Rettiwt

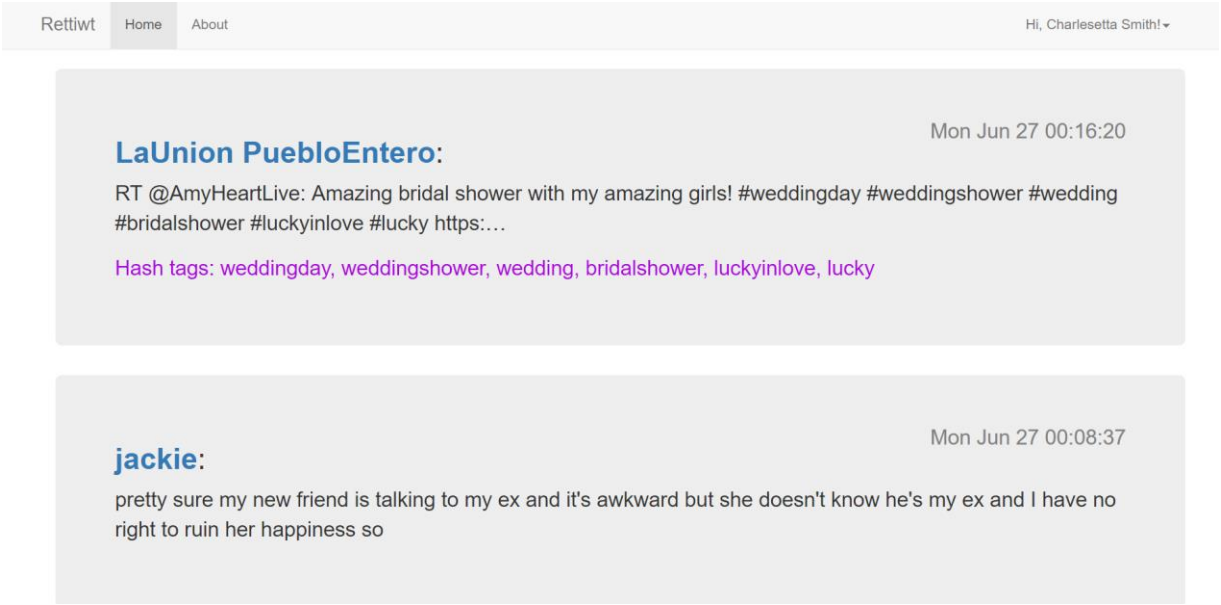
Rebuilding the Twitter Platform

Daoyu Tu
Insight Data Engineering, New York
Summer 2016

Why reinvent the wheel?

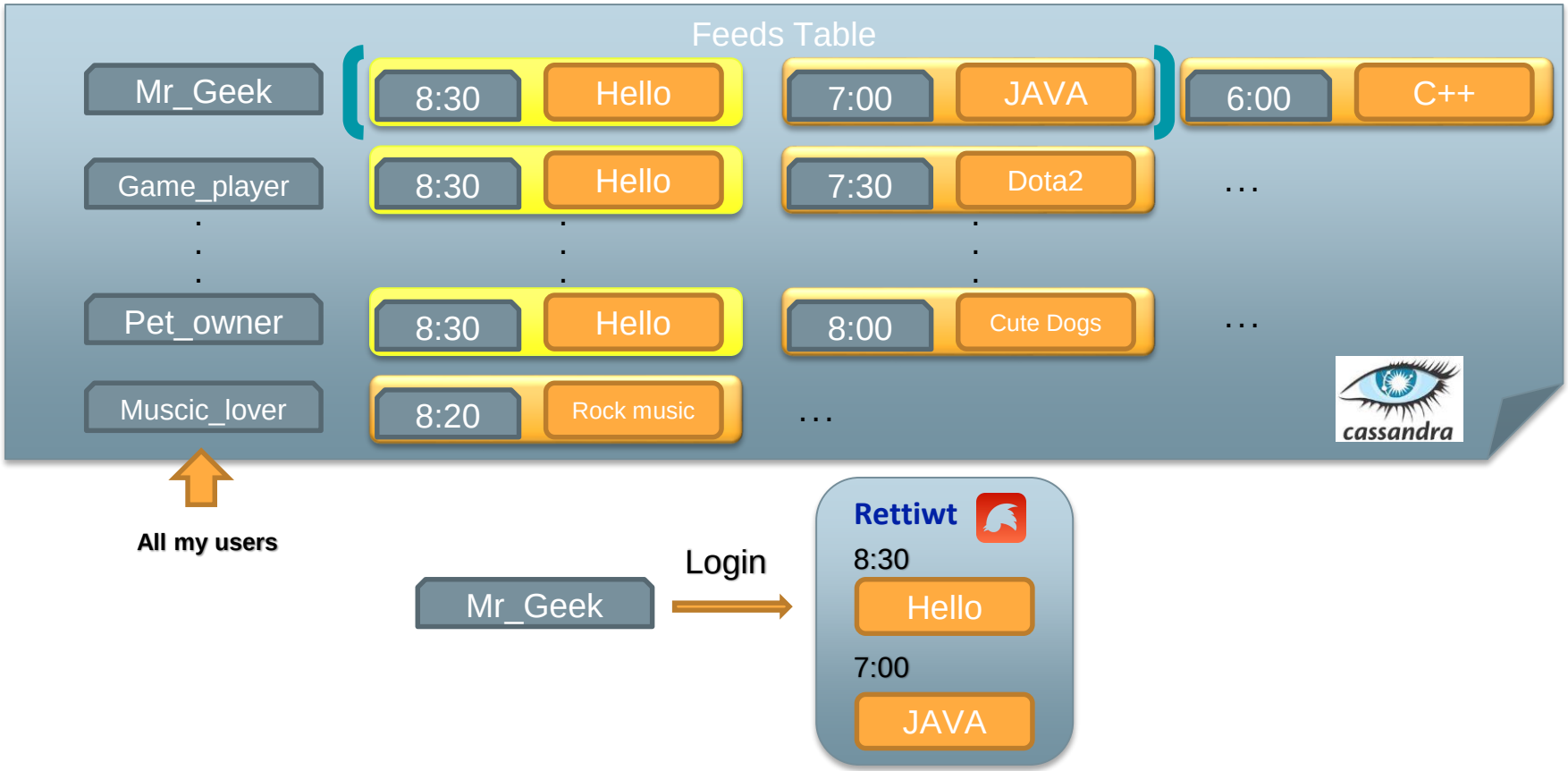
- Building a scalable & robust social media system
- Delivering real-time data to all the users

