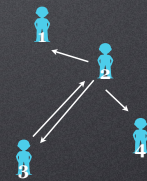
Just tweet text!

LOTS OF DELIVERIES



Digraph

Need to represent this



Matrix

Naïve implementation is not scalable

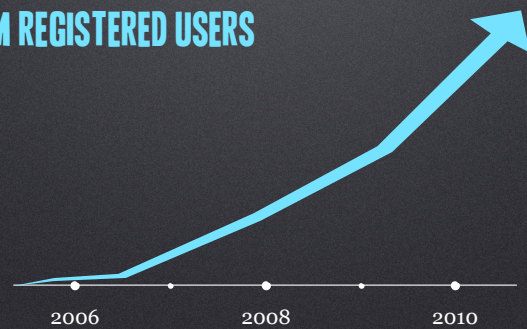
	1	2	3	4
1		✓		
2			✓	
3		✓		
4		✓		



flockdb

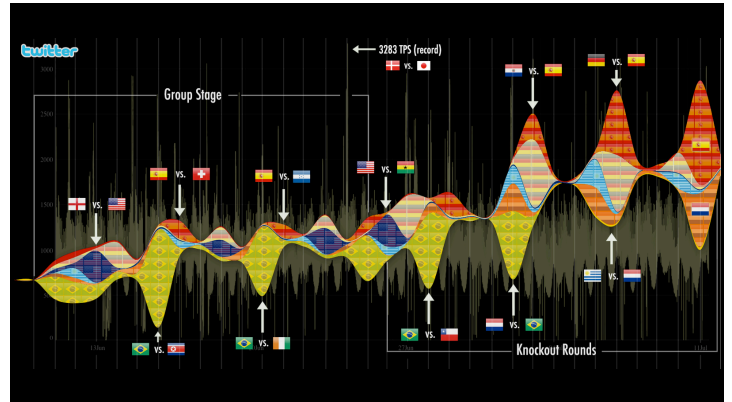
Distributed graph database
High rate of CRUD operations
Complex set arithmetic queries
<http://github.com/twitter/flockdb>

160M REGISTERED USERS



USING THE SYSTEM

6B API calls per day \approx 70,000 calls per second



WHERE DO WE WANT TO BE?

Today - 160M people generate ~1000 TPS

Tomorrow - we want to support half the world and all its devices

(right now, there are 6B people and 5B phones)

QUESTIONS?

Follow me at
twitter.com/raffi

twitter 