Demo


Rettiwt
Home
About
Hi, Charlesetta Smith!
Mon Jun 27 00:16:20
LaUnion PuebloEntero:
RT @AmyHeartLive: Amazing bridal shower with my amazing girls! #weddingday #weddingshower #wedding #bridalshower #luckyinlove #lucky https:…
Hash tags: weddingday, weddingshower, wedding, bridalshower, luckyinlove, lucky
Mon Jun 27 00:08:37
jackie:
pretty sure my new friend is talking to my ex and it's awkward but she doesn't know he's my ex and I have no right to ruin her happiness so

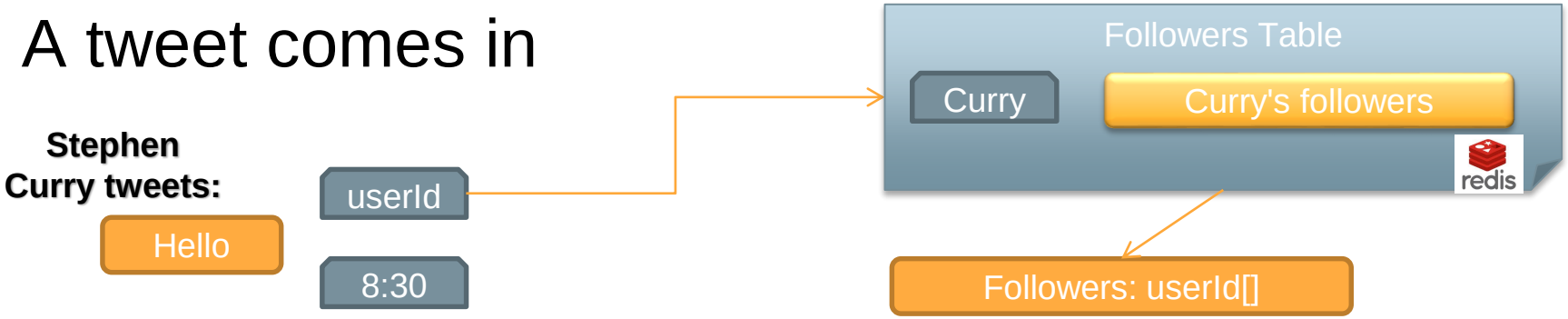

daoyu.online
10
Rettiwt
Please sign in ;^)
User id(0 ~ 99999)
Password
Remember me
Sign in

Feeds Table

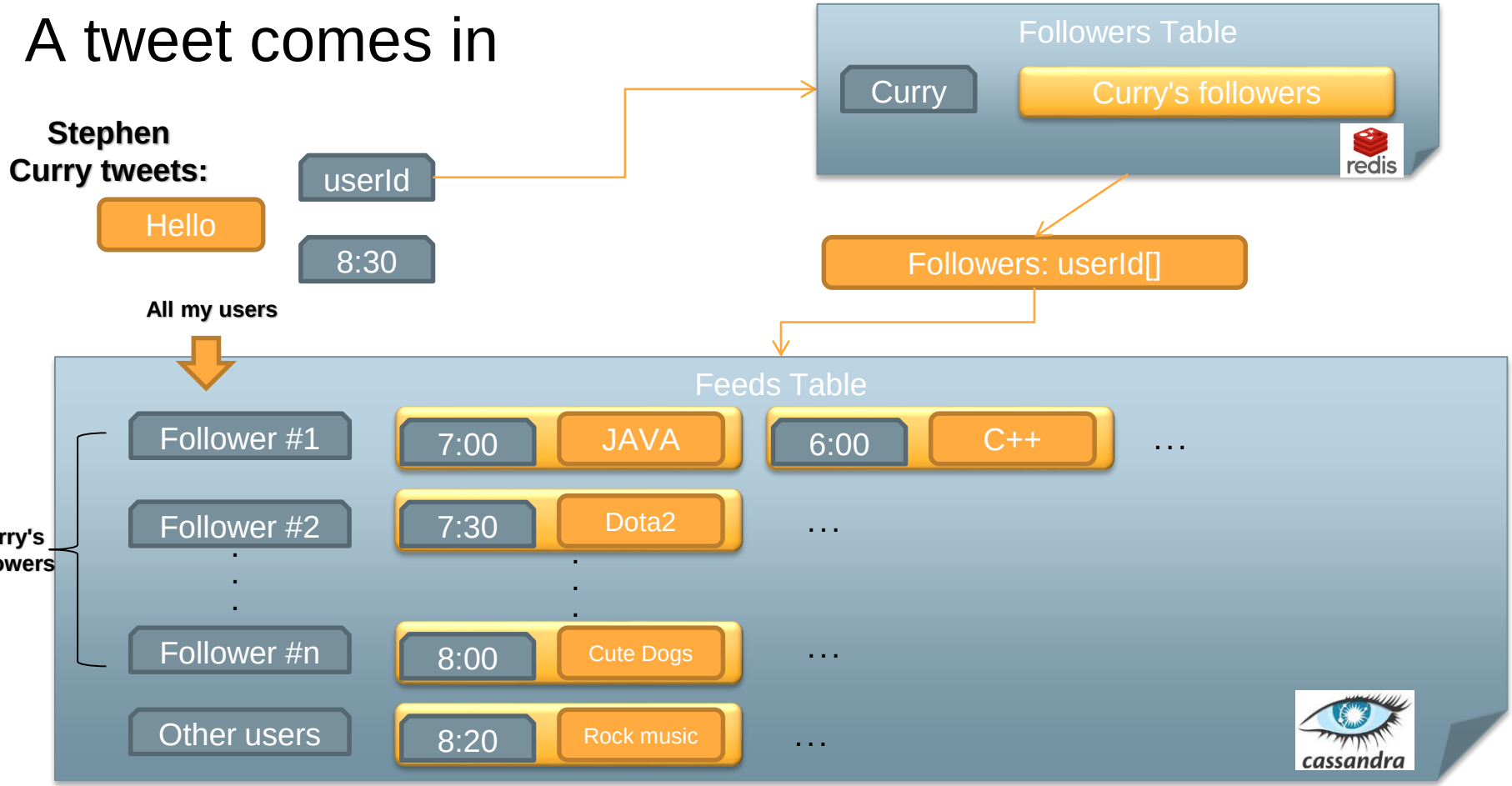

Feeds Table
Mr_Geek
8:30
Hello
7:00
JAVA
6:00
C++
Game_player
8:30
Hello
7:30
Dota2
...
Pet_owner
8:30
Hello
8:00
Cute Dogs
...
Muscic_lover
8:20
Rock music
...
cassandra
All my users
Mr_Geek
Login
Rettiwt
8:30
Hello
7:00
JAVA

A tweet comes in

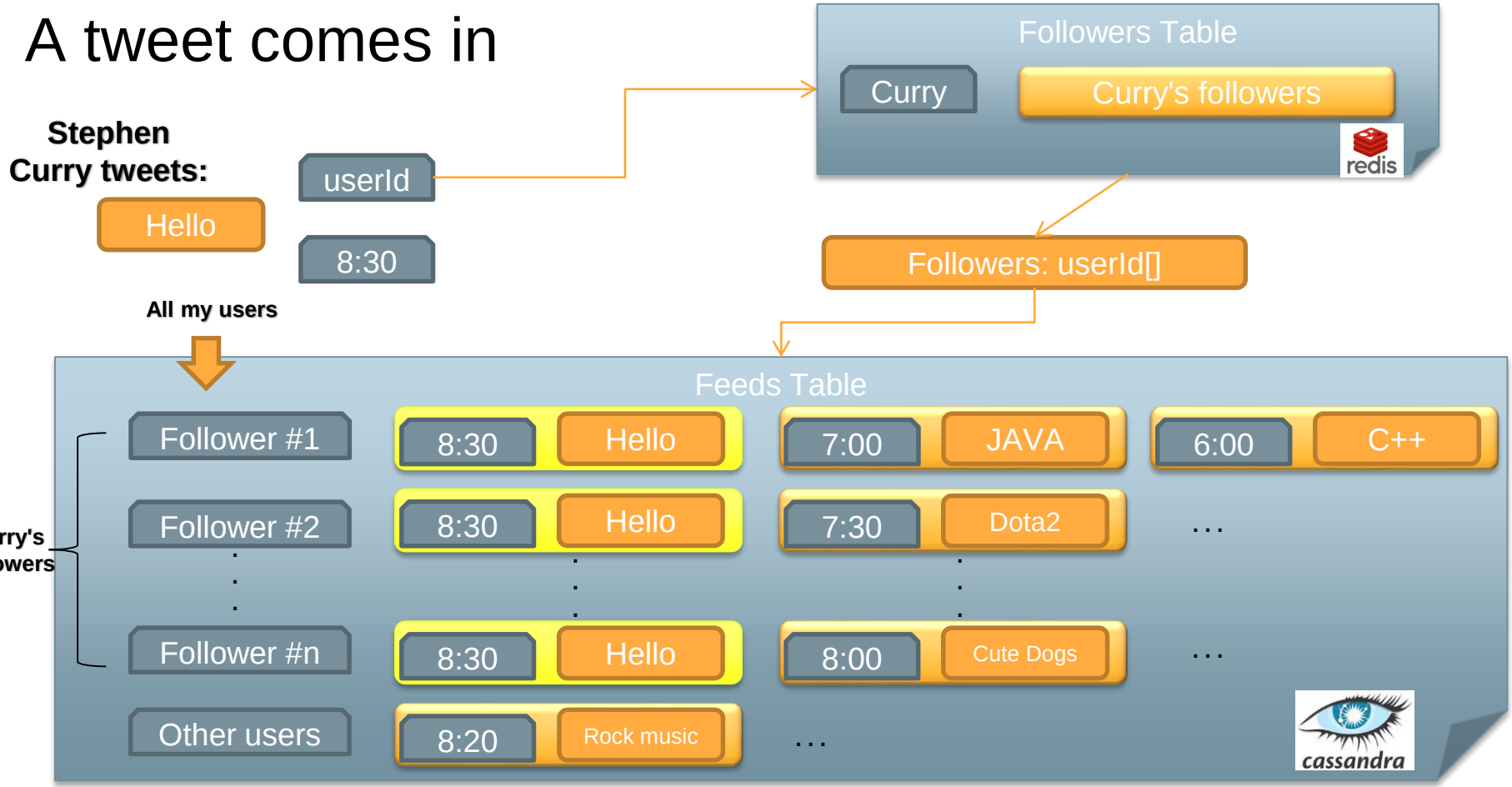
A tweet comes in



A tweet comes in



A tweet comes in



Pipeline


Spark
Streaming
kafka
Flask
web development,
one drop at a time
redis
cassandra
Twitter
Rettiwt

Benchmarking

- Save the tweet or tweet Id?
- Redis or Cassandra?

Denormalized Schema

Feeds

- **uid: Long**
- tweet: String
- time: timestamp

VS

Normalized Schema

Feeds

- **uid: Long**
- tid: Long
- time: timestamp

Tweets

- **tid: Long**
- tweet: String

Results	Norm	Denorm
Writing	1974 recs/s	691 recs/s
Reading	14ms for 20 tweets	10ms for 20 tweets

Benchmarking

- Save the tweet or tweet Id?
- Redis or Cassandra?

Redis	Cassandra
• Fast in-memory reads • Pure key-value storage	• Fault tolerance • More storage

Results	Redis	Cassandra
Tweets	24.4 tweets/s	21.9 tweets/s

About Me

- Daoyu Tu
- Computer Science B.S & M.S.
- Web Developer

Some Stats

- 2 clusters of 4 m4.xlarge nodes on AWS, one for Cassandra, one for Kafka, Spark & Redis
- 100,000 users with average degree of 94
- Writing 1974 records, processing 24.4 tweets per second